# GRAVITY & ELECTROMAGNETISM

YARON HADAD

## 1. Introduction

Thomas Kuhn once argued[6] that unlike a scientist, “a student in the humanities has constantly before him a number of competing and incommensurable solutions to these problems, solutions that he must ultimately examine for himself.” While a critic in the humanities, for example, can choose to adopt an old theory of poetics to explain his ideas, a scientist does not have the luxury to use an old dismissed theory in his research. So in the presence of two competing scientific theories that both agree with current experiments, what theory should we choose? What theory is 'better'? Or even - which theory is the 'correct' description of nature?

A general rule of finger for such cases is to apply Occam's razor, attributed to the 14th-century English logician, William of Occam. In its modern form, this principle states that in order to distinguish between two equally explanatory theories, one should choose the simplest one. For both practical and aesthetical reasons, picking the simplest theory will make our life easier as it makes the explanation easier to follow and work with, and sometimes make the natural world seem more appealing. What is it that makes us expect nature to be simple, beautiful and aesthetically appealing is a whole different matter which we will not address here, but for practical reasons having our theory as simple as possible is essential.

There is another generally accepted rule for distinguishing between two equally explanatory theories: Unification. We not only want our theory to be as simple as it gets, we also want it to account for as many phenomena as possible. Different phenomena in nature often seem to be closely related, so why for example, should we expect to have a different theory for the microcosmos than the one we have for the macrocosmos? We would like a theory that explains the universe as a whole!

There have been many attempts to unite the concepts in physics into one consistent theoretical framework. In the 17th century Isaac Newton[9] explained the motion of all heavenly bodies in our solar system, but also cannonballs on earth using his law of universal gravitation. Later on in the 19th century, James Clerk Maxwell[7] completed the development of his theory of electromagnetism, which united the electric and magnetic fields into one entity - called the electromagnetic field. Maxwell even demonstrated that light is a manifestation of such an electromagnetic field. This fact that was later confirmed experimentally by Heinrich Hertz in 1887, essentially made optics into a subfield of the electromagnetic theory. In 1905, Albert Einstein[2] fully explained the unity of electricity and magnetism into one field. Hermann Minkowski[8] showed that it can be given a simple mathematical description if we unite space and time into a single entity, which we now call spacetime. There have been many other attempts to unite other theories and seemingly distinct concepts in physics into one, single idea.

The main goal of this paper is to review the relationships, the differences and some of the attempts to unite the most stubborn of all physical concepts: gravity and her younger brother the electromagnetic field. A secondary goal of this paper is to convince the reader that despite a few apparent differences between the two fields, the remarkable resemblance between them that appears in so many different forms must mean that they are the manifestation of the same notion. Though in the present time this unification has not yet taken place, we can only hope that it is inevitable.

## 2. A first evidence

In the following section we shall consider the first mathematical description of the gravitational field and of the electric field, leading to several preliminary resemblances and differences.

In Newton's work Philosophiae Naturalis Principia Mathematica[9], Newton's law of universal gravitation is stated as follows:

*"Every particle of matter in the universe attracts every other particle with a force that is directly proportional to the product of the masses of the particles and inversely proportional to the square of the distance between them."*

We may state the law mathematically as[1]:

$$\vec{F}_g = -G\frac{Mm}{|\vec{r}_{12}|^2}\hat{r}_{12}$$

where $\vec{F}_g$ is the gravitational force exerted on body #1 due to body #2, $G$ is the gravitational constant, $m$ and $M$ are the masses of the first and second body respectively, and $\vec{r}_{12}$ is a vector pointing from the second to the first body. The French physicist Charles Augustin de Coulomb developed an analog of Newton's law of gravity for the electric force, in a law that bares his name. Coulomb's law for the electrostatic force may be stated as:

*"The magnitude of the electrostatic force between two point electric charges is directly proportional to the product of the magnitudes of each charge and inversely proportional to the square of the distance between the charges."*

This law may be stated mathematically as:

$$\vec{F}_e = K\frac{Qq}{|\vec{r}_{12}|^2}\hat{r}_{12}$$

where $\vec{F}_e$ is the electrostatic force exerted on body #1 due to body #2, $K$ is the electrostatic constant, $q$ and $Q$ are the charges of the first and second body respectively, and $\vec{r}_{12}$ is a vector pointing from the second to the first body.

Before discussing the properties of these two laws, I would like to tell about my first encounter with Coulomb's law. After studying Newton's law of gravity for a few months in high-school, we started discussing electromagnetism, starting of course with Coulomb's law. When my teacher wrote Coulomb's law on the blackboard for the first time, I remember getting utterly excited about the striking resemblance between Coulomb's law to Newton's law, so I announced to the entire class "Wow! It is exactly the same as Newton's law!". Unfortunately my teacher was less enthusiastic about it and completely ignored my remark...

[1]Throughout the paper, we use SI units.

So is there any connection between the two, other that the apparent similarity between the two equations? Even though that neither Newton's law nor Coulomb's law are considered to be a complete mathematical description of the gravitational and electric forces, they provide us with a very good approximation of nature in the case of slow velocities relatively to the speed of light. Mathematically, Newton's law and Coulomb's law are almost completely identical. Both forces are inversely proportional to the distance between the two bodies, a fact that can be explained as a conservation of flux, as the surface area of a sphere in three-dimensional space being $4\pi r^2$. Both forces are proportional to the property of the body that affects that specific force: the gravitational force is proportional to the product of the masses while the electrostatic force is proportional to the product of the electric charges. But these fascinating similarities also reveal a few fundamental differences between the two forces. Since the constants in the two forces, $G$ and $K$ have been observed to be positive, the appearance of a minus sign on the gravitational force has a fundamental affect of its behavior. Since the mass of a body is conjectured to be always positive, the gravitational force always attracts. On the other hand, the electrostatic force might attract or repel, depending on the charges at hand. For the electrostatic force, like charges repel each other (the opposite of the gravitational force) while opposite charges attract each other.

Keeping in mind Newton's second law $\vec{F} = m\vec{a}$ reveals an even more essential difference between the two forces. Since the mass of the body appears on the right-hand-side of Newton's law, the acceleration due to the gravitational force is:

$$\vec{a}_g = -G\frac{M}{|\vec{r}_{12}|^2}\hat{r}_{12}$$

while the acceleration due to the electrostatic force is:

$$\vec{a}_e = K\left(\frac{q}{m}\right)\frac{Q}{|\vec{r}_{12}|^2}\hat{r}_{12}$$

We can see that the mass of the first body does not affect its motion in a gravitational field, a fact that plays an important role in Einstein's equivalence principle and consequently on the theory of general relativity. The acceleration due to the electrostatic force, on the other hand, does depend on the mass of the first body and also on its charge. In fact, it only depends on the ratio between its charge and its mass, namely $\frac{q}{m}$. Considering the relative strength of the two forces might dim the early enthusiasm about their similarities. We consider the forces exerted on an electron due to the presence of a proton. Computing the ratio between the magnitudes of the two forces yields:

$$\begin{aligned}
\frac{F_e}{F_g} &= \frac{Ke^2}{Gm_em_p} \\
&= \frac{9\times 10^9\frac{Nm^2}{C^2}\cdot\left(1.6\times 10^{-19}C\right)^2}{6.67\times 10^{-11}\frac{m^3}{kg\cdot s^2}\left(9.1\times 10^{-31}kg\right)\left(1.67\times 10^{-27}kg\right)} \\
&= 2.27\times 10^{39}
\end{aligned}$$

so the electrostatic force between an electron and a proton is 39 order of magnitudes stronger than the gravitational force between the two! The gravitational force is nearly irrelevant if you consider the motion of two charged particles!

Yet still, though the three differences we discussed between gravity and the electrostatic force seem to make the two be completely distinct, in the following sections we will see how the development of Maxwell's equations and later on the theory of general relativity strengthen the first impression - there should be some deep intrinsic connection between the two forces.

## 3. Maxwell's Equations

Since the introduction of Coulomb's law, our understanding of the electric force changed dramatically in less than a century. Faraday, Henry, Helmholtz, Ampère and others revealed a strong connection between the electric and the magnetic forces through experimental evidence, and deduced the mutual effect of the two forces on each other. This led Faraday to coin the term 'field', which today is considered to be a more fundamental concept than that of a 'force'. Fields do not only dictate the motion of bodies, but they also have independent physical reality since they carry energy and may interact with each other. These ideas led to the creation of the first unified field theory in physics, when James Clerk Maxwell[7] corrected Ampère's law and provided four equations that together with the Lorentz force law, produce a complete description of all electromagnetic phenomena known in his time. Maxwell's equations may be written as:

$$\begin{aligned} \nabla \cdot \vec{E} &= \frac{\rho}{\varepsilon_0} \\ \nabla \cdot \vec{B} &= 0 \\ \nabla \times \vec{E} + \frac{\partial \vec{B}}{\partial t} &= 0 \\ \nabla \times \vec{B} - \mu_0 \varepsilon_0 \frac{\partial \vec{E}}{\partial t} &= \mu_0 \vec{J} \end{aligned}$$

where $\vec{E}$ and $\vec{B}$ are the electric field and the magnetic field respectively, $\rho$ is the total charge density, $\vec{J}$ is the total current density, and $\varepsilon_0$ and $\mu_0$ are the electric and the magnetic constants, respectively. The Lorentz force law gives the force $\vec{F}$ exerted on a point charge $q$ due to the electromagnetic fields, and is given by:

$$\vec{F} = q\left(\vec{E} + \vec{v} \times \vec{B}\right)$$

where $\vec{v}$ is the velocity of the particle.

The first pair of Maxwell's equations are Gauss's law (equivalently, Coulomb's law) and Gauss's law for magnetism. The second pair of Maxwell's equations are Faraday's law of induction and Ampère's law which give the evolution of the electric and magnetic fields in time. One should notice that even though Maxwell's equations are a set of eight scalar equations, the electromagnetic field has only six components ($\vec{E}$ and $\vec{B}$). The reason for that is that the two Gauss's laws give a restriction on the plausible physical fields, and not really determine their evolution.

Even though the electric field and the magnetic field seemed to be so strongly related, they were considered to be two separate fields in the time of Maxwell. Albert Einstein's work[2] in 1905 on the special theory of relativity revealed that this is merely a misconception. Einstein postulated that the speed of light in vacuum is fixed and is independent of the relative velocity between any (inertial) observer and the speed of the light source. This postulate led to a derivation of the Lorentz transformation and concluded that the electric and magnetic fields are in fact the same field, perceived differently by different observers depending on their relative velocity. This justifies the term: the 'electromagnetic field', instead of two separate 'electric field' and 'magnetic field'.

In the next section we will describe general relativity. We will see that just as Newton's law and Coulomb's law were closely analogous, Einstein's field equation is very similar to Maxwell's equations in the right limit. This will reveal another similarity between the electromagnetic field and the gravitational field.

## 4. General Relativity

4.1. **Introduction.** After Albert Einstein finished his special theory of relativity, which united the electric and magnetic fields into a single 'electromagnetic' field, he considered the force of gravity. Newton's theory of gravitation is not consistent with special relativity since it invokes instantaneous influence of one body on another. This motivated Einstein to seek a new theory of gravity. However, instead of modifying Newton's theory of gravitation and making it compatible with special relativity, Einstein decided to follow a completely different path.

The key idea is that all bodies are influenced by gravity. In fact, all bodies are influenced by gravity in precisely the same way! This notion, known as the equivalence principle is manifested in Newton's theory of gravitation by the statement that the gravitational force on a body is proportional to its inertial mass. Therefore, instead of describing gravity as a force 'acting at a distance', a notion that even Newton himself found to be unsatisfactory, the equivalence principle suggests that the force of gravity might have to do with the structure of spacetime itself. This led Einstein to dismiss the the notion of a 'gravitational field', and instead, his theory of general relativity postulates[3] that:

(1) Spacetime curves in the presence of matter according to Einstein's field equation,
$$G_{ab} = \frac{8\pi G}{c^4} T_{ab}$$
where $G_{ab} = R_{ab} - \frac{1}{2} g_{ab} R$ is the Einstein tensor; $R_{ab}$ is the Ricci curvature tensor, $R$ is the scalar curvature, $g_{ab}$ is the metric of spacetime and $T_{ab}$ is the stress-energy tensor that represent the distribution of matter and energy in spacetime.
(2) In the absence of any forces (remember, gravity is not considered to be a force anymore!) matter travels along geodesics. Geodesics are the curved-space analogs of straight lines.

It should be stressed that even though general relativity revolutionized our notions of space, time and gravity, in some sense it actually got us further away from a real unification of electromagnetism and gravity. Before general relativity, both gravity and the electromagnetic field were considered as two fields existing independently of the structure of space and time, however, general relativity separates the two notions. Gravity is not a force anymore, but merely a result of the curvature of spacetime in the presence of matter. On the other hand, electromagnetism is still a field, separated from the structure of spacetime. We will come back to this issue when we will discuss Kaluza's theory of gravity and electromagnetism.

4.2. **'Weak' gravity.** Next we consider the approximation in which gravity is "weak." This means that the spacetime metric is nearly flat. In practice, this is a pretty good approximation in nature except for phenomena dealing with very strong gravitational fields (e.g. black holes) or dealing with the large scale structure of the universe.

We write the spacetime metric as

$$g_{ab} = \eta_{ab} + h_{ab}$$

where $\eta_{ab} = diag\,(-1,1,1,1)$ is the Minkowski 'flat' metric and the deviation from flat spacetime is "small." This means that the components of $h_{ab}$ are much smaller than 1. Though in the following analysis we will keep writing equal signs, we will only be interested in the first order approximation of the metric, and thus will only retain terms that are linear in $h_{ab}$. Furthermore, as a convention, we will raise indices with the help of the flat Minkowski metric $\eta_{ab}$ with one exception: the inverse of the metric $g_{ab}$ will still be denoted with $g^{ab}$. Thus, up to the first order in $h_{ab}$, the inverse metric is:

$$g^{ab} = \eta^{ab} - h^{ab}$$

Since $\eta_{ab}$ is constant, the Christoffel symbol (to the first order) is:

$$\begin{aligned}
\Gamma^c_{ab} &= \frac{1}{2} g^{cd} \left(\partial_a g_{bd} + \partial_b g_{ad} - \partial_d g_{ab}\right) \\
&= \frac{1}{2}\left(\eta^{cd} - h^{cd}\right)\left(\partial_a h_{bd} + \partial_b h_{ad} - \partial_d h_{ab}\right) \\
&= \frac{1}{2}\eta^{cd}\left(\partial_a h_{bd} + \partial_b h_{ad} - \partial_d h_{ab}\right)
\end{aligned}$$

Since the Christoffel symbol is already first order in $h$, the Ricci tensor (to the first order in $h$) is:

$$\begin{aligned}
R_{ab} &= \partial_c \Gamma^c_{ab} - \partial_a \Gamma^c_{cb} + \Gamma^c_{ab}\Gamma^d_{cd} - \Gamma^c_{ad}\Gamma^d_{bc} \\
&= \partial_c \Gamma^c_{ab} - \partial_a \Gamma^c_{cb} \\
&= \frac{1}{2}\eta^{cd}\partial_c\left(\partial_a h_{bd} + \partial_b h_{ad} - \partial_d h_{ab}\right) - \frac{1}{2}\eta^{cd}\partial_a\left(\partial_c h_{bd} + \partial_b h_{cd} - \partial_d h_{cb}\right) \\
&= \frac{1}{2}\partial^d\partial_b h_{ad} - \frac{1}{2}\partial^d\partial_d h_{ab} - \frac{1}{2}\partial_a\partial_b h^d_d + \frac{1}{2}\partial_a\partial^c h_{cb} \\
&= \frac{1}{2}\left[-\Box h_{ab} + \partial_b\left(\partial^c h_{ac} - \frac{1}{2}\partial_a h\right) + \partial_a\left(\partial^c h_{bc} - \frac{1}{2}\partial_b h\right)\right]
\end{aligned}$$

where we denoted the trace of $h_{ab}$ by $h = h^d_d$, and the wave operator (the d'Alembertian) by

$$\Box = -\frac{1}{c^2}\frac{\partial^2}{\partial t^2} + \nabla^2$$

It is easy to see that if we define

$$\xi_a = \partial^c h_{ac} - \frac{1}{2}\partial_a h$$

then the Ricci tensor is

$$R_{ab} = \frac{1}{2}\left[-\Box h_{ab} + \partial_b \xi_a + \partial_a \xi_b\right]$$

The scalar curvature is

$$\begin{aligned}
R &= R^a_a \\
&= -\frac{1}{2}\Box h + \partial^a \xi_a
\end{aligned}$$

and therefore, the Einstein tensor is

$$\begin{aligned} G_{ab} &= R_{ab} - \frac{1}{2} g_{ab} R \\ &= R_{ab} - \frac{1}{2} \eta_{ab} R \\ &= \frac{1}{2} \left[ -\Box h_{ab} + \partial_b \xi_a + \partial_a \xi_b \right] - \frac{1}{2} \left[ -\frac{1}{2} \Box h + \partial^a \xi_a \right] \\ &= -\frac{1}{2} \Box \left( h_{ab} - \frac{1}{2} h \right) + \frac{1}{2} \left( \partial_b \xi_a + \partial_a \xi_b - \partial^a \xi_a \right) \end{aligned}$$

Since the geometry of spacetime is completely determined by the metric $g_{ab}$, and for any diffeomorphism $\phi$ the metrics $g_{ab}$ and $\phi^* g_{ab}$ represent the same geometry, there is a gauge freedom in general relativity. It can be shown[10] that to the first order in $h_{ab}$, any transformation of the form

$$h_{ab} \to h_{ab} + \partial_a \xi_b + \partial_b \xi_a$$

will leave the geometry unchanged. Using this gauge symmetry, we may choose our coordinates such that

$$\xi_a = 0$$

so that the Einstein tensor simplifies to:

$$G_{ab} = -\frac{1}{2} \Box \left( h_{ab} - \frac{1}{2} h \right)$$

Next we define $\bar{h}_{ab} = h_{ab} - \frac{1}{2} \eta_{ab} h$. Componentwise, the non-diagonal terms of $\bar{h}_{ab}$ and $h_{ab}$ are exactly the same but the trace is reversed, i.e. $tr\left(\bar{h}\right) = -tr\left(h\right)$. Finally, the linearized Einstein equation is:

$$\Box \bar{h}_{ab} = -\frac{16\pi G}{c^4} T_{ab}$$

Notice that this is nothing other than the wave equation for the coefficients $\bar{h}_{ab}$ with a source of $-\frac{16\pi G}{c^4} T_{ab}$. This already reveals a strong similarity between Einstein's equation and Maxwell's equations, namely, that in the limit of a 'weak' gravitational field Einstein equation is nothing other than a wave equation for a wave propagating in the speed of light. Next we will see that if one considers the geodesic equation, then this similarity is even stronger than what it appears to be.

4.3. **Maxwell again?** The stress-energy tensor for a perfect fluid with no pressure is

$$T_{ab} = \rho u_a u_b$$

where $\rho$ is the mass density and $u_a$ is its 4-velocity. If we consider the linearized Einstein equation for non-relativistic speeds ($v \ll c$), then up to the first order in $\frac{v}{c}$, the stress energy tensor becomes

$$T_{ab} = \begin{bmatrix} \rho c^2 & \rho c \vec{u} \\ \rho c \vec{u} & 0 \end{bmatrix}$$

We denote $\phi = -\frac{\bar{h}_{00}}{4}c^2$ and $\vec{A} = c^2\left(\bar{h}_{01}, \bar{h}_{02}, \bar{h}_{03}\right)$ (this notation will become clear immediately) and the linearized Einstein equation becomes:

$$\begin{aligned} \Box\phi &= 4\pi G\rho \\ \Box\vec{A} &= -16\pi G\rho\frac{\vec{u}}{c} \\ \Box\bar{h}_{\mu\nu} &= 0 \end{aligned}$$

where $\mu, \nu = 1, 2, 3$, a convention that we'll keep using in this section. The unique solution to the third equation that is well behaved at infinity is $\bar{h}_{\mu\nu} = 0$. (The solution $\bar{h}_{\mu\nu} = \textit{constant}$ is also valid, but it can be eliminated by another gauge transformation.) In order to obtain the first-order correction to the metric, namely $h_{ab}$, we reverse the trace again to obtain:

$$h_{ab} = \bar{h}_{ab} - \frac{1}{2}\eta_{ab}\bar{h}$$

Since $\bar{h} = \frac{4\phi}{c^2}$, we have

$$\begin{aligned} h_{00} &= -\frac{4\phi}{c^2} - \frac{1}{2}(-1)\frac{4\phi}{c^2} = -\frac{2\phi}{c^2} \\ h_{0\nu} &= \bar{h}_{0\nu} = \frac{\vec{A}}{c^2} \\ h_{ab} &= -\frac{1}{2}\eta_{ab}\frac{4\phi}{c^2} = -2\eta_{ab}\frac{\phi}{c^2} \end{aligned}$$

So to the first order, the metric is

$$g_{ab} = \begin{bmatrix} -1 - \frac{2\phi}{c^2} & \frac{A_x}{c^2} & \frac{A_y}{c^2} & \frac{A_z}{c^2} \\ \frac{A_x}{c^2} & 1 - \frac{2\phi}{c^2} & 0 & 0 \\ \frac{A_y}{c^2} & 0 & 1 - \frac{2\phi}{c^2} & 0 \\ \frac{A_z}{c^2} & 0 & 0 & 1 - \frac{2\phi}{c^2} \end{bmatrix}$$

It is worthwhile to notice that for $\vec{A} = 0$, this is the same metric that leads to Newtonian gravity. In order to interpret this result, we will have to consider the motion of a body in such a gravitational field. In general relativity, bodies travel along a geodesic with respect to proper time. The geodesic equation is

$$\frac{d^2x^a}{d\tau^2} + \Gamma^a_{bc}\frac{dx^b}{d\tau}\frac{dx^c}{d\tau} = 0$$

Since we compute everything to the first order in $\frac{v}{c}$, $\frac{dx^a}{d\tau} = \gamma(c, v) \approx (c, v)$. Thus, in this approximation $\frac{d}{d\tau} \approx \frac{d}{dt}$, and the geodesic equation is:

$$\frac{d^2x^\mu}{dt^2} = -\Gamma^\mu_{00}c^2 - 2\Gamma^\mu_{0\nu}cu^\nu$$

where we have dropped second-order terms. Computing the Christoffel symbols, we see that (no summation over $\mu$ and $\nu$ here):

$$\begin{aligned} \Gamma^\mu_{00} &= \frac{1}{c^2}\left(\frac{\partial\phi}{\partial x^\mu} - \frac{\partial A_\mu}{c\partial t}\right) \\ \Gamma^\mu_{0\mu} &= -\frac{\partial\phi}{c^3\partial t} \\ \Gamma^\mu_{0\nu} &= \frac{1}{2c^2}\left(\partial_\nu A_\mu - \partial_\mu A_\nu\right) \end{aligned}$$

so the geodesic equation yields:

$$\begin{aligned}\frac{d^2\vec{x}}{dt^2} &= -\nabla\phi - \frac{\partial\vec{A}}{c\partial t} + \frac{\vec{v}}{c}\times\left(\nabla\times\vec{A}\right) + 2\frac{\partial\phi}{c\partial t}\frac{\vec{v}}{c} \\ &\approx -\nabla\phi - \frac{\partial\vec{A}}{c\partial t} + \frac{\vec{v}}{c}\times\left(\nabla\times\vec{A}\right)\end{aligned}$$

Remembering the gauge condition $\xi_a = \partial^c h_{ac} - \frac{1}{2}\partial_a h = 0$, we have (for $a = 0$):

$$\frac{1}{c}\frac{\partial\phi}{\partial t} + \frac{1}{2}\nabla\cdot\vec{A} = 0$$

If we denote $\vec{E}_g = -\nabla\phi - \frac{1}{2c}\frac{\partial\vec{A}}{\partial t}$ and $\vec{B}_g = \nabla\times\vec{A}$, corresponding to the gravitational field $\vec{E}_g$ and the *gravitomagnetic field* $\vec{B}_g$, then the gauge condition and the linearized Einstein equation gives:

$$\begin{aligned}\nabla\cdot\vec{E}_g &= -4\pi G\rho \\ \nabla\cdot\vec{B}_g &= 0 \\ \nabla\times\vec{E}_g + \frac{1}{2c}\frac{\partial\vec{B}_g}{\partial t} &= 0 \\ \nabla\times\left(\frac{1}{2}\vec{B}_g\right) - \frac{1}{c}\frac{\partial\vec{E}_g}{\partial t} &= -4\pi\frac{G}{c}\vec{J}\end{aligned}$$

where $\vec{J}$ is the mass current density, $\vec{J} = \rho\vec{v}$. These equations are almost completely identical to Maxwell's equations. Namely, the gravitational field $\vec{E}_g$ and the gravitomagnetic field $\vec{B}_g$ evolve just as the electric field $\vec{E}$ and the magnetic field $\vec{B}$. Therefore, both fields A body will travel along a trajectory $\vec{x}$ that satisfies:

$$m\frac{d^2\vec{x}}{dt^2} = m\left(\vec{E}_g + \frac{\vec{v}}{c}\times\vec{B}_g\right)$$

Once again we see a remarkable similarity between the mathematical law that governs the gravitational force and the law that governs electromagnetism. Both gravity and electromagnetism satisfy similar evolution equations. Both fields will radiate, and the speed of propagation is the speed of light (this is just a result of the wave equation). In both cases, if the magnetic part of the field doesn't vary with time, the law reduces to either Newton's law in the case of gravity or Coulomb's law in the case of electromagnetism. This shows that the resemblance between Newton's law and Coulomb's law was not a coincidence. Just as in the case of the electromagnetic force, the force acting on a body due to a gravitational field has the same form as the Lorentz force, so bodies will travel in a similar way under the influence of electromagnetic force and a gravitomagnetic force (with the one exception, the affect of a negative charge on the trajectory.)

However, this reveals a few fundamental differences. The three differences we mentioned when we discussed Newton's law and Coulomb's law remain in the case of general relativity and Maxwell's equations. Moreover, the process by which we derived the equation reveals a few more differences. The gravitational Maxwell equations are *only an approximation* for the true nature of gravity according to general relativity. General relativity is a tensor theory, but electromagnetism is described by vector fields. This fact can be used to show[3] that the leading term that causes gravitational radiation is quadrupolar, while in Maxwell's theory the leading term that causes electromagnetic radiation is a dipole. The tensorial nature of general relativity is also the cause for some additional factors of 2 in the linearized equations that do not exist in Maxwell's equations. These are merely remnants because the metric is a tensor of rank 2. And of course, Einstein's field equation is very nonlinear, while Maxwell's equations are linear.

## 5. Kaluza-Klein Theory

5.1. **Enter Kaluza.** As mentioned in the previous section, Einstein's theory of relativity treats gravity as the curvature of spacetime, but electromagnetism is still treated as a field similarly to its description in Maxwell's theory, separated from the geometry of spacetime. This fact, together with the similar structure of the linearized Einstein equation and Maxwell's equations, led the German physicist and mathematician Theodor Kaluza to try and unify the two fields by introducing extra dimensions. His key idea was that just as gravity is treated as the manifestation of curved spacetime, the electromagnetic force should be treated similarly. He noticed that since the electromagnetic potential has four independent components, the scalar potential and the vector potential $\vec{A}$, they can both be accommodated in the metric if one adds exactly one more spatial dimension to the theory (since then the metric has 15 independent components, 10 belong to the four dimensional metric and one is a diagonal term, altogether $15 - 10 - 1 = 4$ components).

Kaluza's theory is based upon three postulates[4]:

(1) Spacetime is five dimensional and the metric satisfies the five-dimensional vacuum Einstein field equation:

$$\hat{G}_{AB} = 0$$

where $\hat{G}_{AB} = \hat{R}_{AB} - \frac{1}{2}\hat{g}_{AB}\hat{R}$ is the five-dimensional Einstein tensor. Throughout this section, we denote five dimensional tensors with a hat, e.g. $\hat{G}_{AB}$ is the five dimensional Einstein tensor, vs. $G_{ab}$ which is the four dimensional part of the Einstein tensor. Aside from that, capital Latin letters run over 0-4 and non-capital Latin letters run over 0-3.

(2) The metric has the following form:

$$\hat{g}_{AB} = \begin{bmatrix} g_{ab} + \phi^2 A_a A_b & \phi^2 A_a \\ \phi^2 A_b & \phi^2 \end{bmatrix}$$

where $g_{ab}$ is the 4-dimensional metric we perceive in four dimensional spacetime, $A^a$ is the electromagnetic 4-potential and $\phi$ is some scalar field (called the 'radion' or the 'graviscalar'. Notice that this is not the scalar potential of the electromagnetic field!)

(3) The cylinder condition:
Since we observe nature as having three spatial dimensions, Kaluza tried to explain adding the extra spatial dimension to his theory by requiring that nature is independent of the fifth coordinate. Therefore all derivatives with respect to the fifth coordinate are zero, namely $\frac{\partial}{\partial x^4} = 0$.

It should be emphasized that one of the most beautiful aspects of Kaluza's theory is that spacetime in five dimensions is actually empty (there is no stress-energy tensor). Matter, energy and radiation as we interpret them are just the manifestation of empty five-dimensional spacetime, as we will see next.

The inverse five-dimensional metric is

$$\hat{g}^{AB} = \begin{bmatrix} g^{ab} & -A^a \\ -A^b & \frac{1}{\phi^2} + A^2 \end{bmatrix}$$

where $g^{ab}$ is the inverse four-dimensional metric, and we raised the index of the electromagnetic using the four-dimensional metric, namely $A^a = g^{ab}A_b$. Next we compute the five-dimensional Christoffel symbols by the formula (remembering the cylinder condition)

$$\hat{\Gamma}^C_{AB} = \frac{1}{2}\hat{g}^{CD}\left(\partial_A \hat{g}_{BD} + \partial_B \hat{g}_{AD} - \partial_D \hat{g}_{AB}\right)$$

and then the five-dimensional Ricci tensor by

$$\hat{R}_{AB} = \partial_C \hat{\Gamma}^C_{AB} - \partial_B \hat{\Gamma}^C_{AC} + \hat{\Gamma}^C_{AB}\hat{\Gamma}^D_{CD} - \hat{\Gamma}^C_{AD}\hat{\Gamma}^D_{BC}$$

If we separate the vacuum five-dimensional Einstein field equation $\hat{R}_{AB} = 0$ to the $ab$ components, $a4$ components and the 44 component, we obtain the following equations[4]:

$$\begin{aligned} G_{ab} &= \frac{\phi^2}{2}T_{ab} - \frac{1}{\phi}\left(\nabla_a\left(\partial_b \phi\right) - g_{ab}\Box\phi\right) \\ \nabla^a F_{ab} &= -3\frac{\partial^a \phi}{\phi}F_{ab} \\ \Box\phi &= \frac{\phi^3}{4}F_{ab}F^{ab} \end{aligned}$$

where $G_{ab} = R_{ab} - \frac{1}{2}g_{ab}R$ is the four-dimensional Einstein tensor, $T_{ab} = \frac{1}{4}g_{ab}F_{cd}F^{cd} - F^c_a F_{bc}$ is the electromagnetic energy-momentum tensor, and $F_{ab} = \partial_a A_b - \partial_b A_a$ is the tensor for the electromagnetic field. Kaluza set $\phi = 1$, and then we can see that the first two equations are nothing other than the Einstein equation and covariant Maxwell's equations:

$$\begin{aligned} G_{ab} &= \frac{1}{2}T_{ab} \\ \nabla^a F_{ab} &= 0 \end{aligned}$$

The four-dimensional matter (or at least electromagnetic radiation) arises purely from the geometry of empty five-dimensional spacetime!

Though this seems as a pure miracle, this formulation does have a few inherent problems. Requiring $\phi$ to be constant is consistent with the third field equation only when $F_{ab}F^{ab} = 0$, a condition which the electromagnetic field does not necessarily satisfy. If we allow $\phi$ to vary then this issue is resolved. Nevertheless, the scalar field $\phi$ was one of the reasons that Kaluza's theory was eventually abandoned, as its effect could not be observed. In the next section we discuss Klein's attempt to explain the fact that Kaluza's theory is independent of the fifth dimension by 'compactifing' the fifth dimension.

5.2. **Oskar Klein's hidden dimension.** The contrived nature of Kaluza's assumption, that all physical quantities do not depend upon the fifth dimension seems very 'unnatural.' Klein arrived on the scene in 1926, when a lot of excitement surrounded the birth of quantum mechanics. Since most quantum mechanical effects seem to appear on very small scales, Klein tried to explain this in his theory. Klein assumed that the fifth coordinate in Kaluza's theory is a length-like one, and assigned to it two properties[5]:

(1) It has a circular $\mathbb{S}^1$ topology, and is thus also compact.
(2) It has a very small scale (i.e. the 'radius' of the fifth dimension as a circle is very 'small.')

The first property implies that all quantities become periodic in the fifth dimension, and thus may be written as a Fourier series:

$$\begin{aligned} g_{AB}\left(x^0,x^1,x^2,x^3,x^4\right) &= \sum_{n=-\infty}^{\infty} g_{AB}^{(n)}\left(x^0,x^1,x^2,x^3\right)e^{i\frac{n}{r}x^4} \\ A_A\left(x^0,x^1,x^2,x^3,x^4\right) &= \sum_{n=-\infty}^{\infty} A_A^{(n)}\left(x^0,x^1,x^2,x^3\right)e^{i\frac{n}{r}x^4} \\ \phi\left(x^0,x^1,x^2,x^3,x^4\right) &= \sum_{n=-\infty}^{\infty} \phi^{(n)}\left(x^0,x^1,x^2,x^3\right)e^{i\frac{n}{r}x^4} \end{aligned}$$

where $r$ is the scale parameter (the 'radius' of the fifth dimension) and the superscript $^{(n)}$ refers to the n-th Fourier mode. In quantum mechanics the momentum of a wave function of the form $e^{ikx}$ is of the order of $k$, and thus Klein concluded that also in our case the momentum in the $x^4$-direction is of the order of $\frac{|n|}{r}$. By the second property of the fifth dimension, since $r$ is very small the $x^4$-momentum of all modes other than $n = 0$ will be so large as to put it beyond the reach of current experiments. Therefore the only mode that will actually be observable is $n = 0$, which is independent of the fourth coordinate. This gives a possible explanation for the independence of the fifth coordinate $x^4$ in Kaluza's theory.

As mentioned, at the time of Kaluza-Klein theory, physicists were much more interested in probing the microscopic world with the new quantum mechanics and didn't pay any significant attention to Kaluza-Klein theory. Furthermore, the extra scalar field $\phi$ that appears in the theory caused a few predictions of Kaluza-Klein theory that were in conflict with experimental data. This led to the abandonment of Kaluza-Klein theory in favor of quantum mechanics, though by the early 1970s the interest in the theory led to the development of more modern extra-dimensional theories, such as string theory, who also tried to unify gravity and electromagnetism, together with the strong and weak nuclear interactions. Up to the present time, there is no accepted theory that unifies gravity and electromagnetism.

## 6. The Future

As we saw in the previous sections, gravity and electromagnetism have many similar properties. However, at present one of the most serious problems in theoretical physics is interrelating these two forces. The same problem appears in experimental physics as well: finding an experimental evidence for an interchange between the electromagnetic and gravitational field, just as in the case of the electric and magnetic fields. In fact, even the great experimentalist Michael Faraday spent more than 11 years trying to show such a relation between the two fields unsuccessfully [1].

My personal belief is that gravity is an effect of the electromagnetic field, just as the electric field appears as a magnetic field depending on the frame of reference. The electromagnetic field seems to be more fundamental as it can both attract and repel, while gravity only attracts. On the microscopic scale of an atom the gravitational field seems to be completely irrelevant to the motion of particles, as the electromagnetic field is 39 order of magnitudes stronger. These two facts and our inability of measuring the gravitational field on such a small scale, seem to me as if the gravitational field might just be the residual effect of some mutual influence between negative and positive charges in the atom.

I'll conclude by an optimistic note from Faraday's dairy, after his failure in detecting such a gravelectric effect, which strongly describes my own opinion about the subject[1]:

*'It was almost with a feeling of awe that I went to work, for if the hope [of interrelating gravity and electricity] should prove well founded, how great and mighty and sublime in its hitherto unchangeable character is the force I am trying to deal with, and how large may be the new domain of knowledge that may be opened up to the mind of man... [The experimental results] do not shake my strong feeling of the existence of a relation between gravity and electricity... '*

## References

[1] Geoffrey Cantor. *Faraday's search for the gravelectric effect.* Phys. Educ. 26, 1991.
[2] Albert Einstein. *On the Electrodynamics of Moving Bodies.* Annalen der Physics, 1905.
[3] James B. Hartle. *Gravity: An Introduction to Einstein's General Relativity.* Benjamin Cummings, 2003.
[4] Theodor Kaluza. *Zum Unitatsproblem in der Physik.* Sitzungsber. Preuss. Akad. Wiss. Berlin. (Math. Phys.), 1921.
[5] Oskar Klein. *Quantentheorie und funfdimensionale Relativitatstheorie.* Zeitschrift fur Physik a Hadrons and Nuclei, 1926.
[6] Thomas S. Kuhn. *The Structure of Scientific Revolutions.* University Of Chicago Press, 1996.
[7] James Clerk Maxwell. *A Dynamical Theory of the Electromagnetic Field.* Philosophical Transactions of the Royal Society of London, 1865.
[8] Hermann Minkowski. *Space and Time.* Assembly of German Natural Scientists and Physicians, 1908.
[9] Isaac Newton. *Philosophiae Naturalis Principia Mathematica.* London, 1687.
[10] Robert M. Wald. *General Relativity.* University of Chicago Press, 1984.