The Grand Synthesis

(Or How the Speed of Light Entered into and Emerged from Maxwell's Equations)

M. P. Silverman

[Adapted from M P Silverman, *Waves and Grains*, (Princeton University Press, 1998, Chapter 10)]

> *Hence the velocity of light deduced from experiment agrees sufficiently well with the value of v deduced from the only set of experiments we as yet possess. The value of v was determined by measuring the electromotive force with which a condenser of known capacity was charged, and then discharging the condenser through a galvanometer, so as to measure the quantity of electricity in it in electromagnetic measure. The only use made of light in the experiment was to see the instruments.*
>
> James Clerk Maxwell (1864)[1]

The Problem and the Solution

Despite the enormous success of Augustin Fresnel's wave theory of light, there remained a fundamental and thorny issue: If light were a kind of undulation, then what, precisely, was "waving"?

The answer to this question was discovered by Scottish physicist James Clerk Maxwell not in optics but in the study of remotely connected phenomena of electricity and magnetism. Deeply impressed by the experimental researches of Michael Faraday, Maxwell set out to give mathematical structure to Faraday's geometrical conception of a continuum of lines of force permeating the space between electrified and magnetised bodies. It is perhaps hard to imagine today—when nearly all theoretical physics is a study in field theory—how bold was Faraday's idea and how much resistance it incurred from contemporary scientists and mathematicians. "I was aware," wrote Maxwell in the preface to the first edition of his *Treatise on Electricity & Magnetism,*[2], "that there was a difference between Faraday's way of conceiving phenomena and that of the mathematicians, so that neither he nor they were satisfied with each other's language."

> "For instance, Faraday, in his mind's eye, saw lines of force traversing all space, where the mathematicians saw centres of force attracting at a distance: Faraday saw a medium where they saw nothing but distance: Faraday sought the seat of the phenomena in real actions going on in the medium, they were satisfied that they had found it in a power of action at a distance impressed on the electric fluids."

---

[1]J. C. Maxwell, "A Dynamical Theory of the Electromagnetic Field", reprinted from *The Transactions of the Royal Society* **155** (1865) in *The Scientific Papers of James Clerk Maxwell* (Dover, New York, 1952) 526-597; quotation from p. 580

[2]J. C. Maxwell, *A Treatise on Electricity & Magnetism* , Vols 1 and 2 (Dover, New York, 1954) republication of unabridged 3rd Edition, published by the Clarenden Press in 1891

Like Fresnel, Maxwell lived a regrettably short time, dying of cancer at the age of 48 in 1879. He began his investigations of electromagnetism in the mid 1850's with the resolve to read Faraday's *Experimental Researches in Electricity,* and by the mid 1860's, through what is perhaps the most remarkable application of modeling by analogy of which I know, he arrived at his definitive dynamical theory of the electromagnetic field. In contrast to André-Marie Ampère, whom he greatly admired and of whom he wrote

> "We can scarcely believe that Ampère really discovered the law of [magnetic] action by means of the experiments which he describes. We are led to suspect...that he discovered the law by some process which he has not shewn us, and that when he had afterwards built up a perfect demonstration he removed all traces of the scaffolding by which he had raised it." ,

Maxwell revealed all the interim steps in his own progress toward perfection.

In the first step[3], convinced of the conceptual fertility of Faraday's lines of force, Maxwell likened them to streamlines in a hydrodynamic model of incompressible tubes of electric and magnetic fluids. One insightful outcome of these considerations, which was to become a seminal part of the final theory, was Maxwell's distinction between a vectorial "quantity", associated with the flux of a field through an area, and a vectorial "intensity", associated with the circulation of a field about a closed path.

In the second step[4], he devised an extraordinary mechanical model whereby space-filling molecular vortices and idle wheels transmitted pressures and tensions representative of electrical and magnetic interactions, in particular the interaction between current-carrying wires (Ampère's law) and the induction of an electromotive force by a change in magnetic flux (Faraday's law). From this hypothetical construction Maxwell deduced the existence of transverse waves propagating at a speed determined by the elasticity and density of the matter of the vortices (the aether)—which, when evaluated, very nearly equaled the speed of light as it was then known ($\sim 3.15\infty 10^{8}$ m/s).

In the final step[1], Maxwell displayed the architecture of his own edifice with the scaffolding removed (as Ampère had done from the outset without benefit of "blueprints"), and presented in their awe-inspiring entirety the mathematical relations that account for all (non-quantum) electromagnetic phenomena. "The agreement of the results," wrote Maxwell, "seems to shew that light and magnetism are affections of the same substance, and that light is an electromagnetic disturbance propagated through the field according to electromagnetic laws." Thus, within the span of roughly 10 years Maxwell effected a theoretical synthesis that not only brought electricity, magnetism, and optics under the same set of mathematical laws, but which was to provide the exemplar for virtually all other field theories of modern physics.

To the question 'What is "waving"?', Maxwell's theory provided the answer: an electromagnetic field. But of this, two things must be said.

First, though Maxwell's answer is correct and complete and nothing further is actually required, the question still remained troubling, for it was just as inconceivable to Maxwell and his contemporaries, as it had been to Huyghens, Newton, Young, and Fresnel, that an undulation could propagate through

---

[3]J. C. Maxwell, "On Faraday's Lines of Force", reprinted from the *Transactions of the Cambridge Philosophical Society,* Vol **10**, Part 1 (1855-56) in *The Scientific Papers of James Clerk Maxwell* (Dover, New York, 1952) 155-229

[4]J. C. Maxwell, "On Physical Lines of Force", reprinted from *The Philosophical Magazine* **21** (1861-62) in *The Scientific Papers.. op. cit..* 451-513

space in the absence of a *material* medium. The search for a luminiferous aether consequently remained a pressing issue throughout Maxwell's lifetime and in fact until well after Einstein had effectively disposed of it in his 1905 paper on special relativity[5]. Imagine a substance that had to pass freely through the atoms of matter (to account for the aberration of starlight viewed through a telescope), yet must be a nearly incompressible elastic solid (if it were to transmit high frequency transverse waves); a substance for which was claimed a negative modulus of compression so that it expanded under pressure and contracted when relaxed (!); a substance that now, to satisfy electrodynamics, was riddled with tubes of electric and magnetic flux. If you can not, then you are in good company, for neither could anyone else.

Second, to the distress of many future generations of physics students, Maxwell's theory did not give rise to a *single* field—or even to two (electric and magnetic)—but rather to at least four which in modern notation and terminology are: **E** (electric field), **D** (electric displacement), **B** (magnetic induction), and **H** (magnetic field). Actually, the symbolic designations are exactly those Maxwell had chosen (although he expressed them in Gothic letters), and the nomenclature is only slightly modified from the original; Maxwell termed **E** the "electromotive intensity" and **H** the "magnetic force".

Although **E** and **D**, and likewise **H** and **B**, are conflated in the absence of matter, their conceptual distinctions as "intensities" and "quantities" are vital, and their properties *in matter* are sufficiently different that one must always exercise caution when referring to "transverse" waves of light. For example, in an anisotropic dielectric medium devoid of free charge and current, "quantities" **D** and **B** are transverse to the wave vector **k** (which is normal to the wavefront), whereas "intensities" **E** and **H** are transverse to the Poynting vector **S** (which gives the direction of power flux). **S** and **k**, however, need not be parallel to one another, with the consequence that the wavefronts are *not* transverse to the direction in which light energy is transported. Little things like this make crystal optics a fascinating subject for the devoted—or an ordeal to the geometrically impaired.

The Confusing Matter of Units

Having accomplished his immortal work, Maxwell retired from a professorship at King's College, London, in 1865 to write *A Treatise on Electricity and Magnetism* that should have "for its principal object to take up the whole subject in a methodical manner, and which should also indicate how each part of the subject is brought within the reach of methods of verification by actual measurement." It is this concern of Maxwell's, not only with theoretical foundations, but with concrete experimental details (without which all science reduces to mere opinion) that has influenced so profoundly my own scientific education and research.

Although a significant fraction of my physics pursuits fall within the purview of electrodynamics, I have never had a course in the subject beyond an elementary introduction. Instead, as with most of the physics I learned, I studied the principles on my own—in this case with Maxwell's *Treatise* as both my inspiration and textbook. This is *not* an experience that I would necessarily recommend to others. For all his legendary gentleness, Maxwell is a demanding teacher, and his *magnum opus* is anything but coffee-table reading. He wrote at a time when vectors—introduced by William Rowan Hamilton as part of a long and largely opaque work on quaternions—were understood by only a handful of physicists. And, although Maxwell was among this select group and introduced vector terminology into his *Treatise*—indeed it was he who created such familiar terms as "curl", "gradient", and "divergence"—he nonetheless preferred to express vectorial relations in their Cartesian components, each component distinguished by a different letter rather than a subscript. This did not make reading Maxwell any easier.

[5]A. Einstein, "On the Electrodynamics of Moving Bodies", *Annalen der Physik* **17** (1905) 891-921

All the same, the experience was greatly rewarding in that I had come to understand, as I realised much later, aspects of electromagnetism that are rarely taught at any level today and that reflect the unique physical insight of their creator.

One of the most important of these, in regard to the relationship of electromagnetic waves to light, concerns the delicate subject of electromagnetic units. Few topics, I have found, seem more obscure or less interesting to students and professional physicists alike (except perhaps to those at standards laboratories), and yet more likely to trigger heated discussion over preferences. At the introductory level, it would appear from perusing any number of general physics textbooks that the SI *(Système Internationale)* set of units has swept all others from the field, and so students begin their studies of electricity and magnetism by cluttering their minds with mysterious symbols—epsilon-naught (ε_0) and mu-naught (μ_0)—which are purely concocted numbers that have, in fact, no basis in natural law. Some time later perhaps, students will encounter in more advanced treatments the Gaussian form of Maxwell's equations overflowing with *c*'s—but no explanation is ever given as to *why* the speed of light should occur in relationships between electric and magnetic fields, or between magnetic fields and currents. The absurdity of this situation must surely strike even the least perceptive student, for, as Maxwell wryly relates in the quotation that opens this essay, the only role of light in the measurement of electric and magnetic parameters is to permit experimenters to see their instruments!

Why then should there be any surprise that Maxwell's equations yield electromagnetic waves propagating at the speed of light? In the first case (SI) one has simply contrived to make $1\big/\sqrt{\varepsilon_0\,\mu_0} = c$, and in the second case (Gaussian) an abundance of *c*'s were inserted explicitly by hand at the outset. Students learning the subject from a modern textbook may well be excused if they are not impressed.

But *I*, who learned this marvelous result directly from Maxwell, was impressed indeed. To appreciate the fact that Maxwell's theory makes an extraordinary prediction, and not merely renders what is inserted beforehand, one must first understand how electric charge and current are measured—and I know of no account better than Maxwell's.

"The only systems of any scientific value," Maxwell states in his *Treatise*, "are the electrostatic and the electromagnetic systems."

According to the first (esu) system, a unit of charge is operationally defined by Coulomb's law: Two point charges q attracting or repelling one another with a force F of 1 dyne at a distance d of 1 cm each comprise 1 esu of charge. From $F = q^2\big/d^2$ and its equivalence to mass $\times$ acceleration, the dimension of charge, expressed in terms of the fundamental quantities of mass [M], length [L], and time [T], is readily seen to be

$$\left[q\right]_{esu} = \left[M^{1/2}L^{3/2}T^{-1}\right]\ . \tag{1}$$

It is not electric charge, however, but electric current that is primary in magnetism and consequently the basis for the second (emu) system of units. In this case the unit of current is operationally established by Ampère's law: Equal currents I in two straight segments of wire of length l and separation d attract or repel one another with a force $F = I^2 l\big/d$. It then follows that current has the dimension of the square root of force, or

$$\left[I\right]_{emu} = \left[M^{1/2}L^{1/2}T^{-1}\right]\ . \tag{2}$$

In the emu system charge is a secondary quantity obtained by measuring the passage of current over a period of time; thus

$$[q]_{emu} = [I]_{emu}[T] = \left[M^{1/2}L^{1/2}\right]. \tag{3}$$

Eqs (1) and (3) illustrate that the two systems of units are incompatible, leading to designations of the same quantity (electric charge) that differ not only in their magnitude, but also in their *dimensions*. The number of esu units of charge that make up one emu unit of charge thus takes the form of a velocity

$$q_{esu}/q_{emu} = k\left[LT^{-1}\right] = k[v] \tag{4}$$

where the numerical factor k is a universal constant to be determined experimentally.

It is often quite surprising to those who encounter this basic feature of electromagnetism for the first time—and there are many, I suspect, who, not having read Maxwell, never encounter it. From primary education upward one learns to express measurable quantities in different systems of units—for example, to convert between cgs and mks measures of mass, length, volume, speed, etc. by a simple relocation of the decimal point. Those unfortunate enough to have been brought up in the English system of units learn to convert between feet and miles, pints and quarts, bushels and pecks, ounces and pounds, etc. But, so long as mechanical quantities *only* are involved, the conversion, however awkward, is simply a matter of a dimensionless numerical factor. A metre and a yard are both measures of length, and both have the *dimension* of length. Not so in electromagnetism; an esu of charge has an entirely different dimension from an emu of charge.

Unlike the SI parameters, the conversion factor $k[v]$ is not just a "pencil and paper affair" (to borrow the expression of Nobelist Percy Bridgeman), but is amenable to direct measurement—and Maxwell, who, like Newton and Fresnel, was thoroughly conversant with experiment, described in his *Treatise* at least four ways to measure it. Of these, the method due to Weber and Kohlrausch is especially simple in principle.

A Leyden jar was charged with a certain quantity of electricity, determined in electrostatic measure as the product of the capacitance (*C*) of the jar and the potential difference (*V*) between its coatings. *C* was ascertained beforehand by comparison with the capacitance of a metal sphere suspended in an open space away from other bodies. In the esu system the capacitance of a sphere is given by its radius, and thus *C* for the Leyden jar could be expressed as a certain length. Correspondingly, *V* was measured by connecting the coatings to the terminals of a calibrated electrometer. The two measurements thus furnished $q_{esu} = CV$. The jar was subsequently discharged through the coil of a galvanometer. The transient current caused a small magnet to rotate, and from its extreme angular deviation the quantity of charge q_{emu} was deduced from the appropriate formula. The ratio of the two charges was found to be $k = 3.1074\times10^8$ m/s.

By the time Maxwell composed his *Treatise*, he was aware of at least three direct measurements (in m/s) of the speed of light: 3.14×10^8 (by Fizeau), 2.98×10^8 (by Foucault), and 3.08×10^8 (by measurement of aberration and the Sun's parallax); and three measurements of the ratio of electric units (likewise in m/s): 3.11×10^8 (by Weber), 2.82×10^8 (by Thomson), and 2.88×10^8 (by Maxwell, himself).[6] Although Maxwell was not a person to jump to conclusions, I am nevertheless struck by the subdued expression of his satisfaction:

> "It is manifest that the velocity of light and the ratio of the units are quantities of the same order of magnitude. Neither of them can be said to

---

[6] J. C. Maxwell, *Treatise*, Vol 2, p 436.

be determined as yet with such a degree of accuracy as to enable us to assert that the one is greater or less than the other. It is to be hoped that, by further experiment, the relation between the magnitudes of the two quantities may be more accurately determined."

Any other scientist who had just deduced from theory one of nature's fundamental constants might well have become jubilant, if not euphoric. Consider, for contrast, Einstein's response to a question of what he would have done if Eddington's close, but hardly precise, measurement of the gravitational bending of starlight did not bear out the general relativistic prediction: "Then I would feel sorry for the good Lord. The theory is correct." In any event, thus was impressed upon Maxwell the association of electromagnetic waves and light.

The system-dependent dimensions of charge have consequences for every electrical and magnetic quantity: current, potential, permittivity, permeability, resistance, capacitance, inductance, and more—and especially for the various electromagnetic fields. Within the esu system, the electric field **E** is defined by the force law $\mathbf{F} = q\mathbf{E}$. Correspondingly, the force $\mathbf{F} = I\mathbf{l}\times\mathbf{B}$ on a current-carrying wire of length *l* can be used to define the magnetic induction **B** within the emu system[7]. Since the dimension of force $[MLT^{-1}]$ is that of mass × acceleration irrespective of the nature of the force, one has $[q]_{esu}[E]_{esu}$ = $[I]_{emu}[B]_{emu}[L]$ = $[q]_{emu}[B]_{emu}[v]$, or

$$[E]_{esu} = [B]_{emu} = \left[M^{1/2}L^{-1/2}T^{-1}\right] . \qquad (5)$$

From the definitions of the fields and their interrelationships through Maxwell's equations, one can establish that each field (**E**, **D**, **B**, **H**) in the emu system is related to the corresponding field in the esu system by a velocity. For example, from Maxwell's expression of Faraday's law of induction it follows that $B_{emu} = k[v]\, B_{esu}$.

Now in the formulation of the Maxwell equations, one is free to chose any system of units, so long as relations are expressed consistently. In the Gaussian system, for example, **E** is expressed in the esu system and **B** is expressed in the emu system, and the factor $k[v]$ enters each field equation as a conversion factor to maintain this consistency. The resulting wave equation (whose form is independent of the system of units) for a medium with permittivity ε and permeability μ—both constants being dimensionless numbers in the Gaussian system—leads to a phase velocity $u = k\big/\sqrt{\varepsilon\mu}$ that contains *not* the speed of light *c*, but rather the universal constant *k* giving the number of esu's of charge to one emu of charge. The fact that *k* turns out by measurement to have the same numerical value as *c* is wondrous indeed, and strongly supports the belief that light is a form of electromagnetic wave. But nowhere is *c* built into Maxwell's equations at the outset, as one might infer from modern textbooks.

In the calculation that Maxwell himself made, the phase velocity of the resulting wave equation took the form $u = 1\big/\sqrt{\varepsilon\mu}$. For vacuum (or air, to good approximation), the esu values of the material parameters are $\varepsilon = 1,\ \mu = 1/k^2$, whereas the emu values are just the reverse.

What is the best system of electromagnetic units to use? That depends on one's needs. I have always found the Gaussian system particularly suitable for theoretical analysis, for it manifests the intrinsic symmetry of Maxwell's equations, an especially attractive feature in light of relativity theory

---

[7]The magnetic induction can also be defined in terms of the Lorentz force on a moving charged particle. Maxwell did not refer to the Lorentz force, which was introduced after his death. Rather, he defined **B** and **H** in terms of the force on a hypothetical unit magnetic pole, a construct that as far as we know, still has no realisation in nature.

(unknown, of course, to Maxwell working in the 1860's). Thus, a phenomenon interpretable in terms of **B** in one inertial reference frame may be ascribed to the effect of **E** in another inertial frame. Since **E** and **B** are intimately related by a Lorentz transformation, it is physically significant that they have the same dimensions, as indicated in Eq. (5). On the other hand, the Gaussian system is not particularly convenient for practical work.

The reader might be interested to learn that Maxwell played a significant part in establishing the familiar set of electromagnetic units used in the laboratory—and he discussed this, too, in his *Treatise*. If one adopts the standard metric units of length (cm or m), time (sec), and mass (g or kg), then the units of resistance and electromotive force are too small to express laboratory measurements conveniently, and the units of charge and capacitance are correspondingly too large. The system of practical units (volt, ohm, farad, coulomb...) was initially based on selection of a unit of length of 10^7 m (the length of a quadrant of a meridian of the earth, according to Maxwell[8]) and a unit of mass of 10^{-14} kg. It is to achieve this convenience of scale, therefore, that the familiar, yet mysterious, constants ε_0 and μ_0 are inserted into the fundamental equations, dimensional consistency requiring that $1/\sqrt{\varepsilon_0\,\mu_0} = c$.

The "Electrotonic State" and Maxwell's Perspicacious Insight

Students and physicist colleagues have occasionally asked me why the electromagnetic fields are designated by their particular letters. I have seen letters to editors of various physics journals also pose this question from time to time. In a few cases the choice is self-evident, as in **E** for electric field and **D** for displacement field. But what is one to make of **B** and **H** for magnetic fields? Is there some scientifically significant language (Latin, Greek, German, French, ...) for which the names of these fields correlate with the choice of symbol? I suspect not, but speculate instead that Maxwell simply designated all his electromagnetic variables in an alphabetical order: **A** (vector potential), **B** (magnetic induction), **C** (electric current), **D** (displacement), **E** (electric field: Maxwell's electromotive intensity), **F** (mechanical force), **H** (magnetic field: Maxwell's magnetic force), etc. If so, herein lies another significant insight into Maxwell's thinking—an insight that until recent times has largely been lost in modern presentations of the subject.

That the *first* letter of the alphabet is assigned to the vector potential is not, I believe, an arbitrary choice, but signifies instead the signal importance which Maxwell attached to this function. **A** is the mathematical embodiment of what Faraday, in his qualitative but perceptive reasoning, had termed the "electrotonic state", a "peculiar electrical condition of matter" whereby an isolated circuit remains unaffected by a constant electromagnetic field, but produces a current if the same state of the field were brought into existence suddenly. "The whole history of this idea in the mind of Faraday," Maxwell wrote,

> "...is well worthy of study. By a course of experiments, guided by intense application of thought, but without the aid of mathematical calculations, he was led to recognise the existence of something which we now know to be a mathematical quantity, *and which may even be called the fundamental quantity in the theory of electromagnetism*.." (The italics are my own, not Maxwell's.)

To Maxwell this fundamentality lay in the distinct dual purposes **A** served in the dynamical theory of the electromagnetic field. First, he introduced **A** as the potential function (hence the name: vector potential) from which the magnetic induction was obtained by $\mathbf{B} = \mathrm{curl}\mathbf{A}$, in analogy to a scalar magnetic potential of which **H** was the gradient. Later in the *Treatise*, upon investigating induction in a secondary circuit by current changes in a primary circuit, Maxwell showed that **A**—termed at that point

---

[8] J. C. Maxwell, *Treatise,* Vol. 2, p 268.

the "electrokinetic momentum"—was related to the time integral of the electric field. In this capacity **A** not only determined an induced electric field by the relation $\mathbf{E} = -\partial\mathbf{A}/c\partial\mathbf{t}$ (in the Gaussian system), but was also interpretable as a linear momentum of the electromagnetic field which could be communicated to a charged particle in the secondary circuit if the primary current were suddenly stopped. This dual significance of **A**, as both a potential and a form of momentum, was to have profound implications in the quantum theory of matter and electromagnetic fields.

Potentials, Fields, Forces, and Measurements

In the form of Maxwell's theory that one presently encounters, condensed and simplified by Heinrich Hertz and Oliver Heaviside with full employment of the vector analysis of Willard Gibbs, it is the four electromagnetic fields, not the electromagnetic potentials, that are conceptually fundamental. The fields are regarded as physically real, for they are directly related to observable forces. The vector and scalar potentials, by contrast, are merely auxiliary functions from which these fields are determined by derivative operations. The potentials, unlike the forces, are not unique, but can be modified by a so-called gauge transformation[9] that leaves Maxwell's equations and the force laws invariant.

The 20th century discovery that electrons have wavelike properties, however, has brought to light (no pun intended) an unexpected physical possibility, the implications of which lend strong support to Maxwell's own views of the interpretation and fundamentality of the vector potential. Is there any way the state of motion of a charged particle passing through a region of space where **A** is non-vanishing, but **E** and **B** are strictly null, can be influenced? To this question classical physics yields an unequivocally negative answer, for under such circumstances there can be no force on the charged particle. Nevertheless, as recognized first by British electron microscopists W. Ehrenberg and R. E. Siday[10] in 1949, and independently by quantum theorists D. Bohm and Y. Aharonov[11] ten years later, such a phenomenon is indeed conceivable.

The Aharonov-Bohm or AB effect, as it is known today, entails the modification of an electron self-interference pattern by passage of a coherently split electron "wave" round (but *not* through) an excluded region of space within which is confined a magnetic field **B** (as illustrated in **Figure 1**). In the figure, the magnetic field is produced by an ideally infinitely long solenoid perpendicular to the page; the only electromagnetic influence at the location of a moving electron is the external vector potential **A**. Quantum theory predicts—and a number of independent experiments have confirmed[12]—that the fringes of the electron interference pattern received on a distant screen should be shifted by the angle $\alpha = \Phi/\Phi_0$ relative to the pattern for **B** = 0, where

---

[9]A gauge transformation of electromagnetic potentials consists of the following. From a given pair of vector and scalar potentials $(\mathbf{A}, V)$ one constructs a new pair $(\mathbf{A}', V')$ by means of an arbitrary gauge function $\Lambda(x,t)$: $\mathbf{A}' = \mathbf{A} + \nabla\Lambda;\ V' = V - \partial\Lambda/c\partial t$. For the case of fields coupled to charged particles, as treated within the framework of quantum mechanics, a gauge transformation also entails a unitary transformation of the particle wave function. See, M. P. Silverman, *A Universe of Atoms, An Atom in the Universe,* (Springer, NY, 2002) pp. 334-342.

[10]W. Ehrenberg and R. E. Siday, "The Refractive Index in Electron Optics and the Principles of Dynamics", *Proc. Phys. Soc. (London)* B**62** (1949) 8-21

[11]Y. Aharonov and D. Bohm, "Significance of Electromagnetic Potentials in the Quantum Theory", *Phys. Rev.* **115** (1959) 485-491

[12]See, for example, N. Osakabe *et. al.,* "Experimental Confirmation of the Aharonov-Bohm Effect Using a Toroidal Magnetic Field Confined by a Superconductor", *Phys. Rev.* A**34** (1986) 815. In this experiment the desired field configuration is achieved with a toroidal ferromagnet covered with a superconducting outer layer. The Meissner effect expels the magnetic flux from the layer thereby confining it to the toroidal interior. See M. P. Silverman, *Quantum Superposition: Counterintuitive Consequences of Coherence, Entanglement, and Interference* (Springer, NY, 2008) pp 11-21 and 239-246

$$\Phi = \oint \mathbf{A} \cdot \mathbf{ds} = \iint \mathbf{B} \cdot \mathbf{dS} \tag{6}$$

is the magnetic flux through the excluded domain, and $\Phi_0 = hc/e = 3.9 \times 10^{-9}$ gauss-cm^2 is the quantum unit of magnetic flux, or fluxon. The contour of the line integral of **A** in Eq. (6) can be any closed path round the excluded domain connecting the points of electron emission and detection.

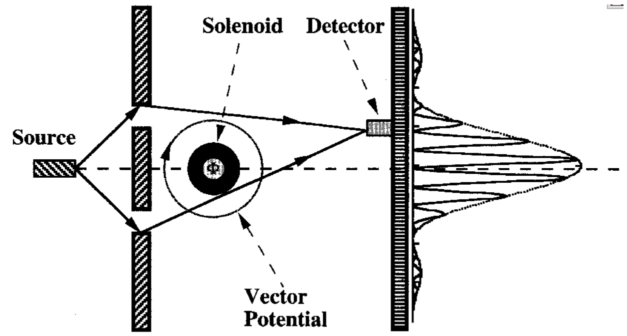

Figure 1. Geometry of the two-slit Aharonov-Bohm effect. Electrons diffract around a long solenoid (perpendicular to the page) and are counted at the detector, giving rise to interference fringes whose locations depend on the magnetic flux Φ confined to the solenoid interior. In the exterior region there is only a cylindrically symmetric vector potential field. The single-slit diffraction envelope is not displaced since, under the circumstances, there is no Lorentz force acting on the electrons.

The startling nature of the effect is illustrated in the displaced fringe pattern of the figure which indicates a value of B such that no electrons at all are received in the forward direction (i.e. along the optic axis). From a classical perspective, the trajectories of the electrons seem to have been shifted despite the absence, under the circumstances, of any Lorentz magnetic force. Quantum mechanics, however, does not permit us to think of electron trajectories in the context of an interference experiment. Correspondence with classical physics is established by the fact that the overall beam, as defined by the single-slit diffraction envelope (and not the two-slit interference pattern), is unaffected by the isolated magnetic field.

As an atomic and nuclear physicist I have given much thought to the implications and novel experimental extensions of the AB effect, which I have discussed in some of my books[13]. Concerning the fundamental quantities of electrodynamics, however, one can say the following. Although the AB fringe shift is dependent on the confined magnetic flux **Φ**—which, according to Eq. (6), can be expressed in terms of either **A** or **B**—the *derivation* of the AB effect, indeed the fundamental starting point for treating *all* electromagnetic interactions of a quantum particle, begins with potentials, not fields or forces. It is the vector potential **A** and scalar potential *V* that enter the Lagrangian and Hamiltonian formulations of quantum mechanics that prescribe the time-evolution of a quantum system. **A** in particular occurs in association with the canonical linear momentum **p** in expressions that reveal its dual role as both a potential and momentum-like quantity—as, for example, in the Hamiltonian operator

$$H = \frac{1}{2m}\left(\mathbf{p} - \frac{q}{c}\mathbf{A}\right)^2 + V \tag{7}$$

of a single non-relativistic charged particle. It is worth noting in this regard that it was Maxwell who introduced Hamiltonian ideas (based on energy and potential) into electrodynamics, thereby sidestepping methods based on force which would have required detailed knowledge of the electromagnetic medium (aether) and nature of electric charge (electron). Maxwell, it should be remembered, knew nothing of the electron, whose discovery in 1897 took place well after his death in 1879.

A question that frequently arises, especially in view of the AB effect, is whether or not the vector potential is a measurable quantity. Maxwell, I believe, certainly thought so, as he associated it with a transferable linear momentum. In quantum mechanics quantities designated as "dynamical observables"—i.e. accessible to measurement—must satisfy two requirements. First, they must be representable by Hermitian (i.e. self-adjoint) operators, since these have real-valued eigenvalues, and the

[13]M. P. Silverman, *And Yet It Moves: Strange Systems and Subtle Questions in Physics,* (Cambridge, NY, 1993); *More Than One Mystery: Explorations of Quantum Interference,* (Springer, NY, 1995); also see footnotes 9 and 12.

outcome of measurement can only be expressed by real numbers.[14] And second, they must be invariant to a gauge transformation, since the latter is analogous to selecting a local coordinate system, and the outcome of a physical measurement can not depend on the arbitrariness of such a choice. Since **A** and V are *not* gauge invariant, they are not, strictly speaking, measurable. In electrostatics, for example, it is always a potential *difference*, ΔV, that is measured.

Having said this, however, it is important to note that a gauge transformation is not entirely arbitrary, for the result of any such transformation must not change the physical presence of the magnetic field $\mathbf{B} = \mathrm{curl}\mathbf{A}$. Now any vector field can be decomposed into transverse and longitudinal components defined by $\mathbf{A} = \mathbf{A}_{\perp} + \mathbf{A}_{\parallel}$ such that $\mathrm{div}\mathbf{A}_{\perp} = 0$ and $\mathrm{curl}\mathbf{A}_{\parallel} = 0$. The actual decomposition takes the form

$$\mathbf{A}_{\perp} = \frac{1}{4\pi}\nabla\times\left(\nabla\times\int\frac{\mathbf{A}(\mathbf{r}')d^3r'}{|\mathbf{r}-\mathbf{r}'|}\right) \tag{8a}$$

$$\mathbf{A}_{\parallel} = -\frac{1}{4\pi}\nabla\int\frac{\nabla'\cdot\mathbf{A}(\mathbf{r}')d^3r'}{|\mathbf{r}-\mathbf{r}'|}\quad , \tag{8b}$$

and one can show with a little effort that the sum of Eqs. (8a) and (8b) does indeed reduce to **A**(**r**). A gauge transformation modifies only the component , leaving the component and the physical field **B** = curl unchanged. The transverse part of the vector potential is a measurable quantity.

Confirmation of Light as an Electromagnetic Wave

Maxwell did not live to see the confirmation of his greatest prediction—the existence of electromagnetic waves—which was demonstrated experimentally by Heinrich Hertz in 1889, some ten years after Maxwell's death. Hertz's investigations are ingeniously simple and of profound significance—an inspiration to anyone who, like me, takes pleasure in small-scale table-top experiments. And yet I can not recall ever seeing a discussion of these seminal experiments in any of the optics or electrodynamics textbooks from which I have studied or taught. Fortunately, early in my career I picked up for a pittance a collection of Hertz's papers[15] in a secondhand bookshop, and have had the pleasure of following Hertz's thoughts and actions in his own words.

Hertz, interestingly enough, did not set out to find electromagnetic waves. On the contrary, his investigations were initially motivated by an entirely different objective brought to his attention by his mentor, the leading German physicist Hermann von Helmholtz. Hertz writes:

> "The general inducement was this. In the year 1879 the Berlin Academy of Science had offered a prize for research on the following problem:—To establish experimentally any relation between electromagnetic forces and the dielectric polarisation of insulators—that is to say, either an electromagnetic force exerted by polarisations in non-conductors, or the polarisation of a non-conductor as an effect of electromagnetic induction."

---

[14]One might inquire why a complex number—in effect, a coupled pair of real numbers—can not serve as a measurement outcome. The answer is that two measurements would be required to furnish the two numbers, and the order of these observations would matter, since quantum measurements are not, in general, commutative.

[15]Henrich Hertz, *Electric Waves,* (Dover, NY, 1962), an unabridged republication of the work first published in 1893 by Macmillan and Company. Quotations in the text are taken from the Introduction, pages 1-20, and from his papers, "On the Finite Velocity of Propagation of Electromagnetic Actions", p. 109, and "On Electromagnetic Waves in Air and Their Reflection", p. 136.

The term "polarisation" here has no bearing on light, but refers instead to the separation of electrical charge (although, again, it must be remembered that in 1879 discrete units of electrical charge were still unknown). In brief, the focus of attention of the Berlin Academy, and ultimately of Hertz, was on Maxwell's predicted *displacement current*, the existence of which was by no means generally accepted at the time even in Britain, let alone in Germany. Representable (in the Gaussian system) by a density

$$\mathbf{J}_d = \frac{1}{4\pi}\partial(\varepsilon\mathbf{E})/\partial t \tag{9}$$

dependent exclusively on the temporal variation of a *neutral* "substance" (the electric field), the displacement current raised troublesome questions concerning the nature of electricity and the closure of electrical circuits.

According to Eq. (9), one might be able to detect the effects of a displacement current in an insulator if he had the means to generate rapidly oscillating electric fields. In this endeavour Hertz had good fortune, for in the collection of physical instruments at the Technical High School at Karlsruhe, where he carried out his investigations, he had earlier found—and used for lecture purposes—a pair of so-called Riess or Knochenhauer spirals. The discharge of a small Leyden jar through one of the spirals, Hertz discovered, amply sufficed to produce sparks in the other, "provided it had to spring across a spark gap." Upon optimising conditions, he eventually succeeded "in obtaining a method of exciting more rapid electric disturbances than were hitherto at the disposal of physicists." Thus did Hertz auspiciously embark upon his researches.

But the work did not proceed well—i.e. as Hertz hoped it would. Actually, it proceeded only too well, although he did not at first recognise it. Having a means of generating rapidly oscillating electric sparks between the terminals of a spark gap in a primary circuit, Hertz attached to each terminal a conducting plate, inserted between the two plates an insulator, and endeavoured to determine the effect of the insulator on electrical oscillations induced between the gap of a separate loop antenna as schematically shown in **Figure 2**.

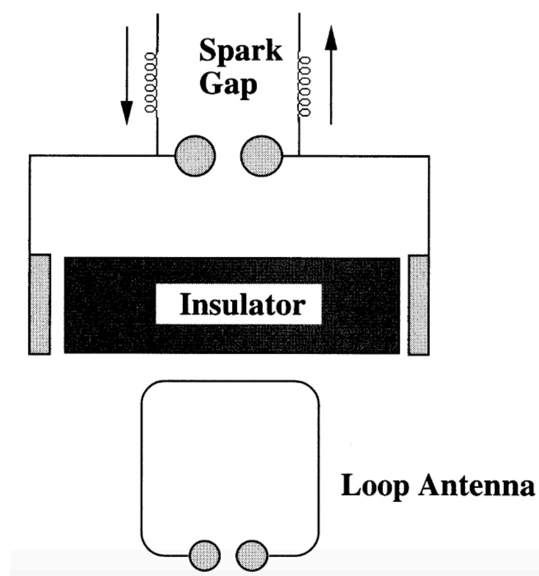

Figure 2: Diagram of Hertz's apparatus designed to test the effect of an insulator on the sparks induced in a secondary circuit (loop antenna) as a result of electrical oscillations in a primary circuit (spark gap and conducting plates).

The expectation was that "when the block was in place very strong sparks would appear in the secondary, and that when the block was removed there would only be feeble sparks." The basis for this expectation was that electrostatic forces—forces derivable from a potential and which ordinarily diminish rapidly with distance from the source—could not induce sparking in a nearly closed secondary circuit (since their integral over a nearly closed contour ought to be vanishingly small). Any sparking, therefore, would have to be induced by the displacement current in the dielectric block. But this was *not* what occurred. "The experiment," Hertz wrote, "was frustrated by the invariable occurrence of strong sparking in the secondary conductor" whether the insulator was present or not.

Gradually it became clear to Hertz that he was not dealing with static or quasi-static fields, as had been commonly the case among electrical investigators up to that time, but with a field of such high frequency that only the laws of a true *electrodynamics* would be applicable. "I perceived that I had in a sense attacked the problem too directly," Hertz concluded somewhat understatedly. More importantly, he also perceived that the particular problem of the Academy, which till then had served as his guide, could be approached in an entirely different and more productive way. Since air—and indeed empty space—according to Maxwell's theory ought to behave like all other dielectrics, there was no need really to look for the effects of the displacement current generated in a solid dielectric. A more worthy and attainable goal, Hertz decided, would be to look for the direct transmission of an electrical signal through the air and to measure its rate of propagation.

The rest, as one says, is history. Connecting a powerful induction coil to a spark gap between two large square brass conducting plates (providing capacitance), Hertz probed the presence of a transmitted field at various distances along a horizontal baseline perpendicular to the gap and the plane of the plates (**Figure 3**). Employing as his detector a circular or square loop antenna with spark gap resonant with the primary circuit, Hertz records, "I was able to observe the sparks [in the antenna] along the whole distance (12 metres) at my disposal, and have no doubt that in larger rooms this distance could be still farther extended." At any given distance, however, the induced sparking could be terminated by rotating the antenna; indeed it is part of the ingenious simplicity of the experiment that, not only did Hertz detect the transmission of electromagnetic waves, but he simultaneously confirmed their transverse polarization by appropriately orienting the plane of the secondary circuit. "The reason is obvious:" Hertz wrote of the cessation of sparking, "the electric force is at all points perpendicular to the direction of the secondary wire."

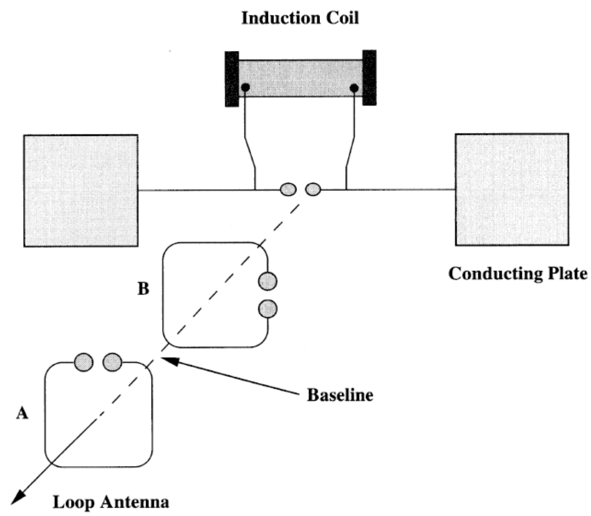

Figure 2: Diagram of Hertz's apparatus designed to test the effect of an insulator on the sparks induced in a secondary circuit (loop antenna) as a result of electrical oscillations in a primary circuit (spark gap and conducting plates).

There remained the intriguing question of the speed of propagation of these newly discovered waves. To this end, Hertz modified his apparatus, adding an additional brass plate parallel to, and a short distance behind, one of the original plates, and extending from that plate a long straight copper wire parallel to the baseline. The wire passed through the window of his laboratory for a distance of some 60 metres and ended freely in the air. By passing one of his tuned loop antennas nearby along the length of the wire, Hertz observed the periodic increase and decrease of sparking characteristic of a standing wave pattern. From the electrical properties of the primary circuit he estimated the period of oscillation to be 0.14 ns, later shown by Poincaré to be an overestimate by $\sqrt{2}$. Placing paper riders at the nodal (no-sparking) positions on the wire, Hertz determined a wavelength of 2.8 m. Dividing the wavelength by the corrected period of 0.10 ns led to a wave velocity of 2.8×10^8 m/s. With evident satisfaction and pleasure, Hertz concluded

> "...it is clear that the experiments amount to so many reasons in favour of that theory of electromagnetic phenomena which was first developed by Maxwell from Faraday's views. It also appears to me that the hypothesis as to the nature of light which is connected with that theory now forces itself upon the mind with still stronger reason than heretofore. Certainly it is a fascinating idea that the processes in air which we have been investigating represent to us on a million-fold larger scale the same

> processes which go on in the neighbourhood of a Fresnel mirror or between the glass plates used for exhibiting Newton's rings."[16]

What more can one say?

Actually, there *is* one more thread to Hertz's story that must be mentioned, an ironic and adventitious twist of fate such as occurs rarely, but nevertheless does occur. Shortly after devising his system of producing rapid electrical oscillations, but before he could apply it to the examination of displacement current in insulators, Hertz had first to free himself from an earlier and somewhat frustrating investigation.

> "Soon after starting the experiments I had been struck by a noteworthy reciprocal action between simultaneous electrical sparks. I had no intention of allowing this phenomenon to distract my attention from the main object which I had in view,"

he lamented,

> "but it occurred in such a definite and perplexing way that I could not altogther neglect it. For some time, indeed, I was in doubt whether I had not before me an altogether new form of electrical action-at-a-distance."

The puzzling effect that riveted Hertz's attention was the apparent diminution in intensity of the spark in his loop antenna when *nonconducting* materials (glass, paraffin, ebonite,...) were interposed between it and the primary oscillator. This shielding effect persisted irrespective of the distance between the two circuits. By contrast, coarse metal gratings, which in principle are excellent electrostatic screens, showed no shielding effect at all.

After groping in the dark (somewhat literally as well as metaphorically), Hertz established that it was *absorption* of the ultraviolet component of the light from the primary spark that degraded the intensity of the induced spark; reciprocally, the ultraviolet light from the secondary spark *sustained* the spark of the primary circuit when the latter was adjusted sufficiently close to misfiring.

Hertz, who died in 1884, never understood the implications of these observations, experiments undertaken as a side study to the more important task set by the Berlin Academy. He had discovered, in fact, the photoelectric effect—and therefore, even *before* his definitive confirmations of the existence of electromagnetic waves, had provided the first (albeit unrecognised) experimental evidence of the granular nature of light.

---

[16] H. Hertz, *Electric Waves, op. cit.* p. 136