What is *Second Quantization* and where does it fit in the overall scheme of Physics? A foray into the history of the early days of Quantum Mechanics and Quantum Field Theory

Wagner L. Truppel

December 15, 2005

Second quantization, which amounts essentially to the quantization of *field* degrees of freedom, as opposed to *particle* ones, is an odd topic of sorts in the Physics curriculum. Perhaps because it's considered a bit too advanced, it's usually not covered in the undergraduate curriculum and, when taught at the graduate level, little time, if any, is spent on its conceptual underpinnings, or on the history behind it.[1] It's also sufficiently different when compared to the "regular" quantum mechanics encountered at the undergraduate level that, by the time it's studied, it's found to be rather obscure. Yet, it has a fascinating history and an important place in our modern understanding of quantum physics, and is at the very core of our best understanding to date of the structure of matter.

The best books on quantum field theory at the introductory level[2] all cover second quantization, but not a single one of those I know of spends more than a few paragraphs on its history. This essay is an attempt to remedy that, although it ended up going much deeper than I originally intended. I should warn the reader, however, that there may be historical inaccuracies here and there since, with the exception of biographical information (which I searched the web for), all of the text is the result of my own understanding of what went on. I'm no historian of science and do not intend this essay to be of professional historical quality. The scientific content, on the other hand, is accurate, to the extent of my own understanding, of course.

[1] It seems to be unfashionable to discuss the history of science in a graduate-level course. But if a subject is too advanced for the undergraduate curriculum and is, therefore, ignored at that stage, shouldn't its underlying conceptual context and history be given at least some time at the graduate level?

[2] My favorites are *Quantum Field Theory*, by Mandl and Shaw, and *Advanced Quantum Mechanics*, by Sakurai.

Classical mechanics of particles and waves

Ironically, the roots of the quantum theory of fields go back all the way to the time of Newton, that is, to the *classical* theory of fields, although the notion of *fields*, as it's known in modern-day Physics, didn't quite exist back then. Back then, there were only *waves* and, of course, *particles*.

So I'll start there, in the latter half of the 17th century, when Newtonian mechanics perfectly explained almost all phenomena known at the time, namely, all particle motion, and also all wave motion, except the motion of *light*. Newton himself took a stab at optics and was a fierce proponent of the idea that light was made of particles. The competing wave theory, under his opposing weight, didn't gain much credit until 76 years after Newton's death, with the work of Thomas Young, who in 1803 showed that light can produce interference and diffraction patterns, just like water waves and sound waves. Clearly, light could not be a stream of particles.[3]

Controversy aside, the theory of wave motion was as firmly established as the mechanics of particles and was, in fact, "derived" from it since all wave motion understood at the time was the result of some underlying particle motion: sound waves, for instance, are the result of air particles moving closer together, then farther apart, then closer together again, and so on, in an ever-expanding cycle of alternating compressions and rarefactions. The same is true of water waves and waves in elastic solids. And therein lies the distinction between those classical waves and the more modern notion of *fields*; fields do not require an underlying elastic medium to support wave-like motion.[4]

In any case, the prototypical classical wave-like motion is *Simple Harmonic Motion* (SHM). Its ubiquity in classical wave-like motion is the result of any well-behaved potential energy function being very accurately expressed as a parabola near any of its minima. It's not surprising that SHM is of paramount importance in the quantum mechanics of *particles* (which, in hindsight, we might call *first quantization*). After all, elastic media — the supporting material for SHM — is made of atoms, which behave according to quantum mechanics. What *is* surprising, at least at first, is that SHM also plays a fundamental rôle in the quantization of *fields*, which is essentially what *second quantization* is. Or maybe it isn't all that surprising, for it's hinted by de Broglie's wave-particle duality. But I'm getting ahead of myself now. The point is, though, that second quantization is rooted on our old friend, simple harmonic motion. To understand why, we need to look more closely at the classical theory of light and find out why and how it came to be quantized.

[3] Alas...

[4] Then, again, we know now that empty space is not entirely empty, being full of so-called *virtual particles*.

Classical electromagnetism

We now move forward about 200 years, to the time of Maxwell, late 19th century. Classical Newtonian mechanics was then more solidly established than ever, given its successes in predicting astronomical phenomena and its rôle as the foundation of *Statistical Mechanics*, which enabled many of the less formal results of *Thermodynamics* to be derived and understood in more fundamental ways. Additionally, some very fundamental formal developments had been made since Newton left the stage, namely, the *Lagrangian* and *Hamiltonian* formulations of classical mechanics (Lagrange, 1788; Hamilton, 1833). But there was still one pesky little phenomenon left — *light*. Young's experiments of 1801-02 had shown light to have a wave nature but there was no underlying theory as to why that should be the case. And there was also a wealth of "new" phenomena, unknown to Newton, namely, *electricity*, *magnetism*, and *radioactivity*.

Faraday was the first to propose the idea that electrical phenomena might be the result of vibrating "lines" of electric flux connecting electric charges but it was Maxwell who formalized that idea and provided a synthesis of all electric and magnetic experimental results known then. And, as a bonus, all of optics was explained too, for Maxwell theorized that light was essentially an electromagnetic phenomenon, the vibrations of *field* lines which Faraday had thought of, thus providing a theory for the origin of light and settling the controversy in favor of its wave nature once and for all.[5] The experimental verification came with the work of Heinrich Hertz, in 1887, fifteen years after Maxwell's theory was published.

In retrospect, Maxwell's theory had two major theoretical impacts: the first was a change in perspective which now considered *fields* to be entities with real existence, and not merely mathematical constructs. After all, electric and magnetic fields may carry energy, momentum, and angular momentum, and may propagate themselves across empty space, *just as particles can*.[6] Gravitational fields had been studied prior to Maxwell's work, but it wasn't until his theory of electromagnetism became accepted that fields took a life of their own. This new understanding of fields was carried a step further by Einstein in his relativistic theory of gravity (the *General Theory of Relativity*, where the very fabric of space and time come alive) and is also the backbone of today's understanding of particle physics. The second major impact of Maxwell's theory was the realization that perhaps all physical phenomena could be unified into one single theory. Maxwell succeeded in bringing together all of electricity, all of magnetism, and all of optics under one single framework. In the decades that followed, many great minds attempted (unsuccessfully, however) to unify electromagnetism with gravity and, later, with considerable more success, other interactions

[5] Or so it seemed...

[6] Yet another clue to the dual nature of particles and waves in general, and of light in particular.

as well. Even today's grand unification attempts have their roots in Maxwell's theory.[7]

However, Maxwell's theory wasn't accepted right away. It was bold and with new mathematical ideas which were somewhat unfamiliar to the physicists of the time. Its biggest barrier was that it predicted the seemingly nonsensical idea that light was a wave-like phenomenon *needing no elastic medium to support its motion.* Even Maxwell himself tried to work out mechanical models for a medium through which light would propagate. Such models for the so-called *æther* were artificial and complicated, and couldn't convincingly explain some simple facts, such as why the Earth didn't seem to drag the æther while in its motion around the Sun. The resolution of that issue had to wait until 1905, with the advent of Einstein's *Special Theory of Relativity*, which altogether dropped the requirement of an æther and simply accepted the reality that electromagnetic waves need no underlying medium to propagate.

As I've already mentioned, at about the same time another set of new phenomena were being discovered and studied, namely, *radioactivity* (Becquerel, 1896).[8] Apparently, matter isn't as stable as it had been thought so far. Pierre and Marie Curie[9] found that certain elements on the periodic table, such as *uranium* (as well as *polonium* and *radium*, which they discovered), would transmute themselves into other elements and, in so doing, would produce what seemed to be radiation, except that this radiation didn't seem to be electromagnetic in nature. It was Rutherford who clarified the nature of these emanations. He also named them *alpha*, *beta*, and *gamma* rays.[10] Clearly, there was more to the nature of the world than mechanics and electromagnetism.

Meanwhile, much progress had been made in the field of *spectroscopy*. Back in 1666, Newton had already discovered that sunlight passing through a transparent prism is sep-

[7]This is true not just from a conceptual perspective; electromagnetism is the simplest example of the so-called *gauge theories*, which play a prominent rôle in the unification of three of the four known interactions (electromagnetism, weak nuclear force, and strong nuclear force).

[8]The history of how radioactivity was discovered is interesting on its own. In 1895, Wilhelm Roentgen discovered what he dubbed *X-rays* when he accidentally saw the bones of his own hand projected on a screen he had been holding against his experimental apparatus, while studying light discharged inside evacuated tubes. A year later, Henri Becquerel, studying Roentgen's X-rays, accidentally discovered that certain fluorescent minerals can affect photographic plates without first being stimulated by any external source of light. It seems that, in science, luck favors the intelligent and the well-prepared.

[9]Marie Curie has the distinction of being the first woman to win a Nobel prize, as well as the first person and only woman to have won it *twice*, once in Physics (1903, shared with Pierre and Becquerel) and again in Chemistry (1911, for her discovery and isolation of *radium*). Following Pierre's death in 1906 (run over by a horse-drawn wagon), she replaced him as the Professor of General Physics in the Faculty of Sciences at Sorbonne, the first time a woman held that position. Her daughter, Irene, also won the Nobel prize in Chemistry, in 1935, for the discovery of new radioactive elements. Marie was also a fervent activist for human rights. Talk about a rôle model for women in the sciences!

[10]Rutherford later showed that *alpha* rays were positively charged and were as heavy as *helium* atoms (they are, in fact, *helium* atoms stripped off of their electrons). *Beta* rays are just electrons (and positrons), and *gamma* rays are electromagnetic radiation. So are X-rays, by the way, but of smaller frequencies.

arated into beams of distinct colors, the so-called *spectrum* of sunlight. In 1800, William Herschel studied the heating effects of different colors of light by placing thermometers in the path of different colors. As a control, he also had a thermometer placed in the dark region next to the red band, but that thermometer registered a *higher* temperature than the illuminated ones: Herschel had accidentally discovered *infrared* radiation. In 1801, Young discovered that light can produce interference and diffraction patterns. In 1802, William Wollaston discovered the existence of dark lines in the spectrum of sunlight. In the years between 1814 and 1823, Joseph von Fraunhofer performed a detailed study of the spectrum of sunlight, after accidentally discovering the absorption D-line of *sodium*. In 1819, Augustin Fresnel won a contest sponsored by the French Academy of Sciences on the subject of the wave theory of light.[11] Herschel's son and a collaborator found, in 1826, that each substance, when heated, gives off its own characteristic set of spectral lines. In 1832, David Brewster (of "Brewster's angle" fame) suggested that the dark lines observed in the spectrum of sunlight might be due to the selective absorption of spectrum lines emitted by other substances. In 1840, Herschel's son discovered that the lines Fraunhofer found in the spectrum of sunlight also existed in the infrared region (which was discovered by his father, as already pointed out). In 1842, Becquerel's father extended the sunlight spectrum into the ultraviolet. In 1871, Anders Jonas Angström measured the wavelengths of the four visible lines in the spectrum of *hydrogen*. In 1885, Johann Jacob Balmer found a simple formula relating the wavelengths of those lines found by Angström, henceforth known as the *Balmer series*.

And here's where Maxwell's theory of electromagnetism, and its extension to optics and the nature of light, found its second barrier to widespread acceptance. Maxwell's electromagnetism predicted that light would be emitted and absorbed as multiples of well-defined frequencies, much like the vibrations of a string in a musical instrument produce notes that are harmonics of the string's fundamental frequency. Clearly, a wave theory of light based on Maxwell's theory was in contradiction with various observed spectra. For instance, the lines in the Balmer series of *hydrogen* did not have frequencies that were multiples of a common frequency.

As if that wasn't enough, much trouble was also lurking in the study of the thermal properties of hollow cavities filled with radiation. Such cavities had been shown by Gustav Kirchhoff to be good models for the study of the radiation emitted by heated bodies, known by the odd name of *black-body radiation*. It's a well-known empirical fact that the

[11] This is another piece of science history that is interesting on its own. Fresnel's 135-page mathematical essay led one of the judges, the famous French mathematician Siméon Poisson, who was a firm believer of Newton's corpuscular theory of light, to predict the bizarre result that the shadow cast by an opaque circular disk should, nevertheless, have a bright spot at its center. Poisson thought that this clearly indicated that Fresnel's theory was wrong. François Arago, the committee chairman, performed the experiment suggested by Poisson and, lo and behold, a bright spot was observed at the center of the shadow. That settled it; light was a wave, and Fresnel won the contest.

frequency of the light emitted most prominently by a heated body increases with the body's temperature. For instance, a low-temperature body will emit light mostly in the infrared part of the spectrum, while a piece of charcoal in a grill presents a reddish appearance, and a welding tool glows brightly white. In fact, an empirical linear relationship between the frequency of the most intense component of light emitted and the temperature of the radiating body was already known by 1893, the so-called *Wien's Displacement Law*. Yet, all theoretical attemps to explain this result, and others related to black-body radiation, had failed. The Rayleigh-Jeans model, in particular, predicted that the amount of radiation emitted by a heated body should be *infinite*, since there are an infinite number of oscillation modes in which the light emitted can oscillate and since each mode carries the same amount of energy, proportional to the body's temperature, according to the *Law of Equipartition of Energy*, a basic result derived from *Statistical Mechanics*. This was known as the *ultraviolet catastrophe* and was a major outstanding problem in the physics of the end of the 19th century. After all, heated bodies most definitely do not radiate infinite amounts of energy.

Another source of trouble for the theoretical physicists of the time was the so-called *photoelectric effect*: shine light on a metal and off comes a stream of electrons, ejected from the metal's surface. Since electromagnetic waves were, well, waves, the energy they carried had to be distributed in space. As a result, in order to transfer enough energy to an electron in a metal for it to be ejected, a significant amount of time — of the order of minutes — had to elapse. Yet, all experimental evidence pointed to an instantaneous effect. Moreover, increasing the *intensity* of the light shining on the metal only increased the number of ejected electrons, but not the energies with which they were ejected. Since the energy carried by a wave is proportional to the wave's intensity, where did all the extra energy go? Additionally, increasing the *frequency* of the light being shone *did* increase the energies of the ejected electrons. Go figure...

New century, revolutionary new physics

With the exception of radioactivity, the outstanding problems at the turn of the century were all directly related to the nature of light. The very existence of light spectra, the fact that they followed certain regularities (such as the Balmer series), but not the ones expected on the basis of Mawxell's theory, the thermal properties of the light emitted by heated bodies, and the surprising properties of the photoelectric effect were all unaccounted for by the theories of the time. Yet, the official theory of the origin of light at the time — Maxwell's theory — had been shown to be accurate, by the experiments of Hertz and others, and by a wealth of experimental results not directly related to light but involving electric and magnetic systems. There was also the issue of the æther, whether or not it really existed and, if it did, what it was made of and what its mechanical and electromagnetic properties were.

The physics of light was clearly in the middle of a crisis. The resolution of this crisis started with Max Planck's solution to the ultraviolet catastrophe, in 1900, but was much more heavily influenced, in its early stages, by Albert Einstein.

Planck had discovered that he could easily fit the empirically observed distribution of energy in the black-body radiation of a hollow cavity, as a function of frequency, if only he assumed that different modes of the oscillating electric and magnetic fields inside the cavity did not have the same equilibrium energy but, instead, an amount of energy proportional to their frequencies. He then labored heavily to try and understand why that was the case and concluded that, for some reason he could not fathom, the emission and absorption processes that bring the walls of the cavity in thermal equilibrium with the radiation inside of it work in 'jumps,' that is, that the energy absorbed or released by the wall when interacting with light of a given frequency could not have an arbitrary value, but had to be proportional to the frequency in question. That assumption solved the ultraviolet catastrophe because, even though there are modes of every frequency (thus, an infinite number of modes), the higher-frequency modes now demand more energy than the lower-frequency ones. However, since the available thermal energy per mode is finite ($kT/2$, as mandated by the equipartition theorem of Statistical Mechanics), only so many modes can have their requirements satisfied, and higher-frequency modes (extending all the way to infinite-frequency ones) are therefore not excited at all. As a result, the cavity — and any black-body it models — radiates a finite amount of energy.

Planck's work still left unanswered the question of why the absorption and emission processes taking place on the walls of the cavity caused a quantization of the energy absorbed or released. An answer to that, though not the final word on the subject, had to wait for Einstein's 1905 paper on the photoelectric effect.

1905 is known in Physics as the *Miraculous Year* for it was then that Einstein, in four separate papers, laid out the foundations to two new branches of Physics (*Relativity* and *Quantum Mechanics*) and helped solidify the acceptance of two others (Maxwell's electromagnetism and *Statistical Mechanics*). I think it's fair to say that Einstein was a genius by any measure.

His paper on a mathematical model of the so-called *Brownian Motion* helped put to rest the last misgivings that anyone might have had about Statistical Mechanics and the atomistic nature of matter. And misgivings there had been. In fact, Ludwig Boltzmann, the primary architect of Statistical Mechanics, had committed suicide over the agony of having his work publicly attacked.

Einstein's 1905 main paper on Relativity[12] used electromagnetism as a framework for a *tour-de-force* analysis of the rôles of symmetry and measurement in physics.[13] As such, it

[12]There was a second one, where he proved the now-famous expression $E = mc^2$.

[13]Quantum mechanics, of course, later added a few twists of its own to these rôles.

was also a clear vote of confidence on Maxwell's theory. Facing the inconsistencies brought about by two distinct principles of relativity, one satisfied by Newtonian mechanics and the other satisfied by Maxwell's theory, and having to choose which theory to 'fix,' Einstein sided with Maxwell. Along the way, he dropped the requirement of a medium for the propagation of light (indirectly solidifying the idea that fields are first-class entities), clarified the rôle of light as a means to propagate information, shattered commonly held notions about space and time, and created a new mechanics to replace Newton's.

Ironically, it was his paper on the photoelectric effect, however, that landed Einstein the Nobel prize in 1921. In this paper, he argued that the empirically observed properties of the photoelectric effect were easily understood if one accepted the idea that light behaved as a particle when it interacted with the electrons of the metal. Borrowing on Planck's idea, he concluded that if light of frequency ν behaves as a particle of energy $E = h\nu$, where h is a constant introduced by Planck in his 1900 paper, then increasing the frequency of the light shone on the metal will indeed increase the ejected electron's energy. Moreover, increasing the light's intensity amounted only to increasing the number of such quanta of light, causing more collisions with electrons in the metal, thereby increasing the number of ejected electrons. And, by being localized in a particle, the energy carried by the light quantum would be instantaneously absorbed by the electron colliding with it, without any detectable time delay.

Einstein's explanation of the photoelectric effect pushed back the question left by Planck's work to an even more fundamental level. No longer are just the emission and absorption processes in the cavity wall in the black-body radiation problem quantized, but the very nature of light seemed to be quantized. Why and how to derive that result from Maxwell's theory were questions that had to wait a little longer to be answered.[14]

Still, it could be argued that light behaved as a particle only when it interacted with electrons bound to atoms. After all, such was the case with electrons on the wall of a cavity and electrons inside of a metal, at the light frequencies involved in the black-body radiation problem and the photoelectric effect. But, in 1923, Arthur Compton published a paper reporting his experiments on the scattering of X-rays by elements with small atomic numbers. He observed a shift in the wavelength of those rays as he tried different scattering angles. Now, X-rays carry considerably more energy than the binding energy of electrons in light elements so, for all practical purposes, the process involved was essentially the scattering of X-rays by free electrons. Moreover, the nucleus, even of a light element, is much too heavy to significantly recoil upon scattering X-rays, so any change in the wavelength of the scattered light was due to the recoil of electrons and, by momentum conservation, implied a recoil of the light quantum itself. This interpretation, of course, leads to the inescapable conclusion that light does behave as a particle.

[14] As we shall see, those questions reside at the core of second quantization.

Now, pause for a moment to consider what these developments implied. First, Newton argued that light was made of particles (mostly because it moves in straight lines and seems to form sharp shadows). Then Young and, later, Fresnel and others, showed that light interferes and difracts like any wave, which contradicted the idea that light is made of particles and lent weight to a wave interpretation. Then Maxwell unifies electricity and magnetism into a theory that is clearly a theory of fields, not of particles, and deduces that light is an electromagnetic phenomenon, thereby settling the issue in favor of the wave nature of light. But then Planck and Einstein argue that light behaves as a particle when interacting on an individual basis. Finally, Compton's results settles the issue in favor of the corpuscular nature. I think it's difficult for physicists today, especially students, to appreciate how schizophrenic physicists of that day must have felt.

The old quantum theory

So, in the first couple of decades of the 20th century, Physics was turned on its head. Einstein had replaced nearly 250 years of solid Newtonian mechanics with a new kind of mechanics (Relativity) that violated everyone's intuition in regards to how time behaved. Equally as troubling, he and others had now reinstated the controversy regarding the nature of light, which had been thought to have been settled on the side of waves with the works of Young, Fresnel, Maxwell, and Hertz, among others.

Meanwhile, trouble was also brewing in the realm of bona-fide particles. Recall that much evidence prior to the turn of the century had been accumulated in favor of an atomistic explanation of matter. Extensive empirical work in Chemistry had led to the conclusion that matter is made of molecules, which are combinations of simpler *elements*, which were slowly identified and catalogued into the so-called *Periodic Table*, so named because it reflected certain recurring chemical properties depending on the position occupied by the various elements. Moreover, theoretical work such as the development of Statistical Mechanics, lent theoretical support to such an atomistic interpretation. It was natural to ask, then, what atoms were made of.

The discovery of the electron, by J.J. Thomson in 1897, plus the acceptance that matter was made of atoms, naturally led to the question of whether atoms are made of electrons. They are, but they're also electrically neutral, whereas the electron is negatively charged. So, in 1906, Thomson proposed his "plum pudding" model of the atom, whereby atoms were spheres of uniformly-distributed positively-charged "stuff" sprinkled with electrons in such a way as to make the atom electrically neutral.

One of Thomson's students, Ernest Rutherford, who had been interested in radioactivity for some time,[15] set out to study the deflection of *alpha* rays by thin slices of matter, in

[15] Recall that he identified three kinds of radioactive emanations, *alpha*, *beta*, and *gamma* rays.

order to verify Thomson's pudding model. The vast majority of his *alpha* particles went straight through his slices of *gold* nearly undeflected but, to his surprise, some of them occasionally bounced *backwards*. Now, *alpha* particles are fairly heavy (recall that they're actually *helium* atoms stripped off of their electrons) and positively charged. They were also quite energetic, since Rutherford had accelerated them with the help of external electric fields. In fact, they were sufficiently heavy and energetic that their collisions with the comparatively much ligher electrons could be completely ignored. So, imagine throwing the analogue of a brick, at high speed, into a thin sheet of paper, and finding out that the brick bounces back. Clearly, whatever atoms were made of, they weren't puddings! Rutherford's conclusion after a detailed analytical treatment of the scattering problem of charged particles by thin slices of matter, assuming different models for the slice itself, was that the atom must be made of a very small and concentrated lump of positively charged material, much heavier than electrons, and surrounded by them, much like our solar system is made of planets surrounding the Sun. In other words, the atom is mostly empty space. Rutherford published his results in 1911.

Rutherford's discovery prompted an exciting time for theoretical physicists at the time. With a fairly well established empirical model for the atom — a positively charged, heavy, and tiny nucleus surrounded by negatively charged electrons — several questions needed explanations:

- Since the nucleus is so small, positively charged, and very heavy compared to electrons, what is it made of? *Alpha* particles, perhaps? Since there are atoms much heavier than *alpha* particles, if nuclei are indeed made of them how can they co-exist in such a tiny volume when their positively-charged nature compells them to fly apart due to their mutual electromagnetic repulsion? Clearly, if nuclei have positively-charged constituents of their own, there must exist a new kind of attractive interaction, much stronger than the electromagnetic interaction between charges — to keep the nucleus together — and very short-ranged, since we don't experience it outside the nucleus.[16] We know today that the nucleus is made of *protons* and *neutrons*, protons being positively charged while neutrons are neutral. The interaction referred to above is then called the *strong nuclear force*, and is sufficiently stronger than the electromagnetic repulsion between protons to keep most nuclei together. Neutrons, by the way, play

[16] The gravitational attraction between those constituents is much too weak to be responsible for the stability of the nucleus, even at such short distances. Since the gravitational attraction and the electrostatic repulsion have the same dependence on distance, the distance factor ends up having no effect when comparing the strengths of those interactions — gravity is completely negligible compared to the electromagnetic interaction. Besides, it's a long-range force while, as pointed out, we need a short-range force to account for the fact that this new interaction is confined to the nuclear volume. *Pop-quiz*: if gravity is negligible compared to electromagnetic forces, why is it that it is the dominant force in the macroscopic scale of bulk matter?

an important rôle in this process of making the nucleus stable.

- Speaking of the stability of nuclei, how can radioactivity be explained? Clearly, some atoms are not as stable as others. Recall that radioactive elements transmute themselves into other elements, emanating *alpha*, *beta*, and *gamma* rays. Since *alpha* rays are particles that might be the constituents of the nucleus, and given the energies released in the radioactive decays, it was an inescapable conclusion that radioactivity is a *nuclear* phenomenon, not one associated with the electrons surrounding the nucleus. We know today that radioactivity is the result of yet another interaction, called the *weak nuclear force*. It's also short-ranged, since it's confined to the nuclear volume.

- Since oppositely charged particles attract themselves, it was clear that electrons were bound to the nucleus by electromagnetic forces. But Maxwell's theory predicted that accelerating electric charges should lose energy by producing electromagnetic radiation, so how could the atom be stable? In fact, a simple calculation showed that electrons should lose all their energy and collapse onto the nucleus almost instantaneously.

- How does light interact with atoms in order to produce the emission and absorption lines that had been observed in the spectra of the various elements? Here, as opposed to radioactivity, the energies involved suggested that the interaction was with the electrons surrounding the nucleus, rather than with the nucleus itself.

The questions raised in the first two items above gave birth to the field of *Nuclear Physics*, which evolved into today's *Particle Physics*, but I want here to focus on the last two items, which fall in the realm of *Atomic Physics*. Recall that Rutherford published his solar model of the atom in 1911. Two years later, Niels Bohr proposed a theoretical explanation which combined equal parts of brilliance and lunacy. By proposing that electrons in an atom were allowed only very specific orbits, and assuming that classical electromagnetism didn't apply to them in those orbits so that electrons wouldn't produce any electromagnetic radiation while in those orbits, Bohr was able to "explain" the stability of atoms and to derive analytical expressions for both the binding energy of *hydrogen* and for its line spectrum, in *perfect* agreement with the Balmer series. The key idea, of electrons being confined to a discrete set of orbits, resulted from Bohr's bold but unjustified requirement that the electron's *angular momentum* in those orbits be an integer multiple of $\hbar$, Planck's old h constant divided by 2π. This quantization of the electron's angular momentum resulted in the quantization of orbital radii and, naturally, a quantization of the electron's energy as well. Then, he argued, an electron would emit or absorb electromagnetic radiation only when jumping from one orbit to another, the amount of electromagnetic energy absorbed or released being equal to the difference in energy between the two orbits involved.

Bohr's model was astonishing. It agreed with all the spectroscopic data available for *hydrogen* and provided a mechanism for the absorption and emission of spectral lines in other elements, provided that one accepted the idea that electrons were subjected to energy quantization, just as light was since Planck and Einstein. But why the quantization of angular momentum? And how to explain that electrons in allowed orbits were excused of having to radiate, despite being accelerated? Was the energy of electrons *really* quantized, as Bohr's model suggested?

In 1914, a year after Bohr published his model, James Franck and Gustav Hertz[17] discovered that *mercury* atoms possess excited states, which could only be explained as *mercury* atoms with their outer electrons in higher orbits, just as the Bohr model had predicted for the *hydrogen* atom. Their results lent a significant amount of credibility to Bohr's theory, but the questions mentioned above remained.

Some theoretical progress was then made towards answering those questions, through the work of Arnold Sommerfeld and others, using the idea of *adiabatic invariants* and some formal results from Hamiltonian mechanics, all of which paved the way for next major chapter in the fascinating history of quantum mechanics.

First quantization

About 10 years after Bohr published his model, in 1923-24, Louis de Broglie[18] made the interesting conjecture that perhaps light wasn't the only entity to display both wave-like and particle-like properties. Perhaps, he suggested, electrons behaved that way too. He argued on the basis of relativity and the light quantum idea, as follows. The energy of a light quantum is $E = pc$, where p is its momentum. This follows from the relativistic expression $E = \sqrt{p^2c^2 + m^2c^4}$, with $m = 0$ as it had to be the case for a particle moving at the speed of light. Using the Planck-Einstein's suggestion that the energy of a quantum of light of frequency ν is $E = h\nu$, he concluded that, for a particle of light, it must be the case that $p = h\nu/c$. But ν/c is the wavelength λ of the associated light wave, so de Broglie concluded that the light quantum's momentum p, a *particle* property, was related to its wavelength λ, a *wave* property, by $p = h/\lambda$. Could it be, he asked, that electrons (and, more generally, all matter) also have an associated wave whose wavelength would be given by $\lambda = h/p$?

This argument might have been considered a mere coincidence if it had not been for the fact that when de Broglie used for p the momentum of an electron in one of the allowed orbits of Bohr's model of the *hydrogen* atom, he obtained a wavelength λ that fit precisely an

[17]Not to be confused with, nor related to, Heinrich Hertz, who verified Maxwell's prediction that light is an electromagnetic phenomenon

[18]Both a physicist and a prince. Yes, a prince!

integer number of times in the circumference of the given orbit. He interpreted this result by concluding that the allowed orbits are those for which the electron's associated wave forms a *standing wave*, much like standing waves in a stretched string. In fact, just as in Bohr's model, standing waves in a stretched string produce a discrete set of frequencies. It just so happens that, in the motion of electrons and particles of light, energy is proportional to frequency so a discrete set of frequencies implies a discrete set of energies, that is, energy quantization, just the result that followed from Bohr's model.[19]

As tantalizing a suggestion as de Broglie's proposition was, it still needed experimental confirmation, like all other theoretical ideas in science. That confirmation came with the results of the so-called *Davisson-Germer experiment*, a mere three years later, in 1927. Their experiment was essentially a diffraction-grating kind of experiment, using a *nickel* crystal for a diffraction grating and a beam of collimated electrons as the incident beam. They then recorded the intensity of electron charge observed at different angles off the grating and observed that the intensity distribution was just like that of a wave, with alternating maxima and minima. de Broglie's *wave-particle duality* conjecture had been proven correct and, as a result, he won the Physics Nobel prize, in 1929.

Let's pause for a moment to take stock of the situation circa 1925. Both light and "regular" matter had been established as possessing both wave and particle properties. There existed already a classical wave theory of light (Maxwell's electromagnetism), but not a quantum one that truly explained why light interacted as a particle whose energy was quantized. On the other hand, there existed a half-baked theory of how electrons interacted in a quantum manner (Bohr's model) but not a classical wave theory describing them. Both problems were being attacked, but I'll focus on the second one here and leave the problem of quantizing the electromagnetic field to the next section (*second quantization*, at last!).

It was in 1925 that Werner Heisenberg published his theory of *Matrix Mechanics*. He started from the empirical result that spectral lines formed a discrete set of frequencies and used Bohr's idea that they corresponded to electronic transitions from one orbit to another, say from orbit i to orbit j, to obtain equations relating discrete sets of quantities of the form ν_{ij} (spectral frequencies), E_{ij} (transition energies), r_{ij} (differences in orbit radii), and so on. All these quantities looked very much like the components of matrices (of infinite order, however), hence the name *matrix mechanics*. It succeeded in reproducing Bohr's results but was quite intricate.

A year later,[20] Erwin Schrödinger, working independently from Heisenberg, published a series of papers that followed up on de Broglie's suggestion that particles should have associated waves. If there were waves, he reasoned, those waves should satisfy a wave equation, just like electromagnetic waves satisfy Maxwell's equations (from which one can,

[19] Incidentally, note how simple harmonic motion is beginning to trickle into the affairs of quantum physics.

[20] 1926, the year the American chemist Gilbert Lewis coined the name *photon* for the quantum of light.

in fact, deduce a wave equation for light). His first attempt at obtaining a wave equation that agreed with relativity failed, however, so he resigned himself to obtaining one that was correct only in the non-relativistic regime of low speeds. At that he triumphed brilliantly and his wave equation reproduced all of Bohr's results, and more, without having to make the same sort of *ad hoc* assumptions. For instance, he showed that the distribution of electric charge in any of Bohr's allowed orbits was, in fact, independent of time and *that* was the reason why electrons in those orbits don't radiate. Moreover, for his wave equation to have sensible solutions, the angular momentum of the particle associated with the wave *had* to be a multiple of $\hbar$ and the particle's energy *had* to be quantized in a particular way. Schrödinger also realized that his *Wave Mechanics*, as it was called, was perfectly equivalent to Heisenberg's Matrix Mechanics.

The contributions from Schrödinger and Heisenberg mark the start of the more formal development of the non-relativistic quantum mechanics of particles. Essentially, it was realized that, due to their wave-like alter-ego, particles don't have well defined classical properties such as position, velocity, momentum, energy, and so on, not in general, that is. Under certain circumstances, particles *can* be in *states* of well defined energy and momentum, for instance, but those states could not simultaneously have well defined positions or certain other dynamical variables. Following the lead from Heisenberg's matrix mechanics, dynamical variables were then interpreted as infinite-dimensional matrices or, as they are now called, *operators*. The fact that certain pairs of dynamical variables cannot have well defined values at the same time was then due to the non-commutativity nature of matrix operators. This result became known as *Heisenberg's Uncertainty Principle*.

As we've seen, the path to a consistent quantum theory of non-relativistic particles was anything but straight. However, once the physical ideas behind the Bohr model and de Broglie's duality were understood, once Schrödinger's wave equation came into existence, and once Heisenberg's interpretation of dynamical variables as matrix operators was in place, much formal progress was made. In particular, a connection to Newtonian mechanics was established, not directly, but through the dressing of Hamilton's theory. For instance, it became relatively simple to quantize a given dynamical problem by first writing down its classical Hamiltonian, then obtaining its canonical variables, and finally by imposing commutators between quantum operators corresponding to classical dynamical variables, using their classical Poisson brackets as guides.

At this point, one might naïvely think that non-relativistic quantum mechanics was considered to be wrapped up. Of course, nature is richer than was thought, as it always is, and it had several more surprises under its cloak, such as the discovery of the electron's intrinsic angular momentum (dubbed *spin*). The existence of spin had already been hinted at by the discovery of the hyperfine structure of spectral lines[21] (Michelson,[22] 1881) and was finally verified unambiguously in 1921 with the *Stern-Gerlach experiment*, which showed that a beam of silver atoms passing through a non-uniform magnetic field is split in *two* sub-beams, rather than being simply widened, which was the classical expectation. It was, in fact, Schrödinger's inability to account for spin that led him to abandon his attempts to obtain a relativistic wave equation for de Broglie's matter waves. The development of a relativistic equation that accounted for spin in a natural way had to wait for Dirac, who succeeded in that task two years after Schrödinger published his equation. In the meantime, Pauli found a workaround that allowed spin to be introduced "by hand" into Schrödinger's equation.

The attentive reader may have noticed that Schrödinger's equation, albeit a wave equation, is not a field equation for electrons or any other particles. The situation is similar to that of Maxwell's equations. Maxwell's equations are field equations for the electromagnetic field *from which* we can derive a wave equation, but they are more fundamental than the wave equation itself. What is the analog of Maxwell's equations for the electron? Reciprocally, what's the analog of Schrödinger's equation for the electromagnetic field? It's not the wave equation that can be obtained from Maxwell's equations, since that wave equation is classical and does not give us the quantum behavior that light exhibits.

The paragraph above suggests that, despite its character as a wave equation, Schrödinger's equation doesn't really describe the quantum behavior of any wave. In other words, none of the quantum mechanics associated with it is really the quantum mechanics of *fields*, but is the quantum mechanics of *particles*.

Second quantization, part I

As I mentioned earlier, Schrödinger had attempted to derive a relativistic wave equation for quantum physics but abandoned his efforts and contented himself with a non-relativistic version instead. In his attempts, however, he discovered an equation that later would be rediscovered by, and named after, Oskar Klein and Walter Gordon. Yet, even Klein and Gordon couldn't make that equation work as a quantum wave equation, even for spinless particles, the reason being that it did not conserve the flow of probability or, equivalently, it didn't conserve the number of particles. The argument goes as follows. It's easy to show that the Schrödinger equation for a spinless particle of mass m leads to a so-called *equation*

[21] Namely, what appeared at first to be single lines were, under more refined inspection, discovered to be multiple lines very close together.

[22] The same Michelson who, in 1887, performed the famous Michelson-Morley experiment which failed to show any signs of the æther that was so much sought after in the wake of Maxwell's theory of light as an electromagnetic phenomenon.

of continuity,

$$\boldsymbol{\nabla} \cdot \boldsymbol{j} + \frac{\partial \rho}{\partial t} = 0\,,$$

where $\rho(\boldsymbol{r}, t) = [\psi(\boldsymbol{r}, t)]^* \, \psi(\boldsymbol{r}, t)$ is the probability density of finding the particle in question near the point $\boldsymbol{r}$ at time t, and $\boldsymbol{j}$ is the associated probability current,

$$\boldsymbol{j} = -\frac{i\hbar}{2m} \left(\psi^* \, \boldsymbol{\nabla} \psi - \psi \, \boldsymbol{\nabla} \psi^* \right).$$

In the equations above, $\psi(\boldsymbol{r}, t)$ is the wave function associated with the particle, evaluated at the point $\boldsymbol{r}$, at time t. The equation of continuity guarantees that the probability of finding the particle in question somewhere doesn't just appear or disappear at any point. Rather, if the probability density increases or decreases at some point, it's because the probability of finding the particle "flowed" to or away from that point, much like a fluid. In essence, it says that a particle can't appear out of nowhere nor disappear all of a sudden; the particle has to *move.*[23]

Now, as it happens, the Klein-Gordon equation also allows for a continuity equation to exist, one where $\boldsymbol{j}$ is the same as for the Schrödinger equation but where

$$\rho = \frac{i\hbar}{2mc^2} \left(\psi^* \, \frac{\partial \psi}{\partial t} - \psi \, \frac{\partial \psi^*}{\partial t} \right) = \frac{E}{mc^2} \, \psi^* \psi \,.^{24}$$

However, since the relativistic expression for the square of the energy of a particle of mass m is $E^2 = p^2 c^2 + m^2 c^4$, we see that ρ may be either positive or negative. But if ρ is supposed to represent a probability density, what meaning should be ascribed to it when it's negative? If anything, what that means is that probability isn't conserved anymore, that is, a particle satisfying the Klein-Gordon equation as its quantum wave equation might simply disappear at some point and suddenly appear somewhere else. The root of that difficulty

[23]Here's the formal proof of that statement. Consider a closed surface S surrounding the particle and evaluate the total probability of finding the particle, at time t, inside the volume V contained in that surface:

$$\mathrm{Prob}(V, t) = \int_V d^3\boldsymbol{r} \, \rho(\boldsymbol{r}, t)\,.$$

Now use the continuity equation and Gauss' theorem to compute the rate at which that probability <u>de</u>creases with time:

$$-\frac{d}{dt}\mathrm{Prob}(V, t) = -\int_V d^3\boldsymbol{r} \, \frac{\partial \rho}{\partial t} = \int_V d^3\boldsymbol{r} \; \boldsymbol{\nabla} \cdot \boldsymbol{j} = \oint_{S(V)} d\boldsymbol{\sigma} \cdot \boldsymbol{j} = \text{outward flux of } \boldsymbol{j} \text{ through } S\,.$$

Thus, the only way for the probability of finding the particle inside the volume to decrease is for the particle to cross the surface, moving out of the volume in question.

[24]The second equality follows from $\psi^* \, i\hbar \, \partial\psi/\partial t$ being the expected energy density of the state with wave function ψ, since $i\hbar \, \partial/\partial t$ is the derivative representation of the Hamiltonian operator.

is literally the root appearing in the expression for the relativistic energy so as long as we try to use $E = \sqrt{p^2c^2 + m^2c^4}$ as the starting point for a wave equation, we are not allowed to interpret ρ as a probability density. That is not to say, however, that the Klein-Gordon, or any other such equation based on $E = \sqrt{p^2c^2 + m^2c^4}$, is useless. They can be useful but we need to interpret ρ as something other than a probability density. For instance, charge density.

All of these developments took place around the time when Schrödinger published his equation, circa 1926. In 1928, however, Paul Dirac succeeded in getting around the problem just discussed in the paragraph above, by proposing a theory as crazy as it was brilliant. The brilliant part was to "factor" the expression $E^2 = p^2c^2 + m^2c^4$ so that it was linear in both E and $\boldsymbol{p}$. Linearity in E would guarantee that the time derivative appearing in the expression for ρ in the case of the Klein-Gordon equation would not appear in the case of Dirac's equation. Linearity in $\boldsymbol{p}$ had to follow that of E, then, to satisfy the relativistic requirement that energy and momentum be treated equally, since they form the components of a *four-vector*.

It's doubtful that Dirac had anticipated the extra bonus that his equation presented him with, namely, *spin*. Yes, the spin of the electron came out naturally from Dirac's equation,[25] with the right value for the so-called *gyromagnetic ratio*, a measure of the electron's magnetic dipole moment, which was directly dependent on the electron spin. However, as successful as his equation was in predicting spin, it failed to solve the negative-energy issue discussed above. In fact, it predicted *four* independent solutions, two of positive energies and two of negative energies, each pair having one solution with spin up and the other with spin down.

So, what's the big deal with negative-energy solutions? Can't they be ignored? The problem is that if there are states of negative energy, then particles of positive energy would gladly give up their extra energy and cascade down to states of lower and lower, more and more negative, energies. In other words, every particle would produce an infinite amount of radiation as they descended into the state of negative infinite energy. Moreover, states of positive energy alone do not form a complete set of states, which is an essential requirement in quantum mechanics, so we can't simply ignore the states of negative energy.

It is at this juncture that the crazy part in Dirac's theory comes in, when he suggested that *all* electronic states of negative energy are already filled. Mind you, there are an infinite number of such states, which means that there must exist an infinitely large, infinitely negatively charged, infinitely massive, *sea* of electrons all around us. Absurd? Not really, if

[25]The essential idea is that the linearization of $E^2 = p^2c^2 + m^2c^4$ can only be accomplished if the wave function (the solution to the equation Dirac was looking for) is no longer a single function, but a set of related functions. That set of functions then behaves in a well-defined manner with respect to rotations of the coordinate system in which the equation is written, and that's what characterizes angular momentum (of which spin is an example).

you consider that a uniform background of anything behaves just like a completely empty background. The math works, by the way, so Dirac wasn't really insane in proposing such radical idea.

Now, why would an electron of positive energy not cascade down and join one of the negative-energy ones in its negative-energy state? After all, you can have more than one particle in the same state, can you not? No, not for an electron. As it turns out, electrons satisfy the so-called *Pauli Exclusion Principle*, which states that no two electrons can share the exact same quantum state. Wolfgang Pauli[26] was led, in 1925, to propose this principle to account for the fact that different atoms have different sizes and different chemical properties. If electrons could all cozy up in the same quantum state, the atoms of all chemical elements would have pretty much the same size and behave pretty much the same way. Now, as it happens, Pauli's Exclusion Principle is more general than he had originally thought, and applies to any particle whose spin is half-integer (that is, 1/2, 3/2, 5/2, etc). Particles of that nature are collectively referred to as *fermions*. It just so happens that this connection between the statistical nature of a particle (whether it likes to cozy up or not) and its spin is very deep and can be proven using very fundamental assumptions about the nature of the world. Incidentally, particles of *integer* spin (0, 1, 2, etc), called *bosons*, do prefer to cozy up in the same quantum state, unlike fermions; the more identical bosons in the same quantum state, the merrier they are.[27]

Now, wild as Dirac's suggestion of a sea of electrons was, it worked. In fact, it worked so well that it allowed him to predict the existence of a particle similar to the electron in every respect, except that it would be *positively* charged. His argument was as follows: you can't add an electron to an already occupied state of the Dirac sea but you can give enough energy to an electron already there to bring it to the realm of positive energies. What happens then is that a *hole*, an empty state of negative energy, appears in the sea. Now, an empty "bubble" in the middle of an otherwise uniform sea of negative charge would stand out as some sort of entity with *positive* charge. Moreover, when other electrons filled the hole, the hole would appear to move around, just as a real particle, of real and positive mass. Thus, for all practical purposes, the holes — which are negative-energy solutions of negative charge — behave as positive-energy solutions of positive charge.

So, the bottom line is that an energetic enough beam of light could excite electrons

[26]Who, at the tender age of 18, wrote a textbook on Einstein's general theory of relativity which is, to this day, considered to be one of the best written accounts of that theory.

[27]Note the use of the word *identical*. Only identical fermions satisfy the Pauli Exclusion Principle and only identical bosons like to share the same quantum state. Thus, for example, electrons and protons, both fermions, are not limited *amongst themselves* by Pauli's Exclusion Principle, although electrons amongst themselves and protons amongst themselves, separately, are. Likewise, photons (the spin-1 particle associated with light) and pions (a spin-0 particle involved in the strong nuclear force) do not like (nor dislike) sharing the same quantum state with one another, although they do, separately, amongst themselves.

filling negative-energy states in the Dirac sea to positive-energy states, leaving behind holes in the sea which behave like positive-energy states of positive-charge. This is the process of *pair-creation*. Similarly, a positive-energy electron could cascade down into an unoccupied negative-energy state in the sea, releasing its positive energy in the form of electromagnetic radiation as it fills the hole. This is the opposite process of *pair anihilation*, whereby an electron and its positive-charge counterpart “collide” and destroy one another, resulting in light. Once again, we see that the requirement of relativistic compliance implies the non-conservation of the number of particles — in both pair creation and pair anihilation, the number of electrons isn’t conserved.

Dirac published his equation in 1928. Four years later, Carl Anderson found experimental evidence of a particle with the same mass as the electron, but with positive charge: Dirac’s prediction was confirmed and the *positron*, the first *anti-particle*, was born.

Now, could the same trick be used for the Klein-Gordon equation, namely, could particles satisfying it as their quantum wave equation, with their negative-energy solutions and all, be part of another infinite sea? No, because the Klein-Gordon equation describes particles of spin 0, which are bosons and, therefore, don’t satisfy Pauli’s Exclusion Principle.

As it turns out, however, it’s possible to reinterpret the negative-energy solutions in such a way as to deal away with the Dirac sea. That interpretation, known as the *Stückelberg-Feynman interpretation*, then applies across the board to every particle, regardless of their spin. It also predicts the existence of anti-particles, so it agrees with all predictions made by Dirac’s sea hypothesis.

Second quantization, part II

So let me pause once again and look at the state of affairs at this point in history. It had become clear that marrying relativity to quantum mechanics necessarily implied the existence of particle states of negative energy, which cannot be ignored and which can be interpreted as *anti*-particle states of *positive* energy. Moreover, particles and their respective anti-particles anihilate themselves upon contact with one another, which is indicative once again that particle numbers aren’t necessarily conserved in relativistic quantum mechanics.

Now, much of the above developments happened while attempting to obtain a relativistic quantum theory of *particles*, in particular a relativistic quantum theory of electrons, but advancements in the quantization of the electromagnetic field were also being made and, in fact, contributed considerably to the understanding of the problems discussed above.

We’ve seen *why* the electromagnetic field must also be quantized: the results from black-body radiation experiments, the photoelectric effect, and the Compton effect clearly indicate that electromagnetic radiation may also behave as a particle of fixed energy.

But *how* do we quantize the electromagnetic field? We don't have a particle mechanics for photons to guide our efforts, as we did for electrons.[28] Instead, we only have a classical *field* theory. If we were to attempt something akin to first quantization, we'd have to write something like Schrödinger's equation, but what potential function should be used? And how are we to define $\boldsymbol{x}$ and $\boldsymbol{p}$ for the electromagnetic field? Moreover, light is as relativistic a phenomenon as it can be and we've seen that relativistic compliance results in non-conservation of particle numbers, which first quantization cannot account for in a natural way.

So, evidently, first quantization doesn't look like the right approach to quantize the electromagnetic field. The solution (second quantization), as shown in the appendices, is to draw on an alternative formulation of the quantum theory of the harmonic oscillator, one which describes states in terms of their *occupation numbers*. Also, even though it doesn't *require* relativity, it's also suitable for the quantization of relativistic fields.

Beyond second quantization

The history of quantum mechanics and quantum field theory doesn't end with second quantization, of course. I could write another tome on what lies *beyond* second quantization, from the Lagrangian formulation of quantum field theory to Feynman diagrams to the so-called *Standard Model*. But enough is enough. Suffice it to say that second quantization isn't important just because it allowed for the quantization of the electromagnetic field to be performed, which finally explained why the photon is what it is, but also — and more importantly — because it opened the door to a more complete understanding of the wave-particle duality, and of its rôle in a fully-relativistic quantum theory of nature. The entire field of particle physics is indebted to second quantization and the modern understanding and unification of the weak, strong, and electromagnetic forces would not have been possible if it wasn't for it.

[28] It's interesting to ponder why that is. Photons have spin 1 and, therefore, are bosons. As such, they *like* to stick together in the same quantum state, producing a *coherent* state which extends to macroscopic scales. Their particle character is, therefore, masked away much like the molecular nature of water is masked away by the great number of water molecules in any macroscopic sample of water, resulting in a fluid that behaves, and which can be described, as a classical *field*. Electrons, on the other hand, carry spin 1/2 and, so, are fermions, which cannot be in the same quantum state. Thus, we don't get to see electronic states of macroscopic size, which is to say, we don't get to see a coherent state of electrons that can be interpreted as a classical field. Moreover, elecrons have mass, which makes their de Broglie wavelength very *small* in ordinary situations (pop-quiz: what about electrons at rest? Their momentum is zero, which corresponds to an *infinite* wavelength. How come we don't observe those as waves?); photons, on the other hand, have zero rest mass, which makes their de Broglie wavelength very *large* for low-energy photons. Since low-energy photons are easier to produce than high-energy ones, it's not a surprise that photons were first observed as waves, while electrons were first observed as particles.

So, in ending this essay, I'll try to address the questions: what *are* particles, waves, and fields, and where does the wave-particle duality fit in quantum field theory?

The modern formulation of relativistic quantum field theory (the next step beyond second quantization) starts with a Lagrangian formulation. Lagrangians aren't interpreted as operators and they don't single out energy, like the Hamiltonian. They are, therefore, relativistic scalars, which makes them suitable for a relativistic theory. It's also easy to include interactions in a Lagrangian formulation and to obtain the equations of motion.[29]

But the Lagrangians in relativistic quantum field theory aren't functions of *particle* properties such as position and velocity. Instead, they're functions of *fields*, which themselves now become variables in the sense of canonical and/or conjugated variables. Position and time, on the other hand, are now simply parameters on which the fields depend. These fields then obey, in the absence of interactions, the field equations that we know and love: Maxwell's equations in free space (the equations of motion for photons), Dirac's equation (the equation of motion for particles of spin 1/2), Klein-Gordon's equation (the equation of motion for particles of spin 0), and other, more exotic equations of motion.

The Lagrangian also, of course, lets us get a Hamiltonian and once we have a Hamiltonian, second quantization can be carried out as is done in the appendix (although that's not the most convenient way to quantize a field theory based on a Lagrangian, but I won't go into that here). As we've seen, then, particles are simply the quanta of the quantized field. If the field represents a boson and there are many quanta in the same state, then a coherent state of the field is created and the field behaves as a classical wave.

So, in the end, we could say that the question of whether light is a particle or a wave is answered by saying that it's neither. It is, in fact, a field. And the same is true of electrons and any other fundamental particle.

Appendix I:
The operator approach to the quantization of the harmonic oscillator

Recall the Lagrangian for a one-dimensional harmonic oscillator of mass m and fundamental angular frequency ω_0,

$$\mathcal{L} = \frac{m\dot{x}^2}{2} - \frac{m\omega_0^2}{2}\,x^2\,.$$

Using x as the generalized coordinate, we then obtain its associated generalized momentum,

$$p \equiv \frac{\partial\mathcal{L}}{\partial\dot{x}} = m\dot{x}\,,$$

[29]It's easy to *obtain* the equations of motion but, in almost every case, it's *impossible* to solve them exactly. That's why *perturbative methods*, such as that based on *Feynman diagrams*, are essential.

in terms of which we can now write a Hamiltonian:

$$\mathcal{H} \equiv p\,\dot{x} - \mathcal{L} = \frac{p^2}{2m} + \frac{m\omega_0^2}{2}\,x^2\,.$$

This can be put in a more convenient form by defining yet another set of generalized variables (which can be shown to be canonically conjugated, just as x and p are):

$$\begin{aligned} a &\equiv \frac{1}{\sqrt{2\hbar m\omega_0}}\,(m\omega_0\,x + ip)\,, \\ a^* &\equiv \frac{1}{\sqrt{2\hbar m\omega_0}}\,(m\omega_0\,x - ip)\,. \end{aligned}$$

It might look like we have only one generalized variable, since a^* is the complex conjugate of a, but in fact they are independent quantities, since x and p are independent quantities. Writing x and p in terms of a and a^*, and substituting into $\mathcal{H}$, we find

$$\begin{aligned} x &= \left(\frac{\hbar}{2m\omega_0}\right)^{1/2}(a + a^*) \\ p &= -i\left(\frac{\hbar m\omega_0}{2}\right)^{1/2}(a - a^*) \\ \mathcal{H} &= \frac{\hbar\omega_0}{4}\Big[\,(a+a^*)^2 - (a-a^*)^2\,\Big] = \frac{\hbar\omega_0}{2}\,(a^*\,a + a\,a^*)\,. \end{aligned}$$

So far, everything has been done classically. Now we promote x and p — and, therefore, also a, a^*, and $\mathcal{H}$ — to the status of operators[30] and impose the canonical commutation relation $[\,\hat{x},\hat{p}\,] = i\hbar$:

$$i\hbar = [\,\hat{x},\hat{p}\,] = -i\,\frac{\hbar}{2}\Big[\,(\hat{a}+\hat{a}^\dagger),(\hat{a}-\hat{a}^\dagger)\,\Big] = -i\,\frac{\hbar}{2}\,(-2)\,[\,\hat{a},\hat{a}^\dagger\,] = i\hbar\,[\,\hat{a},\hat{a}^\dagger\,]\,,$$

from which we deduce the commutation relation between $\hat{a}$ and $\hat{a}^\dagger$: $[\,\hat{a},\hat{a}^\dagger\,] = 1$. Using this result, we may rewrite the Hamiltonian, which is now also an operator, as:

$$\hat{\mathcal{H}} = \frac{\hbar\omega_0}{2}\,(\hat{a}^\dagger\,\hat{a} + \hat{a}\,\hat{a}^\dagger) = \hbar\omega_0\left(\hat{a}^\dagger\,\hat{a} + \frac{1}{2}\right) = \hbar\omega_0\left(\hat{N} + \frac{1}{2}\right),$$

where $\hat{N} \equiv \hat{a}^\dagger\,\hat{a}$ is the so-called *number operator*. Note that both $\hat{N}$ and $\hat{\mathcal{H}}$ are positive-definite, that is, their eigenvalues are non-negative. We next compute the commutators of

[30] $\mathcal{L}$, on the other hand, is not considered an operator because it's not a function of canonically conjugated variables. Note also that a^* turns into the *Hermitian conjugate* of the operator $\hat{a}$, namely, $\hat{a}^\dagger$.

$\hat{N}$ with $\hat{a}$ and $\hat{a}^\dagger$:

$$\begin{aligned}
[\,\hat{N}, \hat{a}\,] &= [\,\hat{a}^\dagger\,\hat{a}\,\hat{a} - \hat{a}\,\hat{a}^\dagger\,\hat{a}\,] = [\,\hat{a}^\dagger\,\hat{a} - \hat{a}\,\hat{a}^\dagger\,]\,\hat{a} = -\hat{a}\,, \\
&\Rightarrow \qquad \hat{N}\,\hat{a} = \hat{a}\,(\hat{N} - 1)\,, \\
[\,\hat{N}, \hat{a}^\dagger\,] &= [\,\hat{a}^\dagger\,\hat{a}\,\hat{a}^\dagger - \hat{a}^\dagger\,\hat{a}^\dagger\,\hat{a}\,] = \hat{a}^\dagger\,[\,\hat{a}\,\hat{a}^\dagger - \hat{a}^\dagger\,\hat{a}\,] = \hat{a}^\dagger\,, \\
&\Rightarrow \qquad \hat{N}\,\hat{a}^\dagger = \hat{a}^\dagger\,(\hat{N} + 1)\,.
\end{aligned}$$

Now imagine a state $|m\rangle$ which is an eigenstate of $\hat{N}$ with eigenvalue m, that is, a state for which $\hat{N}\,|m\rangle = m\,|m\rangle$. As pointed out earlier, m is necessarily non-negative. Whatever the value of m may be, imagine applying the compound operator $\hat{N}\,\hat{a}$ to the state $|m\rangle$:

$$\hat{N}\,\hat{a}\,|m\rangle = \hat{a}\,(\hat{N} - 1)\,|m\rangle = \hat{a}\,(m - 1)\,|m\rangle = (m - 1)\,\hat{a}\,|m\rangle\,.$$

The net effect is, then,

$$\hat{N}\,\Big(\,\hat{a}\,|m\rangle\,\Big) = (m - 1)\,\Big(\,\hat{a}\,|m\rangle\,\Big)\,,$$

that is, $\hat{a}\,|m\rangle$ is an eigenstate of $\hat{N}$ with eigenvalue $(m - 1)$. This implies

$$\hat{a}\,|m\rangle = c(m)\,|m - 1\rangle\,,$$

where $c(m)$ is a non-operator value which may depend on m. Now, successively applying $\hat{a}$ to $|m\rangle$, we could bring the resulting eigenvalue to any negative number we wished, unless m is an *integer* and the state $|0\rangle$ is such that $\hat{a}\,|0\rangle \equiv 0$. Thus, since we know that m can never be negative, we've proved that it also must be an integer. Moreover, we've identified the so-called *ground state* as being the state $|0\rangle$. Note that it's also an eigenstate of the Hamiltonian, with energy $\hbar\omega_0/2$.

Next, applying the operator $\hat{N}\,\hat{a}^\dagger$ on $|m\rangle$, we get:

$$\hat{N}\,\hat{a}^\dagger\,|m\rangle = \hat{a}^\dagger\,(\hat{N} + 1)\,|m\rangle = \hat{a}^\dagger\,(m + 1)\,|m\rangle = (m + 1)\,\hat{a}^\dagger\,|m\rangle\,,$$

with the result that

$$\hat{N}\,\Big(\,\hat{a}^\dagger\,|m\rangle\,\Big) = (m + 1)\,\Big(\,\hat{a}^\dagger\,|m\rangle\,\Big)\,,$$

that is, $\hat{a}^\dagger\,|m\rangle$ is an eigenstate of $\hat{N}$ with eigenvalue $(m + 1)$. Again, this implies

$$\hat{a}^\dagger\,|m\rangle = c'(m)\,|m + 1\rangle\,,$$

where $c'(m)$ is another non-operator value which may depend on m. Since we've established that m is a non-negative integer, we see that there is no upper bound to the eigenvalues of $\hat{N}$, nor to those of $\hat{\mathcal{H}}$.

All that's left now is to normalize the states. Since $\langle m|\hat{N}|m\rangle = m\,\langle m|m\rangle$, imposing a normalization condition on $|m\rangle$ amounts to setting

$$\begin{aligned} 1 &= \langle m|m\rangle = \frac{\langle m|\hat{N}|m\rangle}{m} = \frac{\langle m|\hat{a}^\dagger\,\hat{a}|m\rangle}{m} = \frac{||\,\hat{a}\,|m\rangle\,||^2}{m} = \frac{||\,c(m)\,|m-1\rangle\,||^2}{m} \\ &= \frac{|\,c(m)\,|^2\,||\,|m-1\rangle\,||^2}{m} = \frac{|\,c(m)\,|^2}{m} \qquad \Rightarrow \qquad c(m) = \sqrt{m}\,. \end{aligned}$$

Similarly, it's easy to show that $c'(m) = \sqrt{m+1}$.[31] Summarizing, then, the quantum harmonic oscillator in one dimension has an infinite, but discrete, set of states $|m\rangle$ such that $m \geq 0$, $\hat{N}\,|m\rangle = m\,|m\rangle$, with energies $E_m = (m+1/2)\,\hbar\omega_0$, and with a ground state $|0\rangle$ such that $\hat{a}\,|0\rangle = 0$. Note that nowhere have we talked about a wave function, although we could easily obtain one for the ground state by re-expressing $\hat{a}$ in terms of $\hat{x}$ and $\hat{p}$, using $\hat{p} = -i\hbar\,\partial/\partial x$ and solving for $\hat{a}\,\langle x|0\rangle = 0$. Once we have the wave function for the ground state, we then use the $\hat{a}^\dagger$ operator to obtain the wave function for the next state and so on.

Alternatively, we can stick to a description in terms of particle numbers, also called *state occupation numbers*. We then speak of $|m\rangle$ as a state with m quanta of the oscillator, each carrying the energy $\hbar\omega_0$. This description is convenient because it allows the number of quanta to change. More specifically, the operator $\hat{a}^\dagger$, now known as a *creation operator*, increases the occupation number of a state by 1, while the operator $\hat{a}$, the *destruction operator*, decreases the occupation number of a state by 1.

At last, then, we see the direct connection between simple harmonic motion and relativistic quantum mechanics; this framework is exactly what's needed to attain a relativistic formulation of quantum mechanics, since the combination of relativity and quantum mechanics implies the non-conservation of particle numbers.

This appendix is a very brief introduction to the quantization of the harmonic oscillator using the operator method. For a more detailed treatment, the reader should consult any good book on quantum mechanics.

Appendix II:
Second quantization of the electromagnetic field, in the Coulomb gauge

We shall now endevour to show how electromagnetic waves are to be quantized, that is, how the photon picture of the electromagnetic field emerges. This is done by first showing that electromagnetic waves are essentially composed of an infinite number of degrees of freedom,

[31] Strictly speaking, there's an arbitrary choice of phase available here, that is, we could write $c(m) = \sqrt{m}\,\exp(i\varphi)$ with an arbitrary value of φ, and similarly for $c'(m)$. This is an issue only in the presence of degeneracy, which doesn't exist in the one dimensional harmonic oscillator; all its states have distinct energies.

each of which is a harmonic oscillator, and then by quantizing those harmonic oscillators in exactly the same way the one dimensional harmonic oscillator is quantized.

We start from Maxwell's equations in free space, since we want to describe *free* photons, that is, photons that are not interacting with any other field or matter:[32]

$$
\begin{array}{lrcl}
\textit{Gauss' law} \text{ for } \boldsymbol{E} & \boldsymbol{\nabla}\cdot\boldsymbol{E} & = & 0\,, \\
\textit{Gauss' law} \text{ for } \boldsymbol{B} & \boldsymbol{\nabla}\cdot\boldsymbol{B} & = & 0\,, \\
\textit{Faraday's law of induction} & \boldsymbol{\nabla}\times\boldsymbol{E} & = & -\dfrac{\partial\boldsymbol{B}}{\partial t}\,, \\
\textit{Ampère-Maxwell equation} & \boldsymbol{\nabla}\times\boldsymbol{B} & = & \mu_0\varepsilon_0\,\dfrac{\partial\boldsymbol{E}}{\partial t}\,.
\end{array}
$$

It's possible to automatically satisfy two of the four equations by defining two more fields, a scalar field ϕ (called the *electrostatic potential*) and a vector field $\boldsymbol{A}$ (called the *vector potential*), as follows. From Gauss' law for $\boldsymbol{B}$, it follows that $\boldsymbol{B}$ can be written as the curl of a vector field, so we set $\boldsymbol{B} \equiv \boldsymbol{\nabla}\times\boldsymbol{A}$. Then, from Faraday's law, it follows that

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

which implies that $\boldsymbol{E}$ plus the time derivative of $\boldsymbol{A}$ must be the gradient of some scalar function. We thus set

$$\boldsymbol{E}+\frac{\partial\boldsymbol{A}}{\partial t}\equiv-\boldsymbol{\nabla}\phi\,.$$

The negative sign is just a historically motivated convention, having to do with the inconvenient assignment of negative charge to the electron. Had Ben Franklin chosen to call negative charge what he chose to call positive charge, and vice-versa, then electrons would be positively charged and would move in the direction of the electric field, rather than opposite to it, which would have changed the negative sign in front of $\boldsymbol{\nabla}\phi$ to a positive one. Regardless, we now have:

$$
\begin{aligned}
\boldsymbol{E} &= -\boldsymbol{\nabla}\phi-\frac{\partial\boldsymbol{A}}{\partial t}\,, \\
\boldsymbol{B} &= \boldsymbol{\nabla}\times\boldsymbol{A}\,.
\end{aligned}
$$

Note that ϕ and $\boldsymbol{A}$ are not uniquely defined by $\boldsymbol{E}$ and $\boldsymbol{B}$. In fact, the fields

$$
\begin{aligned}
\phi' &= \phi-\frac{\partial f}{\partial t}\,, \\
\boldsymbol{A}' &= \boldsymbol{A}+\boldsymbol{\nabla}f\,,
\end{aligned}
$$

[32]The quantization of the interacting electromagnetic field is more easily done using a Lagrangian formulation and is beyond the scope of this essay.

where f is an arbitrary (but differentiable) scalar field, lead to the same $\boldsymbol{E}$ and $\boldsymbol{B}$ fields. This freedom in choosing the scalar and vector potentials is called *gauge invariance* and plays a fundamental rôle in the theories of the strong and weak nuclear forces, as well as in their unification with the electromagnetic interaction. Here, however, it suffices to point out that gauge invariance allows us to impose certain constraints on the potentials, in order to facilitate our task of quantizing the electromagnetic field. We shall choose the so-called *Coulomb gauge*, in which the vector potential satisfies $\boldsymbol{\nabla}\cdot\boldsymbol{A} \equiv 0$. It's possible to show that, for given $\boldsymbol{E}$ and $\boldsymbol{B}$ fields, this condition uniquely determines $\boldsymbol{A}$. We'll deal with ϕ in a moment.

With $\boldsymbol{E}$ and $\boldsymbol{B}$ written in terms of the potentials, we can now attempt to solve the second pair of Maxwell's equations,

$$
\begin{aligned}
\boldsymbol{\nabla}\cdot\boldsymbol{E} &= 0 && \Rightarrow && \nabla^2\phi + \frac{\partial}{\partial t}\boldsymbol{\nabla}\cdot\boldsymbol{A} = 0\,, \\
\boldsymbol{\nabla}\times\boldsymbol{B} &= \mu_0\varepsilon_0\,\frac{\partial\boldsymbol{E}}{\partial t} && \Rightarrow && \boldsymbol{\nabla}(\boldsymbol{\nabla}\cdot\boldsymbol{A}) - \nabla^2\boldsymbol{A} = -\mu_0\varepsilon_0\left(\frac{\partial}{\partial t}\boldsymbol{\nabla}\phi + \frac{\partial^2\boldsymbol{A}}{\partial t^2}\right).
\end{aligned}
$$

With the choice of the Coulomb gauge, the scalar potential is seen to satisfy Laplace's equation. But since there are no charges anywhere, the scalar potential must vanish at infinity and the only solution to Laplace's equation satisfying that requirement is $\phi \equiv 0$. Hence, we obtain:

$$
\begin{aligned}
\boldsymbol{E} &= -\frac{\partial\boldsymbol{A}}{\partial t}\,, \\
\boldsymbol{B} &= \boldsymbol{\nabla}\times\boldsymbol{A}\,, \quad \text{with} \quad \nabla^2\boldsymbol{A} - \mu_0\varepsilon_0\,\frac{\partial^2\boldsymbol{A}}{\partial t^2} = 0\,.
\end{aligned}
$$

The equation satisfied by $\boldsymbol{A}$ is none other than the vector form of the *wave equation*, and shows that the vector field $\boldsymbol{A}$ propagates as a wave of speed

$$c \equiv \frac{1}{\sqrt{\mu_0\varepsilon_0}}\,.$$

When Maxwell inserted the known values of μ_0 and ε_0 into the expression above, he obtained a value very near the measured speed of light. It was this fact that led him to suggest that light is an electromagnetic phenomenon.

In order to solve for $\boldsymbol{A}$, and hence for $\boldsymbol{E}$ and $\boldsymbol{B}$, we notice that the wave equation is a *linear* equation, which means that any superposition of solutions is a solution. Since a *plane wave* is a solution of the wave equation, the most general solution can then be written as a superposition of plane waves. Formally, this result is obtained as follows. First, write $\boldsymbol{A}$ as a Fourier integral,

$$\boldsymbol{A}(\boldsymbol{r},t) = \int d^3\boldsymbol{k}\, e^{i\boldsymbol{k}\cdot\boldsymbol{r}}\,\boldsymbol{a}(\boldsymbol{k},t)\,,$$

then insert it into the wave equation. Since the only dependence on $\boldsymbol{r}$ is inside the exponential, the wave equation requires

$$\int d^3\boldsymbol{k}\, e^{i\boldsymbol{k}\cdot\boldsymbol{r}} \left[|\boldsymbol{k}|^2 c^2\, \boldsymbol{a}(\boldsymbol{k},t) + \ddot{\boldsymbol{a}}(\boldsymbol{k},t) \right] = 0\,,$$

from which we obtain $\boldsymbol{a}(\boldsymbol{k},t) = e^{-i|\boldsymbol{k}|c\,t}\, \boldsymbol{a}(\boldsymbol{k})$, with $\boldsymbol{a}(\boldsymbol{k}) \equiv \boldsymbol{a}(\boldsymbol{k},0)$. Thus,

$$\boldsymbol{A}(\boldsymbol{r},t) = \int d^3\boldsymbol{k}\, e^{i(\boldsymbol{k}\cdot\boldsymbol{r}\,-\,|\boldsymbol{k}|c\,t)}\, \boldsymbol{a}(\boldsymbol{k})\,,$$

which is the afore-mentioned superposition of plane waves, each moving along the direction $\boldsymbol{k}$ with an angular speed $\omega(\boldsymbol{k}) = |\boldsymbol{k}|c$. Note that, since $\boldsymbol{k}$ is a continuous variable, an infinite number of such plane waves contribute to $\boldsymbol{A}$.

Now, can we determine anything about $\boldsymbol{a}(\boldsymbol{k})$? Yes, we can. First, since the electric and magnetic fields are *real* quantities, so must be $\boldsymbol{A}$, which implies that the solution written above isn't quite correct as it stands. In order to enforce $\boldsymbol{A} = \boldsymbol{A}^*$, we should then write[33]

$$\boldsymbol{A}(\boldsymbol{r},t) = \int d^3\boldsymbol{k} \left[e^{i(\boldsymbol{k}\cdot\boldsymbol{r}\,-\,|\boldsymbol{k}|c\,t)}\, \boldsymbol{a}(\boldsymbol{k}) + e^{-i(\boldsymbol{k}\cdot\boldsymbol{r}\,-\,|\boldsymbol{k}|c\,t)}\, \boldsymbol{a}^*(\boldsymbol{k}) \right].$$

We're still not done, however. Recall that we're solving Maxwell's equations in the Coulomb gauge, which requires $\boldsymbol{\nabla}\cdot\boldsymbol{A} = 0$. This translates into a requirement on $\boldsymbol{a}(\boldsymbol{k})$, namely, that $\boldsymbol{k}\cdot\boldsymbol{a}(\boldsymbol{k}) = 0$, that is, each plane wave's amplitude vector must be perpendicular to the plane wave's corresponding wave vector. Since each plane wave is travelling along the direction of its wave vector, we see that this condition is nothing but a proof that light is a *transverse* wave.

So, let's choose, for each $\boldsymbol{k}$, a set of two unit vectors $\boldsymbol{\varepsilon}_r(\boldsymbol{k})$, not necessarily real,[34] and where $r = 1$ or 2, such that $\boldsymbol{k}\cdot\boldsymbol{\varepsilon}_r(\boldsymbol{k}) = 0$. In fact, we can choose $\{\boldsymbol{k}, \boldsymbol{\varepsilon}_1(\boldsymbol{k}), \boldsymbol{\varepsilon}_2(\boldsymbol{k})\}$ to form a right-handed coordinate system much like $\{\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}}\}$ do. Then we can write $\boldsymbol{a}(\boldsymbol{k})$ as a linear combination of those vectors,

$$\boldsymbol{a}(\boldsymbol{k}) = a_1(\boldsymbol{k})\, \boldsymbol{\varepsilon}_1(\boldsymbol{k}) + a_2(\boldsymbol{k})\, \boldsymbol{\varepsilon}_2(\boldsymbol{k}) = \sum_{r=1}^{2} a_r(\boldsymbol{k})\, \boldsymbol{\varepsilon}_r(\boldsymbol{k})\,.$$

Note that $a_r(\boldsymbol{k})$ is not a vector. Combining these results, we have

$$\boldsymbol{A}(\boldsymbol{r},t) = \sum_{r=1}^{2} \int d^3\boldsymbol{k} \left[e^{i(\boldsymbol{k}\cdot\boldsymbol{r}\,-\,|\boldsymbol{k}|c\,t)}\, \boldsymbol{\varepsilon}_r(\boldsymbol{k})\, a_r(\boldsymbol{k}) + e^{-i(\boldsymbol{k}\cdot\boldsymbol{r}\,-\,|\boldsymbol{k}|c\,t)}\, \boldsymbol{\varepsilon}_r^*(\boldsymbol{k})\, a_r^*(\boldsymbol{k}) \right].$$

[33] Alternatively, we might have started with a Fourier integral which already satisfied $\boldsymbol{A} = \boldsymbol{A}^*$, in which case we'd have arrived at the solution shown without having to 'fix' anything.

[34] The unit vectors $\boldsymbol{\varepsilon}_r(\boldsymbol{k})$ are not necessarily real to allow for the possibility of *circularly polarized* waves, which can be conveniently represented by having $\boldsymbol{\varepsilon}_1 = (\hat{\boldsymbol{x}} + i\,\hat{\boldsymbol{y}})/\sqrt{2}$ and $\boldsymbol{\varepsilon}_2 = (\hat{\boldsymbol{x}} - i\,\hat{\boldsymbol{y}})/\sqrt{2}$, when $\boldsymbol{k}$ is along the z direction.

Let's pause to interpret what we've done. We've managed to solve Maxwell's equations in free space for the scalar and vector potentials, which then give us the electric and magnetic fields, in terms of the linear superposition of an infinite number of plane waves of the form $e^{i(\boldsymbol{k}\cdot\boldsymbol{r}\,-\,|\boldsymbol{k}|c\,t)}\,a_r(\boldsymbol{k})$, each plane wave moving along its own direction $\boldsymbol{k}$, with corresponding wavelength $\lambda(\boldsymbol{k}) = 2\pi/|\boldsymbol{k}|$, frequency $f(\boldsymbol{k}) = c/\lambda(\boldsymbol{k})$, speed c, complex amplitude $a_r(\boldsymbol{k})$, and with a particular *polarization state*, determined by its value of r. From $\boldsymbol{A}$, we then obtain $\boldsymbol{E}$ and $\boldsymbol{B}$:

$$\begin{aligned}
\boldsymbol{E} &= i\sum_{r=1}^{2}\int d^3\boldsymbol{k}\,|\boldsymbol{k}|c\left[e^{i(\boldsymbol{k}\cdot\boldsymbol{r}\,-\,|\boldsymbol{k}|c\,t)}\,\boldsymbol{\varepsilon}_r(\boldsymbol{k})\,a_r(\boldsymbol{k}) - e^{-i(\boldsymbol{k}\cdot\boldsymbol{r}\,-\,|\boldsymbol{k}|c\,t)}\,\boldsymbol{\varepsilon}_r^*(\boldsymbol{k})\,a_r^*(\boldsymbol{k})\right],\\
\boldsymbol{B} &= i\sum_{r=1}^{2}\int d^3\boldsymbol{k}\,\boldsymbol{k}\times\left[e^{i(\boldsymbol{k}\cdot\boldsymbol{r}\,-\,|\boldsymbol{k}|c\,t)}\,\boldsymbol{\varepsilon}_r(\boldsymbol{k})\,a_r(\boldsymbol{k}) - e^{-i(\boldsymbol{k}\cdot\boldsymbol{r}\,-\,|\boldsymbol{k}|c\,t)}\,\boldsymbol{\varepsilon}_r^*(\boldsymbol{k})\,a_r^*(\boldsymbol{k})\right].
\end{aligned}$$

These expressions look formidable but they're quite easy to interpret if we write them as follows:

$$\begin{aligned}
\boldsymbol{E} &= \sum_{r=1}^{2}\int d^3\boldsymbol{k}\left[\,\boldsymbol{E}_r(\boldsymbol{k}) + \boldsymbol{E}_r^*(\boldsymbol{k})\right],\\
\boldsymbol{B} &= \sum_{r=1}^{2}\int d^3\boldsymbol{k}\left[\,\boldsymbol{B}_r(\boldsymbol{k}) + \boldsymbol{B}_r^*(\boldsymbol{k})\right],
\end{aligned}$$

where

$$\begin{aligned}
\boldsymbol{E}_r(\boldsymbol{k}) &= i\,|\boldsymbol{k}|c\;e^{i(\boldsymbol{k}\cdot\boldsymbol{r}\,-\,|\boldsymbol{k}|c\,t)}\,\boldsymbol{\varepsilon}_r(\boldsymbol{k})\,a_r(\boldsymbol{k})\,,\\
\boldsymbol{B}_r(\boldsymbol{k}) &= i\,\boldsymbol{k}\times e^{i(\boldsymbol{k}\cdot\boldsymbol{r}\,-\,|\boldsymbol{k}|c\,t)}\,\boldsymbol{\varepsilon}_r(\boldsymbol{k})\,a_r(\boldsymbol{k})\,.
\end{aligned}$$

In other words, each constituent $\boldsymbol{E}_r(\boldsymbol{k})$ and $\boldsymbol{B}_r(\boldsymbol{k})$ is a plane wave. Note the result that

$$\boldsymbol{B}_r(\boldsymbol{k}) = \frac{1}{c}\,\frac{\boldsymbol{k}}{|\boldsymbol{k}|}\times\boldsymbol{E}_r(\boldsymbol{k})\,,$$

that is, the constituent magnetic and electric fields are perpendicular to one another and to their mutual direction of motion. Once again, this simply means that each consituent electromagnetic wave is transverse.

So far, everything we've done has been classical. In order to quantize the electromagnetic field, we need first to obtain a set of canonically conjugated variables to describe the field with, then promote them to operators, then impose on them the canonical commutator, and finally obtain the quantized energies by looking at the spectrum of the Hamiltonian operator.

As it happens, we already have a set of canonically conjugated variables; the pair $\{\boldsymbol{a}_r(\boldsymbol{k}), \boldsymbol{a}_r^*(\boldsymbol{k})\}$ can be shown to be, *for each* $\boldsymbol{k}$ *and* r, canonically conjugated. This makes sense intuitively, since *each* plane wave of wave vector $\boldsymbol{k}$ and polarization state r is independent of every other plane wave of other wave vectors and polarization states. In other words, each such plane wave forms an independent *mode* of oscillation of the electromagnetic field, much like normal modes in a vibrating mechanical system. Can anyone smell simple harmonic motion sneaking in?

Now, before we impose on them the canonical commutator, let's obtain the classical Hamiltonian. It's simply the energy stored in the electromagnetic field, which is the volume integral of the field's energy density. As we know, the energy density of the electromagnetic field is given by

$$u(\boldsymbol{r},t) = \frac{\varepsilon_0}{2}\,|\boldsymbol{E}(\boldsymbol{r},t)|^2 + \frac{1}{2\mu_0}\,|\boldsymbol{B}(\boldsymbol{r},t)|^2\,.$$

Rather than show the algebra in detail here, let's work conceptually. The constituent electric fields for each $\boldsymbol{k}$ and r are independent from each other and the same is true of the constituent magnetic fields. Thus, it should not surprise anyone that

$$\begin{aligned}
|\boldsymbol{E}(\boldsymbol{r},t)|^2 &= \sum_{r=1}^{2}\int d^3\boldsymbol{k}\,\Big[\,\boldsymbol{E}_r^*(\boldsymbol{k})\cdot\boldsymbol{E}_r(\boldsymbol{k}) + \boldsymbol{E}_r(\boldsymbol{k})\cdot\boldsymbol{E}_r^*(\boldsymbol{k})\,\Big] \\
&= c^2\sum_{r=1}^{2}\int d^3\boldsymbol{k}\,|\boldsymbol{k}|^2\,\Big[\,a_r^*(\boldsymbol{k})\,a_r(\boldsymbol{k}) + a_r(\boldsymbol{k})\,a_r^*(\boldsymbol{k})\,\Big]\,, \\
|\boldsymbol{B}(\boldsymbol{r},t)|^2 &= \sum_{r=1}^{2}\int d^3\boldsymbol{k}\,\Big[\,\boldsymbol{B}_r^*(\boldsymbol{k})\cdot\boldsymbol{B}_r(\boldsymbol{k}) + \boldsymbol{B}_r(\boldsymbol{k})\cdot\boldsymbol{B}_r^*(\boldsymbol{k})\,\Big] \\
&= \sum_{r=1}^{2}\int d^3\boldsymbol{k}\,|\boldsymbol{k}|^2\,\Big[\,a_r^*(\boldsymbol{k})\,a_r(\boldsymbol{k}) + a_r(\boldsymbol{k})\,a_r^*(\boldsymbol{k})\,\Big]\,.
\end{aligned}$$

Combining these, we find

$$u(\boldsymbol{r},t) = \varepsilon_0\sum_{r=1}^{2}\int d^3\boldsymbol{k}\,|\boldsymbol{k}|^2 c^2\,\Big[\,a_r^*(\boldsymbol{k})\,a_r(\boldsymbol{k}) + a_r(\boldsymbol{k})\,a_r^*(\boldsymbol{k})\,\Big]\,.$$

Note that it's independent of time, as it should be for a free system. The Hamiltonian is, then, the volume integral of the energy density above. We now seem to run into a problem. The energy density above is not a function of $\boldsymbol{r}$, so that when we integrate it over all space we obtain an infinite amount. This isn't really a problem, however, since in any practical application, the fields are confined within some finite volume (for instance, inside a cavity whose walls are in thermal equilibrium with the radiation inside). Besides, in any practical problem, the radiation field isn't free, but interacts with matter and other fields. That causes the field to be finite at infinity, rather than plane waves propagating indefinitely.

If we confine the radiation field to a finite volume, however, then not every $\boldsymbol{k}$ vector is allowed, but only those which cause standing waves to appear inside the volume (the other ones interfere destructively and don't contribute to the Hamiltonian). If the volume is a rectangular box of lengths equal to L, then the available wave vectors are of the form

$$\boldsymbol{k} = \frac{2\pi}{L}\,(n_x, n_y, n_z)\,,$$

where n_x, n_y, and n_z are integers. The Hamiltonian, then, replaces the integral in $\boldsymbol{k}$ with a sum over all possible values of n_x, n_y, and n_z:

$$\mathcal{H} = \varepsilon_0\, V \sum_{r=1}^{2} \sum_{n_x \in \mathcal{Z}} \sum_{n_y \in \mathcal{Z}} \sum_{n_z \in \mathcal{Z}} |\boldsymbol{k}|^2 c^2 \left[\, a_r^*(\boldsymbol{k})\, a_r(\boldsymbol{k}) + a_r(\boldsymbol{k})\, a_r^*(\boldsymbol{k}) \,\right],$$

where $V = L^3$ and $\boldsymbol{k}$ given as above. With this understanding, we'll go back to using the integral form simply because it simplifies the various equations. Thus, we'll write:

$$\mathcal{H} = \varepsilon_0\, V \sum_{r=1}^{2} \int d^3\boldsymbol{k}\, |\boldsymbol{k}|^2 c^2 \left[\, a_r^*(\boldsymbol{k})\, a_r(\boldsymbol{k}) + a_r(\boldsymbol{k})\, a_r^*(\boldsymbol{k}) \,\right].$$

We're now ready to promote $a_r(\boldsymbol{k})$ and $a_r^*(\boldsymbol{k})$ to operators and to impose a canonical commutator between them. This commutator takes the form

$$\left[\, \hat{a}_r(\boldsymbol{k}), \hat{a}_s^\dagger(\boldsymbol{k}') \,\right] = \frac{\hbar}{2\varepsilon_0\, V\, |\boldsymbol{k}| c}\, \delta_{rs}\, \delta(\boldsymbol{k} - \boldsymbol{k}')\,.$$

The Kronecker and Dirac deltas guarantee that modes of distinct wave vectors or distinct polarization states are independent. The extra factor is chosen for convenience and could have been introduced in the original Fourier expansion of $\boldsymbol{A}(\boldsymbol{r}, t)$, in which case the commutator here would have no factors other than the deltas. Using this commutator, the Hamiltonian becomes

$$\mathcal{H} = \sum_{r=1}^{2} \int d^3\boldsymbol{k}\, \hbar|\boldsymbol{k}|c \left[\, \hat{a}_r^\dagger(\boldsymbol{k})\, \hat{a}_r(\boldsymbol{k}) + \frac{1}{2}\, \delta(\boldsymbol{0}) \,\right].$$

We then seem to have run into yet another problem; $\delta(\boldsymbol{0})$ is an infinite quantity. This isn't a problem either, however, because energies can always be measured with respect to an arbitrarily set origin and so we can safely ignore the infinite term and write, instead:[35]

$$\mathcal{H} = \sum_{r=1}^{2} \int d^3\boldsymbol{k}\ \hbar|\boldsymbol{k}|c\ \hat{a}_r^\dagger(\boldsymbol{k})\, \hat{a}_r(\boldsymbol{k}) = \sum_{r=1}^{2} \int d^3\boldsymbol{k}\ \hbar\omega(\boldsymbol{k})\ \hat{a}_r^\dagger(\boldsymbol{k})\, \hat{a}_r(\boldsymbol{k})\,,$$

[35] Except in the presence of a gravitational field! Recall that the relativistic theory of gravity, Einstein's general theory of relativity, is non-linear, in the sense that gravitational energy contributes to the gravitational field, since all energy and matter distributions do so. Thus, energies cease to have an arbitrary origin and the infinity obtained here does become a true problem. That's only one of the reasons why a consistent quantum theory of gravity hasn't been found yet.

with $\omega(\boldsymbol{k}) = |\boldsymbol{k}|c$. This result looks very much like that which was obtained for the harmonic oscillator, except that here we have an infinite number of independent modes, rather than just one, and the extra term relating to the energy of the ground state has been shifted away. What this means is that the operators $\hat{a}_r^\dagger(\boldsymbol{k})$ and $\hat{a}_r(\boldsymbol{k})$, respectively, create and destroy photons of momentum $\hbar\boldsymbol{k}$ and polarization state r.

References

- Most biographical bits of information were found searching *Wikipedia*:
  http://en.wikipedia.org
- The Discovery of Radioactivity
  http://www.lbl.gov/abc/wallchart/chapters/03/4.html
- Roentgen And The Discovery Of X-rays
  http://www.xray.hmc.psu.edu/rci/ss1/ss1_2.html
- Marie Curie - Biography
  http://nobelprize.org/physics/laureates/1903/marie-curie-bio.html
- Marie Sklodowska Curie - Physicist
  http://www.lucidcafe.com/library/95nov/curie.html
- Female Nobel Prize Laureates
  http://www.almaz.com/nobel/women.html
- Double Nobels
  http://almaz.com/nobel/double.html
- A Brief (Incomplete) History of Light and Spectra
  http://dbhs.wvusd.k12.ca.us/webdocs/Electrons/Spectrum-History.html

■