20071108

The Durability of Maxwell's Electromagnetic Theory

Maxwell's Electromagnetism and its bearing on Modern Optics

[Colloquium, Optics Department, Reading University, 1994]

0.0.1 Introduction

About 18 months ago I was asked to talk at a day-long seminar being arranged by the History of Physics Group of the Institute of Physics, to be held in the Edinburgh house where James Clerk Maxwell was born in 1831. The seminar was to deal mainly with Maxwell's ancestry and early life, but the planners felt there should be some account of part of Maxwell's scientific work as well, and they invited me to talk about any aspect I chose of the electromagnetic theory. The other speakers, and the audience, would include professional historians and philosophers of science, and I quite soon learned that there would also be a number of very eminent theoretical physicists – though, come the day, I was both surprised and alarmed to see just how many of them there were. On the other hand, I had realised from the start that the organisation of the Edinburgh International Science Festival, of which this seminar formed part, was such that there might be quite a number of lay persons just coming in off the street – and sure enough, there were. This meant that the audience would be a pretty motley crew. After some hard thought I decided that it might be suitable to give a simple review of some of the changes in physics since Maxwell's day, and to see what one could conclude about the status of the Maxwell theory today, in the light of almost 100 years of relativity and quantum theory.

0.0.2 Classical and Modern Physics

Some 30 or 35 years ago, the Scottish Education Department collected a small group of people to design a possible A-level syllabus in Physics for use in Scottish schools. There were two university physicists in the group – John Gunn, professor of theoretical physics in the University of Glasgow, and myself. At one of the early meetings, I remember John Gunn saying weightily that : "The Physics of the twentieth century is based on two new ideas – the Relativity Idea, and the Quantum Idea, and of the two the Quantum Idea is incomparably the more difficult".

Probably most of Gunn's hearers had until then thought of Relativity as very obscure, mystical, and profound; and of quantum theory as a relatively commonplace matter of energy levels, photons, atoms, electrons – things which, of course, everybody is now completely familiar with ! But Gunn's point was that relativity is the completion – or indeed the culmination – of classical physics, whereas quantum physics is something completely new: relativity, like the rest of classical physics, is continuum physics - in this case the physics of the space-time continuum – whereas quantum physics is the physics of the 'discontinuum'; relativity, like the rest of classical physics, is deterministic, whereas quantum physics is probabilistic right at its roots. It was Richard Feynman, the American Nobel Prize winner, who said that "In the fundamental laws of physics there are odds" – and it was quantum physics he was talking about.

Quantum physics dates, I suppose, from 1900, when Planck published his great paper on the thermodynamics of black-body radiation; Einstein's momentous paper *On the electrodynamics of moving bodies* was published 5 years later. The Planck paper launched what we nowadays call ‘modern physics’; the Einstein paper belonged firmly to ’classical physics’. Both these papers are commonly held to herald ‘revolutionary’ ideas, and I’m going to look at the question : What has been the effect, on the status of Maxwell’s Theory of Electromagnetism, of the two revolutions in thought associated with the special theory of relativity and with quantum theory?

So, first, **what did relativity do to the Maxwell theory ?** which, remember, was presented by Maxwell in a famous 2-volume treatise in 1873. The answer is – **relativity enlarged the scope of the Maxwell theory**. In the introduction to his paper *On the electrodynamics of moving bodies* Einstein stated two postulates – first, “that the same laws of electrodynamics and optics will be valid for all frames of reference for which the equations of mechanics hold good”, and second, “that light is always propagated in empty space with a definite velocity c which is independent of the state of motion of the emitting body”. Then he goes on to say that “These two postulates suffice for the attainment of a simple and consistent theory of the electrodynamics of moving bodies based on Maxwell’s theory for stationary bodies.” In other words: the Maxwell theory applies to electric and magnetic fields and their interactions with stationary bodies, and Einstein’s two postulates extend the applicability of the Maxwell theory to moving bodies. Others – including Poincaré and Lorentz – had already realised this, so that E T Whitaker, in volume 2 of his great *History of the Theories of Aether and Electricity* actually titles one of his chapters, “*The relativity theory of Poincaré and Lorentz*”. The advent of relativity left the Maxwell theory unscathed and enlarged in scope, and gave it new and deeper meaning. The interpretation of the magnetic field as a relativistic effect engendered in the middle of this century a new and unified approach to the teaching of relativity and electro-magnetism, and the technical remark that ‘the Maxwell equations are invariant under the Lorenz transformation’ must surely sound familiar to most physics undergraduates now.

But when we ask: **what did quantum theory do to the Maxwell theory ?** the answer is not so simple, and may still be evolving. That there was a problem was clear before Planck published his 1900 paper, and Planck gave the beginnings of an answer in that paper. He certainly didn’t believe he had written the last word on it – indeed, he hoped he hadn’t, because some of the implications of his paper weren’t much to his liking. Briefly the problem was this:

0.0.3 The failure of the equipartition law

In most branches of physics there are phenomena which are collected under the portmanteau title : **NOISE**; this generally refers to random fluctuations associated with the structure of a material, or of a physical system. Let’s consider for a moment a simple, familiar system – the air enclosed in a pipe closed at one end and open to the atmosphere at the other. The air molecules in the pipe are in random motion. striking against one another and against the walls of the pipe. Because of this there are at all times fluctuations in the number of molecules contained within any volume inside the pipe, and so there must be fluctuations in the local air pressure in any volume. We know that sound waves are periodic fluctuations in the pressure of the air, and these random fluctuations of pressure also constitute a sound. This is different from what we normally think of as sound waves in an organ pipe or other musical instrument because in those latter cases the sound has a definite pitch or frequency, whereas the random motions of the air generate sound which has no definite pitch. If you want to hear this, hold a cup or tumbler – or a seashell – close to your ear: you will hear a rushing sound. The more nearly the vessel you hold against your ear resembles a pipe, or musical wind instrument, the more does this rushing noise contain a preponderance of sound with a recognisable pitch, because a pipe resonates to a family of frequencies which characterise its possible performance as a musical instrument; the frequencies of this family are called the natural frequencies of the system, and holding the cup or pipe or seashell against the ear 'tunes' the ear, weakly, to the natural frequencies of the cup or whatever. In fact, the rushing sound associated with the thermal agitation of the air molecules, and referred to as 'thermal noise', is below the threshhold sensitivity of the human ear by a factor $2\times$. The tuning effect of the cup, or whatever, increases the sensitivity of the ear by just enough to make this noise audible.

Any elementary textbook on physics, or on sound, or on vibrations and waves, derives the natural frequencies of the pipe closed at one end and shows that the nth natural frequency is $f_n = (2n-1).V/4L$ for $n = 1, 2, 3, \ldots$ There is no upper limit to n , so the number of natural frequencies – or 'normal modes' – is infinite. The difference between any two neighbouring members of this sequence is $V/2L$, so the number of natural frequencies per unit frequency range is the reciprocal of this, i.e., $2L/V$.

Classical statistical mechanics tells you – through what is called the 'principle of equipartition of energy' – that when this system is in thermal equilibrium at the temperature T , each of these vibrations, whatever its frequency, should have the same mean energy, kT, where k is a universal constant – Boltzmann's constant. But since the number of these frequencies is infinite the total amount of thermal agitation energy in the pipe must be infinite. And further, since the number of natural frequencies per unit bandwidth is the same across the whole of the infinitely wide spectrum, the noise energy per unit bandwidth is the same across the whole spectrum – this is therefore called 'white noise'.

This argument applies generally to 1-dimensional systems, e.g., electrical transmission lines; in the electrical case the thermal noise is called 'Johnson noise' (J B Johnson, 1927) who treated it as a consequence of the Brownian motion of electrons. Analogous arguments can be developed for 2-dimensional and 3-dimensional systems. In a 2-dimensional system the noise power per unit bandwidth increases across the spectrum, being proportional to the frequency, and in 3-dimensional systems it increases like $(\text{frequency})^2$. An example of this phenomenon in three dimensions is the random fluctuation in the energy of the electromagnetic waves in a cavity (a hole!) in a material when there is thermal equilibrium between the waves and the containing material.

One of the great problems of physics towards the end of the 19C was to derive the spectrum of this temperature radiation – or 'black body radiation'. Sir James Jeans proved that the only rigorous solution of this problem within classical physics showed the energy density increasing without limit at the high frequency end of the spectrum – like $(\text{frequency})^2$, as we have seen. This meant that all the energy in matter should, long before now, have radiated off into space at very short wavelengths; this so-called 'ultra-violet catastrophe' proved catastrophic for classical physics, and for the principle of equipartition of energy which was an essential component of classical physics. This failure of the equipartition principle received independent confirmation a few years later (1913) from an investigation in the theory of magnetism, which led to 'Miss van Leeuwen's Theorem', according to which the rigorous classical calculation of the magnetic properties of matter shows that all matter must be non-magnetic. This most certainly isn't the case. In the theory of the specific heats of gases and solids, too, application of classical ideas about equipartition was known to lead to incorrect conclusions.

What Planck provided in his 1900 paper on black-body radiation was, bluntly, a fudge which Planck himself described as "....only an interpolation formula found by lucky guess-work....". It predicted a distribution of energy with wavelength, for electromagnetic waves in a box, which tended to zero at very short wavelengths, and also at very long wavelengths, and predicted that the amount of radiant energy in the box was finite. That was a considerable relief, but what was more important was that the distribution of energy with wavelength which emerged from Planck's paper agreed very precisely with the best available experimental results. Examination of Planck's argument showed that it implied that the mean energy per allowed wavelength was not in fact kT, but depended on wavelength, and decreased in value as the wavelength became smaller, so that there was very little energy at the shortest wavelengths. This was a very important outcome, because the equipartition principle had been logically essential to classical mechanics. Its abandonment, in the aftermath of the ultra-violet catastrophe, was a death-knell for classical theory.

0.0.4 Waves and Statistics

The topic of the thermal equilibrium of a radiation field in contact with matter is clearly to do with the statistics of radiation, and certain aspects of this presented problems which weren't solved until the 1960's. One of these problems was contained in the question: What is the appropriate statistical distribution to describe the amplitude of a randomly fluctuating radiation field? This was discussed by Debye in 1910, and later by Einstein and von Laue who had a very interesting correspondence about it in the German scientific literature in 1916-17; but no answer emerged. Then in the 1930's the question of the statistics of light fields began to need a solution, partly in the earlier context of the cavity radiation problem, but also so that certain questions about the nature of optical images could be answered; van Cittert in Germany and Zernike in Holland tackled this problem. But again there seemed at that time no way of arriving at a definitive answer.

Physicists have always tended to form mental pictures of systems they are trying to understand – they dignify this process as 'model building' – but please remember that this refers to mental models, not workshop models! Maxwell built such a model – out of 'vortices' and 'epicyclic gearwheels', in the early days of his electromagnetic theory; but he jettisoned it when it had served its turn, and replaced it with something much more abstract – the field concept, which was one of his two most profound contributions to physical theory. The early atomic physicists built models in the first two decades of this century. They pictured – notice how easily one uses the word "pictured" here! – they pictured the atom as a massive nucleus with very lightweight electrons moving round it in elliptic orbits, rather as the planets move round the sun. This kind of modelling has very real intellectual hazards associated with it – the model, taken literally, has properties not all of which may reasonably be attributed to the real system you are trying to understand. Thus, you may be led astray by attributing too much relevance to details of the 'model'.

The solutions of the Maxwell Equations include – and are usually represented as – waves, like that sketched here. [O/HP] Such a wave has, to the physicist, three characteristics with which he expects to be able to associate numbers as the result of a measurement process. These are the frequency, the amplitude, and the phase, all of which can be identified on the diagram. If you receive a radio signal, and you connect your receiver to a Cathode Ray Oscilloscope, the oscilloscope screen will show a picture very like the one in the diagram and you can read off the frequency, amplitude and phase of the wave. But you can't do this with light-waves. As a student I was told that this was because we didn't yet have instruments which could follow the very rapid oscillations of the light field – about a hundred million million waves per second. But some optical theorists had felt, certainly since the early 1930's, that it should be possible to strip out of the theory of optics those quantities which could neither be observed nor measured, and to construct a theory of optics in terms of observable quantities. That theory is what we now call the theory of optical coherence.

0.0.5 Quantisation

So what are the quantities which can be measured when you have stripped out of the theory of light waves those very quantities which emerge so naturally from the Maxwell equations? They are – or at least they include – the intensities at all points in the field, and the statistical correlations between the field amplitudes at different places, or at different times, or at different places and different times. [With the help of the O/HP we can give an indication of how these correlations are determined.] [O/HP]. These correlations, obtained from experiments which measure phenomena which have been familiar since last century – or in some cases even longer – are the observables on which Fritz Zernike, Harold Hopkins, Emil Wolf and others from the 1930's to the 1950's constructed a new formulation of optics which was thought to include a full account of the statistical properties of real light-beams. But this theory was restricted in one respect, because it was assumed that the probability-distribution-function for the amplitude of the electric field in a light-beam was a Gauss function. Physicists and many other scientists always like to use what is called the Gaussian (or normal) distribution because of its simple and well-understood mathematical properties – and because it quite often seems to be appropriate. So it was assumed – for lack of evidence to the contrary – that the amplitude of the fluctuating electric field in the light wave was a random variable, and that the probability of the field having the magnitude E at any arbitrarily chosen place and time would be calculable from the Gaussian distribution. From that, one can calculate another probability – that a measurement of the fluctuating intensity (or power density) at a point in the light field would yield the value I .

But why should the electric field amplitude, and the intensity, be fluctuating quantities? Why do we have to use statistics at all? (The Maxwell theory didn't.) The answer is actually fairly obvious, and points the way ahead. A light source consists of a very large number of atoms – commonly a million million or more – each of which emits light for about a thousand millionth of a second at a time, and then remains quiescent for a much longer time during which many other atoms are emitting. The number of atoms radiating at any one time is then not a constant quantity, but fluctuates at random around a mean. And the relative phase of the radiations from different atoms also is a random quantity. The result is that the summation of the radiations from all the radiating atoms in the source sometimes is very large, and quite often is very small. The resulting intensity – which is what affects the eye or a photographic plate or photoelectric cell at any instant – fluctuates far more rapidly than the eye can follow. The very fast circuitry which was developed for other purposes in the war years can be used to follow such fluctuations, and experiments in the 1950s produced results which, to various optical physicists at the time, were unexpected, unwelcome, and very controversial.

Clearly, we have to pay some attention to the nature of real light sources when we think about the random nature of light fields. It is possible to make progress using the Maxwell theory if we treat all the random effects as being contained within the probability distribution functions which are postulated to describe the fluctuations of the field strength and the associated intensity. A full description of the light field can then be constructed by adding together a whole lot of the basic solutions of the Maxwell equations, with amplitudes distributed according to the form of the probability-distribution-function. This is implicit in what Zernike, Hopkins, Wolf et al did as they developed the classical coherence theory. But when we want to include in our account of light fields the emission processes in the light source and the absorption processes in the light detectors, we have to enter the domain of quantum theory, because only quantum theory can do this properly. One of the standard techniques of formulating a problem quantum-mechanically is to construct, in the first instance, a complete expression, in terms of observables, of the corresponding classical problem; and there are then rules for translating this classical expression in terms of the relevant observables, into the corresponding quantum form. So, from Wolf's classical coherence theory, describing light propagation, diffraction, interference, polarization, correlation and so on all in terms of the optical observables, the way was clear for the development of the quantum theory of light. This was effected by an American, R J Glauber, in two great papers published in 1963. The first of these papers was titled – with elegant acknowledgement of its origins – *On the Quantum Theory of Optical Coherence*, and into it Glauber coupled the pre-existing quantum theory of the emission and absorption of light by atoms. These papers by Glauber founded what we now call "quantum optics".

Right away, new insights began to emerge from the new theory. One of them was the justification for the use of the Gaussian probability function to describe the characteristics of the light coming from the tungsten filament lamps and gas discharge lamps with which all optical experimenters had worked for the half-century before the advent of the laser; this also explained why, and in what respects, laser light differed from that of traditional sources. In this context another of the fresh insights related to a matter of frequent debate between the engineers and physicists who were thrown together in joint enterprises during the 1939-45 war years, namely, why is it that two separate and independent light sources apparently could not produce interference effects whereas independent radio aerials could. The physicists weren't surprised about the first because they'd always known that independent light sources were incoherent (which in fact simply restates that they didn't interfere!); the engineers took the second almost for granted because they'd always believed that separate aerials radiating at the same frequency should cause interference! But it wasn't clear to either group why the shift from an optical frequency to a radio frequency should make the phenomena so different; it was the merging of electromagnetic theory and quantum theory in the 1960's that resolved this apparent puzzle by showing that the two classes of events which were being compared – interference between light radiation from atoms, and interference between signals from different aerials – are not at all analogous – the processes going on inside the sources are entirely different in the two cases.

In a gas discharge lamp, for instance, the spacing between any two simultaneously-radiating atoms is usually greater than the linear size of the induction field around either, so the atoms are virtually uncoupled, and their radiation fields are uncorrelated – the individual atoms are in fact incoherent radiators. In the case of a radio aerial the radiations come from a vast number of conduction electrons in a metal – probably much more than 1023 – which have separations much less than the dimensions of the induction field at the desired frequency. The electrons are therefore rather strongly coupled and their coupling leads to the establishment of a macroscopic current oscillating at the coupling frequency; this frequency is determined by the macroscopic dimensions of the aerial (or, for microwave frequencies, the cavity resonator). The energy of the radio-frequency quanta is very much smaller than kT in terrestrial situations, but the thermal radiations are uncoupled and cause low-level noise, which is – in a properly designed aerial – completely swamped by the radio-frequency field. An aerial therefore acts as a coherent source.

And many other, more subtle, revelations followed.

It was hardly surprising that the new field of quantum optics should contain a rich and exciting diversity of new phenomena. But what could not have been altogether anticipated was that quantum optics as it developed would give us back precisely the results of the Maxwell theory as long as one remained within a clearly defined boundary – a boundary which contained just those areas of optics which had been accessible to experimenters for the past three hundred years, and in which there was a firm, and indeed unchallengeable, experimental background.

0.0.6 Back to Maxwell ?

How does this emerge from quantum optics? It is quite common in calculations in classical physics to replace the real functions which represent the measurable quantities you're interested in – like positions, speeds, or electric fields, – with complex quantities; this is a convenient trick in certain types of calculation, and it doesn't imply anything about the physical nature of the quantities you're describing. Now, complex numbers come in pairs, say $a + ib$ and $a - ib$. You can picture these as a pair of pointers, one rotated clockwise and the other rotated anticlockwise from a horizontal reference line, as on the diagram. [O/HP] When you want to represent oscillating quantities – like the electric field in a light wave – you can do this using the same arrangement of two pointers, but now you assume that they are actually rotating at the frequency of the wave, and that it's the projection of either pointer on the horizontal reference line that represents the real number you would get if you could measure the instantaneous magnitude of the electric field (say). Since each of the rotating pointers in the figure has the same projection on the horizontal line it doesn't matter which one you use.

But when you do this in the electromagnetic theory, and apply the rules which translate your equations into the quantum form, you find that those contra-rotating complex numbers translate into entities – called operators – which have substantially different meanings, according to whether they correspond to the classical pointer rotating clockwise or anti-clockwise. One of these operators turns out to be associated with the physical process of the emission of light, and the other is associated with the opposite process of absorption. The first process, emission, transfers energy from the source into the light field. The other process, absorption, leaves the light-field with less energy, and the detector, therefore, with more energy.

Now, we know that the detectors for light – photocells, photographic emulsions, and so on – don't detect radio waves nor long-wave infrared, because the energy of the quanta in these radiations isn't enough to affect the electrons in the photo-cell, or the crystals in the photographic emulsion. There has to be some match between the energy packets in the light-wave and the characteristics of the detector, before absorption can take place. All the atoms, molecules, or whatever in a suitably chosen detector will be in their lowest possible state – the ground state – before the light falls on them, because the only energy they could have picked up from their environment would be approximately kT , which is only about 1/40 of the excitation energy of the higher state to which light excites the system. [O/HP] This, then, is the mechanism associated with one of those two operators – the absorption process. In the other process, an atom or crystal initially in the upper energy state when the light fell on it would emit light, ending down in the ground state. This doesn't happen naturally, in optics, because as we've already seen there's normally no reason why there should be any significant number of highly excited atoms there to start with.

Since light waves are normally detected when their energy is absorbed by atoms which were initially in their ground state, these detection processes can be described quantum mechanically by equations which contain only the absorption operators, and not the emission operators. In among the ensuing mathematics there is embedded a function which corresponds to the probability distribution function which I spoke of earlier as describing the spread (or uncertainty) in the intensity of the light beam. In classical theory probability functions can have only positive values (it wouldn't make a lot of sense to say that a horse has a -1 in 8 chance of winning a race! – except, possibly, in last year's Grand National). As long as we stick to equations which use only the absorption terms, and as long as we provide that the quantum analogue of the probability function can't have negative values, the quantum calculations give us results which can be associated with possible solutions of the Maxwell equations. But it seems that the quantal quasi-probability function need not be restricted to positive values; and if we include the emission terms, and/or allow this funny quantum quasi-probability function to have negative values – then the quantum calculations predict new and previously unsuspected phenomena. Some of these newly-predicted effects have been found, in experiments of great subtlety, and they can't be described by the classical Maxwell theory.

So we believe we now understand why the Maxwell theory was – and continues to be – so successful in supporting the design and analysis of all optical experiments and all optical instruments known before 1960, the date, remember, of the invention of the laser; and why the Maxwell theory will continue to be useful in much of optical and most of radio science in the future. But in addition we have glimpses of a wider range of phenomena, still largely unexplored and unexploited – and still relying on the Maxwell theory for the classical foundation on which quantum optics has to be built.

Perhaps I should have titled this talk: *The Durability of Maxwell's Electromagnetic Theory*!

R M Sillitto, 1st March, 1994

R M Sillitto – biographical notes

Born and educated in Scotland; spent three years during the 1939-45 war working for the Admiralty, mainly on problems of visual and infra-red signalling. Returning to Edinburgh after this, was for 16 years a low-energy nuclear physicist, before turning back to optics which had always been his long-term intention. Has worked in quantum optics, and on imaging theory, lens assessment, classical diffraction, and most recently on the optical characteristics of opto-electronic devices.

Is a Fellow of the U.K. Institute of Physics, of the Royal Society of Edinburgh, and of the Optical Society of America.