On the unification of forces of nature

A. K. T. Assis

Permanent address: Instituto de Física ‘Gleb Wataghin’, Universidade Estadual de Campinas - Unicamp, 13083-970 Campinas, São Paulo, Brasil.
Institut für Geschichte der Naturwissenschaft
Universität Hamburg, Bundesstr. 55, D-20146 Hamburg, Deutschland
E-mail: assis@ifi.unicamp.br, Homepage: www.ifi.unicamp.br/~assis

ABSTRACT. Different aspects of physical interaction are considered: inertia, gravitation, electrostatics, magnetostatics and galvanism. It is discussed the experimental and theoretical reasons leading to the unification of some of them. It is then explored by analogies what might lead to the unification of gravitation with electromagnetism.

Key Words: Unification of gravitation with electromagnetism, force laws, unified theories, Weber’s electrodynamics, relational mechanics.

P.A.C.S.: 04.50.+h, 41.20.-q, 01.65.+g

1 The Basic Forces of Nature

The aim of this work is to analyse the unification of the forces of nature. The procedure followed here is different from the standard approaches which have been utilized in this century. To this end different branches of physics are considered here, namely: inertia, gravitation, electrostatics, magnetostatics and galvanism. In this work it will not be discussed nuclear interactions (strong and weak forces), chemical forces, nor contact forces like friction, elastic interactions etc. What is considered is the interconnections in the electromagnetic interactions, the interconnection of gravitation with inertia, and what might be expected if there is a fundamental connection of electromagnetic forces with gravitation, as expected. The basic goal of this paper is to suggest possible routes for the unification of gravitation with other branches of physics, following what was accomplished in electromagnetism.
One basic property of bodies is inertia, their resistance of changing their state of motion relative to the earth or to the fixed stars. According to Newton’s second law of motion (1687), the change of motion is proportional to the motive force impressed and is made in the direction of the right line in which that force is impressed. Representing this force vectorially by $\vec{F}$, the inertial mass of the body by $m_i$, and its velocity relative to an inertial frame of reference by $\vec{v}$, then this law can be expressed as:

$$F = K_1 \frac{d(m_i \vec{v})}{dt},$$

where $K_1$ is a constant of proportionality which depends on the system of units to be employed.

Beyond this passive property of a body there are also active properties like the forces of gravitation, electrostatics, magnetostatics and galvanismus. These four kinds of interaction may be expressed by force laws connecting bodies of the same nature. They are presented in order of historical origin (gravitational force by Newton in 1687, magnetostatic and electrostatic forces by John Michell, Tobias Mayer and Augustin Coulomb between 1750 and 1785, and force between current elements by Ampère between 1820 and 1826). A force exerted by a body $j$ on a body $i$ is represented by $\vec{F}_{ji}$. Bodies $i$ and $j$ are considered as particles, namely, with negligible dimensions (or negligible maximal diameters) compared with their separation.

According to the law of universal gravitation the force exerted by a gravitational mass $m_{g2}$ on $m_{g1}$ is given by:

$$\vec{F}_{21} = -K_2 \frac{m_{g1} m_{g2} \hat{r}}{r^2} = -\vec{F}_{12}.$$

Here $K_2$ is a constant of proportionality, $r$ is the distance between the bodies and $\hat{r}$ is the unit vector pointing from 2 to 1. Although Newton in the *Principia* dealt with only of one type of mass, what is being called here the inertial mass of a body, it is better to distinguish the two types of mass for the moment as they are conceptually quite different from one another. While the inertial mass is a measure of the resistance of a body to change its state of motion due to a force of any nature, the gravitational mass is related with a specific property of a body, its gravitational interaction. In this sense it could also be called a gravitational charge (in
analogy with the electrical charge, responsible for the electrical force). The relation between $m_i$ and $m_g$ will be further discussed later on.

The electrostatic force between the electrical charges $q_1$ and $q_2$ is given by (with a constant of proportionality $K_3$):

$$ F_{21} = K_3 q_1 q_2 \frac{\hat{r}}{r^2} = -F_{12}. \quad (3) $$

The magnetostatic force describes the interaction between magnets. Coulomb worked with thin and long magnets, so that the magnetic poles might be considered as concentrated on their ends. Performing experiments with a torsion balance he could find an expression describing the interaction between these magnetic poles (by convention a north pole is considered as positive and a south pole as negative). His force between the magnetic poles $q_1^\ast$ and $q_2^\ast$ is given by

$$ F_{21} = K_4 q_1^\ast q_2^\ast \frac{\hat{r}}{r^2} = -F_{12}. \quad (4) $$

Here $K_4$ is a constant of proportionality.

What is being called here by galvanism is the interaction between circuits carrying electrical currents $I_1$ and $I_2$. Ampère's force between current elements $I_1 d\vec{\ell}_1$ and $I_2 d\vec{\ell}_2$ is given by

$$ d^2 F_{21} = -K_5 I_1 I_2 \frac{\hat{r}}{r^2} \left[ 2(d\vec{\ell}_1 \cdot d\vec{\ell}_2) - 3(\hat{r} \cdot d\vec{\ell}_1)(\hat{r} \cdot d\vec{\ell}_2) \right] = -F_{12}. \quad (5) $$

Here $K_5$ is another constant of proportionality and $d\ell_1$ ($d\ell_2$) is the length of the element 1 (2), respectively. Integrating this expression for the closed circuit 2, $C_2$, Ampère could obtain the force exerted by this circuit on $I_1 d\vec{\ell}_1$ as

$$ dF_{21} = I_1 d\vec{\ell}_1 \times \left( K_5 \oint_{C_2} \frac{I_2 d\vec{\ell}_2 \times \hat{r}}{r^2} \right). \quad (6) $$

With another integration over the closed circuit 1, $C_1$, he was then able to obtain the ponderomotive force exerted between them as given by
The four expressions (2) bis (5) are extremely similar. All of them fall as $1/r^2$, are along the straight line connecting the particles and follow Newton’s action and reaction law. Moreover, all of them depend on the product of a property of each body describing the kind of interaction, namely: gravitational mass, electrical charge, magnetic pole and electrical current. Only Ampère’s force is more complex than the others, depending also on the relative orientation of the bodies (current elements) and the straight line connecting them.

Despite these similarities it cannot be said only for this that they are unified or that they could be derived from one another. More things than this are necessary in order to find out their inner connections, to discover what expression is more basic than the others etc. In principle there might be no relations whatsoever between a gravitational mass, an electrical charge, a magnetic pole and a current element. They might be independent properties of each body. Only experiments or a clear theoretical derivation may connect these concepts.

The constants $K_1$ to $K_5$ can be chosen conveniently depending on the system of units which is being employed. Following the procedure first presented by Gauss in 1832 and later on followed by Weber, it is possible to put all of them equal to 1 unitless and then adapt the dimensions of $m_i$, $m_g$, $q$, $q^*$ and $I$ appropriately. But in this work all of them will be left unspecified for the sake of generality.

2 On the Unification of electrostatics, magnetostatics and galvanism

The first clear connection between any two of these four branches of physics was Oersted’s discovery of the deflection of a magnetized needle by a current carrying wire. This was accomplished in 1820 and motivated the works on electromagnetism by Ampère. But historically Ampère might had discovered independently that two current carrying wires attract (repel) one another when the currents are in the same (opposite) direction. From this he might arrive at his equation (5) before 1820. As a matter of fact, Oersted discovery suggested to Ampère only one basic fact: That the magnetostatic interactions (or the interactions
between current carrying wires and magnets) could be due to an interaction between currents, although this last interaction had not yet been observed. From this insight Ampère was led to work directly with currents and found (5) and all the rest. What will be discussed here is not what came first or how these expressions were originated, but only how their connections arise.

Ampère could then reproduce equation (4) beginning with equation (5). This can be seen as follows: Consider two long cylindrical solenoids with \( N_1 \) and \( N_2 \) turns carrying each one of them currents \( I_1 \) and \( I_2 \), respectively. Each solenoid has a cross sectional radius \( a_1 \) and \( a_2 \), with total lengths \( l_1 \) and \( l_2 \), respectively. Suppose now the closest distance between one extremity of one solenoid and one extremity of the other solenoid be \( r \) such that \( l_1 \gg r \), \( l_2 \gg r \), \( r \gg a_1 \) and \( r \gg a_2 \). Then Ampère's expression (5) integrated for both solenoids yields an expression analogous to (4) with \( \frac{K}{s} \) replaced by the expression \( \frac{K}{s} \left( N_1 I_1 \pi a_1^2 / l_1 \right) \), where \( s = 1 \) or \( s = 2 \). This is then an example of an unification between magnetism and galvanism as the magnetic poles can be replaced by current carrying circuits. Equivalently we might say that a magnetic dipole \( q \) \( d \) (\( d \) being the distance between the north and south poles, which is considered to be much smaller than the point of observation to the magnet) is equivalent to a small loop of area \( A \) carrying a constant current \( I \), namely: \( q \) \( d \) \( \leftrightarrow IA \), the axis of the small magnet corresponding to the orthogonal to the loop area. Essentially what Ampère showed was that magnetostatic might be explained in terms of interacting current carrying loops. To this end he needed to postulate the existence of microcurrents in each molecule of a ferromagnet.

A clear connection between electric charges and magnetostatics arose with Rowland's work of 1876. He charged a dielectric disc, rotated it and showed that it influenced a magnetic needle at rest in the laboratory. This proved that charges in motion are equivalent to a magnet (or to a current loop, following Ampère's unification of magnetism and galvanism). The opposite effect of a magnet deflecting the motion of charges or of an electric discharge had been discovered in 1821 by H. Davy and this was also investigated by J. Plücker in 1858. Hall's discovery of an electromotive force in a current carrying strip with a magnetic field orthogonal to its plane in 1879 may also be considered along these lines.

None of these effects might be derived from eqs. (3), (4) or (5) considered alone. Something else was missing. In order to unify electrostatics
with magnetism or galvanism (these two last areas had already been unified by Ampère's work) it was necessary to have a generalized force between charges depending on their velocities. With the further supposition that an electric current is due to charges in motion, an unification might be accomplished.

The first to take this step was Gauss in 1835, but his work was only published posthumously in 1867. The first published equation generalizing the electrostatic force to take into account the effect of the velocity of the charges in their force is due to Weber in 1846, [1, Vol. 3, p. 25]. For modern discussions of Weber’s law applied to electromagnetism and gravitation with many references see the following works: [2], [3], [4], [5], [6] and [7]. Weber’s force is given by:

$$F_{21} = K \frac{q_1 q_2}{r^2} \left(1 - \frac{\dot{r}^2}{2c^2} + \frac{\ddot{r}}{c^2}\right) = -F_{12}.$$  \hspace{1cm} (8)

Here \(\dot{r} = \frac{dr}{dt}\) and \(\ddot{r} = \frac{d^2r}{dt^2}\). Moreover, \(c = 3 \times 10^8 m/s\) is the ratio of electromagnetic and electrostatic units of charge. This quantity appeared here for the first time in physics. Its first measurement was accomplished by Weber and Kohlrausch in 1855, [1, Vol. 3, pp. 591, 597 and 609]. The obtained value equal to the light velocity suggested a possible unification of electromagnetism and optics.

Weber’s force can also be derived from Weber’s potential energy, which he presented in 1848:

$$U = K \frac{q_1 q_2}{r} \left(1 - \frac{\dot{r}^2}{2c^2}\right).$$  \hspace{1cm} (9)

Electrostatics is recovered from Weber’s force when there is no motion between the charges, namely, when \(\dot{r} = 0\) and \(\ddot{r} = 0\). Weber was able to unify electrostatics and electrodynamics with this expression. The main idea is to derive Ampère’s force from Weber’s one. This was the only thing that was needed, after all magnetism and galvanism had already been unified by Ampère. The idea is to consider each current element \(I_s \, d\vec{r}_s\), where \(s = 1\) or \(s = 2\), as composed of positive \(dq_{s+}\) and negative \(dq_{s-}\) charges moving with velocities \(\vec{v}_{s+}\) and \(\vec{v}_{s-}\), respectively, relative to an inertial frame of reference. Weber’s forces (four of them) exerted by the positive and negative charges of one current element on the positive and negative charges of the other current element are then added to one another. Although the current elements in general are
not electrically neutral it is possible to assume this charge neutrality as the electrostatic forces are usually much smaller than the ponderomotive forces between current carrying wires, [8]. Supposing the charge neutrality of the current elements, \( dq_{s-} = -dq_{s+} \), where \( s = 1 \) or \( s = 2 \), the summation of the four Weber’s force yield the force exerted by the current element 2 on 1 as given by:

\[
d^2 \vec{F}_{21} = -K_3 dq_{s+} dq_{s-} \frac{\vec{r}}{r^2} \left\{ 2 \frac{(\vec{v}_{s+} - \vec{v}_{s-}) \cdot (\vec{v}_{s+} - \vec{v}_{s-})}{c^2} \right\}
\]

This expression is analogous to Eq. (5) with \( \sqrt{\frac{1}{\epsilon_0} \mu_0} dq_s \) replaced by the expression \( \sqrt{\frac{K_3}{c^2}} dq_{s+} (\vec{v}_{s+} - \vec{v}_{s-}) \), where \( s = 1 \) or \( s = 2 \). And in essence this is the unification of electric forces with galvanism because it is then possible to replace all electric currents by charges in motion.

Weber was also able to derive Faraday’s law of induction of 1831 from his expression.

Electrostatics may then be considered as a zeroth order effect. That is, it does not depend on the velocity of the charges. The main conclusion from the unification obtained above is to show that galvanism or electrodynamics, magnetism and Faraday’s law of induction are all second order effects. That is, they depend on \( v_1 v_2 / c^2 \) or to \( r \vec{r} / c^2 \), where \( v_1 \) and \( v_2 \) are the velocities of the interacting charges relative to the laboratory or to one another, \( r \) is the separation between them and \( \vec{r} \) is the relative radial acceleration between them.

It is now considered gravitation.

3 Unification of Gravitation with Inertia

The property of bodies to fall in the direction of the earth’s center when released is called gravity. According to Newton this property is also responsible for keeping the planets in their orbit etc. It can be supposed that in each body there is a certain amount of gravitational mass, \( m_g \), which is responsible for this interaction. Operationally this can be measured by the process of weighing the body (by definition the ratio of the gravitational masses of two bodies is the ratio of their weights at the same spot near the earth’s surface, \( m_{g1} / m_{g2} = P_1 / P_2 \)). The gravitational force can be expressed by Eq. (2).
Another property of a body is its resistance to change its state of motion (being at rest or moving with a constant linear velocity in relation to an inertial frame of reference). This is represented by its inertial mass, \( m_i \). Operationally this can be measured by the acceleration the body suffer when acted on by a force (the ratio of the inertial masses of two bodies is the inverse ratio of their accelerations relative to an inertial frame of reference when acted upon by equal forces, \( m_{i1}/m_{i2} = -a_2/a_1 \)). The inertial force can be expressed by Eq. (1).

As the force or interaction responsible for the acceleration of the test body (with which it is possible to know its inertial mass) can be of any nature (electric, magnetic, gravitational, elastic etc.), there is in principle no connection between the gravitational and inertial masses of a body (they are conceptually different and measured by unrelated processes, gravitational mass by a static procedure of weighting and inertial mass by a dynamic result of its interaction with other bodies). On the other hand it is experimentally known that both masses are proportional to one another. The easiest way to see this is to consider the acceleration of free fall for a test body near the surface of the earth, disregarding the effects of air friction. Combining Eqs. (1) and (2) yields for a constant inertial mass:

\[ K_1 m_{i1} a_1 + K_2 \frac{m_{g1} m_{gE}}{r_E^2} = 0, \]  

(11)

where \( m_{gE} \) is the gravitational mass of the earth and \( r_E \) its radius. This yields \( a_1 = -(K_2/K_1)(m_{g1}/m_{i1})(m_{gE}/r_E^2) \). The ratio of the free fall acceleration of body 1 to the free fall acceleration of body 2 at the same spot on the earth’s surface is then given by

\[ \frac{a_1}{a_2} = \frac{m_{g1}/m_{i1}}{m_{g2}/m_{i2}}. \]

(12)

It is an experimental fact known since Galileo that two bodies fall freely from the same height with the same acceleration near the earth’s surface, no matter their weight, form, chemical composition etc. This means that \( a_1 = a_2 \), or that \( m_{g1}/m_{i1} = m_{g2}/m_{i2} \). As this is valid in general, no matter the form, weight or chemical composition of the bodies, it is possible to write

\[ m_i = K_6 m_g, \]

(13)
where $K_6$ is a constant for all bodies.

Another way of obtaining this fact is to observe that the oscillation of two pendulums of the same length filled with matter of different nature located at the same spot near the earth's surface happens with the same frequency, as first determined by Newton. This means that the inertia of a body can be measured by its weight, a highly non trivial finding. Utilizing Eq. (13) allows Eq. (2) to be written in the form

\[ F_{21} = -\frac{K_2}{K_6^2} m_1 m_2 \frac{\ddot{r}}{r^2} = -F_{12}. \] (14)

This shows that the gravitational force is proportional to the product of the inertial masses of the attracting bodies. It was in this form that Newton presented the law in the Principia.

The free fall and pendulum experiments have been known since Newton's time, indicating that the inertial and gravitational properties of a body are proportional to one another. But until recently there was no theoretical explanation for this remarkable fact. What was necessary was to derive Eq. (13), or the derivation of inertial forces ($m_i \ddot{a}$, centrifugal, Coriolis etc.) from gravitational ones. That is, a quantitative implementation of Mach's principle. This was accomplished by a Weber's law applied to gravitation. For references and a detailed discussion, see [9], [2, Chapt. 6], [10, Chapt. 3], [5] and [6].

Relational mechanics begins with a Weber's law for gravitation and with the principle of dynamical equilibrium: the sum of all forces of any nature (gravitational, electric, magnetic, elastic, nuclear etc.) acting on any body is always zero in all frames of reference. From this it is derived that the equation of motion describing the gravitational interaction of body 1 with the earth (supposing that the test body is falling with velocity much smaller than light velocity) and with the distant universe is given by, [5, Chapters 8 and 9]:

\[ \frac{H_g M_{go}}{R_o^2} m_1 a_{1U} + H_g \frac{m_1 m_2 E}{r^2} = 0, \] (15)

where $H_g$ is a constant of proportionality, $H_o$ is Hubble’s constant, $M_{go}$ is the gravitational mass of the known universe (that is, the gravitational mass inside Hubble’s radius $R_o = c/H_o$, with $c$ being light velocity in vacuum), $a_{1U}$ is the acceleration of body 1 relative to the universal frame of reference $U$ (the frame in which the set of distant galaxies is seen as
at rest without acceleration). Moreover, $\alpha = 6$ dimensionless if Weber’s gravitational force is integrated until Hubble’s radius or $\alpha = 12$ dimensionless if Weber’s gravitational force is coupled with an exponential decay and integrated until infinity.

Comparing Eqs. (11) and (15) yields:

$$m_i = \frac{K_2 \alpha M_{g0}}{K_1 H_0^2 R_0^3} m_g.$$  \hspace{1cm} (16)

This means that relational mechanics yields the constant $K_6$ in Eq. (13). It shows that the inertia of a body is due to its gravitational interaction with the distant universe by means of Weber’s gravitational law. Moreover, it shows that this inertial mass of a body does not need to be a constant, as it is directly proportional to the average density of gravitational mass around this body. Changing the environment around a body should change its inertial properties (acceleration of free fall etc.)

4 Unification of Gravitation with Electromagnetism

This unification can be considered from the experimental or theoretical points of view.

Theoretically the main idea would be to derive Newton’s law of gravitation from a generalized electromagnetic force. This would be analogous to derive Ampère’s force from a generalized electrostatic force. The gravitational mass of a body could then be derived as a statistical summation of charges in motion, as was the case with the identification of $Id\ell$ with $qv$. Works along this line were published in [11] and [12]. The basic assumption is to begin with a generalized sixth order electromagnetic potential energy and force given by, respectively:

$$U = K_3 q_1 q_2 \frac{1}{r} \left[ 1 - \alpha \left( \frac{\dot{r}}{c} \right)^2 - \beta \left( \frac{\dot{r}}{c} \right)^4 - \gamma \left( \frac{\dot{r}}{c} \right)^6 - \ldots \right],$$  \hspace{1cm} (17)

$$\vec{F}_{21} = -\dot{r} \frac{dU}{dr} = K_3 q_1 q_2 \frac{\dot{r}}{r^2} \left( 1 - \frac{\dot{r}^2}{c^2} - \frac{2 \dot{r} \ddot{r}}{c^2} - \frac{\dot{r}^4}{c^4} - \frac{4 \dot{r}^2 \ddot{r}}{c^4} - \frac{\dot{r}^6}{c^6} - \frac{6 \dot{r}^4 \ddot{r}}{c^6} - \ldots \right).$$  \hspace{1cm} (18)
In these equations $\alpha$, $\beta$ and $\gamma$ are dimensionless constants of the order of unity (the precise value is yet to be determined based on the consequences of the law and on experimental results).

The interaction between two groups of neutral electrical dipoles is then considered. In each dipole it is supposed that the negative charges oscillated around the positive ones. The electromagnetic forces (four of them) exerted by the positive and negative charges of one dipole on the positive and negative charges of the other dipole were then added to one another. Performing averages for the directions of oscillation and in time, the zeroth and second order terms went to zero. But it was possible to show that it would remain a small attractive fourth and sixth order electromagnetic force between the dipoles. The fourth order term could then be interpreted as Newton’s law of gravitation. The final average result was given by:

$$
\bar{F} = -K_3 \frac{7\beta}{18} q_1 q_2 \frac{\tilde{R}}{R^3} \frac{A_1^2 - \omega_1^2 A_2^2 - \omega_2^2}{c^4} \left(1 + \frac{\gamma}{\beta} \frac{45\tilde{R}^2 - 18RR}{7c^2}\right). \quad (19)
$$

Here $q_{s+}$ is the positive charge of the $s$ dipole ($s = 1$ or $s = 2$), $\tilde{R}$ is the position vector connecting the dipoles, $R = |\tilde{R}|$, $\dot{R} = dR/dt$, $\ddot{R} = d^2R/dt^2$ and $A_1$ is the amplitude of oscillation of the negative charge of the $s$ dipole around the positive one with frequency $\omega_s$.

The reason for the analogy of the fourth order term with Newton’s law of gravitation was that it did fall as $1/r^2$, was along the straight line connecting the dipoles and had the correct order of magnitude. In order to derive the law of gravitation, Eq. (2), it is necessary essentially only to identify $\sqrt{K_2 m g}$ with $\sqrt{K_3 r^3 \beta / 18c^4 q_{s+} A_2^2 - \omega_2^2}$. And this represents a theoretical unification of gravitation with electromagnetism, as all gravitational masses can then be replaced by electric dipoles in which the negative charges oscillate around the positive ones. The essential result was that gravitation might be considered as a fourth order electromagnetic effect, as magnetism, galvanism and induction are considered to be second order effects.

The sixth order term would then be responsible for inertia, as was shown elsewhere: [9] and [5].

On the other hand, it may not yet be said that gravitation has been unified with electromagnetism as no experimental effect has been discovered connecting these two branches. Faraday was one of the first trying
to find this relation. He believed that as a neutral body was being accelerated towards the earth due to their gravitational attraction, electrical currents might be developed in the body, in the earth or in surrounding matter. Although he did not find any positive effect, this negative outcome did not shake his belief that some relation must exist between gravity and electricity. Had him found a positive result, the opposite effect might be the setting up of a current in a loop affecting the weight or the rate of free fall of a neutral body nearby.

In analogy with what was obtained by Rowland and Hall, it might also be expected a relation between electrostatics and gravitation. For instance, a neutral body might become electrically polarized (or change its electrostatic polarization) when its distance to the surface of the earth is changed (changing its potential energy), or when it is moved relative to the earth with a constant velocity or constant acceleration, or when it rotates or oscillates relative to the earth.

None of these effects has yet been found. But here I want to stress the importance of Faraday’s experiments and to suggest further analogies. The idea is to motivate others to repeat and improve on Faraday’s approach, as there are much better techniques and apparatus nowadays. I believe these experimental routes will unlock the mystery of the probable connection between gravitation and electromagnetism.

Acknowledgments:
The author wishes to thank the Alexander von Humboldt Foundation, Germany, for a research fellowship during which this work was completed.

References
On the unification of forces of nature


(Manuscrit reçu le 1er mars 2001)