

# A Century of General Relativity

**Malcolm A.H. MacCallum** FIMA  
Queen Mary University of London

## 1 The birth of a theory

In 1900 physics was in a quandary. Well-tested Newtonian physics was incompatible with Maxwell's laws for electromagnetism, which implied a fixed speed of light,  $c$ . In Newtonian kinematics measured speed was relative to the observer's motion. It was thus natural to assume, based on knowledge of other waves, that light waves had to be supported by a physical medium, the 'aether', at rest in Newton's fixed space, and that the derivation of  $c$  applied only in a frame at rest in the aether.

Stellar aberration, discovered by Bradley in the 1720s, showed that the earth's surface had to be moving through any aether. The 1880s Michelson-Morley experiments gave the contradictory result that it did not, creating the quandary mentioned. This could be resolved, rather unsatisfactorily, by assuming that measuring devices ('rods and clocks') in motion altered their size or clock rate in the specific way given by the Lorentz transformations. Then light always appeared to have the speed  $c$ .

Artist's conception of a black hole.  
© Henning Dalhoff | Science Photo Library



The need for those effects on rods and clocks was removed by Einstein's special relativity (SR). SR and the related deduction of the famous  $E = mc^2$  were part of Einstein's 1905 *annus mirabilis*, along with key papers on the photoelectric effect and on Brownian motion. SR kept the Lorentz transformation formulae but did away with the aether. In its place came Minkowski space (see equation (3) of Chris Linton's accompanying article on pages 229–233). In the new picture Newton's laws of mechanics were modified only by taking rates of change with respect to the proper time of an object (the time along its worldline) rather than some universal time.

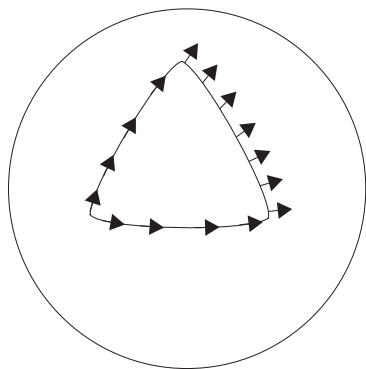
Unfortunately SR's success gave a new problem. Newtonian gravity acts instantaneously at a distance, but in SR there is no invariant concept of simultaneity. As Linton describes, Einstein set out in 1907 to find a gravity theory that was compatible with both special relativity and Newton's theory, and incorporated the equivalence principle, gravitational redshift, light-bending and an explanation for Mercury's perihelion shift. Max Planck, 'As an older friend', advised Einstein against this pursuit, 'for, in the first place you will not succeed, and even if you succeed, no one will believe you.'

The steps that followed, and how the objectives were met, are set out in Linton's article: more details can be found in many places e.g. [1]. Linton also gives a brief description of the key parts of the theory itself, the Einstein field equations, EFE, (his (7)) and the geodesic equation (his (4)).

To understand why Riemannian curvature is needed, one can consider dropping two objects from a height but at relative rest, as in Galileo's legendary experiment. In SR these are moving on parallel lines in spacetime. But since both will fall towards the centre of the Earth, they will move towards one another and, were the Earth not in the way, would eventually meet.

Initially parallel lines that meet are familiar on the Earth's surface. Consider travelling due North from separated points on the equator: such tracks meet at the pole. Using our three-dimensional understanding, we would say this is because the Earth's surface is curved and not flat. To be able to say the same in spacetime without going into a fifth or higher dimension, we need an intrinsic definition of curvature, one which does not depend on being able to 'see' the curved spacetime from outside.

This can be obtained by considering transporting a vector parallel to itself round a closed curve. Doing this on the Earth's surface, we would find on arrival back at the starting point that the vector had turned through an angle proportional to the area of the curve, as shown in the example in Figure 1.



**Figure 1: Parallel transport of a vector round a right-angled triangle on a sphere, starting at bottom left. The vector comes back at 90° to its original position.**

Carrying out such calculations in a general manifold leads to the formula for Riemannian curvature in Linton's article, parallelism being defined by the connection (so-called because it relates parallel vectors at neighbouring points).

General relativity agrees with SR in neighbourhoods small compared with the square root of the magnitude of curvature components. Hence there cannot be a purely local definition of gravitational energy, since no such energy appears in SR. One can still define global energy integrals for isolated bodies.

GR also agrees with Newtonian gravity theory in a weak-field limit. Here 'weak' means comparing the actual size of a body with the size it would have were it to form a black hole of the same mass. Such ratios are typically small for real objects in the Universe, which is why Newton's theory worked so well. For example, for the Sun the ratio is about 1 to 500,000.

General relativity was developed against the backdrop of the First World War, whose centenary, though hardly a cause for celebration, is also being observed currently. The war affected GR's dissemination. Eddington's report introducing GR to the English-speaking world relied on information from de Sitter in neutral Holland. Inevitably, the theory's adherents were caught up in the conflict, most notably Karl Schwarzschild, the discoverer of the eponymous solution of Einstein's equations, equation (8) of Linton's article, now recognised as describing a spherically symmetric 'black hole' (a term attributed to a journalist, Ann Ewing, writing in 1964). Schwarzschild died in 1916 while serving on the Russian front.

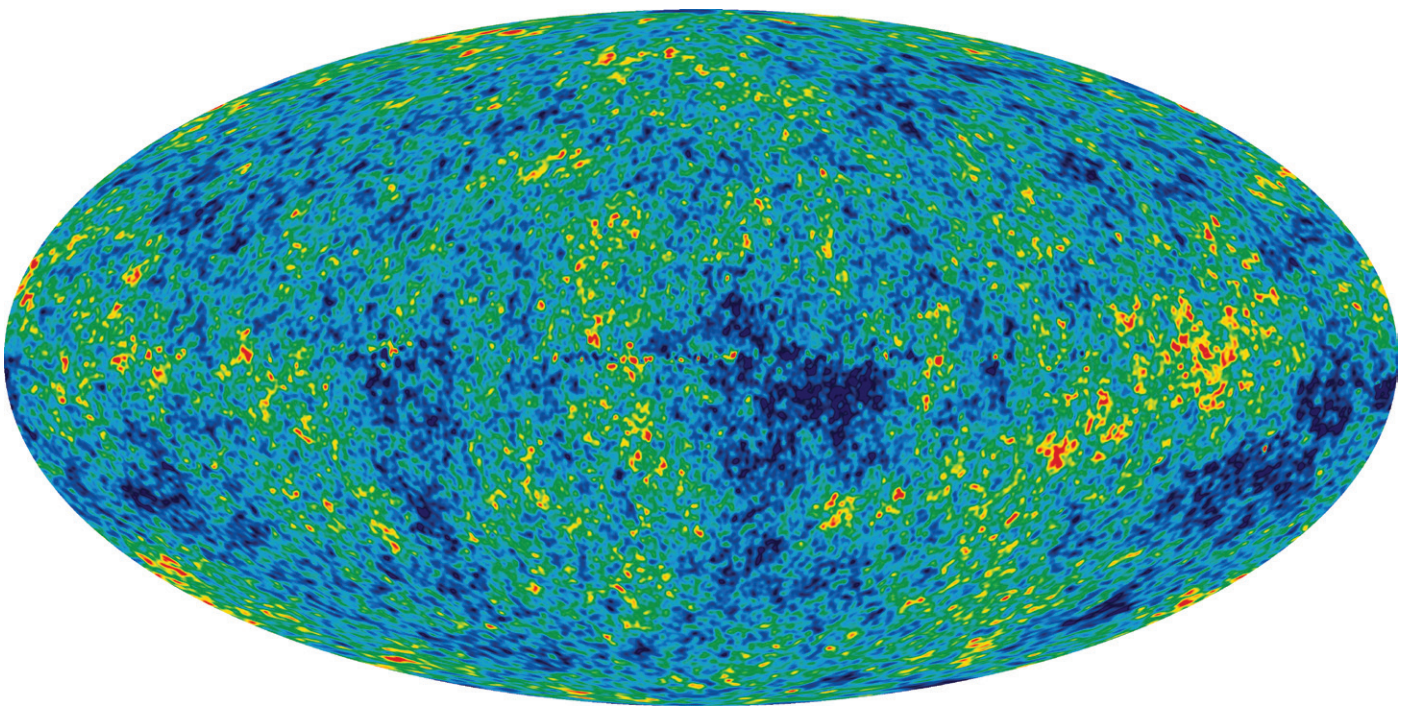
GR resolved the theoretical incompatibility of SR and gravity, and explained, as Linton's article describes, the long-standing mystery of the anomalous precession of the perihelion of Mercury's orbit. After the 1919 confirmation of light-bending, Einstein came to be seen as the personification of scientific genius (perhaps belatedly in the light of the importance of his 1905 papers!). It was not until the 1920s that the first observations of gravitational redshift were made.

For over 25 centuries, spacetime had been a stage on which the dynamics of matter unfolded. With GR, the stage suddenly joined the troupe of actors. As decades passed, new aspects of this revolutionary paradigm have continued to emerge. There are so many fascinating developments in both theory and applications that selection for this article has been very difficult. The theory remains the most nonlinear of the generally-accepted theories of physics, and is still prompting development in analytic and numerical techniques.

As well as the 1990 Fields Medal for Ed Witten, in part for his 'proof in 1981 of the positive energy theorem in general relativity', a series of Physics Nobel Prizes have related to GR and its applications; the prize of 1974 to Ryle and Hewish for pulsars, of 1978 to Penzias and Wilson and 2006 to Mather and Smoot for work on the cosmic microwave background (CMB – see Figure 2), of 1993 to Hulse and Taylor for their observations of a binary pulsar, and of 2011 to Perlmutter, Riess and Schmidt for the supernova observations suggesting the universe's expansion is accelerating.

## 2 GR's 100-year growth

GR is now applied to Solar system motions, the astrophysics of compact bodies with strong gravitational fields, and cosmology. Its early years saw slow growth of the theory and these applications, but key steps in each were taken. Mercury's orbit,



**Figure 2: Nine Year Microwave Sky (see Section 4). The detailed, all-sky picture of the infant universe created from nine years of WMAP data. The image reveals 13–14 billion year old temperature fluctuations (shown as colour differences) that correspond to the seeds that grew to become the galaxies. The signal from the our Galaxy was subtracted using the multi-frequency data. This image shows a temperature range of  $\pm 200$  microKelvin. Credit: NASA | WMAP Science Team.**

light-bending, gravitational redshift, and the Schwarzschild solution, have already been mentioned. Gravitational waves were derived in the 1920s, but whether they were real or just coordinate effects remained unclear.

In cosmology, the most important models are the spatially homogeneous and isotropic cosmological ones (the Friedmann-Lemaître-Robertson-Walker or FLRW models). These can be written, in ‘geometrised units’ in which  $c = 1$ , as

$$ds^2 = a^2(t)[dr^2 + \Sigma^2(r, K)(d\vartheta^2 + \sin^2 \vartheta d\varphi^2)] - dt^2, \quad (1)$$

where the form of  $\Sigma(r, K)$  depends on the constant curvature  $K$ . Einstein and de Sitter had found the static cases very early on. By 1927 Friedman and Lemaître had found the first of the expanding cases now taken as the standard models of the Universe. In these,  $a(t)$  and  $\mu(t)$  obey

$$3\dot{a}^2 = \kappa\mu a^2 + \Lambda a^2 - 3K, \quad (2)$$

$$\dot{\mu} + 3(\mu + p)\dot{a}/a = 0, \quad (3)$$

where the dot denotes a derivative with respect to time,  $\mu$  and  $p$  are the energy density and pressure of the Universe’s matter content, and  $\Lambda$  is the cosmological constant, which Einstein had introduced in 1917 in order to obtain a static solution to (2). Defining the density parameter  $\Omega_m := \kappa\mu/3H^2$  where  $H = \dot{a}/a$ , one finds that if  $\Lambda = 0$ , then  $\Omega_m = 1$  is the boundary between ever-expanding models ( $K \leq 0$ ) and models that contract again ( $K > 0$ ).

After Hubble had discovered a linear distance-velocity relation for galaxies in 1929, the expanding FLRW universes, which start with a ‘Big Bang’ from a singular origin at  $t = 0$ , became more widely accepted, although the value of  $H$  he found, due to an erroneous distance scale, gave discordant results, as described below.

General relativistic dynamics and fluid mechanics, electrodynamics and thermodynamics were developed from 1915 onwards, and their foundations were well laid by World War II. Oppenheimer and Snyder gave the first simple models of gravitational collapse in 1939, study of FLRW perturbations began

with Lifshitz’s 1946 paper, and Gödel found his eponymous solution in 1949. And a number of still significant other solutions of, and theorems about, Einstein’s equations were discovered before 1950.

However, it was only from about 1950 onwards that GR blossomed to its present extent. The reasons for the hiatus in the 1930s and 40s include the political situations in Germany and the Soviet Union, the great attraction of quantum mechanics for young physicists, and the apparently esoteric nature of the theory. In particular the singular origin of expanding universes, and the singularities in black holes, seemed to be merely pieces of abstract mathematics. The 1949 discovery of the Gödel solution, rotating but invariant in time and containing closed timelike lines, reinforced that perception. Einstein himself was focused, unsuccessfully, on creating a unified theory for physics.

Renewed interest arose from improvements in mathematical techniques and understanding, and from technological innovations. Improved optical telescopes and access to other frequency ranges (microwave, infrared, X-ray and  $\gamma$ -ray) have led to us having vastly more evidence in cosmology and astrophysics, revealing objects and processes for whose explanation GR is essential. Spacecraft, the burgeoning of electronics since the invention of the transistor in 1947 (e.g. computers, CCDs, and lasers), and atomic clocks, have also played parts in making GR the important and accepted theory it now is.

Exquisite precision is now achieved in tests and applications of GR through modern devices, but the interaction with technology, especially metrology, has not just been one-way. The satellite experiment ‘Gravity Probe B’, which tested predictions of effects of GR on orbits round the rotating Earth, and the construction of interferometric gravitational wave detectors, provide examples. Moreover, accurate navigation using the GPS system, and accurate planning of satellite trajectories, would be impossible without taking relativistic corrections into account. In what follows I shall first sketch some of the mathematical developments and then briefly outline various applications, starting at the cosmological scale. Readers wishing to know more can find an excellent non-technical account in [2] and a more technical survey in [3].



### 3 Mathematical developments

The first mathematical innovations of the 1950s were Taub's use of symmetry groups generated by Killing vector fields, Choquet-Bruhat's proof that the Einstein equations can be treated as a well-posed Cauchy (initial value) problem and Petrov's algebraic classification of the Weyl tensor, the part of the Riemannian curvature not fixed locally by the EFE. All three still resonate today.

Earlier solutions, like the Schwarzschild and FLRW metrics, have symmetries but Taub's method for finding such solutions has led to many more. These have played important roles in developing our understanding of the whole space of solutions of the EFE and the solutions' possible pathologies, for instance the famous Taub-NUT solution later characterised by Charles Misner as a 'counter-example to almost everything'.

Choquet-Bruhat's proof showed that the initial configuration and motion determine the future evolution in GR as in other classical physical theories, and it paved the way for numerical relativity. The approach splits the problem of understanding and solving the EFE into two parts, an initial data problem of finding values on a spacelike surface satisfying the elliptic constraint equations, four of the ten EFE, and the evolution problem of using the remaining hyperbolic equations to obtain the geometry to the past and future of the initial surface.

Petrov's classification provided the main alternative [4] to the symmetry route to solutions, the consideration of 'algebraically special' spacetimes, and prompted the 1960s work on the geometry of 'congruences' of lightlike curves in spacetime, which played a key role in understanding gravitational radiation and spacetimes' global structures.

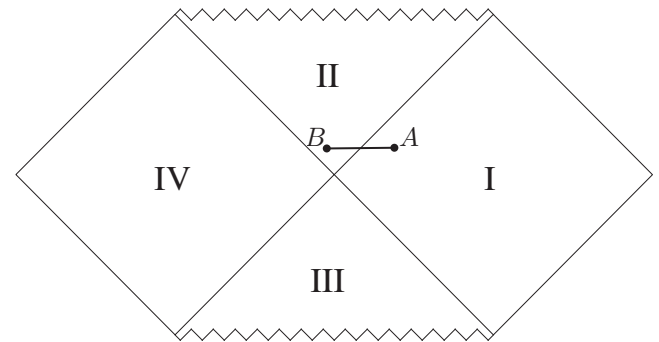
That GR did predict gravitational radiation was clarified from the late 1950s in exact and approximate solutions by Bondi and co-workers, which showed that energy could be transported and transmitted by gravitational waves.

Further remarkable developments of the 1960s and early 1970s included Roy Kerr's rotating black hole solution, the singularity theorems of Hawking and Penrose and the understanding of global structures of spacetime, all of these linked with the mathematics and physics of black holes.

Black holes are regions from which light, and any other known form of matter, cannot escape. Stellar structure calculations imply that the endpoint of stellar evolution must be a white dwarf (maximum mass  $1.44M_{\odot}$  where  $M_{\odot}$  denotes the mass of the Sun), a neutron star (maximum mass below  $5M_{\odot}$ ) or a black hole. Since many stars with masses above  $5M_{\odot}$  are known, black holes are predicted. The surface of a black hole forms an 'event horizon', a boundary to the set of events visible to an external observer. In the Schwarzschild solution (equation (8) in Linton's article) the horizon is at  $r = \alpha$ , and was referred to as the Schwarzschild singularity because the coordinates of (8) become singular there.

The horizon's true nature emerged after Finkelstein's publication in 1958 of coordinates which followed falling particles smoothly across the horizon. It was comprehensively clarified by the 1960 papers of Kruskal and of Szekeres which gave the Schwarzschild solution's complete analytic extension. Global properties of this and other solutions are typically depicted in the conformally transformed form of 'Penrose diagrams' in which infinity is mapped to points at finite distances and light rays travel along lines at  $45^\circ$ . Figure 3 shows such a diagram for the Kruskal metric: region I here is  $r > \alpha$ , exterior to the black hole, and

region II is the black hole interior. Regions III and IV are respectively a white hole, where matter can get out but not in (the opposite of a black hole), and a second exterior region. The jagged lines represent the singularities. Regions III and IV are not expected to occur in nature because matter collapsing to form the black hole would replace the depicted vacuum in an area starting at the bottom of region I and falling into region II. The figure only depicts two of the four dimensions of spacetime, and each point represents a sphere. Drawing a circle of this sphere for each point of the line  $AB$  leads to the picture on the title page of this article.



**Figure 3: The Penrose diagram of the maximal analytic extension of the Schwarzschild solution.**

The uniqueness of the known solutions led to the result that black holes 'have no hair' (i.e. are completely characterised, in the absence of exotic types of additional field, by their mass, charge and spin). Collapsing stars must radiate away their higher moments as gravitational waves. That led to the 'Cosmic Censorship' conjectures, that the singularities that GR predicts in collapsing objects would always be hidden inside an event horizon: rather special counterexamples exist, and generic theorems so far fall short of proving the general case.

The 'laws of black hole mechanics' were found to be of the same form as those of thermodynamics, and Bekenstein, considering information loss by infall, gave reasons for the identification of the surface area of a black hole with entropy, and of temperature with the inverse of mass (with appropriate constant factors). In 1974 Stephen Hawking proved, by considering GR and quantum theory together, that this correspondence is not just formal. Black holes radiate when hotter than their surroundings. Unfortunately, we are unlikely to observe this phenomenon. At the present time it would involve black holes of about twice the mass of the Moon, and none such are expected to have formed except perhaps very early in the Universe's history.

The mathematical work on black holes also spawned 'black hole physics', the study of the effects on nearby matter. By definition, we cannot receive radiation from inside a black hole, but we can (and almost certainly do) see the effects it has on neighbouring stars and gas. In particular we see the effects of accretion of surrounding matter into disks whose inner regions heat up and produce intense radiation, while the disks can funnel jets along the perpendicular axis.

The black hole work interacted with work on singularities. Roger Penrose showed that if GR remained valid as the density of matter increased, then the region inside a 'trapped surface' from which light cannot escape (i.e. a black hole), must contain a singularity, implying the singularities had to be taken seriously. Hawking adapted Penrose's ideas to a proof that cosmological singularities exist, and Hawking and Ellis pointed out that its

conditions were met in realistic models of our Universe. The outcome corrected inferences drawn from calculations by Lifshitz and Khalatnikov that only fictitious (coordinate) singularities occurred in cosmological solutions. (Their calculations, revised in collaboration with Belinskii, are themselves still relevant in cosmology.)

The black holes and singularities work, combined with the study of asymptotics of isolated bodies, as summarised in [5], provides the basis of our present understanding of the global structure of spacetime.

#### 4 GR in cosmology

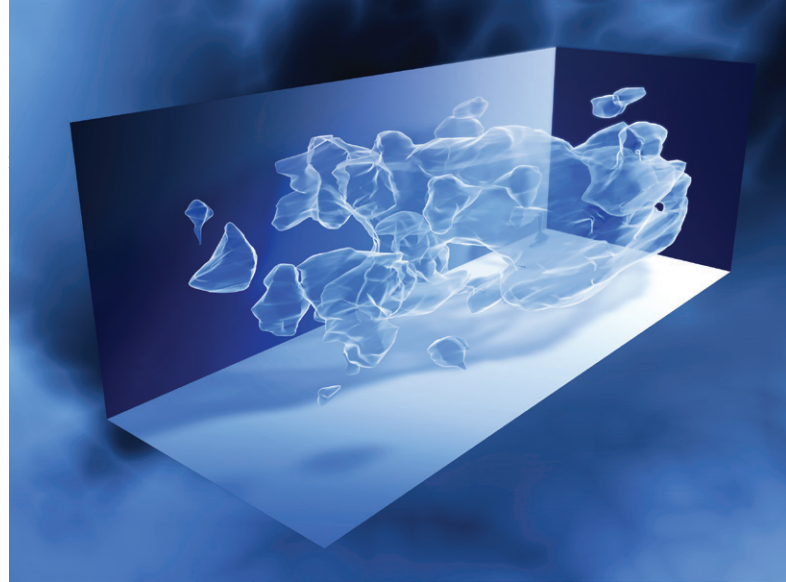
By the 1960s the FLRW models had given an understanding of the formation in the hot big bang of those chemical elements which could not be formed in stars, limits on the number of neutrino types, and a prediction of a cosmic microwave radiation background (CMB), which Alpher and Herman calculated in 1949 would have a temperature of 5K. However, Hubble's erroneous distance scale, which underestimated the Universe's size, age and expansion rate by a factor of about 7, implied that an expanding FLRW Universe must be younger than the geologically-known age of the Earth and that our Galaxy, the Milky Way, was the uniquely largest in the Universe, which encouraged consideration of alternative theories. The error was first corrected in 1952, due to Baade's observations using the Palomar 200" telescope (commissioned in 1950), and by 1958 Sandage was giving essentially the modern values for the distances and expansion rate.

The CMB prediction was revived in the mid-60s, independently by Doreshkevich and Novikov and by Peebles and collaborators, just as the first observations were being made, using a new type of receiver. This evidence and the counts of radio sources, in particular of quasars, tilted the balance against the Steady State theory, in which the Universe was unchanging in space and time but nevertheless expanding, the resulting reduction in matter density being compensated by continuous creation of new matter. By 1968 the radio source counts clearly showed increased source numbers in the recent past, and then a drop at still larger look-back times. The counts are consistent with formation after a hot big bang and a later running out of energy, but not consistent with Steady State.

The dipole in the CMB arising from our motion relative to the rest of the Universe was measured in 1969, but only in 1992 did the first measurements of smaller scale fluctuations (by the COBE satellite) appear. Subsequent satellite and ground-based measurements have given us very precisely the amplitude, polarisation and angular spectrum of fluctuations of temperature on the last scattering surface, the region where photons and gas decouple as the Universe cools after a Big Bang (see Figure 2).

The cosmological models of the 70s fitted most properties of the observed Universe, but they did not give the obviously-required lumpiness. Inflation, the theory in which some field (the 'inflaton') in the early universe caused accelerated expansion and provides large enough seeds for the observed lumps, was independently found by Starobinsky and by Guth in 1980.

From observations of the dynamics of galaxies and clusters of galaxies, made from the 1980s onwards, we knew the Universe had to contain 'dark matter' as well as the luminous stars, galaxies, and radiating gas clouds. Observations led to  $\Omega_m \sim 0.3$ , more than 80% of it non-luminous and cold (meaning composed of particles moving at speeds well below  $c$ ). Moreover, inflation



**Figure 4: The distribution of dark and baryonic matter, mapped via weak gravitational lensing using the COSMOS survey. Credit: NASA, ESA and R. Massey (California Institute of Technology).**

theorists argued, assuming that the inflaton would turn its energy into normal matter after the inflationary period, that  $\Omega_m \approx 1$ , and the CMB measurements agreed with this.

The linearised perturbation theory of FLRW models required improvement, in particular to overcome the issue of 'gauge freedom', the indeterminacy of the map between the real lumpy universe and its smoothed-out model. This was achieved from 1980 onwards. It links the fluctuations from inflation with the observed CMB spectrum, and thence, via nonlinear processes, to the currently observed large scale structures. The evolution of the background FLRW model successfully relates the peak of the CMB spectrum to the peak seen at later times in separations of galaxies, in so-called 'baryon acoustic oscillation' (BAO) measurements. However the detailed calculations of structure only worked if 70–80% of  $\Omega \approx 1$  is really a  $\Lambda$  term (or matter with some similar equation of state). Such a component is called 'dark energy'.

It was therefore good for the theory when measurements of the velocity-distance relation for supernovae, first announced in 1998, unexpectedly showed that the universe's expansion is accelerating. This (from the derivative of equation (2), coupled with the CMB and BAO measurements) gave  $\Omega_m \approx 0.3$ ,  $K \approx 0$  and  $\Omega_\Lambda := \Lambda/3H^2 \approx 0.7$ . The two dark constituents are distinguished by their equations of state. For both, it is GR's models which give faith in their existence, which is why one way to account for them is to adopt a different gravity theory. Perturbed FLRW models with both dark constituents included are now the target of 'precision cosmology', the placing of tight limits on FLRW models' parameters by combining several sources of data.

Two more aspects of modern cosmology are rooted in the early years of GR. The first is gravitational lensing. Light-bending by galaxies was first observed in 1979. Galactic and stellar gravitational lensing are now routine tools for astronomy, used for example to infer the distribution of mass within the lensing galaxies, to study the distribution of dark matter implied by its lensing of distant galaxies (see Figure 4), and to observe individual distant lensed galaxies. Lensing was predicted to affect the CMB's polarisation and angular distribution spectrum, and measurements agree well. On a smaller scale, some of the many exoplanets now known were found by their lensing effects.

The second is that quasars and active galactic nuclei are now considered to contain black holes, bearing out Lynden-Bell's 1969 proposal that accretion disks round very massive black holes

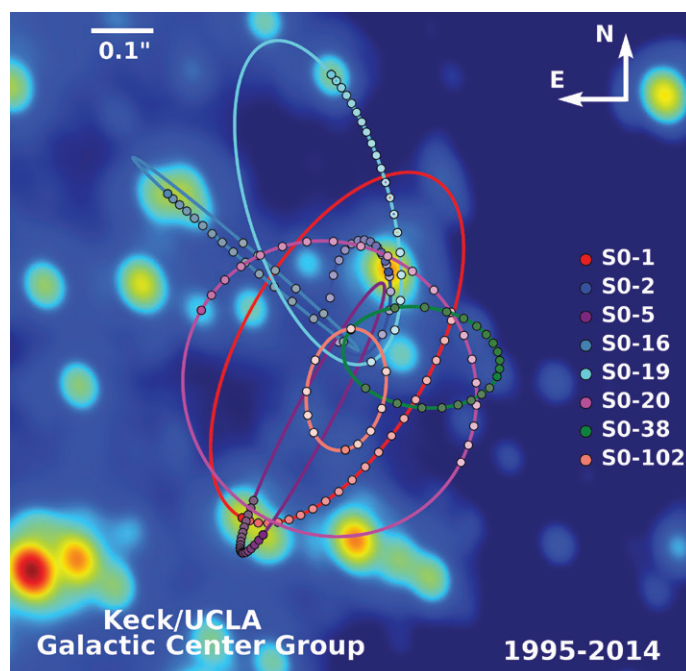


could lead to the jets seen in radio sources. Astrophysicists now believe that virtually all galaxies, or at least all large ones, like the one we live in, contain supermassive black holes at their core. The masses go from a million  $M_{\odot}$  upwards.

The most compelling example is our own Milky Way, the Galaxy. In recent years two groups, those of Reinhard Genzel and Andrea Ghez, have followed the trajectories of stars near the Galactic centre, using the modern technologies of adaptive optics, which employ laser-generated ‘stars’ and rapidly adjustable multi-mirror telescopes. Observations are made in the infrared. Figure 5 shows the observed orbits. One star, SO2, was found to be orbiting the Galactic centre with a period of 15 years, and more recently another, SO-102, was found to have a period of 11.5 years. The stars move at up to 1400 km/s (0.5% of  $c$ ) and their orbits imply a mass  $> 2.6$  million  $M_{\odot}$  inside 0.01 pc (1/30th of a light year).

One surprise from more detailed studies of the black hole masses inferred from such observations is that there is a linear relation between the overall luminosity of a galaxy and the mass of its central black hole. This was first noticed by Magorrian and confirmed in a recent study of 49 nearby galaxies. The analogous relation of black hole mass and the velocity dispersion of stars in galactic bulges seems accurate enough to be used to infer the black hole size from the dispersion. It therefore appears that the central black hole exerts a strong influence on the growth and evolution of a galaxy, even though it has only about 0.5% of the total mass and an even smaller proportion of the physical size (but it has been estimated to have, in Newtonian terms, as much binding energy as the rest of its host galaxy). The mechanisms, and the impact of galactic mergers and collisions, are still being worked out in detail.

Modern cosmology shows clearly how GR has developed into a multi-faceted and accurate modelling tool. It not only supplies the background FLRW solution, whose perturbations match the CMB and BAO fluctuations with high precision, but its lensing and black hole aspects also feature strongly.

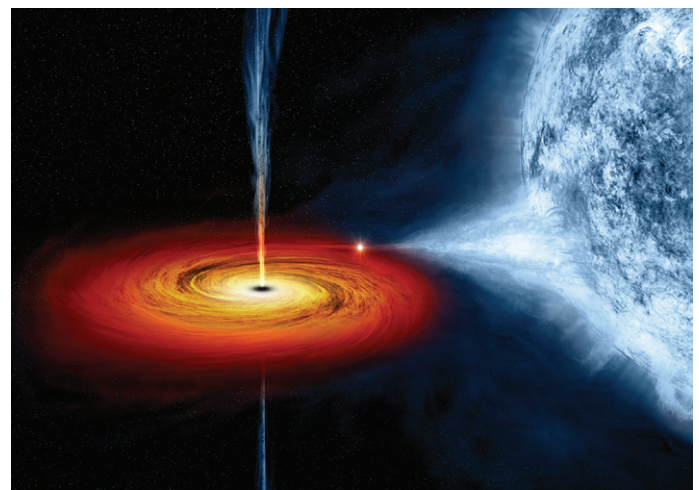


**Figure 5: Orbits of stars near the Galactic centre. This image was created by Prof. Andrea Ghez and her research team at UCLA and is from data sets obtained with the W.M. Keck Telescopes.**

## 5 GR in astrophysics

New types of compact objects such as quasars and pulsars started to be discovered in the 1960s, in particular by radio astronomy. Their compactness was implied by their high energies and rapid variability. Although relativistic stars, and in particular collapse models, had been thought about earlier, the relevance of models with strong gravitational fields became much greater. We now deal with the formation and structure of neutron stars and black holes, and binaries thereof, as well as of  $\gamma$ -ray burst sources (GRB). Neutron stars arise as the residues of supernovae.

Pulsars are rotating neutron stars. The properties of binary pulsars, such as J0737-3039A/B discovered in 2003, enable tests of GR in a strong field situation with enormous precision. The results support the reality of gravitational waves, since the predicted radiation output exactly accounts for the observed period decreases.



**Figure 6: An artist's drawing of a black hole named Cygnus X-1. It formed when a large star caved in. This black hole pulls matter from the blue star beside it. Credit: NASA | CXC | M.Weiss.**

X-ray emitting binaries are now generally thought to consist of a black hole and a companion star from which matter is stripped by tidal forces. Figure 6 shows an artist's impression of one. The X-rays come from the accretion disk. The binaries' orbital data implies black hole masses of  $6M_{\odot}$  up to over  $20M_{\odot}$ ; moreover, they often have spins close to the maximum for a Kerr black hole. This supports the prediction that all stars with mass more than  $5M_{\odot}$ , of which we see many, form black holes eventually.

The leading candidate model of short GRB is a double neutron star or black hole/neutron star merger. Long GRBs, found in galaxies where massive stars are forming, are thought to be due to core collapse of stars: some are associated with unusually bright supernovae, and they are seen out to redshifts of 8 or 9. There are jet outflows, as with active galactic nuclei. Some other GRB may be due to a galactic centre black hole swallowing a star.

## 6 GR in the Solar system

After Sputnik's launch in 1957, GR became a practical theory: one had to include the relativistic effects when predicting and setting spacecraft trajectories. Satellites made GPS navigation ('sat-navs') possible: many millions of devices, down to smartphones, therefore incorporate calculations of SR and GR effects.

The tests of GR other than the binary pulsar measurement are tests within the Solar system, often using spacecraft, and some

are purely terrestrial. Light-bending by the Sun is now measured using radio sources and very-long-baseline interferometry. The gravitational redshift effect has been measured very accurately to the height of a tower using the Mossbauer effect, and also using clocks on board rockets. The orbital effects are tested by the fact that GR calculation of spacecraft trajectories works. In addition to these modern versions of the so-called classical tests, observations are made of time delays on signals within the Solar system, of diurnal variation in superconducting gravimeter readings, and of the rotation of the Earth. That gravity acts on itself is shown by the very accurate position measurements of the Moon, using laser ranging to a reflector placed by astronauts. Two perpendicular rotations of axis of an orbiting gyroscope predicted by GR were tested by Gravity Probe B. That showed agreement to 0.3% with GR for geodesic precession, and agreement for frame dragging, but only with limits of the order of 10–20%.

A number of SR and GR effects enter into, and are therefore tested by, GPS. The GPS system uses 24 satellites (plus spares) in 6 orbital planes. They contain clocks stable to about 4 ns over 1 day. At the speed of light, a 1 ns error is about 30 cm. If light speed varied it would upset GPS measurements by this much or more: so the first aspect GPS depends on is the constancy of  $c$  as incorporated in SR. (For extra details of GPS see [6]).

The gravitational redshift from the satellites to the ground is about  $10^4$  times the clock errors. So the gravitational redshift effect could be in Km. The actual calculation includes quadrupolar and centripetal terms, and the clocks are adjusted to compensate for this effect before launch (not doing this in 1977 gave a 1% test of GR).

The time dilation due to the motion of the satellites (known in a rotating frame as the Sagnac effect) is calculated to contribute a correction of 207.4 ns. In tests it accumulated up to 350 ns. So failing to allow for this special relativistic effect would lead to errors of around 60m, enough to cause problems for car satnavs.

The remaining effects are smaller. The leading one arises from the fact that the satellites are in elliptic rather than purely circular orbits. For an orbit of eccentricity  $e = 0.01$  GR predicts a contribution of about 23 ns (i.e. 7 m), calculated using Linton's equation (9). Detailed measurements on one of the satellites (SV#13), which has  $e = 0.01486$ , showed deviations up to 10m. (The prediction of this effect was tested in 1996, in order to show that a GPS management decision not to include this automatically in the next generation of satellites was misguided. The tests included calculation and observation of satellite positions to within 1 mm.)

In 2012 I saw, while on a farm visit, a £150 K seed drill and tractor which is steered to within 4 inches using GPS. I regard this as one of the most unexpected and remarkable consequences of GR.

## 7 Recent theoretical research in GR

The mathematics has also developed, and has influenced other areas of mathematics and physics. We now have a deep understanding of the possible global techniques. Connection and curvature have become fundamental in modern gauge theories. Connections with other areas of mathematics, for example integrable systems (see [7]) have developed. The technical level of many other exciting results sadly precludes discussion here: a small selection of buzzwords could include gluing constructions, regular conformal field equations, the hoop conjecture and



**Figure 7: Seed drill and tractor with GPS in cab (inset). Photograph courtesy of Nicola Crawley.**

the Penrose inequality theorem, isolated and dynamical horizons, and marginally outer trapped surfaces.

Numerical relativity had always been known to require the full power of modern supercomputers. One incidental side-effect was that the supercomputer centre led by Larry Smarr produced the first web browser with a graphical interface, Mosaic. But the GR codes crashed rather than achieving desired simulations. Only in 2005 was it realised that the formulations of GR used were only weakly hyperbolic and hence numerically unstable. New choices of coordinates, and rearrangements of the equations, avoid that now, and have at last enabled us to simulate strong field gravitational effects. Perhaps the most immediately important use is to obtain detailed template predictions of the gravitational waves arising during the inspiral of a compact binary system. Such templates can then be used to enable experiments to dig the signal out of the noise in the very delicate gravitational wave detectors coming on stream.

Two unexpected outcomes of numerical simulations deserve mention. One is the discovery of critical phenomena in gravitational collapse by Choptuik. This arose from considering the evolution of a one-parameter family of initial data sets. A boundary was found between those data sets evolving to form black holes, and those dispersing. At the boundary value, the evolved solution showed special features including certain types of (continuous or discrete) time symmetry, described by scaling laws. Such critical behaviour arises for essentially all such choices of one-parameter families of data ('universality') and for Einstein's equations coupled to a very wide variety of source fields.

The other is the refinement of the Belinskii et al modelling of anisotropic cosmological models, relevant to the initial Big Bang singularity of the Universe. Following matter back to the singularity, their model has a succession of epochs in which spatial derivative terms in the evolution equations are dominated by time derivative terms, punctuated by 'bounces' leading from one such epoch to another. Numerical work by several authors during the 1990s supported the existence of such behaviour, but with the added feature of 'spikes', where the parameters of the epochs no longer vary secularly in space. Exact solutions with spikes are now known, and a recent paper discusses the statistics of spikes' occurrence.

## 8 The future

It would have been hard to predict the relevance of relativity to farming in 1915, but predicting the gravity theory of 2115 is probably harder. There are, however, a number of aspects of GR and



its uses where effort will clearly be concentrated, because they concern problems already encountered.

There are experiments underway, and more planned, to give additional information on most of the cosmological and astrophysical applications of GR. Even the availability of funding for those is a tribute to GR's success. I shall not attempt a comprehensive list, but just mention a few. Some are directed at characterising the dark matter, dark energy and other unseen constituents: for example, measuring neutrinos to clarify the expected neutrino background left by the Big Bang, using terrestrial detectors for various possible constituents of dark matter, or checking the equation of state of dark energy by detailed observations of the CMB, galaxy distributions and lensing. Further refinement of CMB results may confirm or disprove the presence of detectable amounts of B-mode polarisation (found in 2014 by BICEP2 but possibly due to dust): if confirmed, this could come from gravitational waves generated during inflation.

One of the most important experiments may prove to be that of the set of large laser interferometric gravitational wave detectors. Early attempts to directly detect gravitational waves used very large cylindrical bars with piezoelectric readouts, later ones being cooled to just above absolute zero to eliminate thermal noise. The more recent interferometric experiments (LIGO in the US, and soon India, Franco-Italian VIRGO near Pisa, the UK-German GEO600, and Japan's TAMA) work rather like the classic Michelson-Morley experiment: they look for interference fringes between laser light reflected back and forth in giant (up to 4 km long) vacuum tubes, arising as a gravitational wave stretches and compresses the arms.

Such experiments have achieved their initial design sensitivity. No positive result has been obtained. However, the lack of observed waves has for example, enabled inferences about potential sources. LIGO will soon come back on stream with upgrades that will push the sensitivity to the level where, if our understanding of GR and our knowledge of the formation of compact binaries is correct, we can expect a positive result (experts leapt in when a betting firm was offering good odds against such a result). If we did see nothing, that would perhaps be of even greater interest as showing that our physics or astrophysics is wrong.

I began this article by revisiting how SR resolved the tension between Newtonian kinematics and electromagnetism, and GR that between SR and gravity. The tension now, and since the 1920s, is between GR and quantum theory. The latter is an amazing tool for predicting effects on microscopic scales, and we seem to have a rather good theory, the 'Standard Model', covering all

particle physics experiments so far, most recently reinforced by the verification by the Large Hadron Collider of the reality of the Higgs boson. However, GR cannot be treated by similar methods: it is not renormalisable. We so far have no agreed theory of quantum gravity.

Despite this, combinations of GR and quantum physics have been used in a range of problems – in modelling white dwarfs and neutron stars, in predicting Hawking radiation by black holes, and in predicting tensor perturbations in the CMB arising from quantum generation of gravitational waves in inflation. The rationale is like that of using Newtonian theory rather than GR to calculate the flight of a tennis ball: the corrections should be negligibly small for the problem tackled.

Attempts to resolve the GR/quantum tension have been many and various. By far the most popular at present is (super)string theory, in which the point particles envisaged in the Standard Model are replaced by excited strings in higher dimensions. In string theory it is fairly natural to have at low energies a massless spin 2 field which can be identified with gravity. A second active approach is loop quantum gravity, which builds on a spinorial formulation. My own perception is that attempts tend to favour the usual formulation or approach of one of the two theories, whereas one probably needs a theory that treats these conceptual foundations on a more equal footing. Maybe by 2115 we will know.

## REFERENCES

- 1 Pais, A. (1982) *Subtle is the Lord*, Oxford University Press, Oxford.
- 2 Ferreira, P.G. (2014) *The Perfect theory: A century of genuises and the battle over general relativity*, Little, Brown London.
- 3 Ashtekar, A., Berger, B.K., Isenberg, J. and MacCallum, M.A.H. (2014) *General Relativity and Gravitation: A Centennial Perspective*, arXiv:1409.5823.
- 4 Stephani, H., Kramer, D., MacCallum, M.A.H., Hoenselaers, C.A. and Herlt, E. (2003) *Exact solutions of Einstein's field equations*, 2nd edition, Cambridge University Press, Cambridge.
- 5 Hawking, S.W. and Ellis, G.F.R. (1973) *The large scale structure of space-time*, Cambridge University Press, Cambridge.
- 6 Ashby, N. (2003) Relativity in the Global Positioning System, *Living Reviews in Relativity*, vol. 6, no. 1, <http://relativity.livingreviews.org/Articles/lrr-2003-1/>.
- 7 Mason, L.J. and Woodhouse, N.M.J. (1996) *Integrability, Self-duality and twistor theory*, vol. 15 of *London Mathematical Society Monographs*, Oxford Science Publications, Oxford.

Artist's conception of Superstring theory.  
© David Parker | Science Photo Library