Puzzles and Mysteries of Physics in Everyday Life

Adventures in Quantum Mechanics

by Lajos Rimai
with the assistance of Larry Weinberg

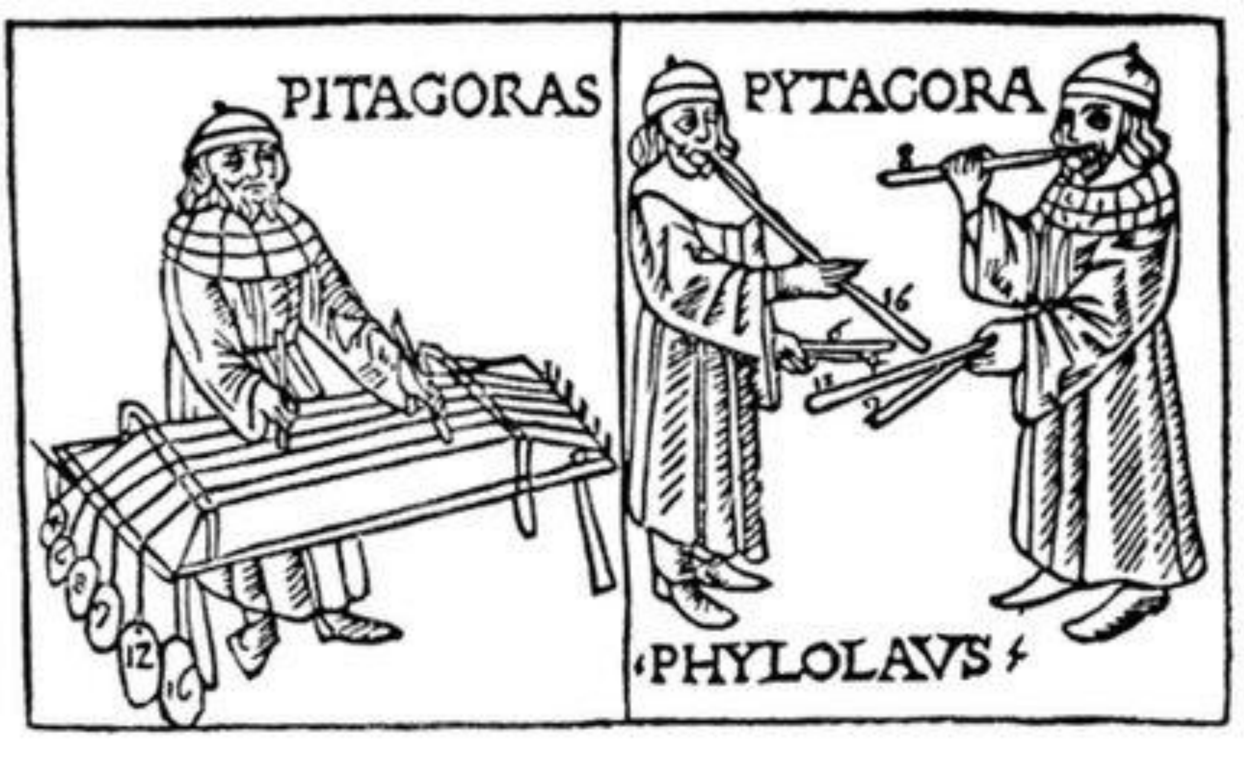


Table of Contents

Puzzles and Mysteries of Physics in Everyday Life 1

Prologue. Why write it? 6

Section 0. Physics of the Cave Dwellers, an introduction. 8
Figure 1. An Inspiration, The Atom, Democritus' insight. 10

Section 1. The contradictory properties of light and radiation 13
Figure 2. Rutherford scattering. 19

Section 2. Waves 21
Figure 3. Color and wavelength 21
Figure 4. Circular motion and a wave 22
Figure 5. A Linearly Polarized Wave Moves as a Rigid Fence. 23
Figure 6. A circularly polarized light wave moves as a right or left-handed screw. 24
Figure 7. Radio Waves 25
Figure 8. The Thomas Young experiment, with photons. 26
Figure 9. Light paths in the double slit experiment. 27
Figure 10. Interfering waves from a common source. 28
Figure 11. Double Slit Experiment schematic illustration and plot of the strengths of the recorded signal. 29
Figure 12(A). Interference fringes as recorded on a photosensitive film. 31
B 32
Figure 13. (B) show interference fringes as recorded by a photon counting array 32
Figure 14. From photons to waves 35

Section 3. Particles and Waves. 38
Figure 15. A multiple slit interference experiment 38
Figure 16. Pythagoras’s Harp and a modern harp. 40
Figure 17. The superposition of harmonics. 44
Figure 18. High resolution spectrum of sunlight showing the Fraunhofer absorption lines of hydrogen and some other components of the solar atmosphere. 45
Figure 19. Comparison of absorption and emission spectra of Hydrogen 46
Figure 20. The emission spectrum of Hydrogen (top) and emission spectrum of Helium (bottom). 47

Section 4. Atoms: The Inside and Outside 50
Figure 21. Illustration of cyclic boundary conditions. 50
Figure 22. The photoelectric effect. 52
Equation 1. Time dependent and time independent Schrodinger equation for the wave function of an electron in a one-dimensional space. 55
Figure 23. This is an attempt to picture the orbitals for the various states of the Hydrogen atom. 59

Figure 24. The energy levels of the hydrogen atom. 60
Figure 25. Comparison of the highest Energy levels of the Helium atom (right) and the Hydrogen atom. Here we are comparing escape levels. 61
Figure 26. Escape energy, as a function of nuclear charge, atomic number, or number of electrons. 63
Figure 27. The Shell Model. 64
Figure 28. The Mendeleev periodic table. 65
Figure 29. The Lithium energy level scheme. 67

Section 5. Bonds, Molecules 68
Figure 30. Schematics of the formation of a covalent bond, The hydrogen molecule. 69
Figure 31. The bonding in the water molecule. 70
Figure 32. The phosphate ion. 72
Figure 33. atoms and phosphate ions assemble to build a phospholipid membrane mostly by covalent bonding. 73
Figure 34. How ions, metabolites, energy (ATP) enter the cell across its membrane. 74
Figure 35. Illustration of the cell membrane from outside(left) and from inside(right). 75

Section 6. Metals, semiconductors and band structure. 87
Figure 36. Conduction electrons in a metal wire without-applied-voltage (top), and with-applied-voltage(bottom) 90
Figure 37. illustrates the positions of the 2 outermost bands in metals, semiconductors and insulators. 95
Figure 38. The semiconductor band gap with thermally excited electrons and holes. 97

Section 7. The p-n junction and semiconductor devices 99
Figure 39. A p-n junction. 99
Figure 40. Current rectification in a p-n junction, easy current direction, electrons move from cathode to anode, holes move in the opposite direction. 101
Figure 41. schematics of PNP and NPN transistors 102
Figure 42. Binary logic gates implemented with transistors. 104

Section 8. Light Emitting Diodes, Diode Lasers and lasers in general 105
Figure 43. The LED mechanisms 105
Figure 44. The general energy level scheme enabling laser action is on the left. On the right is a slightly more complicated scheme. See if you can figure it out before you read the next paragraph. 108
Figure 45. The structure of a laser diode that enhances the stimulated emission by multiple passes inside a resonant structure. 112

Section 9. More about Magnetism, Magnetic Moments and Nuclear Spins 114
Figure 46. The magnetic field lines of a circular loop of electric current showing aligned magnetic needles. 116
Figure 47. Electric field lines for a positive test charge generated by a charge dipole. 117

Figure 48. The magnetic field of a bar magnet. 117

Section 10. Nuclear spins and Nuclear Moments 123
Figure 49. Setup of an NMR experiment 124
Figure 50. The torque on a current loop (Feynman Volume2),@ as predicted by diamagnetism is illustrated for a rectangular loop. 125
Force $F = qv X B$, Torque $\tau = r X F$ 126
Figure 51. Comparison of magnetic dipole moment orientation between bar magnets, current loops, electron spins, nuclear spins. 127
Figure 52. NMR Absorption frequency vs Magnetic field. 128
Figure 53. Diagram of the NMR setup used to record the NMR spectrum by Purcell, Torrey and Pound. 129

Section 11. The MRI magnetic field, DC and RF 137
Figure 54. The various coils positioned around the “sample”. 137
Figure 55. Small longitudinal gradient superposed to the large field, for slice identification, the essence of tomography 138
Figure 56. If a field gradient is applied along a body diagonal and then it is rotated its component in a transverse plane will rotate in this plane with constant magnitude. It becomes a rotating field gradient. 139
Figure 57. Pulsed NMR Experiment 140
Figure 58. Reproduced here to illustrate Fourier analysis. How it applies to pulsed NMR data analysis. The top graph displays the Fourier components of the decay that is displayed in the bottom graph. 141

Section 12. NMR in Chemistry. Exploring molecular structure. MRI mapping of chemical species in the human body 142
Figure 59. The proton NMR spectrum of ethanol. Six Protons in 3 different sites within the molecule. 142
Figures 60(A) and (B). The ethanol molecule in stick imitations. 143
Figure 61. NMR of protons in water obtained after averaging thousands of noisy single pulse spectra 145
Figure 62. Single Pulse Spectrum of protons inside a red blood cell. 146

Section 13. The MRI Scans 147
Figure 63. The 3 red spots form a right triangle. 147
Figure 64. The MRI Trick 148

Section 14. Use of radioactive labeling in medicine, and the Physics of PET scans 151
Figure 65. The sensing signal generation process in a PET scan. 153
Figure 66. The Compton scattering effect, the major source of inaccuracy in the location determination from PET scans 156
Figure 67. Schematics of the positron annihilation process detected by a ring of scintillator detectors. 157

Section 15: Conclusions 157

A due acknowledgement 160

Appendix A 161
  Figure A1. Temperature dependence of the resistance of mercury. 162

Appendix B 168
  Equation 1. Time dependent Schrodinger equation. 175
  5. Diamagnetism, Para magnetism and Ferromagnetism 179

Epilogue 181

Prologue. Why write it?

The motivation of the human species to pursue scientific research is, in a very basic way, curiosity. Questions form. The possibility of useful applications follows after some of the answers are found. This sequence of two steps is present throughout the history of science. Mystery and curiosity are mother and child but in the continuous chain of evolutionary events it is hard to figure which proceeded and which followed. The ideas in this document were part of a presentation to a group of retirees at a senior citizens campus where I also reside. Interesting discussion ensued, and regular science discussion meetings followed.

The first hot topic was the mystery of black holes as the news was reporting on the observation of the center of our galaxy where a black hole was lurking. A young astrophysics graduate student had just entertained us with a discussion of mysterious phenomena at the remote corners of the universe, such as cosmic rays and black holes. This was the seed of the idea of exploring the mysteries at the opposite corner of this universe of ours, where microscopes rather than telescopes are the instruments.

Therein lies the source of this document, a rambling PowerPoint presentation. If such a discussion was interesting to a bunch of seniors it might be of interest to other groups, in particular science educators working with preprofessional students. It might spark curiosity and interest in science. In particular, it would be an advertisement of physics, my field of interest. So, there you have it, the motivation.

For additional supporting information, there are two appendices following the main document:

- Appendix A discusses relevant technologies supporting the applications of quantum mechanics in medicine.

- Appendix B provides some simple explanations to some of the more complex topics presented.

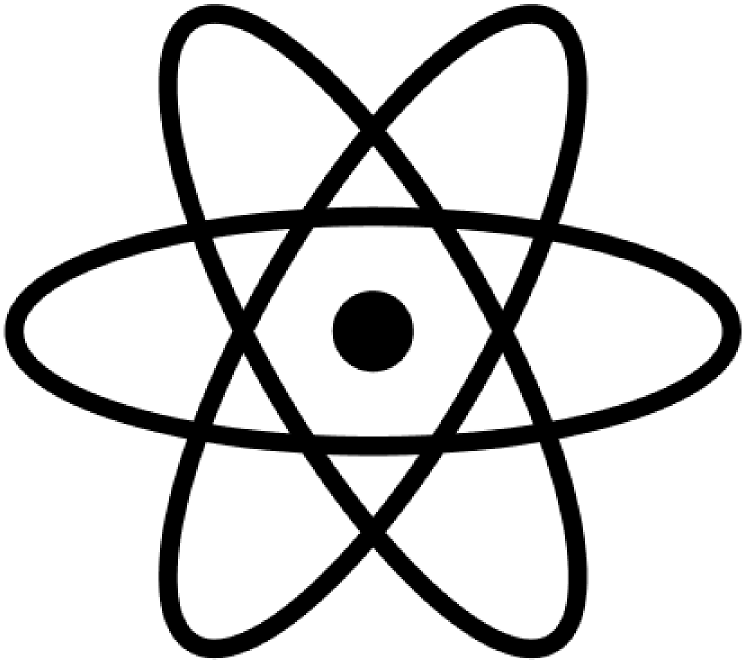
We hope the information we present here creates connections in your knowledge of how Quantum Mechanics underlies our world.

Section 0. Physics of the Cave Dwellers, an introduction.

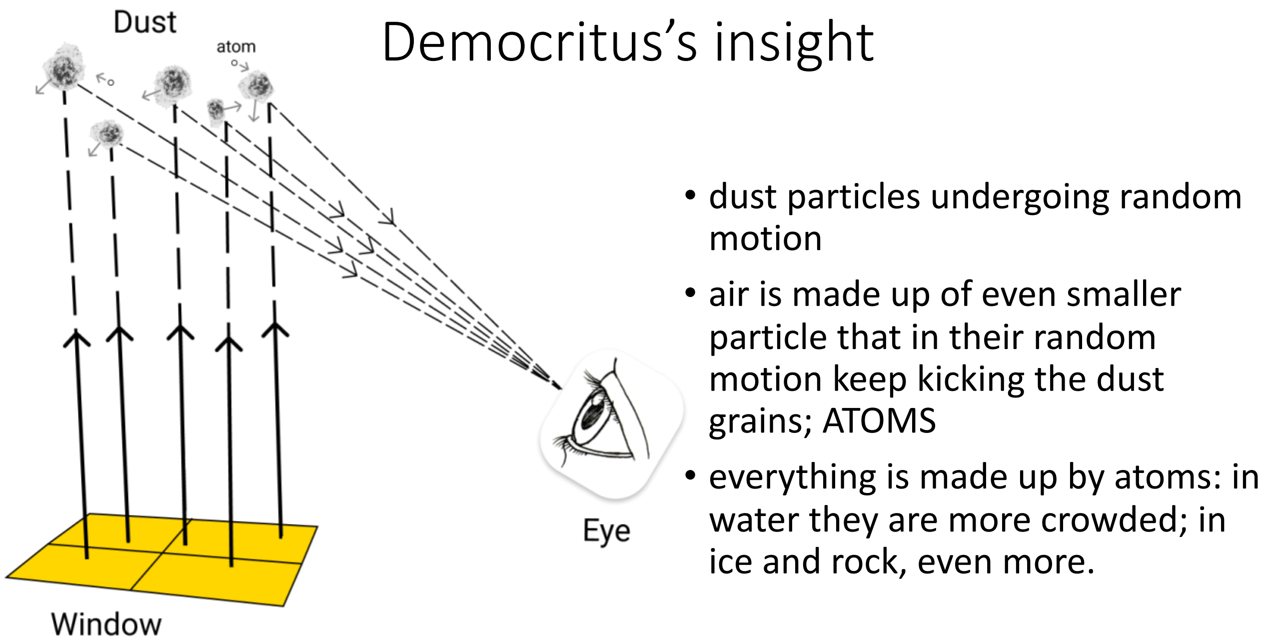
People often believe that science should be able to explain all puzzling and mysterious events. In truth there is no such guarantee. We have done well to find rules and laws that seem to explain most of our measurements, but anything we know today could be proven insufficient with the next observation. Such discoveries prompted Newton to propose his laws of motion and gravitation, Maxwell his equations of electricity, Einstein the theory of special relativity, and, at the dawn of the 20th century, Bohr and Schrodinger needed quantum mechanics to explain what was being revealed in the microscopic world of atoms and molecules.

This process of discovery and modeling will continue. The human species might disappear without having all the answers, but as curiosity survives, new explanations will be needed. We now know that Quantum Mechanics is with us all the time and everywhere. There is no hiding place. In what follows, we try to ease its acceptance and raise your awareness about its impact on your daily life.

Other writings discuss the rules of nature governing phenomena at large scales -- events happening over large distances in space, or long intervals in time. In this work we describe the phenomena at the opposite limits of scale. The rules of quantum mechanics, which operate at the level of molecules and atomic particles, can be more puzzling than those of stars and planets.



In quantum mechanics we have a model for how we think these things work. This model may never be intuitive, but it has proven itself quite useful for improving everyday life. Physics has expanded the range of our senses with Doppler spectroscopy towards the reachable limits of the universe, with optical microscopes towards the too-small-to-distinguish, like cells and bacteria, with X ray and electron microscopes towards molecules, atoms, electrons, nuclei and their parts. If we had not made the observations about Quantum Mechanics that we have, you would not be able to watch your favorite Netflix show on your mobile phone today!



Democritus’s insight

- dust particles undergoing random motion
- air is made up of even smaller particle that in their random motion keep kicking the dust grains; ATOMS
- everything is made up by atoms: in water they are more crowded; in ice and rock, even more.

Figure 1. An Inspiration, The Atom, Democritus' insight.[5]

When we focus on the very small, mysteries appear. So, let us just do that.

If an early cave dweller were to find a modern-day cellular telephone in her time, she would not get much use from it. The cellular services were particularly poor in those days! Today, any citizen can communicate almost instantly with any other citizen anywhere in the world. The cave dwellers couldn’t begin to imagine the laws of nature that could lead to such power – but by investigating their own early mysteries they laid the foundations of physics. Fire and light were there for them to discover and challenge their imagination.

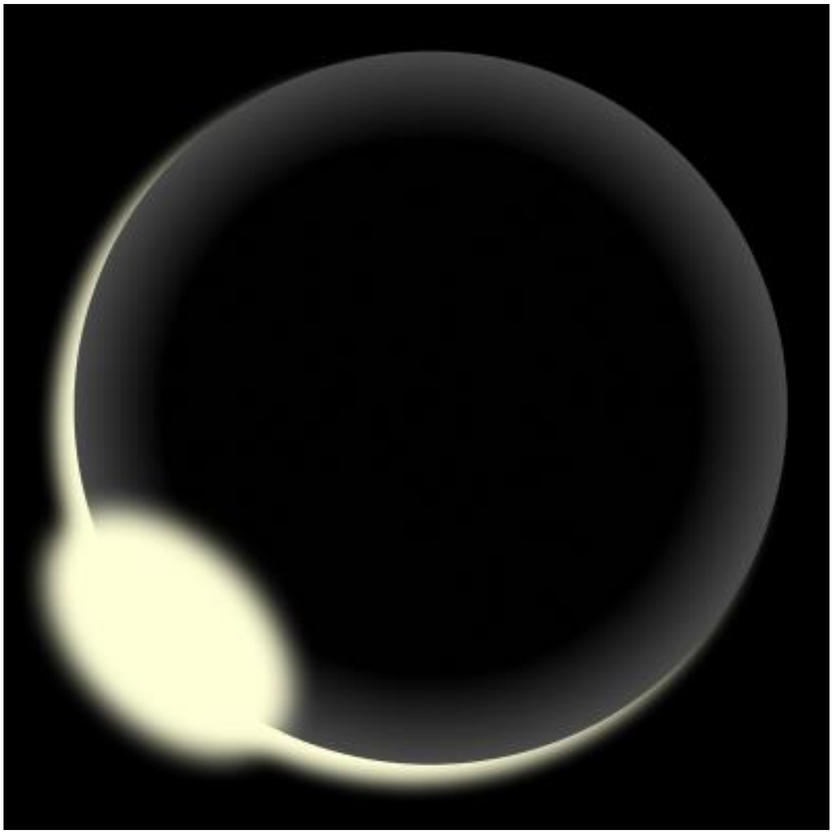
The very bright object in the sky created the daylight and the warmth. With fire came both light and warmth. With light they could walk easily

---

[5] See Democritus in Wikipedia

past each other. Their sensors, their eyes, allowed for this. The fire allowed for this ability at night – even inside the cave.

Putting these things together meant something. Why and how are warmth and light produced? The first double mystery of physics was presenting itself and asking to be solved. Fire seemingly produced two emanations: warmth, and light. Realizing this, that there were two emanations that were related but different, became one of the first fundamental discoveries of science. There were many more dots to connect in the years to follow. Some very tiny dots.



Before science could progress, language grew from the need to store and exchange information. Language had its own effect on thinking, which increased the need for memory. Over time this need encouraged evolutionary directions which improved our abilities to use the input from our sensors, store observations, share with others, and grow, over time, a collective intelligence.

Add to this progression of intelligent life, an ability for imagination -- the ability to develop a memory without specific sensory input – using other memories as new inputs – thus opening the door to scientific discovery. Out of this comes theory. Theory gives us something to try, to use, to accept, or reject. In theology, in philosophy, in physics, theories are put to the test for us to improve our knowledge and find our next creative inspiration to solve real-world problems.

The machinery of science consists of performing experiments repeatedly. And then repeating them. Then repeating them again. And a few more times. Then scientists use imagination to consider underlying mechanisms that could explain how the results were possible – and how they could be predicted – aided by mathematical formulas.

With enough experimental investigation, a theoretical model can be chosen. This becomes the working theory. The working theory can then suggest possibilities to consider by thinking about past known results and extrapolating into the unknown. This leads to further experimentation. Further experimentation leads to verification, yet more experimentation, a pinch of modification, and then you are microwaving on high for 6 minutes. Let it sit for one minute before serving.

Section 1. The contradictory properties of light and radiation

Our collective learning **--** our ability to remember what others have discovered before -- empowers us to have insights and moments of intuition. These sparks of theoretical enlightenment can suddenly allow us to make sense of observations that never made sense before -- and predict new experiments for us to attempt. For instance, we can now reconcile these seemingly contradictory properties of light:

- Light travels on straight lines.

- Light travels in all directions simultaneously.

- Light propagation implies transport of energy but *apparently not of mass.*

We know, now, that this third observation is not exactly true**.**
Energy density does exert gravitational attraction, so it is associated with a gravitational mass density given by **E=mc2,** or **m=E/c2**, the now famous Einstein equation from his ***Theory of Special Relativity***. The equally famous solar eclipse observations of **1918** give us experimental evidence. However. the details of the data do more. They point towards the general theory he later published.

The discovery of the particulate nature of light, i.e. the photon, eases the intuitive acceptance of these ideas. The rest mass of a photon is zero because it disappears when stopped. Its kinetic energy is absorbed by the body that stopped it. If the body does not move, its gravitational mass, the rest mass, increases. Momentum conservation will determine its motion. Therefore, propagation of light must involve transport of some gravitational mass other than the rest mass, the so-called rest mass, representing the internal energy of a moving body.

By the 4th century BCE, the Mayan, Egyptian and Greek astrologers followed the motion of the planets, the wonderers in the sky, as they became illuminated by the sun during the nights. About 20 centuries later Tycho Brahe[6] did the same with a scientific instrument and made a map of their paths. His apprentice, Johannes Kepler, converted these maps into data and algebraic formulas which, in the next century Isaac Newton interpreted as his laws of motion and invented calculus in the process.

Another Greek natural philosopher known as Democritus followed the motion of dust particles in the air as they became illuminated by the light of the sun (see Figure 1). He noticed that, although they would drift towards the floor slowly, they did so in a very irregular manner -- in very short and rapid displacements in all directions. That observation sparked the insight that air might be made up from a very large number of extremely small invisible particles he called Atoms. Imagining that these particles could be moving randomly they might kick the dust particles also in a random manner. This deduction might be the earliest use of observing phenomena in the visibly small domain of nature to inform of phenomena in the invisibly small domain.

This process of sensing the invisibly small by observing the visibly small has since been part and parcel of scientific activity. It is illustrated by the

---

[6] A Danish nobleman, Tycho Brahe (1546-1601), made important contributions by devising the most precise instruments available before the invention of the telescope for observing the planets

following more recent, yet classic, experiments:

- By following the trajectories of pollen particles suspended in water, as they perform Brownian motion, Einstein deduced the number of molecules present in a unit volume of water.
- By following the trajectories of microscopic drops of oil in a vacuum under gravity and a known electric field, Millikan was able to deduce the mass of a single electron.

Through the 17th, 18th, and 19th centuries chemistry matured as a science. Atoms became recognized as the elementary particles of pure elements like hydrogen, oxygen, copper and carbon. Pure elements could be combined to form more complex substances such as water, carbon dioxide, or salt, whose elementary particles became the molecules, built by combining atoms. A pure element by itself could display different identities by forming different multiatomic molecules, as exemplified by oxygen, which we breathe, and ozone that poisons us. Likewise, carbon can form graphite that we burn in coal, or diamond that we can’t afford. Similar associative transformations occur with agglomerated molecules giving birth to the infinite variety of materials found in nature as liquids and solids.

At the sunset of the 19th century, however, it was discovered that atoms were not the ultimate elementary particles. This discovery, or more fittingly, set of discoveries, marked the start of a revolution still happening today. Atoms are assembled from smaller, more elementary parts, which in turn, are made up of even more elementary parts and so on, perhaps. How far this chain of events might progress, we don’t know.

What we do know so far is that atoms are assembled from three kinds of more elementary particles, positively charged protons, negatively charged electrons and uncharged neutrons. The fact that oppositely charged particles attract each other to help assemble the atom, while the repulsion of same sign particles oppose assembly, means that the assembly process involves energy exchanges in a manner such that an assembled atom has potential energy that is released on disassembly.

This is the rest energy of the atom. If the atom moves it acquires kinetic energy. So, the net energy will depend on the reference system. By the equivalence rule of special relativity $\mathbf{E=mc^2}$, so does its mass. The rest mass represents only its internal energy. The kinetic energy also contributes to the gravitational mass. As the total mass is the gravitational mass this makes a connection of special relativity with the general theory of gravity, which is the essence of what the general theory became, the theory of gravity.

Democritus implanted into the human mind the concept that as you break up matter into progressively smaller parts a limit size will be reached (Figure 1). He called it the atom. This idea survives, but only in an adjusted form. The statements of the laws of nature must be reviewed and kept in tune with the “facts”, that is, the latest experimental findings. The idea that at very small distances the world must be described differently was born in that place and at that time.

Until the late 19th century, atoms were THE elementary particles. An experiment performed by Ernst Rutherford at the dawn of the 20th century demonstrated that atoms are not the ultimate elementary particles.

It turned out that an atom is made up of an extremely massive but minuscule nucleus, surrounded by a cloud of extremely light electrons.

At that point in time the experiment performed by Ernst Rutherford definitively confirmed the existence of particles more elementary than the atoms (Figure 2.)

This experiment, performed on a small piece of gold foil, opened up a world of progressively smaller and smaller elementary particles with no end in sight.

His findings were preceded, and specifically enabled, by the discovery of radioactivity and the characterization of the **α, β,** and **γ** radiations from polonium. The Greek letters are traditionally used to label them, so we adopt this notation in this document. In all that follows we will use Greek letters to label variables and constants whenever this is the common usage. The positive **α** radiation consists of two protons and two neutrons, the equivalent of the helium-4 nucleus, whereas **β** rays consist of the much lighter negative electrons. The **γ** rays are neutral high energy photons.

Sources of such energetic radiation had recently (for that time) been uncovered from more conventional sources such as electrical discharges in cathode ray tubes. These discharges sometimes produced beams of radiation (X-rays), other times beams of charged particles. The charged particles could be characterized by electromagnetic determination of their algebraic charge-to-mass-ratio, the X-rays by their wave properties. Higher-energy, uncharged radiation was at that time, detected and characterized by the calorimetric effect as they destroyed targets, but remained a mystery until quantum mechanics. However, Rutherford used **α** particles from Polonium as they were suspected to be more massive than the others.

A sketchy description of the experiment is as follows:

1. The more massive **α** radiation from a polonium source is electromagnetically isolated and pointed at the target.

The target is a very thin film of gold.

2. The **α** particle source is aimed at the gold film in varying directions.

3. The largest fraction of the **α** particles in the beam pass through the film un-deflected but a measurable fraction gets scattered in all directions, some even directly back towards the source.

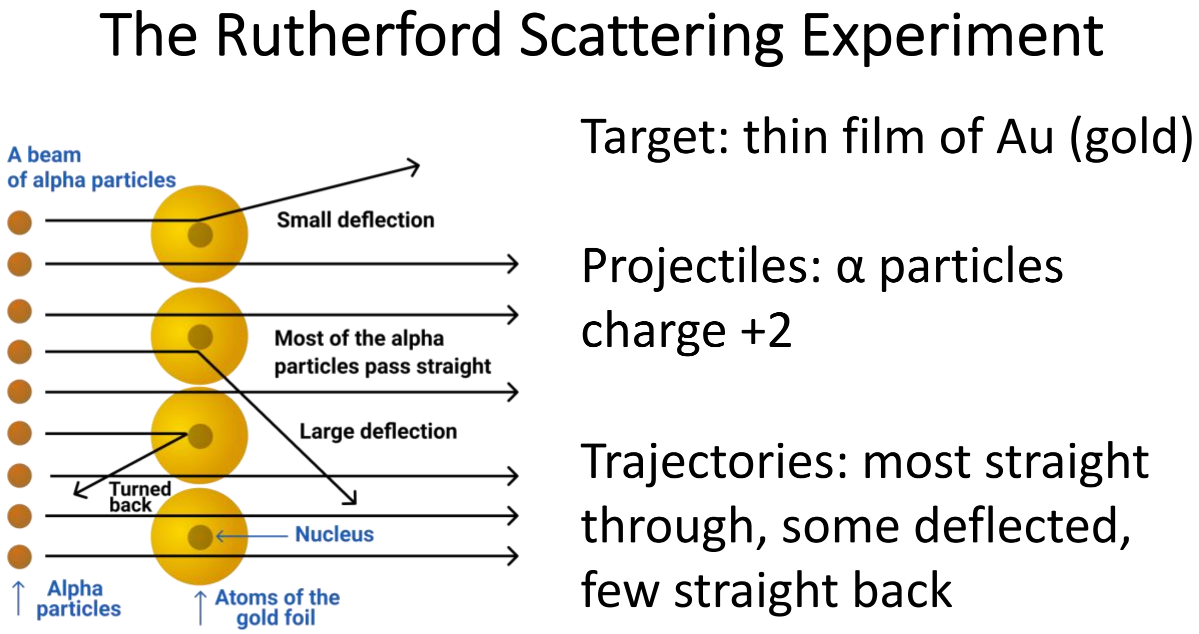
4. A fraction of radiation transmitted seems to pass through a transparent region of the film. The rest seems to hit an opaque region.

5. From the measurement of the ratio of deflected-to-transmitted particles as it varies with the direction angle it can be deduced that the total area of the opaque region is much smaller than that of the transparent region. This leads to conclusion that the size of the opaque region must be very small. The experiment when repeated in different areas of the same film, or in different films, has the same results.

The most plausible explanation was that the gold film consisted of a distribution of massive but very small particles interspaced with an equal number of very light but oppositely charged particles. This experiment and the picture of matter that it implied, are at the root of all further progress in the theory and applications of the natural sciences.

So, very tiny but very massive particles, positively charged nuclei, are distributed uniformly in a sea of very light negative particles, the electrons. Detailed calculations due to Rutherford yielded, for gold, the estimate of internuclear distance of 10 nm=10^{-8} m. This agreed with the charge to mass ratio of the nuclei obtained from beam deflection experiments.

Such experiments, at the atomic level, have become more or less routine since the early 20th century when Ernst Rutherford first performed it.



The Rutherford Scattering Experiment

Target: thin film of Au (gold)

Projectiles: α particles charge +2

Trajectories: most straight through, some deflected, few straight back

http://physicsopenlab.org/2017/04/11/the-rutherford-geiger-marsden-experiment/

Figure 2. Rutherford scattering.

Isaac Newton proposed that a light beam is made up of moving corpuscles that have a velocity dependent on the medium in which they are moving.

He was foretelling photons perhaps. Color is a property of the corpuscles of light, and white light is a superposition of corpuscles of many colors.

In vacuum or thin air all corpuscles move at the same speed. In dense matter the speed depends on the color. In this manner Newton explains reflection, refraction, the prism and the rainbow.

If the matter is too dense it can trap the corpuscles, thus absorbing all colors. Furthermore, his model satisfies Fermat’s principle that light, as it propagates from one point to another, will follow the shortest path. This means that in a homogeneous medium it propagates in a straight line but when it hits a flat boundary between two different media it follows Snell’s laws. These laws apply in the macroscopic world, but they can be justified even in the molecular world by classical models of the interaction of molecules with electromagnetic waves, if we stick to refraction and reflection. More recently discovered phenomena forced quantum mechanics into the picture.

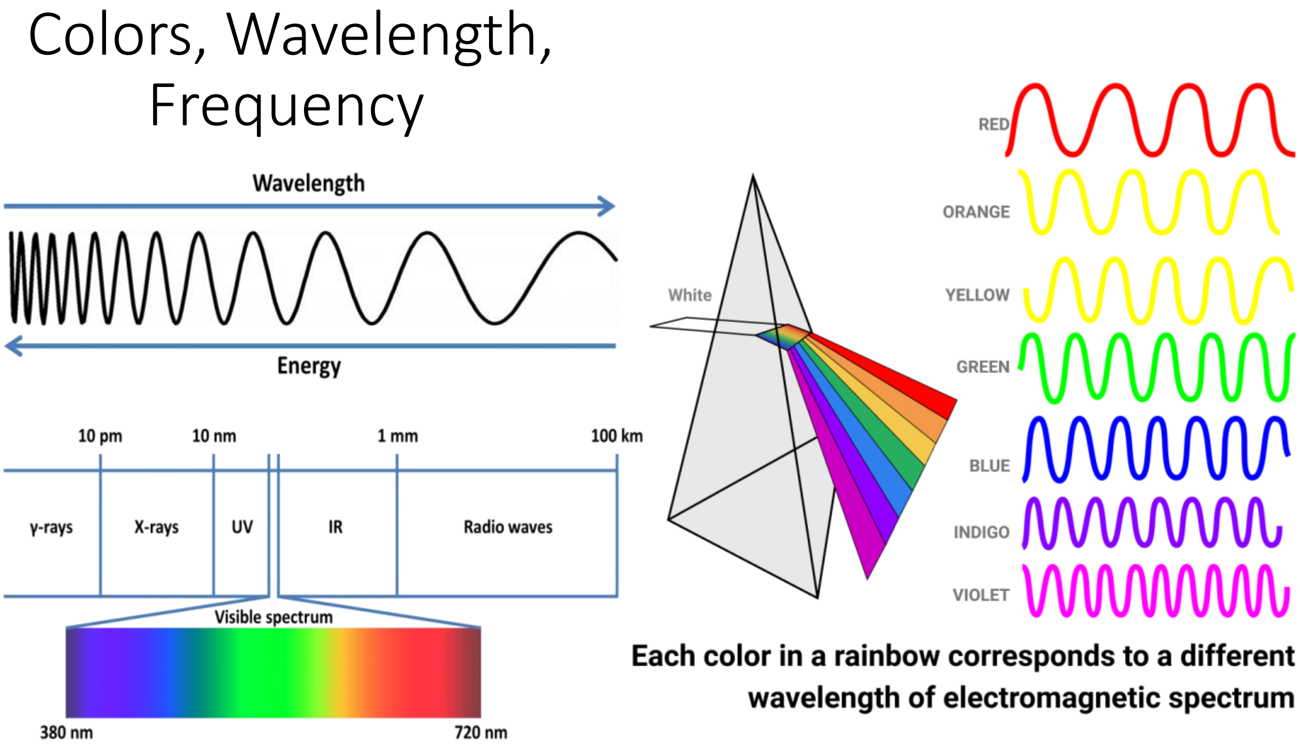


Figure 3. Color and wavelength.

Electromagnetic waves are characterized by their frequency **ν,** and their propagation velocity **c** which relates wavelength **λ** to frequency **$\nu = c/\lambda$.** You will see these symbols often in this document so it's good to get used to them.

A uniformly rotating charge generates a circularly polarized wave propagating in the direction perpendicular to the plane of rotation. However, along either in-plane coordinate axes, it generates a linearly polarized wave.* A circularly polarized light wave propagates as a rigid

---

* on Google Understanding circular and linear polarized light
https://en.wikipedia.org/wiki/Circular_polarization

rotating screw, right-handed or left-handed. The handedness defines a relationship between a rotation and a translation by establishing their connection.

Uniformly Rotating Charge Generates a Circularly Polarized Wave
Viewed on Edge it Generates a Linearly Polarized Wave

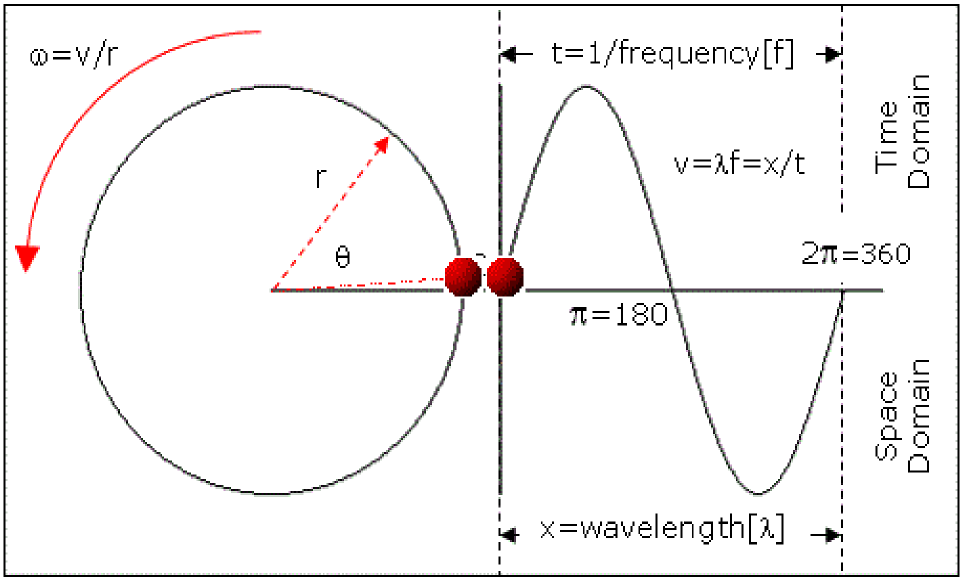


Figure 4. Circular motion and a wave.

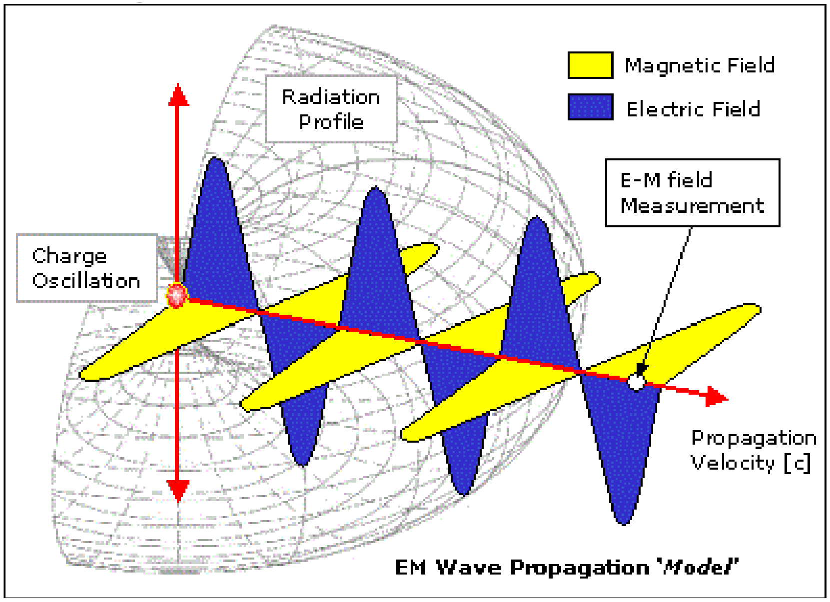
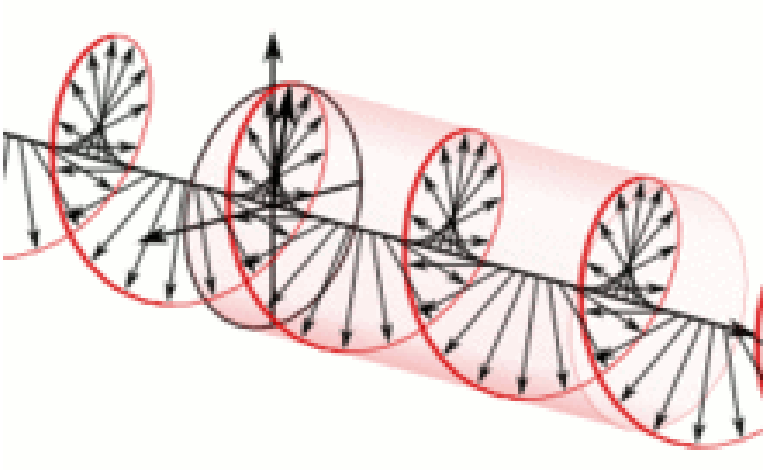


Figure 5. A Linearly Polarized Wave Moves as a Rigid Fence.

Viewed from a fixed location it vibrates at a frequency inversely proportional to its wavelength.

A light Wave Propagates as a Rigid Screw. Right Handed Circularly Polarized Wave



Wavelength=spacing λ
Between neighbor turns

Frequency=number of turns
Moving by per second ν

νλ=c speed of wave

Figure 6. A circularly polarized light wave moves as a right or left-handed screw.

To emphasize the connection between circular motion and wave motion, as you contemplate the figures, the generated circularly polarized wave propagates along the normal to the plane of the circle. Two linearly polarized waves generated by the oscillatory components of the circular motion will propagate along the coordinate axes. The wave polarized along x propagates along y, whereas the wave polarized along y propagates along x.

A Radio Wave Propagates
This Represents Reception
Replace R by Oscillator=Transmission

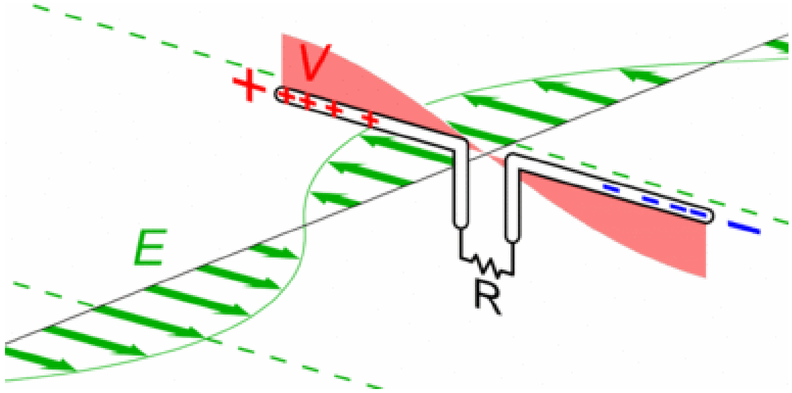
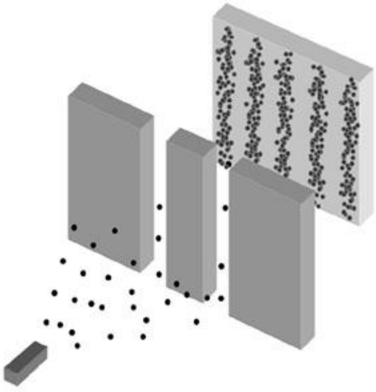


Figure 7. Radio Waves.

A linearly polarized radio wave generated by an oscillating current through the resistance **R**, feeding a dipole antenna, can also be received by such an antenna where the incoming wave generates the same kind of current. Just replace the resistance **R** by a radio receiver. Radio waves propagate from antenna[&] to antenna. The product of frequency and wavelength is the speed of propagation $\nu\lambda=c$, c being the speed of propagation in vacuum, of all electromagnetic waves.

The wave nature of light was suspected by scientists through the 18th century but positively confirmed by the double slit interference experiment of Thomas Young in 1801. The basic interference demonstration setup is described next, in modern day language, mixing wave and particle (photons) aspects of light. As we learned more recently,

[&] Lookup Dipole antenna in Wikipedia, good article electrons have wave properties as well, that are reflected in analogous experiments.



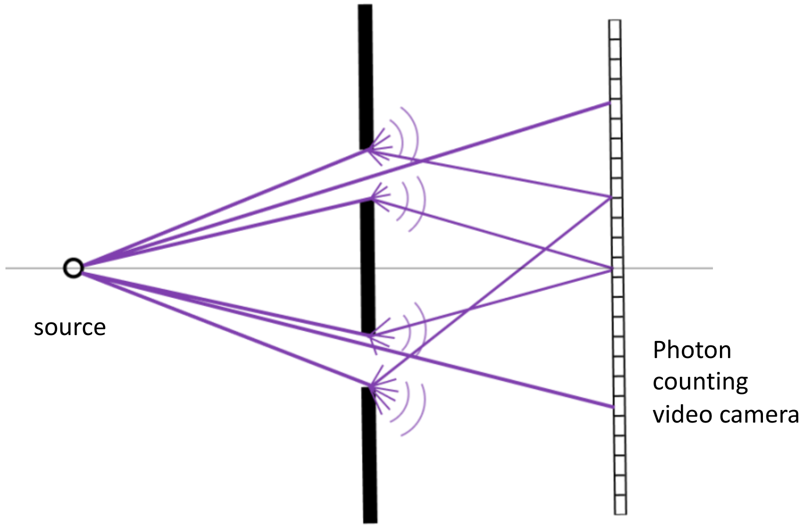
This Photo by Unknown Author is licensed under CC BY-SA

Figure 8. The Thomas Young experiment, with photons.

A small point-like light source sprays photons in many directions, in the form of a spherical wave. Ahead we interpose a flat screen with two parallel vertical slits. Some of the photons move right through the openings. These photons do not contribute to interference, only to a background signal. Those that hit the edges will be scattered in many directions, in the form now of a “cylindrical wave”.

Sufficiently narrow slits, only scatter radiation from the edges, as a pair of expanding cylindrical waves. The wave that started at the common point source is split into two waves as if emanating from the slits. That is the case for sufficiently narrow slits so that only a minimal fraction of the light can flow directly through. A light sensitive detector plane is used to sense for the light distribution beyond the slits. It could be a photographic plate, a video camera, a scanning photon counting detector, or nowadays, a photon counting detector array.

Double Slit Interference with Photons
Corpuscles
“Explaining” Spherical Wave Fronts



Source: photons in many directions.
Some move right through the openings. Those that hit the edges get scattered in many directions; “spherical waves”.
For narrow slits approximate the scattered radiation from both edges as a single spherically expanding wave

Figure 9. Light paths in the double slit experiment.@

Interference effects will then show up in the image on the detecting plane as vertical illuminated streaks, parallel to the slits. (Figure 9) Two light beams originating simultaneously from the same source point but scattered by the edges of both slits can superpose on the detecting plane thereby displaying the interference effects. The difference in the path lengths as it relates to the wavelength determines the strength of the registered signal. Superposition “in phase” doubles the strength. In opposition cancels it to zero. For a given slit separation the measured geometry of this interference pattern will yield the value of the wavelength of the light.

@ https://opentextbc.ca/physicstestbook2/chapter/youngs-double-slit-experiment/

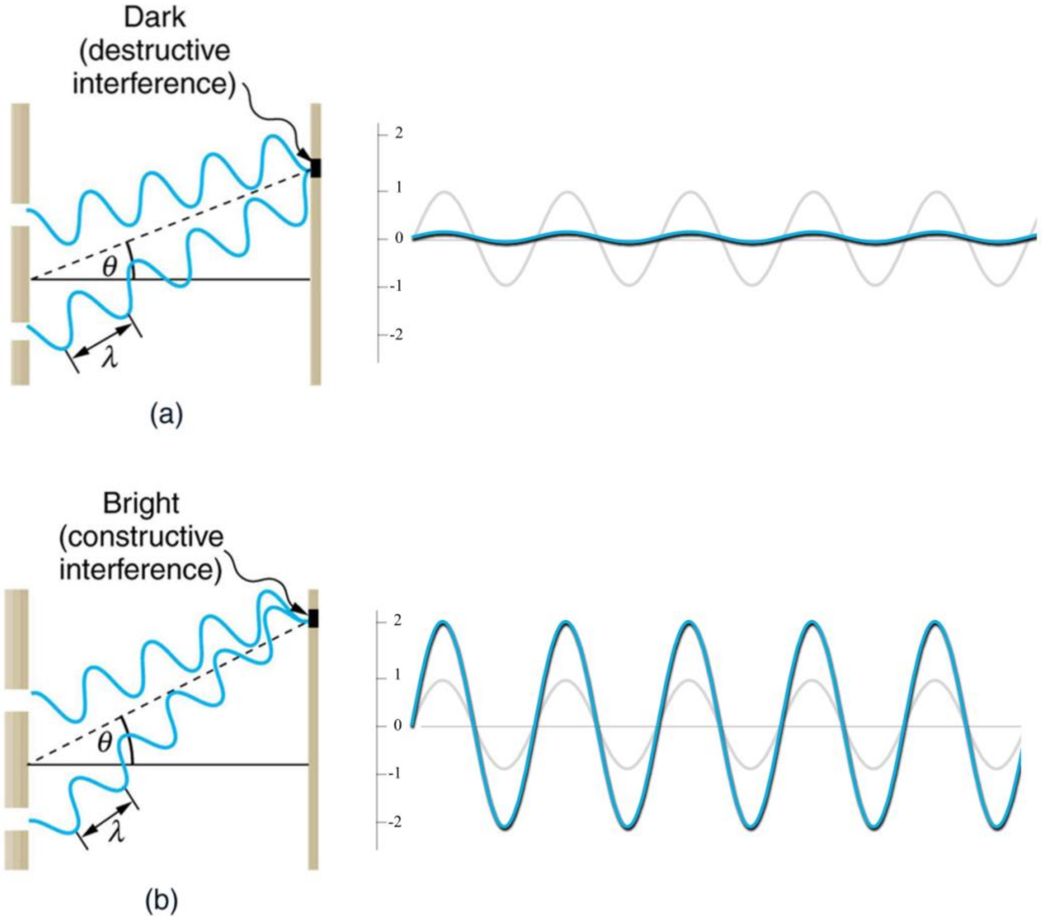


Figure 10. Interfering waves from a common source.

Bottom is in phase.
Top is almost in opposition.

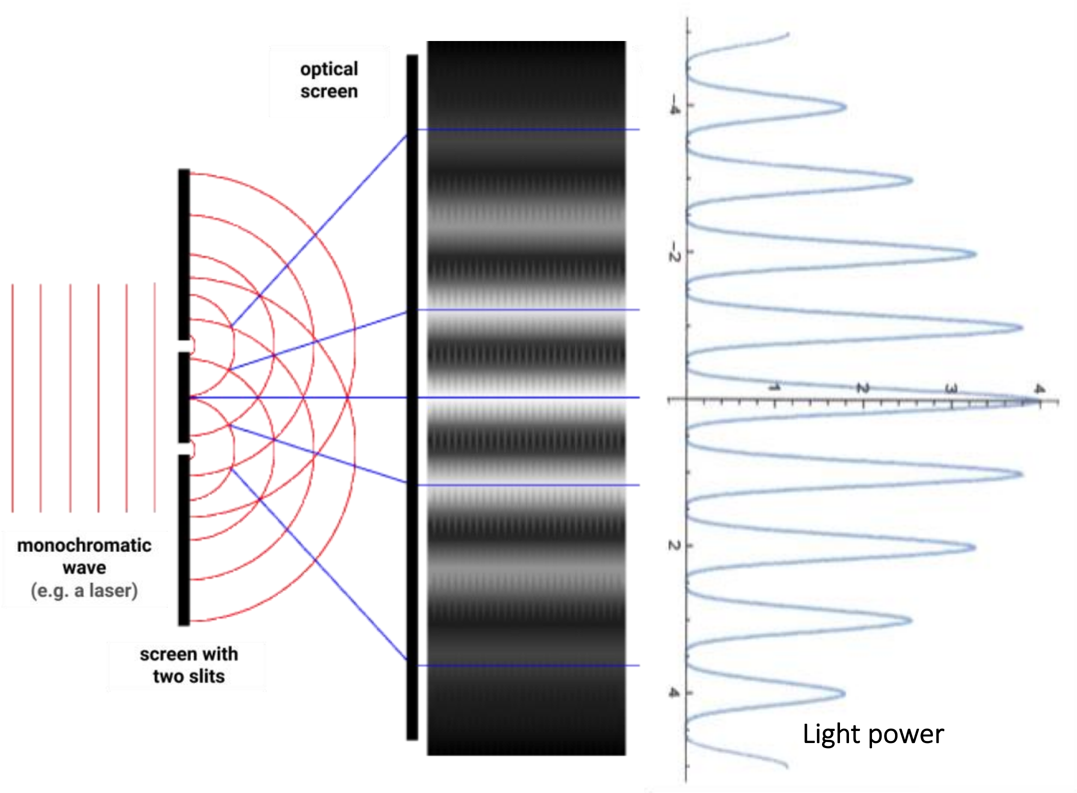


Figure 11. Double Slit Experiment schematic illustration and plot of the strengths of the recorded signal.

An almost-plane wave is split into two cylindrical waves emanating from the slits. On the detection plane the interference fringes appear. The plot on the right shows the corresponding power density distribution.

In order to obtain an exactly parallel beam via lenses and mirrors, no matter how perfect they are, one must start with an exact point source with no spatial extension. That is the reason for the qualifier “almost”. A laser source might come close. This issue will reoccur in the discussion of phase coherence.

Note that this plot is always positive whereas the amplitude of the wave oscillates. This is the basic property of waves. The amplitude of a wave oscillates symmetrically between equal positive and negative extremes. Thus, over multiple cycles its average value vanishes. However, we know that the effects of waves are cumulative. Ocean waves caused serious damage in Fukushima.

Radiation collected by solar panels charge up the lithium ion batteries that power the Tesla. So how do we obtain this average power? Simply by squaring the amplitude which is always positive and by taking the average over many cycles. The wave detector whether it is a photographic plate, a photodiode or a photon counter does the squaring and the averaging. When you square the superposition of two identical waves with differing phases basic trigonometry predicts the interference pattern.

The signal is strongest at the center where the two waves come together in phase as implied by the equality of the lengths of the two paths. However, if they superpose off center the path lengths will differ and if this difference is an odd multiple of half a wavelength they will cancel. If the path difference happens to equal any integer multiple of the full wavelength the two waves will superpose in phase, the signal strength doubles but as the detection pixel moves progressively away from the center the detected signal will weaken as the constant pixel detection area subtends a progressively narrower cone of the arriving radiation.

A

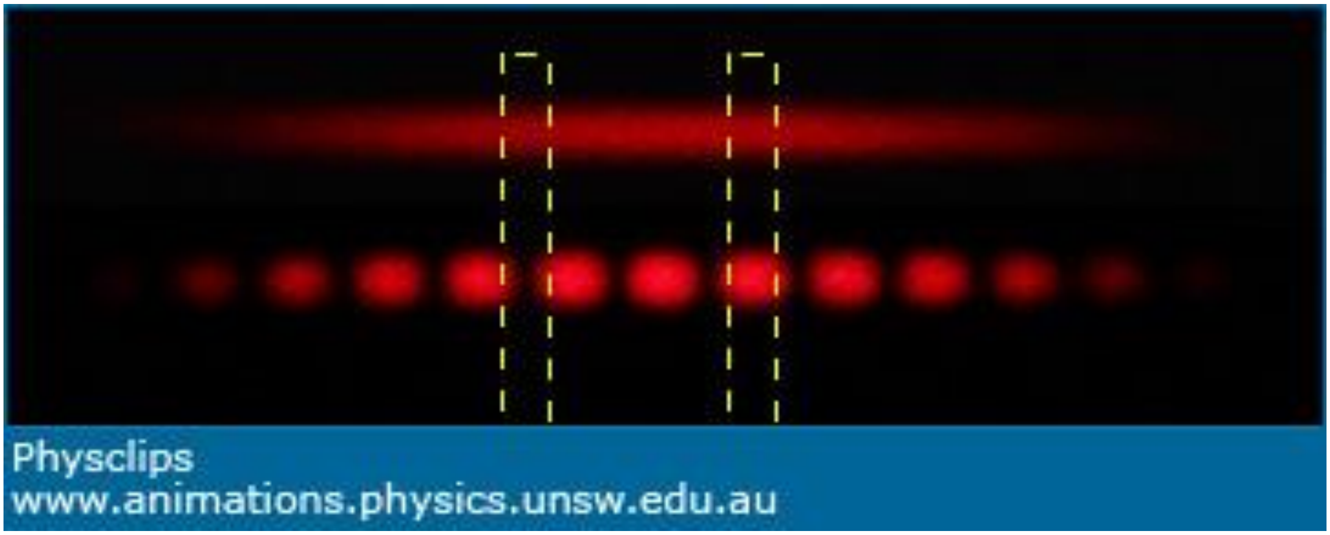


Figure 12(A). Interference fringes as recorded on a photosensitive film. The top spectrum is what you get with one slit open at a time.

B

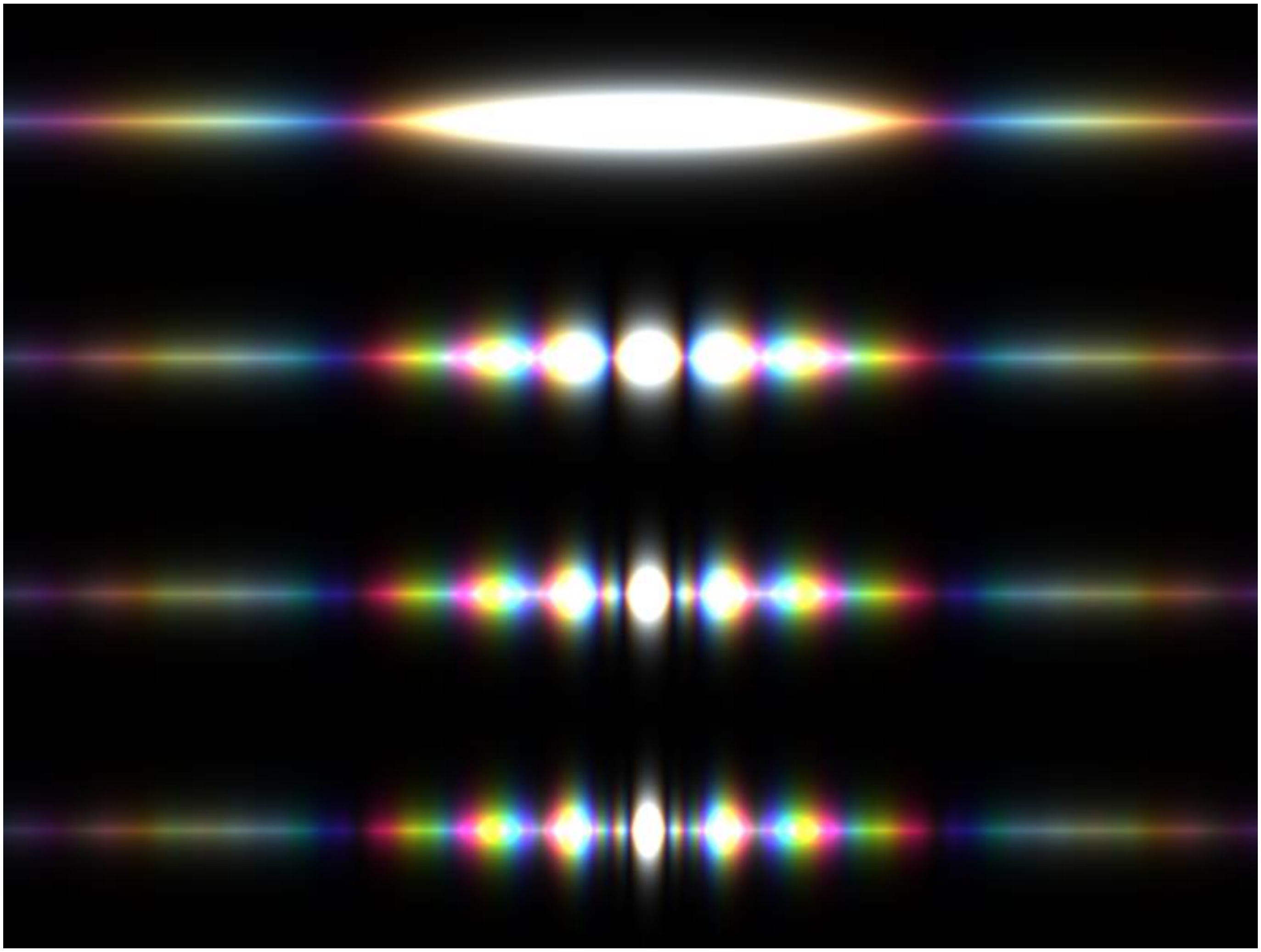


Figure 13. (B) show interference fringes as recorded by a photon counting array.

In order to see interference fringes, both slits must be open simultaneously. When we record the signal with one slit blocked there are no interference fringes, as shown in the top of figures 12(A) and 13(B). With both slits open we obtain the pictures at the bottom, showing the interference fringes.

Figure 13(B) show interference fringes as recorded by a photon counting array.

The first display shows no fringes as the signal was the accumulation of signals from two sequential exposures with alternating slits open one at a time. The next 3 display fringes recorded with the detector array rotating away from being parallel to the plane of the slits. Again, no interference unless both slits are simultaneously open.

This puzzling effect, just by itself, illuminates the dual nature of radiation. The power that is the energy flow carried by the radiation is not the whole story.

The entity that represents radiation is a time dependent field, that carries, besides a magnitude, a phase signature imposed at the instant of first observation. It turns out that the power carried by the radiation is related to its field magnitude by a quadratic relation. When the superposition is applied to the field rather than to the powers the interference shows up. This brings up the notion of phase coherence between radiation signals detected by separate detectors. At this point we assume that the notion of phase of a vibration, or of a wave, is now understood.

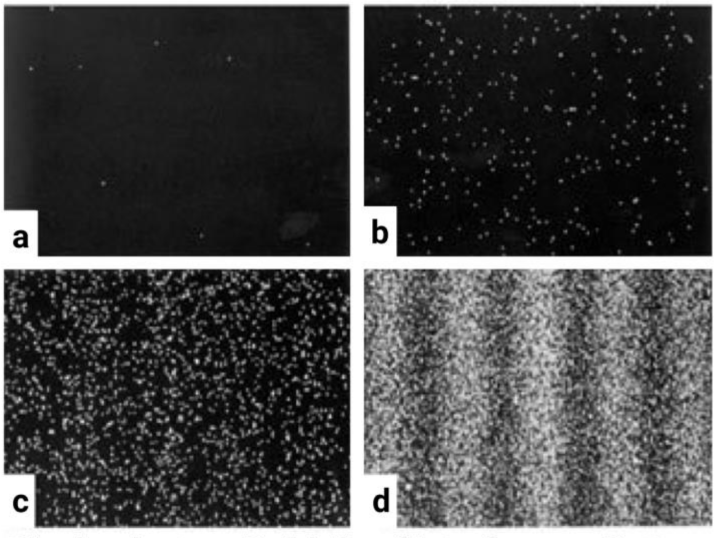
Consider a point light source behind the double slit screen. If on the detector screen we see the interference fringes we label the source coherent and the degree of coherence is the ratio of the intensity of the fringes to the intensity of the background. Perfect coherence requires identical single frequency sources. The best approximation to a coherent source would be a laser focused by a lens through a pinhole.

By observing sunlight light coming off Newton's prism through a narrow slit suitably located, a narrow range of colors, therefore wavelengths or frequencies, can be isolated from the continuum emitted by the sun or by an incandescent light bulb. As shown in the figure the same could be achieved by suitable variants of double slit interference. A laser will do much better. This is because the laser generates light by “Light Amplification by Stimulated Emission of Radiation”. This process was identified by Einstein in 1905 as a consequence of the laws of electromagnetism. A discussion of this process is left for the section on lasers.

In the next paragraphs of this section we illustrate the hidden process that makes the transition between photons and waves. Furthermore, we discuss the idea that the concept of phase can be preserved through this transition.

Figure 14 below illustrates the buildup of interference from the random but correlated arrival of photons from a laser source to a photon counting array detector. The sequence a, b, c, d. corresponds to steps of increasing light intensity. Each dot indicates the arrival of one photon.

Photon Counting Double Slit Experiment



Single-electron Build-Up of Interference Pattern

Figure 14. From photons to waves.

This figure shows how we move from the particle aspect to the wave aspect of radiation. The wave nature manifests itself as a coherence among the photons of common origin that is preserved during propagation.

It is as if each photon carries a clock running at a constant frequency determined by its energy. All photons of the same energy have their clocks running at the same frequency. The only difference can be the phase difference which is determined by the difference in starting times. The clock starts to tic at the instant of its creation, when a particle makes a downward transition between two energy levels. It stops when the photon disappears, giving up its energy to a particle, enabling it to make an upwards transition.

If, shortly after, the same particle makes the reverse transition, another identical photon is created but with a shifted phase. The phase shift is a direct measure of the time delay. This is the quantum definition of an elastic collision.

However, if the circumstances of the particle change during the delay, the energy, and therefore, the clock frequency will differ from that of the original incoming photon, and this is an inelastic collision, usually accompanied by energy absorption by the particle.

For small frequency changes and for short times this looks or feels just like a phase change, but this process starts to wash out their capability to participate in the interference process. When a phase change becomes a frequency change, photons are being replaced by new photons in rapid steps and the frequency spectrum broadens. This is what we call decoherence.

A perfectly elastic collision does not cause decoherence. The narrower the frequency bandwidth of a sharp light beam, the higher its degree of coherence provided all photons follow “equal” paths. Equal, here, means that every photon in the beam must undergo only perfectly elastic collisions. As the beam diverges, generally, inelastic collisions are taking place. However, if it is redirected by highly polished mirrors or lenses a degree of coherence can be recovered. The property of coherence, just as the property of polarization, manifests itself most obviously at sufficiently high intensity that the number density of photons is very large.

These are collective properties of ensembles of photons. How, if at all, do they manifest themselves at weaker intensities where individual photons can be identified? The answer lies in the assumption that each photon carries its own clock which runs at a constant rpm set at birth when the starting phase is also set. So is the polarization, also set at birth depending on the geometry at the instant of birth. Other than that, photons are indistinguishable. Therefore, they can attempt to interfere with each other, successfully to the extent that their frequency band widths overlap. You have to realize that small frequency variations can be understood as small time dependent phase variations.**@**

@ https://en.baltic-labs.com/2012/04/frequency-modulation-vs-phase-modulation/

Section 3. Particles and Waves.

Diffraction and spectroscopy - The birth of quantum mechanics.

Multiple Slits=Diffraction Grating

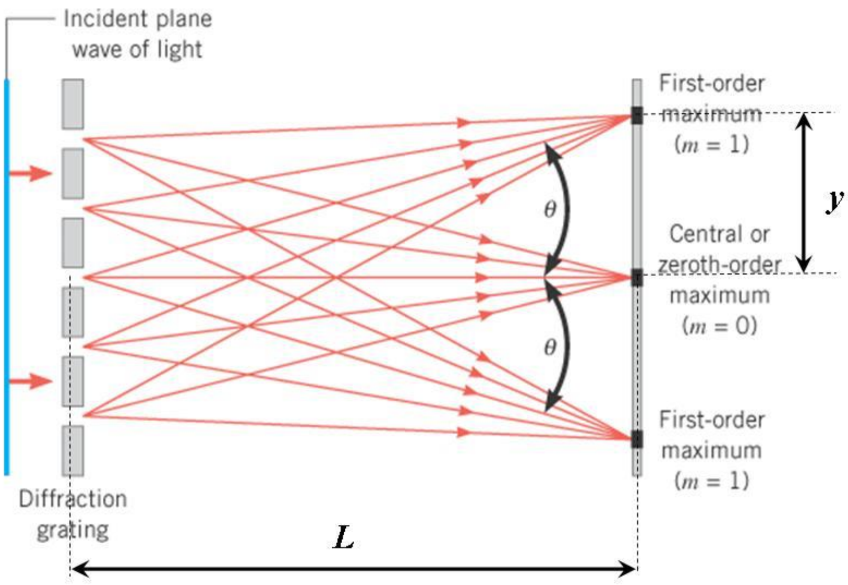


Figure 15. A multiple slit interference experiment

A multiple slit interference experiment with equal inter-slit spacing would simply result in the superposition of similar interference patterns on the detection plane. A scratch on glass scatters light just as the slit edges. Why not replace the multiple slits by a set of parallel scratches on a flat piece of glass or on a flat mirror? Thus, a diffraction grating was born.**@**

The diffraction grating, in combination with a more elaborate system of lenses and curved mirrors, is the basis of high-resolution spectroscopy.

@ https://en.wikipedia.org/wiki/Diffraction_grating

The examination of high temperature light sources (sun, fire) with this technique gave birth to Quantum Mechanics.

Experiments can be “explained” by accepting the following principles. Light provides the energy in proportion to frequency, **$E=h\nu$.** This energy arrives in packets of just this size. There is a minimum packet size to free 1 electron and the current depends on the packet size and the rate of arrival of packets **$I=Nh\nu$, N=packets/second**.

Inspiration for these ideas can be attributed to art and music, the study of how sound is produced by plucking the strings of a guitar or a harp. The mechanism that produces musical sounds involves the notions of Superposition of different frequency vibrations, and of confined vibrations. The notion of confined vibrations was inspiration for the Bohr orbit model and the de Broglie extension thereof.

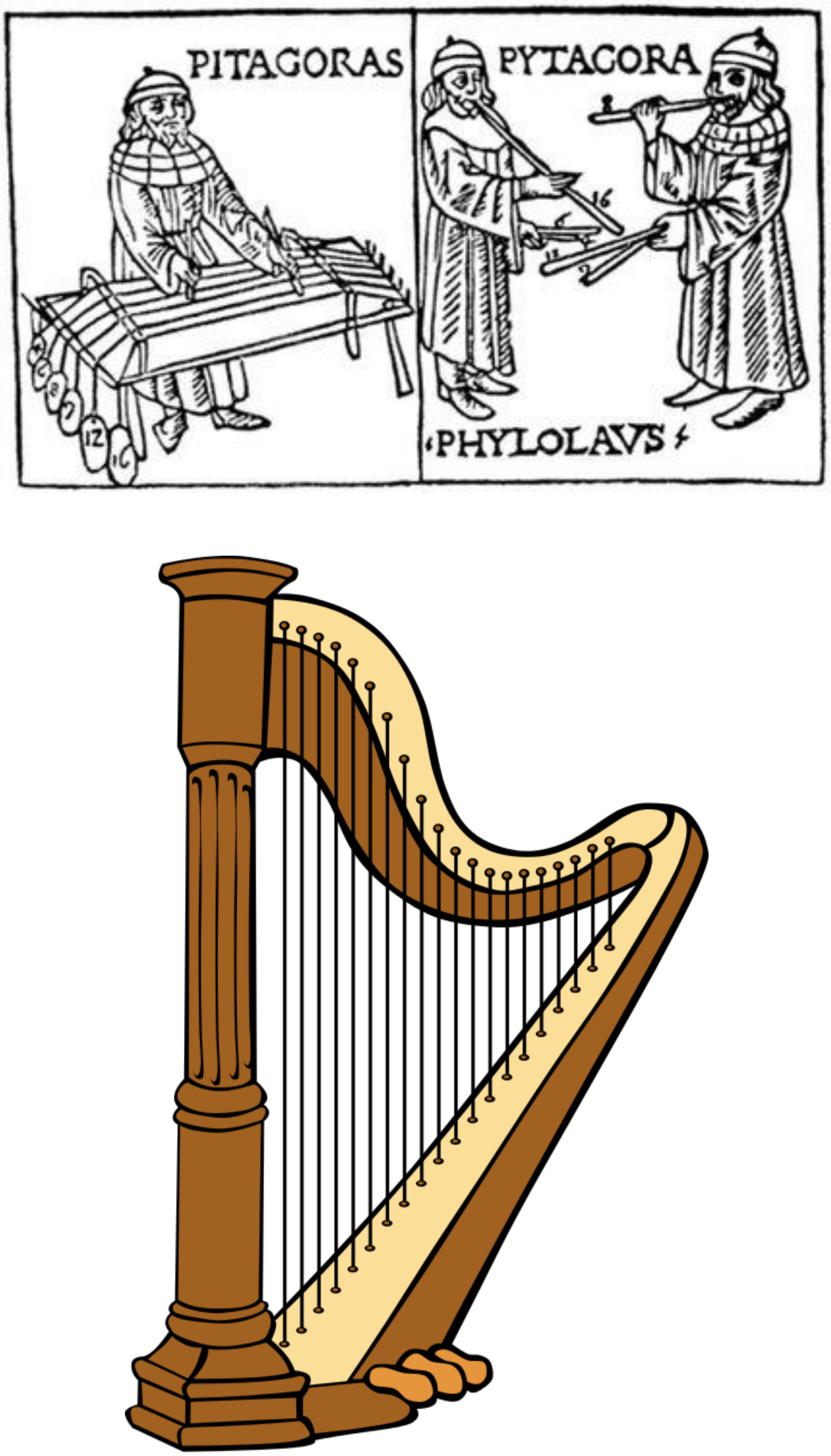


Figure 16. Pythagoras's Harp and a modern harp.

Pythagoras discovered the nature of harmony of sound. A plucked string with length d between secured ends, will tend to vibrate with a fundamental frequency $\boldsymbol{\nu = 2d/c, c = (K/D)1/2,}$ where **K** is the elastic constant and **D** is the mass density of the string's material, c is the speed of sound in the string. Multiple strings vibrating simultaneously generate a chord that can sound agreeable only if the ratio of the frequencies happens to be integers.

A stretched string vibrates at a frequency determined by length and tension. The lowest possible frequency that can satisfy the constraints of the string is called its Fundamental Frequency. The fundamental motion of the string is described by harmonic sine waves both in space and time just as a linear projection of the circular motion of a point. By superposing harmonic motions of different frequencies, any kind of motion might be described. If we fix the endpoints, only those harmonics corresponding to vibrations that do not require motion of the endpoints can form. This is what is called a boundary condition. The endpoints must remain stationary, that is, be at a node. Therefore, it follows that the longest possible wavelength can only be twice the distance between the endpoints. This is the fundamental vibration, with wavelength $\boldsymbol{\lambda_0}$**.** It will have the lowest frequency $\boldsymbol{\nu_0=c/\lambda_0}$ possible for a tone produced by this string.

Any tone produced by such a string has a wavelength related to the length of the string by an integer or half integer ratio. Every fundamental vibration has a countable infinite number of harmonics. The explanation is that, the more common harmonics the chord components possess, the more harmonious the chord sounds. From these experimental facts Fourier Analysis ensued.

The vibrations of such a string are also known as confined waves -- confined by the fixed ends of the string. The confined wave designation implies wave vibrations only at selected wavelengths determined by the nature of the confinement boundary conditions.

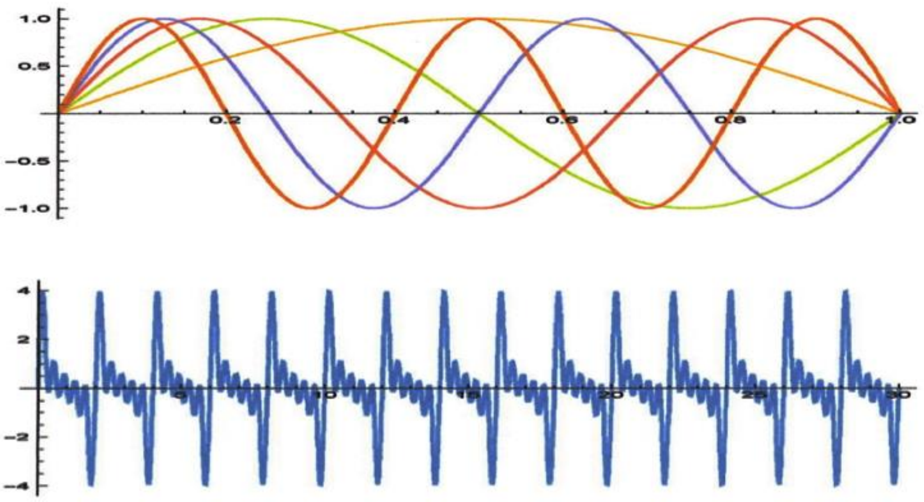
The superposition of waves of different wavelengths subject to the same boundary conditions will produce a train of repeating pulses (repeating plucks) at the fundamental frequency, as demonstrated in Figure 17 below. The peak strength of the pulses can be maximized by putting the phases of all contributing harmonics to $\boldsymbol{\varphi_n = 0}$. This means that all harmonics start at their zero amplitude, the harmonic function for the nth harmonic being $\boldsymbol{\sin(n\omega t-\varphi_n)}$ which includes the phase. This is the time dependent factor, but there is another analogous factor for the positional dependence involving the wavelength rather than the frequency. Wavelength times frequency equals the speed of propagation of the vibration.@ This is so because the phase $\boldsymbol{\varphi_n}$ is a linear function of the length up to this point, as measured in units of the wavelength. The trigonometric expansion:

$$\sin(a+b) = \sin(a)\cos(b)+\cos(a)\sin(b)$$

when applied to $\boldsymbol{\sin(n\omega t-\varphi_n)}$ shows that the combined spacetime dependence of the wave has two same-frequency components de-phased by 90 degrees. (See in the Appendix the section on Exponential notation for Trigonometric functions.) You can look at these components as projections of a rotating vector on two orthogonal axes and the period of rotation is inversely proportional to the frequency. There you have it, the connection of wave to rotation.

@ https://en.wikipedia.org/wiki/Euler%27s_formula#Relationship_to_trigonometry

Confined Waves = Selected Wavelengths Superposition of Different Wavelength = = Train of Pulses



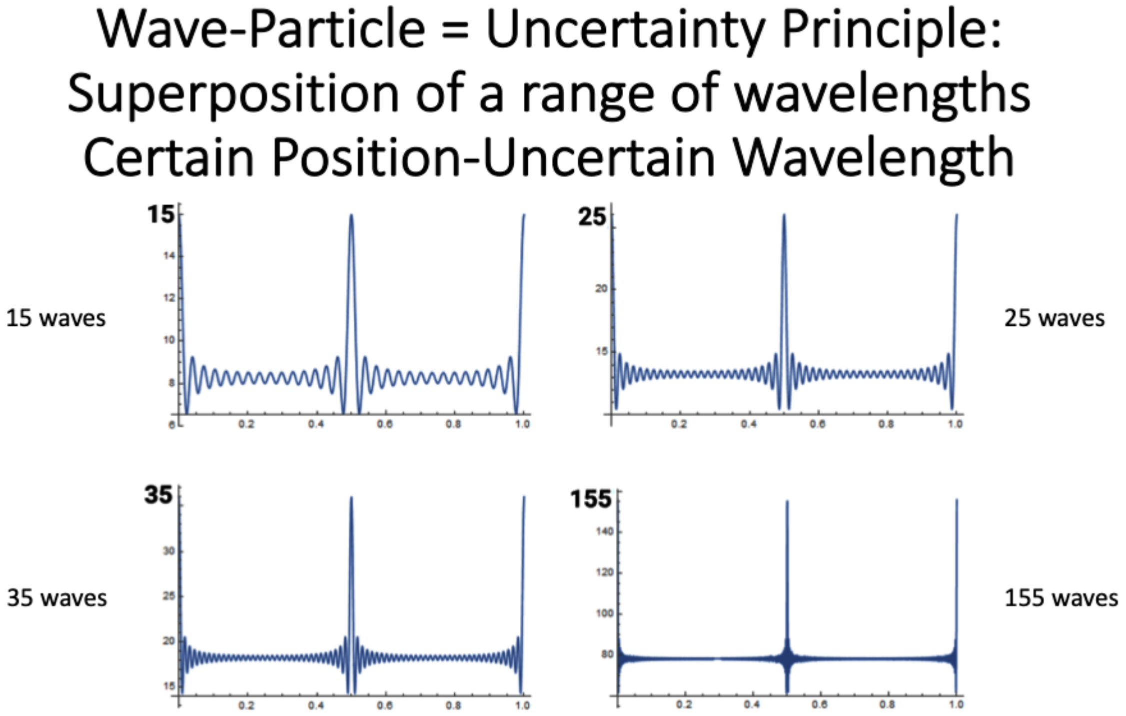

Wave-Particle = Uncertainty Principle:
Superposition of a range of wavelengths
Certain Position-Uncertain Wavelength

Figure 17. The superposition of harmonics.

Figure 17 shows that the superposition of harmonics will generate pulses with oscillatory decaying tails. The number on the plot is the number of harmonics included.

These tails tell the harmonic content of the pulse when analyzed by Fourier transform computation. By superposing more and more harmonics, the peak sharpens, and the tail weakens. That is, the location of the pluck is better determined while the frequency range increases. Doesn’t this remind you of the uncertainty principle?

What motivated the scientific community towards acceptance of the quantum as a basic unit of energy, were the quantitative results of a group of three experiments. In optics measurement of spectra of sunlight and of light from seeded (with different salts) and unseeded flames of a Bunsen burner. In calorimetry measurements of the temperature dependence of specific heats of solids. And then, to top it off, the photoelectric effect. We will focus on the optical experiments here.

Spectrum of Sunlight

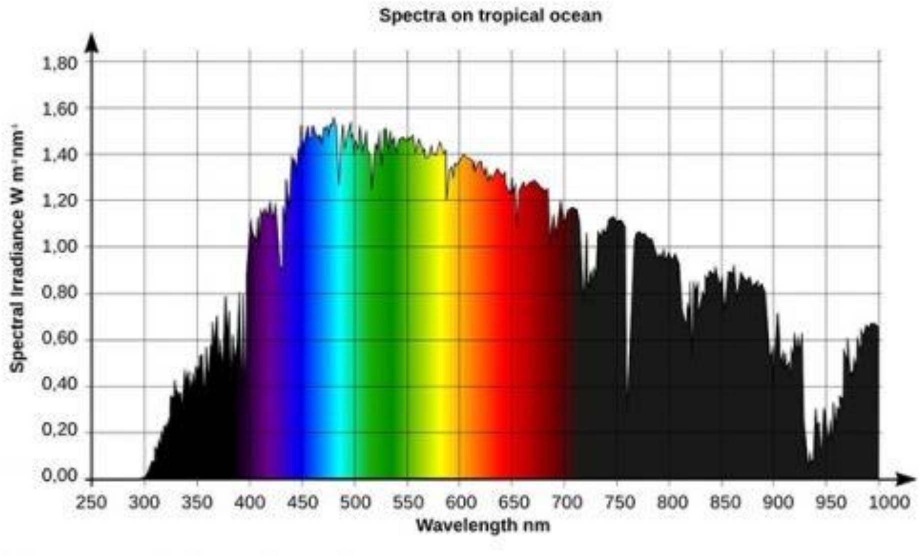
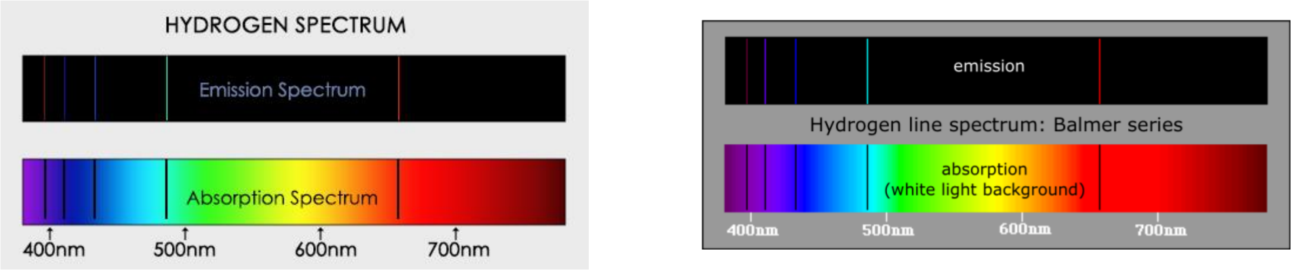


Figure 18. High resolution spectrum of sunlight showing the Fraunhofer absorption lines of hydrogen and some other components of the solar atmosphere.

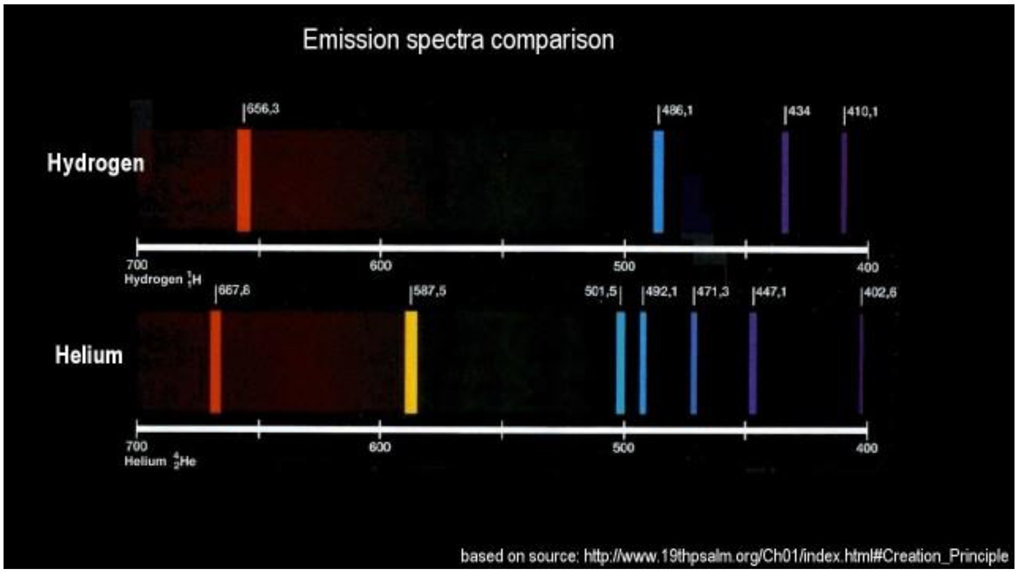
These absorption lines are due to the presence of elements in the form of atomic gases, in the outer atmosphere of the sun, as they absorb radiation at specific wavelengths (or frequencies).



Spectrum of Atomic Hydrogen

Figure 19. Comparison of absorption and emission spectra of Hydrogen. Absorption on left, emission on right

From Hydrogen to Helium



based on source: http://www.19thpsalm.org/Ch01/index.html#Creation_Principle

Figure 20. The emission spectrum of Hydrogen (top) and emission spectrum of Helium (bottom).

Notice the coincidence of frequencies of emission and absorption lines of Hydrogen as you compare with the absorption lines in figure 19.

All “hot” high temperature light sources like the sun emit a continuum of wavelength, and they exhibit sharp dips in emission at certain wave lengths, the Fraunhofer bands. On the other hand, “cool” sources (atoms in gas discharge tubes, LEDs), emit strongly only in selected, characteristic wavelengths. These characteristic wavelengths happen to match Fraunhofer bands, indicating a parallel between absorptions and emissions.

Furthermore, taking the notion of confined vibrations corresponding to discrete frequencies, suggests [Niels Bohr] the idea of selected electron orbits in the hydrogen atom.

The Fourier picture of how a pluck of a string can result from a superposition of smooth vibrations, suggests that whereas a smooth vibration cannot be associated with a specific location, a pluck at a specific location can be specified by a superposition of a large enough range of frequencies. This statement embodies the essence of Heisenberg’s uncertainty principle. A pluck has a specific location on a string. A wave takes a finite length, but then it can accommodate many frequencies, all the harmonics of the fundamental.

As the range of wavelengths of contributing waves to the superposition increases the resulting wave form approaches progressively the shape of a sharp localized spike which now has a well-defined position, but its wavelength is undefined. This is the essence of the uncertainty principle, so basic in formulating the quantum rules which seem to govern events in the domain of the very small and the very fast. There are pairs of properties of real events that, although simultaneously observable are NOT simultaneously measurable. Increased accuracy in the measurement of one of the pair implies decreased accuracy in the measurement of the other. The planetary orbit model had to be modified.

A planetary model will not yield stable orbits as rotating negative charges around a positive nucleus will radiate energy which progressively will shrink the orbit into collapse. To keep an orbit stable, energy has to be supplied at a constant rate. The assumption that energy comes and goes in finite sized packets only, could explain short time stability of such orbits. Upon further elaboration, the following picture emerges: There are discrete orbits, at a discrete set of energy levels.

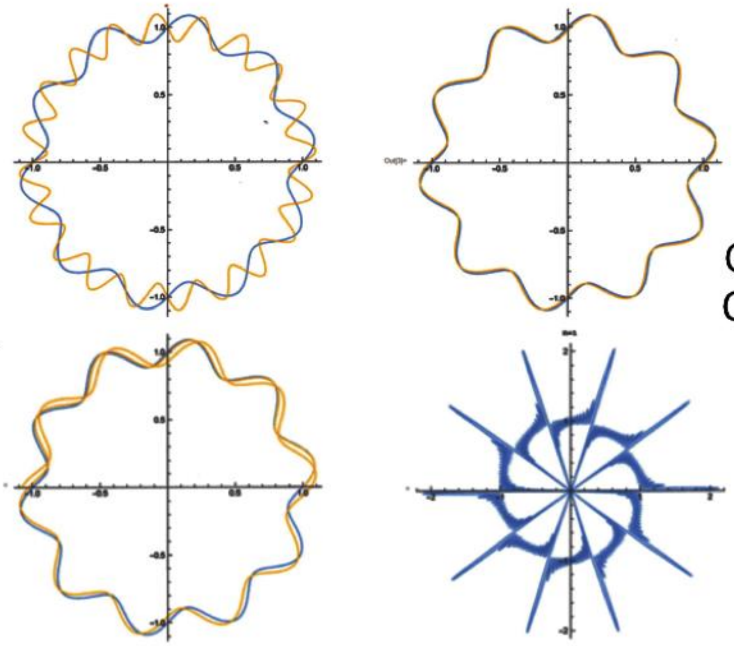
Atoms can jump between orbits either by absorbing or by emitting one packet at a time. And the crucial assumption: **$E=h\nu$,** the energy content of a packet is proportional to the frequency of the radiation. A jump between orbits of energies **$E1$** and **$E2$** requires a packet of frequency **$\nu=(E2-E1)/h$.** This nicely explains sharp emission and absorption lines. It turns out that it also explains the nature of the continuum emission and absorption of energy -- specifically the existence of a frequency maximum beyond which there will be a rapid decrease.

As the packet size increases linearly with frequency, the number of available packets decreases for a given radiation intensity. According to the uncertainty principle, just as the frequency, the energy cannot be measured at an exact time. Therefore, the minimum energy of any elementary particle cannot be exactly zero. It is uncertain in magnitude by the zero-point energy **$0.5h\ \nu$.** Therefore, this amount of energy has to be ascribed to any elementary particle. The energy per elementary particle after **N** deliveries will be **$Nh\nu$ +/- $0.5h$.**

Section 4. Atoms: The Inside and Outside

Confined waves illustrate a concept basic to quantum theory, that boundary conditions can limit the available vibration frequencies to a discrete countable set. Fixed point boundary conditions can be generalized to cyclic boundary conditions for waves propagating around a closed loop like an orbit (De Broglie.)

A wave can propagate along a circular path and be confined by the requirement that maxima in successive cycles exactly superpose. This "cyclic boundary condition" requires that the wavelength be an integral submultiple of the length of the circle $\mathbf{2\pi\rho/\lambda=N}$ an integer.



Waves Confined to Circular Orbit Orbit Length Integral Number of Wavelengths

Superposition of Confined Waves in Circular Orbits, the Classical Limit

Figure 21. Illustration of cyclic boundary conditions.

To introduce the next topic, we will give a brief survey of the photoelectric effect and its role in the birth of the quantum theory.

FACT: Shining light on some materials can cause electrons to be released and flow. Shining on the negative metal cathode of a vacuum tube with positive anode can cause a flow of electrons towards the anode. Current increases with increasing light intensity.

BUT: depending on the metal, there is a minimum frequency threshold for electron emission. Above this frequency the current also increases with increasing frequency.

EXPLANATION: Current carriers are negatively charged electrons bound inside the metal needing influx of a minimum of energy to free them.

Light provides the energy in proportion to frequency**, $E=h\nu$.** This energy arrives in packets of just this size. There is a minimum packet size to free one electron. Beyond this threshold, the current depends on the packet size and the rate of arrival of packets **$I=Nh\nu$. N** is the rate of packet arrival, **packets/second.** The increase with frequency indicates the increase of packet size. So, the current also increases with increasing frequency equivalent to increasing packet size.

We call the packets photons. They are the quanta of radiation.

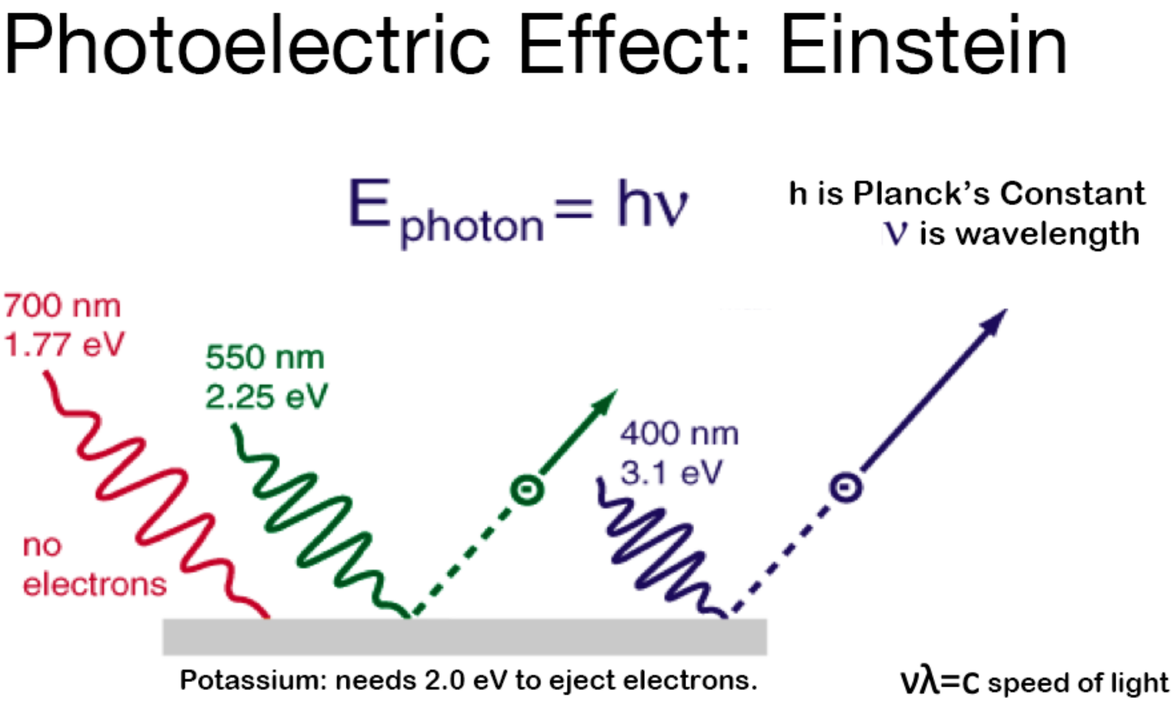


Figure 22. The photoelectric effect.%

At this point we will introduce the language of quantum mechanics by discussing the notion of a wave function and of an operator.

At any point **x, y, z,** in space and t in time, **(x, y, z, t)**, temperature has a value **T**. So, this defines a function:

$$\mathbf{T = f(x, y, z, t)}$$

We can say the same of air pressure:

$$\mathbf{p=g(x, y, z, t)}$$

% https://en.wikipedia.org/wiki/Annums_Mirabilis_papers. Einstein Annums Mirabilis papers (Einstein's miraculous year)

In quantum mechanics we associate a function, the wave function,& **u(x, y, z, t)** with the state of the electron. The meaning of this function is associated with a performance of a measurement of a certain property of the electron at that location in spacetime.

Let us consider the simplest case of an electron confined to a one-dimensional space, say a straight line or a circle, and we want to know if the electron is or is not detectable at a point. So, we position an electron detector at that point and wait for a response, which could be a point in an imagined screen as in Figure 15. If the position is at **x** we can record the time **t** at which it occurs and **u(x,t)** is assigned a large value to that point **u(x0,t0)=**$\boldsymbol{\delta}(\boldsymbol{x0}, \boldsymbol{t0})$ a Dirac delta-function at that point, zero elsewhere.

This is unrealistic because we can never look at a single point or during only a single instant. We must spend a finite, however short, time looking at a small area. We can record how many electrons we have counted. This is always a positive number. Now we also know that this function must reflect the wave characteristic of interference, well documented for photons as well as electrons. By assuming that it is a wave analogous to sound waves, which are in truth, strain waves, the following picture emerges. The wave function has an intensity and a phase which oscillates in time as well as in space. It has to have two components and a positive absolute value. The wave itself can be represented by a complex number (see #2 Appendix B) whose squared absolute value is always positive. It measures the number of electrons counted in the small space-time interval of our experiment. (see #2 Appendix B)

&https://en.wikipedia.org/wiki/Davisson%E2%80%93Germer_experiment
The Devisson-Germer Experiment

This justifies the probabilistic interpretation of the wave function, proposed by Niels Bohr and Max Born. For a state of well-defined constant energy, the basic equation governing that wavefunction is an *eigenvalue equation* as shown below: (see #3 Appendix B for more on eigenvalues):

Time-**Dependent** Schrodinger Wave Equation

$$i\hbar\frac{\partial}{\partial t}\Psi(x,t) = -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}\Psi(x,t) + V(x)\Psi(x,t)$$

PHYSICS NOTATION

Total E term

K.E. term

P.E. term

$$\Psi(x,t) = e^{-iEt/\hbar}\psi(x)$$

Time-**Independent** Schrodinger Wave Equation

$$E\psi(x) = -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}\psi(x) + V(x)\psi(x)$$

Equation 1. Time dependent and time independent Schrodinger @equation for the wave function of an electron in a one-dimensional space.

Since Newton and Huygens, scientists and mathematicians have tried to describe the collective movements of water and air by some extension of Hooke's equation for elastic vibrations. These were designated as wave equations and these quantum equations follow that pattern. Schrödinger’s

@https://chem.libretexts.org/Bookshelves/Physical_and_Theoretical_Chemistry_Textbook_Maps/Book%3A_Quantum_States_of_Atoms_and_Molecules_(Zielinksi_et_al.)/03%3A_The_Schr%C3%B6dinger_Equation/3.04%3A_Operators%2C_Eigenfunctions%2C_Eigenvalues%2C_and_Eigenstates "invention" of quantum mechanics wave equation came out of this long tradition.

The time independent equation applies for stationary states, meaning states in which an electron remains for long times before making a transition to another state. The differential term represents the kinetic energy, **V** the potential energy and **E** is the total energy that the electron would have, were it in this state. $\boldsymbol{\psi}$**(x)** is the wavefunction, **x** is the coordinate of the electron's possible location.

This equation is also the symbolic version of the conservation of energy.

The equation suggests the differential operator interpretation:

The three partial derivatives are the components of the gradient vector representing the spatial variation of the wavefunction. The time derivative is interpreted as the velocity, the second derivative being thus the acceleration.

At this point we have to assume that the reader is familiar with Newton's three laws of motion and the concept of potential energy. An eminently readable account can be found at khanacademy.org online, Newton's laws in Google.

Solving this equation means finding functions $\boldsymbol{\psi}$**(x)** and corresponding values for **E** that satisfy this equation at every value of **x** within a prescribed domain as well as prescribed conditions at every boundary location. The transformation of the time derivative into the multiplicative constant is due to adoption of a complex exponential form for the time dependent factor of the wavefunction.

The value of **E** becomes the energy level of that state. There may be multiple wave functions with the same value of **E.** This means that this is a degenerate state and any linear combination of these functions is also a solution. This is so as this is true of all linear equations.

This equation is valid for stationary states where there is no time evolution of the state other than that implied by the Planck relation which relates frequency to energy by a particular form of the time dependent wave function. For situations where time evolution of the potential is present, we use a more general equation of the form shown above.

The derivative term on the right is replaced by the multiplicative constant **E** term for an exponential form of the time dependence, due to the properties of the exponential function:

$$de^{x}/dx=e^{x} \quad de^{ax}/dx=ae^{ax} \quad d^{2}e^{ax}dx^{2}=a^{2}e^{ax}$$

If $\psi(x,t) = \psi(x,e^{at})$, due to its harmonic nature,
$d\,\psi(x,e^{at}/dt = a\,\psi(x,e^{at})$ with $a = i\omega$.

Therefore, in the right-hand term of the time independent equation we can replace the derivative term by a constant, times the wavefunction itself, making this equation into an eigenvalue equation for a stationary state

Let us now describe the physical features of the solutions of the Schrödinger equation for the hydrogen atom. This is the simplest case: one electron alone in a time independent central electric field, the hydrogen atom.

The solution is obtained by the method of separation of variables, where the solution is initially assumed to be the product of three functions of one variable each, one of each polar coordinate. Because of the linearity of the equation the process now is split into three one dimensional problems, each, due to the boundary conditions, generating a discrete set of confined waves. Usually this solution yields an acceptable if crude picture of the wavefunction which then could be refined by progressive stages of perturbation calculations. We won't go there at this point. We will assume polar coordinates, ρ the radial distance from center, and the two angles θ and φ.

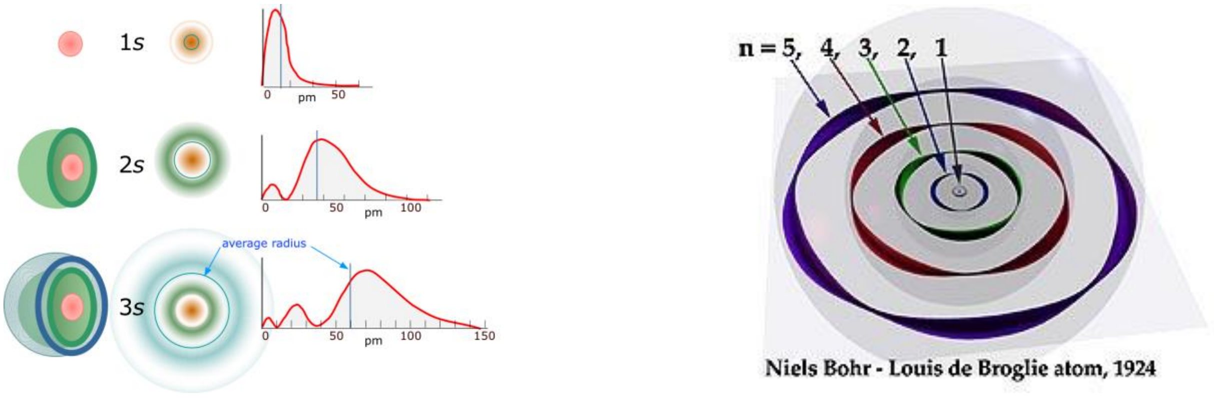
$$\psi(x,y,z) = \psi(\rho,\theta,\varphi)$$

The quantum states of an electron in a hydrogen atom are characterized by three quantum numbers: the principal quantum number n from the radial coordinate refers to Energy and to boundaries at the nucleus and to a faraway spherical surface. The angular momentum magnitude **L**, that for a given **n** can have **2n+1** values (the range), and the magnetic quantum number **m**, which refers to components of **L** projected on a given axis, and which can have **2L+1** integer spaced values between **-L** and **L**, refer to cyclic, de Broglie boundary conditions.

These numbers also characterize the geometry of the wave function, in terms of the position and number of maxima and minima (peaks and valleys). For the one-dimensional space along the radius, particle moving on a straight line, n indexes the maxima. For the boundary condition that the solution should remain finite everywhere the solution has to go to zero as the position moves towards infinity, the total number of maxima has to be finite. Therefore, **n** goes between **0** and **N,** the total number of maxima. This is the principal quantum number for this particular solution. From now on we use **n** and **N** interchangeably.

Position now is specified by three cartesian coordinates **x, y** and **z**, or three polar coordinates $\boldsymbol{\rho}$,$\boldsymbol{\varphi}$ and $\boldsymbol{\theta}$. Thus, the coordinates **x, y** or **z** will be multiples of a suitable unit measure. For the one-dimensional space around the circle cyclic boundary conditions apply and these boundary conditions yield the limitations that are imposed on the possible values of the orbital angular momentum **L** by the quantum rules. (see#4 Appendix B) These cyclic boundary conditions also impose similar limitations on the allowed values of the angular coordinates.

Cartesian symmetry implies points on a rectangular grid whereas polar symmetry implies points placed on regularly spaced concentric spheres, or, in 2 dimensions, circles. The boundary conditions are the source of the quantum numbers for angular momentum and its projection on any chosen axis.



Niels Bohr - Louis de Broglie atom, 1924

Bohr Orbits in Hydrogen

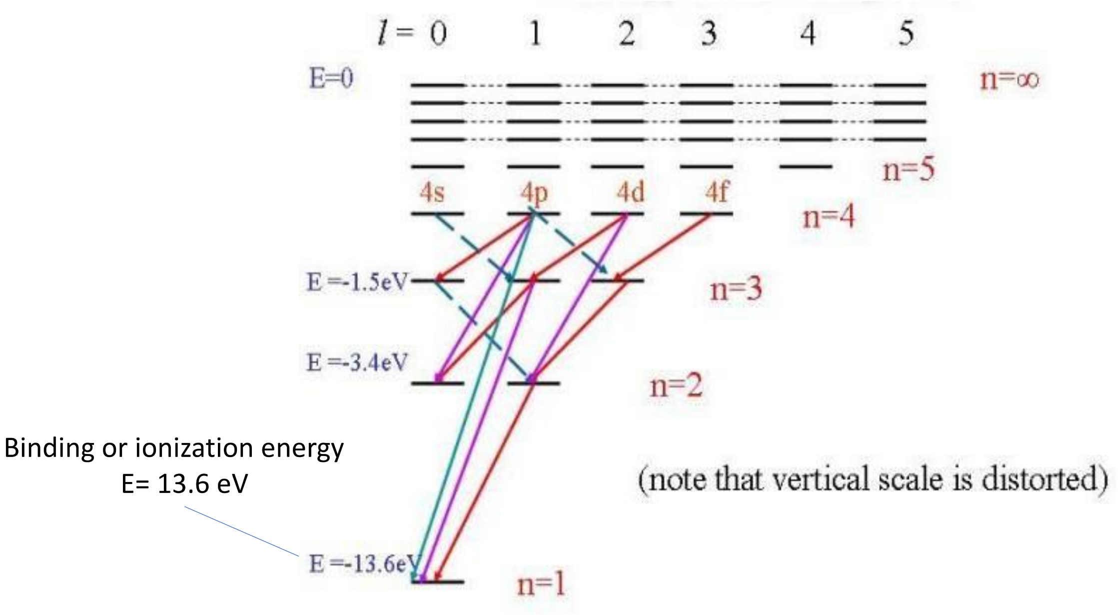
s states p,d,f...states

Figure 23. This is an attempt to picture the orbitals for the various states of the Hydrogen atom.

The pictured orbitals show their basic spatial symmetry, but those that correspond to the same energy level, say two **2s**, and the three **2p** orbitals can be superposed, by linear combination, to form other possible orbitals at the same energy level. Some are asymmetric. This superposition capability into asymmetric orbitals is the quantum mechanical rule that allows the formation of the chemical bonds between, atoms, that generates molecules, proteins, DNA, cells, … us.

Mixed asymmetric orbitals on neighboring atoms overlap and combine to form the covalent chemical bonds of molecules. We will come back to this topic after dealing with atoms containing more than one electron.

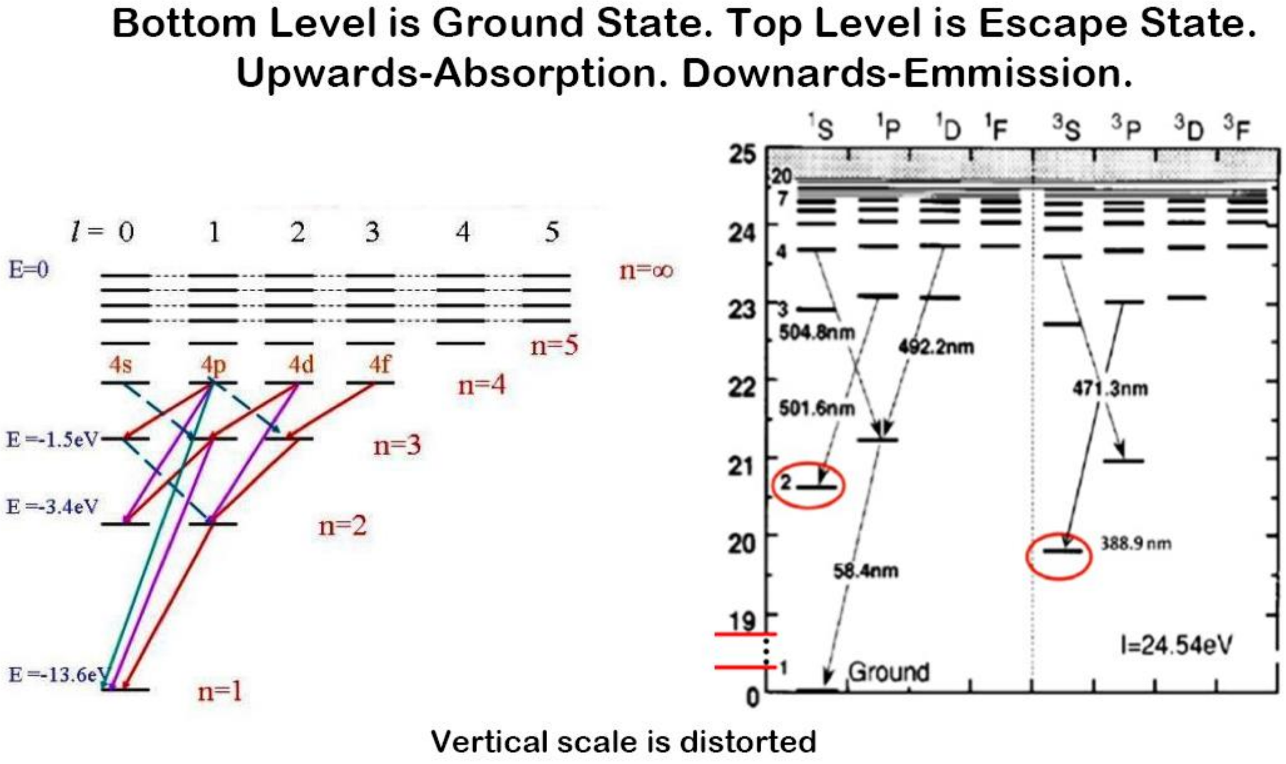
Bottom Level is Ground State
Top Level is Escape State
Upwards-Absorption; Downwards-Emission



Binding or ionization energy
E= 13.6 eV

Figure 24. The energy levels of the hydrogen atom.

So far, we have only dealt with the hydrogen atom for which the Schrödinger equation provides a rigorous basic framework. Beyond hydrogen, for atoms with more electrons, any treatment is approximate. This is so because at small distances the quantum interactions become comparable to, or may even dominate, the electromagnetic interactions and have to be worked out by trial and error from new experimental data.


Bottom Level is Ground State. Top Level is Escape State.
Upwards-Absorption. Downards-Emmission.

Vertical scale is distorted

Figure 25. Comparison of the highest Energy levels of the Helium atom (right) and the Hydrogen atom. Here we are comparing escape levels.

With **n** electrons the nuclear charge is **+n.** If the electrons are added one at a time, the first electron feels the full nuclear charge, so its orbitals will be confined to a region very close to the nucleus and as further electrons are added the confinement region increases outward in volume as the effective charge it sees progressively decreases, until it reaches one. This is the escape level and its wave functions should be somewhat like those of hydrogen. See Figure 25.

Notice though, going from hydrogen to helium, the number of energy levels did not simply double, at least in the absence of a magnetic field. Wolfgang Pauli’s exclusion principle states that two electrons cannot be in the exact same state in the same energy level. This required him to later add the notions of spin and spin degeneracy in such a manner as to explain the magnetic field effect (covered later). The binding energy of this last electron should be comparable to that of hydrogen and its energy level diagram should also show some similarities. The experiments tell a somewhat different story, but they do hint to a possible related scheme, that of periodicity and a so-called shell structure. A large bulk of chemical experiments did strongly suggest some form of periodicity in chemical reactivity which, it was conjectured, was related to the behavior of this least bound n_{th} electron.

Energy required to liberate first electron as function of total number of electrons

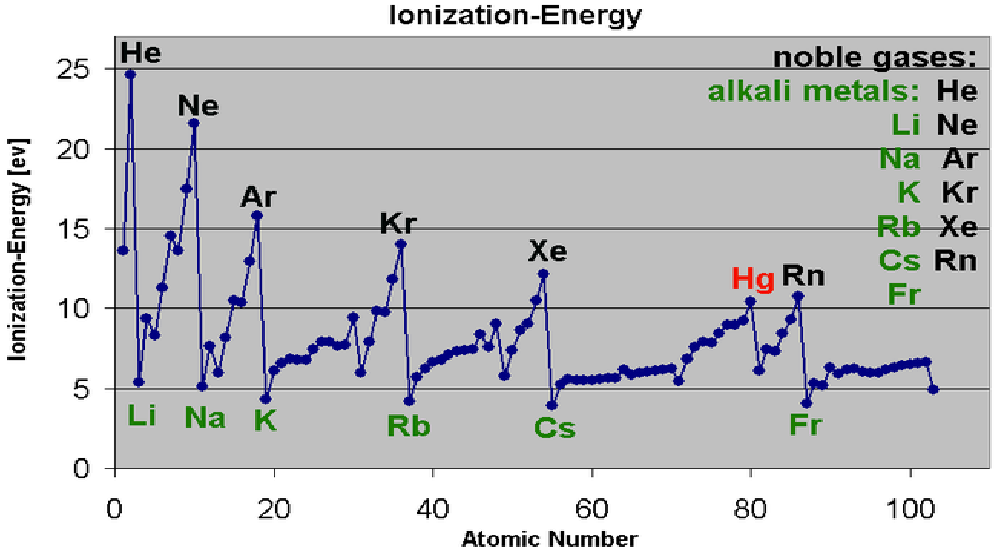


Figure 26. Escape energy, as a function of nuclear charge, atomic number, or number of electrons.

Inspired by this plot, it is reasonable to assume that past Hydrogen the electrons are added in successive shells as illustrated by the periods of the ***Periodic Table,*** Figure 28. The first short period on the table goes between **He** and **Ne,** as the three **2p** and one **2s** orbitals, of the first shell get populated. Beyond **Ne**, the **4d** orbitals are populated ahead of the **3s** orbitals forming the next shell. This is so because the **4d** orbitals have an energy slightly lower than the **3** orbitals. Their respective energies are close, the **4p** being slightly lower. The **4d** shell remains partially filled and the **4s** electron becomes the outer, the escape electron or valence electron. This effect accounts for partially filled inner shells that play a fundamental role in para and ferro magnetism. (see#5 Appendix B.)

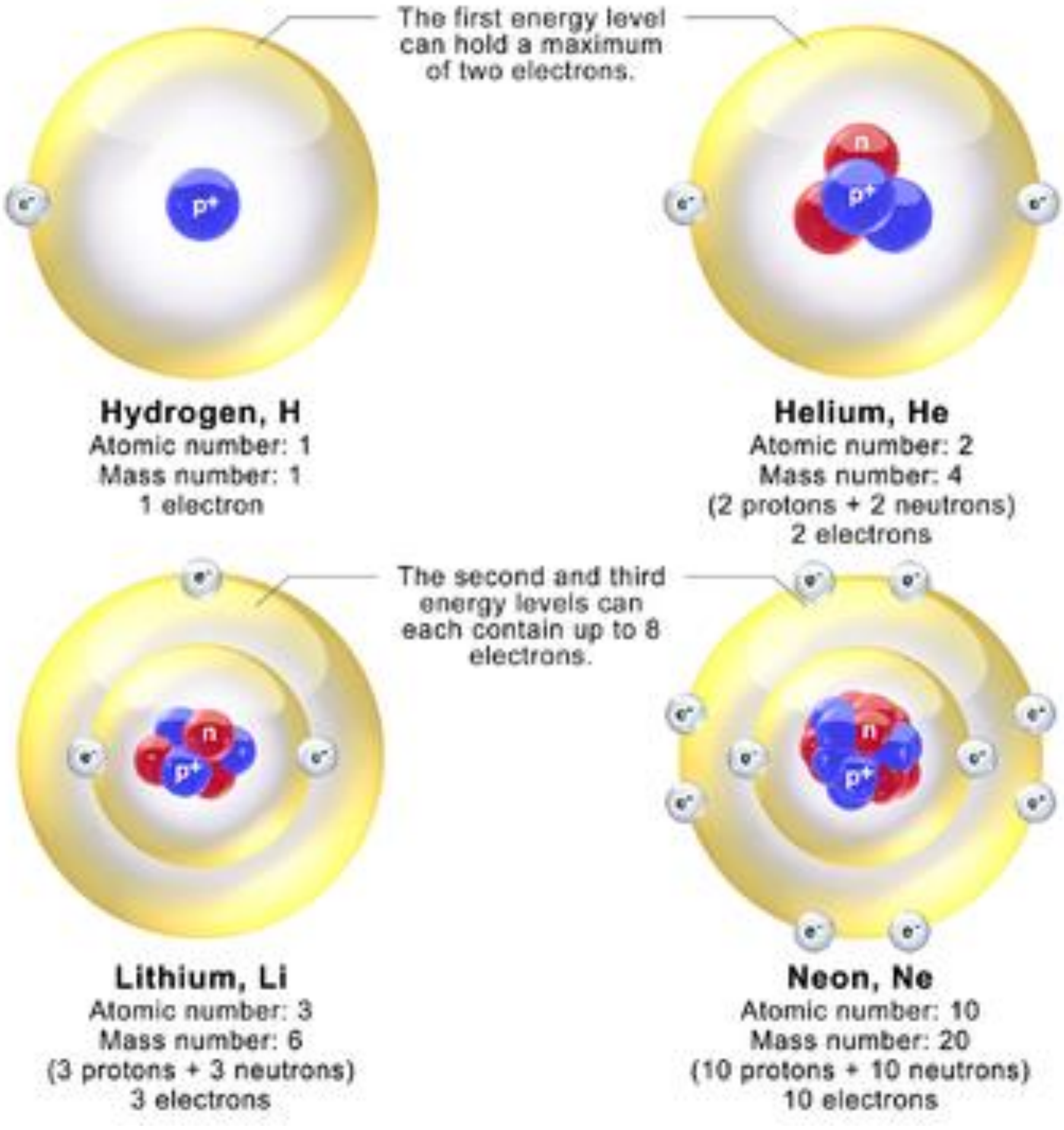


Figure 27. The Shell Model.

As we move further adding nuclear charge and electrons, this sequence of partially filled shells repeats generating all the magnetic atoms, including the rare earth, of present-day notoriety.

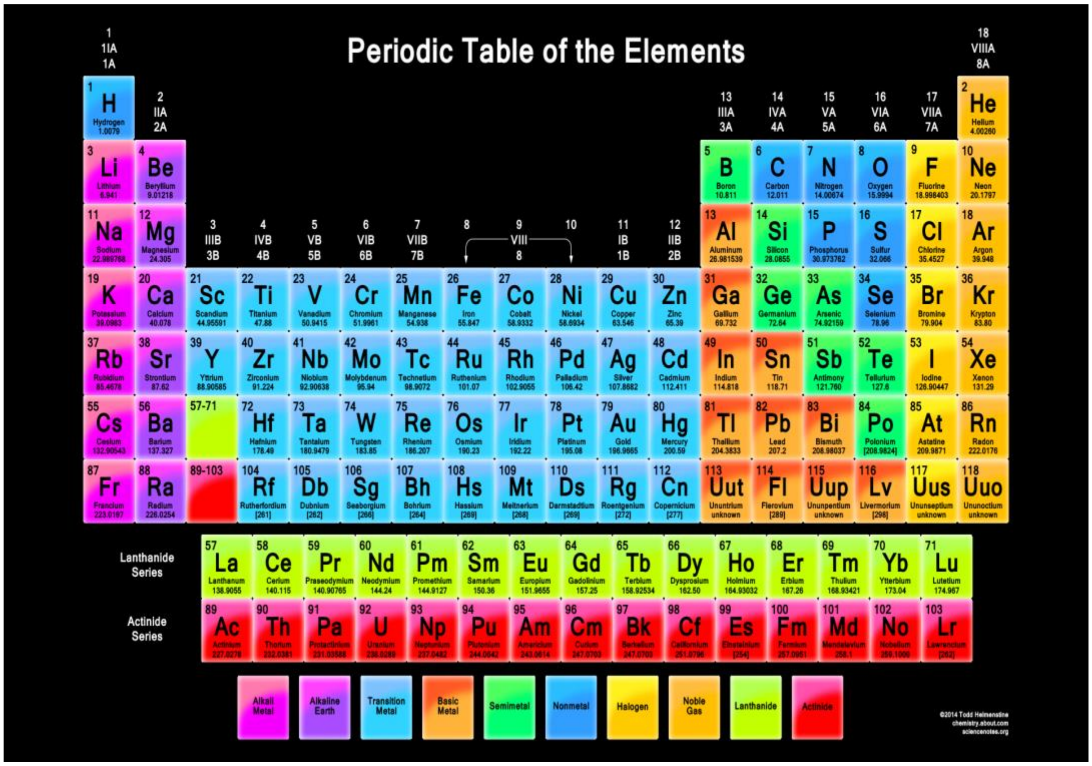


Figure 28. The Mendeleev periodic table.

The buildup process and the exclusion principle, at first without spin, imply one electron per orbit. Then with spin added, it becomes two electrons per orbital, defining the magnetic degeneracy.

For Helium the nuclear charge is +2, it so needs two electrons to be neutral. Add one at a time and stir gently. The orbitals of the first electron should be similar to those in Hydrogen but shrunken in size due to the doubling of the electrical attraction. Then, as the second electron approaches it will, at larger distances, sense attraction of only a net +1 charge, as in the **H** atom. Thus, in further atoms the energy levels of the outer orbital are comparable to those in **H** atom, and so it is as witnessed by the **He** spectral data in Figure 12A.

The energy level scheme for atoms following **He** in the periodic table, for the group of the highest levels, seem to all have a similar hydrogen-like pattern. See Figure 29 for the case of Lithium. This strongly suggests that using hydrogen-like orbitals for individual electrons, at least for the weakest bound electrons, without accounting for the electrical inter-electron repulsion, is a suitable approximation. This gave rise to the buildup principle[%] by starting to assemble the states of many electron atoms by sequentially occupying hydrogenic states. The electron-electron interaction is roughly accounted for by only allowing two electrons to occupy each orbital. The “inner” electrons, those assembled first would be much more strongly bound by the less shielded nuclear charge into a smaller denser core, to which the last added electron would be bound to the net +1 charge in a hydrogen-like manner. This model yields a reasonable facsimile of the experimental chemical properties encrypted in the periodic table.

[%] https://en.wikipedia.org/wiki/Aufbau_principle

Lithium Energy Levels

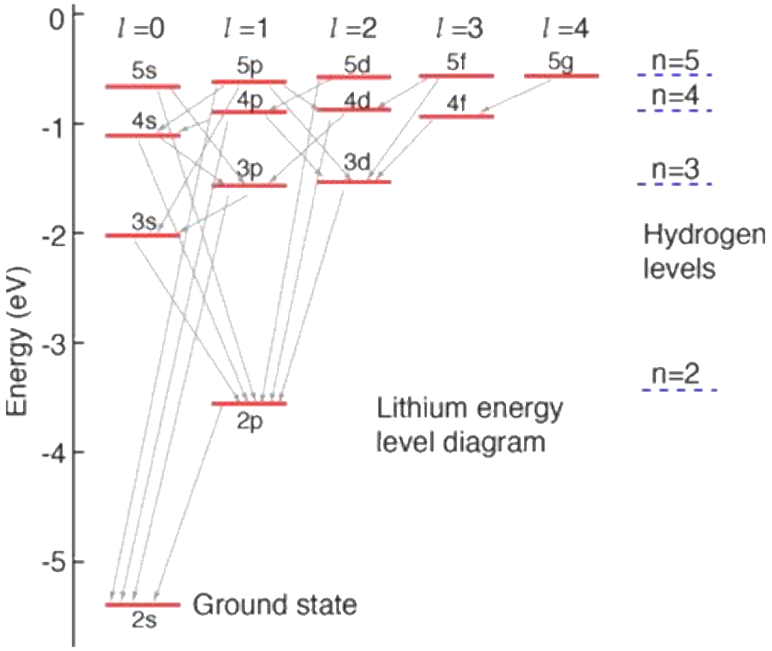


Figure 29. The Lithium energy level scheme.

The new assumption is that as a result of the electron-electron interactions, filled shells become stable configurations. This means two electrons per state. It also predicts that elements with filled outer shells will be chemically stable. These assumptions were sufficient to “explain” atom to atom bonding and the formation of molecules, so we will stick to this subject for the moment.

Section 5. Bonds, Molecules

Consider the formation of the Hydrogen molecule. An individual isolated hydrogen atom has its electron in the 1s orbital all alone. It is therefore unstable. It is what we label a neutral reactive entity, a free radical. Consider what can happen when two such ions approach each other. In the presence of the correct radiation, even if transient, the electron can be excited to any combination of the **2s**, and **2p** states. In these states the charge is distributed in two outward directed lobes, one negative, the other positive. It becomes an electric dipole that will feel a long range electrical attractive force from another such atom, the Vander Waals attraction.

They will come closer and closer, their lobes overlapping, and a short-range quantum mechanical attraction will dominate. The two hydrogen atoms form the covalently bonded hydrogen molecule. The motivation behind this short-range force is the need to have two electrons in the 1s shell of each atom, even if part time.**@**

@ https://en.wikipedia.org/wiki/Covalent_bond

H+H Forms H2 Molecule: How? Orbital Mix, Hint of Spin

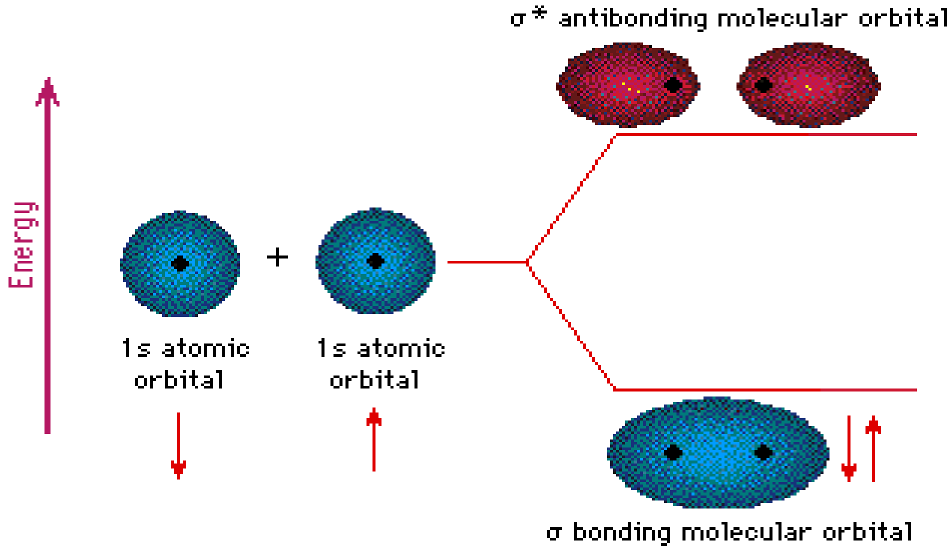


Figure 30. Schematics of the formation of a covalent bond, The hydrogen molecule.

A similar discussion describes the formation of an isolated water molecule. An isolated oxygen atom needs two electrons to fill its outer shell. The two 1s electrons of 2 hydrogen atoms can provide these.

Water Molecule as Electric Dipole
Most Common Polar Solvent of Ions

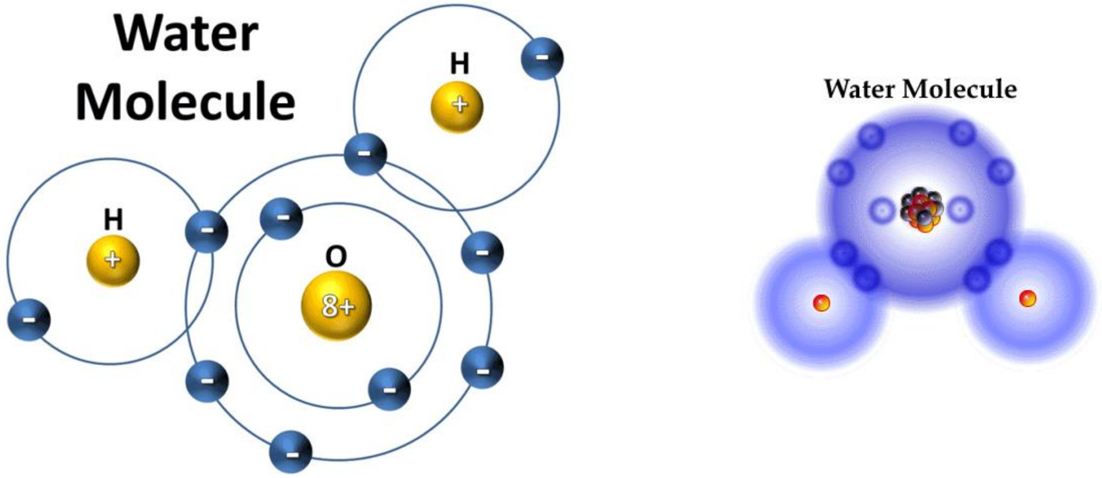


Figure 31. The bonding in the water molecule.

As the single electron in an otherwise empty shell is relatively weakly bound, it might be easily separated from the atom by collisions, say with photons, thereby turning the atom into a positive ion. On the other hand, an atom with an almost full shell will offer a receptive site for a wandering electron turning it into a negative ion. So, if a Na+ and a Cl- pair approach each other, the electric attraction will help the formation of a bond where one can consider both atoms to have filled shells.

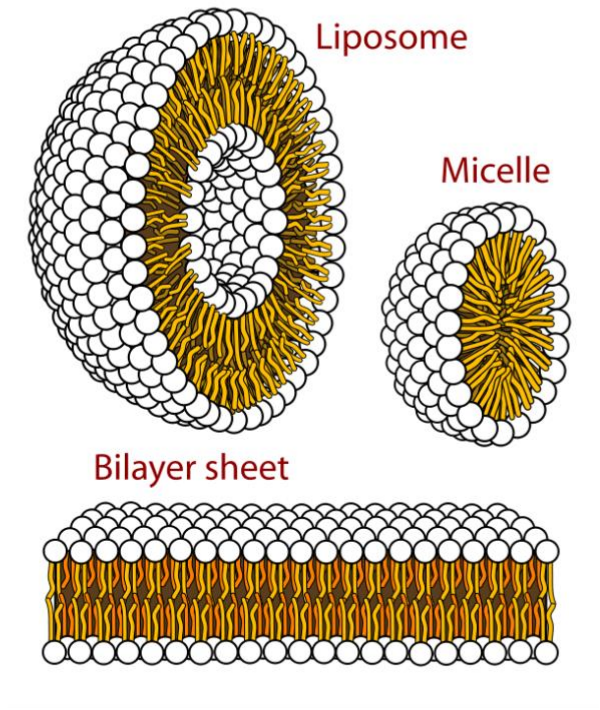
Usually if this happens these molecules will precipitate out. This must also be true for any such corresponding pair with partially filled, but complementary, shells such as **Mg++** and **O--.** This is also the case of half-filled shell atoms such as **C** and **Si** even in the absence of ionization. For carbon, organic chemistry, graphite, diamond follow. For silicon, transistors, laptops, smart phones follow in short order.

The water molecule is actually an electrical dipole, as the center of mass of the positive charges is displaced from the center of mass of the negative charges. Whether an ion is positive or negative, it will surround itself in water with appropriately oriented water dipoles so that it can remain in stable solution even if oppositely charged ions are simultaneously present in the vicinity, e.g. **H+, Cl-** in hydrochloric acid. This property makes liquid water the universal solvent for charged molecules. It is the medium where the most important condensed matter chemical reactions take place, in particular, those basic in sustaining life.

Approaching covalently bonded molecules will tend to assemble into progressively larger insoluble macromolecules. However, when such macromolecules contain charged atoms in their periphery they dissolve, suspended by the electrical forces just as the simple ions. Water is the enabler of biochemistry and life itself.

Proteins are chains of amino acids. They are soluble macromolecules due to the presence of simple ions in their periphery. Lipids are covalently bonded macromolecules perfectly insoluble in water. They can form the internal parts of cell membranes and other organelles that are shielded from water. Phospholipids[+] are covalently bonded macromolecules with strategically located phosphate groups, distributed throughout their periphery. These atomic groups carry a net negative charge thereby solubilizing the macromolecules, the basic constituents of cell membranes.

[+] https://en.wikipedia.org/wiki/Lipid_bilayer
The phospholipid bilayer



The Phosphate Ion at the Periphery of Organic Macromolecules: Hydrophilic, Water Friendly site

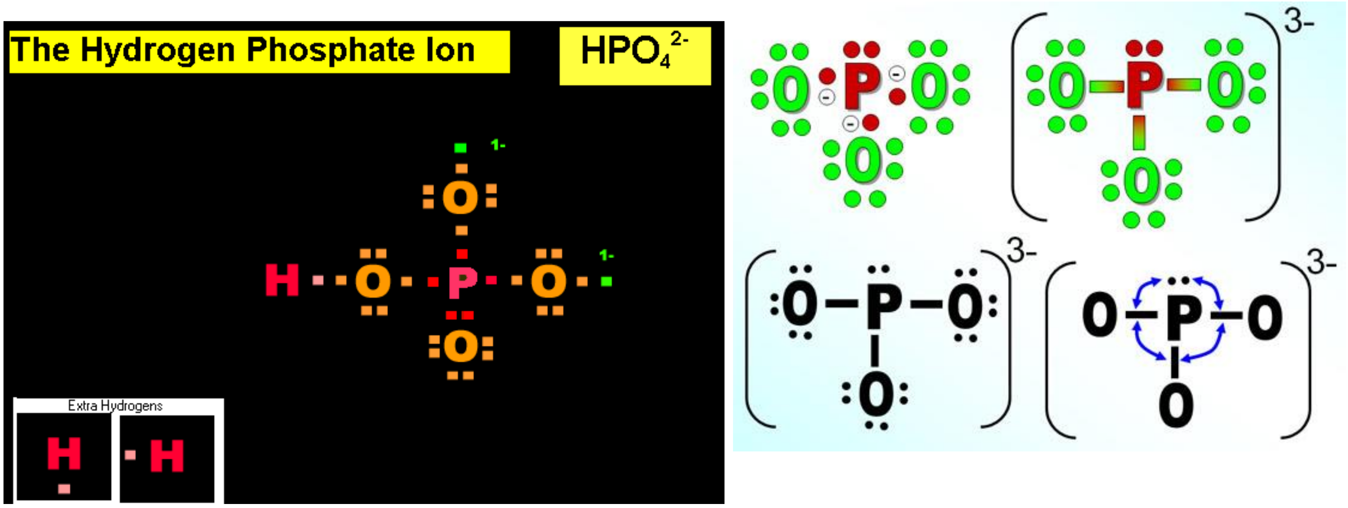


Figure 32. Three main phospholipid structures in solution, and the phosphate ion.

Atoms Assemble to Build a PhosphoLipid

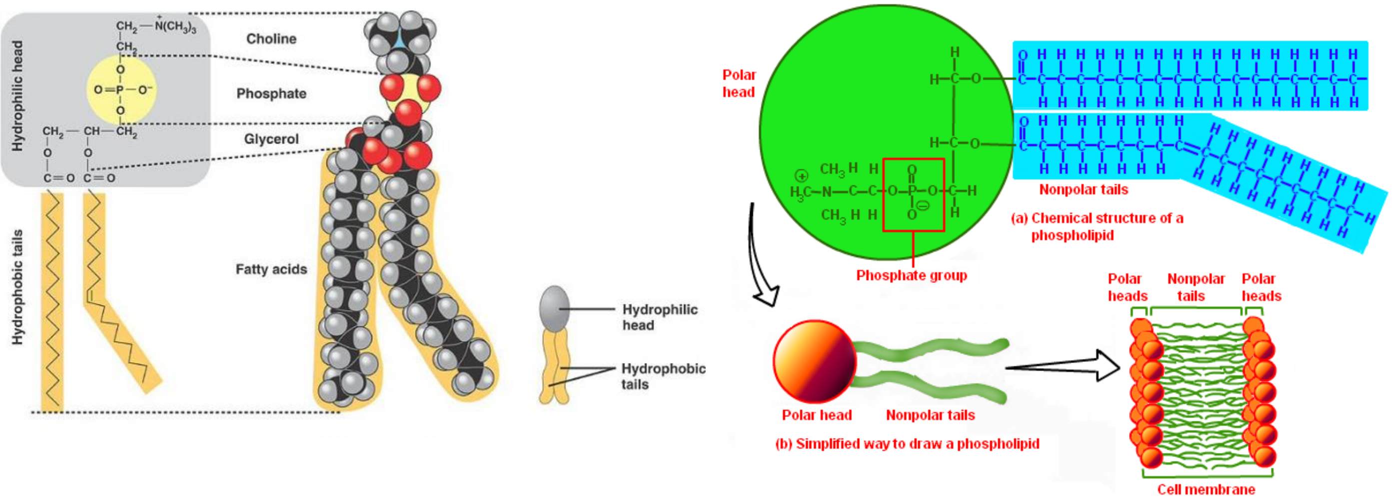


Figure 33. atoms and phosphate ions assemble to build a phospholipid membrane mostly by covalent bonding.

Hydrocarbon molecules and phosphate ions assemble to build a phospholipid macro molecule. The phospholipid membrane is a molecular bilayer formed by the self-assembly of lipid to lipid (tail) attractions combined with phosphate to phosphate (head) attractions in the presence of water. This assemblage is the basis of the biological cell membrane.

Ion, Metabolite, Energy Transport Across Cell Membrane

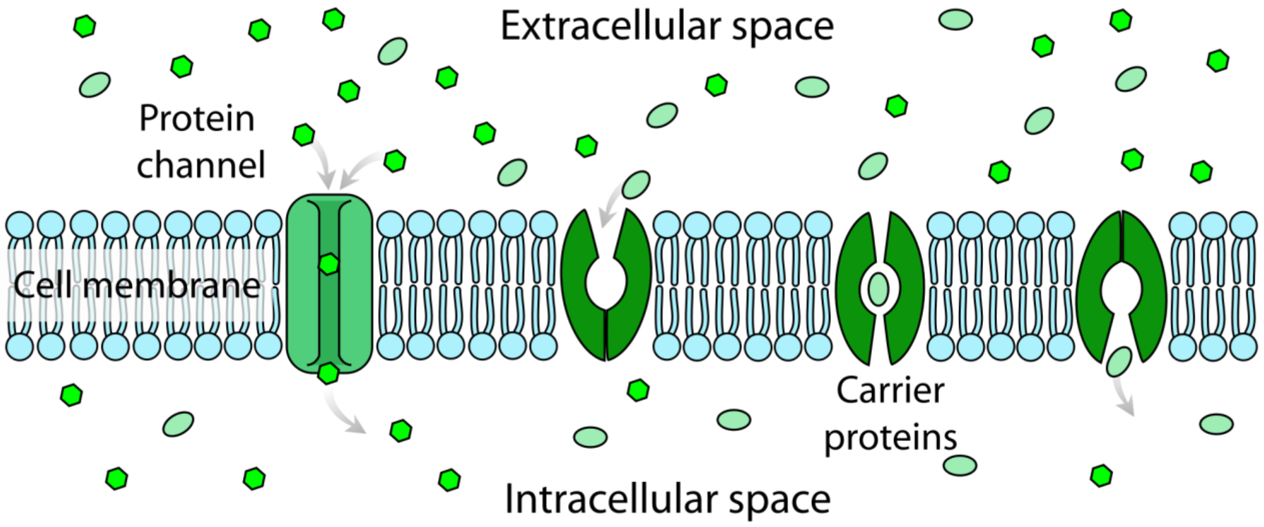


Figure 34. How ions, metabolites, energy (ATP) enter the cell across its membrane.

Waste will exit by a reverse but analogous mechanism. Pictures say more than words. The traffic across the membrane moves through gated protein lined channels as in Figure 34.

The Cell and its Membrane

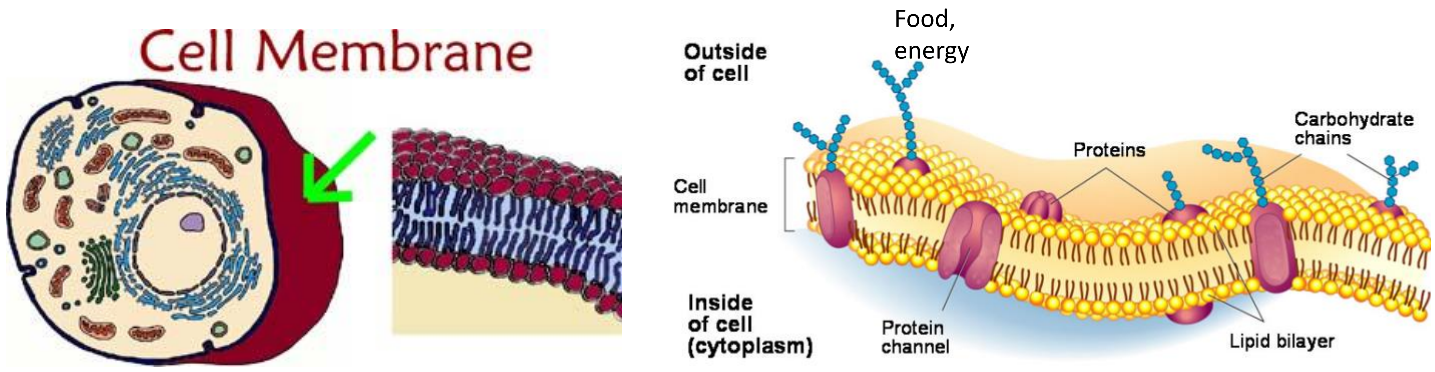


Figure 35. Illustration of the cell membrane from outside(left) and from inside(right).

The Electron Spin

The Zeeman effect is the experimental fact that the spectral lines split upon the application of a large uniform magnetic field. What began as a single spectral line without a magnetic field, could become multiple spectral lines in the presence of a magnetic field. This can happen even with Hydrogen which has only one electron. This splitting could be understood within the Bohr orbit description, as the expected magnetic force acting on the orbital motion of the electron. However, extra doubling of the Hydrogen spectral lines, the anomalous Zeeman effect, was a sign of trouble because of the single electron.@ The basic assumption is that in some form or another the conservation laws and the underlying concepts of energy and momentum, linear and angular, can be extended to the Bohr model.

This splitting of a single Hydrogen electron energy levels suggested to Wolfgang Pauli that electrons, in the absence of any interactions, occupy a ground state that is doubly degenerate – meaning it could be in one of two different states which have the same energy level and produce the same spectral line.

This degeneracy can be lifted by certain interactions, in particular with an applied magnetic field. This means that one of the degenerate states becomes the new ground state, and, for low magnetic fields, the other becomes the first excited state. This energy separation, $\mathbf{\Delta E = \mu B_0}$ is proportional to the magnetic field, $\mathbf{B_0}$ and as it can be measured spectroscopically, by also measuring the magnetic field, the proportionality factor μ can be experimentally determined. It measures the magnetic moment of what would become the electron spin. That magnetic moment means that the electron has an effective circulating charge as a source. This circulating charge represents the spin angular momentum **S**, an axial vector, that can align itself in either sense with any enforced direction, in particular that of an external magnetic field. A detailed and more rational discussion is in the upcoming section on magnetic resonance.

@ https://en.wikipedia.org/wiki/Zeeman_effect

In a magnetic field $\mathbf{B_0}$, the split between the two degenerate ground states is explained, somewhat arbitrarily, by the two values of the components of the angular momentum along the magnetic field $\mathbf{m_s = +/-1/2}$ **.**

This energy splitting, is proportional to the magnitude of $\mathbf{B_0}$, as illustrated in Figure 45. It is given by

$$
\begin{aligned}
\Delta E &= E(m_S = +1/2) - E(m_S = -1/2) \\
&= g_e\beta_e B_0/h \text{ (in Hz)} \\
&= g_e\beta_e B_0 \text{ (in ergs)} \\
&=\mu B_0
\end{aligned}
$$

where $\boldsymbol{\mu}$ is defined as the magnetic moment of the spin.

It is a magnetic analogue of the electric dipole, but the analogy is much less than perfect. In particular, a magnetic monopole, the magnetic analogue of an electric charge, has never been observed. This magnetic dipole moment is associated with the fundamental quantum property: the spin of elementary particles.[7]

The basic laws derived from experiments by Ampere, Oersted, Faraday and some of their contemporaries, where summarized by Maxwell in his equations, with some added subtleties. These subtleties were the actual predictions of the electromagnetic nature of light and of special relativity. The force law which is an integral part of the equations was independently proposed by Hendrik Lorentz at about the same time. These equations predict the existence of the electric and the magnetic force field acting on electrical charges. The electric field is omnipresent, whereas the magnetic field appears only when the charge moves. Of course, this distinction

---

[7] https://en.wikipedia.org/wiki/Magnetic_dipole

depends on the particular reference system (inertial, ignoring gravitation). Actually, this separation into two separate force field is artificial as demonstrated by Einstein when he unified the two separate vector fields into a single, tensor field in spacetime.

Two discoveries required modifications, even in vacuum, namely the photon and the electron spin. Often the magnetic field is defined by the force it exerts on a moving charged particle. Experiments in electrostatics show that a particle of charge ***q*** in an electric field **E** experiences a force **F** = ***q*****E**. Other experiments show that a charged particle experiences a force proportional to its relative velocity to a current-carrying wire. The velocity dependent portion can be separated such that the force on the particle satisfies the Lorentz force law@

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

Here **v** is the particle's velocity, and **x** denotes the cross product. (See#1Appendix B)

The vector **B** is termed the magnetic field, and it is defined as the vector field necessary to make the Lorentz force law correctly describe the motion of a charged particle. This definition suggests a way to determine **B.** "Measure the direction and magnitude of the vector **B** at such and such a place," calls for the following operations: Take a particle of known charge ***q*****.** Measure the force on ***q*** at rest, to determine **E.** Then measure the force on the particle when its velocity is **v;** repeat with **v** in some other direction. Now find a **B** that makes the Lorentz force law fit all these results—that is the magnetic field at the place in question. (See#1Appendix B)

@ https://en.wikipedia.org/wiki/Lorentz_force

The experimentally measured magnetic field around any small magnet looks very much like the electric field of an electric dipole, calculated from the Maxwell equations in the static approximation. The experimentally measured magnetic field of a small current loop looks very much the same, as does the calculated magnetic field from an electric charge moving with constant angular velocity around a circle. These analogies originated the idea of associating these magnetic fields with a source magnetic dipole moment. The analogues of the electric charges would be magnetic charges associated with the poles of the magnet. This is an imaginary concept, as there is no experimental evidence for a magnetic monopole.

According to Maxwell the magnetic field from a current loop is proportional to the charge and to the velocity. The magnetic dipole moment $\boldsymbol{\mu}$ for the current loop can then be arbitrarily defined as

$$\boldsymbol{\mu}=\pi r^2 q.$$

Thus, the magnetic moment is well defined for a current loop, as well as for a single charge moving along a circular path. For the field of a permanent magnet however this is an empirical definition as ferromagnetism cannot be derived from Maxwell. If we try to associate it with current loops the sign, the direction of the flow, comes out wrong. Nevertheless. thinking about both ferro and para magnetism in terms of anomalous current loops facilitates the development of very useful models, so it is accepted. It is another quirk in quantum physics and remains a mystery to solve.

The magnetic moment (or magnetic dipole moment) is a vector. $\boldsymbol{\mu}$. The torque on a current loop can then be written as $\boldsymbol{\tau} = \boldsymbol{\mu} \times \mathbf{B}$, x representing a vector product. The magnitude of the torque is

$$\tau = \mu \, B \sin(\theta),$$ θ being the angle between the two vectors.

For a current loop as well as the charge moving in a circle the magnitude of its magnetic moment is equal to its area multiplied by the current value. This a consequence of Maxwell's equations and the Lorentz force law.&

An externally applied magnetic field **B** exerts a torque on a small current loop $\tau = \mu \, B \sin(\theta)$, where μ is the magnitude of the dipole magnetic moment of the loop. By measuring τ, B, and θ, we can determine the magnitude of the dipole moment. Similarly, we can measure the dipole moment of the circulating charge. for these two cases the result of these measurements can be checked against the values given by theory and they match. This is just another way to verify the validity of Maxwell’s equations. And so, we proceed.

However, if we do the same measurements for a small bar magnet, as ferromagnetism is not covered by Maxwell’s model, what we obtain is an effective magnetic moment for which there are no equations to verify its validity. So, we accept this value and assume that the electrons in the ferromagnet are the same electrons, but the electron spin magnetism does not conform to the rules of magnetism due to the motion of electric charges. And so, we proceed.

The Zeeman effect tells us that atomic energy levels split upon the action of a magnetic field. In particular for the simplest case of the **1s** state. For

---

& https://en.wikipedia.org/wiki/Maxwell%27s_equations non-spherically symmetric states it combines with the diamagnetic moments associated with their orbital angular momenta of the Bohr orbits. However, the energy levels of the hydrogen atom corresponding to spherically symmetric s states still split. The magnetic field dependence of this splitting yields the value of the electron spin magnetic moment. It also yields its sign which is opposite to that for orbital magnetization.

All alkali atoms have a single electron in their outermost shell. Therefore, their energy level schemes serve as good facsimiles for the energy level diagram of the **1s** electron in hydrogen. So, we are dealing with the states of a single electron undergoing orbital motion. The conservation of angular momentum requires that the total angular momentum **J** be a constant. (see#4Appendix B) If electron spin carries an angular momentum **S**, the total angular momentum will be given by the vector equation **J=L+S. L** is the orbital angular momentum. With the added quantum assumption that the components of any angular momentum along any axis can only have discrete equidistant values that are integer multiples of a momentum quantum =**h/2π**, we can work out, kind of, how to add vector quantities such as momenta. The vector quantity, orbital angular momentum is defined by the allowed values for its components along the axis, **m.** Each possible value of **μ**, **mh/2π** specifies an allowed direction of the angular momentum **J**, and, the range of its possible values, which in turn, specifies its magnitude,

m = -l, -l+1,-l+2,…..0,1,2,3…0,1,2,3…l.

Adding two arbitrary oriented vectors means that the range of the vector sum will be the union without overlap, of the two separate ranges **($-l_1-l_2,-l_1-l_2+1,-l_1-l_2+2,\ldots 0,1,2,3\ldots l_1+l_2$).**

If the range of the spin is **(-1/2,1/2)** and the range of **L** is **(-2,-1,0,1,2)** then the range of **J=L+S** will be **(-2.5,-1.5-.5,.5,1.5,2.5).**

What is an Electron ?

An electron is a particle endowed with a unit negative charge e and a unit mass **m**, that inhabits the Bohr orbits, or whatever they might turn out to be in improved models of atomic structure. In Quantum Mechanics we think of the electron as being in a particular state defined by the three Schrödinger quantum numbers. Pauli determined that we could account for the anomalous Zeeman effect, the splitting of the hydrogen levels in a magnetic field, by adding a fourth quantum number he called the spin having two possible values. Each state would be described by four quantum variables then, and his exclusion principle, rules that no two electrons could occupy the same state in a stable configuration. He thus introduced the notion of the spin quantum number as one of the defining variables of quantum state. The others are the radial number, or the principal quantum number, and the two angular quantum numbers. No two electrons in an atom can have the same four values.

The spin having two possible values doubled the number of states available and removed the degeneracy.

The exclusion principle applies to fermions – which an electron is one example of. Fermions have ½ integer spins, **s = +1/2 or -1/2.** Any combination of particles resulting in this kind of spin is a fermion and falls under the exclusion principle. Bosons are particles, or particle combinations, which have integer spin. These are not subject to the Pauli exclusion principle.

How the successive shells are filled up in the buildup process depends on how many have been already filled. That is a consequence of the exclusion principle. The **LS** vs **JJ** methods of filling the shells@ takes into account the exclusion principle.

In a filled shell all electron spins are paired up, so the total spin will be **S** = **0** whereas for a shell with one electron it will be **S** = **1/2**. So it will be for an almost full shell with seven electrons as in the outer shell of Chlorine, six paired up electrons plus one unpaired.

The exclusion principle and Permutation Invariance, Bosons and Fermions: a reminder of the new rules

For any other partially filled shell new rules have to be adopted: they are the Hund Rules. For the shells that are further from the positive nucleus their individual states are constructed in terms of their total orbital angular momentum **L=Σl** where **l** are the angular momenta of individual electrons, and the total spin angular momentum **Σs=S.** These vector quantities are calculated first. Then the total angular momentum characterizing the multielectron atom, **J**, will be the vector sum of **L** and **S: J=L+S.** This is the **LS** coupling scheme.

For electrons in inner shells the coupling between the individual l and s is expected to dominate over any coupling between different electrons, so the individual j=l+s are calculated first, and the total angular momentum will be J = **Σ**(j). This is the **jj** coupling scheme. For electrons in inner shells the calculating rules will be different from those in outer shells. The rules for calculating the total orbital angular momentum are derived from

---

@ https://www.researchgate.net/publication/265469672_Atomic_electronic_states_the_L-S_and_j-j_coupling_schemes_and_their_correlation the known classical rules of adding angular momentum vectors. For the spin, with no classical analogue new rules were devised, the Hund's rules. Subsequent predictions then agreed with spectroscopic experimental results.

As electrons are added, one by one, to an empty shell their spins add for the first four electrons in the following sequence
s = **1/2, 1, 3/2, 2** then subtract **s=3/2, 1, 1/2, 0** for the full shell, the maximum spin being achieved for a half full shell.

The exclusion rule is satisfied by choosing different orbital states. Then, the total spin is added to the total angular momentum to obtain the total shell momentum **J=Σ(j), j=l+s.** This is vector addition, meaning union of ranges for magnetic quantum numbers **m.**

As a consequence, for a shell containing an odd number of electrons, all energy levels are doubly degenerate in the absence of a magnetic field. These rules play a fundamental role in explaining the magnetic properties of the long period elements in the periodic table.

The bond formation mechanisms discussed in connection with water and hydrogen, allow atoms to assemble into progressively larger and larger molecules and eventually to form condensed phase materials, the liquids and solids. We already discussed water, a prime example. Hund's rule may have a role in the explanation of ferromagnetism in the formation of domains with parallel oriented electron spins.

The mathematical basis for the exclusion principle, and the fundamental nature of the different behavior of Fermions and Bosons.

States are understood as defined for entities more generally than just as elementary particles. A nucleus made up by a large number of protons and neutrons can be viewed as an elementary particle in the context of defining quantum mechanical states.

Let us consider a potential two-electron system, say the Helium atom. Let us start with a helium nucleus of charge **+2** with no electrons. In classical physics the concept of states does not exist.

To have electrical neutrality of the atom we see that the helium atom should consist of the nucleus and two electrons. Let us fix our attention on these two electrons, call them **x** and **y.** Assume that they occupy the two states we call **g** and **u.** The product of the two states, or any of their linear combinations, more specifically their sum and difference, also give possible starting states.

Let us pick these, the sum and difference as the starting states and put the two electrons into them:

$$S + (x,y) = g(x) * u(y) + g(y) * u(s)$$

and

$$S - (x,y) = g(x) * u(y) - g(y) * u(s)$$

and let us see what happens when we want to make **x = y:**

$S + (x,x)$ survives as a positive sum,
$S - (x,x) = g(x)*u(x) - u(x)*g(x)$
$= 0 =$ zero identically.

So, if we choose the **S-** rule for electrons, or more generally fermions, we see that we have a rule that represents the Pauli exclusion principle. Results which become zero represent states that can’t exist. For a pair of fermions the starting wavefunction should be **S-**, a difference, the Pauli exclusion principle will be satisfied automatically.

This simple rule for pair permutation can be readily extended to choose the correct alternating sum for permutations of three, four, and five product states as all possible permutations can be reached by successive binary permutations. Going beyond to the general case get us into higher mathematical arguments. As permutations combine through multiplication, any individual state of value zero will invalidate an entire permutation.

For bosons we combine individual states with addition, so we do not have these invalid permutations, and they are not subject to the Pauli exclusion principle. Two or more bosons can populate the same state. For bosons we combine individual states with addition, so we do not have these invalid permutations, and they are not subject to the Pauli exclusion principle. Two or more bosons can populate the same state.

Section 6. Metals, semiconductors and band structure.

Consider an atom, say copper, that has a single electron in its escape shell. A pair of such atoms can covalently bond forming a molecule. A third copper atom can bond forming, by superposition, a three-electron state that extends over the volume of the triatomic molecule. This process can be extended to larger and larger groups of copper atoms. If this is done at sufficiently low temperature and in the absence of a significant number of other atoms or other molecules, a piece of copper is formed. This is not the practical way to make solid copper. The art of metallurgical engineering takes care of that.

Why the requirement of low temperature? The random thermal motion of the approaching atoms prevents the formation of stable bonds. The thermal annealing technique is an exception and deserves a separate discussion.

Why the requirement of absence of other atoms or molecules? The foreign particles would bond to the growing copper piece as impurities. Of course, if you want to create a solid piece of copper oxide you could do that by controlling the mixture of approaching atoms. This idealized process has a practical realization when particular materials are deposited by chemical vapor deposition, or the various variants of beam assisted deposition techniques. This is the only way to obtain materials in the form of uniform thin films, as required by miniaturization.
It turns out that temperature, when properly controlled has a useful role in these material formation processes. A certain amount of thermal motion will encourage the formation of the most stable electronic state of the whole assembly. If you recall the discussion of the Hydrogen molecule, these bonds are formed by superposition of directed asymmetric p states of neighboring atoms. Therefore, the ideal, the most stable state should be

that of a regular perfect crystal where the atomic cores should form a regular lattice. Thus, the origin of the crystalline state of matter is quantum mechanical.

Let us, for the moment, contemplate this situation in a classical manner and we will return to the quantum picture of band structure later in connection with silicon.

This collective state is a superposition of orbits extending through the whole copper crystal. They are very close in energy and the electron can easily jump from orbit to orbit. This is equivalent to the picture of classical free electrons moving in random trajectories due to the thermal motions of the atomic cores. The jumps occur in every possible direction so there is no predominant direction and no place for charge to accumulate.

Consider now a copper wire with a potential difference applied to it by a battery with a very high internal resistance. That means that the electrons will feel a force that favors jumps towards the positive terminal where the electrons will accumulate, stopped by the high internal resistance of the battery--UNLESS they are provided the escape route of a closed circuit, that is a battery with a low internal resistance. This is the fundamental nature of the process of electric conduction in metals.

Heat conduction by electrons via inter orbital jumps is happening at the same time. The random jumps, whether by electrons or by the atoms themselves are what we call heat.

When the frequency of these jumps is higher the temperature will be higher. These random jumps represent thermal energy, or heat. An inhomogeneous distribution of the jump density means temperature gradients -- therefore, heat conduction will take place. However, if the heat conduction is impeded by interposing a thermally insulating layer, a heat gradient will be preserved. This is the mechanism of thermoelectric cooling. In the same manner an inhomogeneous distribution of temperature implies electric charge gradients, therefore, voltages and mechanisms for thermostats.

In the above discussion we focused on collective states formed by the outer electrons of atoms that carry a single electron in the outer shell. Similar considerations can be applied to atoms with two or three such electrons except that now we would be dealing with a set of two or three such states. We call this a band, in particular the conduction band. For four, five, six, seven electrons we have to call it a valence band. This is because the electron quantum interaction enforces the filling of almost full shells locally. Collective states will support electron distributions having maxima close to the source atoms. The collective orbital image looks more like a set of Bohr orbitals around each atom. The jump between orbitals looks like a jump of the electron from one atom to another atom. Or perhaps, from an atom to this multiple atom collective state in the conduction band.

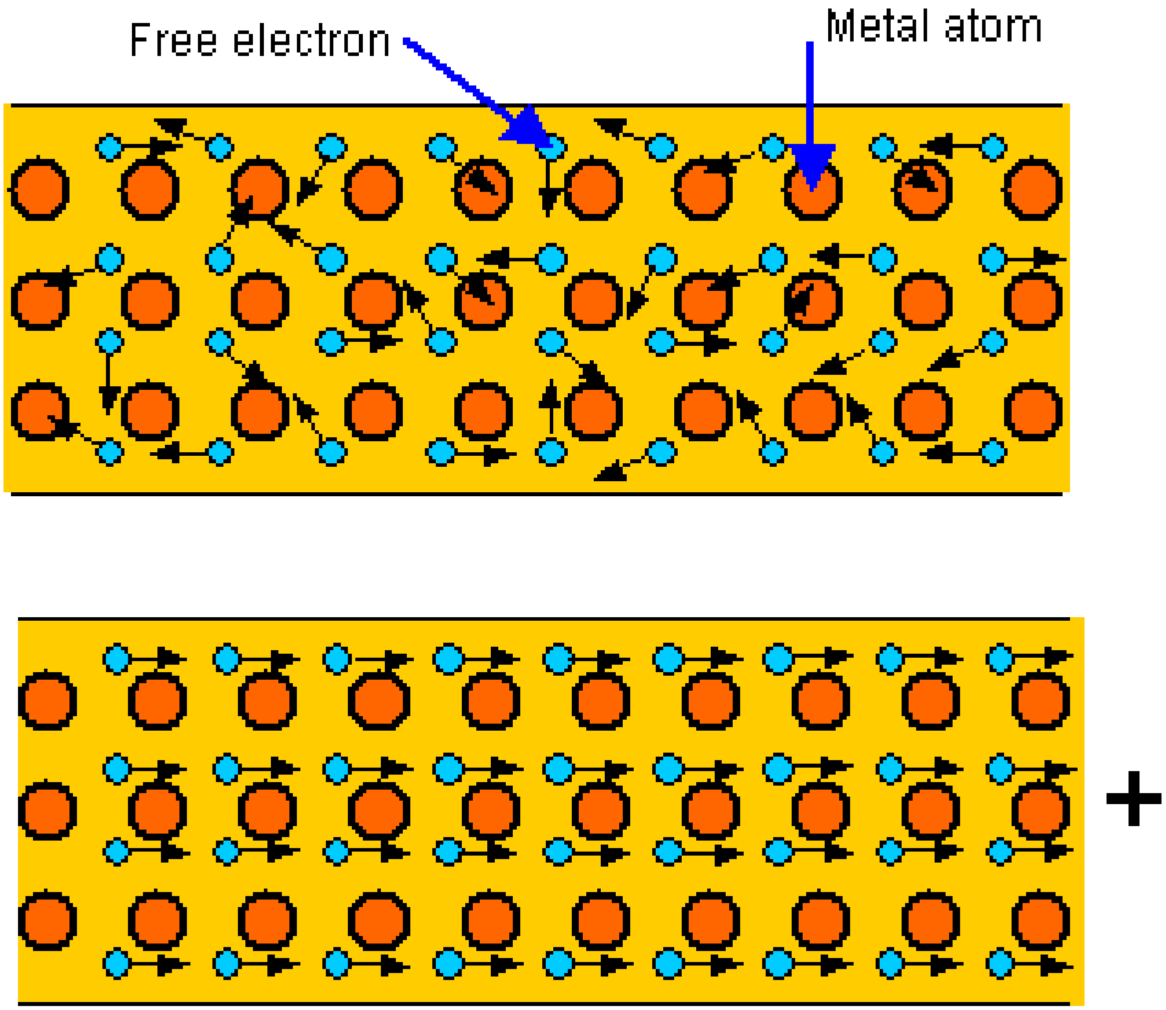


Figure 36. Conduction electrons in a metal wire without-applied-voltage (top), and with-applied-voltage(bottom).

This figure epitomizes the Lorentz free electron model of metallic conductivity. It works remarkably well until we try to explain the experimental values of the mass of these "free" electrons which do generally differ from the mass of really free electrons, as measured by Millikan. These electrons are really moving in the collective attractive positive potential of the net charges of the nuclei partially balanced by the more tightly bound core electrons of the individual atoms.

The metallic bond, conductivity and electrical resistance

If we locate two **Na** atoms near each other, from the point of view of the single electron in the outer shell, it is analogous to a hydrogen molecule where the two outer electrons with opposite spins can form an attractive two electron orbital. Indeed, $\mathbf{Na_2}$ molecules as well as $\mathbf{Li_2}$**,** $\mathbf{K_2}$**…** will form. In the same manner, alkali atoms can be further assembled in close proximity. Their outer electrons will form extended multi electron orbitals that hold the whole metal together. Furthermore, upon application of an electrical potential gradient, these states can sustain an electrical current flow. This is true as long as the ions form a perfectly regular and stationary lattice. The presence of defects, impurity atoms and thermal giggling causes resistance to this flow, which will be overcome by a supply of energy from the battery. Therefore, the total resistance will have a fixed component due to defects and impurities and a temperature increasing component due to thermal giggling, this being the basic character of metals. The detailed mechanism whereby the giggling vibrations as well as the impurities impose resistance to electric flow is also quantum mechanical in nature. Under certain special circumstances these conduction electron orbitals may combine into very special coherent orbitals that can overcome the resistance to flow. Once it is started and the circuit is closed it will continue without the battery. This is superconductivity and it requires a combination of quantum mechanics with electrodynamics to be properly described.

The electronic states in closed shells will also form collective states in any atomic assembly, as the atoms get sufficiently close. You could say that the hydrogen atom has two possible "bands", the bonding and antibonding band, which is the "conduction" band of the molecule.

The Schrodinger equation has distinct terms representing the contributions to the total energy of the electron from its kinetic energy due to its motion, and the potential energy due to the electrical attraction between the negative electron and the positive nuclear cores. (See #3 Appendix B.)

The kinetic energy is represented by a differential operator combining partial derivatives of some function of the spatial coordinates of a point in space. This function is yet to be defined, as we are in mystery territory. The potential energy is represented by scalar function of the coordinates and operates as a multiplicative factor.

For time independent, stationary states, the Schrodinger equation has the general form

$$\mathbf{(K+V)\ f(x,y,z) = Ef(x,y,z)}$$

K is the kinetic energy operator, **V** the potential energy, a scalar function of the coordinates and **E** a multiplicative constant. The process of solving the equation consist in determining a combination of a function **f(x,y,z)** and a constant **E** such that **f** satisfies prescribed boundary conditions. This is an eigenvalue problem. As you may have guessed, **E** is the energy of the electronic state corresponding to the electron distribution represented by **f.** (See #3 Appendix B.)

It turns out that there may be more than one solution $\mathbf{f_1, f_2, f_3}$ **…** say, for the same value of **E**. More importantly, as the **T** operator is linear any linear combination $\mathbf{af_1+bf_2+cf_3}$ will also be a solution. We discussed the geometric features of the solutions **f** for the one electron system of the Hydrogen atom, as orbitals: see Figure 23. The eigenfunction@ **f** is the

---

@ https://chem.libretexts.org/Bookshelves/Physical_and_Theoretical_Chemistry_Textbook_Maps/Book%3A_Quantum_States_of_Atoms_and_Molecules_(Zielinksi_et_al.)/03%3A_The_Schr%

electron distribution function in a state of energy **E.** It is interpreted as a probability distribution function. It is a complex scalar whose squared absolute value gives the probability that a measurement will find an electron with energy **E,** at point **(x,y.z).**

Let us now start to build up the collective states of the crystal in the spirit of the buildup of the shell model for a single atom. First, we start only with the nuclei of, say, Lithium, that by magic are placed at the crystalline sites, and introduce the first electron. It will feel a periodic potential with huge attractive force in the vicinity of every nucleus. It gets stuck near the first nucleus that it approaches. As more and more electrons are introduced, they will balance out the localized attractive force of individual nuclei toward a more delocalized force that will tend to stabilize each new electron in a more and more extended state.

For electrons in the innermost shells these attractive peaks are still large enough to call them atomic core electrons. As we move out from the cores these electronic states become more and more delocalized and for accounting purposes, we group them into bands (rather than shells) labeled by the names of their atomic shells of origin. This intuitive, rather irrational description indicates how the Schrodinger equation would be used to obtain the eigenfunctions for band electron states. (Felix Bloch circa 1930)

The Schrodinger equation really was a gimmick initially – a gimmick that actually worked out. So, let us try another one. Consider a fully assembled crystal with all its electrons occupying states described by Schrodinger,

C3%B6dinger_Equation/3.04%3A_Operators%2C_Eigenfunctions%2C_Eigenvalues%2C_and_Eigenstates like wavefunctions that are to be determined by solving a Schrodinger equation in which the potential was replaced by a trial @periodic potential, invented, with a relatively small number of adjustable parameters. Use the solutions to predict experimental results. Then choose the parameters that give results that most accurately fit the experimental data. Repeat for as large a set of parameters as possible and choose the parameter combination that gives the closest match to the experiment. If you have the guts, call this the real potential. It may not be, but often it is a good start. Doesn’t this feel like AI? Just do it on the laptop. Variants of this process, with complications and difficulties, do work. A couple of Nobel prizes testify to this. What justifies them is not the basic idea, which is obvious and pedestrian, but the solving of the complications and difficulties.

Assuming that we know the wavefunctions and energy values of these states, we have a description of the band structure. So, let us proceed. Figure 37 below depicts the three possible relative positioning of the two highest bands. The top band comes from the partially filled outer shells of the atoms and ls the conduction band. The next below is the valence band and the figure illustrates three possible situations. The fermi level or fermi energy is a parameter of the FERMI-DIRAC distribution function. It marks the energy states that have equal probability of being occupied or not.

$$\mathbf{N(E) = 1/((Exp(E\text{-}E_F)/kT))+1),}$$

@ https://en.wikipedia.org/wiki/Bloch_wave

E_F is the **Fermi level**. The number of electrons occupying states below it equals the number of electrons in levels above it.

The **Boltzmann constant k** is a unit conversion factor that converts Kelvin temperature degrees into energy units.

When the two outer bands overlap, chances are that there will be filled and empty states very close in energy. That means that this sample is a good conductor, probably a metal. If there is no overlap but the band separation is small compared to **kT,** there can still be some conductivity and we have a semiconductor. **kT** measures the average energy that is carried by the random fluctuations present in any dynamical process. This concept of thermal fluctuations was developed by Maxwell and Boltzmann in the kinetic theory of gases, to explain the gas laws of Boyle and Gay-Loussac.

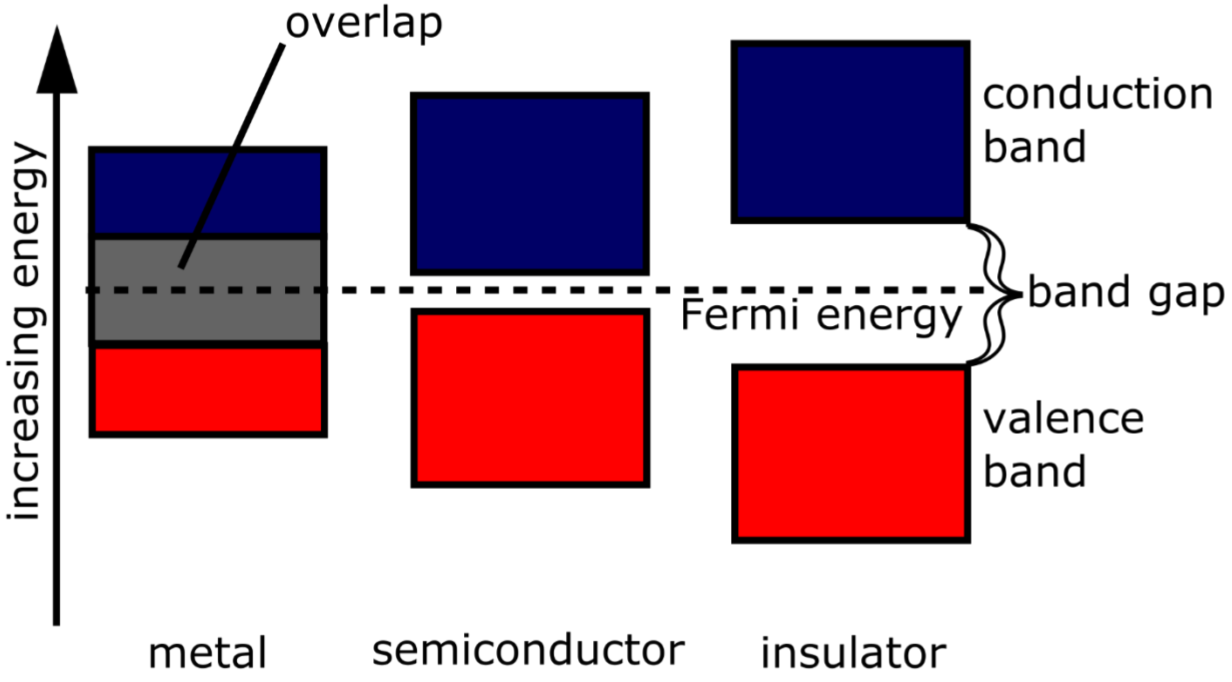


Figure 37. illustrates the positions of the 2 outermost bands in metals, semiconductors and insulators.

In a metal these bands merge so they will have both occupied and empty states. Thus, the electrons can jump under external actions. These can be applied voltages, the thermally excited elastic vibrations of the atoms, or radiation. These all act simultaneously, however, if sufficiently weak, they can be considered one at a time. If a voltage is applied to a copper wire all electrons will tend to occupy states that imply motion along the wire whereas the vibrations favor equally states that have momenta in all directions thus, providing resistance to the current flow.

At this point it is opportune to introduce the concept of charge carrier mobility μ which is especially useful when we talk about the average current carried by a jumping charge. The current carried by charge **q** moving with velocity **v** is **I = qv.** What actually contributes to the flow is the average velocity over the set of jumping charges. For electrons all with the same charge we only have to average the velocity. In the presence of an electric field **E,** this average velocity will be in its direction. Then the mobility is defined by the equation:

$$\mathbf{I = q\mu E,\ \ so\ v = \mu E}$$

where **v** is then the average velocity.

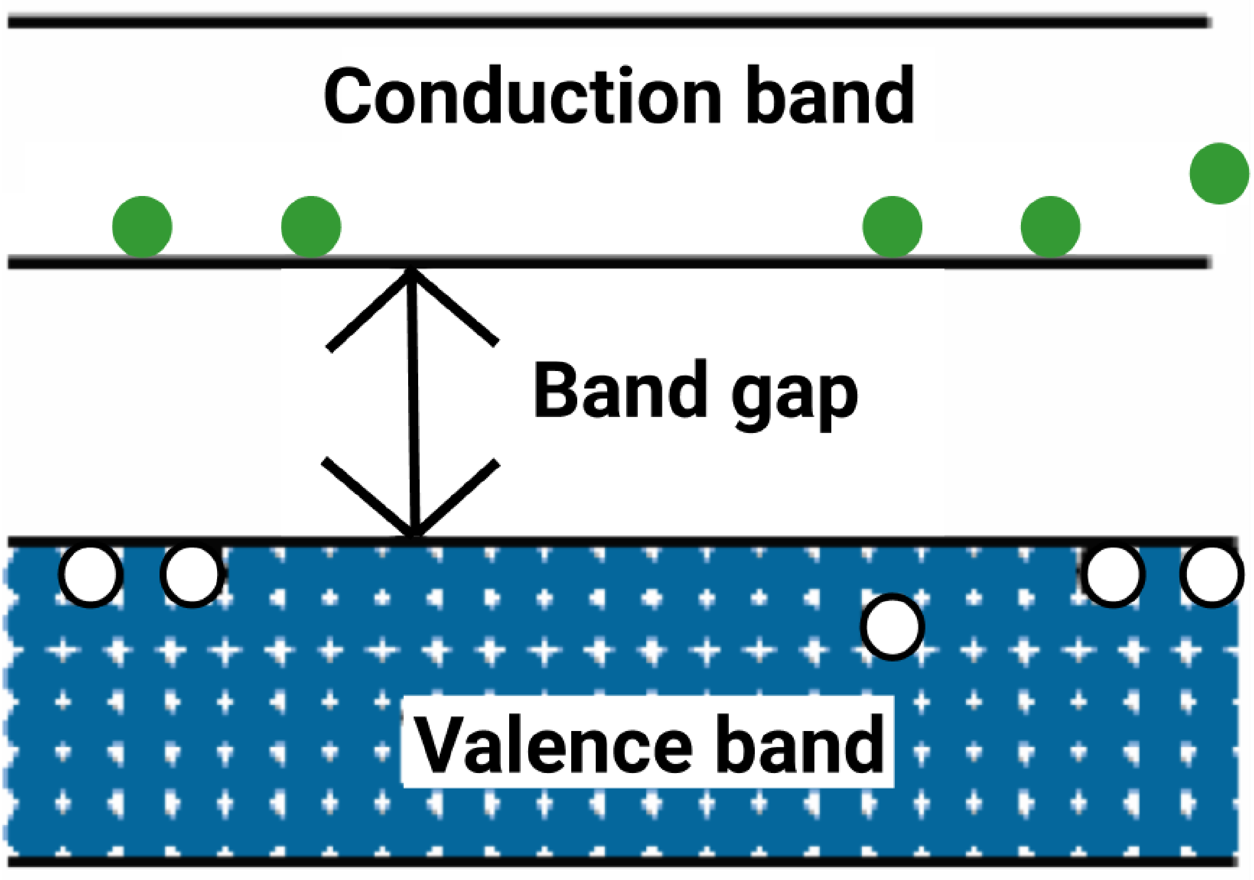


Figure 38. The semiconductor band gap with thermally excited electrons and holes.

If an electron jumps from the valence band to the conduction band it leaves a positive hole behind, which the surrounding electrons try to fill. This means that this hole moves as a positive charge. The charge magnitude is that of the electron but not the mobility. One can introduce extra holes into the valence band as well as extra electrons into the conduction band, but that involves chemistry. It also introduces extra localized states into the bandgap but if they are not too numerous, the thing still works. Electrons and Holes from thermal fluctuations are in equal numbers. However, they have different mobility. So even an intrinsic semiconductor can carry a current at sufficiently high temperatures.

To introduce extra holes in the valence band: replace some silicon by gallium. To introduce extra electrons in the conduction band: replace some silicon by arsenic.

Replace only some, so we can still identify the sample as a Silicon crystal. If you replace every alternate Silicon by a Gallium atom, and those that remain by Arsenic, the Silicon crystal is transformed into a Gallium Arsenide crystal. **GaAs** is another perfect semiconductor. **GaAs** can also be endowed with extra holes and electrons by replacing some **Ga** by say **Mg** for holes and some **As** by **Se** for electrons.

Section 7. The p-n junction and semiconductor devices[$]

The **p-n** junction was the basic structure of all semiconductor devices and it is the simplest structure to examine the physical processes of their workings. The field effect structure common today, although proposed simultaneously, required advances in thin film technology not yet available. The field effect was mentioned in the 1950 book by William Shockley, *Electrons and Holes in Semiconductors* with applications to transistor electronics.

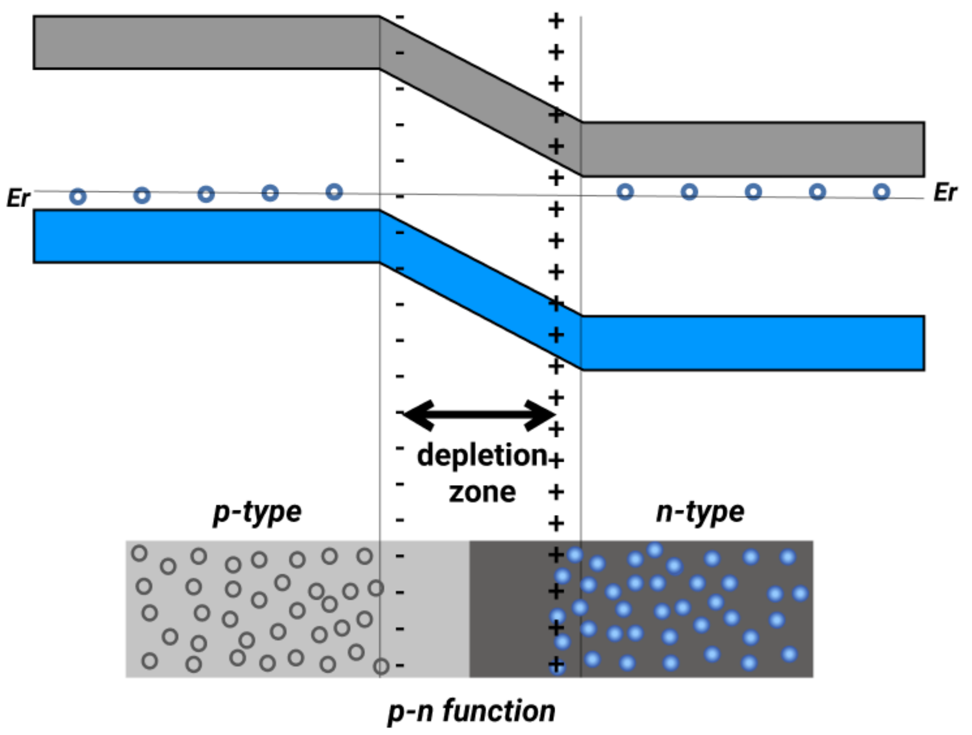


Figure 39. A p-n junction.

[$] https://www.watelectronics.com/difference-between-npn-and-pnp-transistor/, https://en.wikipedia.org/wiki/Electrons_and_Holes_in_Semiconductors_with_Applications_to_Transistor_Electronics

Left, Extra Holes, **Ga** doped **p** type.
Center, Intrinsic,
Right, Extra Electrons, **As** (arsenic) doped **n** type.

Two adjacent regions oppositely doped, and properly annealed, establish, by electron and hole diffusion, an intrinsic, undoped boundary zone between oppositely charged (doped) regions. This structure is called a **p-n junction.** After the diffusion dies out this thin boundary region will support a built-in potential gradient that will locate itself in this region where donors and acceptors are either absent or compensated.

This built-in potential gradient opposes electrical current flowing in one direction, while favoring it in the opposite direction. This is the basic structure incorporated into the silicon rectifier.

This potential barrier is built up as the excess electrons on the **n** side (with **P** or **As**) tend to migrate to the **p** side to fill the holes left by the replacement of **Si** by **Al** or **Ga** atoms. This is quantum mechanics in operation, as discussed in connection with the electron shells and the periodic table. This migration builds up a net dipole charge across the **n/p boundary** implying that there will be a net electric potential difference built in between the two sides of the boundary. Indeed, this is the case and it explains the asymmetry of the current response of such a structure when a voltage is applied across the junction. It is a current rectifier.

Internal Potential Barrier = Asymmetric Current Response

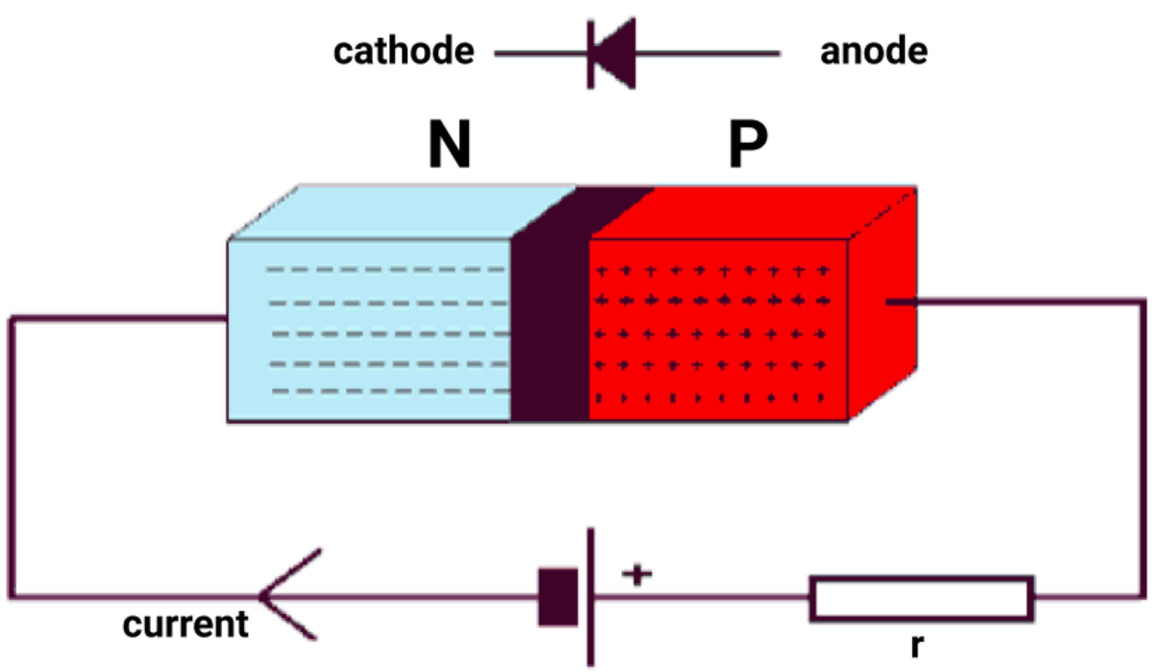
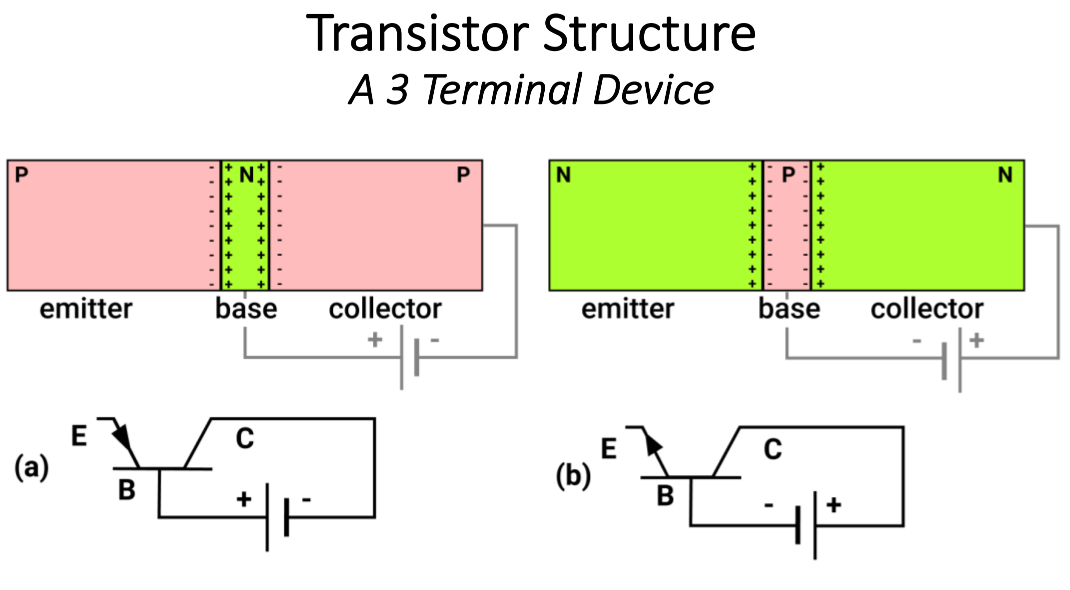


Figure 40. Current rectification in a p-n junction, easy current direction, electrons move from cathode to anode, holes move in the opposite direction.



PNP **NPN**

Figure 41. schematics of PNP and NPN transistors.

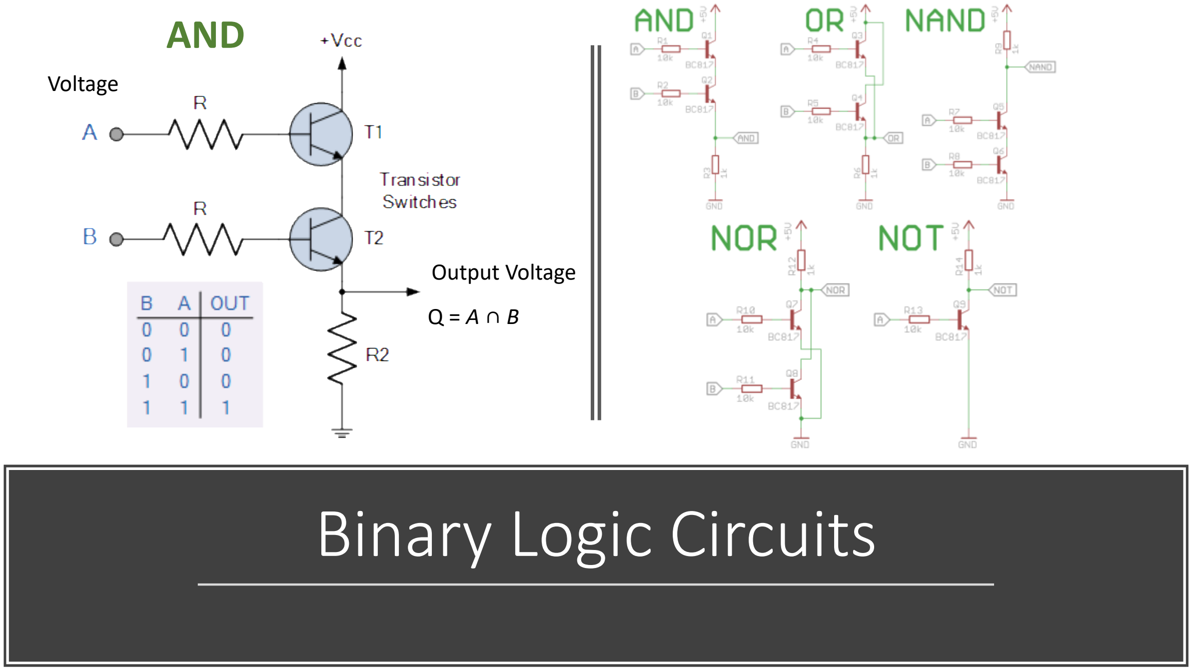
The transistor structure consists of two **p-n** junctions juxtaposed in opposition as in figure 41 above. These two junctions define three zones of different doping in two complementary structures, **PNP** on the left, and **NPN** on the right. The operation of the transistor is based on the rectifying property of the **p-n** junction. When **NPN** is biased as in figure 41, i.e. collector positive, the bias counteracts the first barrier and electrons from the collector pour into the base. Whether they will continue into the emitter depends on the potential difference between the base and the emitter. So, we have built a voltage controlled current switch.
At this point is opportune to give a simple explanation of an FET, or field effect transistor.

On a bar of Phosphorous or Arsenic doped Silicon, with a small metallic contact apply a negative voltage. That will repel the conduction electrons from a small volume surrounding the contact What this does is establish a boundary between a conducting bulk and a carrier depleted insulating base region. With skillful execution we can build a transistor by this process using the thin film technologies developed in the years following the brute force creation of the first working transistors.

These devices, just described, are at the heart, as well as all other organs, of every digital device, from your digital wristwatch to your iPhone, to your laptop, to IBM's Watson supercomputer. Examples of elementary devices constructed using a few transistors are pictured in Figure 41.

A transistor is basically a three terminal device where a current I between emitter **E** and base **B, I(E,B)** can be controlled by a voltage S, applied between collector and base, **S(C,B)**. A continuous variation of **S(C,B)** can yield a corresponding continuous variation of **I(E,B)**. Thus, it is a voltage to current converter, or an amplifier or attenuator. It will also work as a controlled binary (two state) gate if a voltage **S(E,B)** is stepped between two fixed levels that allow, respectively, **I(C,B)** to step between two predetermined levels, and the sign of **I(C,B)** can be easily designed to be opposite to that of **S(E,B)** To achieve this we need to provide a path to the collector to base current, by inserting a resistor in the path between collector and base. The voltage drop across this resistor then becomes the output signal **S(C,B)** of this amplifier.

The implementation of such interconnected gates, by the millions on an area as small as a few square cm. of a **Si** wafer, makes today's super computation a reality but gives the capabilities to common gadgets such as the smart phones and especially relevant to many of us, the hearing aid.



B	A	OUT
0	0	0
0	1	0
1	0	0
1	1	1

Binary Logic Circuits

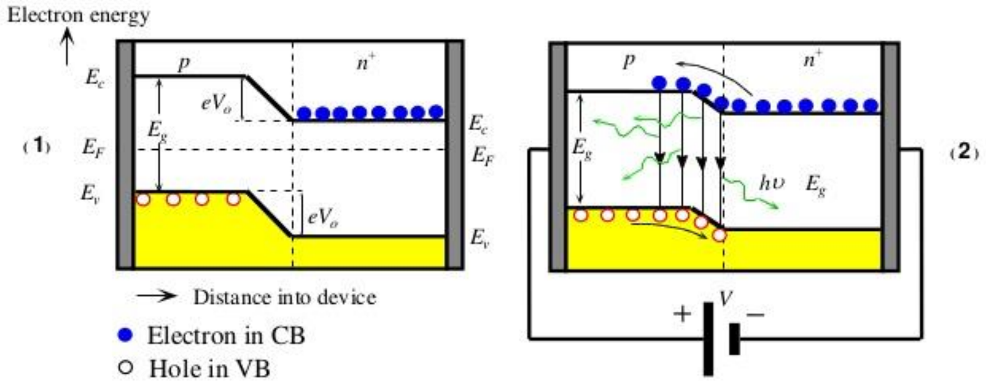
Figure 42. Binary logic gates implemented with transistors.@

@ https://www.instructables.com/id/Logic-Gates-with-NPN-transistors/

Section 8. Light Emitting Diodes, Diode Lasers and lasers in general

The p-n junction also led to various applications of semiconductor devices to optics, either as radiation detectors or as radiation generators.

In the following section our focus is on the more recently discovered applications as sources. The use of silicon point contact diodes as microwave detectors dates back to the time of WWII and the development of RADAR.



(1) The energy band diagram of *p-n*$^+$ (heavily *n*-type doped) junction without any bias. Built-in potential V_o prevents electrons from diffusing from *n*$^+$ to *p* side. (2) The applied bias reduces V_o and thereby allows electrons to diffuse, be injected, into the *p*-side.

Figure 43. The LED mechanisms

LEFT: The energy band diagram of a heavily n doped p-n junction of GaAs without applied bias. The built-in bias **V= E_v-E_c,** prevents electrons from diffusing to the **p** side.

RIGHT: The applied bias reduces **V= E_v-E_c** and thereby freely moves electrons into the higher energy **p** side**.**

As they fall back to the **E_v,** top of the valence band, in direct band gap semiconductors such as **GaAs,** they radiate the energy directly as light photons with very minimal loss to lattice vibrations. Thus, we have an efficient **LED**.

It turns out that when the above devices are implemented on **Si**, the process of photon-electron hole pair interaction are rather inefficient for use. However, an extreme extension of the doping process can be imagined yielding a compound crystalline semiconductor whereby alternate **Si** atoms in the cubic lattice are respectively replaced by **Ga** and **As** or other corresponding pairs of group **III** and **V** elements. This is not the way **GaAs** is actually produced. But the mechanism that holds such a mixed crystal stably in a regular lattice has a lot to do with the formation of closed atomic shells in molecular chemistry. Just as in the formation of **NaCl**'s by transferring the 5th electron from **As** towards the **Ga** so both now will have half-filled outer shells forming a shared 8 electron shell just like in Silicon. #

Then, introducing group II-VI dopants we can develop **n** and **p** regions. It turns out that the inefficiency of current to photon conversion in **Si** results from the large distance between the holes and the electrons. In **GaAs** the electron and the interacting hole, on the average, are in closer proximity. To bring them close requires energy and momentum which can only be supplied by giggling. In **Si**, as the donors and acceptors are present in low concentration and spread randomly, the average separation between an electron and a hole is large. A lot of the energy (phonons) is waisted to bring them together, little remaining to be radiated. In **GaAs** acceptors and

#Gallium arsenide is one such material and it has certain technical advantages over silicon—electrons race through its crystalline structure faster than they can move in silicon

donors are near neighbors to begin the process. Most of the energy is emitted as radiation.

The lattice thermal giggling is a superposition of lattice vibration waves (phonons) each carrying energy and momentum. Photons, as such, also carry both. This process, as well, requires conservation of both energy and momentum.

Thus, in compound semiconductors, like GaAs the current to Photon conversion process does not waste energy. This makes the corresponding LEDs much more efficient and makes the population inversion of electrons between the conduction and valence band required by diode laser emission possible.

In order to enhance the stimulation of the laser coherent emission the interaction opportunities of the generated photons with the laser structure should be increased. This can be done by cleaving the crystal at the two ends of the structure generating accurately parallel surfaces onto which almost, but not quite perfectly reflecting mirrors are deposited. More precisely, the output surface mirror should allow a fraction of the radiation to exit as the generated coherent radiation.

Energy Level Scheme for Laser Action

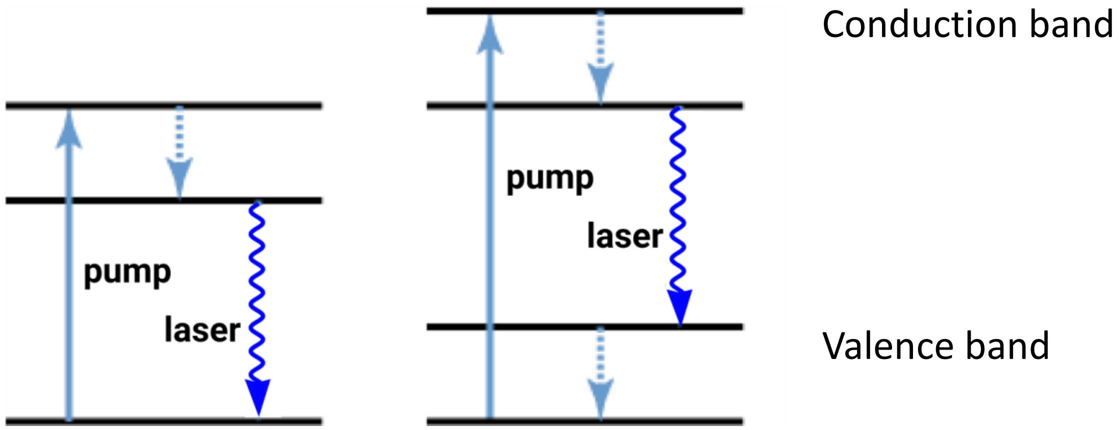


Figure 44. The general energy level scheme enabling laser action is on the left. On the right is a slightly more complicated scheme. See if you can figure it out before you read the next paragraph.

In the diode laser, the top level corresponds to the N side conduction band, on the left, specifically the bottom of the conduction band. On the right, anywhere in the conduction band but rapidly trickling down to the bottom, the third level, so that in both cases the emitted photons have the frequency corresponding to the band gap. Before the advent of the diode laser most masers and lasers used sharp energy levels of atoms or molecules, as represented in Figure 44.

The pump photon pushes an electron to the top level. The thermal giggling drops it to the intermediate level. Before the thermal giggling can kick the electron back to the ground level the photons get multiplied by stimulated emission which drops the electron back to ground before the giggling can do it. Thus, the intermediate level has to be of a state insensitive to the thermal giggling. The search for such a state system involves all the processes that we have been discussing.

In equilibrium at a temperature **T** there will be a smaller number of electrons in the conduction band than in the valence band. As **T** increases this population difference decreases. As the temperature increases this difference tends to zero. With a sufficiently powerful pump this population difference can be inverted causing laser emission by stimulated photon emission. In the case of the diode laser the pumping power comes from the battery as it supplies the reverse current against the built-in voltage across the barrier. There are purely optical lasers where the pump is supplied by a powerful optical source. This source can be a powerful other laser, as in the efficient Neodymium lasers or an extremely powerful broadband source as in the Ruby laser.

In order to enhance the stimulation of the laser coherent emission, the interaction opportunities of the generated photons with the laser structure can be multiplied. This can be done by cleaving the crystal at the two ends of the structure, generating accurately parallel surfaces onto which, almost but not quite perfectly reflecting mirrors are deposited. Every time the photons make a pass, provided the wavelength is resonant with the inter-mirror separation the radiation is resonantly enhanced, and the band width decreases, and the peak intensity increases. The enhancement is due to the process of stimulated emission.

Although the Maxwell’s equations imply the possibility of this effect, it is most conveniently presented in the quantum language of photons. A photon is emitted or absorbed by a particle when the particle makes a transition between two energy levels, upwards in absorption, downwards in emission. Thus, in order to emit a photon, it has to be in the upper state. It can be put there by dropping down from a higher level (a) or by interacting with an identical photon (b). Case (a) is spontaneous emission, case(b) is stimulated emission.

Suppose we could have an isolated atom and we flood it with the correct frequency photons. The first photon that hits it would put it into the upper level from which it would be reemitted, and on and on. This situation, to say the least, is not realistic. We always deal with a finite ensemble of identical atoms in the presence of a large ensemble of all kinds of atoms. If left alone it will come to thermal equilibrium where the number of atoms is always larger in the lower energy state. So, absorption will dominate. One way that this is done is by optical pumping. This is how the first practical maser was implemented. It was called a Maser, M for microwave, which is radiation just as light. The first Lasers operated this same manner except that the pumping radiation covered a wide frequency range, most of it wasted.

In thermal equilibrium the relation of populations of two levels separated by energy difference $\mathbf{E_{12} = E_2 - E_1}$ is given by the principle of detailed balance:

$$\mathbf{N(E_2)/N(E_1) = \exp(-(E_2-E_1)/kT),}$$

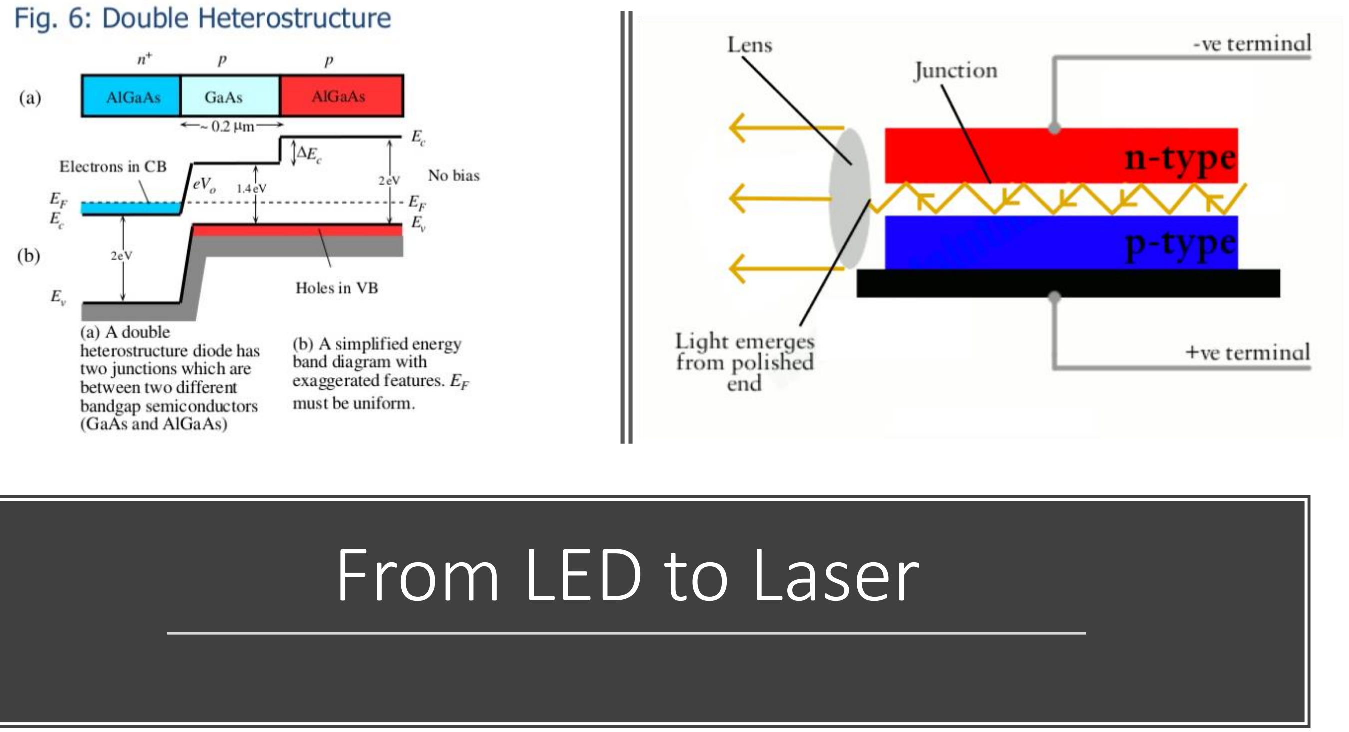
the Boltzmann distribution of classical physics

By setting the starting energy level $E_1=0$ to zero this equation can be integrated up to $E_2,=E$ to yield the dependence of the population distribution on the energy level E or the probability of the particle occupying a given state of energy E, the Bose-Einstein distribution;

$$\mathbf{f(E) = 1/(A\exp(E/kT))-1).}$$

A is the degeneracy number of the level. It is a quantum mechanical concept that has to be included to extend the Boltzmann distribution into quantum mechanics. This is again the cooking necessary to achieve progress in research. Classical mechanics deals with a continuum of states and deals with the continuous function of all possible trajectories of the same energy in terms of a continuous density function. All transitions are differential in nature. Sharp levels and finite step transitions are quantum mechanical. There are no energy gaps in classical physics, only ranges of low density of states.

The elementary particles can be bosons or fermions depending on their total spin being integer (Bosons) or half integer (Fermions).
An infinite number of particles can accumulate in a state, if they are Bosons. This is what is at the root of stimulated emission.



(a) A double heterostructure diode has two junctions which are between two different bandgap semiconductors (GaAs and AlGaAs)

(b) A simplified energy band diagram with exaggerated features. E_F must be uniform.

From LED to Laser

Figure 45. The structure of a laser diode that enhances the stimulated emission by multiple passes inside a resonant structure.

To conclude this section a somewhat personal, historical note is in order. The energy level diagram and the transition scheme on figure 45 were introduced in connection with the idea of the three level Maser in 1956, three years after the ammonia beam maser was introduced to the world by Charles H. Townes in 1953. As a freshly minted graduate student I was "encouraged" to attend a lecture by a freshly minted full professor as part of my beginning graduate education. In this lecture Nicholas Bloembergen@ introduced the idea of the three-level maser with the sentence: if you saturate the transition between 1 and 3 their populations are equal, and one of the two adjacent level pairs will be inverted favoring

@ https://www.nature.com/articles/550458a stimulated emission. Within a year, Bell lab scientists beat the author to the experimental punch. At this point I was just an observer, but this was a learning moment. This is the process whereby Theodore Maiman "invented" the laser in 1960 and is the basic mechanism of operation of the largest fraction of practical Lasers.

Section 9. More about Magnetism, Magnetic Moments and Nuclear Spins

After World War II the field of microwave spectroscopy flourished, based on the radar technology. In particular the work at Columbia University sparked research towards the development of electron spin resonance techniques which allowed the exploration of the anomalous Zeeman effect to confirm the presence of nuclear spin magnetic moments and the possibility of observing them by resonance methods. The difficulty of NMR (Nuclear Magnetic Resonance) is that nuclear moments are about a thousand times weaker than that of the electron spin. It was realized that this limitation might be overcome by detecting NMR in condensed matter where a small sample of water containing billions of protons could be subjected to resonance examination. This pursuit led to the 1947 experiments of E.M. Purcell and F. Bloch, as well as their Nobel prize.

Before discussing the actual NMR technique an introductory historical chapter might be of interest. Most of the pioneering quantitative experiments in magnetism date to the century between 1750 and 1850. These involved the use of permanent magnets and battery driven electric currents flowing in wires. These were static experiments yielding the relationships between electric currents and magnetic fields under static or quasi static conditions. What happens during the time between the closing of a battery switch and the time the current reaches its steady state value, was not the focus. That was the situation until about 1830 when Michael Faraday moved a permanent magnet in the vicinity of a wire loop with a current meter in series. He noticed that as the magnet approached the loop the milliammeter indicated a current. When the magnet moved away it again indicated a current, now flowing in the opposite direction. Next, he took a permanent magnet, and rotated it in front of the loop so that the opposite poles would move in alternating sequence by the loop, generated an alternating current in the loop. He built the first AC generator, the seed of our present-day electrical power network.

James Clerk Maxwell came next. He had already built his reputation as a theoretical physicist with the Maxwell-Boltzmann theory of statistical thermodynamics (remember the Maxwell demon explanation of irreversibility) but now came the biggest thing yet, no doubt inspired by Faraday, the synthesis of most of electromagnetism as known up to his time and a couple of the predictions unknown. That is why we refer to Maxwell so frequently. His theories do not explain ferromagnetism but do serve as the foundation to model some of its effects such as hysteresis loops. The Maxwell equations are rigorously valid only for charges that generate electromagnetic fields in a vacuum where radiation propagates at the speed of light c. Thus, c is a parameter that characterizes vacuum as a medium. However, these equations can be extended to media such as air, glass, by defining a pair of constitutive parameters, the dielectric constant and the magnetic susceptibility, with the effective light velocity given by a relation between them. Of course, this covers up all the quantum mechanical processes of radiation scattering and absorption, but it worked sufficiently well and still does, to be very useful in material science and engineering.

The source of the diamagnetic fields of Faraday and Maxwell are moving electric charges in whatever trajectory they follow. If they move in a circle the magnetic field will be as depicted here:

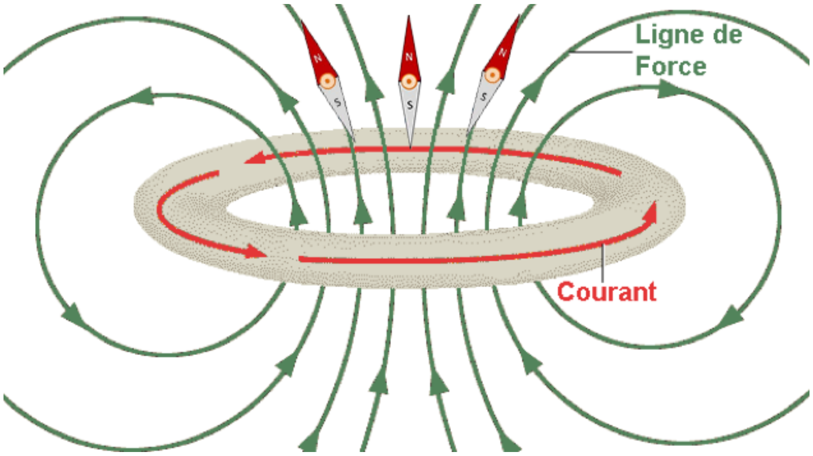


Figure 46. The magnetic field lines of a circular loop of electric current showing aligned magnetic needles.

This idea, of moving charges being the only known source of magnetic field, led to the false assumption that a loop current inside the electron was the source of its magnetic field. With even more evidence -- the similarity of the electric dipole field and the field of a bar magnet shown by iron filings -- the idea took hold. This idea stuck to the point that in 1942 this idea was a prominent part of my introduction to Electricity and Magnetism, magnetic poles and all.

The Electric Dipole and its Electric Field

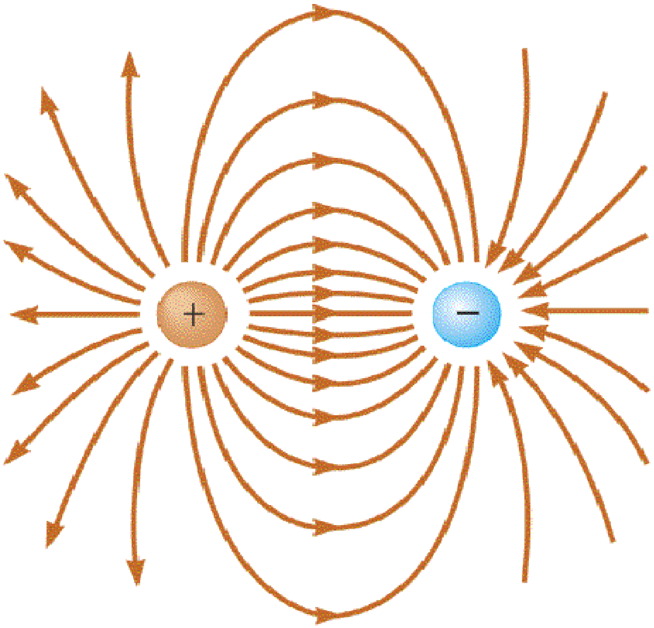


Figure 47. Electric field lines for a positive test charge generated by a charge dipole.

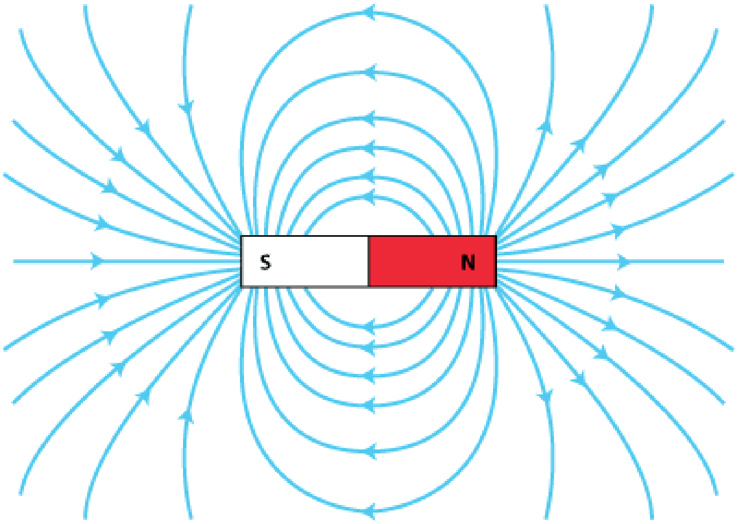


Figure 48. The magnetic field of a bar magnet.

Figures 47 and 48 compare the calculated field lines of an electric dipole (as would be followed by a positive test charge), with the magnetic field lines shown by iron filings sprayed on a sheet of paper superposed onto a small bar magnet.

Before going into a discussion of magnetism consider the two figures. They offer evidence that there are two different kinds of magnetism.

The first is diamagnetism where the field lines do not start or end. They close onto themselves forming closed loops, with the absence of any indication of magnetic charges as sources. There is no analogy with the electric field lines. These are the magnetic fields generated by moving electric charges and are completely explained by the Maxwell theory. Diamagnetism is an integral part of the Faraday-Maxwell electromagnetism where, whether a field is magnetic or electric depends strictly on the reference system.

The sources of diamagnetic fields are moving charges, or, equivalent electric currents, or even just time dependent circulating electric fields (Maxwell in the 2nd edition of his equations in 1862). This dependence of an observation on the reference system contains the theory of special relativity, as Einstein realized later. This is all electromagnetism. For the particular case of diamagnetism, the following important observation applies:

An electron moving in a circle generates a magnetic field in exact opposition to the magnetic field that started it moving in the circle.

This suggests that a circular trajectory would be a stable state for a charge. It suggests that a current circulating in a closed loop could last forever. In reality the current loop has to be supported in some metal wire where the moving electrons get scattered by thermal lattice vibrations and lattice defects. However, this situation is practical today by using superconducting wires, in the closed loops of multiple turn windings.

These are the diamagnetic permanent magnets of today. Just keep them cool. This sort of justifies the hypothesis that the magnetic moment associated with ferromagnets, the second kind of magnetism we mentioned, results from a circulating charge within the electron.

For ferromagnetic fields generated by a permanent magnet, whether it has an additional winding or not, the field lines start at one pole ending on the other pole, but evidence for independent poles is lacking. Furthermore, all ferromagnets seem to be made from metals with atoms that carry an incomplete inner shell with a net finite spin. So, it is reasonable to assume that spins individually are atomic bar magnets associated with the spin angular momentum.

However, this magnetic field superposes and interacts normally with diamagnetic fields even if it cannot be associated with a current loop because it has the wrong polarity. Rather than subtracting from the pre-existing field it adds to it, it lines up with it. Two bar magnets line up, North South, North South. The most basic reason the pole/charge analogy cannot work is because the electric force is a central force, that is, force, velocity, acceleration all directed along the line connecting the two interacting charges. The magnetic, Lorentz force, acts in the perpendicular direction. The magnetic force, the Lorentz force, is not a central force.

So, in order not to get stuck here we accept that there are two different sources for magnetic fields:

1. Moving charges, which covers electric currents and even time dependent electric field loops

2. Spins of elementary particles

From experiments we know that the generated fields are the same in that from measurement on the fields themselves we cannot identify the source.

In explaining the "normal" Zeeman effect the orbital magnetic moment **M** is defined for a closed orbit as a vector proportional to its orbital angular momentum such that **M*B,** the scalar product, is equal to the energy splitting between the neighboring Zeeman levels. This involves the words splitting and levels, so we are in the quantum world. Remember that the angular momentum of a charge moving with velocity v in a circle of radius r is an axial vector normal to the loop $\mathbf{L=rXvm}$, **m** is the mass. For a current $\mathbf{v=2\pi r/T}$ where **T** is the period for a $\mathbf{2\pi}$ turn, so $\mathbf{L=m2\pi r^2/T}$ proportional to area times current.

The magnetic moment is then within a factor of 2, $\mathbf{M=qv\pi r^2}$, the current times the area of the loop. That is the case for diamagnetism as governed by Maxwell and Faraday. The factor of 2 becomes the gyromagnetic ratio which is rigorous for classical diamagnetism but needs the small but important corrections derived by quantum electrodynamics. The value of **g** for the electron spin is near 1 as obtained experimentally, and for nuclear spins is much smaller, and is obtained also from experiments.

For the spin angular momentum, the reverse reasoning has to be used. From the energy difference between the pair of split levels, more precisely its dependence on the magnetic field, we deduce an empirical value for the spin magnetic moment. That is how the spin magnetic moment is defined: by the anomalous Zeeman effect.

Ampère's law with Maxwell's addition

Ampère's law with Maxwell's addition states that magnetic fields can be generated in two ways: by electric current (this was the original "Ampère's law") and by time dependent electric fields of the right configuration (this was "Maxwell's addition", which he called displacement current). In integral form, the magnetic field induced around any closed loop is proportional to the electric current plus displacement current. The Displacement current in turn, is proportional to the rate of change of electric flux through the enclosed surface. The meaning of this statement in less abstract words means that whereas a conducting loop with an insulating gap cannot support a DC current it can support a time dependent current when the battery is turned on, the charging current. Inspired by Faraday he then generalized Faradays idea of AC current in a loop by a time varying magnetic field, by introducing the concept of looping electric fields generated by magnetic fields across the loop, provided they are both time dependent. Then he postulated the reverse, that a looping electric field will generate a magnetic field through the loop. That was the clue to the explanation of the fact that radiation, light in particular, is an electromagnetic phenomenon.

Maxwell's addition to Ampère's law is particularly important: it makes the set of equations mathematically consistent for non-static fields, without changing the laws of Ampere and Gauss for static fields.[2] However, as a consequence, it predicts that a changing magnetic field induces an electric field and vice versa.[1][3] Therefore, these equations allow self-sustaining "electromagnetic waves" to travel through empty space

The speed calculated for electromagnetic waves, which could be predicted from experiments on charges and currents match the speed of light; indeed, light is one form of electromagnetic radiation (as are X-rays, radio waves, microwaves).

Now we really get into NMR, nuclear magnetic resonance

In an external applied magnetic field sufficiently large compared to the interaction fields, for individual electrons as well as protons **(H nuclei) t**he energy separation between the levels of a parallel(+1/2) and an antiparallel(-1/2) proton or electron are proportional to the magnetic field, so by determining the transition frequency between the two levels one can obtain the value of the field. The actual field felt by say the proton, is slightly altered by the bond arrangements among the neighbors. The technique that emerged became basic in the determination of chemical structures especially for organic materials.

The spin of the proton, being 1/2 gives only two energy levels with magnetic quantum numbers +/-1/2.) So does the neutron. Other nuclei contain larger numbers of both that have to be combined to the range defining their total angular momentum, or total spin.

Section 10. Nuclear spins and Nuclear Moments

The idea that the (Nuclear) Magnetic Resonance technique could be used for multi-dimensional (tomographic) imaging of chemical inhomogeneities even in live bodies, by observing and analyzing magnetic resonance spectra, obtained in properly designed and calibrated magnetic fields emerged early.

However, the associated technology took over a decade to come to fruition. It is made possible by the use of the time coherence of the employed radiation at the relative low frequencies used for NMR where the wavelength is long compared to the dimension of the sensing and excitation components and one is dealing with a very large photon flux; coherence effects can be detected directly in the time domain. Furthermore, the discovery of relatively high temperature type II superconducting alloys that could be drawn into fine wires and remained superconducting in rather high magnetic fields was essential.

The proper description of all these effects which is required for equipment design, is deeply rooted in quantum mechanics. Not only that, the realization that any more fundamental "explanation" of the properties of the electron spin, including items such as the **LS** coupling requires the use of concepts of special relativity implies the uncomfortable truth that the electron spin is not only quantum mechanical but a fundamental relativistic concept. This was synthesized by Paul Dirac in his relativistic wave equation, which, on top of it, introduced to the world the positron. At this point we stop this side story to return to our drive towards NMR and MRI.

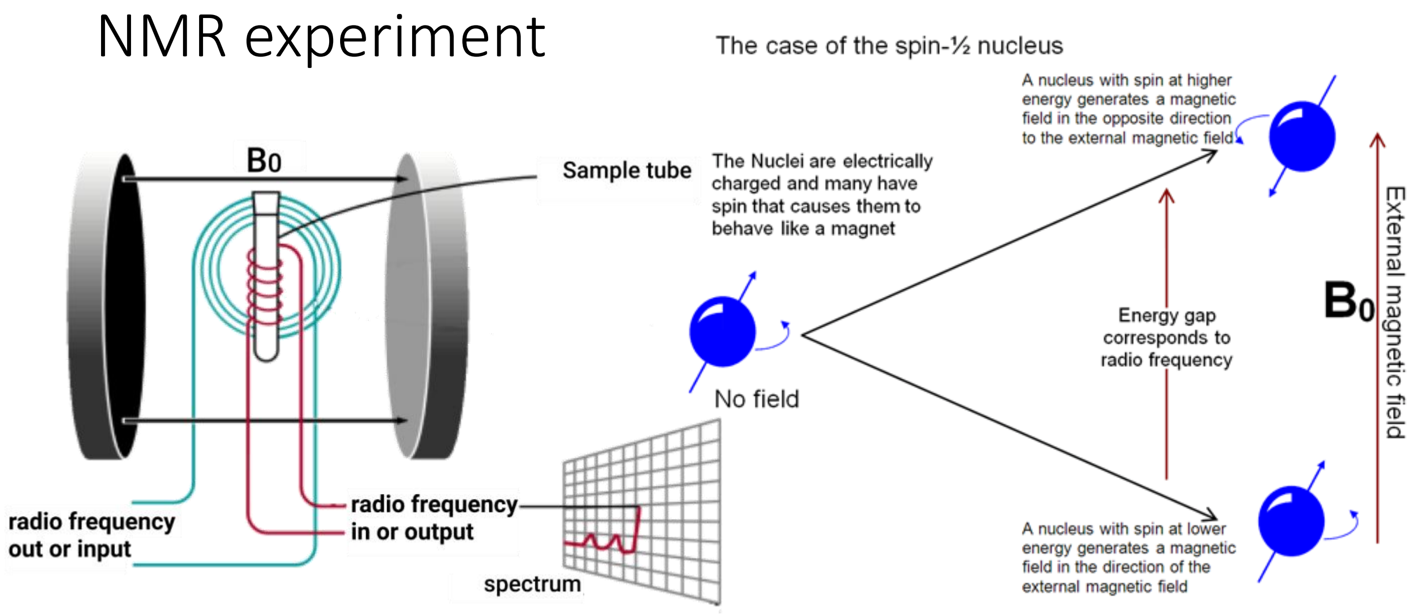


Figure 49. Setup of an NMR experiment.

The following picture illustrates how a magnetic field exerts a torque on a current loop via the Lorentz force

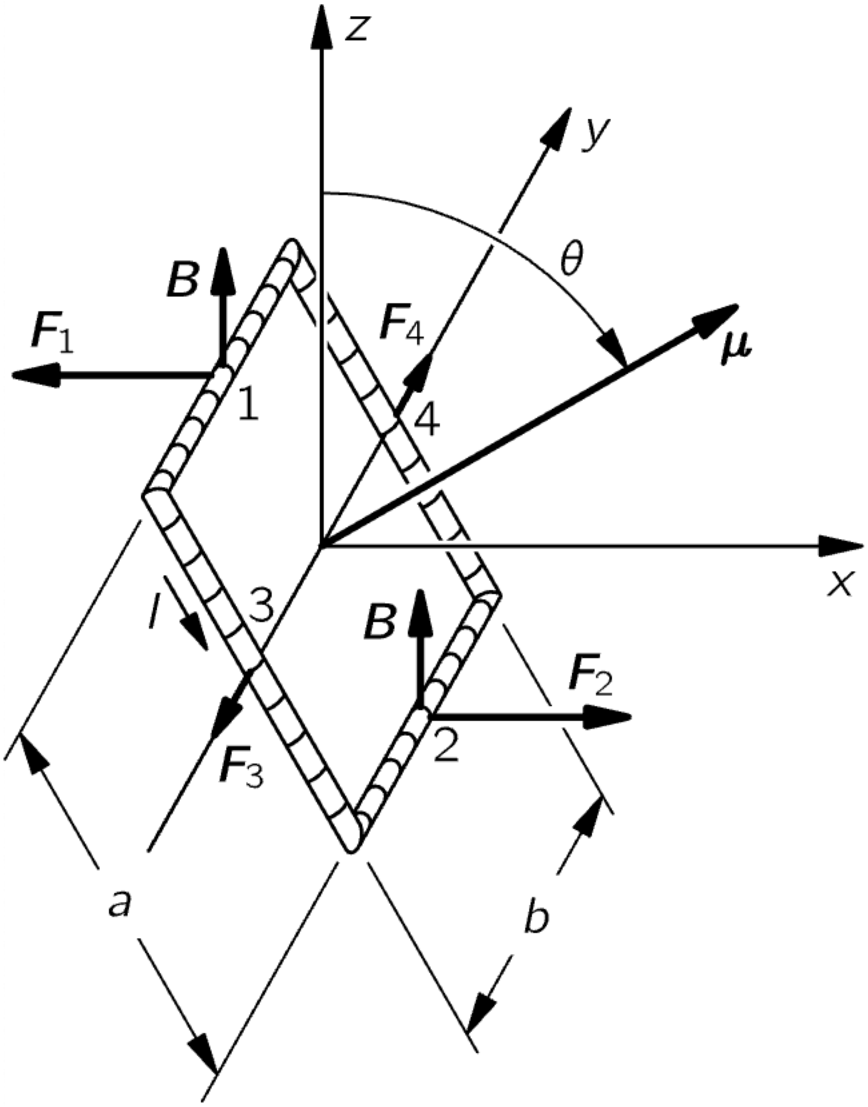


Figure 50. The torque on a current loop (Feynman Volume2),@ as predicted by diamagnetism is illustrated for a rectangular loop.

The magnetic field tries to align the loop to a plane normal to the magnetic field. The magnetic needle or a small bar magnet is also aligned parallel to

@ Richard Feynman, Lectures on Physics, Volume II

Field. This puts the spin's loop, just as the current loop in the plane normal to the field, but in the opposite direction (handedness, remember we are dealing with axial vectors here!).

$$\textbf{Force } \mathbf{F} = \mathbf{qv} \text{ X } \mathbf{B}, \qquad \textbf{Torque } \boldsymbol{\tau} = \mathbf{r} \text{ X } \mathbf{F}$$

So, the spin, just as the bar magnet, cannot be explained by unmodified linear Maxwell equations. This can be achieved, but at the cost of introducing nonlinear terms into the Maxwell equations inside homogeneous materials. These terms represent the electric and magnetic interactions of the molecules making up the materials, the polarizations, but important as they are, they have no role in the discussions of magnetic resonance physics. (And the electromagnets are fast being replaced by superconducting coils).

So, physics sausage makers accepted -- rather invented -- a new rule for the spin. The magnetic torque rotates the spin into alignment BUT along the field, not in opposition, and miracle, this works. Thus it imparts to the SPIN an extra energy of magnitude proportional to the product of the magnitudes of the torque and the magnitude of this spin angular momentum, starting at zero if the loop starts in a plane parallel to the field (**L**, angular momentum, perpendicular to the field **B**), then moving to its maximum value when the angular momentum **L** becomes parallel to the field **B**.

With the magnetic force given by **F** = **qv** X **B** and the corresponding torque given by **τ = r X F,** the magnitude of this energy will be proportional to the product of the magnitude of **B** and the magnitude of the component **mħ** of the angular momentum parallel to **B.** X indicates vector product (see#1Appendix B.)

Spin (1/2) in magnetic field

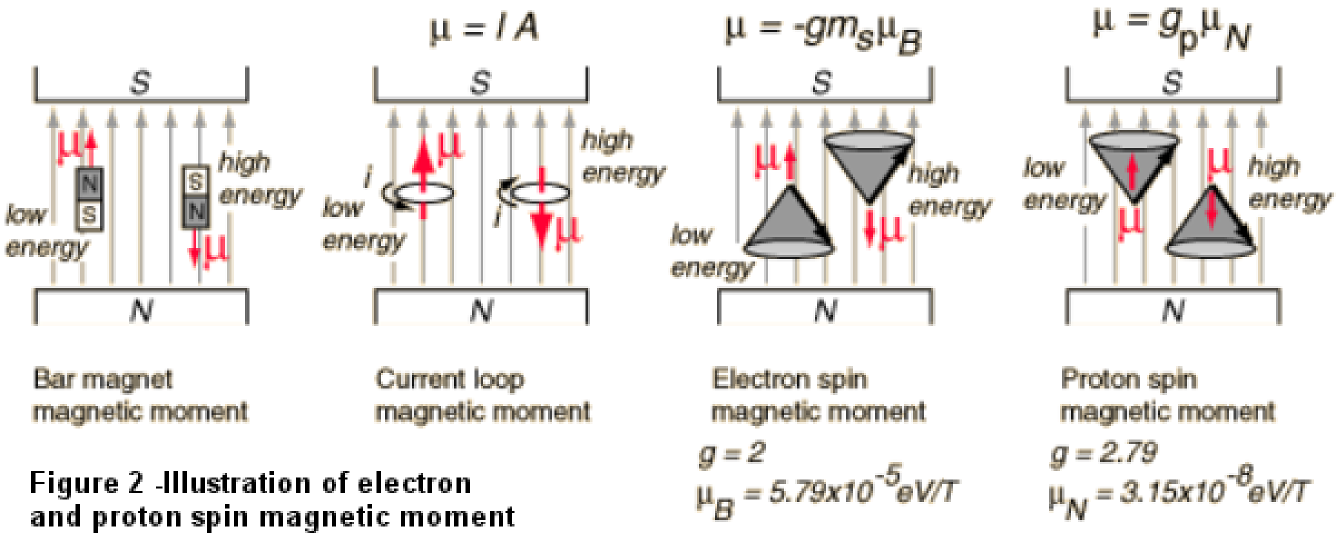


Figure 51. Comparison of magnetic dipole moment orientation between bar magnets, current loops, electron spins, nuclear spins.

The nuclear spins have the same experimentally observed properties as the electron spins, but with much smaller magnitude. The component along an imposed axis will have integer or half integer values that are determined by its angle with the axis. This discussion deals with the proton, the hydrogen nucleus, of spin 1/2, therefore its magnetic moment can be only parallel or anti-parallel to the net magnetic field.

The total field at the site of the nucleus, protons in this case, is **B**(total) = **B**(applied) + **B**(intra molecule atoms) + **B**(near neighbor molecules).

In an applied magnetic field large compared to the interaction fields (they cancel to a large extent), the energy separation between the levels of a parallel and an antiparallel spin is proportional to the magnetic field, so by determining the frequency one can obtain the value of the field.

The actual field felt by say the proton, is slightly altered by the bond arrangements to the neighbors. The technique that emerged became basic in the determination of chemical structures especially for organic compounds.

Spin (Electron or Proton) Aligns Up or Down
in a Magnetic Field
Absorbed Frequency $\nu=\Delta E/h$

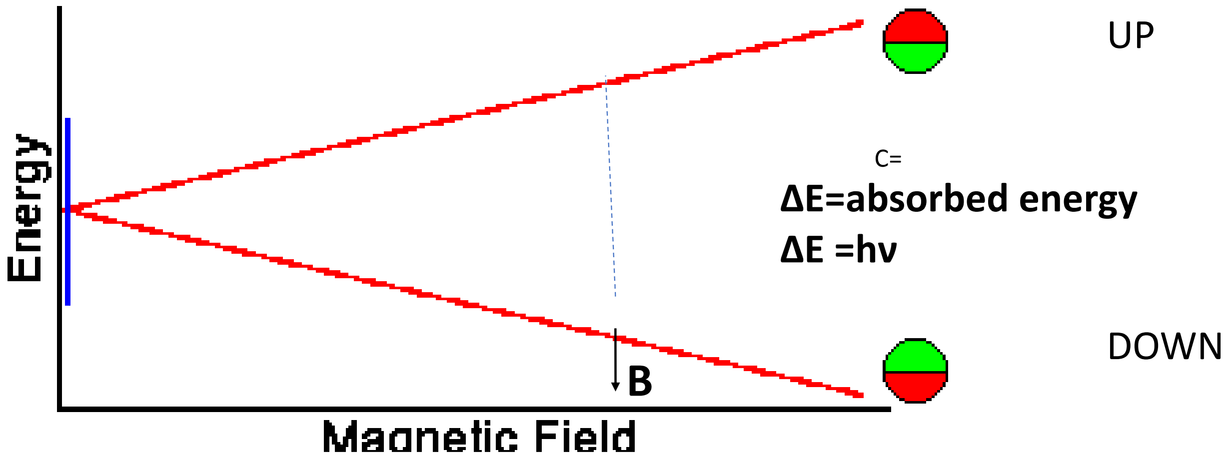


Figure 52. NMR Absorption frequency vs Magnetic field.

If the magnetic field is inhomogeneous, frequency vs position can be calibrated by point by point probing, scanning with a miniature NMR probe.

Two techniques for observing NMR, absorption and induction will be considered.**8**

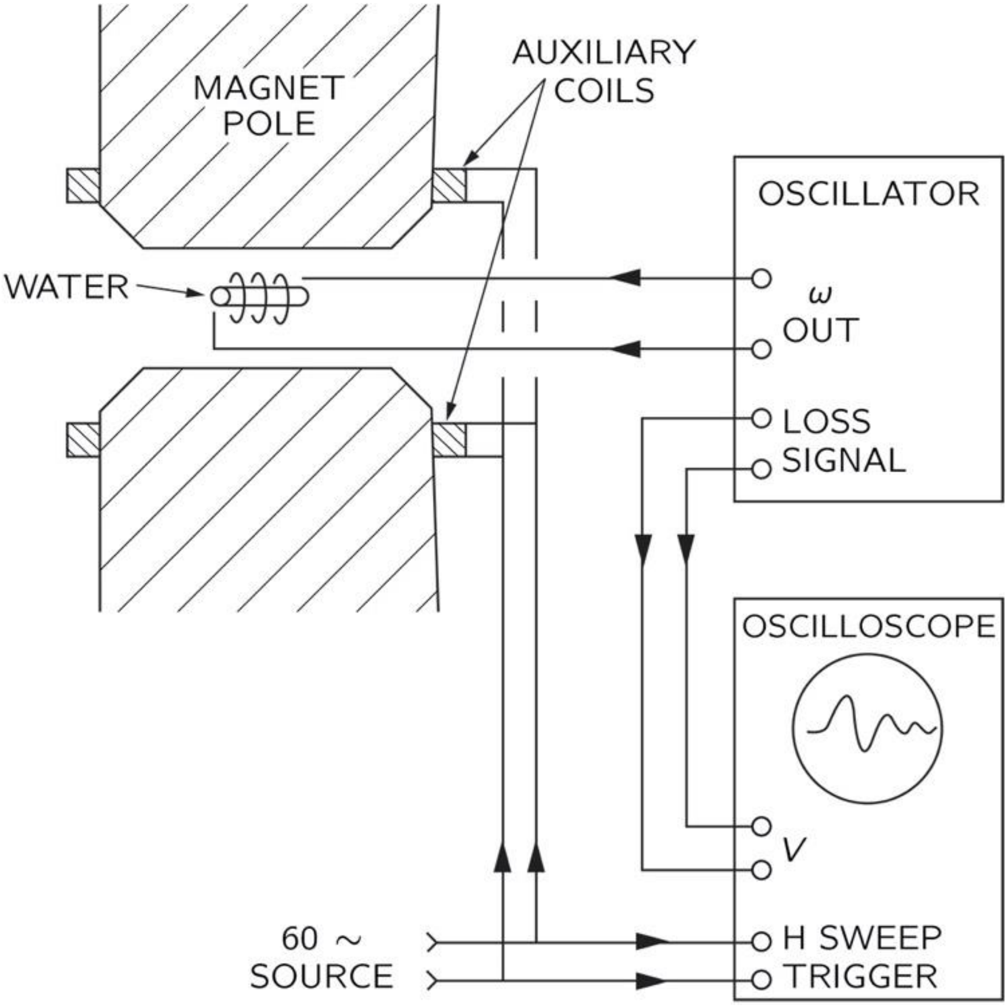


Figure 53. Diagram of the NMR setup used to record the NMR spectrum by Purcell, Torrey and Pound.

8E. M. Purcell, H. C. Torrey, and R. V. Pound, Phys. Rev. **69**, 37 (1946). absorption
F. Bloch, W. W. Hansen, and M. Packard, Phys. Rev. **70**, 474 (1946). induction

Next, we discuss, in some detail how the NMR is detected experimentally. There is some history in what follows. In the constant and homogeneous magnetic field, the nuclear spin will absorb a photon of correct frequency and jump to the upper state. A small enough sample so all spins see the same field, will keep absorbing photons of the same frequency until all spins are in the upper state. At this point the sample stops absorbing. It is saturated.[9] In reality total saturation is never achieved due to mechanisms of relaxation ever present. One mechanism is direct, where thermal giggling directly encourages spins to drop back down. It is called spin lattice relaxation. It is characterized by a relaxation time **T1**.

The second is called spin-spin relaxation caused directly by the magnetic interaction between the spins, with relaxation time **T2**, the spin diffusion time. **T1** is usually much longer than **T2**. This fact allows us to manipulate the spin system while it is out of thermal equilibrium. It is this situation that makes it possible for induction as well as pulse detection work.

One can sense the absorption by generating a tunable radio frequency RF signal by means of a tunable oscillator. The positive feedback loop of this oscillator consists of a series RLC circuits whose resonant frequency can be swept through the resonance region by sweeping the capacitance. At resonance the feedback loop will be essentially a short thus maximizing the positive feedback and the stability of the oscillator.

---

[9] Relaxation Effects in Nuclear Magnetic Resonance Absorption
N. Bloembergen, E. M. Purcell, and R. V. Pound
Phys. Rev. 73, 679 – Published 1 April 1948
An expanded version of this paper was published by Dover and it consists of Bloembergen's doctoral thesis presented to, Leiden University. At that time, many European Universities required the dissertation to be in regular book form accepted for publication by a recognized publisher. His degree was not a Harvard PhD but a Leiden D.Phil., a source of pride.

The sample is enclosed in a small test tube, with the inductance coil wound around it. When the frequency of the oscillator hits the NMR frequency some of the strength of the oscillator will be lost, i.e. the series resistance of the coil increases, and the feedback weakens the oscillation dips and the NMR is detected. The design of such an oscillator, the marginal oscillator was the contribution of R.V. Pound to the Harvard effort towards NMR and was baptized as a Pound box.

The frequency of the **RF** signal is determined by setting the capacitance **C**. Remember this is 1947, using 1940-1950 technology. We sweep the magnetic field, as it crosses the value corresponding to the frequency, the strength of the **RF** signal dips due to extra resonant absorption of resonant **RF** photons. If the rate of arrival of the **RF** photons is too high the **T1** relaxation process cannot keep up, the spin ensemble system is saturated and there is no absorption signal.

In the years after WWII, experiments were unsuccessful in detecting an NMR of protons, even in water, until the saturation mechanism was understood. The cure, when the need for the enhancement of the **T1** relaxation was realized in about 1947, was the construction of stable very low intensity **RF** oscillators (marginal oscillators also called Pound boxes) and dissolving into the sample high spin magnetic ions such as **Fe+++ or Gd ++++.** The first reliable signal was detected!

These magnetic ions provided an enhancement of thermal relaxation processes via strong local fluctuating magnetic fields. However, another less obvious mechanism was developed at Stanford. Whereas the Harvard path to NMR was quantum paved, at Stanford it was paved with classical concepts such as the Lorentz force and the Larmor precession. This more-subtle, and complex, procedure deserves the following extended discussion as it is deeply imbedded in the MRI technology. In contrast to absorption it is labeled induction and it involves the basics of the **T2** relaxation process. It is best described in macroscopic terms that we try to justify.

A strong uniform magnetic field will align most of the proton spins parallel to generate a macroscopic magnetic moment **M**. The **RF** current in the coil generates an oscillating small magnetic field perpendicular to **B**, the large uniform constant field. This small field tips **M** by a small angle from the field, creating a torque that will start to rotate (precess) it around the magnetic field. The torque will generate an angular momentum kick that starts the rotation, that will continue by itself indefinitely in the absence of **T1** relaxation. The frequency of this rotation is the Larmor frequency. If this small field were time independent, it would continue to tip **M** to larger and larger angles to the field lasting times of the order of **T1**. Well as we have learned about polarized waves, a circular motion can be constructed by superposing two linear oscillatory motions. The reverse is also correct. The linear oscillation of the **RF** magnetic field in the coil decomposes into two oppositely rotating components rotating at the **RF** frequency and if this frequency happens to equal the Larmor frequency, GUESS WHAT! The classical Larmor frequency is the quantum mechanical NMR resonant frequency at that field.

By imagining two opposite magnetic charges at the ends of the bar magnet we invented the magnetic dipole. There is no experimental evidence of independent magnetic charges. But in some sense the magnetic dipole is the generator of the magnetic field. It has a magnetic dipole moment which can be associated with an angular momentum. The angular momentum is the real thing. The Zeeman effect justifies the association of a magnetic moment to the electron spin as a consequence of the energy level splitting in a magnetic field. We define the magnetic moment of a spin as the factor by which we have to multiply the magnetic field to obtain the energy splitting at that field. This is analogous to the orbital diamagnetic case, but it is only an analogy of convenience not justifiable by any fundamental argument.

In a way NMR plays the equivalent role for nuclear spins. Accepting this magnetic moment as a fundamental property of nuclear spins the science of magnetism continued its development. New rules, new assumptions are added as old ones are modified or even abandoned as experiments keep piling up.

Now, we take a step back and, for the specific case of NMR, make a connection to the pre-quantum world. This is necessary to properly explain the induction NMR detection method. We can actually listen to the radiation from the oscillator exciting the NMR absorption via the heterodyne technique. It will produce an audible tone whose frequency parallels that of the NMR oscillator.

What happens to this tone, in addition to the loss of intensity as the magnetic field is swept by the resonant value, is a sharp detuning. First, the frequency goes down then jumps very high, or vice versa depending on the direction of the sweep. If we increase the power of the oscillator towards saturation the strength of the signal decreases as we approach saturation, but the detuning persists, indicating that somehow the dispersion that accompanies saturation might be used to detect magnetic resonance. Indeed, this works. The presence of the extra absorption not only adds resistance loss to the resonant circuit but shifts the effective magnitude of the inductance **L** thereby affecting the frequency. These effects can all be reproduced in simple AC circuit theory by working out the impedance of a resonant LC circuit into which a series resistance is added to the inductance, to represent the magnetic resonance absorption by the sample.

Thus, the idea emerges of examining the transient response to a pulse of resonant frequency radiation. The oscillating magnetic field in the coil can be thought off as a superposition of two oppositely rotating component fields. To proceed, start by moving (in thought) to the Rotating Field Reference Coordinate System& rotating with the component of the **RF** field that is in synchrony with the Larmor precession. It won't be an inertial system, but in the presence of the large magnetic field plus a weak radiation field, gravitation can be neglected. This also means that we are neglecting the counter rotating component of the **RF** field, and because our time scale will be short compared to **T1**, thermal giggling effects are negligible. So, in the rotating frame the magnetic field is time independent. But now we pluck it by adding a pulse field of short duration compared to **T1**. This will tip the field by an angle that keeps increasing as long as the pluck is on. As I go off the pluck the **RF** magnetic field is tipped to a fixed angle to the constant field depending on how long it was left on. For example, if the pluck was left on for a full period of the **RF** frequency it would have returned to its initial position parallel with the field. For a quarter period it would put it perpendicular with the field. We used the term pluck to remind you of our discussion of confined states in section 2 (Waves)**.** In more technical language we replace pluck by pulse. I remind you as well as myself that the experiment is never actually performed in this imaginary reference system, as realistic as it sounds. But it is extremely useful to visualize what is being explained. There is, however, a related experiment which can be performed involving this concept -- that of the spin echo using sequences of 180-degree pulses. This technique is an integral part of the pulse process used in many practical NMR detectors.

&NMR in the rotating frame http://www-keeler.ch.cam.ac.uk/lectures/understanding/chapter_3.pdf

After the cutoff of a pulse, the intensity of the received signal will decay with a time constant which is a combination of T1 and T2, such that

1/T = 1/T1+ 1/T2, or T = T1*T2/(T1+T2)

So, if **T1 >> T2, T = T2** approximately. If we Fourier analyze the decay, we obtain all the magnetic field values where protons are present. The designed spatial variation of the magnetic field identifies the location of protons. The strength of the component gives the value of the proton density at this location in relative terms. That is how an MRI scan can be constructed.

Now you know how **T1** and **T2** detection schemes differ. In **T1** we look for low frequency components, so you observe the decay for long times. For **T2 w**e record with much higher time resolution the initial decay. In typical MRI applications, protons are present all over the body and one can select specific areas of observation by using suitably designed **RF c**oils to supply suitably configured radiation. A suitably designed antenna coil will capture the radiated tail signal to a receiver through a time gated input circuit that blocks the pulse from saturating the receiver.

A small designed inhomogeneity of the magnetic field is superposed to a strong homogeneous magnetic field so that the net magnetic field becomes a known function of position.

Then, the large-scale NMR spectrum will consist of a superposition of spectra coming from all locations which can be captured by pulsed methods to be presented, below. Analysis of the pulse response can unscramble the different contributions coming from the different locations. This provides the distribution of the absorbing protons throughout the sample. Analysis of the frequency distribution yields the 3-

dimensional image of the absorbing proton distribution. This is the physical principle of MRI**@**

@ T1 and T2 scans https://casemed.case.edu/clerkships/neurology/Web%20Neurorad/MRI%20Basics.htm

Section 11. The MRI magnetic field, DC and RF

A long, large-diameter superconducting solenoid generates a very large, accurately homogeneous magnetic field in a horizontal direction. The solenoid is enclosed in a liquid helium annular cross section Dewar. The diameter of the central room temperature cylinder has to be sufficiently large to accommodate the specimen (me or you) and additional gradient coils that provide the designed small inhomogeneities to the total field, controllable from external current supplies, and the coil to provide the pulsed radio frequency to be absorbed by the protons. The position of this coil determines the region of the subject that is the object of the scan.

The Magnetic Field in MRI

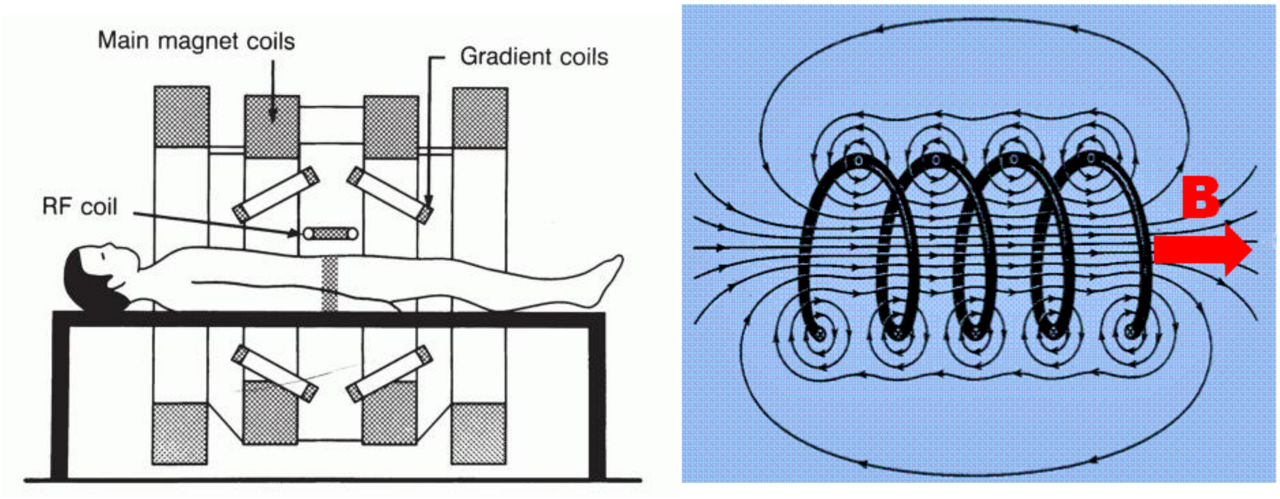


Figure 54. The various coils positioned around the “sample”.

Longitudinal Field Gradient
Selects the Slice by RF Pulse Duration

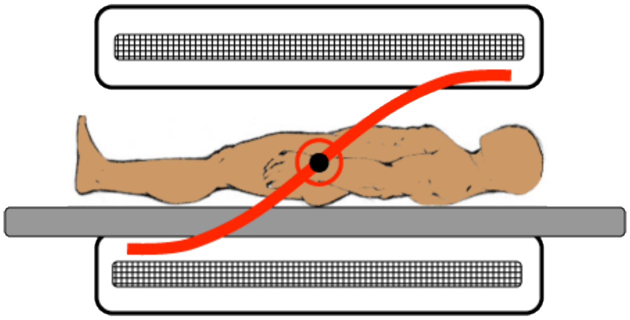


Figure 55. Small longitudinal gradient superposed to the large field, for slice identification, the essence of tomography.@

The slice is specifically selected by timing as the patient is slowly, or by steps, moved in the longitudinal direction, i.e. along the axis of the solenoid.

@ https://study.com/academy/lesson/what-is-tomography-definition-history.html

Rotating Transverse Gradient
Image of Selected Slice by Tomography

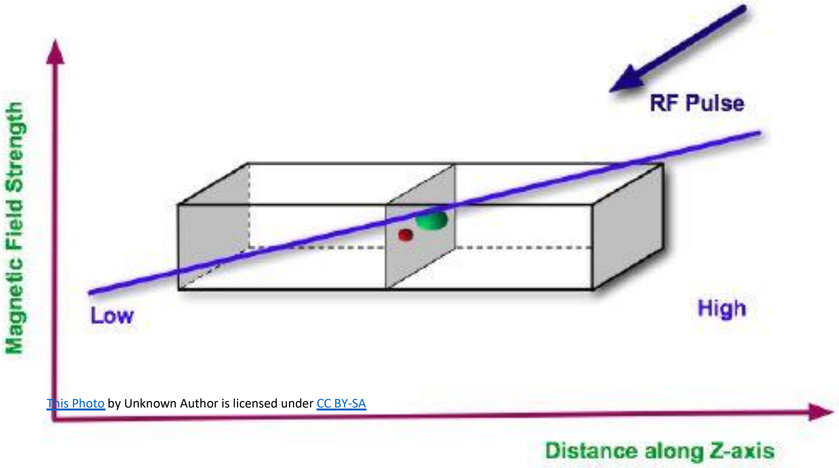


Figure 56. If a field gradient is applied along a body diagonal and then it is rotated its component in a transverse plane will rotate in this plane with constant magnitude. It becomes a rotating field gradient.

This is the simplest moving coil arrangement to realize an approximate rotating field arrangement in the transverse plane.

NMR Experiment. Induction.

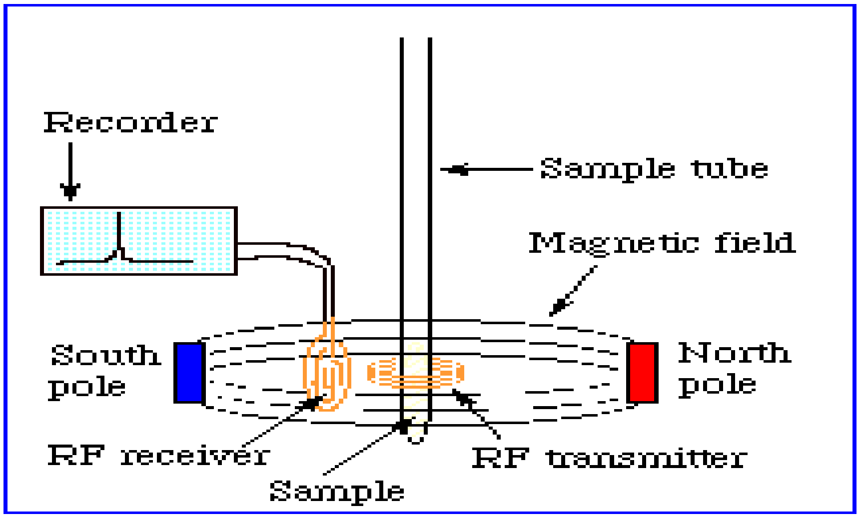


Figure 57. Pulsed NMR Experiment

The transmitter supplies the strong RF pulse to the transmitter coil. The receiver coil picks up both the strong pulse and the weak decay, but the receiver is time gated to block the strong pulse. The signal is processed by a fast Fourier transform algorithm. The recorder displays the Fourier transform of the signal decay.

If the **RF** transmitter applies the current to the coil in the form of a pulse with a sharp end, the induced **RF** current in the receiver coil will not decay so sharply. This is analogous to what happens to the wave response of a violin string to a pizzicato pluck. As illustrated in the next figure, the time evolution of this decay, recorded and input to a computer program, can be analyzed to yield all the different frequency components of the nuclear spin response. Thus, the NMR spectrum can be obtained without scanning either the frequency or the magnetic field.

In this version of this illustration, in the lower plot, each pulse represents a pulse of current applied to the transmitter coil and the subsequent decaying oscillations. The decaying signal in the receiver coil is the superposition of the different frequencies from protons at different locations shown in the upper plot.

Confined Waves = Selected Wavelengths
Superposition of Different Wavelength =
= Train of Pulses

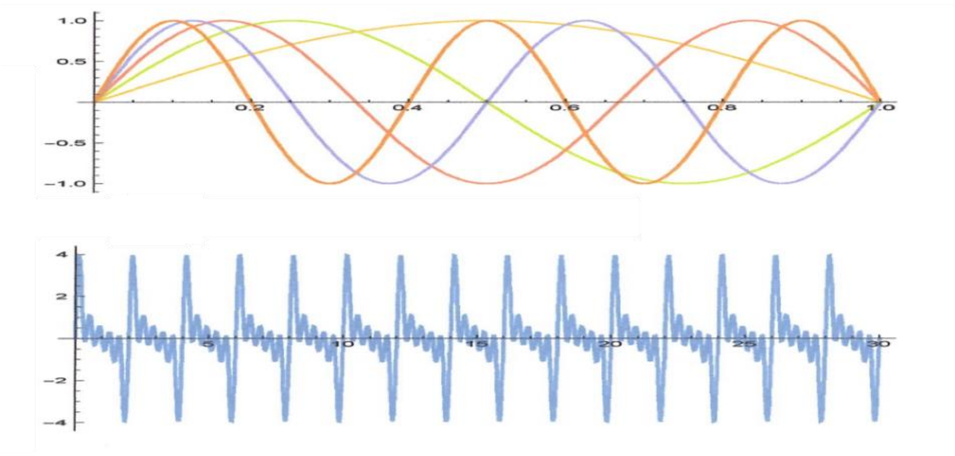
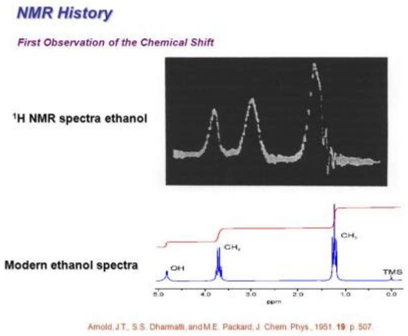


Figure 58. Reproduced here to illustrate Fourier analysis. How it applies to pulsed NMR data analysis. The top graph displays the Fourier components of the decay that is displayed in the bottom graph.

Section 12. NMR in Chemistry. Exploring molecular structure. MRI mapping of chemical species in the human body

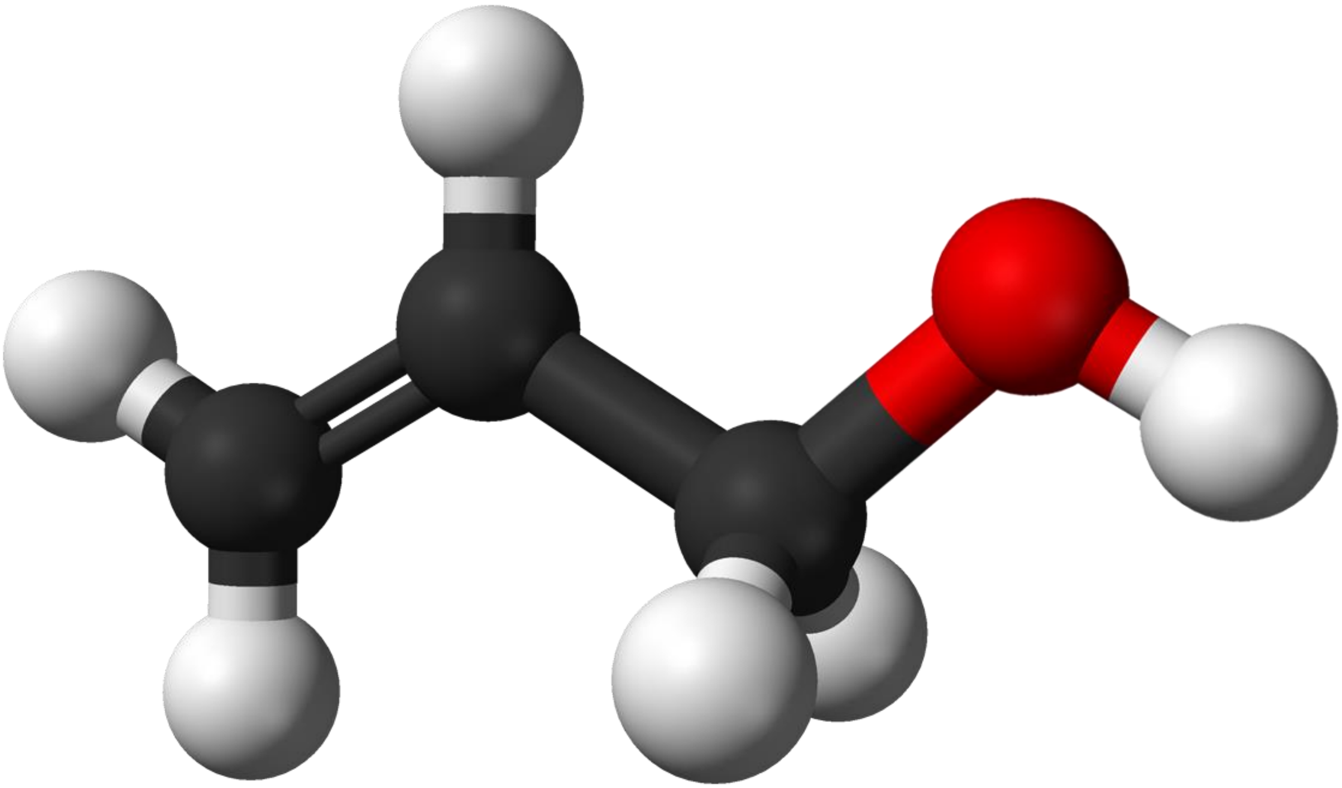
NMR Absorption of Protons at Different Locations in the Same Material



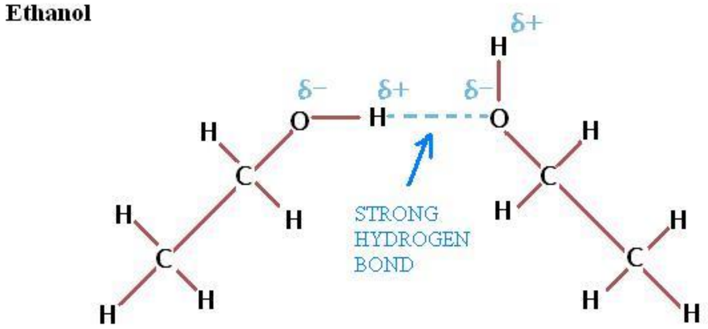
Arnold, J.T., S.S. Dharmatti, and M.E. Packard, J. Chem. Phys., 1951. **19**: p. 507.

Figure 59. The proton NMR spectrum of ethanol. Six Protons in 3 different sites within the molecule.

Two of the sites are almost equivalent but not quite as shown in the lower high-resolution spectrum as the peak on the right seems to be a doublet.



This Photo by Unknown Author is licensed under CC BY-SA



Figures 60(A) and (B). The ethanol molecule in stick imitations.

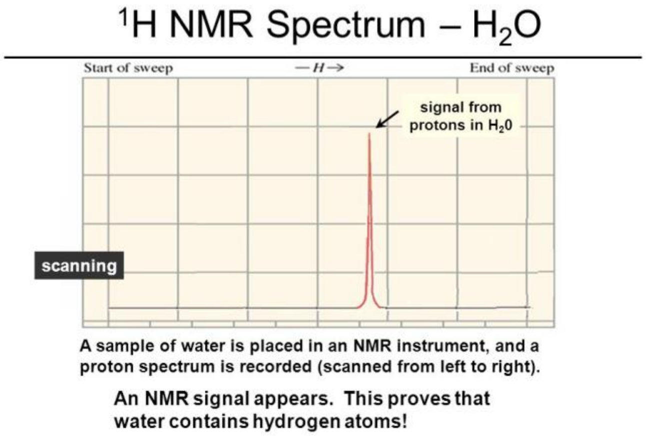
The three major absorption peaks in Figure 59 (top) belong to protons at different locations in the molecule. The higher resolution spectrum below has a fine structure indicating at least three other locations. Indeed, ethanol has six protons. This molecule has a twofold symmetry but the small splitting in the absorption lines indicates that it is only almost perfect. The horizontal scale is in frequency units. This shows the extreme power of NMR as an extension of our senses to observe and measure the invisible small.

Total field = average external field + field inhomogeneities + field due to neighboring, e.g. atoms with partially filled shells **(Fe, Ni, Gd…) +,** small but definite contributions from bonding to neighboring atoms via partially filled covalent bonding (chemical shifts), and weak Para magnetism of conduction electrons in metals.

These magnetic interactions play a central role in equilibrating the population distribution among the spin states (up and down), at a spin temperature different from that of the whole material. This occurs during time **T1.** Then this spin temperature slowly decays to the temperature of the whole material. This takes times in the order of **T2.** Depending on the time scale of the observation, different effects in or out of thermal equilibrium can be recorded providing information about different processes occurring in the material.

These magnetic interactions are enhanced by the presence of atoms such as **Fe** that have unfilled inner shells and thereby net electron spin magnetic moments. They enhance the spin-spin interactions as they introduce into the sample these large magnetic moments. (Remember Hund's rule.) Hemoglobin has iron in its molecule. The presence of iron nearby enhances the thermal equilibration of the populations of protons so that it keeps the number in the low energy states larger than in the high energy state. Thus, high power observation of NMR, without saturation, are possible. The single pulse signal from protons in red blood cells, in Figure 62 indicates the sites of blood flow in the sample.

NMR of Protons in Water



A sample of water is placed in an NMR instrument, and a proton spectrum is recorded (scanned from left to right).

An NMR signal appears. This proves that water contains hydrogen atoms!

Figure 61. NMR of protons in water obtained after averaging thousands of noisy single pulse spectra.

Next, we consider a striking example where proton NMR gives information about its chemical neighborhood which had immediate applications in MRI scanning.

NMR of Protons in Blood
Effect of Hemoglobin

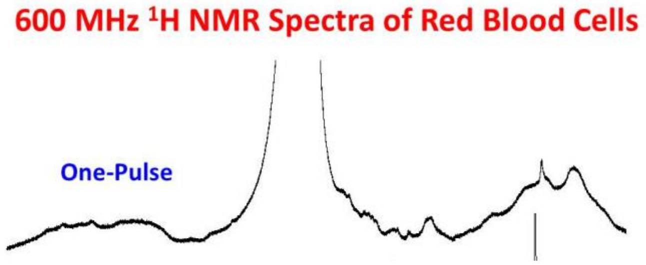


Figure 62. Single Pulse Spectrum of protons inside a red blood cell.

The intensity enhancement is caused by the proximity of the heme and its bound iron atom. Iron has a half full inner shell which has a spin of 5/2 (Hund’s rule) so carries a magnetic moment. By interacting with the nearby protons, thereby shortening **T1**, and minimizing saturation, it strongly enhances their NMR signal.

So, NMR can be used to locate regions of intense blood flow. This was the basis for the first application of MRI in medicine.

Let us see how this application works, in some detail. Take a simple case of a subject that has three spots of high blood flow in the same transverse plane, as in the figure below. And let us designate each location in this plane by its polar coordinates centered on the axis of the rotation of the gradient.

Locating Blood by NMR with Field Gradient

Weaker field-------Stronger field

Weaker field

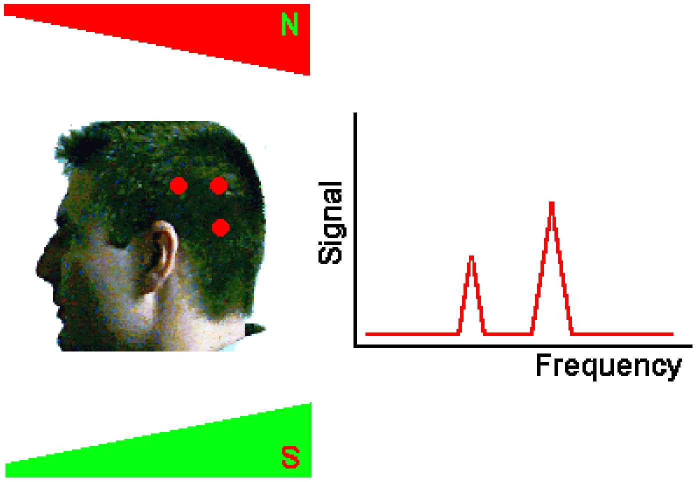


Figure 63. The 3 red spots form a right triangle. They indicate the presence of blood

When applying a field gradient either in the vertical or horizontal direction the field scan will be as in the plot where the frequency is the Larmor frequency representing the value of the field. If we rotate the gradient by

360 degrees, in 30-degree steps, we obtain the scans depicted in the figure 64 below

Describing Slice Image Perception by Nuclear MRI

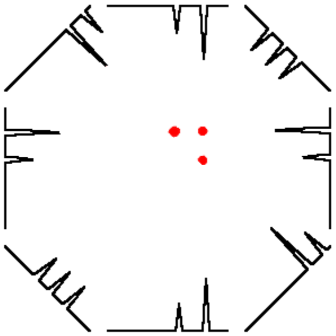


Figure 64. The MRI Trick.

We can increase the number of Rotating steps until this diagram looks like a continuous curve. The question is, how do we interpret such a curve. To do that let us look at the peak values. They double if there are twice as many spots along that direction.

So, the interpretation is obvious. It measures the number of locations along that direction that have the same value of the field. If a field value identifies a geometric location the interpretation is that this process identifies the distribution of protons in space, provided the correspondence between resonant frequencies or field values on one hand and field values and locations on the other hand are unique. This procedure has to be repeated for each value of the radial coordinate to completely specify the proton distribution in the slice, circle by circle.

Because one rotating gradient coil is more practical than two translating coils, we have dealt with polar coordinates in the transverse plane. However, it is easier to explain what follows using cartesian coordinates, so let us turn to the longitudinal gradient.

We don't actually perform a longitudinal scan, just store the three-dimensional data in the form of cylindrical coordinates. The longitudinal cartesian coordinate $\mathbf{z}$ designates the tomographic slice and the polar coordinates as used in the above argument do the reconstruction circle by circle of the distribution of proton densities. The above discussed example shows what has to be done for each value of the radial polar coordinate. The peaks that would appear in the longitudinal scan represent the proton locations which are in the slice that has the value of the magnetic field corresponding to the RF frequency as its Larmor frequency, and the peak intensity measures the number of protons in this slice, in other words, the density of protons as a function of the longitudinal coordinate $\mathbf{z}$.

This implies that if we pulse the RF the decay will consist of the superposition of all Larmor frequencies pertinent to all magnetic fields present in this gradient. An inverse Fourier analysis will therefore give the distribution of proton density as a function of location -- in this case, slice position in the field of this longitudinal gradient. In other words, we have obtained the proton density $\boldsymbol{\delta(z)}$ **as a function** of the longitudinal coordinate $\mathbf{z}$**.** We have done the same thing before, as a function of the radial coordinate $\boldsymbol{\rho}$ and polar angle $\boldsymbol{\theta}$ for the transverse gradients. Thus, we have obtained the distribution of the proton density in cylindrical coordinates. This is for convenience because of the cylindrical geometry of the magnetic field. At the same time, this has outlined the tomographic process of transforming experimental scanned data into spatial distribution of local measurable properties.

Every item around the specimen needed for the NMR detection has to be rotated simultaneously. This is quite a technical achievement!!! The longitudinal gradient allows slice selection and the rotation allows tomographic reconstruction.

The discussion above used the example where the hemoglobin is present only in three distinct locations. Actually, three peaks only clearly show in one view. In the other views one of the peaks is occluded by another peak. In this particular case the three locations are at the vertices of a right triangle. Its high symmetry allows for a number of equivalent viewing directions. As the number of locations increases the distribution of peaks along the circle assumes the aspect of a continuous distribution which represents, in polar coordinates, the distribution of hemoglobin in the chosen slice of the sample. The name tomography is inherited from the technique of actually cutting the sample into sufficiently thin slices so that their anatomy could be examined by transmission microscopy.

The rules governing the behaviors of elementary particles are those of quantum mechanics no matter how many particles are present in the volume under observation. However, as this number increases, the use of these rules becomes extremely complex and the rules of classical, pre-quantum physics become a useful approximation. The description of the NMR phenomenon is the shining example of this situation. Intelligent and lucky guesses gleaned from the classical rules were the essential sources of progress. We reserve the discussion of the physics enabling technique, Superconductivity, to the appendix section at the end of the paper.

Section 14. Use of radioactive labeling in medicine, and the Physics of PET scans

We have discussed quantum mechanics in situations where classical physics provided at least a starting point for understanding if not a quasi-complete explanation. There is no such starting point for radioactivity. So, let us start with a discussion of the short and the long-term stability of molecules, atoms and nuclei, in that order.

Explosives are made up from molecules that upon very minor perturbation can decompose and suddenly release some or all its internal energy in an explosion. That is what Alfred Nobel did in synthesizing dynamite for mining engineers. Realizing what its main application had become he endowed the Swedish academy with the means to sponsor the prize, for the benefactors of society through science.

Well, radioactivity is a somewhat analogous phenomenon in atomic nuclei. Let us not invoke quarks and the bosons of the week force. Just protons, neutrons, electrons, and the positron, but not the other anti-particles. The nucleus is a very compact agglomeration of protons, neutrons, electrons, and positrons -- the last one being a newcomer. They are held together by the gluons of the strong nuclear force but the repulsive electrical force between the protons is there to make certain nuclear arrangements unstable. These are the radioactive isotopes.

They have the same number of protons as their stable brothers therefore their atoms will have the same chemical properties and occupy the same place on the periodic table. But their masses will differ, not just because they have a different number of protons and electrons, but because of the phenomenon of beta decay.

This has to do with the fact that even though a neutron can be observed moving freely, just like the proton, it is inherently unstable and will decay in about 12 minutes into a proton and an electron by the process of (minus) beta decay. The reverse decay of a free proton into a neutron and a free positron (beta+) decay has never been observed but it can happen inside certain nuclei, and it is responsible for the appearance of the free positrons in certain radioactive decays and allowed for their experimental discovery in 1932.

The Dirac' s extension of the wave equation to its relativistic form predicted its existence as the electron's anti-particle, as well as the possibility of electron-positron pair annihilation reactions, with the emission of very energetic gamma photons. It is one specific such reaction, involving emission of two photons, that opened the way to practical implementation of a PET scan by providing a clear filtering procedure, as illustrated in Figure 65. We should emphasize that the rest energy of either an electron or a positron, are so much larger than the atomic or molecular energies, that to a very good approximation they can be neglected in the use of the conservation laws as they pertain to annihilation or creation of electrons or positrons.

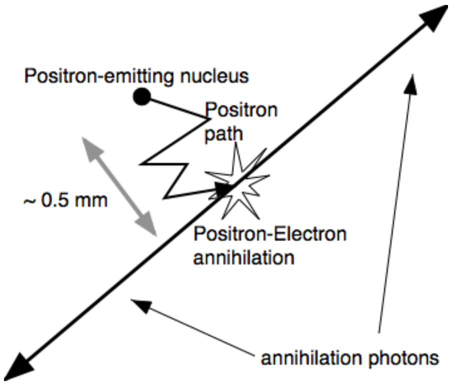


Figure 65. The sensing signal generation process in a PET scan. Pair annihilation transforms the positron into a pair of oppositely directed gamma photons. Suitably positioned and time gated gamma detectors filter out the background of all other such photons.

Certain radioactive nuclei decay by emitting a positron, the positively charged antiparticle of the electron. If such a nucleus is that of an isotope of say nitrogen or carbon, it can be chemically incorporated into the human body via injection or metabolism. So, the positrons will appear in the vicinity of these atoms marking their location on the line connecting the two detectors. Furthermore, if the energies of the two gammas are equal, the location will be at the midpoint of the distance between the two detectors. The positrons will mark the location of the molecules that carry the radioisotopes. This tells us where the carrier molecule is in the sample. As soon as the positron encounters an electron it will combine, disappear generating a pair of gamma ray photons with the combined rest energy of both particles.

The contribution of the kinetic energies, and molecular binding energies to energy conservation can be neglected. By the same argument so can their contribution to momentum conservation. Therefore, the two gammas come off in opposite direction and should be simultaneously detected by a pair of scintillator-photomultiplier combinations properly located. The maximum distance between the event and detection should be short compared to the mean free path of the gamma photons. When a positron emitter isotope is incorporated into molecules of a certain injectable medication that molecule serves as a tracer for that drug through the body. Thus, arose the idea of a Pet scan. Examples of nuclei that can be introduced as tracers into patients are, $_{13}\mathbf{N}$, $_{18}\mathbf{F}$, $_{11}\mathbf{C}$, $_{15}\mathbf{O}$. The location of the molecule for most of the events so detected, as shown in figure 67 below, is midway between the two detectors.

Thus, two detectors facing each other, located across the body to be scanned, will pinpoint the location of the source to the midpoint of the straight line connecting the detectors. This is so because gamma photons, move at the speed of light and rarely get scattered by ordinary matter, that is, if the distance between detectors is shorter than the gamma's mean free path. This immediately suggests the possibility of using this for scanning for the distribution of the tracer molecules in the body. The body can be inserted into a cylindrical tube carrying a distribution of such pairs of gamma detectors time synchronized for simultaneous detection. They would determine the distribution of the tracer molecules inside the body. There is the question of accuracy in timing and position to be answered.

By positioning these detectors in a narrow cylindrical section and translating the patient in successive steps we will have achieved a PET scan. The interpretation of the measurements would go along the same lines as in the case of MRI presented above with the advantage that we have all the data around the section without need of physical rotation. The data, consisting of the spatial coordinates of the midpoints of the lines connecting the detector pairs is direct location. No need for calibration.

The pair annihilation is a high energy process which has no significant background of competing processes with similar filtering conditions of two simultaneous gamma photons in opposite directions and identical energy. As a consequence, it is the most sensitive scanning technique. On the other hand, getting the error limits, which depends on statistics, becomes more problematic.

The alignment between the two photon paths varies depending on the momentum of the electron at the time of the event. The process is controlled by the conservation rules of momentum and energy of all new participants the event created, a replacement electron and the 2 gamma photons. These will participate in Compton scattering whereby momentum and energy will be exchanged among them. In this case we have a relatively complex situation but as an illustration of what happens, the next figure illustrates this balancing act between a single photon and an electron. This is what we call Compton scattering. The exchanges shift slightly the energy, thus the frequency of the photon and the electron's momentum, as indicated in figure 66.

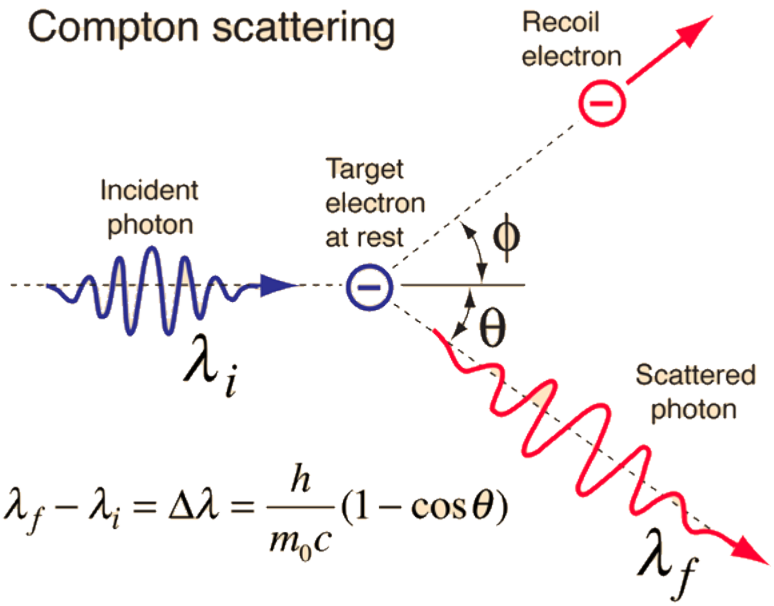


Figure 66. The Compton scattering effect, the major source of inaccuracy in the location determination from PET scans.

Because Compton scattering involves gamma radiation it might affect somewhat location accuracy. One can get an idea of the inaccuracies in the location due to this slightly inelastic scattering by contemplating the diagram and formula in Figure 66. Furthermore, gamma photons have very large energies compared to those of electrons inside molecules.

$$\mathbf{mc_2 >> m_0 v_2}$$

where $\mathbf{m=2m_0}$ is the combined mass of the electron and the positron, v the speed of the electron inside the material and **c** is the speed of light. As the potential energy of the electron is of the same order as its kinetic energy this just means that the energy of the gammas is much, much larger than the intra molecular energy of the disappearing particles, electron and positron. Therefore, the errors committed by neglecting them are small so the location accuracy of this technique, even if not as good as that of MRI is compensated by the absence of background noise. It is especially useful in "full body scans" where the presence of problems that involve multiple organs have to be identified, as for example, heart and lung.

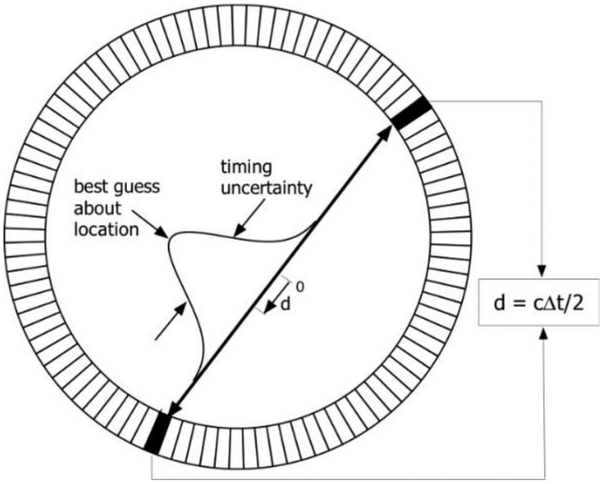


Figure 67. Schematics of the positron annihilation process detected by a ring of scintillator detectors.

The graph indicates the distribution of events near the center of the line connecting the detector pair, thus locating tracer molecules within this circle just by recording pair scintillator outputs as a function of the pair position on the ring. This is similar to the MRI case, without requiring physical rotation.

Section 15: Conclusions

1. A meaningful discussion of quantum mechanics can be carried out mostly without explicit and ugly mathematical language of equations and without the philosophical language needed to explain the unexplainable. A few pictures can help. You can find more of them online.

2. The basic principles of physics have always governed our lives, both individual or collective, whether we knew it or not. Learning about

them enabled the discoveries and developments, that improve, more often than not, the quality of life.

3. Those weird, sometimes crazy physicists actually did a lot of good.

4. The next comment is not really part of the conclusions list to be drawn from the contents of this document. So, forget about the number 4 and read on. In this document the intention was not to demystify, rather to make these mysteries more palatable, by pointing out connections between the apparently diverse mysterious phenomena that actually are part of the same mystery. For example, the electron spin is a mysterious presence in our universe, and it is a basic mystery responsible for the exclusion principle, as well as para magnetism and ferromagnetism. The reality of these connections does not explain the individual mysteries but makes them acceptable, easier to live with so work can go on, fermions and bosons are accepted, and lasers get invented. We are left with the mystery of the exclusion principle and the periodic table, yet new molecules were synthetized, and the human genome was sequenced. The mysteries are all there to take or pick as topics for papers or lectures. The objectives of this document were the connections. And as it discusses only a few, it is reasonable to list missing ones, at least those we could remember. So, here is the list from our memory.

1. Dark matter and its possible particulate nature, wimps. As we have evidence for gravitation waves, and gravitons perhaps we will find evidence for wimps, the quanta of dark matter.

2. The neutrino, it is almost a wimp, but not quite. It has a mass and a spin.

3. Intrinsic handedness (Helicity) of the electron. The weak force and matter-anti-matter interactions.

4. Squids, Cooper pair tunneling and magnetic flux quantization

5. Superconductivity in insulators, high temperature superconductors.

A due acknowledgement

...to all those physicists who created the models and theories and, perhaps even more importantly, imagined the experiments on which the contents of this paper are based. A very few I was fortunate to know; even fewer, if any, are still with us.

Appendix A

Two Scientific developments made the medical application of nuclear magnetic resonance possible, Superconductivity and The Fast Fourier Transform. We try to highlight their basic features in this appendix.

1. SUPERCONDUCTIVITY

At the turn of the 19th to the 20th century as the liquification of gases like oxygen, nitrogen, hydrogen and eventually helium were achieved, the achievable temperatures moved closer to absolute zero. This allowed experimental physicists to measure the resistivity of metals in search of the ultimate lowest resistivity when thermal giggling effects disappear at absolute zero.@

There was a surprise in Leiden, the Netherlands. For some metals the resistivity dropped to zero at a finite temperature.

@ **Liquid Helium** http://www.bibnum.education.fr/sites/default/files/19-onnes-analysis.pdf

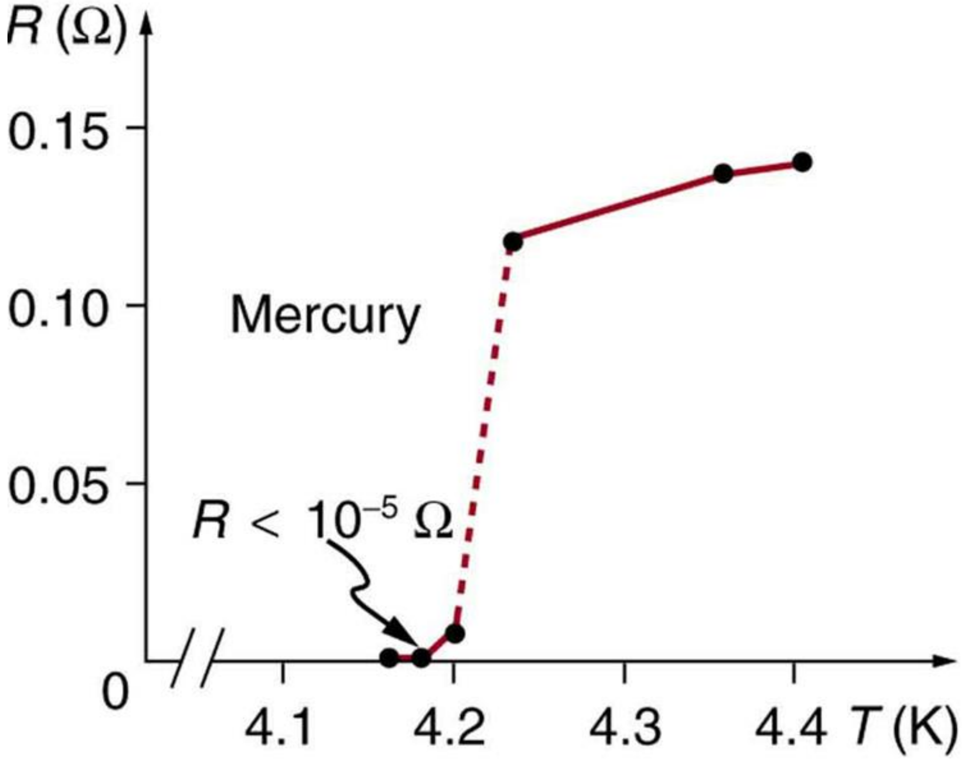


Figure A1. Temperature dependence of the resistance of mercury.@

In a metal like Mercury, if the electrons are subjected to some new attractive interaction mechanism, they can couple and pair up. If this interaction is of sufficiently, long range, the pairing can still manifest itself even if the separation is relatively large. The electron spin and the tendency of opposite spin electrons to pair up into bonding states in which both electrons have the same linear vector momentum, will allow these coupled pairs to carry a current more easily than individual electrons.

Furthermore, considering the pairs as individual particles of zero spin, categorizes them as bosons that can agglomerate into Bose Einstein condensates. If such a condensate would extend along an entire wire loop,

@https://en.wikipedia.org/wiki/History_of_superconductivity the loop could carry a current forever once it starts. These Cooper pairs formed by such an interaction, pair electrons of opposite spin into a state of even spin, 0 or 1, so they are bosons now, not restricted by the exclusion principle.

They can accumulate all in the same state, and into Bose-Einstein condensates. This means that the electrons in these pairs will accumulate in this lowest energy state, separated from the higher energy normal states by the superconductivity energy gap. The sudden appearance of zero resistivity at a finite temperature is evidence of its reality and the transition temperature yields the value of the gap as **kT**, where **k** is the **Boltzmann Constant**.

There are various ifs to overcome to get such an interaction to be credible. This could happen if such a long-range collective interaction could overcome the effect of the thermal giggling, by involving the giggling itself. The technical name of the giggling is "lattice vibrations or phonons". That is exactly what happens in certain metals. As an electron moves, due to its attraction of the core positive charges, it shakes the atomic cores exciting these vibrations. If another electron arrives shortly, it sees the altered configuration of positive charges, therefore interacting with the first electron via atomic giggling.@

This type of interaction, however, is stable only for relatively small amplitude giggling, therefore, it will manifest itself only below a **t**ransition temperature. This transition temperature is material dependent, Again, experiments confirm this situation by measurements on metals made up

@ https://en.wikipedia.org/wiki/BCS_theory of superconductivity
https://en.wikipedia.org/wiki/Cooper_pair

from isotopes of different atomic mass. These coupled electron pairs are the Cooper Pairs.$

If either a constant electric or magnetic field could survive inside a super conductor, it would initiate a continuously increasing current, an unstable situation. So, a stable superconductor, by definition, has to exclude all fields. Therefore, a current carrying superconductor can only exist in the form of a loop where the circuiting current generates a @magnetic field opposing any preexisting magnetic field. By “magnetic field in a superconductor”, we mean that the superconductor has holes filled with normal state material. The smallest such holes are the site of magnetic current loops called magnetic vortices, where the super current curls up around the hole. If a superconductor contains a small enough concentration of non-superconducting impurities the superconducting state will survive with a negligible fraction of the supercurrent being lost to curling.

The only way that we can establish a circulating current in a superconductor wire loop is by a transient process. With the loop in the normal state, above the transition temperature, establish a current with say a battery or a voltage-controlled power generator. Next, lower the temperature through the transition, as slowly as possible. Then, as soon as the current reaches the desired value cut off the power source. Obviously, this is rather tricky.

$https://en.wikipedia.org/wiki/Cooper_pair
@https://courses.physics.ucsd.edu/2018/Fall/physics211a/topic/london.pdf
Magnetic fields and superconductors The London equations
https://www.feynmanlectures.caltech.edu/III_21.html
In Vol III

However, there are certain alloys and intermetallic compounds where the supercurrent survives the introduction of the foreign atoms even if they assemble into finite regions of normal matter. Not only that, but their transition temperature is considerably higher than that of conventional superconductors. They basically can, at the same time, support a supercurrent flowing in certain filamentary regions of the material, that extends end to end of the wire, even when the temperature is above the critical temperature, these filamentary regions will narrow up as the temperature increases, until it reaches a considerably higher transition value, where these filaments disappear.

A niobium tin **Nb_3Sn** intermetallic compound is used in winding most MRI magnets. These are the type II superconductors, vital for the practical use of superconducting coils to generate the large magnetic fields over large volume, that make MRI a useful medical tool.

The magnetic field trapped within such a hole, when integrated over its area, yield the magnetic flux through it, in analogy with a fluid flow. Theoretical and dimensional analysis considerations indicate that this flux is quantized but this suggests a very difficult experiment.

The only way to include a magnetic field is above the transition temperature and then cool it down in a controlled manner. However, by using a superconducting wire loop into which a thin section of non-superconducting wire has been spliced, which will still carry a super current, via tunneling through the thin resistive layer, the confirming experiments can and have been done repeatedly. Incidentally, tunneling is a quantum mechanical property of the electron’s wavefunction.

The process of establishing the magnetic field in a winding solenoid is a tricky operation. It involves having a segment of the winding wire held above the transition temperature to limit the current and protect the current generator against a sudden short as the temperature is lowered through the transition region.

The difficulty is attenuated by the use of type II superconductors which undergo a much more gradual transition. What happens is that just above the transition region the whole metal is in the normal state. As you slowly lower the temperature through the transition region, part of the metal turns superconducting forming filamentary regions that carry the super current in adjacent vortices that form the supercurrent filament. The vortices are formed around islands of normal material whose size goes to zero at the end of the transition when the whole material becomes superconducting and the possible residual magnetic field and the accompanying flux quanta have been expelled.[@]

2. The Fast Fourier Transform

The reason we should mention this strictly computational technique in this appendix, is its vital importance in making possible the application of **NMR**, as well as many other procedures, originally created in physics, to diagnostic techniques, now commonly used in medicine. This algorithm was invented by mathematicians over a century ago, to facilitate paper and pencil calculations, but its sequence of operations fits perfectly into programing sequences for digital computation.

Just contemplate the jargon, **CT, PET, MRI…** and all the large-scale computations involved in gene therapy and cancer research.

[@] https://en.wikipedia.org/wiki/Alexei_Alexeyevich_Abrikosov

Any periodic function can be represented by an infinite but countable sum of certain simple periodic functions. FFT solves the slow convergence of such sums by bypassing intermediate terms and filling in the holes by smoothing algorithms. The Fourier representation of periodic functions by sines and cosines is structurally a digital process because it involves a COUNTABLE number of known functions, each associated with an integer. This is just a Fourier series representation of a periodic function, as discussed in connection with confined waves.

Appendix B

Explanatory comments on the more-opaque or less-familiar topics discussed in the text.

1. The product of two vectors.

There are two ways to calculate the product of two vectors A and B.

A. The ***scalar product,*** or ***dot product,*** is a number calculated as **A** . **B** = **a** * **b** * **cosϕ** where **ϕ** is the angle between **A** and **B, a** is the magnitude of **A**, **and b** the magnitude of **B**.

It produces the signed magnitude of the projection of **A** on **B** or **B** on **A.** If the vectors are normalized to unit length, the result is just the cosine of the angle between them.

B. Unlike the scalar product which generates a single number, the vector product, or cross product, results in a new vector **C** = **A** X **B**, of magnitude **a * b * sinϕ** but with direction depending on the ordering of the factor vectors. The resulting vector is always perpendicular (normal) to the plane formed by the two vectors. As written, **A, B** and **C** directions follow the right-handed rule but if **C** = **B** X **A** the left-handed rule applies. These are axial vectors associated with rotation rather than translation (polar vectors or normal vectors.) If you invert the coordinate axes the components of a polar vector change sign but that is not the case of an axial vector. This is strictly a property of three-dimensional Euclidian space. The angular velocity, the angular momentum, the torque and the Lorentz force are all axial vectors.

In a way we can think of a vector product as a matrix product of a 3-element line array with a 3-element column array with signs reversed yielding a 3 by 3 square skew symmetric matrix with zero diagonal elements, with only 3 independent off-diagonal elements being the 3 cartesian components of the axial vector.

Algebraically, the cross product of a vector (x_1,y_1,z_1) with another vector (x_2,y_2,z_2) is the new vector (X,Y,Z) calculated as:

$$\begin{aligned} X &= y_1 * z_2 - z_1 * y_2 \\ Y &= z_1 * x_2 - x_1 * z_2 \\ Z &= x_1 * y_2 - y_1 * x_2 \end{aligned}$$

2. Complex Exponentials and trigonometry[@]

We introduce here the complex notation for trigonometric functions which simplifies the discussion that follows. It is applicable to harmonic waves such as radiation and sound and a large category of quantum waves such as those describing stationary states which have a harmonic time dependence.

[@] https://web.mit.edu/22.09/ClassHandouts/Charged%20Particle%20Accel/CHAP12.PDF

Define **$I = (-1)^{1/2}$,** the square root of -1, as the imaginary unit.
A complex number **s** is then defined as the linear combination **$s = a + ib$**, a being the real part, **b** the imaginary part. The rules of arithmetic operations are the same as those for real numbers, namely**:**

$$(a+ib) + (c+id) = (a+c) + i(b+d)$$

and

$$(a+ib)\,(c+id) = (ac-bd) + i(ad+bc)$$

and, by definition:

$$i*i=-1.$$

The basic rule for exponentiation of real numbers is to add the exponents when multiplying:

$$(e^{a} * e^{b}) = e^{a+b} .$$

The extension of this rule to complex numbers reads:

$$(e^{a+ib}) * (e^{c+id}) = e^{a+c+i(b+d)}.$$

If we define the sine and cosine functions arbitrarily by the Euler relation:

$$e^{iw} = \sin w + i\cos w \text{ then } e^{-iw} = \sin w - i\cos w$$

we can solve for the trigonometric functions in term of the two exponentials.

$$\sin w = (e^{iw} + e^{-iw})/2$$
$$\cos w = (e^{iw} - e^{-iw})/2i$$

This mathematical trick also has physical meaning for waves as they relate to circular motion as we discussed in section 2 of the paper. So**,** we now can write the wavefunction in exponential notation this function will contain both components.

This exponential notation is extremely convenient when dealing with rotations in a plane and in general, circular motion in polar coordinates. However, its usefulness goes far beyond that. Remember rotations in two dimensions commute but not in three dimensions. However, when transforming position, and keeping the radial distance uniform, any rotation in three dimensions can be composed by two sequential plane rotations if we shift the direction of the axis by a finite amount. Let us consider the situation of a solid object undergoing continuous rotation with a continuous shifting of the axis direction. This is equivalent to a continuous sequence of differential plane rotations, but which allows the immediate recognition of parameters such as instantaneous radius of curvature and angular velocity or angular momentum for that matter. This is analogous to the way we define a slope or the derivative for a curve. These physical notions are no doubt there in the language of matrices and tensors but for a large fraction of the older audience I think this inclusion would be helpful.

Furthermore, complex wavefunctions are part of the jargon of quantum mechanics, as well as the classical treatment of all types of oscillatory processes. That includes electric current resonances in LC circuits, mechanical resonances in spring connected mass arrays, the plucked springs of confined vibrations, sound waves in organ pipes, and even chemical kinetics. However complex algebra is no more than a method to deal with variables and functions that include frequency and phase in their physical nature

3. Regarding Eigenfunctions and Wave Equations

Applying Newton's second law to a simple harmonic oscillator obeying Hooke’s law we get

$$m(d^2x/dt^2) = -Kx$$

m=mass, **K=spring constant** with solution depending only on **K/m**, **x(t)=Csin(f*t) , f=K/m** for ANY VALUE of **C** .

This defines it as a linear equation: the sum of different solutions is also a solution. Specifying the value of **dx/dt** at **t=0** defines **x** by initial condition. Giving the value of **x** at t_1 and t_2 defines **C** by a boundary condition. With this boundary condition we can rephrase solving Newtons equation as an eigen value problem

(dx(t) /dt)=-fx is the equation.

x(t)=Csin(f*t) will solve the equation for any value of **C**. The boundary conditions x(t1)=x1, x(t2)=x2 generate the pair of simultaneous equations

$$x(t1)=C\sin(f*t1),\quad x(t2)=C\sin(f*t2)$$

which can only be solved if **sin(f*t1)= sin(f*t2)** which solves for the eigenvalues of the frequency **f** in terms of the time interval $t2-t_1$.

For a harmonic wave we have to deal with two variables **x** and **t**, and two partial derivative operators that operate on the same original function of two variables. Therefore, two new functions of the same two variables will be generated as possible solutions:

$$\mathbf{df(x,t)dt = C_x{*}f(x,t)} \quad \text{and} \quad \mathbf{df(x,t)dx = C_t{*}f(x,t)}$$

where the two constants are to be determined by a pair of boundary conditions at (x_1,t_1) and $(x_2,t_2.)$

Again by using the linearity argument we can determine the two coefficients by solving now a pair of linear equations. The boundary conditions so stated actually determine the value of **f(x,t)** on a rectangular boundary line consisting of alternate straight segments parallel to the **x** and **t** coordinate axes. In this case the derivatives mean partial derivatives and the meaning is given by context as the functions are multivariate.

In general, the boundary conditions are specified on some more general curve. For two-dimensional space, **x** is replaced by **x** and **y** and the special rectangular boundary becomes an actual curve and the time boundary combined with this new curve becomes a trajectory, the boundary trajectory.

This is how generalizations are cooked up. Now, if we try to generalize further, the spatial part of the boundary becomes a surface where the intuitive concept of a curve as a boundary on a plane does not apply. The boundary is just a time dependent surface enclosing a volume that changes shape as a function of time. On a curve we can choose an infinity of pairs of points. Analogously, on a surface we could choose an infinity of curves, but we better stop here before trouble sets in.

For a wave propagating in one spatial dimension the equations would be:

$$\mathbf{df(x,t)/dx = Lx = (2\pi / \lambda)x,} \text{ and } \mathbf{df(x,t)/dt = -Kt = -\nu t.}$$

The solution of the first equation has to be expressed in terms of a cosine function as **L** is positive. If however we use the complex number language to write the solution using the Euler formula,

$$e_{iw} = \sin w + i\cos w \text{ and } e_{-iw} = \sin w - i\cos w$$

then we can solve for the sine and the cosine functions in terms of the real and imaginary parts of the exponentials:

$$\sin w = (\mathrm{Re}^{iw} + \mathrm{Re}^{-iw})/2, \quad \cos w = (\mathrm{Ime}^{iw} - \mathrm{Ime}^{-iw})/(2i),$$
$$w = Lx \text{ or } -Kt \text{ for sin or cos.}$$

Actually the sine and the cosine are the same function, and if we think in terms of the Euler relation, and the product rule for exponentials the following complex exponential, w, has all the features of a two variable wave function,

$$w(x,t) = Ce^{\,i(Lx-Kt)} = C(e^{\,i(Lx)})*(e^{\,(-Kt)})$$

where now **C** is an arbitrary constant that may be complex. The pair of differential equations for this wave function are now:

$$dw(x,t)/dt = (e^{\,(-Kt)})C_t w(x,t) \quad \text{and} \quad dw(x,t)/dx = C_x(e^{\,i(Lx)})$$

with two sets of boundary conditions, as discussed above.

The operation of taking the derivative of a function generates a new function, so we label **d/dt** or **d/dx** as operators that change a function into another function, its derivative.

Functions themselves are multiplicative operators, as **$f(x)*w(x)=g(x)$**, a different new function. This concept applies to partial derivatives as well. We label the pair of partial derivatives of a function of 2 variables a two-dimensional vector operator **$\{\partial/\partial x, \partial/\partial y\}$** operating on functions of two variables yielding a vector pair of different functions **$\{\partial g(x,y)/\partial x, \partial g(x,y)/\partial y)\}=\text{grad } g$**, the vector gradient of the scalar function **g**.

The generalization to three space dimensions is natural. The gradient operator represents the slopes, along which the electron can move without external force applied. It is taken as representing velocity as an observable. So, the second derivative is the acceleration that requires application of a force. The source of the force is the potential energy V of the wave function.

The term that has the second derivative is taken as the kinetic energy operator **T**. So, **T+V** is the total energy, just as in classical mechanics. The next statement completes definition of the Schrodinger equation in operator language.**Ψ Y**

The total energy operator applied to spatial coordinates is equivalent to the time derivative operator applied to the time coordinate

$$((-\hbar^2/2m)\nabla^2 + V)\Psi = i\hbar\partial\Psi/\partial t$$

where $\nabla^2 = (\partial^2\Psi/\partial x^2+\partial^2\Psi/\partial y^2+\partial^2\Psi/\partial z^2)$,

$\hbar=h/2\pi$ (h being Planck's constant)

Equation 1. Time dependent Schrodinger equation.

The inverted delta means the sum of the 3 spatial second derivatives.

For stationary states, the time dependence of the wavefunction is that of a complex exponential:

$$\Psi\ (x,y,z)e^{i\omega t}\ \text{As}\ /\partial\Psi\ (x,y,z)e^{i\omega t}/\partial t = i\omega\Psi\ (x,y,z)e^{i\omega t}$$

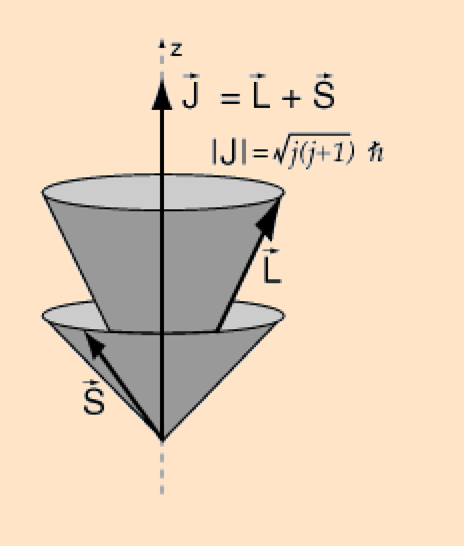
The time dependent equation becomes a time independent eigenvalue equation with stationary state eigen function solutions paired with eigenvalues being the angular frequencies or energies by Planck's relation. The angular velocity ω in radians/sec is related to frequency ν by $\omega= \nu/2\pi$ and frequency is related to energy $E=h\nu$.

Applying the first derivatives in sequence keeps generating multiple combinations of functions that form new tensors rather than vector functions. However, if we only apply derivatives with respect to the same variables used in the first round, i.e. $\{d^2f/dx^2,\ d^2y/dy^2\}$ and so on, we keep generating two-dimensional vector functions.

However if we add the two vector components at this point we obtain a new operator called the divergence, which produces a scalar out of a vector after producing the second derivative vector pair. The time independent version is displayed bellow and the various terms actually operate on the wavefunction Ψ as operators transforming a function into another function either by differentiation or by multiplication.

Note that two multiplications always commute. Multiplication never commutes with differentiation on a common variable. This is at the root of the Heisenberg commutation relations. But two differentiations on a common variable do commute.

4. Angular momentum in Quantum Mechanics



The angular momentum in quantum mechanics is fundamentally quantized in terms of its component or projection along a certain axis.

When it is parallel to this axis the projection should be equal to its magnitude, therefore its magnitude should also be quantized. As implied by the cyclic confinement required by De Broglie the resulting requirement of compatibility imposes quantum restrictions on possible magnitude values.

The expressions for the components of the angular momentum are products of coordinates and their derivatives, they do not commute with each other. Therefore, only one component can be measured, and the magnitude of the angular momentum has to be inferred from measurements of the possible values of this single component. The picture illustrates the meaning of all this. If $\mathbf{L}$**,** the angular momentum would be exactly parallel to the axis its projection would give its magnitude. this would be incorrect as this would not be an integer-only for certain angles. It will be an integer-only for certain angles, the minimum angle yielding the quantum mechanical magnitude of $\mathbf{L}$**.**

The Physical interpretation suggested by the figure is that random magnetic fields from neighbor particles exert torques that put the angular momentum into precession. There is one conserved quantity that is inherently quantized into packets. The quantization of momentum whether linear or angular is really a consequence of the quantization of energy combined with boundary conditions. This is true as well of the quantization of magnetic flux in superconductors.

The classical angular momentum of a particle moving in a circle is an axial vector $\mathbf{L}=\mathbf{r} \text{ X } \mathbf{mv}$ where $\mathbf{v}$ is the instantaneous tangential velocity (Remember the Lorentz force.) If the magnitude of $\mathbf{v}$ is constant, the angular momentum of mass $\mathbf{m}$ on this circular trajectory will be a vector of constant magnitude $\mathbf{rmv=mr*r\omega}$**,** $\boldsymbol{\omega}$ being the angular velocity in radians/second.

It also defines the angular momentum per unit mass and unit area, within a factor of $\boldsymbol{\pi}$ **.** Define $\mathbf{h=h/2*p}$**,** it corresponds to $\mathbf{h}$ but for angular frequencies $\mathbf{E=h*\omega}$**.** Thus the quantization of energy leads to the quantization of angular rotation around a given axis in packets of size $\boldsymbol{\pi}$ **.**

The whole issue of the angular momentum resurfaces in the section on NMR, especially in connection with the induction detection method and the transition from the quantum to the classical picture on Page 129.

The absorption method, including its saturation can be well understood in quantum terms. It is basically a low photon rate experiment. The induction method starts by the application of a very intense pulse of **RF** radiation, obviously a high photon rate phenomenon well suited for a classical analysis. This very strong **RF** current generates an **RF** magnetic axial field in the coil wound around the sample which is inhabited by billions of protons, each with its own spin magnetic moment. Before the application of the **RF** pulse, the spins will be processing approximately in unison around the constant large and uniform magnetic field, at the Larmor frequency corresponding to the field value. **RF** field at the Larmor frequency will alternate at this frequency. This alternating field can be decomposed into two counter-rotating circularly polarized components and one will be, of necessity resonant, in phase with the precessing spins. This is equivalent to saying that the macroscopic sum of all spins is precessing around the field, which is a classical statement. This picture is only correct for time intervals shorter than either **T1** or **T2**. Beyond that time the synchronism of the rotations will be broken by spin-spin interactions and eventually by thermal fluctuation. But during this time the macroscopic concepts of precession and the rotating frame can be assumed to apply.

5. Diamagnetism, Para magnetism and Ferromagnetism

Diamagnetism is the magnetism generated by the motion of electric charges. Therefore, it depends on the reference system used for its description. Para and Ferromagnetism are the magnetism generated by an intrinsic property of the electron and other elementary particles which is labeled the spin. It mimics the properties of the diamagnetic field described by Maxwell's equations in that it interacts with electric currents and moving charges, described by the Lorentz force but its source cannot be attributed to electric currents. It is the consequence of the property of spin degeneracy, a strictly quantum mechanical property where elementary particles behave as miniature permanent magnets.

6. The idea of the magnetic dipole, a false analogy.

By using Maxwell's equations, the electric field lines around a nearby pair of oppositely charged particles can be calculated. The result is shown on Figure 47. It is astonishingly similar to the picture of the magnetic field lines around a bar magnet obtained by an iron filing experiment. This similarity created the idea that there are such things as magnetic charges of opposite charge, the north and the south, and they are located at the opposite ends of the bar magnet, thus the name, the magnetic dipole. Actually, despite Maxwell, this idea persisted for a century, to the point that I was taught to think about magnetism in this manner in middle school.

Epilogue

As the ancient statement says, "Everything that goes around, comes around." Quantum Mechanics is back, as it were, in our everyday life, at least in terms of language and popularity. It’s behind every discussion of neutron stars, the photons of the cosmic microwave background, a residue of the big bang, and the protons of the cosmic rays. As we said in the beginning of the paper, the practice of physics is the making of connections. This is a very long-distance connection that physics provides.

The End