Chapter 1

Introduction to electrodynamics

May 25, 2021

An introduction to classical electrodynamics is given, placing it in context within modern physics.

Contents

1 Introduction

Why do we bother to learn electromagnetism (EM) or classical electrodynamics (CED)?

1.1 Applications

First, there are many applications.

Problem 1
Cite three applications of CED (preferably from different domains) and explain which laws or principles are involved and how. §

1.2 Fundamental interactions

More importantly for the present course: CED is the first fundamental interaction to be extensively understood, and leads to many concepts at the core of theoretical physics.

interaction	strength	range
strong	1	10^{-15} m
EM	10^{-2}	∞
weak	10^{-5}	10^{-18} m
gravitation	10^{-40}	∞

Table 1. Fundamental interactions

Fundamental interactions are of four types[1] (Table 1). For an introductory account on these types of forces, many reference [1] are available.

Range
For all these interactions, one can roughly speak of a potential

$$V(r) \quad \sim \quad r^{-1}\, e^{-r/a} \tag{1}$$

[1]There are now some differences in nomenclature concerning the strong interactions. Some authors restrict this term to the interaction between quarks and gluons (i.e., between objects with color degree of freedom). The nuclear force is a left-over effect between two colorless objects (i.e., color singlets), like the van der Waals force between two electrically neutral objects. Here we adopt a looser (and the historical) convention, and refer to nuclear forces also as part of the strong interaction.

and the range cited in Table 1 corresponds to a, beyond which the potential and the force are cut off exponentially. When we say that EM and gravity have infinite ranges, that means $a = \infty$, i.e., there is no exponential factor, so $V \propto r^{-1}$, $F \propto r^{-2}$ — the familiar inverse-square forces.

Yukawa was the first to realize that the range of interaction a is determined by the mass m of the mediating particle (**Figure 1**), through the relationship

$$\boxed{a = \frac{\hbar}{mc}} \tag{2}$$

The argument goes as follows.

- A mass m has to be created, costing an energy $E = mc^2$. Energy conservation is violated by this amount.
- This is only possible if there is an uncertainty in the energy, ΔE, also of this amount. Hence $\Delta E \sim mc^2$.[2]
- Such energy uncertainty is only possible for a short duration Δt, and by the uncertainty principle $\Delta t \cdot \Delta E \sim \hbar$. Thus, the intermediate state can only persist for $\Delta t \sim \hbar/mc^2$.
- During this time, the mediating particle can at most travel $a \sim c\,\Delta t \sim \hbar/mc$.

EM and gravity are described by r^{-1} potentials and r^{-2} forces because the mediating particles are massless — respectively the photon and the graviton.

A word about gravitons, which is at the moment purely hypothetical.

- If we believe Einstein's equations, then there should be wave solutions — gravitational waves. There is indirect evidence for gravitational waves, through the change in the orbital period of binary pulsars due to energy radiated by gravitational waves [2, 3, 4]. In 2016, direct evidence for gravitational waves was announced by LIGO [5, 6, 7].
- If we further believe in the universality of wave–particle duality, then these waves should exist in quanta called gravitons. It will be a long time before gravitons can be found experimentally.

Problem 2

Using the above relationship, we can understand the ranges of the strong and weak interactions.

(a) The strong interaction (binding protons and neutrons together) is mediated by the π meson, with mass $m \approx 0.14$ GeV/c^2.[3] Estimate the range a and hence explain the size of nuclei.

(b) The weak interaction (responsible for β decay for example) is mediated by intermediate bosons with masses ~ 90 GeV/c^2.[4] Estimate the range a. §

Range and classical physics

Because of the very short range of the strong and weak interactions, they are not (directly) visible at macroscopic distances. Their experimental study requires techniques for probing such short distances. Their theoretical study requires quantum mechanics (indeed quantum field theory). Turning this statement around, it means that EM and gravity are the only forces that can be studied classically, and were the first two forces to be understood.

Relative strengths

The story about the relative strengths (at least for the first three) is more complicated.

First, the main difference between the EM and weak interactions is that the latter has a short range. In low energy experiments (say below mc^2 for the mass m in Problem 2(b), ~ 90 GeV), there is a low spatial resolution and heuristically one only senses some average over a larger distance, thus causing a dilution. Above this energy range, the EM and weak interactions have approximately the same strengths, the essential idea behind *electroweak unification*.

Second and more subtly, the strengths have a slow (in fact logarithmic) dependence on the energy scale in question, and probably converge to the same magnitude at about 10^{15} GeV, at which the three interactions are likely to be unified — usually referred to as *grand unification*.

EM and gravity

EM and gravity are very similar, with inverse-square forces mediated by massless particles. But their strengths are very different, as shown in Table 1 and as illustrated in the next Problem.

Problem 3

Consider two protons. There is an electrostatic force F_E between them, and also a gravitational force F_G. Their ratio $\rho = F_G/F_E$ is independent of distance. Find ρ. §

An important difference is that charges can be positive or negative, and electrical forces can be either attractive or repulsive. Therefore, in bulk matter most of the electrical forces cancel and there is only a small net force due to a small unbalanced charge. On the other hand, masses are always posi-

[2] For example, the statement "6 = 7" is certainly wrong, but "$(6 \pm 1) = (7 \pm 1)$" can be regarded as correct.

[3] The mass of $\pi^{\pm}$ is 0.1396 units, while that of π^0 is 0.1350 units.

[4] The mass of the charged bosons $W^{\pm}$ is 80.4 units, while that of the neutral boson Z^0 is 91.2 units.

tive, and gravitational forces are always attractive; they can only add up. So in bulk matter (and even more so in astronomical matter), gravitational forces become very important.

A more fundamental way of viewing the difference is that EM is described by a vector field A^μ, whereas gravity is described by a tensor field $g^{\mu\nu}$. The spin-2 nature of the latter is ultimately the reason why gravity has only one sign. More importantly, it is the reason why gravity, when quantized in the standard way, does not lead to a finite theory. It is not expected that students would understand these remarks at this point. The hope is, however, that students would at least understand the vector nature of EM as a theory in spacetime at the end of these lectures, as background for going on to study the tensor theory of gravity (i.e., general relativity) — and perhaps one day going on to the challenges posed by trying to incorporate gravity into quantum mechanics and to unify it with the other three interactions.

2 The laws of EM

2.1 Conservation of charge

Electricity and magnetism are due to charges and currents — the sources. A key property about the sources is that charge is conserved, not only globally but locally. Consider a certain volume V. The total charge inside this volume is

$$Q = \int_V \rho \, dV$$

where ρ is the charge density. Its rate of increase must be equal to the rate of influx of charge into this volume, namely

$$-\int_S \mathbf{J} \cdot d\mathbf{S}$$

where $\mathbf{J}$ is the current density, i.e., the rate of flow of charge per unit area per unit time, and S is the surface bounding the volume V. The minus sign arises because $d\mathbf{S}$ is defined to be pointing outwards, so the surface integral gives the rate of outflux of charge.

Putting these together we get

$$\frac{d}{dt}\int_V \rho \, dV + \int_S \mathbf{J} \cdot d\mathbf{S} = 0$$

We then (a) in the first term bring the time derivative inside the integral; (b) in the second term convert the surface integral to a volume integral by Gauss' theorem; and (c) peel off the integral common to both terms, and obtain:

$$\boxed{\frac{\partial \rho}{\partial t} + \boldsymbol{\nabla} \cdot \mathbf{J} = 0} \tag{3}$$

as the statement for the local conservation of charge.

2.2 Two sides to electrodynamics

Electrodynamics consists of two parts.

- How EM fields act on charges and currents. The answer is the Lorentz force law.
- How charges and currents generate EM fields. The answer is the Maxwell equations.

The latter is much more complicated, and we shall be spending a large part of the course on advanced aspects of the Maxwell equations, especially in relation to radiation. The former is relatively easy, and is reviewed in Section 2.3.

These two aspects, like action and reaction, must be tied together in a consistent way. We shall deal with this at two levels: through the energy and momentum of the field (showing that their rates of change just compensate for those of the charged particles), and in a more sophisticated way, through the action formalism — in which both come from the same interaction term $J_\mu A^\mu$.

2.3 Lorentz force law

Force on point charge

The action of EM fields on point charges is given by the Lorentz force law, which can be taken to be the definition of the electric and magnetic fields:

$$\mathbf{F} = q\,(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \tag{4}$$

The units in the MKSA system are[5]

$$\begin{aligned}[\mathbf{E}] &= \mathrm{N\ C^{-1}} \equiv \mathrm{V\ m^{-1}} \\ [\mathbf{B}] &= \mathrm{N\,s\ m^{-1}\,C^{-1}} \equiv \mathrm{T}\end{aligned}$$

The Lorentz force law is so familiar that you may no longer be intrigued by its form: Why should $\mathbf{E}$ and $\mathbf{B}$ (especially the latter) appear in this particular way, which is by no means intuitive or "natural"? We hope that this question will be answered when you have completed this course.

Force on a current

Consider a thin wire segment $\Delta\boldsymbol{\ell}$, in which a current I flows. The force acting on the wire segment

[5] In some systems of units, the factor $\mathbf{v}$ is replaced by $\mathbf{v}/c$, which would have the convenience that $\mathbf{E}$ and $\mathbf{B}$ would have the same units. Units are discussed in the Appendix.

is

$$\mathbf{F} \quad = \quad I\,\Delta\boldsymbol{\ell} \times \mathbf{B} \tag{5}$$

coming from the second term of (4).

Problem 4
By considering the motion of the point charges which make up the current, prove (5). §

2.4 Maxwell's equations

You should be familiar with Maxwell's equations, typically written in the following form, with obvious meanings for the symbols:

$$\begin{aligned}
\boldsymbol{\nabla} \cdot \mathbf{E} &= \frac{\rho}{\epsilon_0} \\
\boldsymbol{\nabla} \times \mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t} \\
\boldsymbol{\nabla} \cdot \mathbf{B} &= 0 \\
\boldsymbol{\nabla} \times \mathbf{B} &= \mu_0 \mathbf{J} + \epsilon_0 \mu_0 \frac{\partial \mathbf{E}}{\partial t}
\end{aligned} \tag{6}$$

There is a surprising amount of information in these four lines, and since the whole course is about learning all the implications, not much will be said at this point, except to stress that there are (at least) three different ways of thinking about these equations.

Perspective 1: The divergence and the curl
The way Maxwell's equations are written in (6) is of the form

$$\begin{array}{ccc}
\boldsymbol{\nabla} \cdot \mathbf{E} = ? & , & \boldsymbol{\nabla} \times \mathbf{E} = ? \\
\boldsymbol{\nabla} \cdot \mathbf{B} = ? & , & \boldsymbol{\nabla} \times \mathbf{B} = ?
\end{array}$$

for which the implied story line is: A vector field is determined by its divergence and its curl; so given these, it should be possible to solve for **E** and **B**. This is the perspective most familiar from (and most useful for) electrostatics and magnetostatics.

Perspective 2: Evolution equations
Another way is to focus on the time derivative, namely the two curl equations, and write them as

$$\begin{aligned}
\frac{\partial \mathbf{E}}{\partial t} &= ? \\
\frac{\partial \mathbf{B}}{\partial t} &= ?
\end{aligned}$$

Now the implied story line is: Given the time derivative, it is possible to find **E** and **B** a bit later. This perspective is appropriate for initial value problems.

But what about the divergence equations? They should now be understood as constraints on the given initial conditions, say at $t = 0$.

But there seems to be a problem. Given such initial conditions, all later **E** and **B** are determined. We are not allowed to impose further conditions on them; in other words, we are not allowed to further require the divergence equations at later t — unless they are satisfied automatically.

Problem 5
Show that

$$\begin{aligned}
\frac{\partial}{\partial t}\left[\boldsymbol{\nabla} \cdot \mathbf{E} - \rho/\epsilon_0\right] &= 0 \\
\frac{\partial}{\partial t}\left[\boldsymbol{\nabla} \cdot \mathbf{B}\right] &= 0
\end{aligned} \tag{7}$$

so that if the square brackets vanish at the initial time, then they are guaranteed to vanish for all later times. The proof depends on the conservation of charge. §

Perspective 3: Fields produced by sources
The third perspective requires potentials, and only a schematic account is given here. First consider the case of electrostatics, for which it should be well known that the scalar potential Φ satisfies

$$-\nabla^2 \Phi \quad = \quad \frac{\rho}{\epsilon_0}$$

The analogous formula for the other components of the potential (denoted schematically as A in general) is of the form

$$\boxed{D\,A \sim J} \tag{8}$$

where D is some second-order differential operator (similar to $-\nabla^2$), J is the charge density or the current density, and some constants are omitted.

The solution to the problem in electrostatics is obtained by superposition:

$$\Phi(\mathbf{r}) \quad = \quad \int \frac{1}{4\pi\epsilon_0} \frac{1}{|\mathbf{r} - \mathbf{s}|}\, \rho(\mathbf{s})\, d^3 s$$

or schematically

$$\Phi \quad \sim \quad G\rho$$

where G is an operator representing the integral. Obviously we can think of G as the inverse of the operator $-\nabla^2$. In just the same way, the solution to (8) is schematically

$$\boxed{A \sim G\,J} \tag{9}$$

where G is the inverse of D in some sense

$$\boxed{G = D^{-1}} \tag{10}$$

All the above is schematic at this point, and it is not expected that students will understand them in detail now. But the point is that the potential (and therefore the field) is now given in terms of J (the source, namely the charge density and the current density). So in this perspective, we think of the fields as being produced by the sources. This point of view is appropriate to problems in radiation.

2.5 Different vectors

In many texts, especially those dealing with material media, one will see two other vectors, which for our present purpose (namely, EM in vacuum) can be taken simply to be

$$\begin{aligned} \mathbf{D} &= \epsilon_0 \, \mathbf{E} \\ \mathbf{H} &= \mu_0^{-1} \, \mathbf{B} \end{aligned} \tag{11}$$

These are mentioned only to cross reference to texts in which these symbols appear. The complications, i.e., additional terms on the RHS of (11) in the presence of material media, will not be much discussed in this course.

2.6 "Classical" electrodynamics

The title of this course and this set of lecture notes is *classical* electrodynamics, but the adjective is used with some licence. Certainly we shall not quantize the EM field (at least not formally, though we shall feel free to think about an EM wave as a bunch of photons). But we do allow the charges to be described by quantum mechanics, e.g., the electrons of an atom being described by the Schrödinger equation. It turns out that the ability to consider such situations will lead to interesting conditions, e.g., in terms of gauge transformations. In this connection, Planck's constant *will* appear.

3 Covariance

A key concept permeating these lectures is *covariance*:

There is no preferred frame. All physical phenomena can be explained by the same laws of physics in any frame.

There is a deep connection between CED and covariance. On the one hand relativistic covariance was first discovered through CED. On the other hand one can turn the argument around: the requirement of covariance places strong constraints on the laws of CED. The latter perspective is more interesting, and is now routinely used to constrain the possible laws of physics for other interactions — thus reducing these to very few possibilities to be distinguished experimentally.

Covariance will be discussed more systematically later. Here we simply analyze one phenomenon in two different frames and see what conclusions can be drawn.

Problem 6
Two parallel long wires, each of length ℓ, are a distance r apart. Each has a mass per unit length $\rho = m/\ell$ and carries a charge per unit length λ. One wire is held fixed, and the other accelerates in a transverse direction as a result of the EM force.
(a) Using Gauss' law, find the electric field E as a function of radius from the first wire and then the electric force $F_E = (\lambda\ell)E$ acting on the second wire. What is the acceleration $a = F_E/(\rho\ell)$ of the second wire?
(b) From this point onwards, go to a coordinate frame moving at a velocity v in the direction of the wire. Transverse distances are the same in the two frames, and $t' = \gamma t$ by time dilation. What is the corresponding value of the acceleration, a'? Hint: there should be a factor

$$\gamma^{-2} \quad = \quad \left(1 - v^2/c^2\right) \tag{12}$$

coming from the Lorentz transformation. §

Problem 7
Continue the previous problem, but now solve it in the moving frame.
(a) What is the length ℓ' and the charge per unit length λ' in this frame?
(b) What is the electric force F'_E acting on the second wire? (You can use the same formula as in the previous problem, but everywhere using $'$ variables.)
(c) What is the current I' due to each wire?
(d) Using Ampere's law, find the magnetic field B' as a function of radius, and hence the magnetic force $F'_B = -I'\ell'B'$. The minus sign indicates attraction between like currents.
(e) Hence find the total force F' and thus the acceleration $a' = F'/(m\gamma)$. Explain the factor γ here. Hint: The answer will involve a factor

$$\left(1 - \mu_0\epsilon_0 v^2\right) \tag{13}$$

where the second term is the ratio of the magnetic force ($\propto \mu_0$) to the electric force ($\propto \epsilon_0^{-1}$), and the magnetic force moreover involves the product of the two currents, each $\propto v$.
(f) By comparing (12) and (13), find the relation between ϵ_0, μ_0 and c. Unlike many other derivations, we have *not* used the time-dependent Maxwell equations. §

These two Problems involve a number of laws of CED: the Lorentz force law, Gauss' law and Ampere's law. These must be intimately related to one another in order to satisfy covariance, i.e., it is not possible to contemplate arbitrarily changing one of them. For example, the $1/r$ nature of the electrical force in this configuration requires the $1/r$ nature of the magnetic force, which is an essential feature of Ampere's law (e.g., in its integral formulation).

This example verifies how *one* phenomenon is consistently described in two reference frames. But we need a much stronger condition — that *every* phenomenon is consistently described in two reference frames. This harder task cannot be accomplished by examining phenomena one at a time; eventually we have to examine the *theory* and show that it has the right properties, and this will be one important focus of this course.

The more familiar derivation of $\mu_0\epsilon_0 = c^{-2}$ involves the time-dependent Maxwell equations, whose wave solutions have speed c. But the above used only electrostatics and magnetostatics.

The velocity of light

Although we have avoided EM waves and the time-dependent equations, you may object: the appearance of c (through the factor γ in the Lorentz transformation) surely relies on the concept of light, i.e., EM waves, which in turn relies on the time-dependent Maxwell equations. This line of reasoning is correct as a historical account, but not necessarily as a logical account. The privileged role of c in the theory of relativity actually has very little to do with light or EM; it is simply the maximum possible velocity of *any* particle or signal.

In fact, c is a property of spacetime itself — which defines the stage on which EM (and all other interactions) are played out. Thus it is more useful to regard it as logically prior to the theory of EM. This point is further developed through an analogy in Appendix C.

To emphasize this point, it may be useful to refer to the following (apparently outrageous) statement:

The "velocity of light", c, does not need to have anything to do with light.

This statement can be appreciated through two hypothetical scenarios.

- Suppose light and photons do not exist, or have never been discovered because we do not have instruments (and eyes) sensitive to them; in fact, suppose there are only protons and electrons in this world. We observe the collisions of protons, and find (contrary to Newton) that the sum of $m\mathbf{v}$ is not conserved; rather, careful measurements show that the sum of $m\mathbf{v}(1-v^2/c^2)^{-1/2}$ is conserved, for some constant c. This is then the parameter that appears in (12). If there are extremely accurate measurements, this discovery could in principle be made at low velocities.
- Suppose photons exist but have a non-zero mass m. (We shall discuss this possibility at length later in the course.) Then the energy E and the momentum $\mathbf{p}$ of a photon would be related as

  $$(E/c)^2 = p^2 + m^2c^2$$

  There is the extra term, but c is still the same as before. However, these particles do not travel at the speed c.

The name "velocity of light" is so common that we shall continue to use it, despite these caveats.

Electricity and magnetism

The situation described in these Problems illustrates another point: typically, magnetic effects compared to electric effects are reduced by a factor of $O(v^2/c^2)$, where v is the typical velocity of the charges producing or responding to the magnetic fields. In relation to this, several points can be made.

- We should regard c as coming first logically, setting the stage for EM. Electrical forces go as ϵ_0^{-1} and magnetic forces as relativistic corrections go as $\epsilon_0 c^{-2}$, which is then defined as μ_0. Do not think of ϵ_0 and μ_0 determining c. (The latter point of view is unfortunately reinforced by the analogous situation in a material medium, which is not covariant in the naive sense.)
- In typical laboratory situations, v^2/c^2 is extremely small.
- But in many situations, the net electrical effect is cancelled because of overall charge neutrality, so the tiny magnetic effect becomes apparent.
- There is a similar "magnetic", i.e., velocity-dependent, effect in gravity. But in most cases it tends to be a tiny correction to the dominant "electric" effect, which is never cancelled because gravity has only one sign.

Some of these aspects are illustrated by the following Problem.

Problem 8

A current of 1.0 A flows in a copper wire, whose cross section is circular, with a radius of 2 mm.
(a) What is the density n of conduction electrons? Useful data for copper: density $= 8.92 \times 10^3$ kg m^{-3}; atomic weight = 63.5; number of valence electrons = 1 per atom.

(b) Find the velocity v of the electrons. (This is the average drift velocity, and there is a larger random velocity.)
(c) Hence find v^2/c^2.
(d) Explain why the relatively small magnetic force between two current-carrying wires is so easily detected. §

4 Unification

Much of the progress of theoretical physics for the last two centuries can be described within the theme of unification, how phenomena that used to be regarded as unrelated have been brought within the same framework.

4.1 Within CED

Electricity and magnetism
Covariance points the way to the unification of electricity with magnetism. Electric fields are produced by charges; magnetic fields are produced by currents. But currents are just moving charges — and whether charges are moving or not is merely the point of view of particular observers. Thus the two must be related, and the example above (see Problems 6 and 7) is one that is easy to analyze.

Induction
However, this line of argument does not reflect the historical sequence. What happened was that Faraday discovered that a changing magnetic flux causes an EMF, i.e., some sort of electric field with a non-zero line integral around closed loops.

Displacement current
Then the reverse effect was discovered, but this time theoretically. Maxwell deduced that, in order for the theory to be consistent, a changing electric flux must also produce a magnetic field with a non-zero line integral around closed loops — in this case not much different from a current; hence a changing electric flux is called a *displacement current*. This effect is actually very small, except for situations with high rates of change, which is why the effect was not observed experimentally earlier.

With the addition of the displacement current, the Maxwell equations are complete.

EM waves
It was then discovered that the complete Maxwell equations have wave solutions, with wave speed given by $(\epsilon_0\mu_0)^{-1/2}$, whose value turned out to be the same as that of light. This led to the conjecture that light is simply EM waves, later confirmed in a number of ways.

With this step, the three phenomena of electricity, magnetism and optics are all brought under CED.

4.2 From CED to QED

The next unification is to bring electrodynamics together with quantum mechanics (QM), resulting in *quantum electrodynamics* (QED). The story here is long.

First consider classical EM fields (including EM waves) interacting with atoms and molecules described by QM. One can certainly obtain transition rates induced by an external classical field, and by a clever extension also arrive at transition rates due to the absorption or emission of a single photon.

Next incorporate wave–particle duality for photons in a systematic way. Thus, each classical oscillation mode j of the EM field is regarded as a harmonic oscillator, and its amplitude a_j is turned into an operator: essentially the lowering or raising operator for the eigenstates of a harmonic oscillator, now becoming the annihilation or creation operator for a photon state.

Then the charged particles, for example electrons, must also be treated relativistically, and in that context, one must also deal with antiparticles and processes involving them, e.g., the creation of e^+e^- pairs. To allow for the change of particle number, the amplitude for describing the matter waves likewise needs to be regarded as an operator. That is the essence of quantum field theory (QFT). The QFT involving electrons (or any other lepton with only EM interactions) is called QED.

QED is usually studied in perturbation theory, with all amplitudes expanded in powers of the fine structure constant[6]

$$\alpha = \frac{e^2}{4\pi\epsilon_0\hbar c} \approx \frac{1}{137} \tag{14}$$

But beyond lowest order, the amplitude involves a sum (or more precisely integral) over an infinite number of intermediate states, which turns out to diverge. It was found, around 1950, that these infinities[7] can all be hidden by redefining certain constants, a process known as *renormalization*. The first success of renormalization was the calculation of the anomalous magnetic moment of the electron, with the result $g = 2 + \alpha/\pi$. (The coefficient of the $O(\alpha^2)$ term would not be calculable analytically.)

An important ingredient that allows renormalization to be carried out is *gauge invariance*, a con-

[6]The denominator is not exactly an integer. The best experimental value at the time of writing is $\alpha^{-1} = 137.035\,999\,139(31)$.

[7]Importantly only a finite number of them.

cept that will be studied in this course, at least in the classical context.

4.3 Electroweak unification

It had long been speculated that the weak interaction is like EM, but with the mediating particles, called the intermediate vector bosons (the analog of photons), being (a) massive, and (b) of two varieties, charged and neutral. The coupling constant would be about the same as that for EM, i.e., $O(\alpha)$, and the apparently weak interaction is due to the short range implied by the heavy mass. The similarity of coupling constants suggests some possible unification.

While it is easy to construct a theory that works to lowest order in perturbation theory, higher-order contributions would be divergent (even if a renormalization procedure similar to that in QED is implemented) if the mediating particles have non-zero mass, which would break gauge invariance. Only in the late 1960s was a subtle mechanism found, which allows the masses of the mediating particles to be regarded as both zero or nonzero — through the Higgs field. EM and the weak interaction became unified in a renormalizable theory which has been verified to high precision.

4.4 Quantum chromodynamics

In the early 1970s, it was proposed that the strong interaction is due to the exchange of *gluons* between quarks. Quarks come in three *colors* and the gluons couple to the color degree of freedom. The theory, which is the generalization of QED to color, is called quantum chromodynamics (QCD). This is a renormalizable theory much like QED, with the additional property that the effective interaction constant α_s (the analog of the fine structure constant α) decreases logarithmically to zero for high energies or short distances, making perturbation possible. Perturbative QCD has been accurately verified experimentally.

The union of electroweak theory and QCD is now called the *Standard Model*, and works very well for all experimental situations.

All these theories grow out of electrodynamics.

5 Standard references

These lecture notes are certainly incomplete, and students may wish to refer to more extensive texts. The standard one at this level is Jackson's *Classical Electrodynamics* [8], which in earlier times every graduate student would need to learn almost from cover to cover (> 800 pages).

Two other authoritative references are Laudau's texts on *Classical Theory of Fields* [9], and on *Electrodynamics of Continuous Media* [10].

A Metric and units

Sign and unit conventions present an annoying complication to beginning students, especially when they have to consult references adopting different conventions. The teacher of a course such as the present one is torn between two desires:

- To stick to conventions familiar to beginning students — and this means the MKSA system.
- To use a set of units in which the formulas look simpler and more "natural" — and this means, for example, the relativistic Gaussian units in which $c = 1$, $\epsilon_0 = 1$ and $\mu_0 = 1$.

These lectures stick primarily to the MKSA system of units, for a number of reasons.

- Students are familiar with them. It would be disastrous if students learn beautiful formulas but cannot calculate the value of physical variables that can be measured by voltmeters or ammeters.
- The majority of the class will not go on to QED etc., and the greater convenience of Gaussian units in that setting is a false advantage for them.
- It is easy to go from MKSA units to Gaussian units (e.g., by replacing μ_0 with 1), but not the other way round (one can hardly replace 1 with μ_0).

But we shall ease the transition to simpler systems of units by expressing formulas in terms of a single constant μ_0.

These conventions are explained in this Appendix, in terms of how we deal with the three constants c, ϵ_0 and μ_0; they are of course related by

$$c^{-2} = \epsilon_0 \mu_0 \tag{15}$$

Four-vectors

The convention with four-vectors is

Follow the spatial components.

Thus the time component should be multiplied by suitable powers of c to turn it into the same units.

$$\begin{aligned} \text{coordinate} &= \vec{x} = (ct, \mathbf{x}) \\ \text{derivative} &= \vec{\partial} = (c^{-1}\partial/\partial t, \boldsymbol{\nabla}) \\ \text{momentum} &= \vec{p} = (E/c, \mathbf{p}) \end{aligned}$$

$$
\begin{aligned}
\text{current density} &= \vec{J} = (\rho c, \mathbf{J}) \\
\text{potential} &= \vec{A} = (\Phi/c, \mathbf{A})
\end{aligned}
$$

This convention ensurea that, for example, the 1-component of $\vec{x}$ and the 1-component of $\mathbf{x}$ are the same, both denoted as x^1.

A corollary of this convention is that

$$
\nabla \times \mathbf{A} \quad \mapsto \quad \partial^\mu A^\nu - \partial^\nu A^\mu
$$

which will contain components of $\mathbf{B}$ and $\mathbf{E}/c$. Thus in many formulas it is helpful to think in terms of these two quantities rather than $\mathbf{B}$ and $\mathbf{E}$. For example, the Lorentz force law can be written as

$$
\mathbf{F} \quad = \quad q\left[c\left(\mathbf{E}/c\right) + \mathbf{v} \times \mathbf{B}\right]
$$

Metric convention

Again, based on the rule of "following the spatial components", the dot product for 4-vectors are defined as, e.g.,

$$
\begin{aligned}
\vec{x} \cdot \vec{x} &= -(x^0)^2 + \mathbf{x}^2 \\
&= -(ct)^2 + \mathbf{x}^2 \\
\vec{p} \cdot \vec{p} &= -(p^0)^2 + \mathbf{p}^2 \\
&= -(E/c)^2 + \mathbf{p}^2
\end{aligned}
$$

etc. Such a *spacelike metric* is common in texts on relativity, but the opposite is common in texts on particle physics.

Using only one EM constant

Given (15) and having specified how we deal with c, it remains to deal with *only one* of ϵ_0 and μ_0. Whenever necessary, we shall eliminate ϵ_0 and write all formulas in terms of μ_0. For example, in the inhomogeneous Maxwell equations, the source terms always appear multiplied by μ_0 (and suitable powers of c), namely

$$
\begin{aligned}
\nabla \times \mathbf{B} &= \mu_0 \mathbf{J} + \ldots \\
\nabla \cdot (\mathbf{E}/c) &= \mu_0(\rho c)
\end{aligned}
$$

(Check the factors of c.)

Although electrical effects are often more important, we nevertheless choose to use μ_0 rather than ϵ_0 for two reasons.

- We adhere to the rule of "following the spatial component". The spatial component of the four-current couples as $\mu_0 \mathbf{J}$.
- This choice accords with the (old) convention in the definition of the MKSA system of units, which amounts to fixing a value of μ_0. The coulomb and hence the value of ϵ_0 follow as a consequence. (But see below for update in May 2019.)

Note added in 2019

The definition of electrical and magnetic units changed on 20 May 2019.

The *old definition*, probably the one you are familiar with, was to take the exact value

$$
\mu_0 \quad = \quad 4\pi \times 10^{-7}\ \mathrm{H\ m^{-1}} \qquad \text{exact}
$$

This in effect defines the ampere first, and the coulomb as a consequence.

The *new definition* takes the elementary charge (charge on a proton) to be

$$
q_0 \quad = \quad 1.602\,176\,634 \times 10^{-19}\ \mathrm{C} \qquad \text{exact}
$$

The value of $1/(4\pi\epsilon_0)$ is then experimentally determined, in principle by measuring the force between two electrons a given distance apart.

Unless you work to extremely high accuracy, the two approaches can be regarded as equivalent. In this course we shall continue to take the point of view of the old definition, namely, we think of μ_0 as coming first.

B Index notation

This Appendix introduces the index notation extensively used throughout the course. Only a quick review is sketched for the familiar material, mostly by way of introducing notation. Reference is made to 3-dimensional (3D) space, or more generally N-dimensional space.[8]

Vectors

A vector $\mathbf{a}$ has 3 components and can be represented as

$$
\mathbf{a} \quad = \quad (a_1, a_2, a_3)
$$

This is more convenient than the more familiar notation (a_x, a_y, a_z), especially for the obvious generalization to N dimensions.

The indices 1, 2, 3 can be written freely as subscripts or superscripts. This would not be true when we come to Minkowski space.

It will also be convenient to think of $\mathbf{a}$ as a column vector, in which case we write

$$
\mathbf{a} = [a] \quad = \quad \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}
$$

In terms of an index

Such a vector can also be specified in terms of its

[8]Note, however, that 4D space here means (a hypothetical) 4-dimensional Euclidean space, *not* space*time*. The latter is Minkowski space, where the generalization of Pythagoras' theorem has a minus sign in one term.

components in the manner

$$a_i \quad , \quad i = 1, \ldots, 3$$

Here i is a *free index* that can be assigned any value. Sometimes, as a shorthand, we say "the vector a_i", meaning "the vector whose components are a_i".

Dot product
The dot product, also called scalar product or inner product, between two vectors is[9]

$$\begin{aligned} \mathbf{a} \cdot \mathbf{b} &= a_1 b_1 + a_2 b_2 + a_3 b_3 \\ &= \sum_{i=1}^{3} a_i b_i \end{aligned} \tag{16}$$

Summation convention
Purely as a matter of convention, we drop the summation sign in (16), with the understanding that if an index (here i) is repeated in any term, it is meant to be summed over — unless explicitly specified to the contrary. Thus

$$\boxed{a_i b_i \quad \text{means} \quad \sum_i a_i b_i}$$

With the upper limit of summation deliberately omitted, the formula applies in any number of Euclidean dimensions.

With this notation

$$\mathbf{a} \cdot \mathbf{b} = a_i b_i \tag{17}$$

In formulas such as (16) or (17), the index i cannot be freely assigned any particular value; rather, it has to be summed over all possible values. Such an index is called a *dummy index.*

Matrices
Almost without exception, only square matrices will be considered, in fact, $N \times N$ matrices if vectors have N components. The elements of a matrix are denoted as, say for $N = 3$:

$$[M] = \begin{pmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{pmatrix}$$

In M_{ij}, the first index i labels the row, and the second index j labels the column. As before, these indices can be freely written as subscripts or superscripts. Sometimes, as a shorthand, we say "the matrix M_{ij}", meaning "the matrix whose elements are M_{ij}".

[9]Here we assume the vectors are real; if they are allowed to be complex, then it is necessary to conjugate the components of one vector, usually the one on the left.

The Kronecker delta

$$\delta_{ij} = \begin{cases} 1 & \text{if} \quad i = j \\ 0 & \text{if} \quad i \neq j \end{cases}$$

are the matrix elements of the identity matrix $[I]$.

Matrix multiplication
The most elementary example of matrix multiplication is (again specializing to 3D)

$$\begin{aligned} [b] &= [M]\,[a] \\ \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} &= \begin{pmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \\ b_i &= M_{ij}\, a_j \end{aligned}$$

The last form contains 3 equations, one for each choice of the free index i. Each of these contains an implied sum over the dummy index j — which must appear twice, next to each other, in order to represent matrix multiplication. The more compact index notation has the added advantage that it works in any number of dimensions.

For multiplying two matrices,

$$\begin{aligned} [C] &= [A]\,[B] \\ C_{ij} &= A_{ik}\, B_{kj} \end{aligned}$$

Note that i and j are free indices, so there are 3×3 equations. The dummy index k appears twice, right next to to each other and on the "inside", whereas the free indices appear "outside", at the extreme left and right.

What if we encounter an expression such as

$$D_{ij} = A_{ik}\, B_{jk}$$

in which the repeated index appears in positions that are not right next to each other? The trick is to write

$$B_{jk} = B^T_{kj}$$

where the superscript ^{T} indicates the transposed matrix. Then

$$\begin{aligned} D_{ij} &= A_{ik}\, B^T_{kj} \\ [D] &= [A]\,\left[B^T\right] \end{aligned}$$

Basis vectors
The basis vectors are denoted as $\hat{\mathbf{e}}_1$, $\hat{\mathbf{e}}_2$, $\hat{\mathbf{e}}_3$, instead of $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$, $\hat{\mathbf{k}}$ for more convenient generalization to higher dimensions. The notation $\hat{}$ indicates a *unit* vector. These basis vectors are orthonormal:

$$\hat{\mathbf{e}}_i \cdot \hat{\mathbf{e}}_j = \delta_{ij} \tag{18}$$

In this equation, both i and j are free indices, so that (18) contains 3×3 equations (though some of these are dependent, because of the symmetry in the two indices).

Levi-Civita symbol

The Levi-Civita symbol, with 3 indices, is defined as

$$\epsilon_{ijk} = \begin{cases} 1 & \text{if} \quad (ijk) = (123), (231), (312) \\ -1 & \text{if} \quad (ijk) = (321), (213), (132) \\ 0 & \text{if} \quad \text{any two indices are equal} \end{cases}$$

Incidentally, this symbol cannot be defined in ND with $N \neq 3$.[10]

It is left as an exercise to show that for any 3×3 matrix $[A]$,

$$\det\,[A] = \epsilon_{ijk} A_{1i} A_{2j} A_{3k}$$

which is seen to correspond to Cramer's rule: select one element from row 1, one element from row 2 and one element from row 3, and count the product with weights $+1$, -1 or 0. In fact, ϵ_{ijk} can be *defined* as the coefficient that appears when one writes out Cramer's rule.[11]

Cross product

The cross product

$$\mathbf{c} = \mathbf{a} \times \mathbf{b}$$

implies the following for the components

$$c_i = \epsilon_{ijk} a_j b_k$$

which contains 3 equations, one for each choice of i. In each of these, the RHS contains 9 terms (for all the possible choices of j and k), but most of these are zero. Suppose we take $i = 3$. Then on the RHS, (a) each of j and k cannot be 3, or else there is a repeated index in ϵ_{ijk}; (b) j and k must be distinct. Thus there are only two terms left, for $(jk) = (12)$ and $(jk) = (21)$, with the value of ϵ_{ijk} being respectively $+1$ and -1. Thus

$$c_3 = a_1 b_2 - a_2 b_1$$

as expected.

Two identities

In EM, we often have to use the cross product *twice*.[12] Thus, the product of two Levi-Civita symbols appears. There are two important identities. The first is

$$\boxed{\epsilon_{ijk}\epsilon_{imn} = \delta_{jm}\delta_{kn} - \delta_{jn}\delta_{km}} \qquad (19)$$

There are 4 free indices $jkmn$, so in principle there are $3^4 = 81$ equations in (19), of course not all independent. The repeated index i is placed in the same position in the two factors (namely, in the first position).[13] The free indices, if paired in the "natural" or corresponding order (namely, pairing second indices (jm) and third indices (kn)), would be associated with a $+$ sign; the indices, if paired in the opposite order, would be associated with a $-$ sign.

The brute-force way to prove (19) is simply to exhaust the finite number of cases (in fact 81 cases). We shall show just one example. Take $(jkmn) = (1212)$. Then

$$\text{LHS} = \sum_i \epsilon_{i12}\epsilon_{i12} = 1$$

where we have explicitly restored the summation sign and noted that only the $i = 3$ term contributes. On the other hand

$$\text{RHS} = \delta_{11}\delta_{22} - \delta_{12}\delta_{21} = 1 - 0$$

which proves the identity in this case.

The next identity involves summation over two repeated indices:

$$\boxed{\epsilon_{ijk}\epsilon_{ijn} = 2\delta_{kn}} \qquad (20)$$

To prove this, start with (19), and sum over $m = j$.

Vector identity

A trivial consequence of (19) is the vector identity

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = \mathbf{b}\,(\mathbf{c} \cdot \mathbf{a}) - \mathbf{c}\,(\mathbf{b} \cdot \mathbf{a}) \qquad (21)$$

To prove this, let $\mathbf{b} \times \mathbf{c} = \mathbf{d}$ and $\mathbf{e}$ be the LHS of (21). Then

$$\begin{aligned} & e_i = \epsilon_{ijk}\, a_j d_k \\ = \ & \epsilon_{ijk} a_j\ \epsilon_{kmn} b_m c_n \\ = \ & \epsilon_{kij}\epsilon_{kmn}\, a_j b_m c_n \\ = \ & (\delta_{im}\delta_{jn} - \delta_{in}\delta_{jm})\, a_j b_m c_n \\ = \ & b_i(\delta_{jn} a_j c_n) - c_i(\delta_{jm} a_j b_m) \\ = \ & b_i(c_n a_n) - c_i(b_m a_m) \end{aligned}$$

which is the i component of the RHS.

What else?

An important part of the formalism is the properties under rotation of axes, which will be handled in Chapter 13.

[10] A totally antisymmetric symbol with N indices can be defined.

[11] Incidentally, this should give a hint as to how to define a totally antisymmetric symbol in N dimensions.

[12] When the right-hand rule is used an even number of times, there would be no effect if we use the left-hand rule instead; left is the same as right, and parity is conserved. EM does conserve parity, though the weak interaction does not.

[13] Otherwise permute the order of indices in one factor until the repeated index appears in the same position.

C Analogy

The following question has also been asked: In gravity there is a constant analogous to ϵ_0, in fact related to G. If we do very careful measurements, we will find velocity-dependent interactions (the "magnetic" part of gravity) with an interaction strength analogous to μ_0. When we combine these two constants (call them ϵ_0' and μ_0') in the analogous wave equation,[14] how do we guarantee that the wave speed thus determined has the exact same value of c (as almost every theorist would insist)? If that should turn out to be the case, would it not be a huge coincidence? This difficulty arises because these constants (ϵ_0, μ_0 or their gravitational counterparts) are implicitly regarded as logically prior to c — a point of view that is unfortunately emphasized by historical accounts.

In fact c is a property of the "stage" and logically prior to the electrical constant ϵ_0 and the magnetic constant μ_0. In order to explain this point of view, an analogy is presented. A stage has an unknown width x. Two players Mr E and Mr M walk onto the stage, and stand at some point together. Mr E walks $y = 11.0$ m to the left and comes to the edge of the stage. Mr M walks $z = 5.9$ m to the right and comes to the other edge of the stage. Thus we conclude that the width of the stage is

$$x = y + z = 16.9 \qquad (22)$$

- Surely you would not say that the width of the stage is 16.9 *because* of the measurements by Mr E and Mr M.
- Suppose another pair of players G and H (representing Newtonian gravity and the additional velocity-dependent interactions) go through the same measurements, one going to the left, the other to the right, giving measurements y' and z'. Surely you would not query whether the y' and z' would add up to the same $x = 16.9$.
- In both cases, the width of the stage is logically prior, and independent of the players or their measurements. Of the two distances (one to the left and one to the right), it suffices to specify just one.

In fact, in MKSA units

$$\begin{aligned} \log \epsilon_0^{-1} &= 11.0 \\ \log \mu_0^{-1} &= 5.9 \\ \log c^2 &= 16.9 \end{aligned}$$

[14]This becomes less hypothetical now that gravitational waves have been discovered. However, gravity is more complicated than this simple account suggests, but the idea that there are extra constants describing velocity-dependent interactions is correct.

and (22) is just

$$\log c^2 = \log \epsilon_0^{-1} + \log \mu_0^{-1}$$

which completes the analogy.

D Supplement: Departure from Coulomb's law

Each supplement will consist of one or more problems, usually more challenging and of an open nature, to stretch the better students.

D.1 Define the problem

The key property of Coulomb's law is that the force goes as r^{-2} or the potential goes as r^{-1}. How good is the experimental evidence for this?

Power-law parametrization
One way to ask the question is to assume

$$\begin{aligned} F &\propto r^{-2-\varepsilon} \\ \Phi &\propto r^{-1-\varepsilon} \end{aligned}$$

and try to put a limit on ε.

Yukawa parametrization
Alternately assume

$$\Phi \propto r^{-1} \exp(-\mu r)$$

where $\mu = a^{-1}$. In this case define

$$\varepsilon = (\mu R)^2$$

where R is some characteristic length. Again, what can we say about ε? Note that we claim the effect starts at μ^2.

Jackson [8] quotes modern experiments that put a limit on the order of

$$|\varepsilon| < 10^{-16}$$

If interpreted in terms of the Yukawa potential, such results can also be regarded as placing an upper limit on the photon mass; a recent review is given in Ref. [11].

D.2 Null experiment

How is it possible to obtain such accuracy? The trick is to do a null experiment, i.e., to measure an effect which should vanish if $\varepsilon = 0$.

Take two concentric conducting spherical shells, of radii R and $R' = \rho R$, $0 < \rho < 1$. Charge the outer one to voltage V and measure the potential difference v between the two shells. (Sign convention: $v > 0$ if inner shell is at higher potential.) Define the dimensionless experimental number

$$\delta = \frac{v}{V}$$

It is a well-known result (even to freshmen) that

$$\boxed{\varepsilon = 0 \;\Rightarrow\; \delta = 0}$$

Therefore we expect, to first order (and for both the power-law case and the Yukawa case)

$$\boxed{\delta = f_1(\rho)\,\varepsilon}$$

where the proportionality constant f_1 can only depend on ρ for dimensional reasons, and is moreover expected to be $O(1)$. Measurements give

$$V \approx 10^4 \text{ V}\ ,\ |v| < 10^{-12} \text{ V}$$

so that $|\delta| < 10^{-16}$ and hence a similar limit on ε (assuming $f_1 = O(1)$).

Required calculation
The problem is to derive f_1 for both the power law case and the Yukawa case.

E Supplement: Theorem in vector calculus

Each supplement will consist of one or more problems, usually more challenging and of an open nature, to stretch the better students.

It is said in Section 2.4 that a vector field is determined by its divergence and curl. Prove this theorem, assuming that the vector field in question vanishes sufficiently rapidly at infinity.

References

[1] See e.g.,
`http://sciencepark.etacude.com/particle/forces.php`
or
`http://hyperphysics.phy-astr.gsu.edu/hbase/forces/funfor.html`

[2] RA Hulse and JH Taylor, "Discovery of a pulsar in a binary system", ApJ **195**, L51 (1975).
DOI: 10.1086/181708
An online version can be obtained at
`http://adsabs.harvard.edu/full/1975ApJ...195L..51H`

[3] JM Weisberg, JH Taylor, "Relativistic binary pulsar B1913+16: thirty years of observations and analysis", APS Conference Proceedings, (2004). The preprint version can be downloaded from
`https://arxiv.org/pdf/astro-ph/0407149v1.pdf`

[4] Cardiff University, Prifysgol Caerdydd School of Physics and Astronomy, "The Hulse-Taylor pulsar — evidence of gravitational waves" (nd), read on 2 June 2016.
`www.astro.cardiff.ac.uk/research/gravity/tutorial/?page=3thehulsetaylor`

[5] BP Abbott et al. (LIGO Scientific Collaboration and Virgo Collaboration), "Observation of gravitational waves from a binary black hole merger", Phys. Rev. Lett. **116**, 061102 (2016).

[6] BP Abbott et al. (LIGO Scientific Collaboration and Virgo Collaboration), "GW151226: Observation of gravitational waves from a 22-solar-mass binary black hole coalescence", Phys. Rev. Lett. **116**, 241103 (2016).

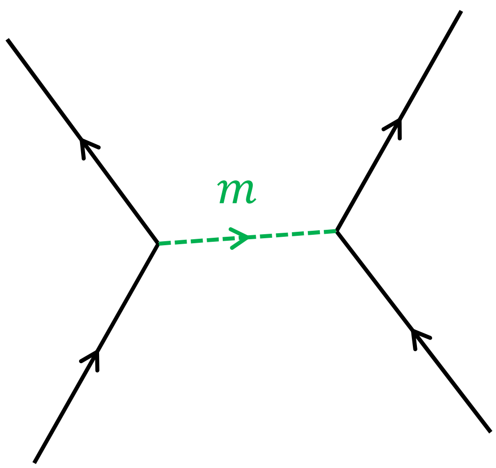
[7] The Shaw Prize Foundation, Press Release (2016).
`http://www.shawprize.org/en/shaw.php?tmp=3&twoid=102&threeid=254&fourid=478`

[8] JD Jackson, *Classical Electrodynamics*, 3rd Edition, Wiley (1998).
ISBN: 978-0-471-30932-1.
An online edition (not sure whether it breaches copyright) is available at
`www.fisica.unlp.edu.ar/materias/electromagnetismo-licenciatura-en-fisica-medica/electromagnetismo-material-adicional/`
or from
`https://www.scribd.com/doc/48520397/Jackson-Classical-Electrodynamics-3rd-edition`
The first edition (1962) can be downloaded from
`https://archive.org/details/ClassicalElectrodynamics`

[9] LD Landau, *The Classical Theory of Fields* (Fourth Edition), Volume 2 in Course of Theoretical Physics, translated from the Russian by M Hamermesh, Pergamon Press (1975).
ISBN: 978-0-08-025072-4
An online version can be obtained at
`http://www.sciencedirect.com/science/book/9780080250724`

[10] LD Landau and EM Lifshitz, *Electrodynamics of Continuous Media* (Second Edition), Volume 8 in Course of Theoretical Physics, translated from the Russian by JB Sykes, JS Bell and MJ Kearsley, Pergamon Press (1984).
An online version can be obtained at
`http://www.sciencedirect.com/science/book/9780080302751`

[11] AS Goldhaber and MM Nieto, " Photon and graviton mass limits", Rev. Mod.Phys. **82**, 940 (2010).
DOI: 10.1103/RevModPhys.82.939



$$\Delta E \sim mc^2$$

$$\Delta t \sim \frac{\hbar}{\Delta E} \sim \frac{\hbar}{mc^2}$$

$$a \sim c\,\Delta t \sim \frac{\hbar}{mc}$$

Figure 1

Range of potential