CHAPTER 1

THE PRINCIPLE OF RELATIVITY AND THE ORIGIN OF INERTIA

CHAPTER SUMMARY:

After sketching the nature of the central problem in rapid spacetime transport – the manipulation of inertia – Mach's ideas on the topic are mentioned. The origins of the concept of inertia and the principles of relativity and equivalence in the 17th century are outlined. But they did not lead to the theory of relativity, in no small part because of Newton's adoption of absolute space and time. Special relativity theory is investigated, leading to Einstein's discovery of the relationship between energy and inertial mass: $m = E/c^2$, where E is the total non-gravitational energy of an isolated object at rest and c the speed of light in vacuum, How general relativity theory bears on this definition of inertial mass is then explored. The role of the Equivalence Principle – particularly, the prohibition of the localization of gravitational potential energy – is examined, preparing the way for a discussion of Mach's principle in Chapter 2. The behavior of light in the vicinity of negative mass matter is mentioned in anticipation of the third section of the book.

INTRODUCTION:

When you think of traveling around the solar system, especially to the inner planets, a number of propulsion options arguably make sense. When the destination involves interstellar distances or larger, the list of widely accepted, plausible propulsion schemes involving proven physical principles drops to zero. If a way could be found to produce steady acceleration on the order of a "gee" or two for long periods without the need to carry along vast amounts of propellant, interstellar trips within a human lifetime would be possible. But they would not be quick trips by any stretch of the imagination. If a way to reduce the inertia of one's ship could be found, such trips could be speeded up as larger accelerations than otherwise feasible would become available. But such trips would still be sub-lightspeed, and the time dilation effects of Special Relativity Theory would still apply. So when you returned from your journeys, all of your friends and acquaintances would have long since passed on.

As is now well-known, wormholes and warp drives would make traversing such distances in reasonable times plausible. And returning before your friends age and die is possible. Indeed, if you choose, you could return before you left. But you couldn't kill yourself before you leave. A wide range of "traversable" wormholes with a wide range of necessary conditions are possible. The only ones that are manifestly practical are, in the words of Michael Morris and Kip Thorne, "absurdly benign". Absurdly benign wormholes are those that restrict the distortion of spacetime that forms their throats to modest dimensions – a few tens of meters typically – leaving the surrounding spacetime flat. And their throats are very short. Again, a few tens of meters or less typically. Such structures are called "stargates" in science fiction. The downside of such things is that their implementation not only requires Jupiter masses of "exotic" matter, they must be *assembled* in a structure of very modest dimensions. Imagine an object with the mass of Jupiter (about 600 times the mass of the Earth) sitting in your living-room, or on your patio.

Even the less daunting methods of either finding a way to accelerate a ship for long intervals without having to lug along a stupendous amount of propellant or reduce its inertia significantly do not seem feasible. Sad to say, solutions to none of these problems – vast amounts of propellant, or inertia reduction, or Jupiter masses of exotic matter to make wormholes and warp drives – are presently to be found in mainstream physics. But Mach effects – predicted fluctuations in the masses of things that change their internal energies as they are accelerated by external forces – hold out the promise of solutions to these problems. To understand how Mach effects work, you first have to grasp "Mach's principle" and what it says about how the inertial properties of massive objects are produced. You can't manipulate something that you don't understand, and inertia is the thing that needs to be manipulated if the goal of rapid spacetime transport is to be achieved. Mach's principle, though, can only be understood in terms of the principle of relativity and Einstein's two theories thereof. While the theories of relativity, widely appreciated and understood, do not need a great deal of formal elaboration, the same cannot be said of Mach's principle. Mach's principle has been, from time-to-time, a topic of considerable contention and debate in the gravitational physics community, though at present it is not. The principle, however, has not made it into the mainstream canon of theoretical physics. This means that a certain amount of formal elaboration (that is, mathematics) is required to insure that this material is done justice. The part of the text that does not involve such formal elaboration will be presented in a casual fashion without much detailed supporting mathematics. The formal material, of interest chiefly to experts and professionals, will be set off from the rest of the narrative either in the body of the text, or placed in appendixes. Most of the appendixes, however, are excerpts from the original literature on the subject. Reading the original literature, generally, is to be preferred to reading a more-or-less accurate paraphrasing thereof.

THE RELATIVITY CONTEXT OF MACH'S PRINCIPLE:

Ernst Mach, an Austrian physicist of the late 19th and early 20th centuries, is now chiefly known for Mach "numbers" (think Mustang Mach One, or the Mach 3, SR71 Blackbird). But during his lifetime, he was chiefly known for penetrating critiques of the foundations of physics. In the 1880s he published a book – *The Science of Mechanics* – where he had taken Newton to task for a number of things that had come to be casually accepted about the foundations of mechanics – in particular, Newton's notions of absolute space and time, and the nature of inertia, that property of real objects that causes them to resist changes in their states of motion. Einstein, as a youngster, had read Mach's works, and it is widely believed that Mach's critiques of "classical", that is, pre-quantum mechanical, physics deeply influenced him in his construction of his theories of relativity. Indeed, Einstein, before he became famous, had visited Mach in Vienna, intent on trying to convince Mach that atoms were real. (The work Einstein had done on "Brownian motion", a random microscopic motion of very small particles, to get his doctoral degree had demonstrated the fact that matter was atomic.) Mach had been cordial, but the young Einstein had not changed Mach's mind.

Nonetheless, it was Mach's critiques of space, time, and matter that had the most profound effect on Einstein. And shortly after the publication of his earliest papers on general relativity theory (GRT) in late 1915 and early 1916, Einstein argued that, in his words, "Mach's principle" should be an explicit property of GRT. Einstein defined Mach's principle as the "relativity of inertia", that is, the inertial properties of material objects should depend on the presence of other material objects in the surrounding spacetime, and ultimately, the entire universe. Framing the principle this way, Einstein found it impossible to show that Mach's principle was a fundamental feature of GRT. But Einstein's insight started arguments about the "origin of inertia" that continue to this day. Those arguments can only be understood in the context of Einstein's theories of relativity, as inertia is an implicit feature of those theories (and indeed any theory of mechanics). Since the issue of the origin of inertia is not the customary focus of examinations of the theories of relativity, we now turn briefly to those theories with the origin of inertia as our chief concern.

Einstein had two key insights that led to his theories of relativity. The first was that if there really is no preferred reference frame – as is suggested by electrodynamics[*] – it must be the case that when

---

[*] The simple case analyzed by Einstein in his first paper on special relativity theory – titled "On the Electrodyamics of Moving Bodies" – is the motion of a magnet with respect to a loop of wire. If the relative motion of the magnet and wire causes the "flux" of the magnetic field through the loop of wire to change, a current flows in the loop while the magnetic field is changing. It makes no difference to the current in the loop whether you take the loop as at rest with the magnet moving, or vice versa. The rest of the paper consists of Einstein's demonstration that the mathematical machinery that gets you from the frame of reference where the magnet is at rest to the frame where the loop is at rest requires that the speed of light measured in both frames is the same, or "constant". This is only possible if space and time are inextricably interlinked, destroying Newton's absolute notions of space and time as physically distinct,

anyone measures the speed of light in vacuum, s/he always get the same number, no matter how s/he is moving with respect to the source of the light. When the implications of this fact for our understanding of time are appreciated, this leads to special relativity theory (SRT). SRT, in turn, leads to a connection between energy and inertia that was hitherto unappreciated. The curious behavior of light in SRT is normally referred to as the speed of light being a "constant". That is, whenever anyone measures the speed of light, no matter whom, where, or when they are, they always gets the same number – in centimeter-gram-second (cgs) units, 3×10^{10} cm/sec. While this works for SRT, when we get to general relativity theory (GRT) we will find this isn't quite right. But first we should explore some of the elementary features of SRT, as we will need them later. We leave detailed consideration of Einstein's second key insight – the Equivalence Principle – to the following section where we examine some of the features of general relativity theory.

THE PRINCIPLE OF RELATIVITY:

Mention relativity, and the name that immediately jumps to mind is Einstein. And in your mental timescape, the turn of the 20th century suffuses the imagery of your mind's eye. The *principle* of relativity, however, is much older than Einstein. In fact, it was first articulated and argued for by Galileo Galilei in the early 17th century. A dedicated advocate of Copernican heliocentric astronomy, Galileo was determined to replace Aristotelian physics, which undergirded the prevailing Ptolemaic geocentric astronomy, with new notions about mechanics. Galileo hoped, by showing that Aristotelian ideas on mechanics were wrong, to undercut the substructure of geocentric astronomy. Did Galileo change any of his contemporaries' minds? Probably not. Once people think they've got something figured out, it's almost impossible to get them to change their minds.[†] As Max Planck remarked when asked if his contemporaries had adopted his ideas on quantum theory (of which Planck was the founder), people don't change their minds. They die. But Galileo did succeed in influencing the younger generation of his day.

Galileo's observations on mechanics are so obvious that it is, for us, almost inconceivable that any sensible person could fail to appreciate their correctness. But the same could have been said of Aristotle in Galileo's day. Arguing from commonplace experience, Aristotle had asserted that a force had to be applied to keep an object in motion. If you are pushing a cart along on a level road and stop pushing, not long after the cart will stop moving. However, even to a casual observer, it is obvious that how quickly the cart stops depends on how smooth and level the road is and how good the wheels, wheel bearings, and axel are. Galileo saw that it is easy to imagine that were the road perfectly smooth and level, and the wheels, wheel bearings, and axel perfect, the cart would continue to roll along indefinitely. Galileo, in his *Science of Mechanics* (published in 1638, a few years before he died), didn't put this argument in terms of carts. He used the example of a ball rolling down an incline, and then along a smooth level plane, and eventually up an incline. From which he extracted that objects set into motion remain in that state of motion until influenced by external agents. That is, Newton's first law of mechanics. Newton got the credit because he asserted it as a universal law, where Galileo only claimed that it worked below the sphere of the Moon. After all, he was a Copernican, and so assumed that the motions of heavenly bodies were circular.

Galileo figured out most of his mechanics in the 1590s, so when he wrote the *Dialog on the Two Chief World Systems* in the 1620s (that got him condemned by the Inquisition a few years later for insulting the Pope in one of the dialogs), he had his mechanics to draw upon. One of the arguments he used involved dropping a cannon ball from the crow's nest on the mast of ship moving at steady speed across a smooth harbor. Galileo claimed that the canon ball would fall with the motion of the ship, and thus land at the base of the mast, whereas Aristotle would have the cannon ball stop moving with the ship when it was released. As a result, according to Aristotle, if the ship is moving at a good clip, the cannon ball should land far from the base of the mast as the ship would keep moving and the cannon ball would not. Anyone who has ever

---

independent entities. The concept underlying the full equivalence of the two frames of reference is the principle of relativity: that all inertial frames of reference are equally fundamental and no one of them can be singled out as more fundamental by any experiment that can be conducted locally.

[†] Galileo himself was guilty of this failing. When Kepler sent him his work on astronomy (the first two laws of planetary motion anyway), work that was incompatible with the compounded circular motions used by Copernicus, Galileo, a convinced Copernican, ignored it.

dropped something in a moving vehicle (and a lot who haven't) knows that Galileo was right. Galileo was describing, and Newton codifying, "inertial" motion. Once Galileo's take on things is understood, Aristotelian ideas on mechanics become features of the intellectual landscape chiefly of interest to historians.

Galileo did more than just identify inertial motion. He used it to articulate the principle of relativity. Once you get the hang of inertial motion, it's pretty obvious that there is, as we would say today, no preferred frame of reference. That is, on the basis of mechanics with inertial motion, there is no obvious way to single out one system as preferred and at rest, with respect to which all other systems either move or are at rest. Galileo's way of making this point was to consider people shooting billiards in the captain's cabin of the ship where the cannon ball got dropped from the crow's nest. He posed the question: if all of the portholes were covered up so you couldn't see what's going on outside the cabin, can you tell if the ship is moving across the harbor at constant speed and direction, or tied up at the dock, by examining the behavior of the balls on the billiards table? No, of course not. Any inertial frame of reference is as good as any other, and you can't tell if you are moving with respect to some specified inertial frame by local measurements. You have to go look out the porthole to see if the ship is moving with respect to the harbor or not. This is the principle of relativity.

Galileo's attack on Aristotelian mechanics didn't stop at identifying inertial motion. Aristotle, again on the basis of casual observations, had asserted that heavier objects fall faster than light objects. It had been known for centuries that this was wrong. But Aristotelians had either ignored the obvious, or concocted stories to explain away "anomalous" observations. Galileo took a cannon ball and a musket ball to the top of the leaning Tower of Pisa and dropped them together. (But not in front of the assembled faculty of the local university.) He noted that the musket ball arrived at the ground "within two fingers' breadth" of the cannon ball. The cannon ball, being more than ten times more massive than the musket ball, should have hit the ground far in advance of the musket ball. Aristotle was falsified. Galileo surmised that the small difference in the arrival times of the two balls was likely due to air resistance, and inferred that in a vacuum the arrivals would have been simultaneous. Moreover, he inferred that the time of fall would have been independent of the compositions, as well as the masses, of the two balls. This is the physical content of, as Einstein later named it, the Equivalence Principle.

Isaac Newton, one of the best physicists of all time,[‡] took on the insights of Galileo, asserted them as universal principles and codified them into a formal system of mechanics. He worked out the law of universal gravitation, and saw that his third law – the necessity of an equal and opposite inertial reaction force for all "external" applied forces – was needed to complete the system of mechanics. He did experiments using pendula to check up on Galileo's claim that all objects fall with the same acceleration in the Earth's gravity field.[§] His synthesis of mechanics and gravity, published in 1687 as the *Principia Mathmatica Philosophia Naturalis* ranks as one of the greatest achievements of the human intellect. But if Newton incorporated the principle of relativity and the Equivalence Principle into his work, one might ask: why didn't he figure out the *theory* of relativity? Absolute space, and absolute time. Newton was nothing, if not thorough. So he provided definitions of space and time, which he took to be completely separate physical entities (as indeed they appear to us today on the basis of our everyday experience of reality).

[‡] Most historians of science would probably name Newton the greatest physicist of all time. Most physicists would likely pick Einstein for this honor (as did Lev Landau, a brilliant Russian theoretical physicist in the mid-20th century). Getting this right is complicated by the fact that Newton spent most of his life doing alchemy, biblical studies, pursuing a "patent" of nobility, and running the government's mint after the mid-1690s. Physics and mathematics were sidelines for him. Einstein, on the other hand, aside from some womanizing, spent most of his life doing physics, albeit out of the mainstream after the late 1920s. It's complicated.

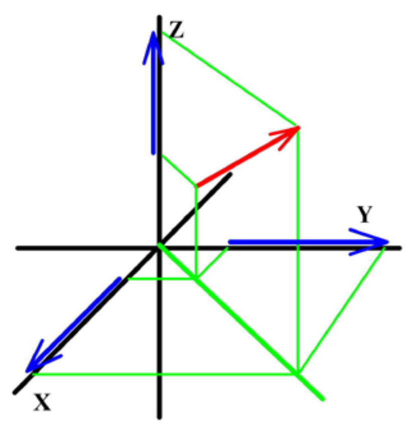
[§] The period of a pendulum depends only on its length, so you can put masses of all different weights and compositions on a pendulum of some fixed length, and its period should remain the same. Newton did this using a standard comparison pendulum and found that Galileo was right, at least to about a part in a thousand. Very much fancier experiments that test this principle have been (and continue to be) done to exquisitely high accuracy.

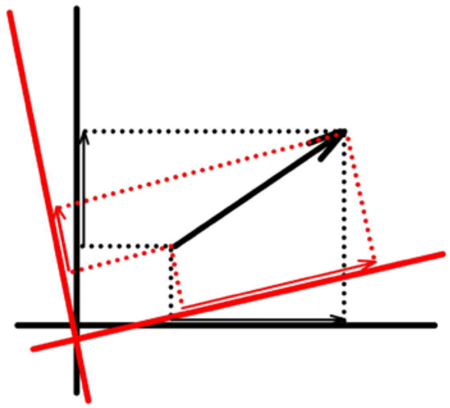
Alas, it turns out that this is wrong. And if you make this assumption, as Newton did, you can't discover the theory of relativity.

SPECIAL RELATIVITY THEORY:

Nowadays, everyone knows that SRT takes the physically independent, absolute Newtonian notions of space and time and inextricably mixes them up together to get "spacetime". That is, in the Newtonian world-view, all observers, no matter where they are or how they are moving with respect to each other (or any other specified frame of reference), see the same space and measure the same time. Einstein's profound insight was to see that if all observers measure the same value for the speed of light (in vacuum), this can't be true, for if one observer measures a particular value in Newtonian space and time, and another observer is moving with respect to him, that other observer *must* measure a different value for the speed of light, *c*. But if this is so, then we can pick out some frame of reference, for whatever reason, and call it the fundamental frame of reference (say, the frame of reference in which nearby galaxies are, on average, at rest, or the frame in which the speed of light has some preferred value in a particular direction), and we can then refer all phenomena to this fundamental frame. The *principle* of relativity, however, requires that such a frame with preferred physical properties that can be discovered with purely local measurements not exist, and the only way this can be true is if the measured speeds of light in all frames have the same value, making it impossible on the basis of local experiments to single out a preferred frame of reference. So, what we need is some mathematical machinery that will get us from one frame of reference to another, moving with respect to the first, in such a way that the speed of light is measured to have the same value in both frames of reference. The "transformation" equations that do this are called the "Lorentz transformations" because they were first worked out by Hendrick Antoon Lorentz a few years before Einstein created SRT. (Lorentz, like Einstein, understood that the "invariance" of the speed of light that follows from electrodynamics required the redefinition of the notions of space and time. But unlike Einstein, he continued to believe, to his death roughly 20 years after Einstein published his work on SRT, that there were underlying absolute space and time to which the "local" values could be referred.)

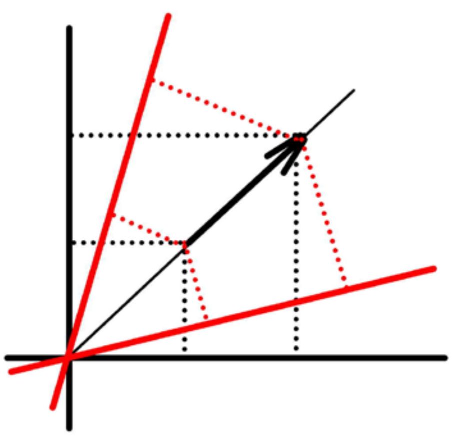
Many, *many* books and articles have been written about SRT. Some of them are very good. As an example, see Taylor and Wheeler's *Spacetime Physics*. We're not going to repeat the customary treatments here. For example, we're not going to get involved in a discussion of how time slows when something is moving close to the speed of light and the so-called "twins paradox". Rather, we're going to focus on the features of SRT that we'll need for our discussion of Mach's principle and Mach effects. Chief among these is what happens to the physical quantities involved in Newtonian mechanics like energy, momentum, and force. The way in which SRT mixes up space and time can be seen by choosing some spacetime frame of reference, placing some physical quantity at some location, and examining how it looks in two different frames of reference. Mathematically speaking, physical quantities come in one of three types: scalars, vectors, or tensors. Scalars are those things that have only magnitude, like temperature or energy, and thus can be specified by one number at every event in spacetime. Vectors are things that have both magnitude and point in some direction, like momentum and force. They are customarily represented by arrows that point in the spatial direction of the quantity. Their length represents the magnitude of the quantity at the point where their back end is located (see Figure 1). In Newtonian physics, with its absolute space and time, that means that they point in some direction in space, and thus require three numbers to be fully specified – the projected lengths of the vector on three suitably chosen coordinate axes, one for each of the dimensions of space. Since time is treated on the same footing as space in spacetime, four-vectors in spacetime need four numbers to specify their projections on the four axes of spacetime. Tensors are used to specify the magnitudes that characterize more complicated things like elasticity which depend on the direction in which they are measured. We'll not be concerned with them at this point.





The things we will be most interested in are those represented by vectors. So we start with a vector in a simple three dimensional space as in Figure 1. We include some Cartesian coordinate axes of an arbitrarily chosen "frame of reference".** The projections of the vector on the coordinate axes are the component vectors shown in Figure 1. The distinctive feature of the vector is that its length (magnitude) shouldn't depend on how we choose our coordinates. That is, the length of the vector must be "invariant" with respect to our choice of coordinates. This will be the case if take the square of the length of the vector to be the sum of the squares of the component vectors, because the vector and its components form a right triangle, and the sum of the squares of the shorter sides of the triangle is equal to the square of the longest side. Note, as the Pythagorean theorem informs us, that this is true even if we choose some other coordinates, say those in red in Figure 2, a two dimensional simplification of Figure 1. The component vectors are different from those in the black coordinates, but the sum of their squares is the same. This is true when space is "flat", or "Euclidean". It is not in general true when space is "curved".

To make the transition to spacetime we need to be able to treat space and time on an equal footing. That is, if we are to replace one of our two space axes in Figure 2 with a time coordinate, it must be specified in the same units as that of the remaining space coordinate. This is accomplished by multiplying time measurements by the speed of light. This works because the speed of light is an invariant – the same for all observers – and when you multiply a time by a velocity, you get a distance. So, with this conversion, we end up measuring time in, say, centimeters instead of seconds. Should you want to measure time in its customary units – seconds – to get everything right you'd have to divide all spatial distances by the speed of light. Spatial distances would then be measured in light-seconds. It doesn't matter which choice you make, but we'll use the customary one where times are multiplied by c.



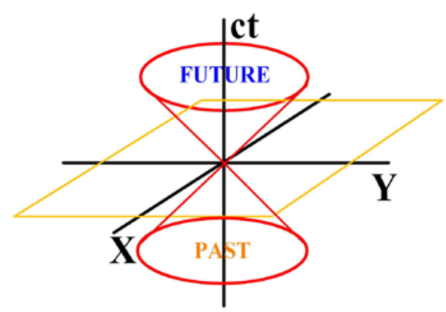
We now consider a vector in our simple two dimensional spacetime in Figure 3. Were spacetime like space, we would be able to specify the length of our vector as the sum of the squares of its projections on the space and time axes. But spacetime isn't like space. The requirement that the speed of light be measured to have the same value in all spacetime frames of reference forces us to accept that spacetime is "pseudo-Euclidean". Pseudo-Euclidean? What's that? Well, as in a Euclidean (or flat) space, we still use the squares of the projections of vectors on their coordinate axes to compute their lengths. And when there is more than one space dimension, the "spacelike" part of the vector is just the sum of the squares of the projections on the space axes. But the square of the projection on the time axis (the "timelike" part of the vector) is *subtracted* from the spacelike part of the vector to get its total length in spacetime.

Why do you subtract the square of the timelike component of the vector from the square of the spacelike part? Because time stops for things traveling at the speed of light (as Einstein discerned by imagining looking at a clock while riding on a beam of light, since moving away from the clock on the light beam carrying the clock time, time stops). Look at the two-dimensional spacetime and ask, how can we

** Named for their inventor, Rene Descartes, these axes are chosen so that they are (all) mutually perpendicular to each other. He got the idea lying in bed contemplating the location of objects in his room and noting that their places could be specified by measuring how far from a corner of the room they were along the intersections of the floor and walls.

construct a path for light in the spacetime so that its length in spacetime is zero, but its distances in space and time separately aren't zero? Well, it's impossible to add two non-zero positive numbers to get zero. And the squares of the component vectors are always positive. So it must be that we have to subtract them. And to get zero, the two numbers must be the same. This means that the path of light rays in our two dimensional spacetime is along the line at a 45 degree angle to the two coordinate axes so that the distance in space along the path is the same as the distance in time. Since a clock taken along this path registers no passage of time, it is called the path of zero "proper" time. A "proper" measurement is one that is made moving with the thing measured – that is, the measurement is made in the "rest" frame of the thing measured.

In Figure 2 we saw how vectors in different sets of coordinates preserved their lengths. You may be wondering, what happens when you rotate (and "translate" if you choose) the coordinates of spacetime in Figure 3? The answer is that you can't do the simple sort of rotation done in Figure 2, because as soon as you rotate the space and time axes as if they were Euclidean, the lengths of the spacelike and timelike component vectors for a vector lying along a light path have unequal lengths, so the difference of their squares is no longer zero – as it must be in spacetime. Evidently, the only way an observer moving with respect to one at rest in our original (black) spacetime coordinates can get zero for the length of a vector along a light path is if the coordinates look like the red coordinates in Figure 3. Oh, and a word about "lightcones". If we imagine our two dimensional spacetime to now have another spacelike dimension, we can rotate our 45 degree light path around the timelike axis, creating a conical surface in which light that passes through the origin of coordinates propagates. That surface is the future lightcone of the event at the origin of coordinates if it lies in the direction of positive time. The lightcone that lies in the direction of negative time is the past lightcone of the event at the origin of coordinates. Events that lie within the past and future lightcones of the event at the origin of coordinates can communicate with the event at the origin at sublight speeds. Those that lie outside the lightcones cannot. They are said to be "spacelike" separated from the origin of coordinates.



The *principle* of relativity forces us to accept that the speed of light is measured by all observers to have the same value, irrespective of their motion. And the invariance of the speed of light in turn forces us to accept that space and time are interconnected, and that the geometry of spacetime is pseudo-Euclidean. The question then is: what does this do to Newtonian mechanics? Well, not too much. The first and third laws of mechanics aren't affected at all. Bodies in motion at constant velocity, or at rest, in inertial frames of reference not acted on by external forces keep doing the same thing (first law). And when forces act on objects, they still produce equal and opposite inertial reaction forces (third law). Customarily, it is said that the second law is only affected in that the mass must be taken to be the "relativistic" mass (as the mass of an object, as measured by any particular observer, depends on the velocity of the object with respect to the observer). This is all well and good, but we want to take a bit closer look at the second law.

The most famous equation in all of physics that $E = mc^2$ replaced 20 or 30 years ago was $\mathbf{F} = m\mathbf{a}$, or force equals mass times acceleration − the simple version of Newton's second law. Boldface letters, by the way, denote vectors, and normal Latin letters denote scalars. ($E = mc^2$, or energy equals mass times the square of the speed of light, is identified by Frank Wilczek as Einstein's first law, terminology we adopt.) The correct, complete statement of Newton's second law is that the application of a force to a body produces changing momentum of the body in the direction of the applied force and the rate of change of momentum depends on the magnitude of the force, or $\mathbf{F} = d\mathbf{p}/dt$.[††] (Momentum is

[††] Since $d\mathbf{p}/dt = d(m\mathbf{v})/dt = m\mathbf{a} + \mathbf{v}\,dm/dt$, we see that force is a bit more subtle than $m\mathbf{a}$. Indeed, if you aren't careful, serious mistakes are possible. Tempting as it is to explore one or two in some detail, we resist and turn to issues with greater import.

customarily designated by the letter "p". The "operator" d/dt just means take the time rate of change of, in this case, **p**.) Now, there is a very important property of physical systems implicit in Newton's second law. If there are no "external" forces, the momentum of an object (or collection of objects) doesn't change. That is, momentum is "conserved". And this is true for all observers since the lengths of vectors in space are measured to be the same by all observers in Newtonian physics. Moreover, you can move vectors around from place-to-place and time-to-time, preserving their direction, and they don't change. (Technospeak: vectors are invariant under infinitesimal space and time translations.[‡‡]) The question, then, is: How do we generalize this when we make the transition to spacetime required by relativity? Evidently, the three-vector momentum in absolute space must become a four-vector momentum in spacetime, and the length of the four-vector momentum in spacetime must be invariant in the absence of external forces.

In Newtonian physics the momentum of an object is defined as the product of its mass m and velocity $\mathbf{v}$, that is, $\mathbf{p} = m\mathbf{v}$. Mass, for Newton, was a measure of the "quantity of matter" of an object. In the early 20th century, the concept of mass was expanded to encompass the notion that mass is the measure of the inertial resistance of entities to applied forces, that is, the m in $\mathbf{F} = m\mathbf{a}$, and m might include things hitherto not thought to be "matter". Mass, by the way, is also the "charge" of the gravitational field. Here, however, we are interested in the inertial aspect of mass. When we write momentum as a four-vector, the question is: what can we write for the timelike part of the four-vector that has the dimension of momentum? Well, it has to be a mass times a velocity, indeed, the fourth (timelike) component of the four-velocity times the mass. What is that fourth component of all velocities? The velocity of light, because it is the only velocity that is invariant (the same in all circumstances as measured by all observers). This makes the timelike component of the four-momentum equal to mc. The definition of the four-force, then, would seem to be the rate of change of four-momentum. What, or whose rate of change? After all, the rate of time depends on the motion of observers, and by the principle of relativity, none of them are preferred. Well, it would seem that the only rate of time that all observers can agree upon is the rate of time in the rest frame of the object experiencing the force – that is, the "proper" time of the object acted on by the force.[§§] So, the relativistic generalization of Newton's second law is: when an external force is applied to an object, the four-force is equal to the rate of change of the four-momentum with respect to proper time of the object acted upon.

You may be wondering: What the devil happened to Mach's principle and the origin of inertia? What does all of this stuff about three- and four-vectors, forces and momenta (and their rates of change) have to do with the origin of inertia? Well, inertia figures into momenta in the mass that multiplies the velocity. Mass, the measure of the "quantity of matter" for Newton, is the quantitative measure of the inertia of a body – its resistance to forces applied to change its state of motion. The more mass an object has, the smaller its acceleration for a given applied force. But what makes up mass? And what is its "origin"? From Einstein's first law, $E = mc^2$, we know that energy has something to do with mass. If we write Einstein's second law, $m = E/c^2$,[***] and we take note of the fact that SRT explicitly ignores gravity, then it would appear that we can define the mass of an object as the total (non-gravitational) energy of the object divided by the speed of light squared. How does this relate to the timelike part of the four-momentum? Well, look at the timelike part of the four-momentum: mc. If you multiply this by c, you get

[‡‡] When something doesn't change when it is operated upon (in this case, moved around), it is said to possess symmetry. Note that this is related to the fact that momentum is "conserved". In 1918 Emmy Noether, while working for Einstein, proved a very general and powerful theorem (now known as "Noether's theorem") showing that whenever a symmetry is present, there is an associated conservation law. Noether, as a woman, couldn't get a regular academic appointment in Germany, notwithstanding that she was a brilliant mathematician. When the faculty of Gottingen University considered her for an appointment, David Hilbert, one of the leading mathematicians of the day, chided his colleagues for their intolerance regarding Noether by allowing as how the faculty were not the members of a "bathing establishment".

[§§] As mentioned earlier, the term "proper" is always used when referring to a quantity measured in the instantaneous frame of rest of the object measured. The most common quantity, after time, designated as proper is mass – the restmass of an object is its proper mass.

[***] We continue to use Frank Wilczek's, enumeration of Einstein's laws.

mc^2. Dimensionally, this is an energy. And since relativistic mass depends on velocity, as the velocity of an object with some rest mass changes, its energy increases because its mass increases, notwithstanding that c doesn't change. So, with this simple artifice we can transform the four-momentum vector into the energy-momentum four-vector. In Newtonian physics, energy and momentum are separately conserved. In SRT it is the energy-momentum four-vector that is conserved. Einstein figured this out as an afterthought to his first work on SRT. And he didn't have the formalism and language of four-vectors to help him. That wasn't invented until a couple of years later – by one of his former teachers, Herman Minkowski[†††]. Einstein had posed himself the question, "Does the inertia of a body depend on its energy content?" Indeed, that is the title of the paper that contains his second law: $m = E/c^2$. His first law doesn't even appear anywhere in the paper.

Distinguishing between Einstein's first and second laws, as they are the same equation in different arrangements, may seem a quibble to you. But as Frank Wilczek points out in his delightful book, *The Lightness of Being*, the way you put things can have profound consequences for the way you understand them. When you write $E = mc^2$, it's natural to notice that m is multiplied by, in everyday units like either the cgs or meter-kilogram-ssecond (mks) system, an enormous number. Since m is normally taken to refer to the restmass of an object, that means that restmass contains an enormous amount of energy, and your thoughts turn to power plants and bombs that might be made using only a minuscule amount of mass. When you write, as Einstein did in 1905, $m = E/c^2$ completely different thoughts come to mind. Instead of ogling the enormous amount of energy present in small amounts of restmass, you appreciate that all non-gravitational energy contributes to the inertial masses of things. Non-gravitational? Why doesn't gravitational energy contribute to inertial mass? Well, it does. But only in special circumstances, in particular, in the form of gravity waves. There are other special circumstances where it doesn't. Gravitational potential energy due to the presence of nearby sources makes no contribution. But all these subtleties are part of GRT, and that's in the next section. For now we need only note that it is energy, not restmass alone, that is the origin of inertial mass. As Wilczek notes, more than 95% of the mass of normal matter arises from the energy contained in the restmassless gluons that bind the quarks in the neutrons and protons in the nuclei of the atoms that make it up.

A caveat should be added here. The foregoing comments about energy and mass only strictly apply to localized, isolated objects at rest in some local inertial frame of reference. The comments are also true in some other special circumstances. But in general things get more complicated when observers are moving with respect to the object observed and when other stuff is in the neighborhood that interacts with the object whose mass is being considered. Moving observers can be accommodated by stipulating that m is the relativistic mass. But nearby interacting entities can be trickier to deal with.

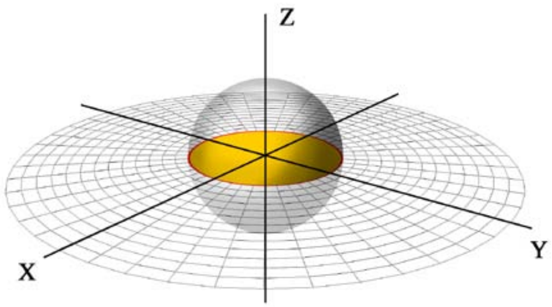
Summing up, the principle of relativity demands that the speed of light be "constant" so that it is impossible to identify (with local measurements) a preferred inertial frame of reference. The constancy of the speed of light leads to SRT which in turn leads to pseudo-Euclidean spacetime. When Newton's second law is put into a form that is consistent with SRT, the four-momentum (the proper rate of change of which is the four-force) multiplied by the object's four-velocity for zero spatial velocity, leads to $E = mc^2$. When this is written as Einstein's second law [$m = E/c^2$], it says that energy has inertia, in principle, even if the energy isn't associated with simple little massy particles that you can put on a balance and weigh. But there is no explanation why energy, be it massy particles or photons (particles of light) or gluons, has inertia. So we turn to general relativity theory to see if it sheds any light on the issue of the origin and nature of inertia.

GENERAL RELATIVITY:

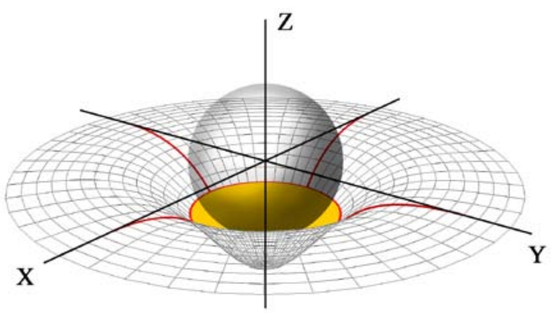
Einstein's first key insight – that the principle of relativity demanded that the speed of light be measured to be the same by all observers, and that this required space and time to be conceived as spacetime – led to SRT. His second key insight – that Einstein called "the happiest thought of my life" – was his so-called "Equivalence Principle" (EP): that the action of a gravity field which causes everything to

[†††] Minkowski characterized Einstein the undergraduate student as a "lazy dog".

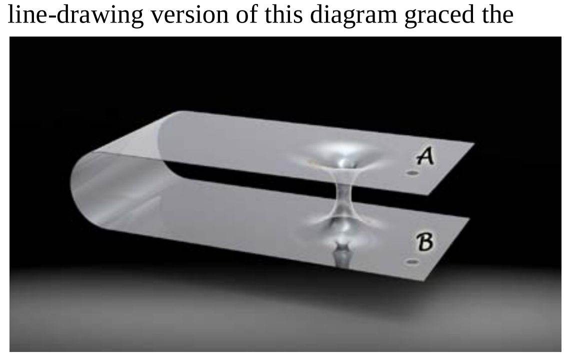
"fall" in the direction of the field with the same acceleration irrespective of their masses and compositions is equivalent to the behavior of everything in the absence of local gravity fields, but located in an accelerating frame of reference – say, in a rocket ship accelerating in deep space. Einstein realized that this equivalence could only be true if local inertial frames of reference – those in which Newton's first law is true – in the presence of a local concentration of matter like the Earth are those that are in a state of "free fall". And for this to be true, it must be the case that local concentrations of matter should distort the geometry of spacetime rather than produce forces on objects in their vicinity. This eventually led Einstein to his general relativity theory (GRT) where, in the words of John Wheeler, "spacetime tells matter how to move, and matter tells spacetime how to bend." "Matter", with its property of inertia and as the "charge" (or source) of gravity, does not simply produce a field in spacetime, the field is the distortion of spacetime itself. This is why GRT is called a "background independent" theory of gravity. It is this fact – that the field is not something in spacetime, but rather the distortion of spacetime itself – that makes possible the wormholes and warp drives that enable serious rapid spacetime "transport", if we can figure out how to build them.



The customary visual rendition that is intended to show how this works is a "hyperspace embedding diagram". Consider the case of a simple spherical source of gravity, say, a star or somesuch that is not changing in time. The warping that this source effects on space is the stretching of space in the radial direction. Space perpendicular to the radial direction is unaffected by the presence of the star. To show this, we consider a two dimensional plane section through the center of the star, as shown in Figure 5. We now use the third dimension, freed up by restricting consideration to the two dimensional plane in Figure 5, as a hyperspatial dimension – that is, a dimension that is not a real physical dimension – that allows us to show the distortion of the two dimensional plane through the center of the star. This is illustrated in Figure 6. Note that the radial stretching that distorts the two dimensional plane through the center of the star has no effect at all on the circumferences of the circles in the plane centered on the center of the star.



As everyone now knows, wormholes are shortcuts through hyperspace between two locations in spacetime separated by arbitrarily long distances through normal spacetime. The now famous embedding diagram of a wormhole is shown in Figure 7. A simple line-drawing version of this diagram graced the pages of Misner, Thorne and Wheeler's classic text on gravity: *Gravitation*. Published in 1973. Indeed, Figure 7 is a shaded version of their classic embedding diagram. The length of the throat of the wormhole is exaggerated in this diagram. But it conveys the point: the distance through the wormhole is much shorter than the distance through the two-dimensional surface that represents normal spacetime. This exaggeration is especially pronounced in the case of an "absurdly benign" wormhole – a wormhole with a throat only a few meters long and with all of the flarings at each end of the throat restricted to at most a meter or two. A rendition of the appearance of such a wormhole is shown in Figure 8. Note that the wormhole is a four-dimensional sphere, so the appearance of circularity is deceptive. It should be noted that the wormhole can



connect both spatially and temporally distant events. That is, wormholes can be designed to be time machines connecting distant events in both the future and past. Famously, Stephen Hawking has argued that the actual construction of such wormholes is prohibited by the laws of physics. But not all physicists share his conviction in this. Hawking's argument depends on making time machines with only smooth deformations of spacetime. That is, tearing spacetime to connect two distant events (and causing "topology change") is prohibited. With only smooth deformations allowed, you always end up at some point creating a "closed timelike curve" (CTC); and if even one measly photon starts endlessly circulating along the CTC, its multiple copies build up infinite energy instantly in the wormhole throat, blowing it up. If we can tear spacetime, need I say that this disastrous situation can be avoided? But enough of such silliness. Back to the EP.

"Jump to Titan City" - © 2010 Nembo Buldrini

I'll bet you won't be surprised to learn that the EP has been a source of criticism and debate since Einstein introduced it and made it one of the cornerstones of GRT. The feature of the EP that many critics dislike is that it permits the local elimination of the local gravitational field by a simple transformation to a suitably chosen accelerating frame of reference, or, equivalently, *to a suitably chosen spacetime geometry*. This is only possible because everything responds to a gravitational field the same way. We know that Newtonian gravity displays this same characteristic (that Galileo discovered), and we can use Newtonian gravity to illustrate this point. Consider an object with (passive gravitational) mass m acted upon by the Earth's gravity field. The force exerted by the Earth is just:

$$\mathbf{F} = \frac{GMm}{R^3}\mathbf{R},$$

where M and R are the mass and radius of the Earth. By Newton's second law, $\mathbf{F}$ is also equal to $m\mathbf{a}$, so:

$$\mathbf{F} = \frac{GMm}{R^3}\mathbf{R} = m\mathbf{a},$$

and since $\mathbf{R}$ and $\mathbf{a}$ point in the same direction we can drop the vector notation and canceling m write:

$$\frac{GM}{R^2} = a.$$

Note that this is true regardless of the value of m, and this is only possible if the Equivalence Principle is correct. That is, the passive gravitational mass that figures into the gravitational force and the inertial mass that figures into Newton's second law, in principle at least, need not necessarily be the same. Only if they are the same does the cancellation that shows all gravitational accelerations of various bodies to be the same carry through.

Gravity is the only interaction or force that satisfies the Equivalence Principle, making gravitation the unique universal interaction. Of the known interactions – gravity, electromagnetism, and the strong and weak forces – it is the only interaction that can be eliminated locally by a suitable choice of geometry. The local elimination of "apparent" forces is also possible for "fictitious" forces: "forces" that appear because of an infelicitous choice of coordinate frame of reference (for example, Coriolis forces), and all sorts of inertial reaction forces. (See Appendix A for a discussion of "fictitious" and gravitational forces.) None of the other forces conveyed by fields have this property. They are all mediated by the exchange of "transfer" particles, photons, gluons, and the like, that pass through spacetime to convey forces. But gravity can be accounted for by the warping of spacetime itself. If you are determined to believe that gravity is just another force, like all the others, you will likely want it to be mediated by "gravitons" that pass through spacetime. But no one has ever seen a graviton.

In no small part, much of the distaste for Einstein's version of the EP, the so-called Einstein Equivalence Principle (EEP), stems from the fact that it forbids the "localization" of gravitational energy (or, strictly speaking, energy-momentum). Gravity waves, considered over several wavelengths, are an exception to this prohibition. But that doesn't change the prohibition significantly. If a gravity field can be eliminated by a transformation to a suitably chosen accelerating frame of reference, or equivalently a suitable choice of geometry, then no energy can be associated with it locally, for in a sense it isn't really there in the first place. If accelerations, in themselves, conferred energy on objects being accelerated, the situation might be different. Why? Because then they might produce energy equivalent to that which would be produced by the action of a gravity field on its sources – local gravitational potential energy. But accelerations *per se* don't produce energy in this way. Accelerations are related to changes in motion and resulting changes in the energies of objects.[‡‡‡] But applying a force to an object in, say, the Earth's gravity field to keep it from engaging in free-fall, a steady force of one "g", does not change the energy of the object, at least after stationary conditions have been achieved. So, for gravity to mimic accelerations as the EP stipulates, localization of gravitational (potential) energy must be forbidden. [See Appendix B for Misner, Thorne, and Wheeler's comments (in their 1973 book, *Gravitation*) on localization of gravitational energy and the EP.]

Einstein's critics based their attacks on the fact that the EP is only strictly speaking true for a "uniform" gravitational field – that is, the gravitational field that would be produced by a plane, semi-infinite mass distribution, something that cannot exist in reality. For any realistic mass distribution, the field is not uniform, and the non-uniformity means that the EP is only approximately true in very small regions of spacetime. Indeed, they argue that no matter how small the region of spacetime under consideration is, "tidal" gravitational effects will be present and, in principle at least, measurable. Tidal effects, of course, are a consequence of the non-uniformity of the field, so arguably their presence in real systems cannot be said to invalidate the EP. But that's not what the critics are after. What they want is to assert that in reality, gravitation is just another field like all others.[§§§] Had his critics been successful, of course, Einstein's accomplishment would have been measurably diminished. Einstein stuck to his guns. An ideal uniform gravity field might be an unobtainable fiction in our reality, but it was clear to him that the E(E)P was correct notwithstanding the arguments of his critics. It is worth noting here that "idealization" has been used successfully in physics for hundreds of years to identify fundamental physical principles.

There is a very important point to be noted here. No matter how extreme the local distortion of spacetime produced by a local concentration of matter might be, the "constancy" of the speed of light at every point in spacetime remains true. That is, SRT is true at every point in spacetime in GRT. The extent of the spacetime around any given point where SRT is true may be infinitesimally small, but it is never of exactly zero extent. While SRT is true at every point – or, correctly "event" – in spacetime in GRT, the speed of light in GRT is no longer a "constant". That is, all observers no longer get the same number for the speed of light in vacuum. All *local* observers still get the same number. But when distant observers

---

[‡‡‡] Changes in internal energies of accelerating objects may take place if the objects are extended and not rigid. As we will see later, this complication leads to the prediction of interesting transient effects.

[§§§] I cannot resist mentioning here that Einstein's critics were (and are) quite happy to use unrealizable conditions when it suits their purposes in other situations. For example, they are quite content to assume that spacetime is Minkowskian at "asymptotic" infinity, or that spacetime in the absence of "matter" is globally Minkowskian. Actually, neither of these conditions can be realized. Their assumption is the merest speculation. Just because you can write down equations that model such conditions does not mean that reality actually is, or would be, that way. What we do know is that at cosmic scale, spacetime is spatially flat. And that condition corresponds to a mean "matter" density that while small, is not zero. In fact, in Friedmann-Robertson-Walker cosmologies (which are homogeneous and isotropic) spatial flatness results from the presence of "crititcal" cosmic "matter" (everything that gravitates) density – about 2×10^{-29} grams per cubic centimeter. That's about one electron per 50 cubic centimeters. Not very much stuff, to say the least.

measure the speed of light near a large local matter concentration, the speed they measure is less than the speed measured by the local observers. (Technically, distant observers measure the "coordinate" speed of light. It is not constant. The coordinate speed of light depends on the presence of gravity fields.)

This curious feature of GRT is especially obvious in the case of light from a distant source as it passes in the vicinity of a massive star. The light is deflected by the star because the star warps the spacetime in its neighborhood. But a distant observer doesn't see the spacetime around the star warp. After all, empty spacetime is perfectly transparent. What s/he sees is the light moving along a path that appears curved; a path that results from the light appearing to slow down the closer it gets to the star. If we are talking about some object with "restmass" – mass you can measure on a balance in its proper frame of reference – the path followed is also curved, though a bit differently as objects with finite restmass cannot reach the speed of light. These free-fall paths have a name: geodesics. They are found by solving Einstein's field equations of GRT for the particular distribution of sources of the local gravitational field. Because gravity warps spacetime so that things you measure depend on the direction in which you measure them, it turns out to be a tensor field – a more complicated thing than a scalar or vector field. Happily, tensor gravity has the property of symmetry, so several of the field components are paired, and only 10 components have independent values. To find 10 components you need 10 equations, which is messier than scalar or vector theories. Often, however, it is possible to simplify things either by choosing simple circumstances, or by making simplifying approximations, to reduce the messiness.

The most famous prediction of GRT is that of "black holes", or as they were known before John Wheeler gave them their catchy name, "frozen stars". These objects have all of their masses enclosed by their "event horizons". For a simple non-rotating spherical star, the radius of the event horizon, also sometimes called the "gravitational radius", is given by $R = 2GM / c^2$, where G is Newton's universal constant of gravitation, M the mass of the star, and c the speed of light in vacuum. As everyone now knows, the event horizon of a black hole is a surface of "no return". Should you have the misfortune to fall to the event horizon, you will inexorably be sucked into the hole – and spagettified by tidal forces too as you approach the singularity at the center of the hole where space and time cease to exist. Books have been written and movies made about black holes and the exploits of those in their vicinities. There is an important point about black holes, however, that sometimes doesn't get made exactly. For distant observers, time stops at the event horizon.

So what? Well, this means that for us distant observers, we can never see anything fall into a black hole. Everything that has ever fallen toward a black hole, for us, just seems to pile up at the surface that is the event horizon. It never falls through. That's why, pre-Wheeler, they were called frozen stars. But what about observers who fall into a black hole? Time doesn't stop for them, does it? No, it doesn't. Indeed, you fall through the event horizon as if there were nothing there at all to stop you. How can both stories be right? Well, as you fall towards the hole, the rate of time you detect for distant observers out in the universe far from the hole speeds up. And at the instant that you reach the event horizon, the rate of distant time becomes infinite and the whole history of the universe – whatever it may be – passes in that instant. So, an instant later when you are inside the event horizon, the exterior universe is gone. Even if you could go back (you can't), there is no back there to go to. For our purposes here, though, what is important is that for distant observers like us, the measured speed of light at the event horizon is zero, because, for us, time stops there.

This is true if the mass of the black hole is positive. Should the mass of the hole be negative, however, time at the gravitational radius measured by distant observers would speed up. Indeed, at the gravitational radius, it would be infinitely fast. This means that if the mass of the hole is "exotic", stuff near the gravitational radius can appear to us to travel much, much faster than the speed of light. This odd behavior in the vicinity of negative mass stars (should they even exist) doesn't have much direct value for rapid spacetime transport. After all, you wouldn't want to hang around an exotic black hole so that you could age greatly before returning to much the same time as you left. But it is crucial to the nature of matter as it bears on the construction of stargates. If the "bare" masses of elementary particles are exotic, they can appear to spin with surface velocities far in excess of the speed of light. And if a way can be found, using only “low” energy electromagnetic fields, to expose those bare masses, stargates may lie in our future. How all this works is dealt with in the last section of this book.

Now, all of this is very interesting, indeed, in some cases downright weird. We normally don’t think of space and time as deformable entities. Nonetheless, they simply are. And the drama of reality plays itself out in space and time that are locally uninfluenced, beyond the effects predicted by GRT, by the action taking place within them. The thing that distorts space and time, or more accurately, spacetime in GRT is mass-energy. How the distortion occurs can be constructed with the Equivalence Principle and the principle of general covariance. The principle of general covariance is the proposition that all physical laws should have the same form in all frames of reference – inertial or accelerated. Einstein noted early on.that he was not happy about this as he thought the distribution of matter and its motions throughout the universe should account for inertia and thus be essential to a correct description of reality. The physical reason why this *must* be the case rests on Mach’s principle, as Einstein suspected. How this works involves subtleties that have made Mach’s principle a topic of contention and confusion literally from the time Einstein introduced it to the present day. We now turn to Mach’s principle, for getting it right is essential to our purpose: making starships and stargates.

APPENDIX A

Excerpt from Adler, Bazin, and Schiffer, *Introduction to General Relativity, pp. 57-59*:

Lagrange's equations in the light of general relativity theory. Suppose one forgets about the physical origin of the generalized coordinates x^i and sees the equations of motion written in the form

$$\ddot{x}^i + \left\{ \begin{matrix} i \\ k\ l \end{matrix} \right\} \dot{x}^k \dot{x}^l = F^i \tag{2.23}$$

When thinking in terms of Newtonian mechanics, one would like to see these equations take the form $\ddot{x}^i = \tilde{F}^i$, where $\tilde{F}^i$ represents external forces according to Newton's law. To be able to make this identification one considers the quantities

$$- \left\{ \begin{matrix} i \\ k\ l \end{matrix} \right\} \dot{x}^k \dot{x}^l$$

as representing fictitious forces (such as centrifugal and Coriolis forces); these visibly depend on the coordinate system used (through the Christoffel symbols) and are often said to appear because one uses the "wrong kind" of coordinate system, for instance, a system attached to a rotating disk. A "right kind" of coordinate system is of course one in which these fictitious forces simply do not appear. However, one can use the alternative approach of treating all forces equally, be they external, fictitious, or due to a constraint, and accordingly write

$$\ddot{x}^i = \tilde{F}^i \qquad \tilde{F}^i = F^i - \left\{ \begin{matrix} i \\ k\ l \end{matrix} \right\} \dot{x}^k \dot{x}^l \tag{2.24}$$

Clearly, the combined force depends very much on the coordinate system.

Obviously, such a viewpoint is not in the spirit of general relativity theory where all kinds of coordinate systems are considered equivalent. From the viewpoint of general relativity, one would instead like to reduce the equations of motion as much as possible to the *geometry* of the configuration space. That is, instead of explaining away wrong geometries by fictitious forces, we should like to explain away forces by proper choices of geometry. This will be possible at least in the case of gravitational forces. The easiest way to do this is to postulate that the gravitational forces F^i can be made to disappear from the above equations of motion by incorporating them into the geometric term $\begin{Bmatrix} i \\ k\ l \end{Bmatrix} \dot{x}^k \dot{x}^l$ just like a fictitious force. This approach is motivated by the fact that gravitational and fictitious forces both act on material bodies in the same way; they communicate an acceleration $\ddot{x}^i$ which is *independent of the body's mass.* (This is not the case for other types of forces; for instance, the acceleration communicated to a body by a spring is inversely proportional to the mass of the body.) The above property is the basis of the *principle of equivalence*, which states that the effect of a gravitational field can be "reproduced" by describing physics in an appropriately accelerated frame of reference without interior gravitational forces present. In such a frame of reference the generalized coordinates will be some y^j (which can be considered functions of x^i and t), and the kinetic energy of the system will be described by a new function $\bar{T}$. In order to bring in the principle of equivalence and incorporate all gravitational forces in the geometric term, one would like the Lagrange equations in the moving frame (with coordinates $y^{\cdot}$ and kinetic energy function $\bar{T}$) to take the form

$$\ddot{y}^j = -\begin{Bmatrix} j \\ k\ l \end{Bmatrix} \dot{y}^k \dot{y}^l \tag{2.25}$$

which are the equations of configuration-space geodesics in the moving frame.

Unfortunately, we can easily show that such an attempt to incorporate the principle of equivalence cannot succeed within the framework of classical mechanics: consider the concrete case of a particle moving under the influence of gravity along a three-dimensional trajectory described by $y^j(t)$ $(j = 1, 2, 3)$ in a moving frame of reference. If Eqs. (2.25) were valid, the acceleration $\ddot{y}^j$ of the particle in that frame of reference would depend quadratically on the velocity $\dot{y}^j$ of the particle; doubling the velocity of a particle submitted to a gravitational field would therefore quadruple its acceleration. We know from experience that the movement of a particle in a gravitational field does not obey such a law in any frame of reference. Equations (2.25) are therefore unacceptable to describe the effects of gravitational forces, and *it is impossible to have gravitational forces take the same mathematical form as fictitious forces within the framework of classical analytical mechanics.*

We shall be able to formulate the principle of equivalence in mathematical terms only when we consider *Euler-Lagrange equations in a four-*

dimensional space which includes time as an ordinary coordinate; in order to consider the solution of gravitational problems by a purely geometrical treatment, it will be necessary to make use of the concepts of special relativity theory and the Lorentz metric. In fact, in a four-dimensional space for which the zeroth coordinate is ct (time multiplied by the speed of light), Eqs. (2.25) are acceptable. When the velocities involved in the problem are small compared with the speed of light, we have

$$\dot{y}^0 = \frac{d}{dt}(ct) = c$$

$$\dot{y}^k \ll c \qquad k = 1, 2, 3$$

Thus Eqs. (2.25) reduce in lowest order to

$$\ddot{y}^j = -\left\{ {j \atop 0\ 0} \right\} c^2 \qquad j = 1, 2, 3 \tag{2.26}$$

The terms quadratic in velocity, which prevented us from making any progress in a three-dimensional framework, do not appear in these lower-order equations in a four-dimensional framework. The only term which survives is the constant c^2. Furthermore, we see that the Christoffel symbols (which are geometric entities) here play the role of forces. These considerations will be taken up in greater detail in Chap. 4.

APPENDIX B

Misner, Thorne, and Wheeler's discussion of localization of gravitational energy in their comprehensive textbook, *Gravitation* (Freeman, San Francisco, 1973):

§20.4. WHY THE ENERGY OF THE GRAVITATIONAL FIELD CANNOT BE LOCALIZED

Consider an element of 3-volume $d\Sigma_\nu$ and evaluate the contribution of the "gravitational field" in that element of 3-volume to the energy-momentum 4-vector, using in the calculation either the pseudotensor $t^{\mu\nu}$ or the pseudotensor $t^{\mu\nu}_{\text{L-L}}$ discussed in the last section. Thereby obtain

$$\boldsymbol{p} = \boldsymbol{e}_\mu t^{\mu\nu}\, d\Sigma_\nu$$

or

$$\boldsymbol{p} = \boldsymbol{e}_\mu t^{\mu\nu}_{\text{L-L}}\, d\Sigma_\nu.$$

Right? No, the question is wrong. The motivation is wrong. The result is wrong. The idea is wrong.

To ask for the amount of electromagnetic energy and momentum in an element of 3-volume makes sense. First, there is one and only one formula for this quantity. Second, and more important, this energy-momentum in principle "has weight." It curves space. It serves as a source term on the righthand side of Einstein's field equations. It produces a relative geodesic deviation of two nearby world lines that pass through the region of space in question. It is observable. Not one of these properties does "local gravitational energy-momentum" possess. There is no unique formula for it, but a multitude of quite distinct formulas. The two cited are only two among an infinity. Moreover, "local gravitational energy-momentum" has no weight. It does not curve space. It does not serve as a source term on the righthand side of Einstein's field equations. It does not produce any relative geodesic deviation of two nearby world lines that pass through the region of space in question. It is not observable.

Why one cannot define a localized energy-momentum for the gravitational field

Anybody who looks for a magic formula for "local gravitational energy-momentum" is looking for the right answer to the wrong question. Unhappily, enormous time and effort were devoted in the past to trying to "answer this question" before investigators realized the futility of the enterprise. Toward the end, above all mathematical arguments, one came to appreciate the quiet but rock-like strength of Einstein's equivalence principle. One can always find in any given locality a frame of reference in which all local "gravitational fields" (all Christoffel symbols; all $\Gamma^\alpha{}_{\mu\nu}$) disappear. No Γ's means no "gravitational field" and no local gravitational field means no "local gravitational energy-momentum."

Nobody can deny or wants to deny that gravitational forces make a contribution to the mass-energy of a gravitationally interacting system. The mass-energy of the Earth-moon system is less than the mass-energy that the system would have if the two objects were at infinite separation. The mass-energy of a neutron star is less than the mass-energy of the same number of baryons at infinite separation. Surrounding a region of empty space where there is a concentration of gravitational waves, there is a net attraction, betokening a positive net mass-energy in that region of space (see Chapter 35). At issue is not the existence of gravitational energy, but the localizability of gravitational energy. It is not localizable. The equivalence principle forbids.

Look at an old-fashioned potato, replete with warts and bumps. With an orange marking pen, mark on it a "North Pole" and an "equator". The length of the equator is very far from being equal to 2π times the distance from the North Pole to the equator. The explanation, "curvature," is simple, just as the explanation, "gravitation", for the deficit in mass of the earth-moon system (or deficit for the neutron star, or surplus for the region of space occupied by the gravitational waves) is simple. Yet it is not possible to ascribe the deficit in the length of the equator in the one case, or in mass in the other case, in any uniquely right way to different elements of the manifold (2-dimensional in the one case, 3-dimensional in the other). Look at a small region on the surface of the potato. The geometry there is locally flat. Look at any small region of space in any of the three gravitating systems. In an appropriate coordinate system it is free of gravitational field. The over-all effect one is looking at is a global effect, not a local effect. That is what the mathematics cries out. That is the lesson of the nonuniqueness of the $t^{\mu\nu}$!