Physics 203 Lecture 1

Electric Field and Potential

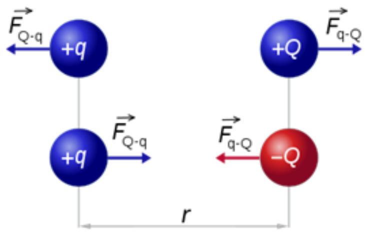
- This term we move solidly beyond Newtonian mechanics.
- Typically "modern physics" means relativity and quantum mechanics, but I think it makes sense to include electromagnetism also.
- Another theme for the term is the microscopic source of force.
- The term splits pretty even down the middle:
  - In the second half we drive into the microscopic and subatomic by picking up quantum mechanics and high energy physics.
  - Now we start with the study of electromagnetism which will drive us to consider relativity as well.

Long Range Forces And Fields

Magnetism **Gravity** **Electricity**

- Physicists have always had a love/hate relationship with long-range forces. There is something deep in the human psyche that resists the idea of any kind of influence that acts "at-a-distance".
- This resistance manifests itself in many ways: "nature abhors a vacuum" (which goes back to Aristotle), Descartes filled the solar system with "fluid" to describe the motion of the planets. We also have "animal magnetism" and telekinetics.
- In modern physics we still have the differential form of physical equations and local conservation laws.
- There are three basic long-range forces for us to consider: magnetism, gravity, and electricity.
- We now know that magnetism (the first to be studied as such) is a relativistic manifestation of electricity. This realization comes late in our story in the 1800s with Faraday and Maxwell. Inspired by this, Einstein reworked Newton's theory of gravity.
- Both electromagnetism and general relativity are based on the idea of the **force field** introduced by Faraday.
- The field allows us to have our cake and eat it too. The force field acts as the intermediate agent between the source of force (an electric charge or gravitational mass) and the object influenced.
- One unforeseen result of this field idea was that these investigations of electromagnetism were unshackled from classical concepts. In this way Maxwell's equations uncovered the weaknesses of Newtonian mechanics and lead to relativity.

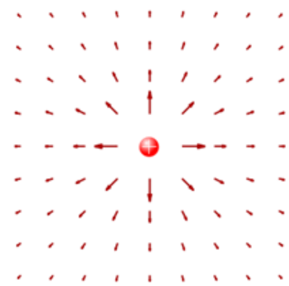
Coulomb's Law for Electric Charge



$$F = \frac{kQq}{r^2} = \frac{1}{4\pi\epsilon_0}\frac{Qq}{r^2}$$

$$F = qE$$

$$E = \sum \frac{1}{4\pi\epsilon_0}\frac{Q}{r^2}$$



- If Einstein was later inspired by Maxwell to discover the gravitational field equations, earlier in history Coulomb was inspired by Newton's law of gravity when measuring static electricity.
- Coulomb found that the force of electricty obeys an inverse-square law:

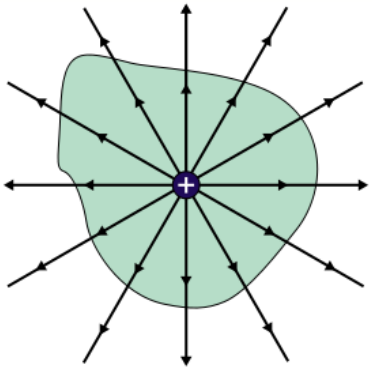
$$F = \frac{kQq}{r^2} = \frac{1}{4\pi\epsilon_0}\frac{Qq}{r^2}$$

- You will see both of these forms of the proportionality constant as we move through the term. The SI unit for electric charge is the **coulomb**. Typical values are in the micro range (10^{-6}).
- If there are multiple sources, we simply add the forces like vectors.
- Notice that in each of these terms will be present the value of the charge under the influence. We can factor this out and the force splits into two pieces: one related to the charge and the second related to the sources.
- This second piece we call the **electric field**. Thus,

$$F = qE \qquad \text{and} \qquad E = \sum \frac{1}{4\pi\epsilon_0}\frac{Q}{r^2}$$

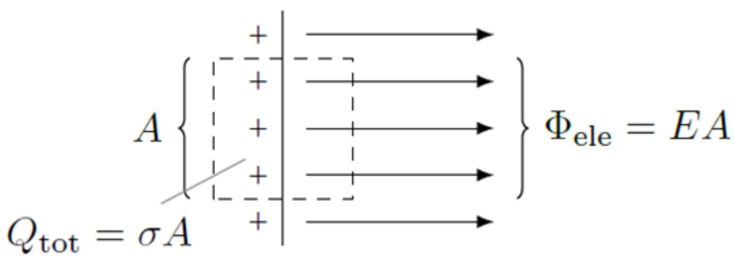
- Unlike gravity, electricty is both attractive and repulsive. As a consequence, we must have at least two kinds of electric charge: positive and negative.
- One nice aspect of Coulomb's law is that it still works: charges with the same sign attract and charges with different signs repel.

Gauss' Law and Electric Flux



$$\Phi_{\text{ele}} = \sum E_{\text{avg}} A \cos\theta$$

$$\sum \Phi_{\text{ele}} = Q_{\text{tot}}/\epsilon_0$$



- The electric field starts to split every electromagnetic interaction into two problems: how the field is created, and how the field influences those things around it. In principle, we already know the answers: $E = kQq/r^2$ and $F = qE$. But as we progress, we will tweak these answers—right now they only work for the special case of charges at rest.
- When Faraday first introduced the idea of the electric and magnetic fields, he was thinking in terms of **field lines**. These are mathematically very similar the streamlines we talked about when studying fluid flow. (They may be conceptually similar also: in quantum field theory, the EM field is modeled as a flood of quantum photons.)
- Each electric charges is either a source or sink of these electric field lines: positive charges are sources and negative charges are sinks. The field lines radiate from these charges and distribute themselves evenly throughout space. The density of the field lines represent the magnitude of the electric field at that point.
- We call the **electric flux** is the "number" of field lines running through any given surface. We have:

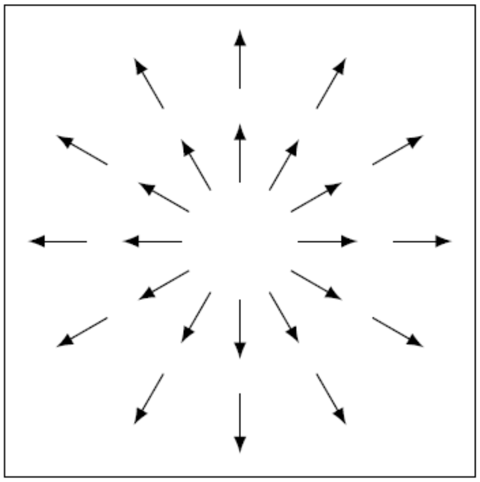
$$\Phi_{\text{ele}} = \sum E_{\text{avg}} A \cos\theta$$

  where Φ is the electric flux and $E_{\text{avg}} \cos\theta$ is the average value of the electric field running perpendicular through the surface A.
- $Gauss'Law$ states that the value of the electric flux through any closed surface is directly proportional to the amount of charge enclosed by the surface:

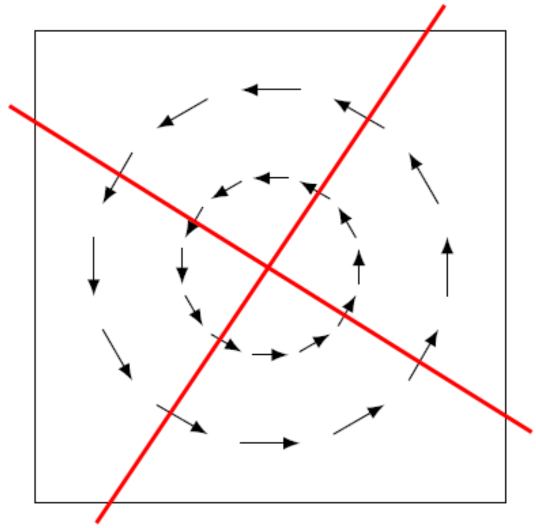
$$\sum \Phi_{\text{ele}} = Q_{\text{tot}}/\epsilon_0$$

- This actually follows from Coulomb's law and that fact that the surface area of a sphere is $4\pi r^2$.

Divergence and Curl



$$\nabla \cdot E = \rho/\epsilon_0$$



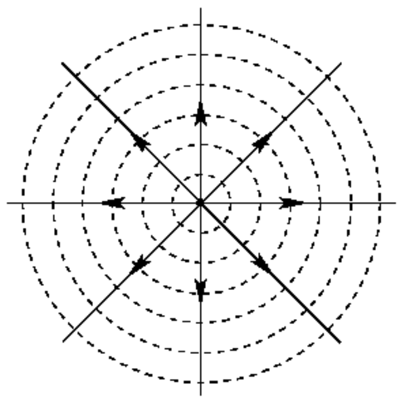
$$\nabla \times E = 0$$

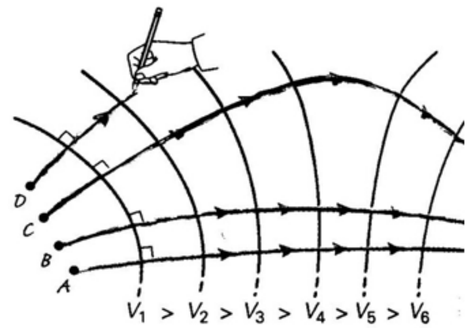
- Back in our study of streamlines in fluid flow, we introduced a number of concepts that are directly applicable to our study of electric and magnetic fields.
  - For example, the mass flow rate in the context of stream lines is equivalent to the electric flux here.
- Two critical tools for the analysis of these fields in general are divergence and curl.
  - **Divergence** describes how much the field spreads. Symbolized by $\nabla \cdot E$.
  - **Curl** describes how much the field circulates. Symbolized by $\nabla \times E$.
- The divergence is closely related to
  - The equation of continuity: if there is no charge, what goes in must come out.
  - Also Gauss' law: the field spreads at positive charge and is collected at the negative charges.
- When the curl is zero (we called this "irrotational" when dealing with fluids), a field function exists which is similar to the potential energy function for a conservative force.
- Coulomb's law written in this new format is as follows:

$$\nabla \cdot E = \rho/\epsilon_0 \qquad \text{and} \qquad \nabla \times E = 0$$

- The curl is zero because Coulomb's law is a central force and does not introduce any angular momentum into the field flux. The divergence is proportional to the charge density, ρ, at that point.
- Maxwell's laws of electromagnetism refine and correct these equations for charges in motion.

Electric Potential





$$\Delta V = -E_{\text{avg}} \Delta x$$

$$V = \sum \frac{1}{4\pi\epsilon_0} \frac{Q}{r}$$

$$PE = qV$$

- Because the curl of the electric field is zero everywhere, the circulation around any closed circuit is also zero. From this it follows that an electric potential function can be used—in exactly the same way that a potential energy function exists for conservative forces.
- We call this electric potential **voltage**. The electric voltage from a set of point sources is

$$V = \sum \frac{1}{4\pi\epsilon_0} \frac{Q}{r}$$

- The overwhelming advantage with using voltage is that it is not a vector. In the same way in which energy problems are mathematically simpler than force problems, the same is true with electric potential. When you can, use it!
- In general, the electric potential is related to the electric field via

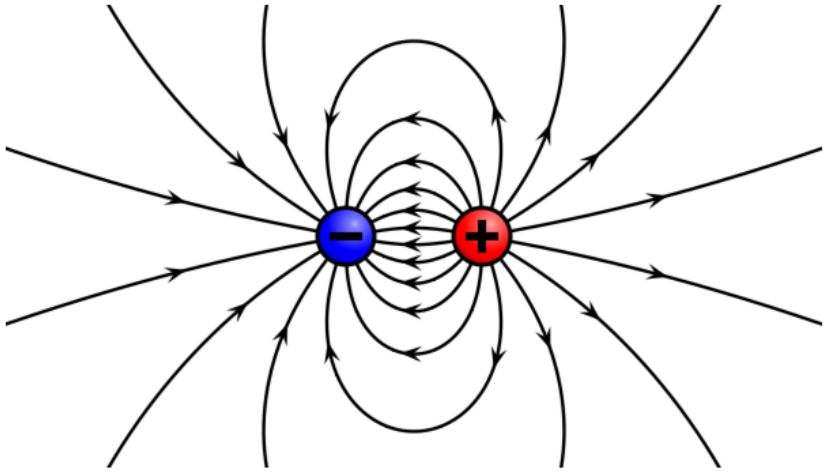
$$\Delta V = -E_{\text{avg}} \Delta x$$

- Notice that, really, it is the potential *difference* that is defined by the field, and this difference is the physically relevant quantity. In other words, the zero level for the voltage is arbitrary—which means we can choose to put it somewhere else if it helps.
- We also have a simple way to talk about electric potential energy. When a charge of $+q$ is in an electric potential V, its potential energy is simply

$$PE = qV$$

  which follows from $F = qE$ and the definitions of the electric potential and potential energy.

Neutral Atoms and Dipole Moments



$$V = \frac{kp}{r^2}\cos\theta$$

$$\tau = pE\sin\theta$$

$$p = qd$$

- Coulomb was originally inspired by Newton's law of gravity when he discovered the law of electric force. However, the magnitude of these forces are vastly different:

$$G = 6.673 \times 10^{-11} \qquad \text{and} \qquad \frac{1}{4\pi\epsilon_0} = 8.9876 \times 10^{9}$$

- The electric force is roughly 10^{20} times stronger than gravity. We experience gravity every day we try and get out of bed. Why then is the electric force only manifest in dramatic bursts of energy like lightning or parlor tricks of static electricity?

- The short answer is that most things are electrically neutral. No electric charge, no electric force.

- But there is more to this story. All three building blocks of matter (two quarks and the electron) are electrically charged. But because there are *two* types of charge, they combine together in an electrically balanced atom.

- In a way, the electric force is shackled with its own strength: these charged particles lock together and shield the rest of the universe from the electric charges running around inside. Gravity accumulates but electricity does not.

- The simplest neutral system of charged particles is called a dipole: one positive and one negative charge of equal magnitude. We define its **dipole moment** as

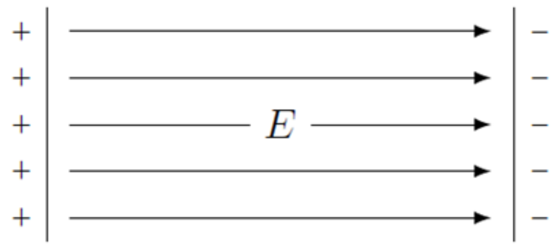
$$p = qd$$

  where q is the magnitude of the charges and d is the distance between them.

- Although dipoles are neutral, they can be attracted to one another if they get close enough: the force is proportional to the separation cubed and their relative orientation.

- Ultimately, it is this dipole-dipole interaction which explains *every* macroscopic force other than gravity: the elastic force, friction, contact forces, cohesion, etc. You can't sit in your chair without it.

Conductors, Insulators, and Capacitors



$$E = \frac{\sigma}{\epsilon_0} \qquad \Delta V = \frac{Qd}{\epsilon_0 A} \qquad C = \frac{Q}{V} = \frac{\epsilon_0 A}{d}$$

- Different materials respond differently in the presence of the electric force. On the one side are **conductors** which allow electric charge to flow with ease. On the other side are **insulators** which strongly resist the flow of electric charge. We can use conductors to transport charge and insulators to hold it.
- Since charge flows easily in a conductor, any excess charge will be pushed to the surface and spread evenly througout. If we take another conductor and touch a charged one, the excess charge will spread to it also. In this way we can "pick up" the charge and move it.
- There is a trickier way to "induce" charge in a conductor. Suppose we place a positively charged object close to a neutral conductor. This will pull the negative charge close and push the postive charge away. If we allow the postive charge to drain away, the conductor is left with a net negative charge.
- One important configuration to consider is the **capacitor**, which is simply two parallel conductive plates separated by an insulator. A capacitor "traps" electric charge: one plate with positive charge and the other with negative charge. These two sets of charge will attract, but the insulator keeps them apart.
- The charge on each plate distributes itself: we end up with a consistent charge per unit area (symbol = σ). There is a strong electric field inside the capacitor due to the strong attractive force between the plates.
- If we assume the plates are infinitely large (practically this means we ignore any edge effects), we can use symmetry and Gauss' law to determine the magnitude of the electric field and resulting potential:

  $$E = \frac{\sigma}{\epsilon_0} \qquad \text{and} \qquad \Delta V = \frac{Qd}{\epsilon_0 A}$$

  for a capacitor with plate distance d with a total charge Q.
- The ratio of the charge required to maintain a particular voltage on a capacitor is called its **capacitance** and is primarily related to its geometry:

  $$C = \frac{Q}{V} = \frac{\epsilon_0 A}{d}$$

Electric Energy

$$PE = \frac{1}{2}CV^2$$

$$u = \frac{1}{2}\epsilon_0 E^2$$

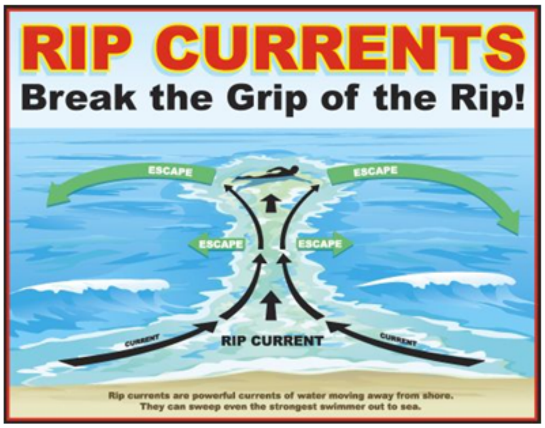
- The previous analysis for the capacitor ignores the presence of the insulator which actually increases the capacitance. This is because the plates attract the bound opposite charges in the insulator effectively decreasing the distance d. The factor of improvement for the insulating material is called its **dielectric constant** (symbol = κ).
- On the other hand, we have ignored the "edge effects" which will be significant for any real capacitor. The SI unit for capacitance is the **farad**. Typical values for capacitors are microfarads (10^{-6}) and picofarads (10^{-9}).
- The amount of potential energy created when charging a capacitor is
$$PE = \frac{1}{2}CV^2$$
- This follows from the formula for potential energy qV and the definition of capacitance $Q = CV$. The factor of one-half is there because the potential energy builds up with each incremental charge. The voltage is not constant rather it increases throughout.
- One nice thing about the capacitor is that on the inside the electric field is constant. This gives us an easy situation to analyze the electric field.
- For example, consider a vacuum filled capacitor without any insulator. We can use our result from Gauss' law to rewrite the potential energy formula as:
$$PE = (\frac{1}{2}\epsilon_0 E^2)(Ad)$$
- But Ad is simply the volume contained within the capacitor (I'm trying to avoid the letter V). We define the **energy density** of the field as this potential energy divided by this volume. Thus,
$$u = \frac{1}{2}\epsilon_0 E^2$$
- Though we only derived this formula for this special case, this formula is true in general. This represents the energy carried by the electric field itself.

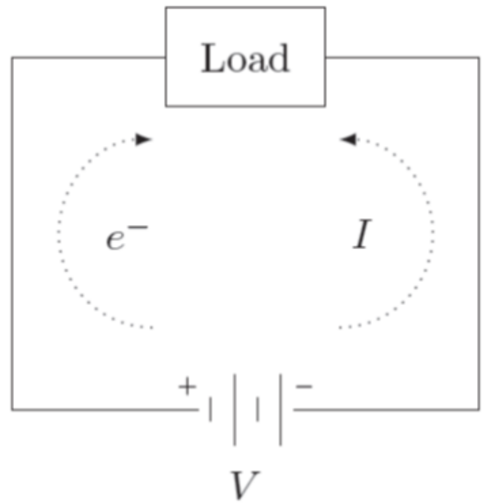
Physics 203 Lecture 2

DC Circuits and Magnetism

- The applications of the electric force are extremely diverse. But—hands down—the most important application is in electricity and electronics. This lecture will introduce us to some basics.
- The central idea of electric current also plays an important role in magnetism. Though the first to be studied in history, magnetism is the least fundamental. We will see that it comes from electric charge in motion.

EMF Drives Current





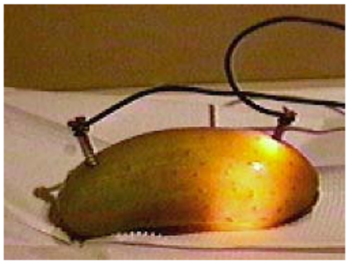
$$I = \Delta q/\Delta t$$

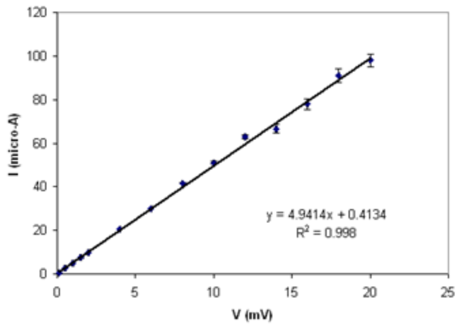
- Electric **current** is defined as the flow of electic charge. The formulas and concepts we have been using for fluid flow (also heat flow and diffusion) are directly applicable here. In general, Maxwell's laws of electromagnetism are best thought of as being driven by a *fluid* of charge.
- But we will almost exclusively be thinking in terms of electric circuits in which the electric current runs through thin wires. In that context it's best to define current as the amount of charge that flows per second:

$$I = \Delta q/\Delta t$$

- Pressure drives fluid flow, and the corresponding quantity for electricity is called the **electromotive force**, or EMF. This is really a misnomer because it is measured in volts, not newtons.
- A simple source of EMF is the common battery (simple in the context of electronics—the chemistry involved can get pretty complex). An ideal battery will provide a constant source of EMF indefintely, no matter what the load. This is called **direct current**, or "DC" for short.
- In a way, one can consider a battery as a capacitor with infinite capacitance: it supports a voltage difference with an infinte supply of charge. This is why in electric schematics a battery is represented as a couple of capacitors back-to-back.
- In the 1700s when people were just starting to investigate electricity, it was not known which direction the charge was flowing. Franklin had to guess—unfortunately he guessed wrong. This means that the **conventional current** (which flows from high to low voltage) flows in the opposite direction of the actual current (the momentum).
- In electronics problems it doesn't really matter which way the current truly flows. As long as we are consistent, we are okay. However, when we step outside of electronics proper we may have to remember this situation—but usually not even then. The negative charge of the electron will flip all the right signs for us. The math will keep track of it.

Resistors and Ohm's Law

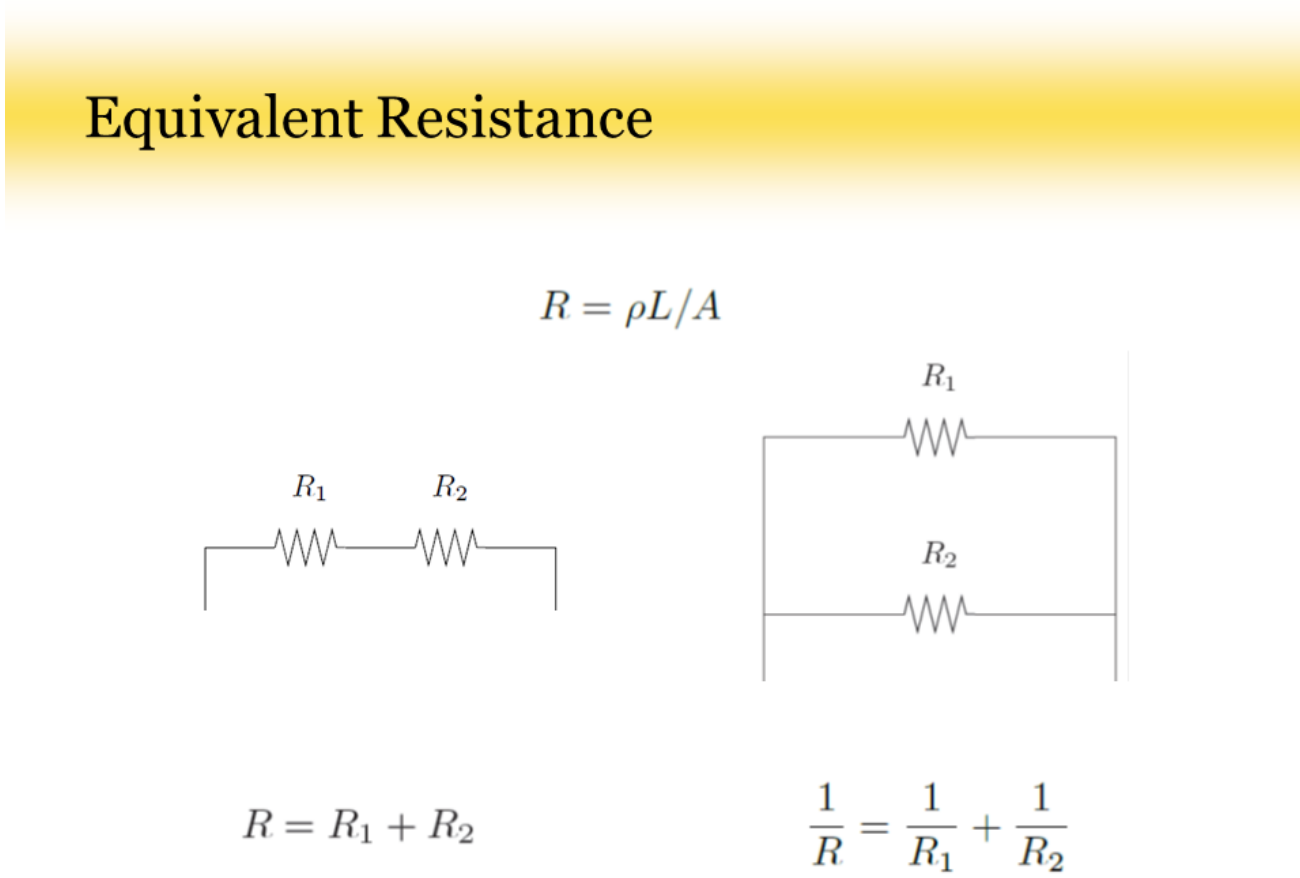




$P = IV$ Always true

$V = IR$ Not always true

- Whenever current flows between two points, it does so because of the voltage difference between them. The power required to maintain this flow of current is
$$P = IV$$
which follows from $PE = qV$ and the definitions of current and power.
- Physically, this is a manifestation of Newton's second law. As such, we expect the charges to accelerate down the electric potential. This is exactly what happens in an old-fashioned CRT computer display. Also, in you local neighborhood particle accelerator (there are more than you think: that is what an MRI and PET scan needs them).
- But in electric circuits, this does not happen for two reasons. The first is that the electrical influence occurs at the speed of light. Not because the electrons move that fast (their drift speed is rather slow) but because the wire is full of electrons. The wire is like a filled water hose: you don't have to wait for the water from the spigot—it generates pressure that pushes through the hose to where the water comes out.
- As such, when we flip a switch in an electric circuit, we can assume that the current appears instantly.
- The other reason that the electrons don't accelerate through the circuit is resisitivity, or electrical friction. Just like a air drag, resistivity produces a kind of terminal velocity for the current. For this reason, the current is stable throughout the circuit.
- Like air drag, the details of how current is related to voltage can be difficult to determine from first principles. The simplest situation is when the current is directly proportional to voltage. An electrical component for which this is true over a wide range of voltage is called a **resistor**.
- By definition, then, **Ohm's law** holds for a pure resistor:
$$V = IR$$
- The SI unit for resistance is the **ohm** (symbol = Ω).



- The instrinsic property associated with resistance is called **resistivity**, which has a tendency to slightly increase with tempertaure. The total resistance is a combination of the resistivity and the geometry of the resistor:

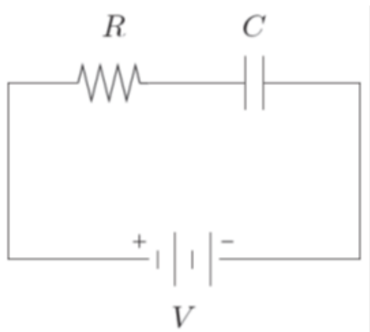
$$R = \rho L/A$$

- If we take two resistors and place them end to end, so that the current from one is forced to go through the next, they are said to be **in series**. Their total resistance is the sum of the resistances since we are essentially adding their lengths. By Ohm's law, the total voltage is split between the resistors in proportion to each resistance.
- The other way to combine two resistors is **in parallel**. In this case, the current is split between them. They share the same voltage because conductors have no resistance, so there is no voltage drop.
- Reasoning again from Ohm's law, one can see that the total resistance of two resistors in parallel is given by

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} \implies R = \frac{R_1 R_2}{R_1 + R_2}$$

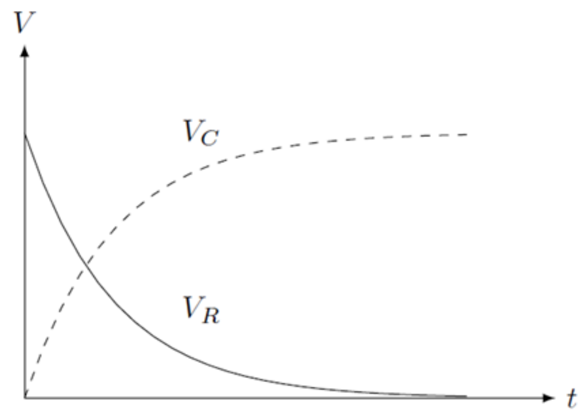
- This formula is also a consequence of the resistivity formula above since parallel resistors are effectively increasing the cross-sectional area for the current.
- Combinations of multiple resistors can often be grouped in pairs either in series or in parallel. In this way one can determine the total resistive load of a circuit.
- But not all circuits can be analyzed this way. The simplest example is the **Wheatstone bridge** with five resistors. When this happens, one must fall back on two fundamental rules called **Kirchoff's laws**:
  - The total current coming out of a split must equal the total current coming in. This is essentially a statement of the conservation of charge.
  - Around every closed loop, the sum of all the voltage drops must equal the sum of all the voltage sources. This comes from the conservation of energy (also because the circulaton of the electric field is zero).

RC Circuits



$$I(t) = I_0 \exp(-t/RC)$$

$$I_0 = Q/RC$$



- Resistors are actually the second electronic component we have studied so far—the first is the capacitor. What happens if we combine them?
- If the capacitor is not charged, nothing will happen since there will be no voltage to drive current. But when a resistor is attached across a charged capacitor, charge will flow.
- Initially, the voltage in the capacitor is given by $V = Q/C$. This voltage across the resistor drives current that follows Ohm's law, so $I = V/R$. The initial current flow is therefore

$$I = Q/RC$$

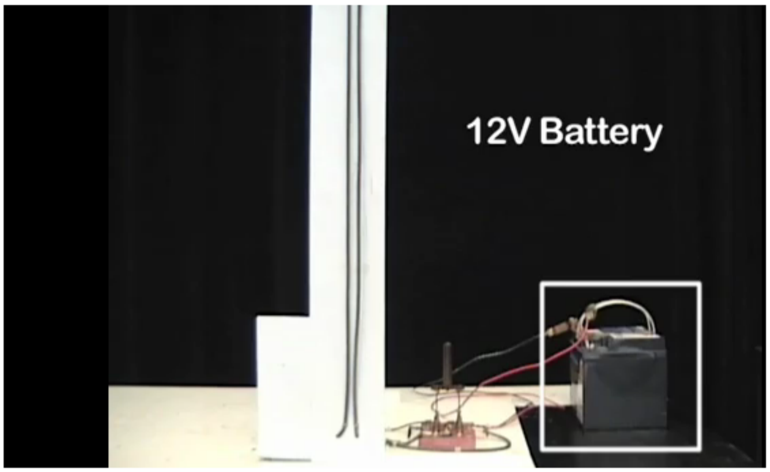
- But the charge on the capacitor drops over time—that is what current means, after all. Because the current is proportional to the charge remaining, the current will also drop over time. Mathematically, we are describing exponential decay:

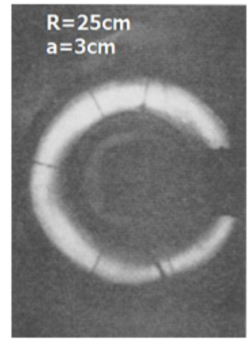
$$I(t) = I_0 \exp(-t/RC)$$

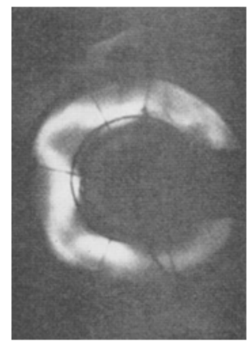
- The quantity RC is sometimes called the **time constant** for the circuit. After three time constants, the current in the circuit is $I_0 \exp(-3)$, or about 5% of its original value. After five time constants, the capacitor is effectively discharged.
- The reverse of all this occurs when we charge the capacitor back up. Our voltage source (presumably a battery) provides the EMF to drive current through the resistor according to Ohm's law.
- As charge accumulates on the plates of the capacitor, its voltage starts to oppose the source voltage. Since less voltage drops across the resistor, less current flows.
- The current again suffers exponential decay as the capacitor charges up. After about five time constants, the capacitor voltage will be nearly identical to the source voltage.

Parallel Currents Attract?

$$F = \frac{\mu_0}{2\pi}\frac{I_1 I_2}{r}$$



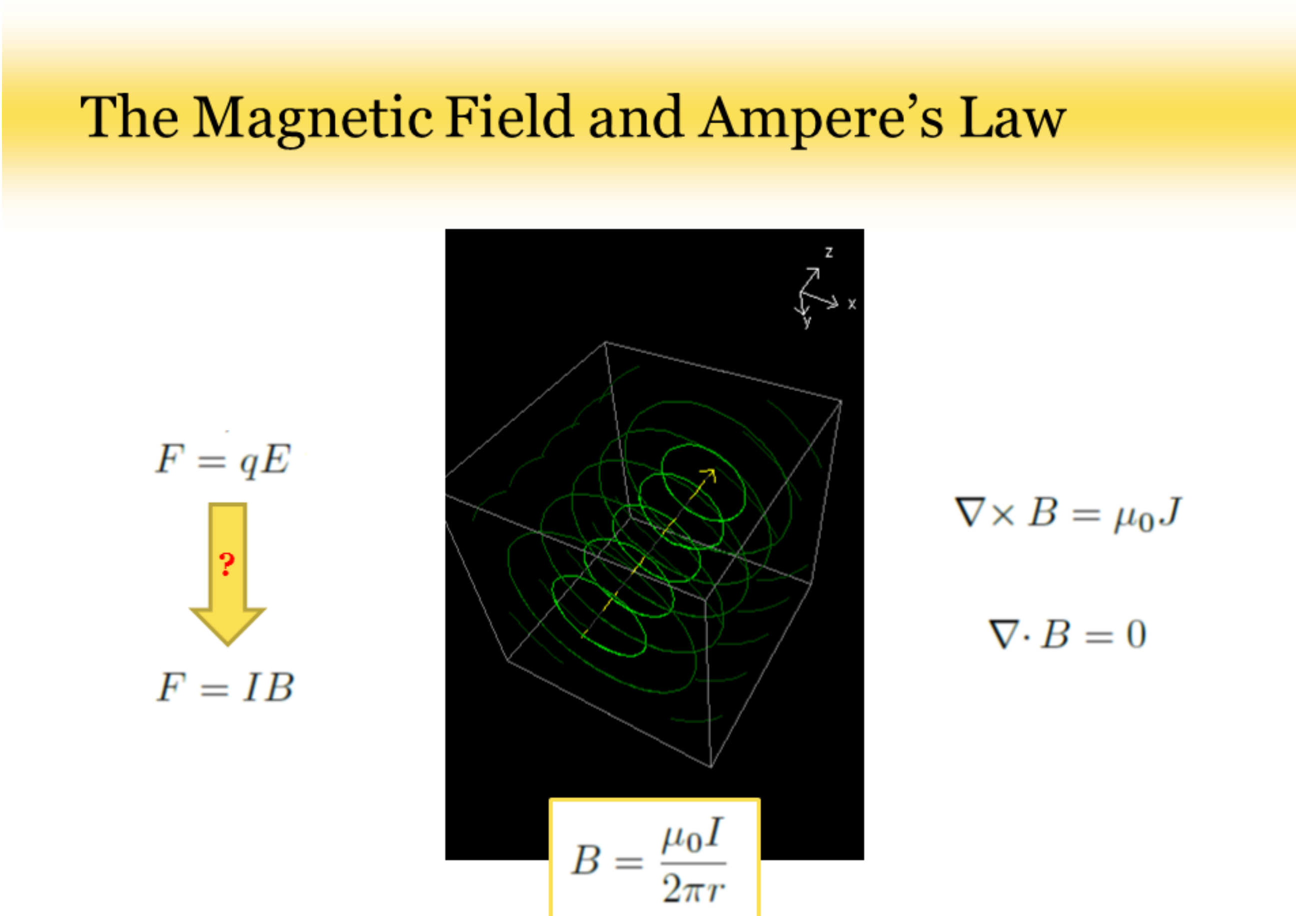




- I've left out one important fact about DC current that is impossible to anticipate. So far, our electric force is identical to Newtonian gravity except for its two types of charge. In the end, this new fact will force us to rebuild electrodynamics and Newtonian mechanics from the ground up. Here it is:
  - Parallel lines of current attract and opposing lines of current repel.
- Frankly, this ought not be. When current flows through a wire, it is still electrically neutral because the flow of charge is steady. This new force is therefore not electric. But we do have a term for it: it is the **magnetic force**.
- The force per unit length between two parallel lines of current is

$$F = \frac{\mu_0}{2\pi}\frac{I_1 I_2}{r}$$

- In fact, we use this specific situation to define the SI unit of current (and the proportionality constant μ_0):
  - If two parallel wires one meter apart each carry a current of one ampere, then the force per unit length on each wire is 2×10^{-7} N/m.
  - The SI unit of charge (the coulomb) is derived from this definition, not the other way around.
- The rule corresponding to Coulomb's law for current is called the **Biot-Savart law**. The problem with approaching the magnetic force this way is that one cannot really talk about a "piece" of current: the requirements of steady flow introduce a number of mathematical difficulties. We will skip it.
- This force happens to be one of the main difficulties to overcome in the development of fusion energy. One of the most promising approaches to fusion is by using high-temperature plasma. Since plasma is electrically active, its flow is extremely unwieldy. Its motion causes the plasma to pinch, twist, and braid itself in complicated ways.



- There is another difference between the magnetic force between current and the electric force between charges. The former involves orientation: if we flip the direction of one current, the force repels.
- This means the magnetic "source" is a vector, and the magnetic force is an interaction between vectors. We seek an analogy with the electrostatic formula $F = qE$ to define the magnetic field. Something like:

$$F = IB$$

- But we cannot "divide both sides" by I because it is a vector. What you are looking at is a linear function between vectors. Mathematically, the magnetic field is an example of a "tensor".
- However, we get a lucky break here. Because of Newton's third law, this tensor is "skew-symmetric". This means we can represent the field as circular rings surrounding the current. The direction of this circulation is arbitrary. By convention we use the **right-hand-rule**:
  - Point your right thumb in the direction of the current. Curl your fingers into a half-fist. The magnetic field lines curl in the same direction as your fingers.
- The magnitude of the magnetic field at a distance r around an infinite line of current is:

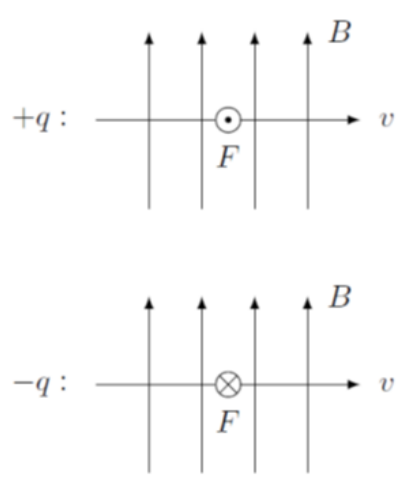
$$B = \frac{\mu_0 I}{2\pi r}$$

- Remember divergence and curl? For the magnetic field we have

$$\nabla\cdot B = 0 \qquad \text{and} \qquad \nabla\times B = \mu_0 J$$

where J is the current density, or the charge flow rate per cubic meter. The first of these shows that there is no such thing as a magnetic "charge". The second is called **Ampere's law**. The circulation in the magnetic field is proportional to the flow of charge.

Lorentz Force and Vector Potential



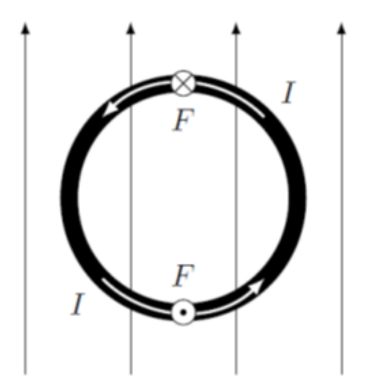
$$F = qvB\sin\theta$$

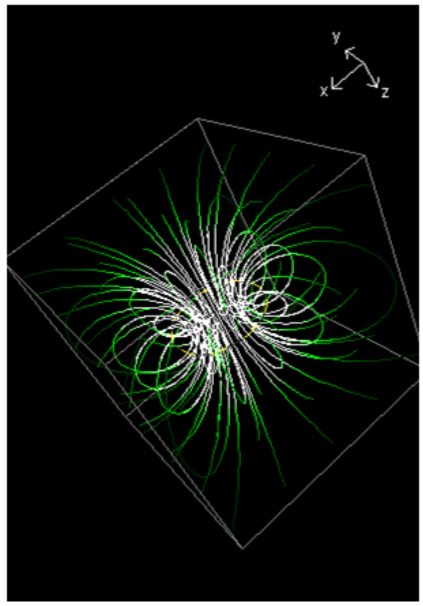
$$B = \nabla\times A$$

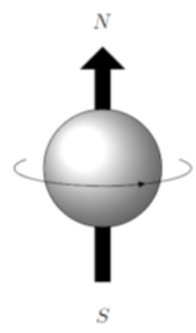
$$L = qvA\cos\theta$$

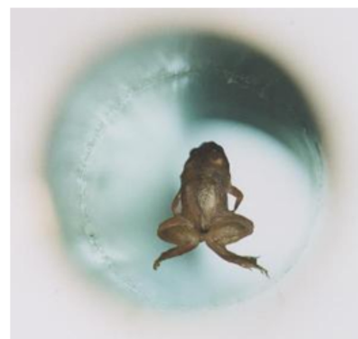
- This is where the book starts.
- When an electric charge moves in the presence of a magnetic field, it experiences a magnetic force. This is sometimes called the **Lorentz force**. The magnitude of this force is proportional to the charge and the magnetic field perpendicular to its velocity:
$$F = qvB\sin\theta$$
- The direction of the Lorentz force is given by another right-hand-rule:
  - Point your right forefinger in the direction of the motion of the charge. Orient your remaining fingers along the magnetic field. Your right thumb points in the direction of the force.
  - This second "right-hand-rule" effectively unwinds the previous one in the definition of the magnetic field. In this way we recover our initial fact that parallel currents attract and opposites repel.
- A circular form of current will experience a torque in a magnetic field. If one side is pushed left, the other side will be pushed right. This is the principle behind a basic DC motor works.
- Notice that the magnetic force is perpendicular to the charge velocity. This means the magnetic field does no work: it deflects the motion without changing its speed. The magnetic force will cause the charge to move in circles.
- Ampere's law guarentees that no magnetic "voltage" exists. But the magnetic force is conservative since it does no work. Therefore, we might expect some sort of potential to exist. And it does—but it is not a scalar. It is called the **vector potential**.
- The vector potential follows the current and the magnetic field "curls" around the vector potential:
$$B = \nabla\times A$$
- Using the principle of least action, the Lorentz force law follows from $L = qvA\cos\theta$.

Natural Forms of Magnetism









- What about our old childhood friend: the bar magnet? Actually, there are three forms of natural magnetism:
  - **Ferromagnetism**: This is one you are familiar with in bar magnets and such. Some materials have a tendency to line up their internal magnetic moments. This is the strongest form of magnetism.
  - **Paramagnetism**: A kind of magnetic induction similar to how an electric dipole induces a dipole moment in an insulator. This is why magnets stick to the refrigerator.
  - **Diamagnetism**: Related to Lenz Law (see next lecture). This is a repulsive form of magnetic induction and very weak. See here: http://www.ru.nl/hfml/research/levitation/diamagnetic/
- These types of magnetism are based on the fact that the electrons in the atoms are in motion. Each electon generates a natural magnetic field. If the motion has a reason to line up, these magnentic fields will accumulate.
- Each electron is like a ring of current spinning around the nucleus. The magnetic field curls around the current in a kind of donut-like shape. This is also the shape of the earth's magnetic field. In the exact center we have:

$$B = \frac{\mu_0 I}{2r}$$

- The magnetic field around a ring of current is similar to the electric field around an electric dipole. Actually, there are many parallels between these two. The magnetic dipole moment for the ring is $m = IA$, where A is the area surrounded by the current (not the vector potential).
- One nice trick is to coil our wire into a long cylinder. The resulting magnetic field still wraps itself around the current but doesn't break through the coil. This effectively traps the magnetic field in the center of the coil. This is how you build an electromagnet, also known as a **solenoid**.
- The field inside an ideal solenoid of length L with N coils is constant:

$$B = \mu_0 N I / L$$

Physics 203 Lecture 3

Electromagnetic Induction

- The previous lecture is sometimes called "magnetostatics" because the resulting magnetic fields are constant over time. This is a result of the current flow begin steady—so although the electrons are in motion, the pattern does not change.
- We know how charges respond to electric and magnetic fields:

$$F = q(E + vB\cos\theta)$$

- We also know how electric charge generates these electromagnetic fields through Coloumb's law

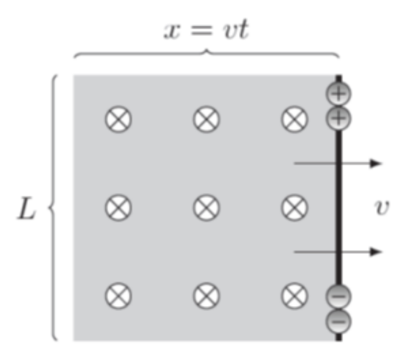
$$\nabla\cdot E = \rho/\epsilon_0 \qquad \text{and} \qquad \nabla\times E = 0$$

  and Ampere's law:

$$\nabla\cdot B = 0 \qquad \text{and} \qquad \nabla\times B = \mu_0 J$$

- What we are about to learn is how the electromagnetic field itself can act as its own source. This was Faraday's and Maxwell's great discovery and our modern way of life depends upon it.

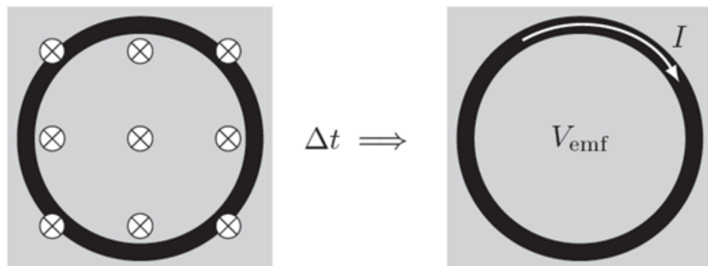
Faraday's Law



$$\text{EMF} = vBL$$

$$\text{EMF} = -N\Delta\Phi_{\text{mag}}/\Delta t$$

$$\Phi_{\text{mag}} = B_{\text{avg}} A \cos\theta$$



- If we drag a conductor through a magnetic field, current will flow. This is a direct consequence of the Lorentz force:
  - Though the conductor is neutral, its electrons are free to flow. When the wire is perpendicular to the field and the motion perpendicular to both, the Lorentz force on the electrons will align with the wire according to the right-hand-rule.
  - This is called **motional EMF**. If the current is blocked, a charge separation will build up and generate an electric field—similar to what happens in a capacitor. Along a wire of length L the EMF is vBL.
  - Notice that the orientation requires the wire to "cut though" the magnetic lines of flux. The magnitude of the EMF is proportional to both the number and rate at which these lines of flux are cut.
- Faraday realized that the same effect can occur if we move the magnetic field instead.
- Of course, moving a constant magnetic field is the same as not moving it—it is the same everywhere. What we need is the motion of a non-constant magnetic field. In fact, "moving" the field doesn't really make sense. What we mean is *changing* the magnetic field.
- The electromotive force from a changing magnetic flux through N coils of wire is:

$$\text{EMF} = -N\Delta\Phi_{\text{mag}}/\Delta t$$

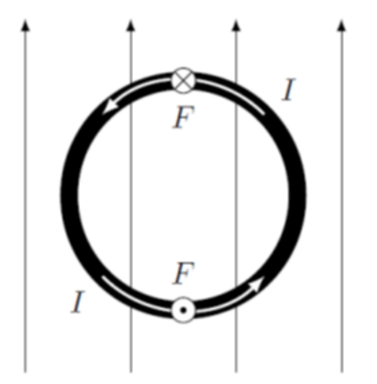
- This is Faraday's law of induction. We'll talk about why there is a negative sign in a bit...
- Remember that the magnetic flux Φ_{mag} through a given surface area A is defined as:

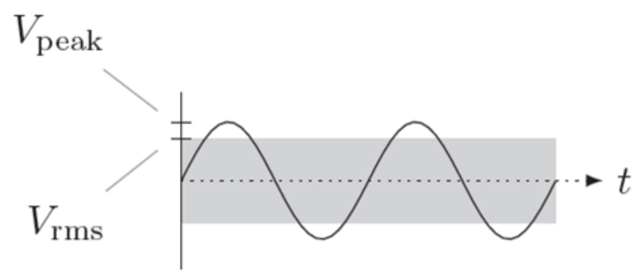
$$\Phi_{\text{mag}} = B_{\text{avg}} A \cos\theta$$

where θ is defined relative to the normal line perpendicular to the surface area.

Electric Generators

$$\text{EMF} = NBA\omega \sin \omega t$$





$$\frac{V_{\text{rms}}}{V_{\text{peak}}} = \sqrt{\frac{1}{2}}$$

- Faraday and others turned his insight regarding EMF and magnetic flux into a practical machine: the electric "dyanamo", which was the first example of an electric generator.

- Every electric generator converts rotational energy into electrical energy. In fact, they are literally electric motors in reverse.

- When the magnetic field rotates through the coil (or *vice versa*), a pulse of current occurs. When the opposite end of the magnetic field rotates through the coil, a pulse of current in the opposite direction occurs. For a rotor in uniform circular motion, the EMF pattern follows from Faraday's law:

  $$\text{EMF} = NBA\omega \sin \omega t$$

  where ω is the angular speed of the rotor.

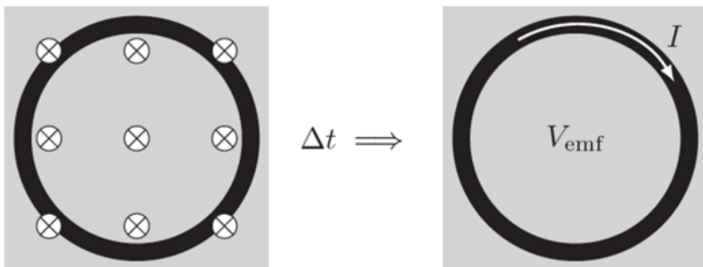
- Originally, this **alternating current** was a difficulty to be overcome. We can now harness this "AC" voltage directly and to our advantage.

- Technically, any pattern of current that oscillates is a kind of alternating current. The sinusoidal pattern typical of electric generators is the most common.

- For any AC pattern, its characteristic voltage is its **RMS** voltage—a kind of average value for the pattern. It's proportional to the "peak" or maximum voltage. For a sine wave:

  $$\frac{V_{\text{rms}}}{V_{\text{peak}}} = \sqrt{\frac{1}{2}}$$

Revised Electric Potential



$$\nabla\times E = -\frac{\partial B}{\partial t}$$

$$\cancel{\Delta V = -E_{\text{avg}}\Delta x}$$

$$\Delta V = -\left(E + \frac{\partial A}{\partial t}\right)_{\text{avg}}(\Delta x)$$

- You may not have noticed, but there is a big theoretical problem with Faraday's law: circular EMF.
- In general, the electric field points from high to low voltage and pushes the current through the circuit.
- But Faraday's law drives the current in a circle, so the electric field must also. In other words, a changing magnetic field twists and curls the neighboring electric field. In symbols:

$$\nabla\times E = -\frac{\partial B}{\partial t}$$

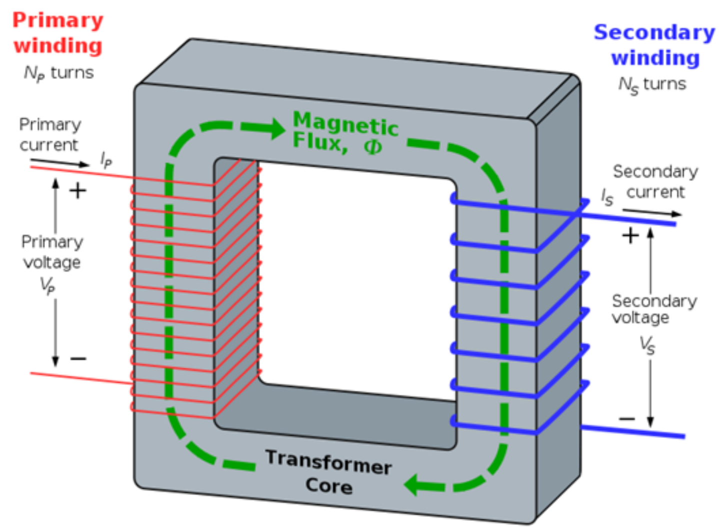
- This is Faraday's law rewritten for the fields directly. Sorry for the calculus notation—the right-hand side represents the rate at which the B field changes. (Recall that the field can be thought of as the flux density.)
- Earlier we used that fact that $\nabla\times E = 0$ to justify our defintion of voltage as $\Delta V = -E_{\text{avg}}\Delta x$. Faraday's law invalidates this argument.
- The situation can be salvaged by using the magnetic vector potential. Remember that $B = \nabla\times A$. So Faraday's law can be written as:

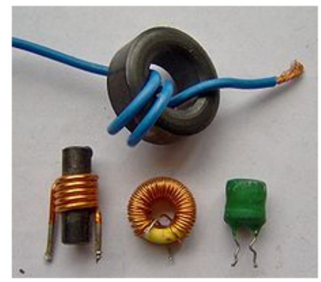
$$\nabla\times E = -\frac{\partial}{\partial t}(\nabla\times A) \implies \nabla\times\left(E + \frac{\partial A}{\partial t}\right) = 0$$

- This combination of the electric field and the magnetic vector potential has zero curl. This motivates a generalization of the electric potential when the magnetic field changes:

$$\Delta V = -\left(E + \frac{\partial A}{\partial t}\right)_{\text{avg}}(\Delta x)$$

Transformers and Inductors





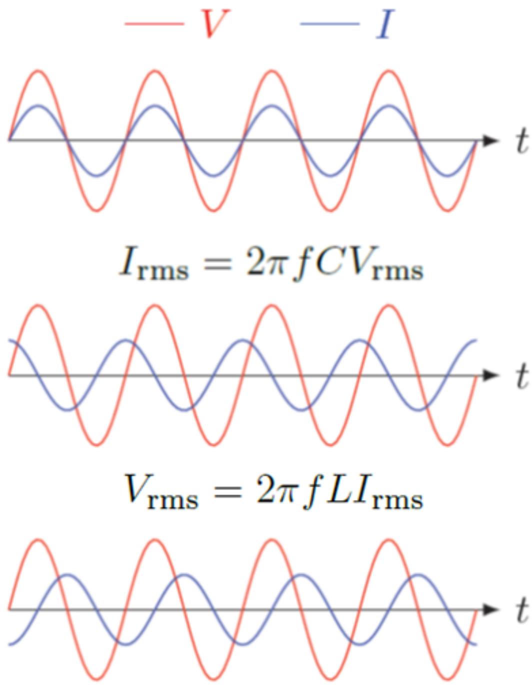
$$V_{\mathrm{emf}} = L\frac{\Delta I}{\Delta t}$$

$$PE = \tfrac{1}{2}LI^2$$

$$\frac{V_1}{V_2} = \frac{N_1}{N_2}$$

- Faraday's law gives us a way to extract energy from the magnetic field with a coil of wire. We also know that we can *generate* a magnetic field with a coil of wire. Perhaps we can put them together?
- A **transformer** just such a thing. When we drive AC current through the primary coil, AC current is induced in the secondary coil even though there is no electrical connection between them. The energy moves through the magnetic field instead.
- The amount of induction between the two coils depends on the geometry: position, orientation, size, etc. Also, there is a lot of magnetic field energy wasted since only the field inside the secondary contributes.
- One common trick is to wrap the two coils around an iron core. This iron core has the effect of channeling the magnetic field thus increasing the efficiency of the transformer.
- One common use of the transformer is to "step-up" or "step-down" voltage. The inductance is proportional to the number of coils on either side. In general,
$$\frac{V_1}{V_2} = \frac{N_1}{N_2}$$
- A transformer that steps-up voltage will step-down current. Energy is still conserved and so is power: $P = VI$.
- One common configuration for a transform is to wind wires around the same core on top of one another. This maximizes their mutual inductance. This begs the question: will a single coil induce EMF within itself?
- The answer is yes and it is called **self-inductance**, L. This is where the negative sign in Faraday's law makes a difference: the self-induced EMF *opposes* the current flow. This **back EMF** reflects the fact that some of the energy is getting pushed into the magnetic field.
- The magnetic energy from current flowing in an inductor is
$$PE = \tfrac{1}{2}LI^2$$

Reactance: ELI the ICE Man



$$V = IX$$

$$X_L = 2\pi f L$$

$$X_C = \frac{1}{2\pi f C}$$

- Ampere's law and Faraday's law gives us a way to use a coil of wire to interact with the magnetic field in our electric circuits. When used as an electrical component we call this an **inductor**. It is quantified by its self-inductance, L. The SI unit is the "henry".
- In many ways the inductor is to the magnetic field what the capacitor is to the electric field. Along with the resistor, these three components are the basic tools used in an AC electronic circuit.
- When we put an AC voltage across an inductor, its self-inductance will oppose the resulting AC current. This opposition is different than the resistance in a resistor, because the energy is not lost—it is being stored in the oscillating magnetic field. When the field collapses, the energy is regained.
- Nonetheless, this opposition causes the current to be a quarter-cycle behind the voltage. The value of the final current depends upon the driving AC frequency:

$$V_{\text{rms}} = 2\pi f L I_{\text{rms}}$$

- For a capacitor, the charge on the plates is proportional to the AC voltage. But current is what charges the capacitor, so the relationship is the opposite of the inductor: the current is a quarter-cycle ahead of the voltage. And the RMS values are related by:

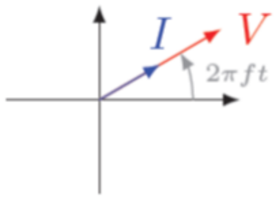
$$I_{\text{rms}} = 2\pi f C V_{\text{rms}}$$

- These equations look similar to Ohm's law: they both have the form $V = IX$. This X is called the **reactance** of the component. This is not a resistance: the energy is not lost but stored in the electromagnetic field. Thus,

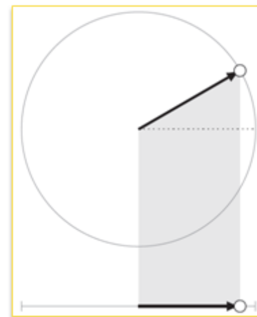
$$X_L = 2\pi f L \qquad \text{and} \qquad X_C = \frac{1}{2\pi f C}$$

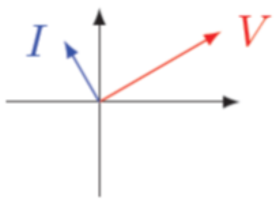
- In order to remember whether the voltage or current leads, use the acronym "ELI the ICE man".

Phasors Again

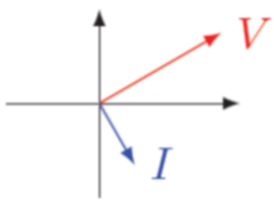


$$\omega = 2\pi f$$





$$V = IZ$$

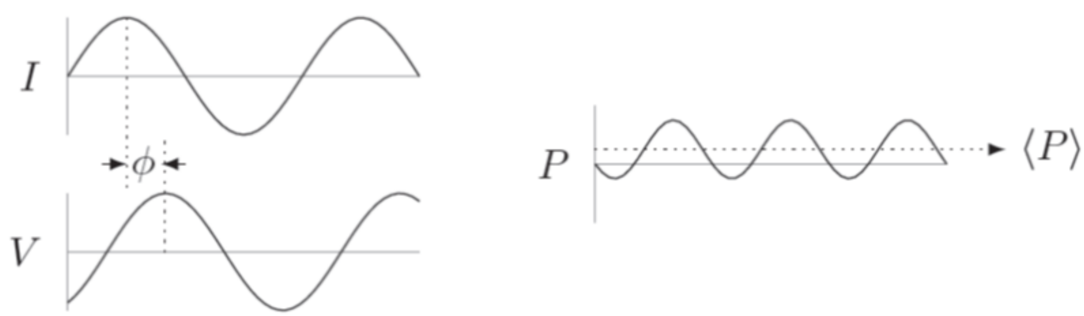


$$Z = \sqrt{R^2 + \left(2\pi f L - \frac{1}{2\pi f C}\right)^2}$$

- Dealing with the phase lag (or lead) in these AC circuits is a mathematical nightmare. But there is a trick. And it's a trick you've already seen: **phasors**.
- The reason this works is that we map the sinusoidal pattern of the voltage (or current) to the uniform circular motion of some imaginary vector. The magnitude of the vector is the peak voltage (or current). The phase differences correspond to different angles for these vectors.
- You can also use complex numbers to work with these phasors.
- For example, consider a circuit with a resistor, a capacitor, and an inductor in series.
  - Since they are in series they all share the same current. The voltage drops will depend on the respective resistance and reactances.
  - The voltage drop across the resistor is in phase with the current and the voltage is given by Ohm's law. We represent this as a vector with a length that corresponds to the peak voltage given by IR and pointing along the x-axis.
  - The voltage drop across the capacitor has a length that corresponds to $I/2\pi fC$. It points in the negative y-direction because voltage lags the current in the capacitor ("ICE").
  - The voltage drop across the inductor has a length that corresponds to $2\pi fLI$. It points in the positive y-direction because voltage leads the current in the inductor ("ELI").
- The *total* voltage drop is the vector sum of these three arrows. Its horizontal component comes from the resistance in the circuit, and its vertical component comes from the reactance.

Impedance and Power

$$P = IV\cos\phi$$



- We can redo the previous logic with the resistance and reactances directly. Because they are in series, the resistance and reactances add—but they add as vectors. The end result is a vector that has both resistive and reactive components.
- The magnitude of this final vector is called the **impedance** of the circuit. The RMS voltage and current obey Ohm's law with the impedance: $V = IZ$.
- For our series circuit, the impedance is given by

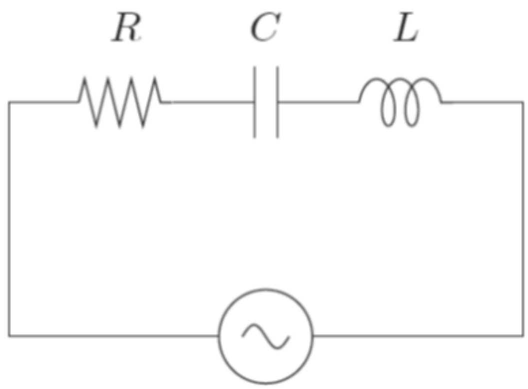
$$Z = \sqrt{R^2 + \left(2\pi fL - \frac{1}{2\pi fC}\right)^2}$$

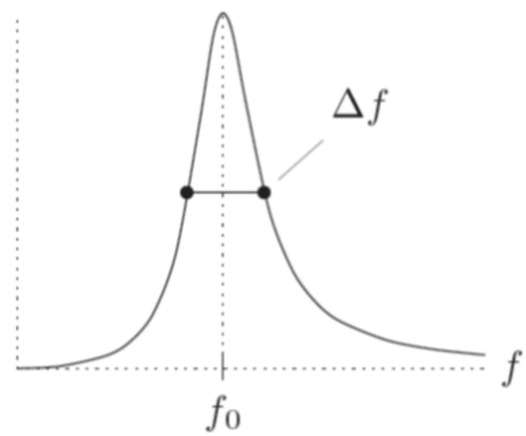
- Remember that the impedance (which is what we measure) has both resistive and reactive components. The power lost in an AC circuit is lost only through the resistor. The formula is

$$P = VI\cos\theta$$

- Although this looks a lot like the fundamental equation $P = VI$ from which it is derived, the symbols have a slightly different meaning in this AC context. The current and voltage are RMS values and the power is the average power consumed over time.
- At any one moment, the power used by the circuit is the product of its instantaneous voltage and current. But some of that power is going to (or coming from) the electromagnetic field. On average, only the resistance consumes power in the circuit.

RCL Resonance





$$X_L = X_C$$

$$f_0 = \frac{1}{2\pi}\sqrt{\frac{1}{LC}}$$

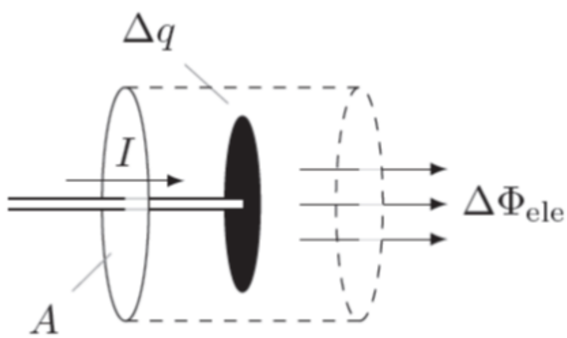
- A plot of the impedance from the series RCL circuit we have been discussing reveals that there is a certain frequency in which the impedance is at a minimum. This is an example of resonance in the AC circuit.
- What is happening is that at low frequencies the components act like they would for direct current: the induction is a simple wire with low impedance, but the capacitor is an opening in the circuit with very high impedance.
- On the other extreme, at high frequencies the roles reverse. The impedance of the inductor is very high due to Faraday's law. The capacitor simply charges and discharges with ease.
- Sometimes we can qualitatively understand where current will flow through a complicated circuit by mentally replacing the components with their respective open or short circuit counterparts.
- But the really interesting thing is in the middle. The width of the dip in impedance is proportional to the resistor and the resonant frequency is given by
$$f_0 = \frac{1}{2\pi}\sqrt{\frac{1}{LC}}$$
- This is the frequency at which the impedance is lowest and the resulting current is highest. Looking back at the impedance formula for an RCL circuit, this happens with the reactance component is zero. In other words, it is the frequency at which $X_L = X_C$.
- Physically this means that the energy from the inductor feeds the capacitor and *vice versa*. Once the circuit is up to speed, energy oscillates between the magnetic field and the electric field, exerting no additional load on the power supply.

Physics 203 Lecture 4

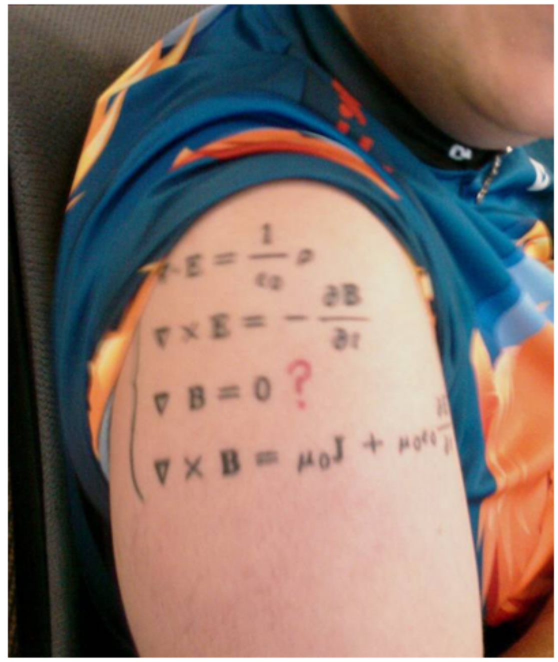
EM Radiation and Relativity

- Einstein's theory of relativity was born from the study of electromagnetism. As such, the theory of electromagnetism is already relativistic.
- What Einstein realized is that the final theory was in conflict with Newton's laws of motion. The special theory of relativity is the correction.
- One consequence is that Newton's law of gravity is unworkable in special relativity. Einstein had to fight hard to find the proper modification for the gravitational field. The final answer is general relativity and in several ways parallels the laws of electromagnetism.

Maxwell's Correction and EM Waves



$$\nabla\times B = \epsilon_0\mu_0\frac{\partial E}{\partial t}$$



- Working from theory, Maxwell found a flaw in the laws of electromagnetism. There are a couple of ways to see this, but one simple way is within a charging capacitor. As current flows into and out of the capacitor, a magnetic field is generated—just like around any current. This is Ampere's law.
- But what about *inside* the capacitor? It seems natural to assume the magnetic field is continuous and exists there too. Insipred by Faraday's law, Maxwell realized that the only way to justify this is to assert that a changing electric field will also curl the magnetic field:

$$\nabla\times B = \epsilon_0\mu_0\frac{\partial E}{\partial t}$$

- This is the final piece of the puzzle. With this, we have finally fully incorporated our original "surprize" that parallel currents attract.
- For the electric field, we have Gauss's law and Faraday's law:

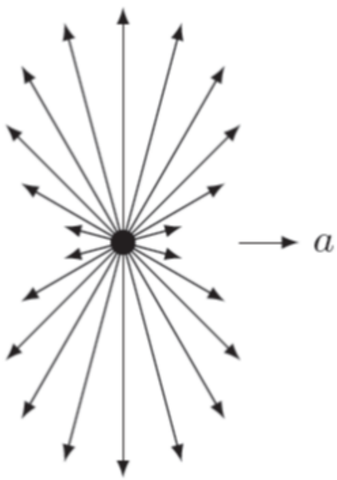
$$\nabla\cdot E = \rho/\epsilon_0 \qquad \text{and} \qquad \nabla\times E = -\frac{\partial B}{\partial t}$$

- And for the magnetic field we have Ampere's law with Maxwell's extra term:

$$\nabla\cdot B = 0 \qquad \text{and} \qquad \nabla\times B = \mu_0 J + \epsilon_0\mu_0\frac{\partial E}{\partial t}$$

- These are **Maxwell's laws** of electromagnetism and represents one of the high-water marks in theoretical physics.

Radiation: Power, Intensity, Momentum



$$c = \sqrt{\frac{1}{\mu_0 \epsilon_0}} = 2.998 \times 10^8$$

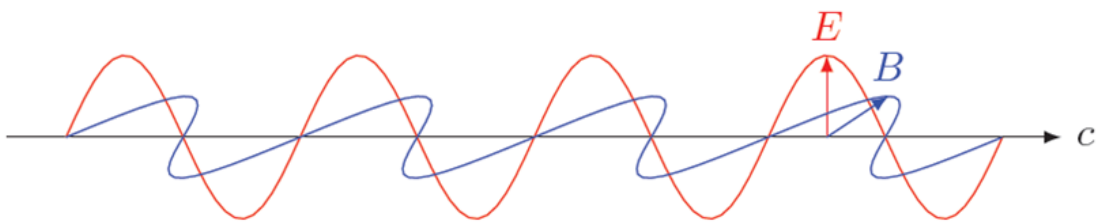
$$u = \tfrac{1}{2}\epsilon_0 E^2 + \tfrac{1}{2} B^2/\mu_0$$

$$P = \frac{a^2 q^2}{6\pi\epsilon_0 c^3}$$

$$I = c u_{\mathbf{avg}} = \tfrac{1}{2} c \epsilon_0 E^2$$

- After Maxwell's discovery, he realized that he had just created an physical model for light. A changing magnetic field can create an electric field and a changing electric field can create a magnetic field. With the right initial conditions, a self-sustaining electromagnetic field is possible. We call this **electromagnetic radiation**.
- Light must move in order to live. The fields must be in continual change, or the self-sustaining conditions won't be right. The simplest form is in an electromagnetic wave. Maxwell verified that the speed of this wave is equal to the speed of light:
$$c = \sqrt{\frac{1}{\epsilon_0 \mu_0}}$$
- On consequence of this model is that visible light is a mere sliver of the wavelengths possible for EM radiation. Radio waves, gamma rays, X-rays, microwaves are all possible because of the elecromagnetic nature of light.
- In general, the total energy density of the electromagnetic field is
$$u = \tfrac{1}{2}\epsilon_0 E^2 + \tfrac{1}{2} B^2/\mu_0$$
- But for a electromagnetic wave, the electric and magnetic are perpendicular to one another and both are perpendicular to the direction of the wave propogation. The fields related by $E = cB$ so the energy density is evenly split between them. The intensity of the radiation flow is given by
$$I = c u_{\text{avg}} = \tfrac{1}{2} c \epsilon_0 E^2$$
- This electromagnetic radiation occurs whenever a charged particle accelerates. Most of the energy is distributed perpendicular to the direction of the acceleration. The total power radiated is given by **Larmor's formula**:
$$P = \frac{a^2 q^2}{6\pi\epsilon_0 c^3}$$
where q is the magnitude of the charge and a is its acceleration.

Relativity and Einstein's Second Postulate



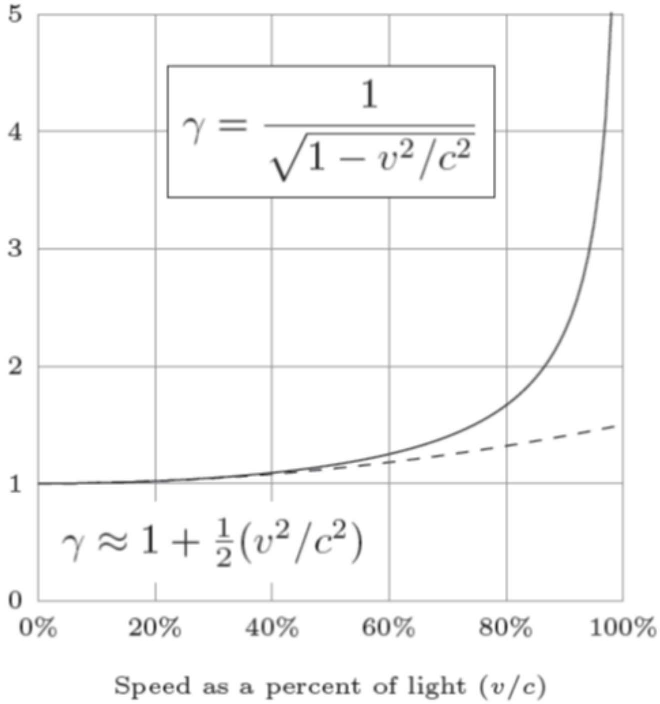
$$v_{ac} = \frac{v_{ab} + v_{bc}}{1 + v_{ab}v_{bc}/c^2}$$

- The **principle of relativity** is the statement that the laws of physics only depend on the relative speed between objects. There is no such thing as "absolute speed". We are always free to choose a reference frame—this choice is not dictated in any fundamental way.
- The laws of electromagnetism demand that light move in order to live. But movement relative to what? The original answer was the electromagnetic "ether",
- Einstein recognized that this cannot be. If one could travel in a rocket along with a light beam, its apparent motion goes away. What then? The light cannot exist. But the existence of the light should not depend on *my* motion. Something is wrong.
- The root of this conflict is between Newton's law of motion and Maxwell's laws of electromagnetism. The special theory of relativity is the result of resolving this conflict. Einstein showed how every element of mechanics must be "tweaked" when speeds are near light.
- Several electromagnetic rules can be used to discover the "tweaks". Einstein found the simplest one: regardless of the speed of the source or observer, the speed of light is always 3×10^8 m/s. The only way this will work is if speed doesn't work the way we expect.
- We have Einstein's velocity addition formula:

$$v_{ac} = \frac{v_{ab} + v_{bc}}{1 + v_{ab}v_{bc}/c^2}$$

- This differs from the simple additon we expect only when the speeds are near light speed. But when they are weird things happen. Consider the extreme: suppose one of them is light speed itself. Regardless what the other one is, the result is still light speed. You cannot catch up to your headlights no matter how fast your car.

Space, Time, Mass Affected By Motion



$$t = \gamma t_0$$

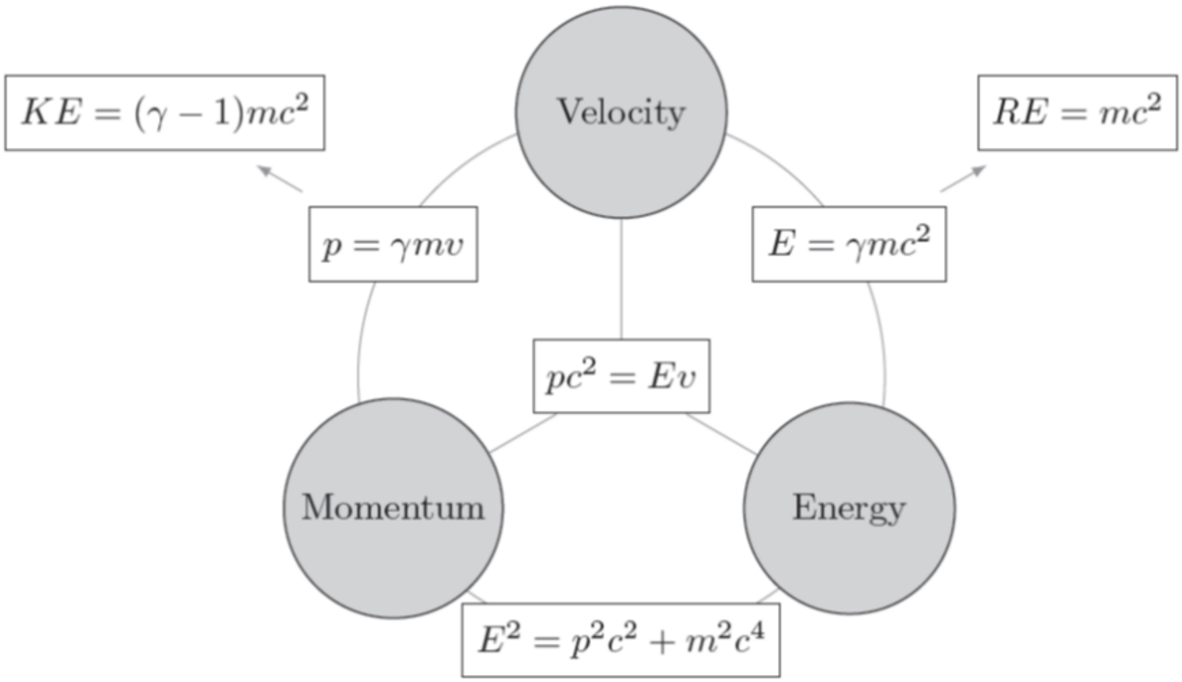
$$\Delta t = vL_0/c^2$$

$$L = L_0/\gamma$$

$$m = \gamma m_0$$

- Einstein's velocity addition formula is a consequence of three more fundamental relativistic effects. These are collectively called the **Lorentz transformation**. They are
  - **Time dilation**. Moving clocks tick slower than normal. The formula is
  $$t = \gamma t_0$$
  - Motion desychronizes separated clocks. The time gap created is proportional to the sepration:
  $$\Delta t = vL_0/c^2$$
  - **Length contraction**. This can actually be shown to follow from the first two. Moving rulers shrink in the direction of their motion:
  $$L = L_0/\gamma$$
- In all of these formula, the subscript "0" represents the quantity when measured at rest.
- The value of γ is $1/\sqrt{1 - v^2/c^2}$ and is sometimes called the **Lorentz factor** for the motion. Notice that it is always a number greater than one.
- Finally, a fourth relativistic effect completes our devastation of Newtonian mechanics. The mass of fast objects increase with speed:
  $$m = \gamma m_0$$
- Frequently this last effect is cited as the "physical" reason for the light-speed limit. The closer one gets to the speed of light, the harder it becomes to accelerate and increase this speed. This is kind of like the situation approaching absolute zero: it gets harder the closer you get.

Energy and Momentum Also



- Actually, the mass adjustment is no longer the "preferred" way to explain the effect of relativity on Newton's laws. It's a bit clearer to say that the formula for momentum is different:

$$p = \gamma mv$$

- You can see that it is basicly a "six of one, half a dozen the other" kind of thing. Whether we associate the Lorentz factor with the mass or the velocity, we must adjust Newton's second law with this version of momentum.
- However, we continue to maintain our defintions of work and energy. But with this redefinition of momentum, our formula for kinetic energy changes:

$$KE = (\gamma - 1)mc^2$$

- This does reduce to the Newtonian formula when the speeds are small. In order to show this, we can use the following approximations:

$$\gamma = 1 + \tfrac{1}{2}v^2/c^2 \qquad \text{and} \qquad 1/\gamma = 1 - \tfrac{1}{2}v^2/c^2$$

when v is small relative to c.

- In a relativistic collision, momentum is still conserved but mass is not. When two particles of equal mass collide, each momentum above mv goes into the mass of the final particle. Considerations like these lead Einstein to the most famous equation in physics:

$$E_0 = mc^2$$

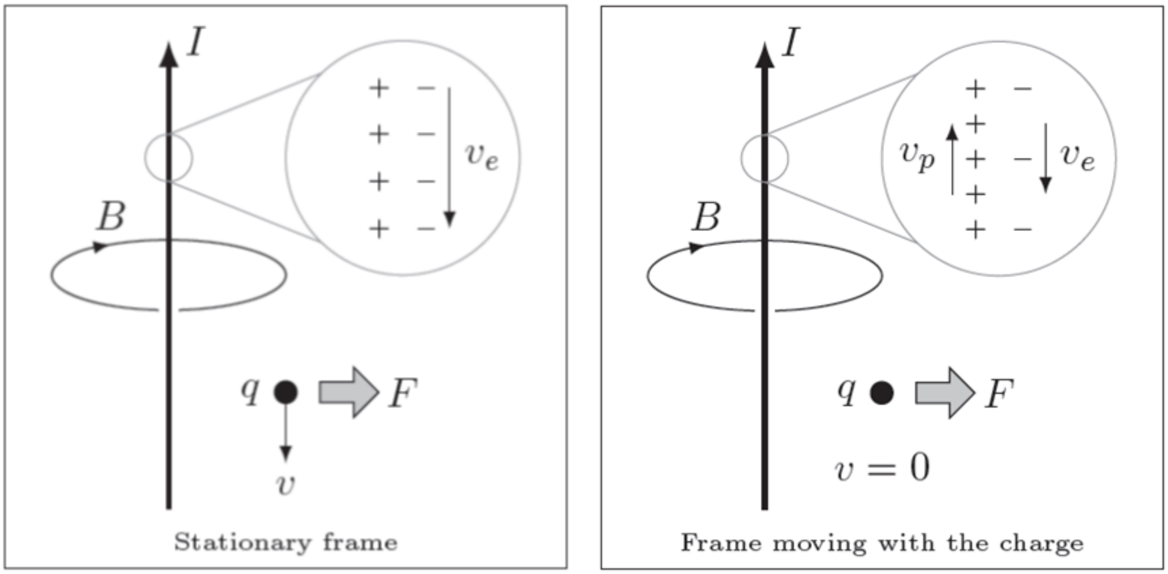
- This formula represents the fact that an object has a kind of energy even when it is not moving. This is called its **rest energy**. The combination of the rest energy and kinetic energy is conserved in any collision (or explosion):

$$E = \gamma mc^2$$

- The relativistic momentum and energy are also connected by two very useful formula:

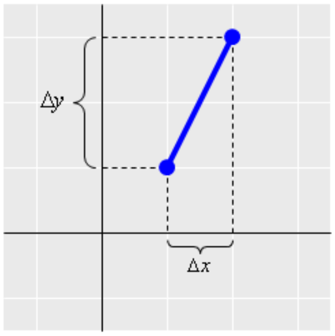
$$E^2 = p^2c^2 + m^2c^2 \qquad \text{and} \qquad p = Ev/c^2$$

Magnetism Comes From Relativity



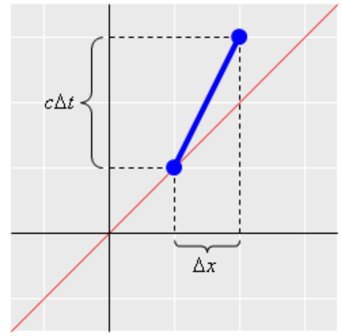
- With this slide we come full circle to view electromagnetism through the lens of relativity. The point here is to show that we cannot have a relativistic theory of the electric force without magnetism.
- The fact that parallel lines of current attract was the starting point for our excursion into magnetism. Previously, we pulled the result from "thin air" (experiementaly justified, or course). Now we can show that this must occur from relativity and Coulomb's law.
- Consider a line of current the flows in the x-direction. We know that physically this means electrons are flowing in the opposite direction and the positive atomic charges stand still. Furthermore suppose an electron is flying by (in a parallel line of current) in the negative x-direction.
- Consider the moving electron's frame of reference. The electrons in the other wire are at rest, but the positive charges in the wire are moving backward (in the positive x-direction). Since they are moving, they suffer length contraction: they are closer together than the electrons in the wire.
- Because the positive charges are more dense than the negative charges, there is a net positive charge on the wire. The moving electron is attracted to this net positive charge according to Coulomb's law. So this is why parallel currents attract: length contraction.
- In one frame we interpret this force as electric (moving with the current) and in another we interpret this force as magnetic (at rest in the lab). This duality justifies the term "electromagnetic" field.
- It may occur to you that Coulomb's law parallels Newton's law of gravity. Is there some sort of "gravito-magnetic" force associated with gravity? Yes, there is: http://en.wikipedia.org/wiki/Gravitomagnetism
- One effect that has been tested is called **frame dragging**. In the same way that an spinning charge will generate a magnetic field, so does the Earth generate this gravito-magnetic field due to its rotation. The effect is extremely small, but has been measured in the Gravity Probe B experiment.

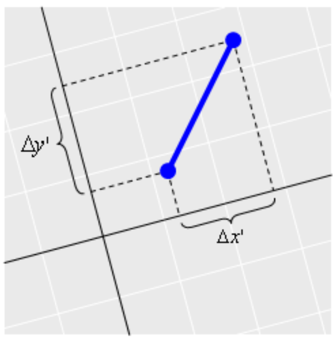
The Rosetta Stone: Space-Time Interval



$$t = \gamma t_0$$

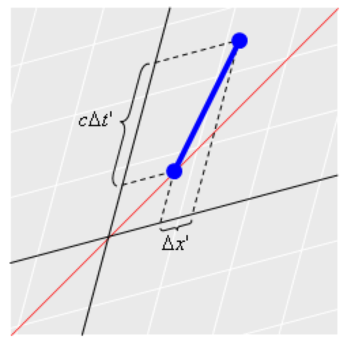
$$(ct)^2 - x^2 = (ct_0)^2$$





$$\boxed{I = (c\Delta t)^2 - \Delta x^2}$$

$$E^2 - p^2c^2 = m^2c^4$$



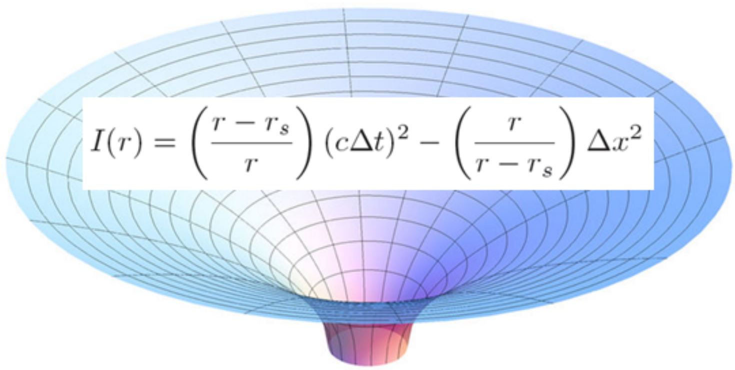
- So far, we have presented the theory of relativity in a way that is self-consistent, but it all has the feel of some cosmic conspiracy designed to thwart light-speed. There is another way to look at this which rearranges them in a more satisfying way.
- We can start by rewriting the time dilation formula in the following way:

  $$(ct)^2 - x^2 = (ct_0)^2$$

  For this I've used the fact that the distance the clock moves is $x = vt$.
- The advantage of writing it this way is that all the terms on the left are related to the motion of the clock and the right hand side is a constant, the proper time. No matter what frame of reference I am in the combination on the left is the same.
- This is not unlike the way the Pythagorean theorem works. No matter what the orientation of my x and y axes, the sum of the squares is the same: the square of the total distance. In relativity we call this difference the **space-time interval**.
- This interval shows how space and time depend on one another: if x goes up, then the measured time must go down. Time slows down. From this interval formula we can also get

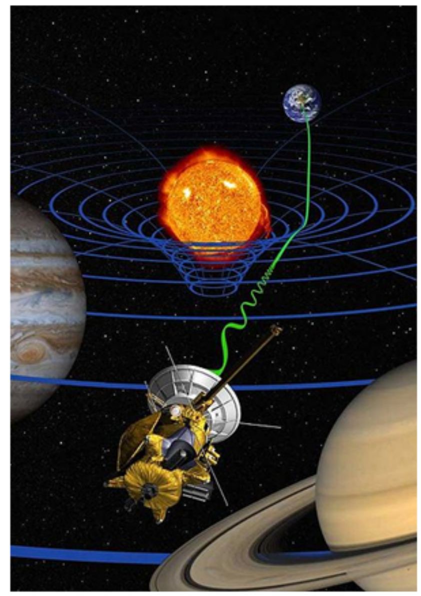
  $$E^2 - p^2c^2 = m^2c^4$$
- This shows a characteristic feature of relativity: it pairs mechanical quantities. One can be thought of as the "time component" of the other. In this case, energy is the time component of momentum. Power is the time component of force.
- The same thing happens in electromagnetism. The electric field and magnetic fields combine. Charge and current pair together. The electric potential and the magnetic vector potential also.

General Relativity



$$r_s = 2GM/c^2$$

$$t = t_0\sqrt{1 - \frac{2GM}{rc^2}} \approx \left(1 - \frac{GM}{rc^2}\right)(t_0)$$



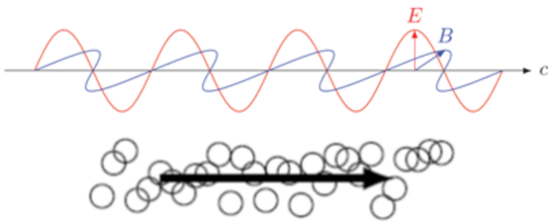
- We've already touched on the fact that we need a relativistic field theory for gravity. The most obvious approach is to mimic Maxwell's equations in the context of gravity. Flip some signs so that the force is attractive, replace $1/4\pi\epsilon_0$ with G, and you are done.
- This actually works pretty well and you can get pretty far with these equations. But it cannot be the whole story. It took Einstein 10 years to crack this nut.
- The issue is that there is energy in the field—just like in electromagnetism. But by $E = mc^2$ the field has mass. Therefore the field itself is a source of gravity. This doesn't happen in electromagnetism because the field is not charged.
- Mathematically this introduces a non-linearity into the equations of general relativity. In the end, the solution involves modifying the space-time interval formula. These modifications are called the **space-time metric** for the gravitational field.
- Schwarzschild was the one who found the solution for the simplest case. The metric around a particle with mass M is:
$$\left(\frac{r - r_s}{r}\right)(ct)^2 - \left(\frac{r}{r - r_s}\right)(x)^2 = (ct_0)^2$$
where $r_s = 2GM/c^2$, also known as the **Schwarzschild radius**.
- From this formula we can see that even if a clock is stationary (so $x = 0$), the clock will tick slower in a gravitational field:
$$t = t_0\sqrt{1 - \frac{2GM}{rc^2}} \approx \left(1 - \frac{GM}{rc^2}\right)(t_0)$$
- The approximation holds when GM/c^2 is much less than r. This is the red-shift formula we discussed back when we first talked about celestial mechanics and Newton's law of gravity.

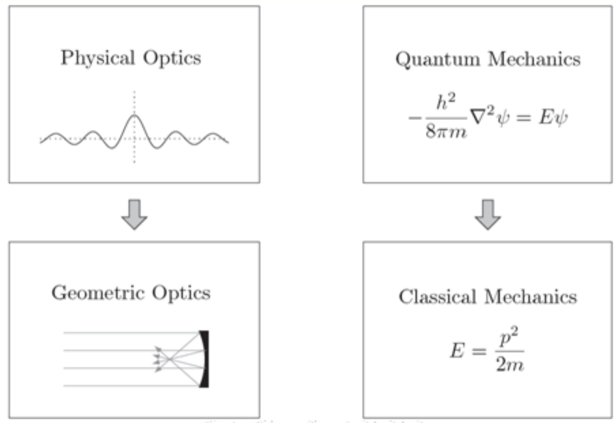
Physics 203 Lecture 5

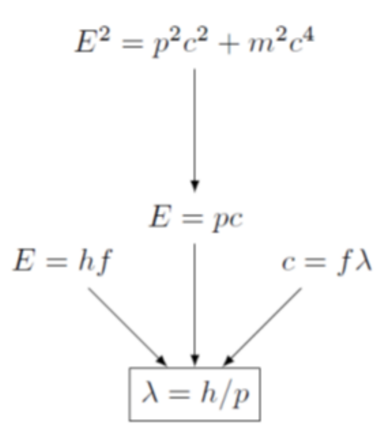
Quantum Mechanics

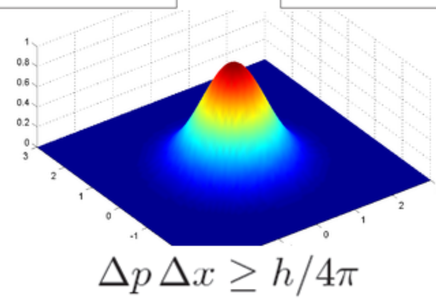
- Here is where we shift gears and move directly into the microscopic realm. At the turn of the 20th century technology in the physical science were allowing people to probe into and underneath the microscopic world around us. From these investigations quantum mechanics was born.
- The main way in which quantum mechanics touches our everyday lives is through microelectronics (setting aside of course the very structure of matter and the world we take for granted). Every DVD player, every iPhone, your car, even some shoes depend on it.
- In this way quantum mechanics is more relevant to us than relativity or the more esoteric topics later in this term.
- Yet quantum mechanics is a bizarre world. It's a weird combination of waves, probability, and complex mathematics that make philosophy professors drool with excitement. How our macroscopic reality springs from this foundation I will attempt to describe, but honestly it is very difficult to grasp.

De Broglie and Heisenberg









- Even today there are a number of open questions regarding the foundations of quantum mechanics. Perhaps the most straight-forward approach is to acknowledge that the fundamental elements of matter have both particle and wave-like properties. This is called **wave-particle duality**.

- We will mostly side-step the question of how this could be—I am just asking you to accept it. If you take this red pill, this new quantum world will open up.

- It is a bit more accurate to say that every elementary particle has a quantum *field* associated with it, but the terminology goes back to 1923. De Broglie was impressed with Einstein's "photon" hypothesis to explain the photoelectric effect. If a wave can act like a particle, why can't a particle act like a wave?

- Based on the photon energy formula, $E = hf$, the relationship between frequency and wavelength for any wave, $v = f\lambda$, and the connection between momentum and energy for electromagnetic radiation, $p = E/c$, one gets

$$p = h/\lambda$$

- De Broglie postulated that every material particle also has a characteristic wavelength based on its momentum:

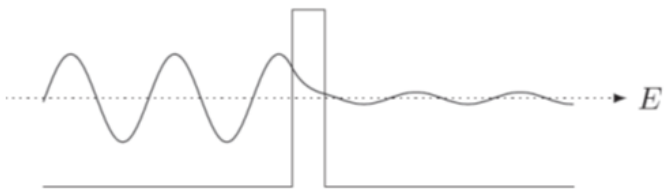
$$\lambda = h/p$$

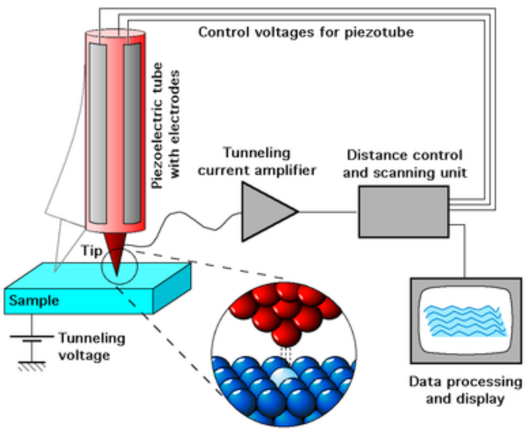
- A few years later, Heisenberg showed how the wave-nature of matter places a restriction on how precisely we can track the trajectory of microscopic particles:

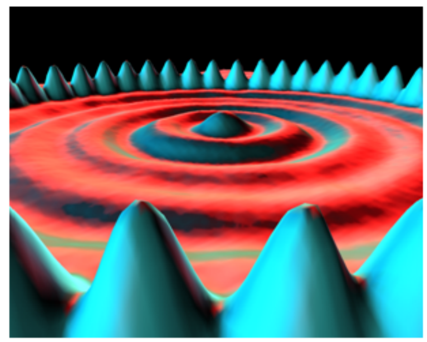
$$(\Delta x)(\Delta p) \geq \tfrac{1}{2}\hbar = h/4\pi$$

- This is called the **uncertainty principle**. The more precisely we know the location of a particle (that is, Δx), the less we know about where it is going (that is, Δp).

Trapped Waves and Tunneling

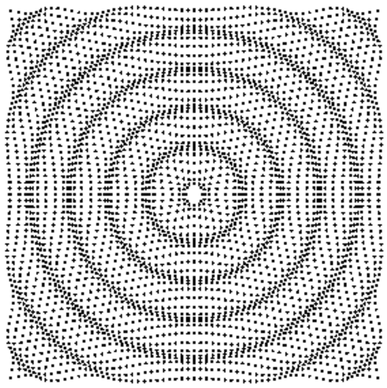






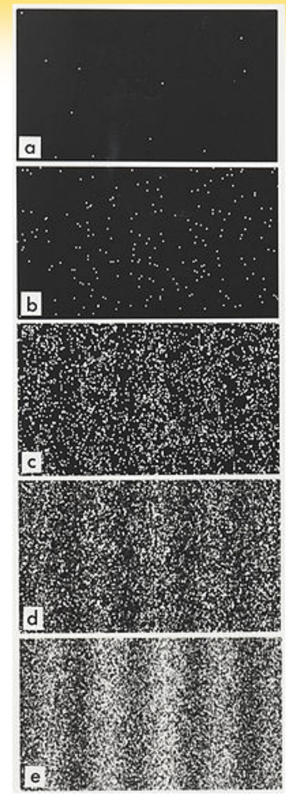
- Let's examine a couple of consequences of the wave nature of quantum particles.
- Consider an electron trapped in a box surrounded by an infinite force field. Within the box the particle is force-free. What happens?
- This is a situation analogous to a string fixed at both ends. The string can only vibrate at certain frequencies because its wavelength is restricted. The same thing happens to our electron: its momentum is restricted to certain values. Therefore so does its energy according to $E = p^2/2m$.
- These **energy levels** are characteristic for any constrained quantum particle. We will see this again when we consider the structure of the atom.
- Another consquence is known as **quantum tunnelling**. This occurs when an electron encounters a thin potential barrier. If the energy of the electron is under this barrier, as a classical particle, the electron will bounce and get pushed back.
- But a wave will bleed into the barrier. This is known as an **evanescent wave**. This can be seen with light and even water waves. The wave extends into the barrier with an exponentially decaying amplitude.
- If the barrier is thin enough, some of the wave energy will propogate through (the rest of the energy reflects back). This happens with electrons too. Due to the exponential decay, the amplitude of the transmitted wave is highly sensitive to the barrier thickness.
- This is the operating principle behind the **scanning tunneling microscope**. As the conductive tip is traced over the sample, any height increase produces a current spike as the electrons tunnel through the gap. The precision of this technique allows one to literally see the atoms in the given sample.

But What Is Waving?



$$nhf/t = \frac{1}{2}c\epsilon_0 E^2$$

$$\boxed{\rho \propto \psi^2}$$



- Although not actually performed directly until 1961, the classic test for the wave-nature of electrons is the double-slit experiement. Feynman has said that it "has in it the heart of quantum mechanics. In reality, it contains the *only* mystery."

- Initially, the results should not surprize you: send an electron beam through the double-slit and we see the same interference pattern as we do with light. Cover up one of the slits and you get a single-slit diffraction pattern.

- The trouble begins when you replace the beam with a single electron. If there is one slit, the electron flies through, but not always in a straight line. Sometimes it goes straight through, sometimes it is deflected. Perhaps you are willing to buy that based on the uncertainty principle.

- Now open both slits and let the electron through. Sometimes the particle strikes here, sometimes there. It is unpredictable. But if we send a bunch through one after the other, a pattern develops. It's the regular interference pattern. (By the way, the same thing happens with low intensity light).

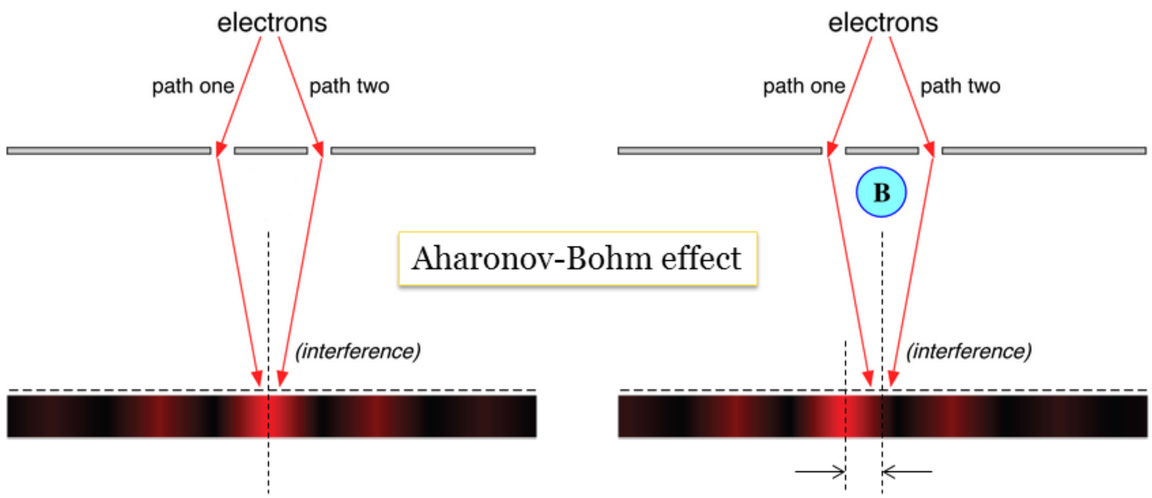
- Okay—maybe you are still willing to accept this because you are getting used to thinking of the electron as a field. The field interacts with the double slit which determines the interference.

- But the question is for *this* electron, what principle determines why it strikes *this* spot on the screen at *this* time? Using Newton's laws we could, in principle, answer the question based on forces, momentum, etc. In quantum mechanics, there is no answer. It's random.

- But random does not mean without any order at all. There is a pattern—eventually. The middle ground we seek is the idea of probability. The quantum field represents the *probability* of finding the particle in that state.

- There are some deep philosophical issues here, but the truth is that it doesn't matter much for the physics. From a pragmatic standpoint, I think most physicists tend to think in terms of quantum fields while recognizing that the particle viewpoint is occasionally necessary.

Electromagnetism in Quantum Mechanics



$$p + qA = h/\lambda$$

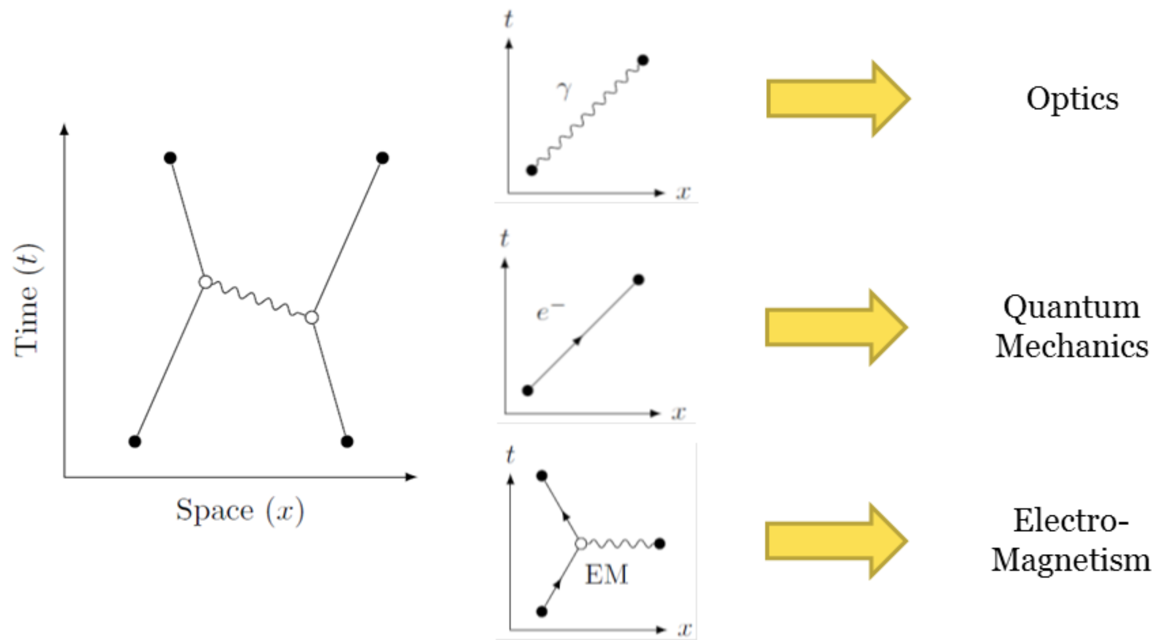
- Feynman wrote: "There are many changes in concepts that are important when we go from classical to quantum mechanics ... Instead of forces, we deal with the way interactions change the wavelengths of waves."

  (From p. 6, "Collective Electrodynamics" by Carver Mead quoting Feynman from p. 15-8 in Ref. 5)
- Previously we used the Lorentz force to see how the fields affect the motion of charged particles. In quantum mechanics, we seek a way to interpret the action of these fields on the wavelengths of the particle-waves.
- We already know that the wavelength is related to momentum. In the presence of the electromagnetic field the de Broglie wavelength formula should be rewritten as

$$p + qA = h/\lambda$$

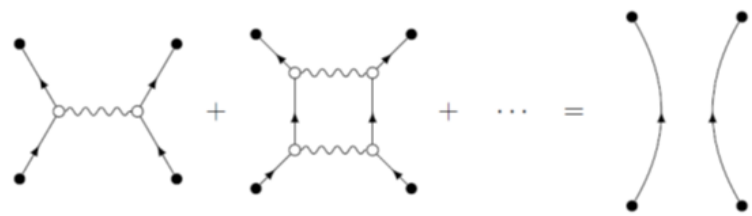
- The term qA/c is sometimes refered to as the **electrodynamic momentum**. It acts as a kind of relativistic generalization of potential energy qV. (See page 244 from Raymond's "Radically Modern Approach, Vol. 2".)
- We could recast electromagnetism from the beginning using the magnetic vector potential. This approach treats electromagnetism as a **gauge theory**. Classically though, this is merely an exercise in mathematics. The electric and magnetic forces are "real".
- But in quantum mechanics, the roles are reversed. The **Aharonov-Bohm** effect shows that it is possible for the magnetic vector potential to affect the outcome of an experiment without the magnetic field by combining a double slit and a solenoid.
- The interference pattern that results depends on the current in the coil even though the magnetic field is concentrated inside the solenoid. The vector potential affects the particle field through the revised de Broglie wavelength and affects the interference.
- In this way we are obligated to recognize the fundamental nature of the magnetic vector potential.

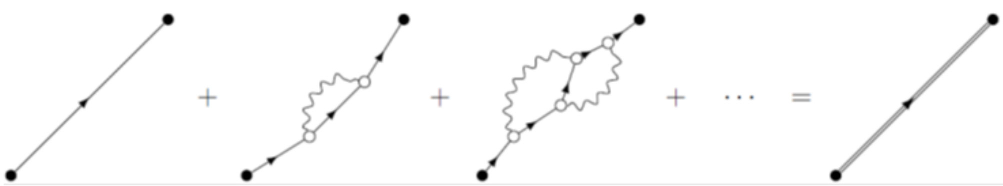
Bosons, Fermions and Feynman Diagrams



- If the electrodynamic momentum is how we introduce the Lorentz force into quantum mechanics, then **quantum electrodynamics** (or QED) is how we introduce Maxwell's laws.
- In the quantum world, every particle field is either a boson or a fermion.
  - The bosons have a tendency to accumulate into the same quantum state. For thermodynamic reasons there is always a distribution spread of the particles across the states, but there is a greater tendency for them to collect than one would expect classically. Photons are examples of bosons.
  - The fermions obey the Pauli **exclusion principle**: there can only be one fermion in any one quantum state for the field. Electrons and quarks are examples of fermions.
- In general, the fermions are what we might call the "material" particles: they are "hard" due to the exclusion principle. The bosons mediate the fundamental interactions between particles. They make up the force fields by their gregarious nature.
- In QED, the electrons exchange photons. The way they do this yields everything we know about electromagnetism. **Feynman diagrams** are the pictures we use to calculate the details.
- The nice thing about these Feynman diagrams is that you can actually see the electrons exchange the photons. However, these diagrams are merely a book-keeping tool. Each diagram represents a whole class of similar diagrams with the internal exchange events located in different places.
- The diagram is used to calculate a complex number with an angle proportional to the total action of the diagram (using the Lagrangian from classical mechanics). Those diagrams representing situations far from the classical prediction typically cancel out.
- In this way, one can justify the classical principle of least action. The diagrams also show us a practical method of correcting our classical predictions by including those diagrams close to the classical result.

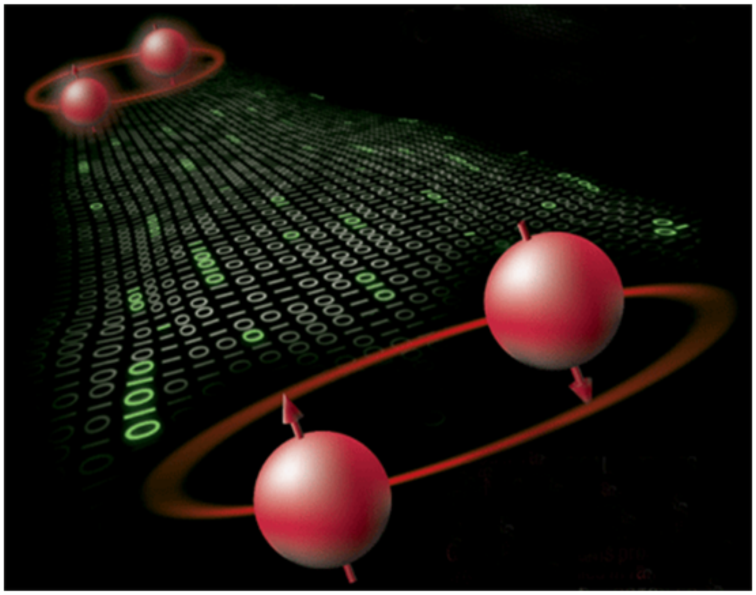
Indistinguishability and Superposition



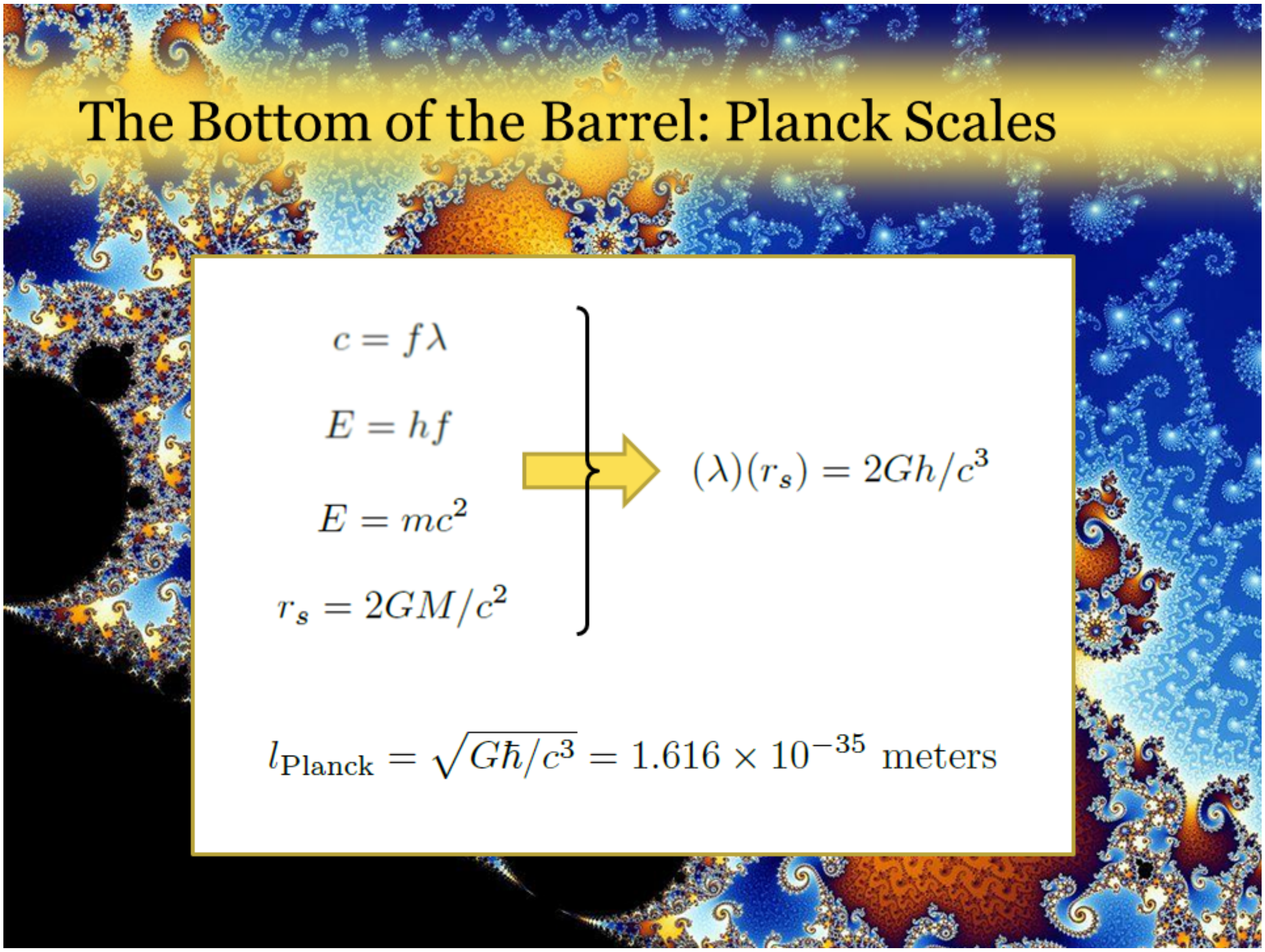


- At its core, atomic theory postulates that everything is made of combinations of elementary particles. The word "atom" comes from the Greek for "indivisible". We now call the 92 elements "atoms", but this is a misnomer—the word subatomic is actually a contradiction: "pieces of the indivisible".
- Though implied in the atomic theory of Democritus, only quantum mechanics fully embraces this axiom through the idea of quantum **superposition**. Let me explain.
- It's a bit unfair to Democritus because we now know something he did not: light is heavy. Specifically, electromagnetic radiation carries both energy and momentum. We cannot simply watch these atoms without disturbing them.
- The light we use must participate in the process we are watching in order to be useful. But light is not a passive participant (this is what Democritus did not know). It will alter the very process we seek to observe. Ultimately, this yields the uncertainty principle.
- A second axiom of atomic theory is that atoms of the same type are indistinguishable—they are *exactly* the same. This is a consequence of the simplicity implied by being small. Take one carbon atom from your cat and another from a nearby supernova. There is no way in principle to determine which is which. They are the same.
- This idea applies also to Feynman diagrams. Two diagrams with the same initial prepared state and final measured states are indistinguishable—regardless of what happens in between. Any "observation" of the internal details would be a different diagram.
- Quantum superposition means that if there are two *indistinguishable* ways for a quantum process to occur, the quantum wave describing that process is the sum of the two individual quantum waves. The overall probability for the process is proportional to the square of the wave amplitude.
- In other words, the true physical process is a kind of quantum average of every possible way we can conceive of it happening, no matter how complex. Anything goes—even moving backward in time.

Entanglement, EPR, and Bell's Inequality



- Quantum superposition is at the root of everything "weird" in quantum mechanics. Fortunately, each quantum wave is driven by the classical formula for action. Through the intermediary of the principle of least action we can connect back to the world we know and love.
- But there are situations in which it's not so simple. We have to work hard to create them, but it can be done. These situations highlight rather than obscure the quantum nature of our world.
- Superconductivity, microelectronics, and lasers are all on the list. But perhaps the most unintuitive thing is called quantum **entanglement**. "Schrodinger's cat" is an example of an entangled system.
- Take two particles and allow them to interact. The final state will be a superposition of possible outcomes based on the internal interaction. For example, consider the momentum of the particles. After they interact, the individual momenta may be different—but I do know the total, because it is conserved.
- This is where the **EPR paradox** comes in. Suppose I measure the position of one and the momentum of the other. Since momentum is conserved, I can calculate the momentum of the first. Haven't I just violated the uncertainty principle?
- There are technical details involving time and relativity here, but the end result is you will not be able to make both measurements. Somehow the measurement on one side "communicates" to the other side.
- This has been tested and verified experiementally, but there is also an argument called **Bell's inequality** that puts an upper limit on the amount of information that can be extracted from entangled systems like this.
- At first, it may appear that we can somehow channel this entanglement to violate the light-speed limit in relativity. Most physicists don't think this can work: you cannot use a *lack* of information to communicate actual information. But all this remains an active area of research.



- A fairly simple argument shows that there is a certain length scale at which the very ideas of space and time break down. The quantities for a photon of light are connected through the following equations:

$$c = f\lambda \qquad \text{and} \qquad E = hf \qquad \text{and} \qquad E = mc^2 \qquad \text{and} \qquad r_s = 2GM/c^2$$

- There is a limit to how precisely we can probe with our photons because precision requires a small wavelength. But energy is required to decrease wavelength. Eventually, this energy is enough to create a black hole. As we dump more energy into the photon to decrease its wavelength, we increase the radius of the black hole.

- We can summarize this length scale conflict by combining the four equations above:

$$(\lambda)(r_s) = 2Gh/c^3$$

- In order to characterize the issue, we define the following Planck scales:

$$l_{\text{Planck}} = \sqrt{G\hbar/c^3} = 1.616 \times 10^{-35} \text{ meters}$$
$$t_{\text{Planck}} = \sqrt{G\hbar/c^5} = 5.391 \times 10^{-44} \text{ seconds}$$
$$E_{\text{Planck}} = \sqrt{\hbar c^5/G} = 1.956 \times 10^{9} \text{ joules}$$

- These are our limits. We cannot physically explore beyond these values *in principle*. Notice how we have used equations from both quantum theory and general relativity. We know that the two theories are incompatible as they stand. What the Planck scales do is gives us a target. These are where some super-theory *must* manifest itself.

- The Planck energy is usually quoted in electron-volts: 1.22×10^{19} GeV. You might wonder whether the Large Hadron Collider will help here. Unfortunately, the LHC is expected to reach energies of (only!) 7500 GeV – nowhere close to the Planck scale. In fact, it's hard to express in words just how far away we really are.

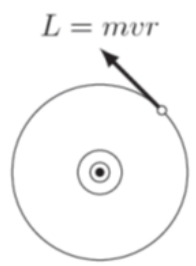
Physics 203 Lecture 6

Atoms and Solid State Physics

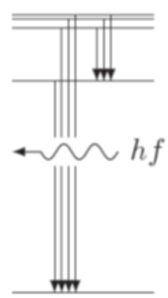
- In the exploration of the quantum world, there are three main categories of evidence:
  - Scattering experiments
  - Particle decay modes and rates
  - Bound states
- The final group is where quantum mechanics started—in particular, the study of atomic structure: electrons trapped in the Couloumb potential of the nucleus.
- So, today we reintroduce quantum theory in a progression that is more historical. We start with the study of the atom, touch briefly on physical chemistry, then segue into solid state physics and microelectronics.

Bohr's Atomic Model

$$\frac{1}{\lambda} = R\left(\frac{1}{n_1} - \frac{1}{n_2}\right)$$



$$E_n = -\frac{13.6 \text{ eV}}{n^2}$$

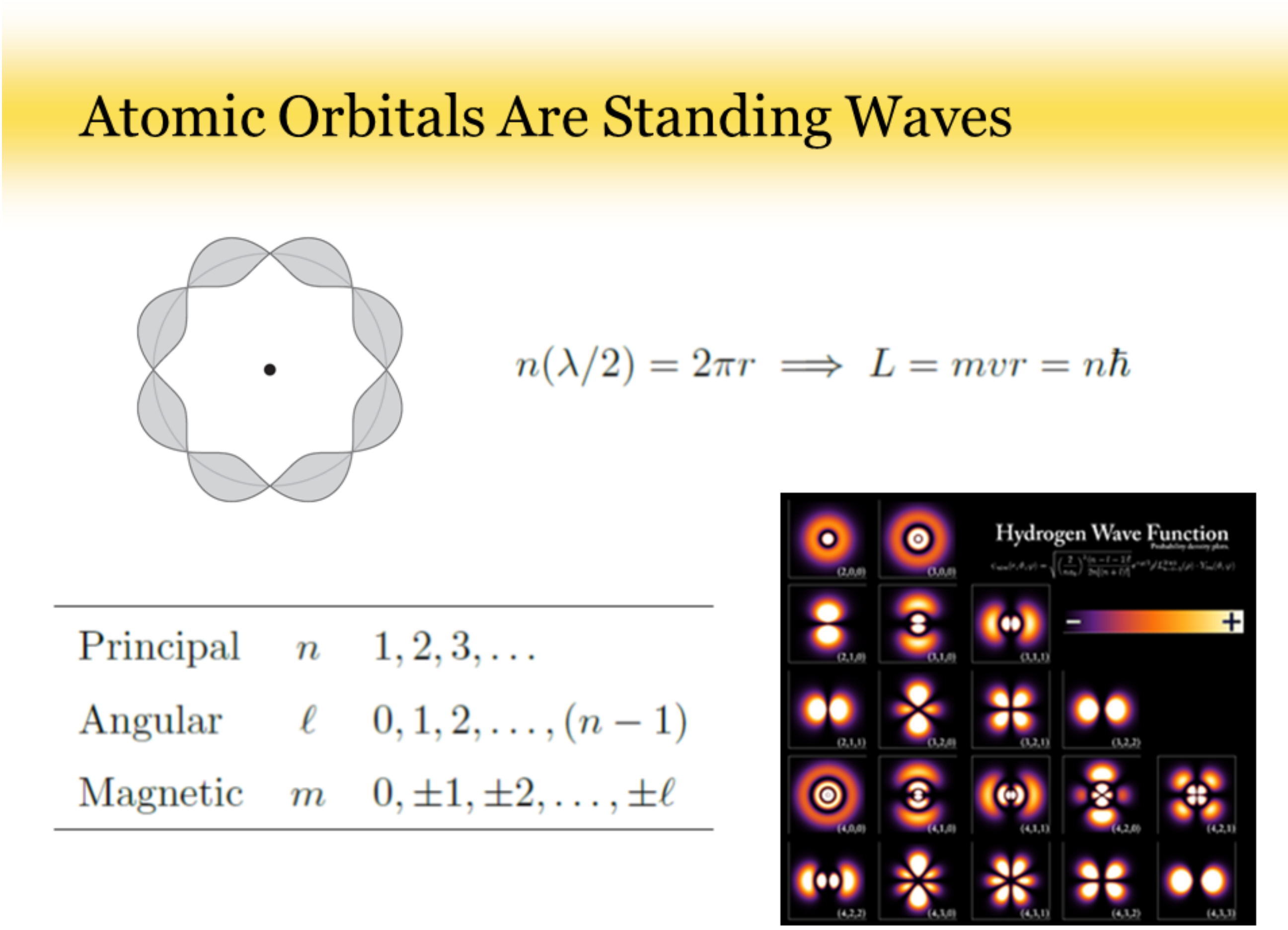


$$\boxed{L = mvr = n\hbar}$$

!?

$$r_n = (0.529 \times 10^{-10} \text{ meters})(n^2)$$

- In 1911 Rutherford discovered (via scattering experiments) that the nucleus of the atom was highly concentrated. A planetary model of the atom immediately suggests itself with the electrons orbiting the positively charged nucleus through electrostatic attraction.
- The issue is that this model is unworkable. Because the electron accelerates as it orbits, it must radiate electromagnetic radiation. The total power loss is given by Larmor's formula. It can be shown that the electron will shed all of its kinetic energy within 10 picoseconds.
- In 1913 Bohr proposed to adapt this model of the atom and simply assumed the atom would not radiate. In order to explain the spectral lines observed in the previous 50 years.
- In the late 1880's a formula for the spectral lines of hydrogen was empirically discovered:
$$\frac{1}{\lambda} = R\left(\frac{1}{n_1} - \frac{1}{n_2}\right)$$
where $R = 1.097 \times 10^7$ and is the most accurately measured physical constant in physics.
- There are four lines that sit in the visible range for hydrogen with $n_1 = 2$ and $n_2 = 3$, 4, 5, and 6. These are called the **Balmer lines** and are useful to detect the presence of hydrogen gas in interstellar space.
- Bohr saw that these wavelengths can be explained if the electron were confined to specific orbital energies. Each line corresponds to a photon emission or absorption during a "quantum jump" from one energy level to another.
- The energy level and corresponding radii for hydrogen must be
$$E_n = -\frac{13.6 \text{ eV}}{n^2} \qquad \text{and} \qquad r_n = (0.529 \times 10^{-10} \text{ meters})(n^2)$$
- These orbits correspond to uniform circular motion of the electron around one proton in which the angular momentum is constrained to be
$$L = mvr = n\hbar$$

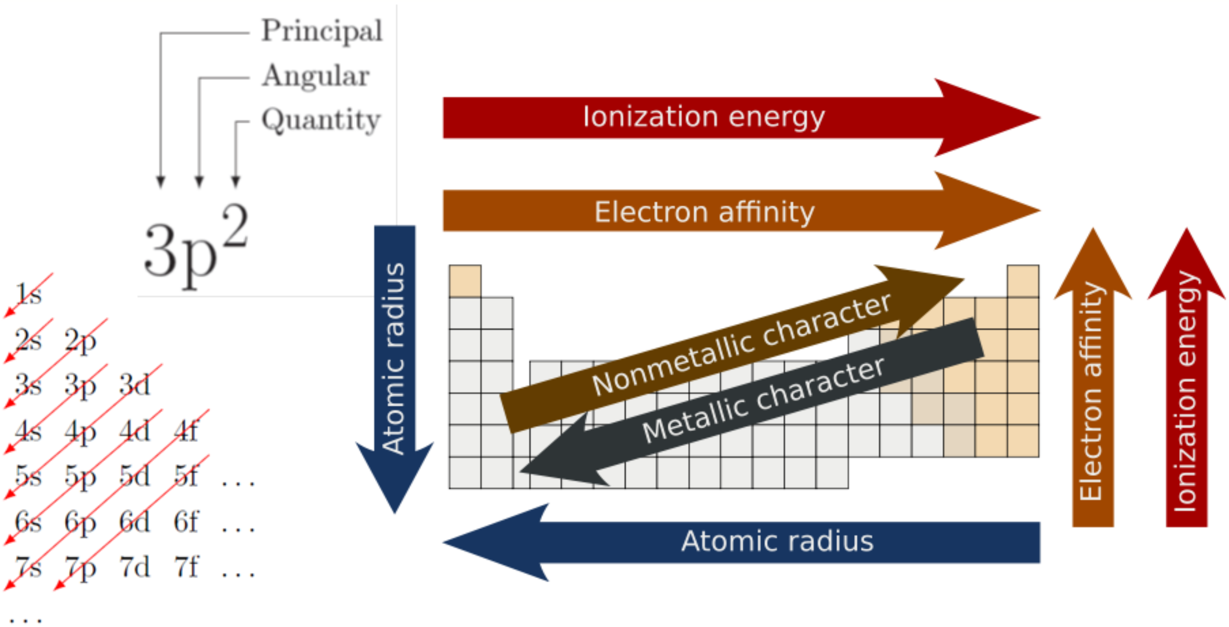


- We can make sense out of Bohr's quantization using the idea of de Broglie matter waves. A circular standing wave around the nucleus will occur when a whole number of half-wavelengths fit the orbit:

$$n(\lambda/2) = 2\pi r \implies L = mvr = n\hbar$$

- Notice that the angular momentum of the standing wave is zero unlike the electrons in Bohr's model. This has a measurable impact (e.g., a magnetic moment) and it is clear that the wave picture is right. Think of the electron surrounding and vibrating around the nucleus instead.
- But even this is too simplistic. The electron field actually vibrates in all three dimensions. In a stable atom, this vibration will occur in a particular vibrational **mode**. These "spherical harmonics" are characterized by whole numbers—one for each dimension of vibration.
- But you've seen them before. These are the electron **orbitals**. Each is defined by three **quantum numbers**:
  - The principal quantum number: n. This number is the same as in the Bohr atom and is responsible for the *size* of the orbital. Values range from one to infinity.
  - The angular quantum number: ℓ. This determines the *shape* of the orbital. These are said to be **degenerate** because they all have the same energy level. Values range from zero to $n - 1$.
  - The magnetic quantum number: m_ℓ. This number describes internal steady *flow* within the orbital. This current gives the orbital a magnetic moment. Values range from $-\ell$ to $+\ell$.
- To complete the picture we must include one more quantum number (four total):
  - The electron's intrinsic spin: m_s. This number only takes on one of two values: $\pm\hbar/2$. I'm not sure how to explain this—it is an extra degree of freedom that doesn't really correspond to anything classical.
- In quantum field theory, this quantum spin is related to the angular momentum of the field and makes the electron a little magnet. In a way, the extra degree of freedom is not surprizing since QFT is a four-dimensional theory.

The Periodic Table



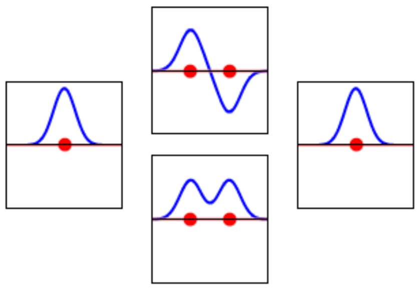
For example: Ca (20 electrons) = $1s^2\ 2s^2\ 2p^6\ 3s^2\ 3p^6\ 4s^2\ 3d^2$

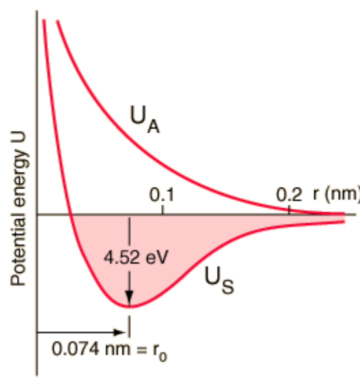
- When one steps back and looks at our modern quantum atomic model, it is amazing to me just how close Democritus was. Not only was he right about the existence of atoms, but his idea was that the shape of each atom determines its properties. We now know that it is the electron orbitals which control the chemistry of each atom.
- For any particular atom, the charge of the nucleus attacts an equal number of electrons to balance the overall charge. However, electrons obey the Pauli exclusion principle: they cannot share the same set of quantum numbers.
- This has the effect that the electrons "stack up" in the orbital chain with the lowest energy orbitals getting filled first. An electron may exist in an **excited state**, but that is temporary—the electron will eventually emit a photon and fall down the chain as far as it can.
- These ideas ultimately explain the structure of the periodic table.
  - Lithium, for example, has three electrons. The first two occupy the $n = 1$ state, each with different values of spin (m_s). The third electron is forced to occupy the higher energy $n = 2$ state.
  - Sodium has 11 electrons. It fills the $n = 1$ and $n = 2$ shells with one left over in the $n = 3$ state. The outer electrons dominate the chemistry and this is why sodium acts so much like lithium.
  - The vertical columns in the periodic table (i.e., the families) align these atoms.
- For historical reasons (going back to spectroscopy), we label represent the angular quantum number with the following letters: s, p, d, f, g. Thus, the electronic structure for lead (82 electrons) is written as:

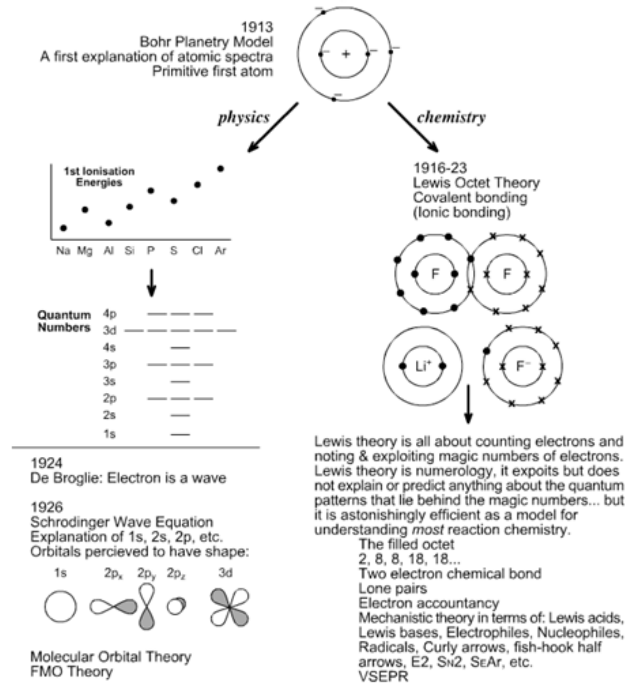
$$1s^2\ 2s^2\ 2p^6\ 3s^2\ 3p^6\ 3d^{10}\ 4s^2\ 4p^6\ 5s^2\ 4d^{10}\ 5p^6\ 4f^{14}\ 5d^{10}\ 6p^2\ 6p^2$$

- The orbitals don't fill in the order you might expect. Interactions between the electrons break the degeneracy of the angular quantum number pushing some energies above others with higher principal quantum numbers.
  - This is also why we have to wait until the fourth period to get efficient conductors (e.g., copper). These are the first elements accessing the "d" orbitals which are much more closely degenerate.

Exchange Energy and the Covalent Bond

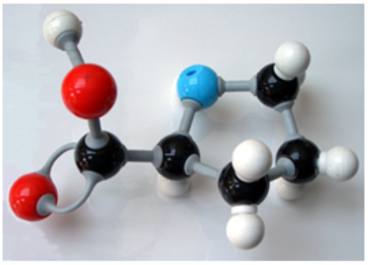


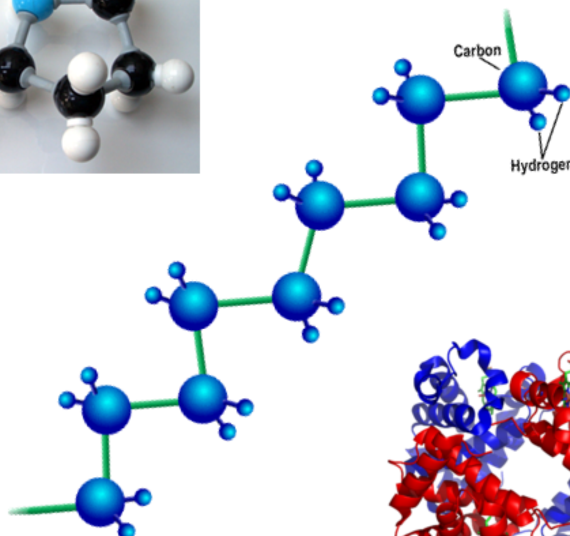


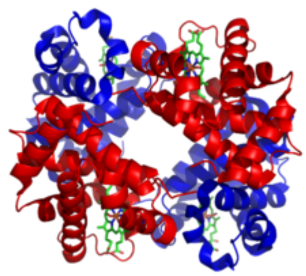


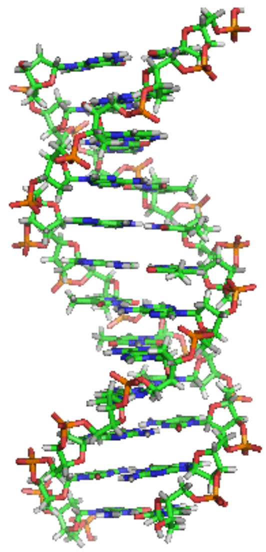
- So much for the atom. What about two? Suppose we pull two hydrogen atoms close to one another. If we pull one close to the other, the electron in the second will push around the energy levels of the first (and *vice versa*). This is what happens to the "p" orbitals in a single atom.
- As the two orbitals start to overlap, they combine into an overall wavefunction describing the pair of electrons. This wavefunction has two states: one in which the spins of the electrons are aligned, and another in which the spins are opposed.
- When the atoms are far apart, these states are degenerate. When the atoms get closer, these energy levels separate: this is how the exclusion principle works. But the important thing here is that energy of the allowed state (opposed spins) is pushed down.
- In other words, the energy of the combination is more favorable than when they are separate. As we continue to push them together, electrostatic repulsion pushes them apart—there is a middle ground in which the energy level is lowest.
- This is how a **covalent bond** works. It is a uniquely quantum mechanical effect rooted in the indistinguishablility of the electrons. The energy decrease here is called the **exchange energy** for the system.
- This hydrogen molecule is called a **sigma bond** because it is made of the overlap of two "s" orbitals. There are "pi bonds" and "delta bonds" from the combination of "p" and "d" orbitals also.
- Ferromagnetism is also a consequence of this exchange energy. You see, in classical mechanics a simple bar magnet cannot exist. Magnetic dipoles always torque one another to cancel out (opposite poles attract). So the atoms in a magnetic material ought to pair up and cancel.
- But in some materials (e.g., iron), the electronic structure is such that a kind of covalent bond occurs. This exchange energy aligns the internal magnetic moments of the atoms. The individual magnetic moments accumulate.

Microbiology and the Size of Life



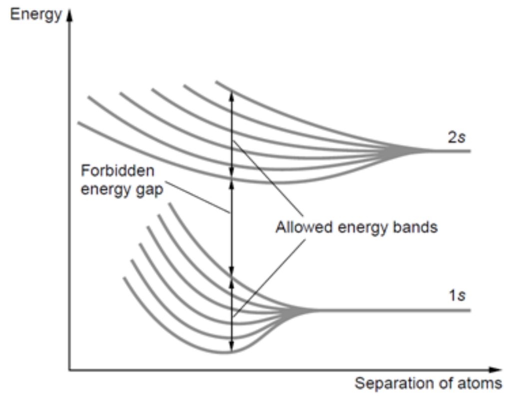


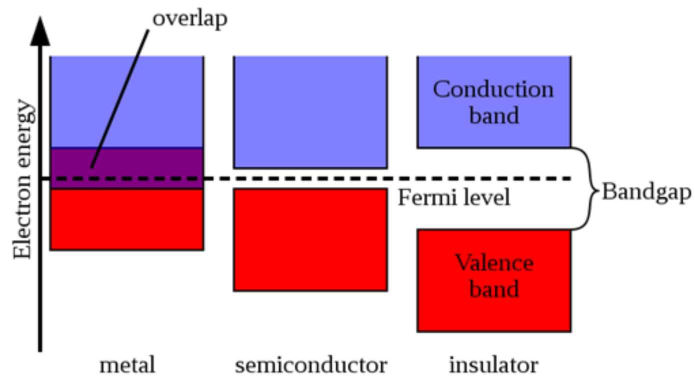




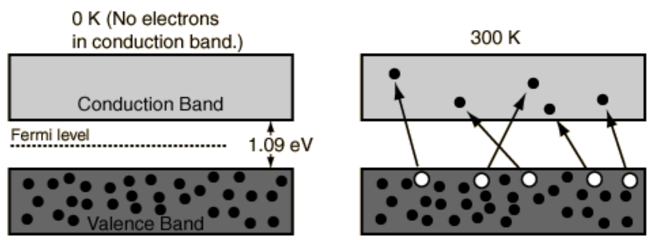
- This slide inspired by "The Force of Symmetry" by Vincent Icke. Highly recommended reading for anyone interested in things that start with a Q. In Chapter 9 he explores some fascinating consequences for microbiology from quantum mechanics.
- The first point is the distinction between ionic and covalent bonding. More gentle and short-range than ionic bonding, the covalent bond means that atoms can act very much like Legos. Molecules have an exactly defined shape and therefore precisely reproducable.
- In addition, molecules do not "wear down". If a molecule is altered for some reason, it becomes an entirely different molecule. There is no ambiguity: they are easy to identify. This means that the chemistry is very specific between them.
- And carbon is special with a valence of four. This makes it possible to form indefinitely long chains of atoms, which form the backbone of organic chemistry. Carbon is also quite abundant in the universe (10x more than silicon).
- Thus, organic chemistry sits between the quantum world of eternal atoms and the classical world of wear and tear. Life is possible in this quantum "twilight zone".
- We can even go farther: genetics also depends on these facts. Since molecules have an exact shape, it is possible for them to replicate. But life requires the ability to interact with the environment to obtain nutrition, repair, defend, etc. For this a larger size will help.
- Of course, we know the answer: DNA. The DNA molecule supports a *symbolic* replication of the organism. For this, we need a lot of storage space. Again the shape of a chain is an ideal comprimise which takes advantage of the quantum world for precise replication.
- This comprimise is at the core of every living creature on Earth and has been so for the last 2.5 billion years. Pretty good for gangly chain of hydrocarbons.

Band Theory of Solids



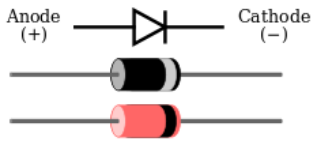


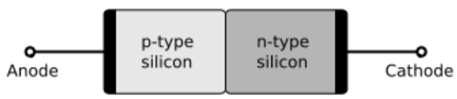
$$\exp(-E_{\mathrm{gap}}/2kT)$$

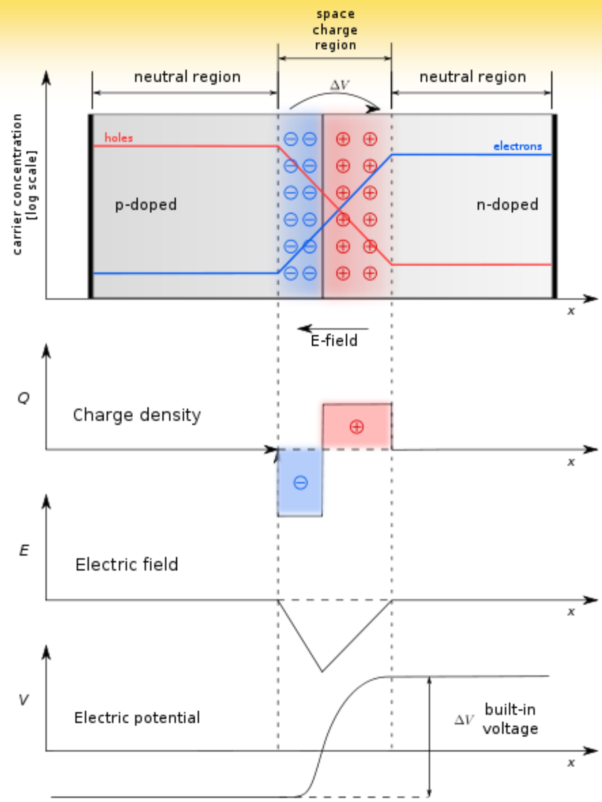


- Consider the description of the covalent bond one more time. Suppose we add another atom to the system. The energy levels will split again. As we continue to add atoms, the individual energy levels form an **energy band** of closely spaced levels.
- Frequently there are gaps between the bands—energy levels inaccessible to the electrons. As the atoms collect into the crystal, the electrons fill the lowest energy bands first. Bands that are filled are called **valence bands** and those that are empty are called **conduction bands**.
- It is possible that the bands overlap—if that occurs between the last valence band and the first conduction band, then the electrons are free to move and we have a conductor. On the other hand, an energy gap creates an insulator.
- This description ignores the effect of temperature. At any temperature, there is a probability that any individual electron will gain enough kinetic energy to overcome the energy gap. The number of electrons that do this is proportional to
$$\exp(-E_{\mathrm{gap}}/2kT)$$
- Because of this exponential factor, this energy gap must be on the order of an electron volt, otherwise this temperature effect is negligible. For example, silicon has a band gap of 1.1 eV, so at 300 kelvin (approximately room temperature) a small number of electrons "evaporate" into the conduction band.
- So even if the bands don't overlap, there is a second way for a material to conduct electricity. Those materials with an energy gap around 1 eV are called **semiconductors**.

The pn-junction

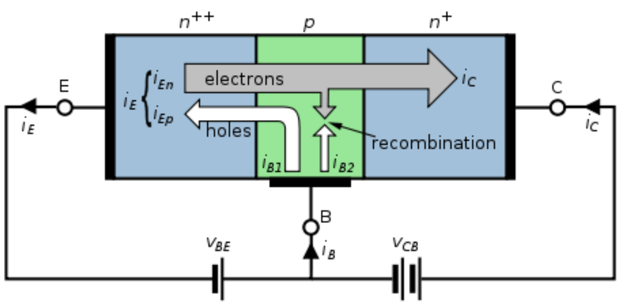


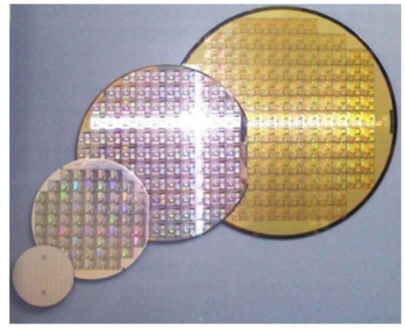


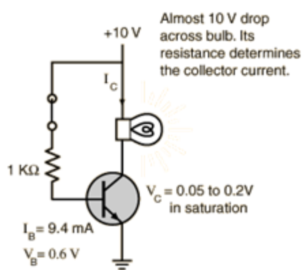


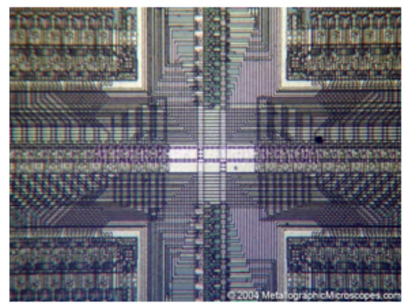
- One of the reasons for the tremendous utility of semiconductors in electronics is the ability to precisely control the electrical properties through **doping**.
- If we diffuse a small amount of phosphorous into silicon we introduce an extra energy level just below the conduction band with extra electrons. This is called an n-type semiconductor because of the addition of negative charge.
- We can create a p-type semiconductor by diffusing boron which introduces an empty energy level just above the valence band. This gives the electrons trapped in the valence band a bit of room to move. If an electron jumps up to this extra energy level it leaves behind a **hole** in the valence band.
- But this is just the beginning. Take p-type and n-type materials and put them together. This is called a pn junction and is fundamental to any integrated circuit. At the interface between the two materials, the free electrons in the n-type materials jump over to holes in the p-type material.
- This charges separation creates an internal voltage across the interface. This opposes the continued flow of electrons and an equilibrium state is reached. This internal voltage is dependent upon the nature of the two materials making the junction.
- If we treat this junction as an electronic component, it is called a **diode**. The diode is said to be **forward biased** if the positive terminal of the battery to the p-type side. The other way is said to be **reversed biased**.
- When forward biased, the battery siphons off the electrons in the p-type material (if the battery voltage is large enough to overcome the intrinsic voltage of the junction, about 0.6 eV for a silicon-based diode). This disturbs the equilibrium and allows current to flow freely.
- When reversed biased, the current is blocked. As the current flows, it contributes to the internal voltage (like a capacitor) and it builds up until it is equal to the battery voltage which stops the current.
- So the pn junction acts like a one-way street for current. When we treat this pn junction as an electronic component it is called a **diode**.

Transistors and Microelectronics









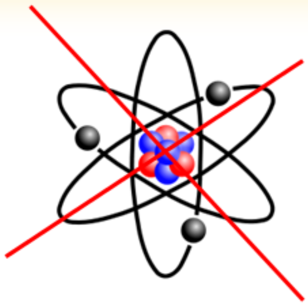
- As current flows in a diode, each electron jumps from the conduction band in the n-type material to the valence band in the p-type material. When the diode is forward biased these levels are separated by an amount equal to the internal voltage at equilibrium.
- This means that each electron drops in energy just like when the electron in an exited atom falls to its ground state. However, not every pn junction will emit photons: the change in angular momentum has to be lined up too.
  - Actually, this is true of the atom also. But typically the angular momentum states are degenerate so there is always a way to get from one energy level to another.
- The band gap is said to be "direct" when this happens and the energy drop is released as a photon (silicon based diodes have indirect band gaps). If the drop creates a photon with a frequency in the visible range, we have a **light-emitting diode**, or LED. Run this process in reverse and you have a solar cell.
- We can keep going too. The transistor which is essentially two diodes placed back-to-back. A *pnp* transistor has an internal voltage pattern that looks like a potential barrier to current.
- Through doping we can control which side has higher and lower potential. The higher end is called the **collector** and the lower is the **emitter**. The central region is called the **base**. By applying a voltage to the base we can vary the height of the potential barrier within the transistor.
- When the base voltage the potential barrier down, current flows from the collector to the emitter (these names are a consequence of the fact that electrons move in the opposite direction because they have negative charge).
- In this way, a transistor can act as a miniature electronically driven electronic switch. And this is how computers are made (multiplied by a few trillion or so).
- Transistors can also be used amplifiers. The base voltage—which uses a small amount of current—can control a large flow of current between the collector and emitter.

Physics 203 Lecture 7

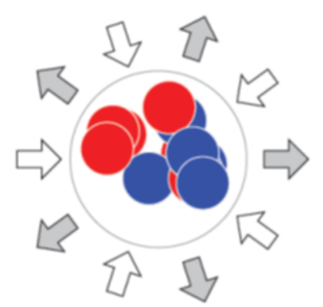
Nuclear Science

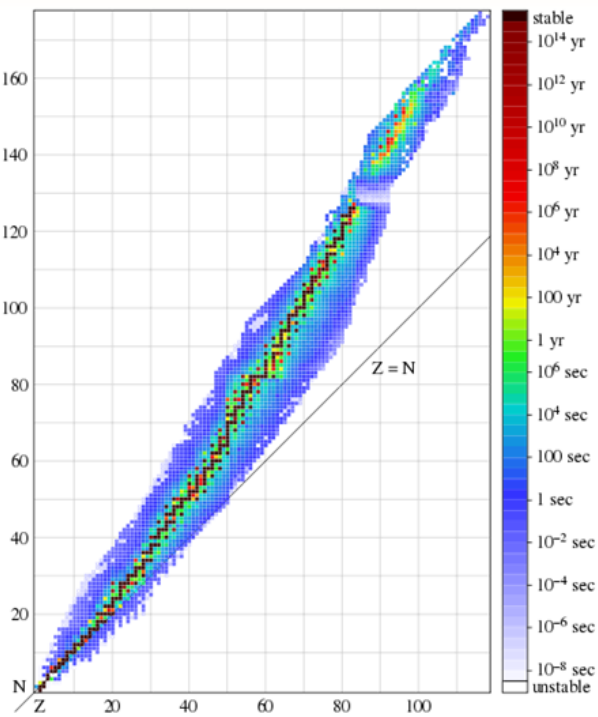
- Coming out of the 1920's the tsunami caused by investigating the atom was somewhat under control. Setting aside the complex mathematics, the picture was quite simple.
- There were three fundamental particles: the electron, the proton and the neutron. The first two are charged and attracted according to the prinicples of (quantum) electrodynamics. Problems related to QED were eventually worked out and the theory takes the form we have today.
- However, the nucleus was a bit of mystery. Clearly there was some nuclear force that hold the protons together against their electrostatic repulsion. Radioactivity shows that the nucleus is slightly unstable. Physicists were starting to see the possibility of somehow harnessing this energy.
- In addition, the particles from cosmic radiation did not seem to fit this simple scheme.
- High energy physics is the result of these investigations. Interestingly, the sub-nuclear world was destined to grow quite confusing before order was restored to the land of physics.
- The last two lectures of the term follow this detective story in a roughly historical way. We will start at the beginning...

Nuclear Structure and Isotopes



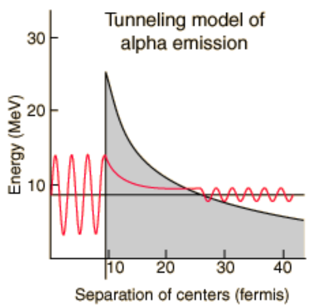
$$r = (1.25 \times 10^{-15} \text{ meters})(A)^{1/3}$$





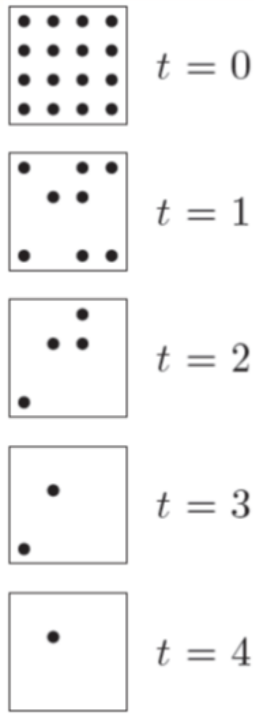
- A casual observation of the periodic table uncovers a problem: the atomic masses do not increase in step with the atomic number. One might expect helium to be twice as massive as hydrogen, lithium three times, etc. But it's roughly 1, 4, 7, 9, 11ish, 12, 14, 16, 19 for the first nine elements.
- Initially it was supposed that, for example, nitrogen had 14 protons in its nucleus (which accounts for the mass) that had somehow swallowed seven of its electrons (which accounts for the charge). However, no arrangement of 21 particles can correctly account for its quantum spin.
- In 1920, Rutherford offered the idea of a third subatomic particle in the nucleus. Without charge and with mass similar to the proton, this **neutron** was later discovered in 1932.
- The neutrons stabilize the nucleus by pitting the strong nuclear force against the electrostatic repulsion of the protons. Although protons do participate in the strong force, their repulsion is too great to allow them to bind together: helium must have at least one neutron.
- But the neutron needs the proton too. In isolation, the neutron is unstable and decays into a proton and an electron with a mean lifetime of about 15 minutes. But in the nucleus the neutron is stable.
- The atomic number of the atom describes the number of protons and drives the chemistry of the element (by attracting the same number of electrons), and the remaining mass is provided by the neutrons.
- We can even explain our "11ish" mass for boron as the average mass of several **isotopes**, or atoms with the same number of protons but a different number of neutrons. Isotopes have different masses but are indistinguishable chemically.
- In the case of boron we have two stable isotopes: boron-10 and boron-11 with five and six neutrons respectively. The relative abundance of the two are 20% and 80% which yields an average mass of 10.8.

Radioactivity



$$N = N_0 \exp(-\lambda t)$$

$$T_{1/2} = \ln 2/\lambda$$



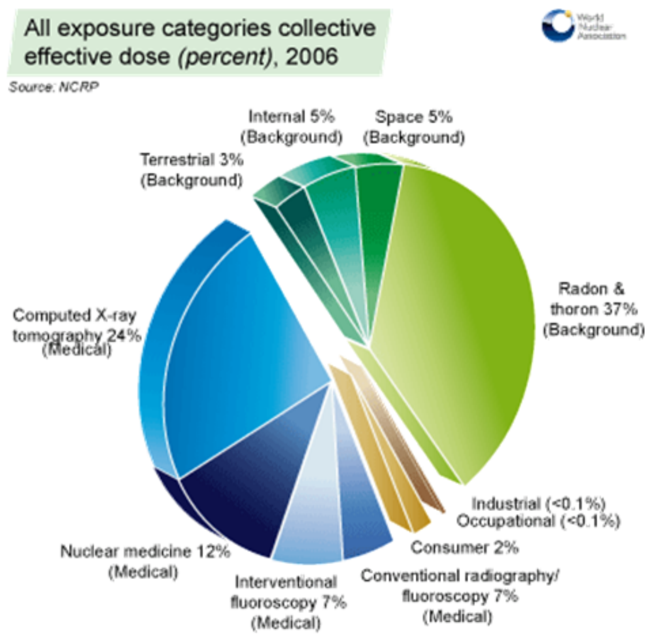
Carbon-14

$$T_{1/2} = 5730 \text{ yr}$$

$$t = \frac{1}{\lambda} \ln\left(\frac{N_0}{N}\right)$$

- Natural radiation is a quantum mechanical effect. The strong and electric forces create a potential energy barrier that holds the nucleus in place. Occasionally, wave functions of the most energetic nuclear material tunnel through this barrier and radiation is the result.
- This probability is the decay constant, λ for the material. It follows that the amount of substance left after a particular time period is
$$N = N_0 \exp(-\lambda t)$$
- Another way to characterize the radioactivity is through its **half-life**, or the time it takes for half of the initial substance to disintegrate. We have
$$T_{1/2} = \ln 2/\lambda$$
- A common way for nuclei to shed excess size is by **alpha decay** which is a pair of protons and neutrons (an helium nucleus). This decay is said to be **transmutation** since it changes the chemical element by reducing the atomic number by two.
- An imbalance between the number of protons and neutrons can also cause instability in a nucleus. If there are too many neutrons, it is typical for one to decay into a proton and eject an electron. This is called **beta decay** and is another transmutation process (adds one to the atomic number).
- If there are too many protons, one may transform into a neutron and eject a positron (subtracting one from the atomic number). This time-reversed, anti-matter twin of neutron decay is also called beta decay. The two particles are often labeled β^- and β^+ to distinguish them.
- Alpha decay is the typical transmutation process for heavy nuclei while beta decay is typical for light nuclei. Other decay processes are possible, but the vast majority are one of these two.
- It is not uncommon for these transmutation processes to leave the nucleus in an excited energy state. Like the atom, the nucleus will emit a photon of electromagnetic radiation to release this energy. These photons are called gamma rays and typically have millions of electron-volts of energy.

Radiation Sickness

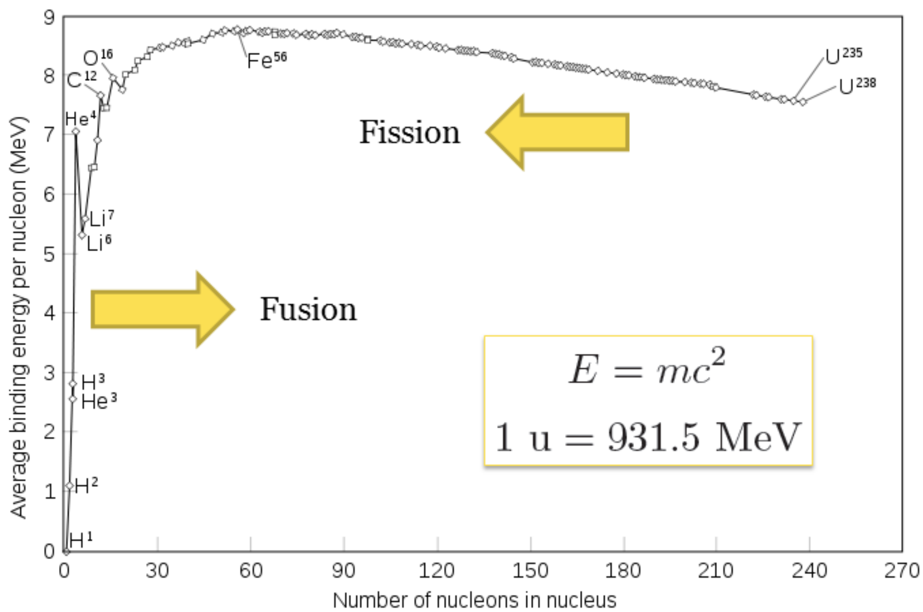


Average American:
6.2 mSv

Dose	Effect
50 mSv	Federal Limit
200 mSv	White Blood
500 mSv	Mild Sickness
1000 mSv	10% Fatal
3000 mSv	50% Fatal
6000 mSv	100% Fatal

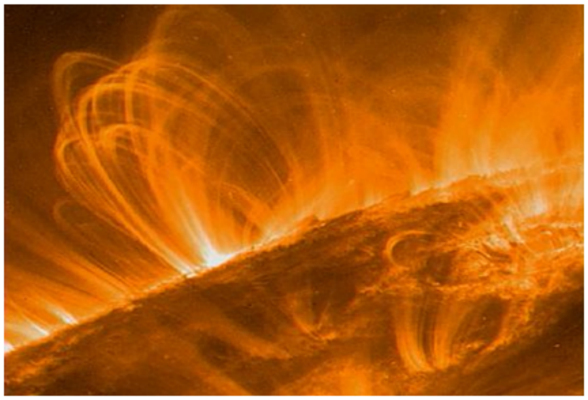
- All three of these types of radiation are classified as **ionizing radiation** because they are energetic enough to ionize an atom (tens of electron-volts). As such they can do significant biological damage in suffcient quantity.
  - Alpha radiation is easily shielded by a piece of paper or even just a few centimeters of air. Beta radiation requires more shielding: five millimeters of aluminum is typical. Three inches of lead is the generally accepted norm for gamma radiation.
- The radiation dose is calculated as the amount of ionizing energy absorbed divided by the mass of the absorbing material. The SI unit of one joule per kilogram is called a **gray**. A more traditional unit is the rad (radioactivity absorbed dose) which is one-hundredth of a gray, so 100 rad = 1 gray.
- But not all radiation has the same effect on our physiology. Dose for dose, alpha particles and other nuclear fragments do 20 times more damage than beta particles and gamma rays. This variation is summarized in a number called the **relative biological effectiveness**, or RBE.
- The absorbed dose multiplied by the RBE gives us the **biologically equivalent dose** for the radiation exposure. Based on the gray, the SI unit is called the **sievert**. (It is much more common, though discouraged, to see this equivalent dose quoted in "rem" which is based on the rad.)
- The rule of thumb is that one sievert will cause nausea and over six is lethal.
- Our natural background radiation is approximately 2.4 millisieverts per year.
  - About half from radon gas with the rest split evenly between
    * Cosmic radiation,
    * Terrestrial radiation, and
    * Food and water.
  - We are also exposed to an additional 0.4 millisieverts through various medical procedures.
  - See http://www.unscear.org/docs/reports/gareport.pdf for details.

Nuclear Energy

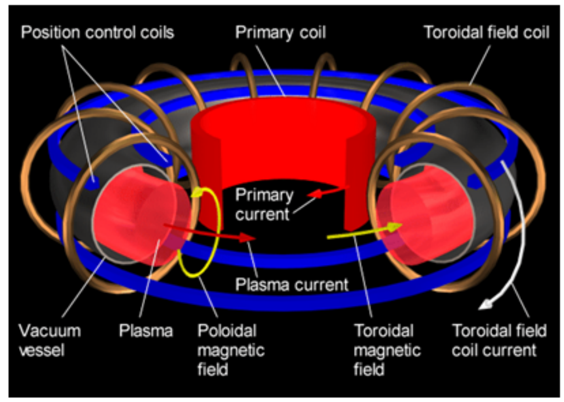


- We now consider how to induce various nuclear interactions. The most important item to consider is the nuclear **binding energy curve**. This curve represents the net energy needed to break apart a stable nucleus into its component parts.
- Iron-56 is the most stable of all nuclei with a binding energy of 8.8 MeV per nucleon (the catch-all term to describe both protons and neutrons). The high binding energy for helium-4 (at 7.1 MeV per nucleon) explains why heavy radioactive nuclei tend to decay using alpha particles.
- Fusion pushes us to the right on the curve, fission to the left. Both are possible sources of power because there is a release of energy on either side of iron-56. Fission is the kind of power used in modern nuclear power plants. Fusion powers the sun.
- In either nuclear fusion or fission, we could use the binding energy curve to calculate the net energy requirements for the process. However, there is an easier way using the so-called **mass defect**.
- A differences in mass between the sides of a nuclear reaction is simply a consequence of $E = mc^2$ and the binding energy involved. Every difference of one atomic mass unit corresponds to 931.5 MeV of energy. When one takes into account the "mass" of the binding energy, the equations balance.
- The first man-made fission reaction involved bombarding uranium-235 with neutrons. On average, the fission of uranium-235 yields 215 MeV of energy with 2.4 neutrons as a byproduct. The amount of energy generated by the fission of one kilogram of uranium-235 could easily provide the energy needs of one person for a lifetime.
- However, the extra neutrons are typically too energetic to initiate a secondary reaction unless the concentration of uranium-235 (relative to uranium-238) is in excess of 90%–which is weapons-grade uranium.
- To compensate, a moderator (like graphite) can be used to slow the neutrons down in order to create a controlled chain-reaction. Nuclear reactors work this way with a smaller concentration around 20% and harness the resulting heat energy through a high-efficiency steam engine.

Nuclear Fusion



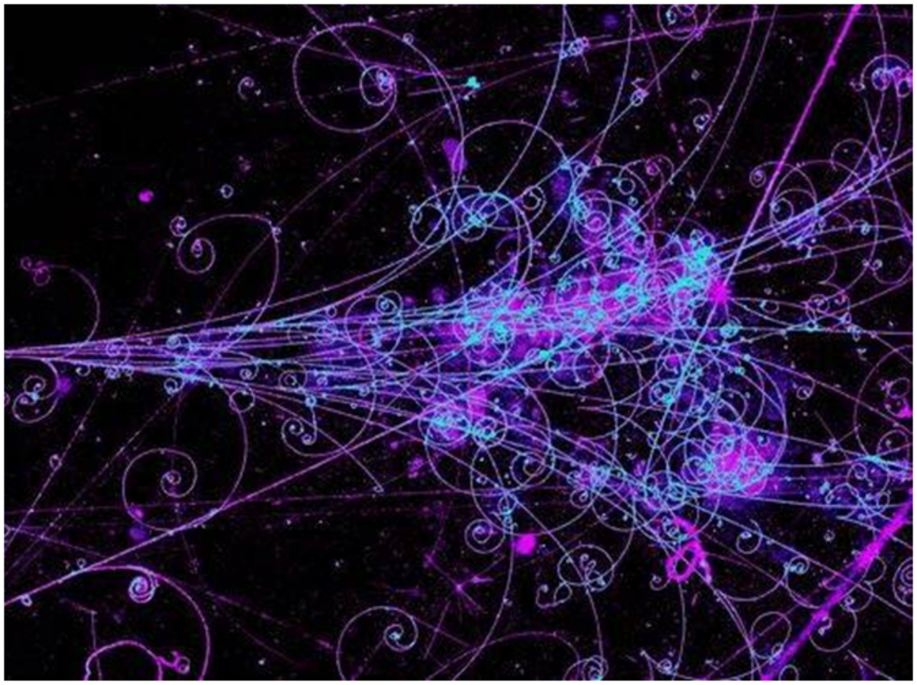
$$^{2}_{1}\mathrm{H} +^{3}_{1}\mathrm{H} \rightarrow^{4}_{2}\mathrm{He} +^{1}_{0}\mathrm{n}$$



$$(1)\ {}^{1}_{1}\mathrm{H} + {}^{1}_{1}\mathrm{H} \rightarrow {}^{2}_{1}\mathrm{H} + {}^{0}_{1}\mathrm{e} + \nu$$
$$(2)\ {}^{1}_{1}\mathrm{H} + {}^{2}_{1}\mathrm{H} \rightarrow {}^{3}_{2}\mathrm{He} + \gamma$$
$$(3)\ {}^{3}_{2}\mathrm{He} + {}^{3}_{2}\mathrm{He} \rightarrow {}^{4}_{2}\mathrm{He} + {}^{1}_{1}\mathrm{H} + {}^{1}_{1}\mathrm{H}$$

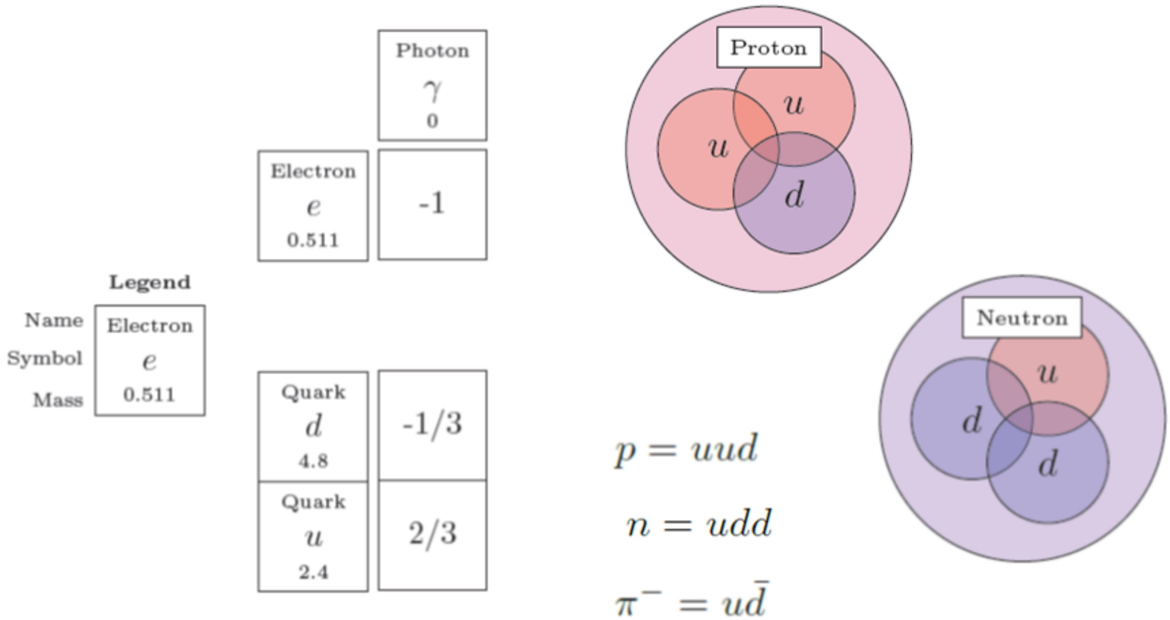
- The problem with nuclear fission as an energy source is the radioactive waste products. On the other end of the scale we have fusion which powers the sun.
- Also, fusion packs more power per kilogram of fuel than fission. This can be seen by comparing the slopes on the left and right of the binding energy curve.
- One of the most promising nuclear reactions for fusion on earth is the combination of deuterium and tritium:
$$^{2}_{1}\mathrm{H} +^{3}_{1}\mathrm{H} \rightarrow^{4}_{2}\mathrm{He} +^{1}_{0}\mathrm{n}$$
- This reaction produces 3.52 MeV per nucleon. The fission of uranium-235 is about 0.915 MeV per nucleon, so pound-for-pound this reaction is almost four times more powerful—with helium as the main waste product.
- There are two problems with this approach. The first is the fuel: deuterium is relatively plentiful and could be extracted from sea water, but tritium has a half-life of about 12 years (decay product is helium-3). It would have to be generated (by bombarding lithium-6 with neutrons). This would take some work, but is possible.
- But the real problem is that no one has yet been able to control these fuels with enough precision to consistently overcome the “activation barrier” to release the fusion energy. Plasma is electrically active which makes it extremely unstable and unwieldy.
- The **tokamak** uses a toroidal (donut) shaped magnetic field to perform the necessary confinement. Several experimental tokamaks have been constructed with some success and research is ongoing in order to produce a workable structure.
- Another approach is to fire a high-energy laser or electron beam at a small pellet of the deuterium/tritium fuel. The beam burns off the outer layer explosively which drives an implosion of the inner layer—just like a tiny atomic bomb.
- And the universe has a third way: we call them stars.

The Particle Zoo



- Our humble neutron, which appeared as an add-on to the structure of the atom, turns out to play a central role in both the creation of nuclear fuel and the unleashing of nuclear energy.
- But I've left out one important fact regarding neutron decay. When the neutron decays three particles are produced not two. The extra particle is called a **neutrino** because it has no electric charge and is nearly massless.
- In fact, the neutrino wasn't the only particle known to exist in the 1930's that really didn't need to be. Studies of cosmic radiation revealed the existence of other short lived particles label muons, pions, etc.
- We now understand this as a consequence of high-energy interstellar particles (90% protons, 9% alpha particles) slamming into the Earth's atmosphere. The resulting collsions create a **particle shower** which include these more exotic particles.
- This "shower" is a result of the kinetic energy of the incoming particle being converted into mass via $E = mc^2$. The distribution of particles created depends on their relative mass and quantum uncertainty.
- In the 1950's, particle accelerators were designed with sufficient energy to study these high-energy nuclear reactions. And the more studies were done the more confusing the situation became.
- Frequently new exotic particles were formed in the collision of electrons or protons. Over time, this continual discovery of new particles stirred up quite a bit of discomfort among physicists.
- Around 1930, it seemed that the world could be explained with only three fundamental particles—by the late 1960s, there were hundreds of "fundamental" particles. Clearly something needed some explaining!

Quarks



- The fact that the neutron can decay seems to imply that there is still some substructure to be found. On the other hand, relativity teaches us that the creation of matter from pure energy cannot be ruled out. However, there is more direct evidence of substructure: the neutron has a magnetic moment.
- This is the clue we need. There is another periodic chart underneath this nuclear particle zoo.
- It took some time, but eventually physicists became convinced that the proton and the neutron are both composed of a triplet of still smaller particles called **quarks**.
- In order to make the scheme work, we need two different particles: one with a charge of $-\frac{1}{3}e$ and the other with charge $+\frac{2}{3}e$. The first is called the "up" quark and the second is called the "down" quark.
- So a proton has two up and one down quarks and the neutron has one up quark and two down:

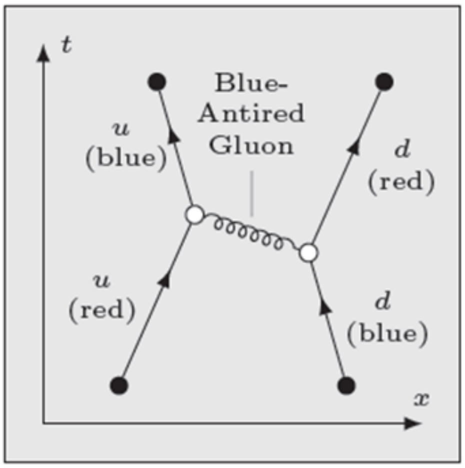
$$p = uud \qquad \text{and} \qquad n = udd$$

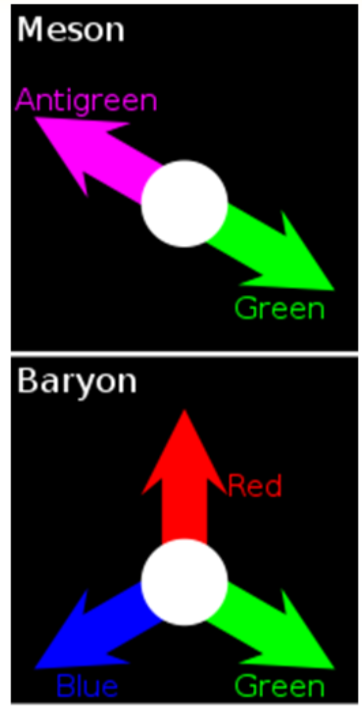
- The down quark is more massive than the up quark. This is why the neutron is slightly heavier than the proton and also why it can decay. The proton is stable. The Pauli exclusion principle prevents it from decay into a uuu combination.
- Almost all the particle zoo can be explained as combinations of these quarks and their antiparticles. For example, the negative pion is

$$\pi^- = u\bar{d}$$

  where the overbar indicates the antiparticle with the opposite charge.
- Actually, there are six quarks: two more massive and unstable copies of the original pair. The first copy is "charm" and "strange". the second copy is "bottom" and "top".
- These copies are called **generations**. Whether there are more than three generations is still an open question (though most think this is the end).

Three Dimensional Color





- These quarks combine in one of two ways. Triplets like the proton or neutron are called **baryons**. Combinations of quark-antiquark pairs like the pion are called **mesons**.
- The quarks all participate in the strong interaction, which is what hold the nucleus together. The quantum theory for the strong interaction is very similar to QED with two exceptions. The first has to to do with the nature of its charge:
  - The strong charge has three dimensions. These dimensions are labeled red, green, and blue. Every quark will possess one and only one of these colors at any one time. Antiquarks carry corresponding anticolors.
  - The method to this madness is that these colors interact in order to produce a "white" particle. This can be done in two ways. First by combining red, green, and blue: these are the baryons. The other is by combining red and anti-red, for example: these are the mesons.
- The second difference is that the strong boson which mediates the interaction (called a **gluon**) exchanges color. This means that it carries color and also participates in the interaction—unlike the photon which mediates but is not subject to electromagnetic force.
  - This gives the theory a non-linearity making it extremely difficult to solve. On the other hand, the strength of the interaction makes it very difficult to approximate.
  - Since the gluon interact with themselves, they should be able to form bound states called **glueballs**. According to theory we should see them with current collider technology, but they will act very similar to the mesons. As of 2011, they have not been identified with certainty.
- One unusual consequence of all this is that the strong force does not get smaller when we separate the quarks. In fact, the force is somewhat constant (around 10,000 newtons). The energy used to overcome this force eventually exceeds the amount required to create another quark/anti-quark pair.
- As a result, one never sees a single color-charged quark. Instead, the result is an explosion or "jet" of color-neutral baryons and mesons. This is called **quark confinement**.

Physics 203 Lecture 8

Loose Ends

- This final lecture will take us to the edge of what we know about physics—and a bit beyond.
- We will finish off the Standard Model by introducing the weak force and its role in the decay processes we have been discussing so far.
- We will follow up with some short-comings of the Standard Model and why physicists continue to work on the "next" theory. String theory being the most famous example.
- Finally I'd like to finish off with a quick word or two about cosmology.

A New Periodic Table

Legend

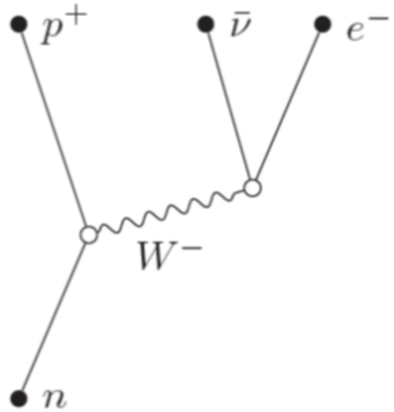
Name	Electron
Symbol	e
Mass	0.511

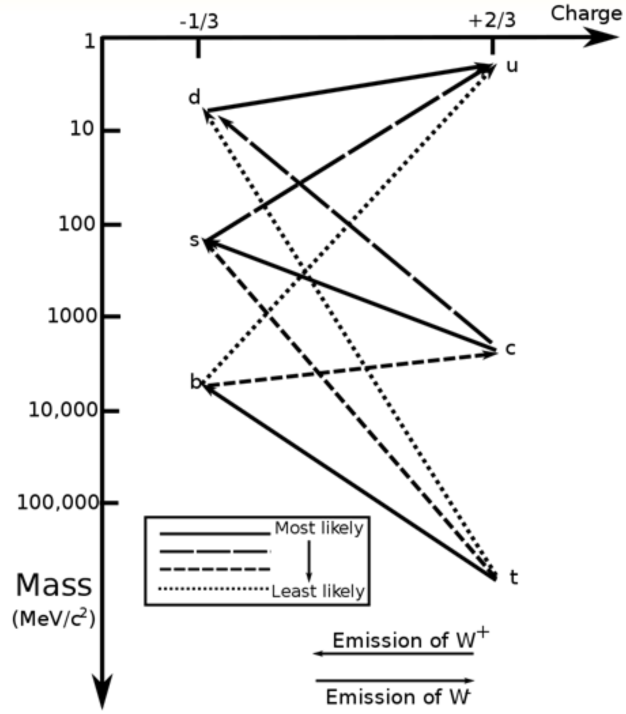
Spin 1/2 particles (Fermions) / Spin 1 particles (Bosons)

Spin 1/2 particles (Fermions)			Photon γ 0	Gluon g 0	Weak W $\sim$ 80000	
Tauon τ 1800	Muon μ 106	Electron e 0.511	-1	0	$-\frac{1}{2}$	Leptons
Neutrino ν_τ > 0	Neutrino ν_μ > 0	Neutrino ν_e > 0	0	0	$+\frac{1}{2}$	Leptons
Quark b 4200	Quark s 100	Quark d 4.8	$-\frac{1}{3}$	g	$-\frac{1}{2}$	Quarks
Quark t 170000	Quark c 1300	Quark u 2.4	$+\frac{2}{3}$	g	$+\frac{1}{2}$	Quarks

- It's been a long and winding road, but we seem to have a good understanding of what the nucleus is and what holds it together. We have six quarks with both electric charge and strong color. Each of these correspond to a different interaction through either photons or gluons.
    - The force that holds the nucleus together is actually the *residual* force from this strong quark interaction similar to the way a magnet sticks to your refigerator.
- However, there is more to the story. The muon (first discovered in 1936) was first thought to be a meson due to its moderate mass: about 10 times lighter than the proton.
- However, it was later determined that the muon possessed every property of the electron except it was 200 times more massive. In 1975, a third twin was found called the "tauon".
- In addition, each of these electron-like particles has a neutrino side-kick. Collectively these six particles are known as **leptons**. They are *not* affected by the strong interaction.
- This is where we stand today. Twelve fundamental particles of matter (and their antiparticles). These form the building blocks of the **Standard Model** of particle physics.
- The interactions between these fermions are mediated by bosons using the rules of quantum field theory. But we have one loose end to tie up: the weak nuclear interaction.

The Weak Force Transforms Particles



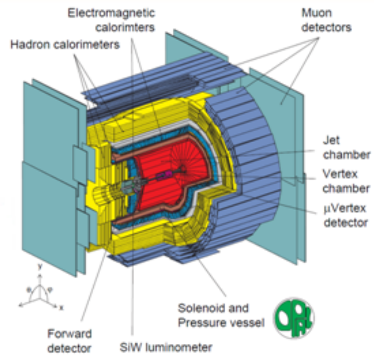


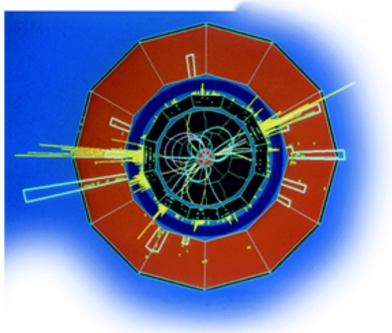
- Neutron decay illustrates that there is some probability for a "d" quark to spontaneously decay into a less massive "u" quark. In the process it sheds an electron and a neutrino.
- We now seek an explanation in terms of quantum field theory for this decay. This means we need to define the vertex—and a new boson to mediate the interaction. We will call this the **weak boson**.
- Again, we want to treat this decay as an interaction. On one side the "d" quark emits the weak boson. The "u" quark is left and the weak boson travels for a bit. It then explodes into an electron and a neutrino.
- If this were a diagram from QED, we would interpret the decay of the weak boson as a matter-antimatter creation event. This is why the neutrino is labeled as its anti-particle.
- But the unusual characteristic of this diagram is that it pairs *different* particles together: the "d" with the "u" quark and the electron with the neutrino. The weak boson *transforms* the particles.
  - Notice that this boson must be electrically charged since it converts a charged object (the electron) into a neutral one (the neutrino). This means it has an anti-particle twin with the opposite charge.
- In order to keep track of these details, the "u", "c", "t" quarks and the neutrinos are said to have a **weak isospin** of $+\frac{1}{2}$ while the "d", "s", "b" quarks and the electron-like leptons have weak isospin $-\frac{1}{2}$.
- The reason for this is to let the weak boson carry one unit of weak isospin. So when the "d" quark decays the W^- boson carries away a negative unit of isospin. This can then interact with a neighboring neutrino by flipping it into an electron, or by simply decaying into an electron and an antineutrino (each with $-\frac{1}{2}$ isospin).
- Usually, the weak boson supports a transformation within generations, but occasionally also across generations. It never mixes quarks and leptons. The probabilities to go from one particle to another is captured in the **CKM matrix**.
- The weak interaction never mixes quarks and leptons.

Massive Bosons and the Higgs Mechanism

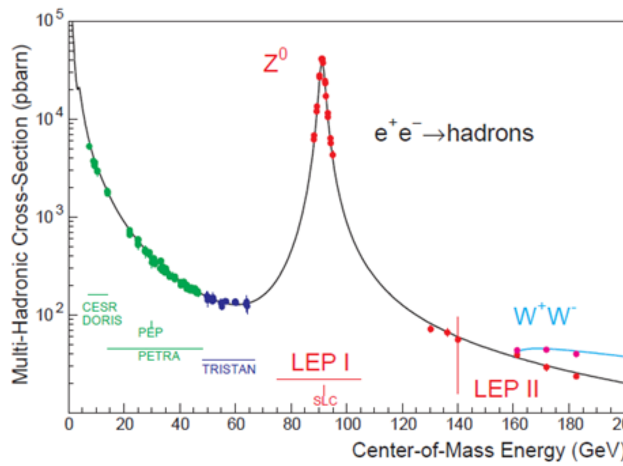
- There is a significant problem in describing these decay processes as a quantum field interaction that we must now confront. The mass of the neutron is slightly larger than the proton. This difference in mass must be carried away by the boson.
- It's possible that this is done purely energetically (like a high energy gamma ray), but the details indicate that the weak bosons are massive. This contributes to the short-lived and short-range nature of the weak interaction.
- The problem is that bosons cannot carry rest mass.
- In 1967 a way out was discovered that involves two independent concepts. The first is the unification of the electromagnetic and weak interactions. In this electroweak theory we have four bosons (W^+, W^-, W^0, and B^0) all of which are initially massless. Not much help so far.
- The second concept is called the **Higgs mechanism** which postulates the existence of yet another entity called the Higgs field. The idea runs like this...
  - At very high temperatures the four electroweak bosons and the Higgs field exist in a kind of equilibrium. But as the temperature drops, a level is reached at which this equilibrium goes unstable and the Higgs field "mixes" with the electroweak bosons.
  - This causes the (massless) weak bosons to "drag along" the Higgs field creating a kind of inertia—like pulling a spoon through syrup. We interpret this as rest mass.
- This approach may look convoluted and unmotivated. It is inspired by the BCS theory of superconductivity. A superconductor is a system involving massive particles (electrons) acting like bosons about which we know quite a bit. So it makes sense that this theory might act as a template to understand the weak boson.
- But the details are still in debate. There may be ways to "break" an initial electroweak symmetry without the Higgs mechanism. The Higgs boson has yet to be observed—this is one of the main objectives of the Large Hadron Collider.

Electroweak Unification



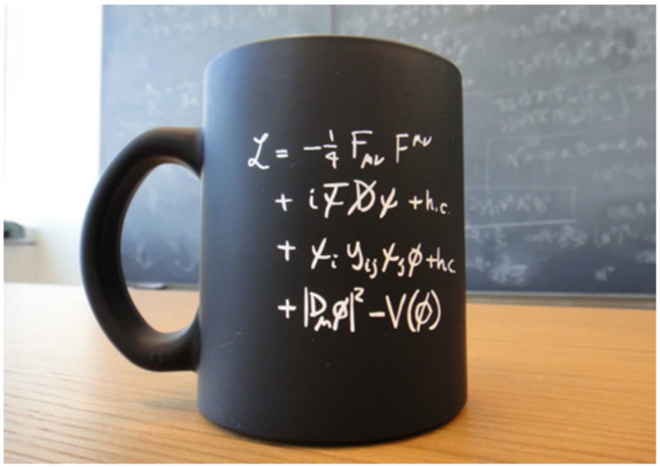


Boson	Measured	Predicted
W	80.398 ± 0.025	80.390 ± 0.018
Z	91.1876 ± 0.0021	91.1874 ± 0.0021



- The whole point of the Higgs mechanism is to explain the "broken symmetry" between the electric and weak interactions thereby giving the weak bosons mass. Most physicists expect the Higgs boson to found eventually—but we've been wrong before.
- Regardless of the mechanism, the "mixing" of the electroweak bosons cause the W^+ and W^- bosons to acquire mass. Current measurements indicate a mass of 80.398 ± 0.025 GeV/c^2. Current theory yields 80.390 ± 0.018 GeV/c^2, which is a great match.
- After mixing, the B^0 boson remains massless and becomes the photon.
- The final W^0 boson is left. It also acquires a mass, but remains without charge. We call it the Z boson (presumably due to its zero electric charge).
- The Z boson has been observed (about 92 GeV/c^2) and acts like a heavy photon. It doesn't carry electric charge or weak isospin, so it doesn't change the nature of the particle like the other weak bosons.
- For this reason, it's hard to observe. Its mass makes it much weaker than its more powerful cousin the photon so the electromagnetic interaction overwhelms its effect. The Z boson is the runt of the lot among the electroweak bosons.
- But neutrinos are leptons without electric charge. They do not participate in either the strong or electric interactions—only the weak interaction. If we fire a beam of neutrinos into a bunch of electrons, any collision that doesn't result in new particles is from the mediation of this Z boson.
- One huge source of neutrions is the sun. We are literally bathed in an enormous amount of neutrinos raining down from the sun from every nuclear reaction. Nearly all simply pass through the earth into space because they *only* interact via the weak bosons.

The Standard Model and Its Problems



The Standard Model WORKS !

- ★ The Standard Model describes ALL experimental observations
- ★ ALL particles of SM have been discovered with the exception of the Higgs.
- ★ Highly predictive theory - tested to high precision at an energy scale of $\sqrt{s} = 100$ GeV

BUT many many questions

- ★ Too many free parameters (over 20): G_F, M_{Z^0}, α_{em}, α_S, 12 fermion masses, quark and ν mixing matrices, the Higgs boson.
- ★ The Standard Model is just that - a model rather than anything more fundamental.
- ★ Do not understand the origin of fermion masses.
- ★ Why 3 generations ?
- ★ Are leptons/quarks fundamental - substructure ?
- ★ The Higgs model has problems $\longrightarrow$ huge cosmological constant.
- ★ Need to unify all forces : GRAND UNIFICATION
- ★ Ultimately gravity needs to be included
- ★ + many other fundamental questions......

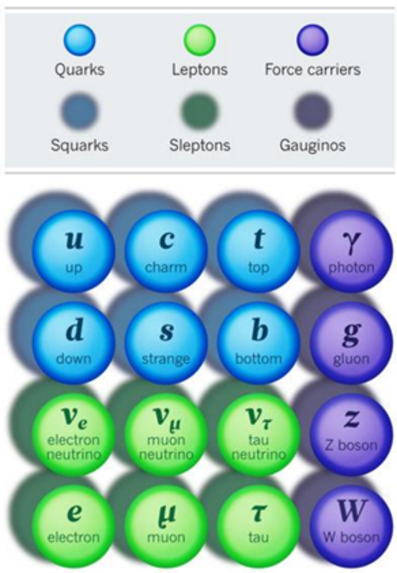
Over the course of the last 30 years out understanding of Particle Physics has changed beyond recognition.

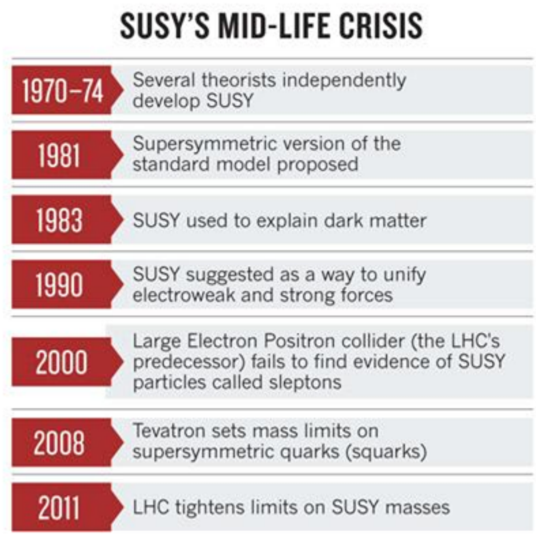
Through precise measurements and powerful theoretical ideas our understanding is still evolving.

In the next few years hope for yet more surprises!

- We now have in place all the pieces that make up the **Standard Model** of particle physics.
  - Twelve fundamental fermions in three generations with two quarks and two leptons each. These particles participate in three interactions each mediated by bosons: the photon, gluons, and the weak bosons.
  - We may also have a Higgs boson. Almost certainly we should consider the photon and the weak bosons as deeply related in some primordial sense.
- The Standard Model works and is extremely successful in predicting the results of high-energy physics. Nonetheless, few physicists would call this pretty and nearly all expect its complexities to disappear in some overarching "Theory of Everything". Why?
- The short answer is that the model appears incomplete. Clearly we need to incorporate gravity yet it is not clear how this can be done. The Higgs particle has yet to be found—what if it's not there?
- Also, there are things that happen in nature that the original model does not anticipate.
  - **Parity violation**. Surprisingly, the weak interaction makes a subtle distinction between matter and anti-matter. This could explain the apparent imbalance in the universe between the two.
  - **Neutrino oscillation**. Discovered in 1998, solar neutrinos apparently bounce randomly between the three generations. Conveniently, this provides evidence that there are *only* three generations. It is also implies the neutrino must have some mass: current upper bound is 0.3 eV.
- These features can be "bolted" on to the model to match experiment by tweaking and fine tuning the formulas, but a complete theory really ought to predict these features naturally.
- And then there are other things that you would like the theory to explain that it does not. For example, is there a pattern to the particle masses? How did the weak symmetry break and why did it fall out in this particular way? Why are there three generations of particles—are there more?

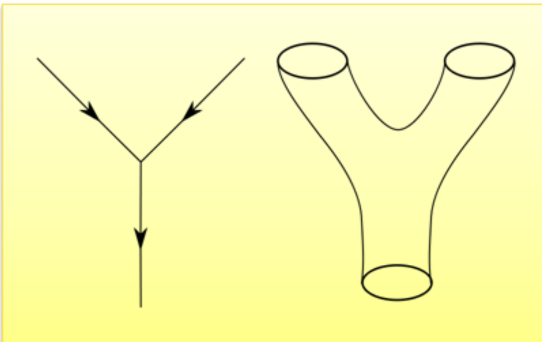
Super-Symmetry Does Not Work

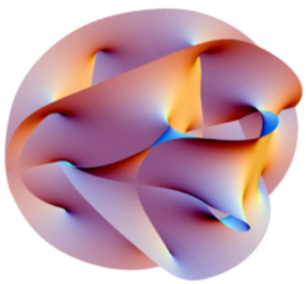
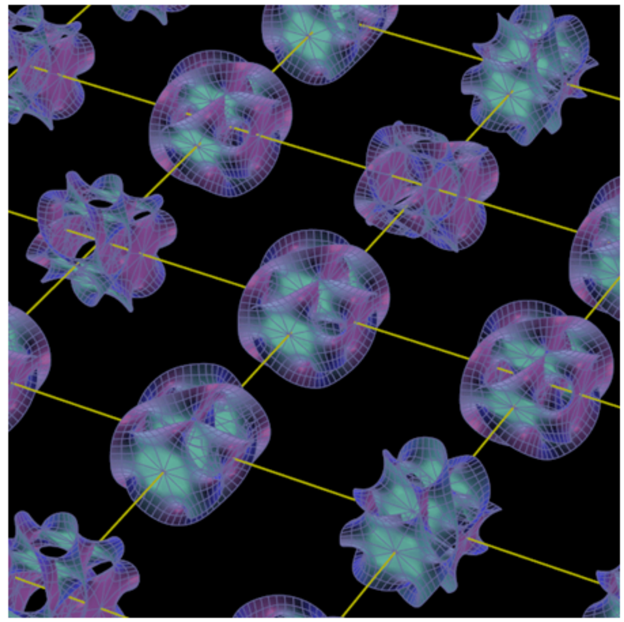




- There are a couple of other suspicious coincidences in the Standard Model:
  - Why three generations of both quarks and leptons?
  - Why is the charge of the electron and proton exactly the same?
- It is these questions that have motivated physicists to guess there may be some sort of "supersymmetry" between the leptons and quarks which unifies the strong and electroweak interactions.
- Inspired by the success of the electroweak theory and even the memory of the unification of the electric and magnetic forces these models are called **grand unified theories** (GUT). None work.
- The main problem that they all have (by design) is that they predict the decay of the proton: just as the weak interaction decays the quarks and leptons down the three generations into less massive particles, so should this "supersymmetry" decay the "u" quark into electrons and neutrinos.
- No one has ever witnessed the decay of the proton—the current lower bound on the half-life of the proton is 10^{33} years and growing every day.

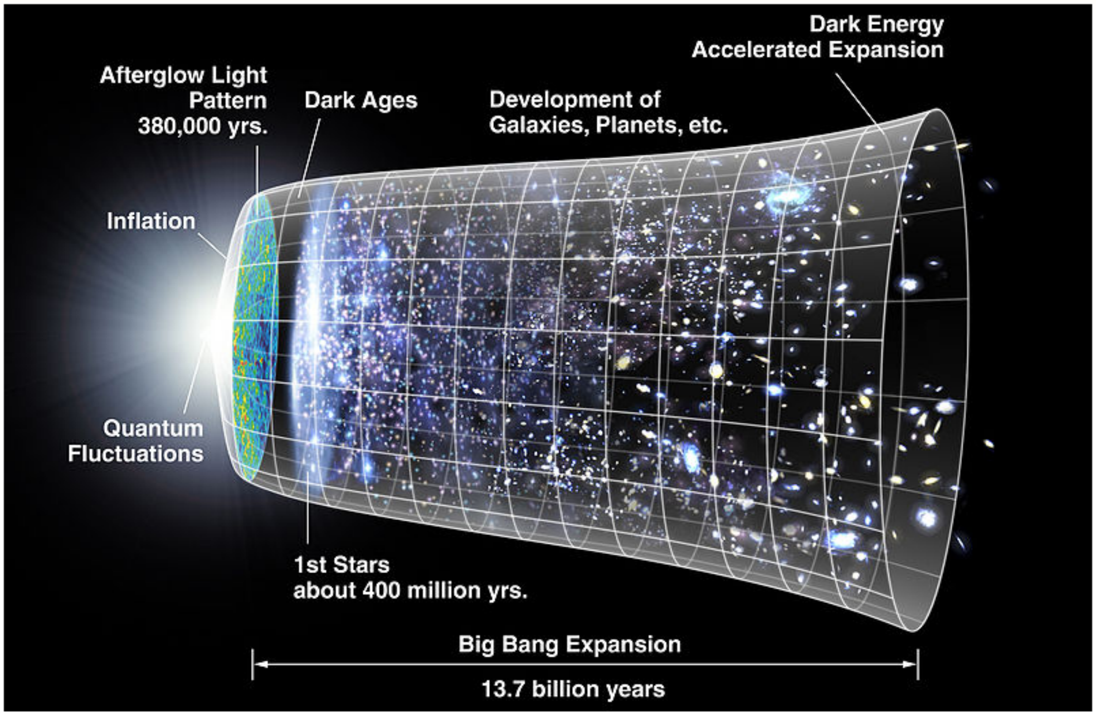
String Theory?



10D +  = 

- And then there is gravity. This is where **string theory** enters the picture.
- But before I explain these speculations let us be very clear: we are deep in the wilderness here and the forest is very dark. We are still waiting to see if we understand the electroweak unification correctly and we know that unifying the strong interaction has fundamental problems.
- The "Hail Mary" approach of string theory is a long shot indeed.
- Nonetheless, the approach has prestigious roots. Einstein spent the last decades of his life working on it. He was inspired by the quirky observation that if one expands general relativity to five dimensions, Maxwell's equations pop out.
- There were problems, of course. There are no fermions, there are unobserved extra bosons, and it doesn't say anything about the strong and weak interactions. Einstein appears to have hoped to see an explanation for quantum mechanics come out of the theory, but it never did work.
- But the "extra-dimension" seed had been planted and fifty years later string theory blossomed. Many theoretical facts about quantum strings vibrating in 10 dimensions were uncovered which yield supersymmetry with a new boson to mediate the force of gravity.
- The mathematics is obscure but in the end there appears to be five different options for a viable string theory. In the mid-1990's a single 11-dimensional version based on vibrating membranes was developed which incorporates the five as special cases.
- Currently much speculation (and controversy) is focused on this "M-theory" as the final unification of fundamental physics. It is hard to be convinced without some specific experimental evidence to guide us.

A Short History of the Universe



- Fundamental facts of cosmology
  - Hubble redshift (1929)
  - Elemental abundancies: Hydrogen 74%, Helium 24%
  - Cosmic microwave background (1964). Studied by Wilkinson Microwave Anisotropy Probe (WMAP).
  - Large scale structure: galaxies, clusters, superclusters
- Original theory offers a straight-forward explanation for these facts, but there are problems (the horizon problem, flatness problem). These can be solved if the bang is initiated by a rapid **inflation** period prior to its current rate of expansion. The mechanism for this inflation is unknown.
- Current Big Bang theory is called the Lambda-CDM model. The Lambda refers to the cosmological constant or "dark energy" while the CDM refers to "cold dark matter" required for many details of the theory.
- Big Bang 2.0 rough timeline:
  - **Inflation**. Time prior to inflation may not have meaning. Ends at 10^{-32} seconds.
  - **Matter forms**. Electroweak symmetry breaks at 10^{-12} seconds. Quark confinement at 10^{-6} seconds. Most hadrons annihilate at one second leaving leptons as bulk of mass-energy. Most leptons annihilate at 10 seconds leaving photons as the bulk of mass-energy. Big bang **nucleosynthesis** occurs between 3 and 20 minutes.
  - **Recombination** occurs at 379,000 years after the Big Bang: electrons combine with nuclei to form neutral atoms. Photons no longer scattered by charged matter—CMB released. Universe becomes transparent and the "dark ages" begin. The 21 cm line may be able to probe this dark age period.
  - **First stars form** around 100 million years. Radiation from these stars reionize the universe and the "dark ages" end. At 600 million years first galaxies form and second generation stars are born. Around 8 billion years our sun forms.
  - 13.75 billion years after the Big Bang we sit here talking about it.

Index