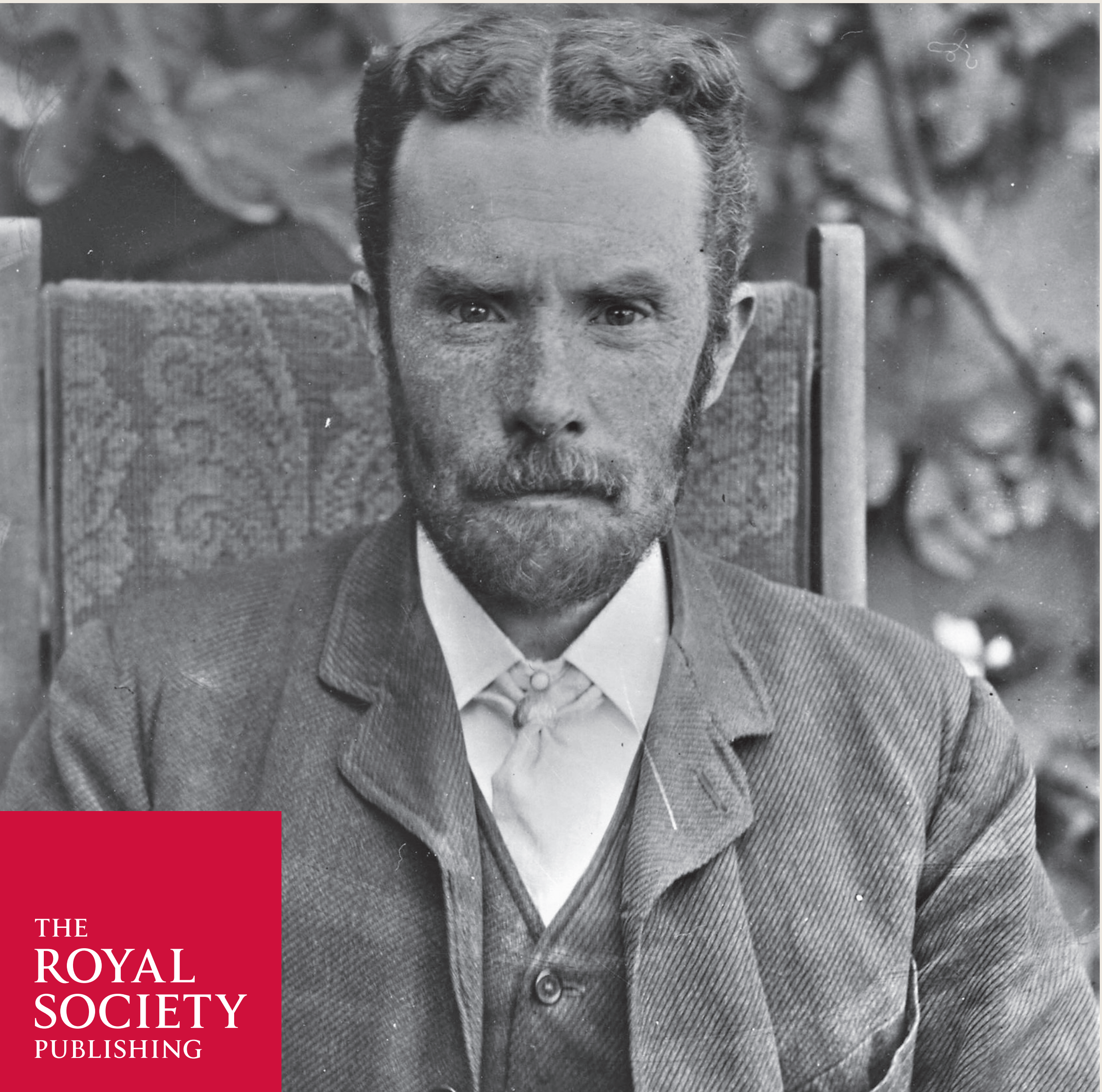
ISSN 1364-503X | Volume 376 | Issue 2134 | 13 December 2018

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A

MATHEMATICAL, PHYSICAL AND ENGINEERING SCIENCES

Celebrating 125 years of Oliver Heaviside's 'Electromagnetic Theory'

Theme issue compiled and edited by Christopher Donaghy-Spargo and Alex Yakovlev



THE ROYAL SOCIETY PUBLISHING

PHILOSOPHICAL TRANSACTIONS A

rsta.royalsocietypublishing.org

Introduction

Cite this article: Donaghy-Spargo C, Yakovlev A. 2018 Oliver Heaviside's electromagnetic theory. *Phil. Trans. R. Soc. A* **376**: 20180229. http://dx.doi.org/10.1098/rsta.2018.0229

Accepted: 14 August 2018

One contribution of 13 to a theme issue 'Celebrating 125 years of Oliver Heaviside's 'Electromagnetic Theory''.

Subject Areas:
electromagnetism

Keywords:
electromagnetism, fellows of the Royal Society, history of engineering, Oliver Heaviside, telegraphy, Victorian science

Author for correspondence:
Christopher Donaghy-Spargo
e-mail: christopher.spargo@dur.ac.uk

Oliver Heaviside's electromagnetic theory

Christopher Donaghy-Spargo[1] and Alex Yakovlev[2]

[1]Department of Engineering, Durham University, South Road, Durham DH1 3LE, UK
[2]School of Engineering, Newcastle University, Merz Court, Newcastle upon Tyne NE1 7RU, UK

CD-S, 0000-0001-5337-4522

The year 2018 marks the 125th anniversary of the first of three published volumes on electromagnetic theory by the eminent Victorian electrical engineer, physicist and mathematician, Oliver Heaviside FRS. This commemorative issue of *Philosophical Transactions of the Royal Society A* celebrates the publication of this work by collecting papers on a broad spectrum across the field of electromagnetic theory, including innovative research papers interspersed between historical perspectives and relevant reviews. Heaviside was a remarkable man, an original thinker with brilliant mathematical powers and physical insight who made many significant contributions in his fields of interest, though he is remembered primarily for his 'step function', commonly used today in many branches of physics, mathematics and engineering. Here, we celebrate the man and his work by illustrating his major contributions and highlighting his great success in solving some of the great telegraphic engineering problems of the Victorian era, in part due to his development and detailed understanding of the governing electromagnetic theory. We celebrate his *Electromagnetic theory*: three volumes of insights, techniques and understanding from mathematical, physical and engineering perspectives—as dictated by J. C. Maxwell FRS, but interpreted, reformulated and expanded by Heaviside to advance the art and science of electrical engineering beyond all expectations.

This article is part of the theme issue 'Celebrating 125 years of Oliver Heaviside's 'Electromagnetic Theory''.

THE ROYAL SOCIETY PUBLISHING

1. 125 years of electromagnetic theory

One hundred and twenty-five years ago, Oliver Heaviside FRS published the first of his seminal masterpiece series on electromagnetic theory. The works contained within this Special Commemorative Issue are scattered across a broad range of electromagnetic topics, including original research articles in the physical and engineering sciences as well as in the history of science and technology, including new perspectives on the work of Heaviside and his subsequent influences. The issue is not intended to provide a comprehensive overview of his work or life, nor is it intended to provide the reader with a showcase of the latest research articles relating to the work of Heaviside. The intention is to act as a mark to commemorate his work *Electromagnetic theory* on the anniversary of the publication of his first dedicated book on the subject in 1893. It is the view of the editors of this commemorative issue, and many learned scholars, that Heaviside was a remarkable man, an original thinker with brilliant mathematical powers and physical insight—his most important work can be found in his three volumes entitled *Electromagnetic theory* and his connected works *Electrical papers*. Here, a brief overview of the man Oliver Heaviside and his main contributions are presented in the broadest of terms. Secondly, we illuminate the main points of interest contained within the three published volumes of *Electromagnetic theory* and discuss the 'missing' fourth volume. These volumes contain his significant works in the subject across a broad spectrum of ideas and it is hoped that the readers of those original works and this special issue are as captivated as the editors by the charm, intrigue and sheer brilliance of Oliver Heaviside FRS—an eminent *electromagnetician*.

2. The man

Oliver Heaviside FRS (portrait in figure 1) was born on 18 May 1850 and died in Torquay on 3 February 1925, at the age of 74. From humble beginnings, he was able to make very many significant scientific achievements to rival the most revered of his university-educated peers in the great Victorian era of scientific discovery. Heaviside was an English self-taught electrical engineer, physicist and mathematician, who changed the face of telecommunications, mathematics and science for years to come, up to and well beyond his death. He adapted complex numbers to the study of electrical circuits, coining such terms as *inductance*, *impedance* and *reluctance*, among many others, all in common use today. Heaviside invented many mathematical techniques, one is his 'operational calculus', a method of solution of systems of differential equations (today's equivalent is the Laplace transform) representing the great telegraphic and electromagnetic problems of his time. He reformulated James Clerk Maxwell's field equations in terms of the electric and magnetic forces ('murdering' Maxwell's potentials along the way, due to their non-observability) while independently co-formulating vector analysis and Prof. Henry Poynting's energy-flux theorem. This impressive catalogue of achievements in the realms of the electromagnetic theory was complemented by his many technical achievements in electrical engineering practice. He also invented the coaxial cable, for which he was granted a patent. He was responsible for the discovery of the necessary conditions required for 'distortionless' transmission of telegraph signals and invented the practical means by which this could be achieved—perhaps among the most important of his achievements. This list is complemented by the many improvements in practical telegraphic and telephonic systems he contributed over his lifetime.

His achievements (with the correct attribution) are not well known, yet they are profound and wide-reaching. The so-called 'Maxwell Equations' or 'General Equations of the Electromagnetic Field' [1], as presented in every undergraduate physics textbook in the world are in fact in the form presented first by Heaviside, in the language that he independently developed for the purpose (alongside W. Gibbs, known for 'Gibbs phenomena' in Fourier theory). This 'recasting' alone is a major achievement, for which Heaviside was given credit by Heinrich Hertz [2]. Maxwell had published his two-volume work *A treatise on electricity and magnetism* in 1873 [3]. Heaviside first came across this seminal work as part of his 'self-study' regime while he was living

rsta.royalsocietypublishing.org Phil. Trans. R. Soc. A 376: 20180229

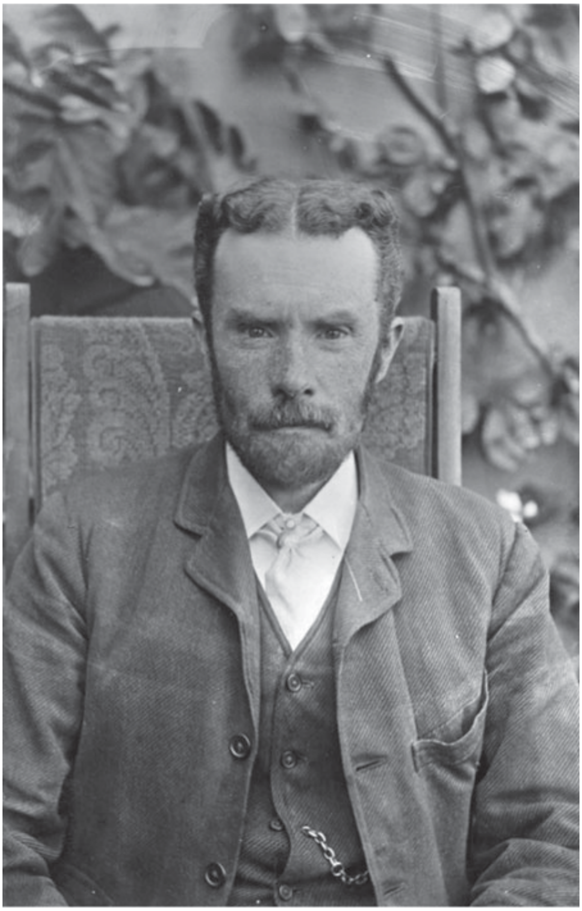


Figure 1. Oliver Heaviside (SC MSS 140 08, courtesy of the Institution of Engineering & Technology Archives).

in Newcastle upon Tyne in late 1873, while working for the Great Northern Telegraph Company (his only paid work, which he left in 1873 at the age of 23 to concentrate on his studies). It was his time in Newcastle that would influence his research interests for the rest of his life. It is the aim of this Special Issue to raise awareness of his very many contributions that have been documented so well in the biographies of Heaviside.

There have been two widely published biographies, the first (and mathematical) biography was written by P. J. Nahin and published in 1988 [4], a second one was written in 2009 by B. Mahon [5]. Both biographies have a broad scope regarding their content covering his ideas, life and works—they are recommended reading for those keen to know the details of Heaviside's fascinating life. The two other biographies are very much less known. The first is a personal account of Heaviside as written by his lifelong friend G. F. C. Searle FRS. The Searle biography was never officially published as a complete manuscript, Searle having written the account in 1949/1950 for the Institution of Electrical Engineers *Heaviside centenary volume* (celebrating 100 years since Heaviside's birth), in which a short version is published [6]. It is understood that the full manuscript was later uncovered in the 1970s by I. Catt, who then self-published *Oliver Heaviside, the man* in 1987 [7], attributing the authorship to Searle with Catt as editor. This 'personal sketch' by Searle, presents a unique insight into the personality of Oliver Heaviside, written, perhaps, by the most fitting person to portray Heaviside an authentic light. H. J. Josephs wrote the second unpublished biography of Heaviside based on his extensive research [8]. The manuscript was unfortunately never published and only two copies are known to exist—one in the Institution of Engineering & Technology Archives [9] and one in the Science Museum Archives in London. Arnold Lynch (1914–2004, known for the optical tape reader which was used in the construction of the Colossus computer during WW2) compiled a list of sources for a biography of Oliver Heaviside [10]. It is left to these references for the reader to indulge in learning the intriguing details of the life, work and times of Heaviside.

He was elected a Fellow of the Royal Society in 1891 for his contributions to the mathematical description of electromagnetic phenomena; this was his first accolade of many. In 1905 he was conferred an honorary doctorate from the University of Göttingen (where K. F. Gauss spent much of his career), in 1908 he received Honorary Fellowship of the Institution of Electrical Engineers, from which he then received the first Faraday Medal in 1921.

3. The first three volumes of *electromagnetic theory*

Heaviside's first installment of his three volumes of *Electromagnetic theory* was published in 1893 (aged 43) by The Electrician Printing and Publishing Co, London, based in Salisbury Court, Fleet Street, London. The publishing company had its own weekly trade journal which was at the forefront of publishing papers relating to all manner of electrical matters, including telegraphy and signalling through the transatlantic cable. *The Electrician* described itself as 'A weekly journal of telegraphy, electricity and applied chemistry' and in May 1895 claimed that it was 'The Oldest and Best English Electrical Journal', a bold claim.

Heaviside was a regular contributor among many eminent electrical men of the time—it is in this journal that Heaviside published many of the articles (in fact he had a long and fruitful relationship with the editors of *The Electrician*) that would ultimately end up published in book form. His *Electrical papers* and *Electromagnetic theory*, which were essentially a collection of previously published papers, are peppered with new insights and humorous (with some serious) commentary. Volume I was published in 1893 [11] and focuses on topics such as a review of current electromagnetic theory from the point of view of Heaviside, the development of his 'Vectorial Algebra and Analysis' and moving these principles to his 'Theory of Plane Electromagnetic Waves'. Here he considers the mathematical description of the guidance of waves by transmission lines, including his distortionless circuit and the practical means by which it is to be achieved. This work sets out Heaviside's major ideas about his electromagnetic theory as interprets Maxwell's prior work and it paves the way for his own and others future studies of the subject.

Electromagnetic theory, volume II, came 6 years later in 1899 [12] (aged 49), the first chapter being a departure from electromagnetic theory, considering not wave propagation, but the *Age of the Earth*—here he uses his mathematics and knowledge of electromagnetic theory to discuss methods of answering such a physical, if not philosophical, question. Heaviside's venture outside the world of electromagnetic theory is perhaps linked to an appendix in volume I where he discusses *A Gravitational & Electromagnetic Analogy*, using his vector language and the 'potential' function to describe the gravitational field and its propagation. The bulk of volume II comprises work relating to the generation, propagation and behaviour of transverse electric and magnetic waves in various media and circuital configurations. He also discusses the existence of compressional electromagnetic waves and devotes pages to his mathematics, particularly his use of divergent series and differential operators arising from the study of natural electromagnetic systems. His non-rigorous use of divergent series was an abhorrence to the pure mathematicians, as discussed in [4]; despite this difficulty regarding his peers' acceptance of his methods, he managed to use them rather successfully in his work and remarked about his general love for mathematical series in his *Electrical papers*, vol. II [13];

> The subject of the decomposition of an arbitrary function into the sum of functions of special types has many fascinations. No student of mathematical physics, if he possesses any soul at all, can fail to recognise the poetry that pervades this branch of mathematics.

The final published installment, volume III, followed in 1912 [14] when Heaviside was 62 years old. Perhaps his most mathematical work to date, this volume comprises two main sections *Waves from Moving Sources* and *Waves in the Ether*. The former deals at great length with the generation of electromagnetic waves by moving sources, including that of the 'electron', as discovered by the British physicist J. J. Thomson PRS in 1896. Here he discusses the acceleration of charged particles, including 'faster-than-light in the medium' particles, leading to predictions of what we now known as Cherenkov radiation and various relativistic effects, including length contraction that he first worked on with G. F. C Searle in 1888 with *motion of electrification through a dielectric* [15]. The latter section presents mathematical descriptions of the movement of energy through the ether, radiation pressure and connected subjects while interspersing the electromagnetic theory with a discussion of *deep-water waves* and another on *the solution of definite integrals by differential transformation*, among other interesting excursions.

These three volumes of 'Electromagnetic Theory' offer a plethora of insights, techniques and understanding from mathematical, physical and engineering perspectives. Despite being published as separate volumes, their combined count exceeding some 1500 pages, the three volumes were later combined and published as a single volume in 1950 as *Electromagnetic theory: the complete & unabridged edition* [16], complete with a *Critical and Historical Introduction* by Ernst Weber (first president of the Institution of Electrical and Electronic Engineers, USA), in which he describes Heaviside:

> Oliver Heaviside, one of the most unusual characters among great modern scientists, could probably be classified best as an outstanding applied mathematician. He was truly a pioneer in this new branch of science.

Later in 1971, another edition of the unabridged three volumes was published again [17], this time with a foreword by Edmund Whittaker FRS, who ranked Heaviside with that of both Poincaré and Ricci, stating that Heaviside's operational calculus is one of the three most important discoveries of the late nineteenth century.

4. The mysterious volume IV

The three volumes of *Electromagnetic theory* contain his most significant works in the subject across a broad spectrum of ideas.[1] That said, at the time of his death in 1925, and for some time before (it was published in 1912, in memory of his friend George Francis Fitzgerald FRS), a fourth volume was in preparation, to which he makes some reference in his preface in volume III:

> Long ago, I had the intention, if circumstances were favorable, of finishing a third volume of this work in about 1904 and a fourth in about 1910. But circumstances have not been favorable.

These unfavourable circumstances are somewhat to do with his health and his disagreements with the new editors of his publisher, The Electrician Printing and Publishing Co [4]. It is understood that Heaviside intended unpublished notes, among some of the articles intended initially for volume III, to be published in this fourth installment; some of which are discussed in the Institution of Electrical Engineers *Heaviside centenary volume* in an insightful article by the General Post Office (GPO) Research Engineer H. J. Josephs [6]. It is claimed that Heaviside had finished the manuscript for volume IV in 1916/1917 and that perhaps a different publisher may have taken on the task of disseminating his work. For whatever reason, this was not to be, and the works intended for volume IV were never published. A reasonable question to ask is, 'Where is the manuscript?', if it was completed and ready for publication. The answer to this question has been sought by many ever since Heaviside's death. Following the purchase of

[1]His other significant work on pure electromagnetic theory is in his two volumes of *Electrical papers* with the multiple part articles on *Electromagnetic induction and its propagation*.

Heaviside's books and papers by the Institution of Electrical Engineers in 1927[2] (the extensive collection now held at the Institution of Engineering & Technology Archives and known as the *Heaviside collection* (see footnote 2)) and extensive searches by some followers of Oliver Heaviside (e.g. H. J. Josephs and Dr W. G. Radley, both of the GPO), a complete manuscript has never been found. However, it is now understood that scattered papers in the *Heaviside collection* were intended to form part of the lost fourth volume [6], with some of these papers being found under the floorboards in Heaviside's old Newton Abbot house in the late 1950s [18]. The fate of the manuscript has been much debated; it has been claimed that it was stolen soon after Heaviside's death, that there is a copy held somewhere in MIT and that there was never a completed manuscript. Perhaps we will never know the full truth, but the loose and scattered unpublished papers contained within the *Heaviside collection* may hold some of the answers and contain even more mathematical and electromagnetic gems—undiscovered since Heaviside first uncovered them. For more information, the reader is directed to the work of H. J. Josephs [18,19] and B. R. Gossick [20] on this intriguing aspect of his *Electromagnetic theory*.

5. Editors' remarks

Each of the papers contained within this special issue has been prepared for this volume. Some of the papers explicitly explore either the work or life of Oliver Heaviside, whereas others have their technical roots founded in Heaviside's electromagnetic theory, the remainder are linked only by their relevance to the topic of electromagnetic theory as a scientific discipline. It is hoped that this diverse mix of articles appeals to the reader, giving historical insight and perspectives interspersed between original research articles and reviews—reminiscent of reading any of Heaviside's five volumes of *Electrical papers* and *Electromagnetic theory*.

Competing interests. We declare we have no competing interests.
Funding. We received no funding for this study.

References

1. Maxwell JC. 1865 A dynamical theory of the electromagnetic field (PDF). *Phil. Trans. R. Soc.* **155**, 459–512. (doi:10.1098/rstl.1865.0008)
2. Hertz H. 1893 *Electric waves: being researches on the propagation of electric action with finite velocity through space*. New York, NY: Dover Publications.
3. Maxwell JC. 1873 *Treatise on electricity and magnetism*, vols. I & II. Oxford, UK: Clarendon Press.
4. Nahin PJ. 1988 *Oliver Heaviside: sage in solitude: the life, work, and times of an electrical genius of the Victorian age*. New York: NY: IEEE Press.
5. Mahon B. 2009 *Oliver Heaviside: Maverick mastermind of electricity*. London, UK: IET.
6. IEEE 1950 *The Heaviside centenary volume*. London, UK: Institution of Electrical Engineers.
7. Searle GFC. 1987 *Oliver Heaviside: the man* (ed. J Catt). London, UK: Westfields Press.
8. Josephs HJ. 1963 *Oliver Heaviside: a biography*. London, UK: H.J. Josephs.
9. Oliver Heaviside, a biography by H J Josephs (UK0108 SC MSS 140/11, IET Archives, London).
10. Lynch AC. 1991 The sources for a biography of Oliver Heaviside. *Hist. Technol.* **13**, 145–149.
11. Heaviside O. 1893 *Electromagnetic theory*, vol. I. London, UK: The Electrician Publishing.
12. Heaviside O. 1899 *Electromagnetic theory*, vol. II. London, UK: The Electrician Publishing.
13. Heaviside O. 1892 *Electrical papers*, vol. II, pp. 201. London, UK: Macmillan and Co.
14. Heaviside O. 1912 *Electromagnetic theory*, vol. III. London, UK: The Electrician Publishing.
15. Heaviside O. 1889 On the electromagnetic effects due to the motion of electrification through a dielectric. *Phil. Mag. S* **27**, 324. (doi:10.1080/14786448908628362)
16. Heaviside O. 1950 *Electromagnetic theory: the complete & unabridged edition*. New York, NY: Spon/Dover.

[2]The Heaviside Collection: 1872–1923 (UK0108 SC MSS 005, IET Archives, London).

17. Heaviside O. 1971 *Electromagnetic theory; including an account of Heaviside's unpublished notes for a fourth volume*'. New York, NY: Chelsea Publishing Co.
18. Josephs HJ. 1959 The Heaviside papers found at Paignton in 1957. *Proc. IEE Part C Monogr.* **106**, 70–76. (doi:10.1049/pi-c.1959.0012)
19. Josephs HJ. 1963 Postscript to the work of Heaviside. *J. IEE* **9**, 511–512.
20. Gossick BR. 1974 Heaviside's 'posthumous papers'. *Proc. Inst. Electric. Eng.* **121**, 1444. (doi:10.1049/piee.1974.0303)

PHILOSOPHICAL TRANSACTIONS A

rsta.royalsocietypublishing.org

Oliver Heaviside and the Heaviside layer

Hugh Griffiths

Department of Electronic and Electrical Engineering, University College London, Gower Street, London WC1E 6BT, UK

HG, 0000-0002-9947-5553

Review

Check for updates

Cite this article: Griffiths H. 2018 Oliver Heaviside and the Heaviside layer. *Phil. Trans. R. Soc. A* **376**: 20170459.
http://dx.doi.org/10.1098/rsta.2017.0459

Accepted: 25 March 2018

One contribution of 13 to a theme issue 'Celebrating 125 years of Oliver Heaviside's 'Electromagnetic Theory''.

Subject Areas:
electrical engineering

Keywords:
Heaviside, Marconi, ionosphere, radio, propagation

Author for correspondence:
Hugh Griffiths
e-mail: h.griffiths@ucl.ac.uk

In an entry in the *Encyclopaedia Britannica* in 1902, Oliver Heaviside had suggested the existence of a reflecting layer in the upper atmosphere to account for long range beyond line-of-sight radio propagation of the type demonstrated by Guglielmo Marconi in 1901, in the first transatlantic radio transmission. In about 1910, William Eccles proposed the name 'Heaviside Layer' for this phenomenon, and the name has subsequently been adopted and used quite widely. This paper describes the basis of Marconi's experiments and various interpretations of the results in the context of Heaviside's wider work. It also describes some later experiments to measure the height of the ionosphere.

This article is part of the theme issue 'Celebrating 125 years of Oliver Heaviside's 'Electromagnetic Theory''.

1. Introduction

Oliver Heaviside was born in Camden, London, in 1850, the son of a wood engraver originally from Stockton-on-Tees. The sister of Heaviside's mother had married Sir Charles Wheatstone (she had actually been Wheatstone's cook, so this must have been an unusual marriage, particularly in Victorian times). Wheatstone took a strong interest in his nephews and encouraged them to learn French, German and Danish, and no doubt sowed the seeds of Heaviside's interest in telegraphy. Heaviside suffered from scarlet fever in his youth, and this left him partially deaf. He left school at 16 and pursued his studies at home. Then, at the age of 18, he took a job (the only paid employment he ever had) with the Great Northern Telegraph Company, working in Newcastle and in Denmark. There seems little doubt that Wheatstone was instrumental in getting him this job. During this time, he had started to study Maxwell's

THE ROYAL SOCIETY PUBLISHING

rsta.royalsocietypublishing.org Phil. Trans. R. Soc. A 376: 20170459

work and began to publish articles in *The Philosophical Magazine* and *The Electrician* on various aspects of circuit and telegraph theory. After six years, in 1874, he left this job and returned to live with his parents. Two years later, the family moved to 3 St Augustine's Road, not far from Heaviside's birthplace. In 1889 (when Heaviside would have been 39), the family moved to Paignton, near Torquay, in southwest England. In the summer of 1908, after his parents had died, he moved to a house called 'Homefield' in Torquay and was looked after by Miss Mary Way, the sister of his brother's wife. In 1916, she left, so from 1916 till his death in 1925 (at the age of 74) he lived entirely alone [1–3].

Heaviside's scientific contributions were extraordinary, especially because he had no university education and was essentially self-taught. The three-volume *Electromagnetic Theory* [4] is especially remarkable, both in its content and in its style. Specific contributions include the formulation of Maxwell's equations in compact vector notation; the development of operational calculus to analyse electric circuits [5,6]; the formulation of the Telegrapher's Equation to analyse transmission lines, the dispersionless transmission line and the role of inductance; and a patent on the coaxial transmission line. These contributions are described in greater detail in the other papers in this Special Issue.

In his work on operational calculus, he defined and used the unit step function which is named after him (the Heaviside step function). The other entity that bears his name is the Heaviside Layer, and the purpose of this paper is to explain the background to this subject and to Heaviside's specific contribution to it. The structure of the rest of the paper is therefore as follows: §2 describes Marconi's experiments in radio transmission, and in particular the first transatlantic radio transmission from Poldhu, Cornwall, to St John's, Newfoundland, on 12 December 1901. Section 3 describes various attempts to explain the success of these experiments in achieving ranges beyond line-of-sight, including Heaviside's own publication on the subject. Section 4 describes two later experiments, by Appleton and Barnett in 1924, and by Breit and Tuve in 1925, to measure the height of the ionosphere, Finally, §5 draws some conclusions.

2. Marconi's experiments

Marconi arrived in England in 1896 at the age of 21, having demonstrated radio transmissions over distances of a few miles, and beyond line-of-sight, at his father's estate near Bologna. He sought funding for his research from the British government and gained the interest and support of William Preece, the Engineer-in-Chief of the Post Office. Marconi made a series of demonstrations, and by March 1897 he had demonstrated transmission of Morse code signals over a range of 6 km on Salisbury Plain. In May 1897, he made the first transmission across the sea, from Flat Holm in the Bristol Channel to Lavernock Point, again at a range of 6 km. In March 1899, he made the first cross-channel transmission, from Wimereux, France, to South Foreland Lighthouse, at a range of about 45 km.

Encouraged by these successes, he turned his attention to the big challenge – the transatlantic path. He engaged John Ambrose Fleming of University College London as a consultant, and Fleming designed and built a transmitter and antenna at Poldhu in Cornwall (figure 1). A similar antenna was installed at Cape Cod, Massachusetts. However, this was damaged in a storm, so instead for the first tests, a simpler system was set up at St John's, Newfoundland, using a wire antenna suspended from a kite (figure 2). The range from transmitter to receiver on this path was approximately 3500 km.

The transmitter at Poldhu was a spark gap, giving a nominal output power of about 15 kW at a frequency of 850 kHz, powered by a 32 bhp engine driving a 25 kW alternator. The output from that was fed to two 20 kW transformers which stepped the voltage up from 2 to 20 kV, thence to a bank of condensers discharged via the spark gap [8]. The receiver was a device known as an 'Italian Navy coherer' or 'mercury coherer' [7]. This used a globule of mercury between iron or carbon plugs in a glass tube. It was invented by J. C. Bose but, controversially, patented by Marconi [9,10].

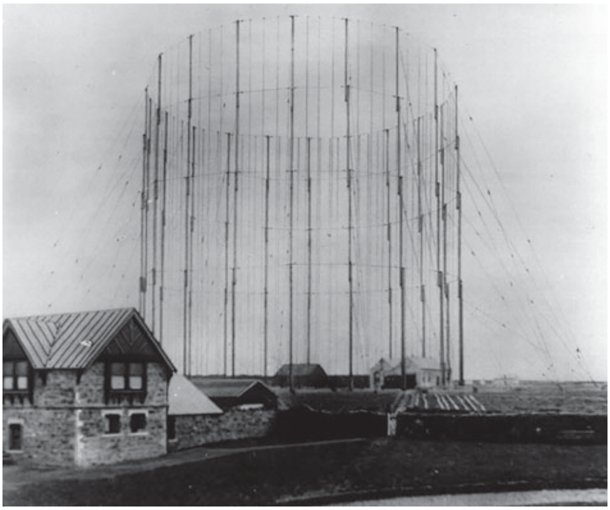


Figure 1. The antenna of Marconi's transmitter at Poldhu [7].

Figure 2. Marconi watching associates raising the kite used to lift the antenna at St. John's, Newfoundland, December 1901 [7].

It should be appreciated that the spark gap transmitter essentially generated an impulse with a broadband spectrum, filtered by whatever selectivity the generating circuit and antenna might possess – so by no means a pure monochromatic tone.

On 12 December 1901 at 12.30, Marconi and his assistant Kemp heard the weak Morse letter 'S', and Marconi recorded in his notebook:

> Sigs at 12:30, 1:10 and 2:20 (local time)

To confirm these results, in February 1902, Marconi sailed from Southampton to the USA on the *SS Philadelphia*, accompanied by several engineers and witnesses, and with a receiver with a 150 foot mast antenna, making measurements of the received signal as the ship progressed. He reported consistent reception of the signals from Poldhu, both by ear and recorded on paper tape, but at longer ranges by night (up to 2500 km) than by day (up to 1120 km).

3. Interpretation of Marconi's results

Heaviside had been engaged to write an entry on 'Telegraphy' for the *Encyclopaedia Britannica* in 1902 [11]. Most of this entry is concerned with the propagation of waves on transmission lines.

However, there is a short passage that considers beyond line-of-sight radio propagation:

> When a wave sent along wires comes to a sharp bend in the circuit, a new wave is generated at the bend. This combined with the old wave forms the wave after passing the bend. There is a rapid accommodation of the wave round the wire to the new direction, but if the bending is continuous, instead of abrupt, the accommodation goes on constantly ... There is something similar in wireless telegraphy. Seawater, though transparent to light, has quite enough conductivity to make it behave as a conductor for Hertzian waves and the same is true in a more imperfect manner for the Earth. Hence the waves accommodate themselves to the surface of the sea in the same way as waves follow wires. The irregularities make confusion, no doubt, but the main waves are pulled round by the curvature of the Earth, and do not jump off.
>
> There is another consideration. There may possibly be a sufficiently conducting layer in the upper air. If so, the waves will, so to speak, catch on to it more or less. Then the guidance will be by the sea on one side and the upper layer on the other.

In a letter to *Nature* in 1927 [12], William Eccles describes how he came up with the name Heaviside Layer, and noted that in early 1902, before the piece in the *Encyclopaedia Britannica*, Heaviside had written a letter to *The Electrician* putting forward the same explanation, but that it was not published. Eccles continues:

> May I explain why I happened to choose the name 'Heaviside Layer' some sixteen years ago?
>
> ... The suggestion [of the existence of the layer] was gradually approved during the years that followed; and about 1910 I used the convenient name 'Heaviside Layer' in a paper, to indicate the portion of the atmosphere that functions so usefully for the purposes of wireless telegraphy.

In the same year, Arthur Kennelly in the USA had put forward an essentially similar explanation [13]:

> At an elevation of about 80 km (50 miles) a rarefaction exists, which, at ordinary temperatures, accompanies a conductivity to low-frequency alternating currents about 20 times as great as that of ocean water. There is well known evidence that the waves of wireless telegraphy, propagated through the ether and atmosphere over the surface of the ocean, are reflected by that electrically conducting surface. On waves that are transmitted but a few miles, the upper conducting strata of the atmosphere may have but little influence. On waves that are transmitted, however, to distances that are large by comparison with 50 miles, it seems likely that the waves may also find an upper reflecting surface in the conducting rarefied strata of the air. It seems reasonable to infer that electromagnetic disturbances emitted from a wireless sending antenna spread horizontally outwards, and also upwards, until the conducting strata of the atmosphere are encountered, after which the waves will move horizontally outwards in a 50 mile layer between the electrically reflecting surface of the ocean beneath, and an electrically reflecting surface, or successive series of surfaces, in the rarefied air above. If this reasoning is correct, the curvature of the Earth plays no significant part in the phenomena, and beyond a radius of, say, 100 miles from the transmitter, the waves are propagated with uniform attenuation cylindrically, as though in two-dimensional space.

While not wanting to detract from Heaviside's contribution, to me at least, Kennelly's explanation is the clearer.

Fleming, too, considered this issue and published a paper [14] on the subject in 1913. However, this was based on the idea that the signal would be refracted round the curvature of the Earth rather than reflected from the ionosphere.

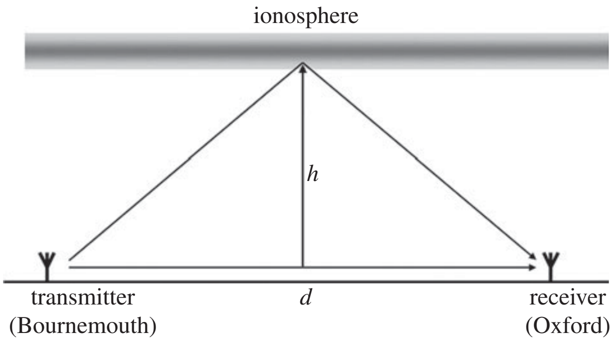


Figure 3. Appleton and Barnett's experiment of 1924.

There is also a suggestion that a Scottish physicist, Balfour Stewart, had proposed the existence of layers in the upper atmosphere some years earlier. He had studied terrestrial magnetism and had made observations of the magnetic storm associated with the Carrington extreme solar event of 1859. However, all of this was well before any experiment or observation of radio propagation.

The fullest analysis of Marconi's results [15] is due to Jack Ratcliffe, a highly respected academic at the Cavendish in the post-war years, and then Director of the Radio and Space Research Station at Slough from 1960 to 1966. He describes the contributions of Heaviside, Kennelly and Eccles noted above. Most significantly, though, he casts doubt on the conventional explanation for Marconi's reception of the transmissions, because the transatlantic transmission path was entirely in daylight when such long ranges would hardly be feasible. Instead, he models the time-domain waveform of the signal from the spark gap transmitter, also taking into account the frequency response of the antenna. From this, he calculates the spectrum of the transmitted signal. This has a peak around 800 kHz, as expected, but also a broad peak around 3.5 MHz. He suggests that it is much more likely that this was responsible for the signals that Marconi heard. A further factor is that the coherer receiver used at St John's was untuned, and the kite-borne receiver antenna would have had very little selectivity and hence would have been able to receive the signals at 3.5 MHz, while a tuned antenna would not.

4. Measurement of the height of the ionosphere

It was not until 1924 that the height of the ionosphere was measured. This was undertaken by Appleton and Barnett in the UK, using an elegant radar technique which is worth noting because it was the first example of frequency modulation (FM) radar (in which the range information is derived from linear FM of the transmitted signal), and also the first example of what is now known as passive radar (where the radar exploits an existing broadcast or communications transmitter [16]). Both of these techniques are now widely used. In this case, the experiment used a BBC broadcast transmitter at a frequency of about 770 kHz, located at Bournemouth on the south coast of England, and a receiver at Oxford at a distance of some 120 km.

The experiment relied on the signal propagating via two paths – the direct (ground) ray and the path reflected from the ionosphere (figure 3). In a letter dated 2 January 1925 to Balthazar van der Pol, Appleton noted:

> I calculate that at about 100 miles from a station [broadcasting] the low rays should be of equal amplitude & we should get fading – and we <u>do</u>.

In order to measure the range difference, the frequency of the transmitter was varied. The description of the experiment [17] reports that:

> . . . Capt. A.G.D. East of the BBC arranged the transmitter so that a known small change of wavelength $\delta\lambda$ (e.g. 5 to 10 m) could be uniformly produced in a given time (e.g. 10 to 30 s) . . . No [amplitude] modulation of the wave took place during the experiment.

The result of the measurement was approximately $h = 90$ km. Taking this value and $d = 120$ km gives a time difference of arrival of the signals via the two paths of approximately 3.2×10^{-4} s. Taking $\delta\lambda = 10$ m and the frequency sweep time to be 10 s gives a beat frequency at the receiver of about 0.6 Hz, which would have been very suitable for displaying on a galvanometer (the detection method used). Also, the resolution of the measurement (determined by the sweep bandwidth) is about 10 km – so quite coarse.

Later, in 1925, Breit and Tuve in the USA used a pulsed radar technique to make similar measurements [18]. The configuration used was similar to that of Appleton and Barnett, with transmitter and receiver separated by a significant distance, and measuring the time difference of arrival at the receiver between the direct signal and that reflected from the ionosphere. The description of the transmitter hardware is quite detailed. In particular, the transmitter at the Naval Research Laboratory (NRL) in Washington DC is described as using a quartz crystal oscillator to give a precise frequency. Several different transmitters were used at different locations in these experiments, at various dates in the spring and summer of 1925, and at various times of the day. The results ranged from 55 to 141 miles (88–226 km), allowing a picture to be built up of the variation with time of day and season of the year and giving insight into the physics behind ionospheric propagation. It is also interesting to observe how far transmitter technology had advanced in the 23 years since Marconi's experiments, from crude spark gap transmitters to stable sinusoidal oscillators and valve power amplifiers.

Through these experiments, and others, the subject of ionospheric radiophysics became well and truly established. We can note, too, that Robert Watson Watt, who led the development of British radar in the late 1930s and was responsible for the British Chain Home air defence radar system in WW2, came from a background in ionospheric physics.

5. Conclusions

Some conclusions can be drawn from these accounts. Firstly, the appellation Heaviside Layer is due to Eccles. Heaviside's first writing on the existence of a reflecting layer was in an unpublished letter, prior to the entry in *Encyclopaedia Britannica* in 1902. Kennelly's publication on the subject was slightly later, but arguably clearer in its explanation, so the names Heaviside Layer or Kennelly–Heaviside Layer are entirely reasonable. Secondly, Ratcliffe's explanation of the success of Marconi's first experiment, in terms of a signal component at a higher frequency, is very likely correct. Thirdly, the measurements of the height of the ionosphere, in 1924 by Appleton and Barnett, and in 1925 by Breit and Tuve, represent some elegant 'firsts' in the history of radar and laid the foundations for the subject of ionospheric radiophysics.

Data accessibility. This article has no additional data.
Competing interests. I have no competing interests.
Funding. This work has not received any funding.
Acknowledgements. I am most grateful to Brian Austin and Bill Liles for their critical and helpful comments.

References

1. Nahin P. 1988 *Oliver Heaviside – Sage in solitude*. New York, NY: IEEE Press.
2. Searle GFC. 1987 *Oliver Heaviside, the man*. St Albans, UK: C.A.M. Publishing.
3. Griffiths HD. 2006 Oliver Heaviside, Ch.7. In *History of wireless* (ed. TK Sarkar). Hoboken, NJ: Wiley.
4. Heaviside O. 1893 *Electromagnetic theory*, vols. 1–3. New York, NY: Chelsea Publishing Company.

5. Kullstam PA. 1991 Heaviside's operational calculus: Oliver's revenge. *IEEE Trans. Educ.* **34**, 155–156. (doi:10.1109/13.81595)
6. Kullstam PA. 1992. Heaviside's operational calculus applied to electrical circuit problems. *IEEE Trans. Educ.* **35**, 266–277. (doi:10.1109/13.168699)
7. Bussey G. 2000 *Marconi's Atlantic leap*. Coventry, UK: Marconi Communications.
8. Belrose JS. 2006 The development of wireless telegraphy and telephony, and pioneering attempts to achieve transatlantic wireless communications, ch. 12. In *History of wireless* (ed. TK Sarkar). Hoboken, NJ: Wiley.
9. Phillips VJ. 1993 The 'Italian Navy coherer' affair: a turn-of-the-century scandal. *IEE Proc. A Sci.* **140**, 175–185. (doi: 10.1049/ip-a-3.1993.0029)
10. Bondyopadhyay PK. 1998 Sir J. C. Bose diode detector received Marconi's first transatlantic wireless signal of December 1901 (the 'Italian Navy Coherer' Scandal Revisited). *Proc. IEEE* **86**, 259–285. (doi:10.1109/5.658778)
11. Heaviside O. 1902 Telegraphy. In *Encyclopaedia Britannica*, p. 214. Edinburgh and London: Adam and Charles Black.
12. Eccles WH. 1927 Wireless communications and terrestrial magnetism. *Nature* **119**, 157. (doi:10.1038/119157a0)
13. Kennelly J. 1902 Research in telegraphy. *Elect. World Eng.* **6**, 473.
14. Fleming JA. 1913 On atmospheric refraction and its bearing on the transmission of electromagnetic waves around the Earth's surface. *Proc. Phys. Soc. Lond.* **26**, 318–333. (doi:10.1088/1478-7814/26/1/333)
15. Ratcliffe JA. 1974 Scientists' reactions to Marconi's transatlantic radio experiment. *Proc. IEEE* **121**, 1033–1038. (doi:10.1049/piee.1974.0242)
16. Griffiths HD, Baker CJ. 2017 *An introduction to passive radar*. Norwood, MA: Artech House.
17. Appleton EV, Barnett MAF. 1925 On some direct evidence for downward atmospheric reflection of electric rays. *Proc. R. Soc. Lond. A* **109**, 621–641. (doi:10.1098/rspa.1925.0149)
18. Breit T, Tuve MA. 1926 A test of the existence of the conducting layer. *Phys. Rev.* **28**, 554–575. (doi:10.1103/PhysRev.28.554)

PHILOSOPHICAL TRANSACTIONS A

rsta.royalsocietypublishing.org

Research

Check for updates

Cite this article: Bruton E. 2018 From theory to engineering practice: shared telecommunications knowledge between Oliver Heaviside and his brother and GPO engineer Arthur West Heaviside. *Phil. Trans. R. Soc. A* **376**: 20170454.
http://dx.doi.org/10.1098/rsta.2017.0454

Accepted: 6 August 2018

One contribution of 13 to a theme issue 'Celebrating 125 years of Oliver Heaviside's 'Electromagnetic Theory''.

Subject Areas:
electrical engineering

Keywords:
telecommunications, history, Oliver Heaviside, electrical engineering

Author for correspondence:
Elizabeth Bruton
e-mail: elizabeth.bruton@gmail.com

From theory to engineering practice: shared telecommunications knowledge between Oliver Heaviside and his brother and GPO engineer Arthur West Heaviside

Elizabeth Bruton

Curator of Technology and Engineering, Science Museum, London, UK

EB, 0000-0002-6981-5705

In May 1900, renowned General Post Office (GPO) engineer Arthur West Heaviside gave the Inaugural Address of the Institution of Electrical Engineers Newcastle local section. With a career spanning the pre-Telegraph Act private telegraph networks as well as the subsequent GPO management and licensing of British inland telecommunications, Arthur Heaviside outlined his innovative and experimental work with all three forms of telecommunication in his various GPO engineering roles based in Newcastle. Omitted from the address was the contribution made by Arthur's younger brother, Oliver Heaviside. Throughout Arthur's career at the GPO, the two brothers exchanged frequent correspondence—some of which has survived in the IET Archives—and Arthur regularly consulted his brother about his experimental work and published papers, incorporating his brother's ideas, suggestions and corrections. The two brothers informally collaborated and published separately upon two key areas of experimentation: duplex telegraphy and the 'bridge system' of telephony. The separate publication of the brothers' work in telecommunications was notable: senior and influential GPO electrical engineer William Preece strongly resisted the theoretical work of Oliver Heaviside and other so-called Maxwellians. It was not until the 'Kennelly–Heaviside layer', independently

THE ROYAL SOCIETY PUBLISHING

proposed by Oliver Heaviside and American electrical engineer Arthur Kennelly in 1902, was experimentally demonstrated in the 1920s that the GPO began to formally engage with the work of Oliver Heaviside. This paper will explore the difficult and complex relationship between Preece and the two Heaviside brothers and how these personal relationships reflect the wider reception of Maxwellian ideas and theorists in British electrical engineering as well as the engineering practice of the GPO, a state institution that could be both innovative and resistant to change in equal measure.

This article is part of the theme issue 'Celebrating 125 years of Oliver Heaviside's 'Electromagnetic Theory''.

1. Introduction

In May 1900, renowned General Post Office (GPO) engineer Arthur West Heaviside (henceforth AWH) gave the Inaugural Address of the Newcastle local section of the Institution of Electrical Engineers where he outlined the 'more recent period [of electrical industry in the Newcastle district] which is scored all over with the honourable marks of discoveries and inventions which have mainly assisted in making electricity the servant of man' from the origins of electrical telegraphy to electric lighting and traction (figure 1) [1]. With a career spanning the pre-Telegraph Act private telegraph networks as well as the subsequent GPO management and licensing of British inland telecommunications, AWH also highlighted his innovative and experimental work with all three forms of telecommunication in his various GPO engineering roles based in Newcastle; the address was also published in an amended form and without illustrations in the GPO staff journal, *St Martin's-le-Grand* in 1901 [2].

In the early 1880s, AWH conducted experiments with an innovative ABC telegraph exchange for mercantile communication in Newcastle, in particular collieries and shipping offices, and devised a metallic circuit system of telephones known as the 'North East (Northern)' system, moving away from rented telegraphy lines. In 1887, AWH and his younger brother Oliver Heaviside (henceforth OH) began working on a paper on the innovative 'bridge' (or parallel circuit) system of telephony on which AWH had been working to improve the clarity of communication on long-distance telephone lines in Newcastle and to which OH had been providing theoretical understanding, in particular his propagation theory. This would have been the two brothers first joint publication and a keystone in their collaborative partnership but for OH's proposal that self-inductance, loading a circuit with inductance, could be a key contribution to decreasing and eventually removing distortion altogether [3]. GPO senior engineer William Preece strongly opposed self-inductance and OH's more theoretical and Maxwellian work more generally and so Preece, both as AWH's professional superior and as a well-respected figure in electrical engineering, blocked the paper's publication. This led to a lifelong rivalry between Preece and OH, and a triangle of tension between the two Heaviside brothers and Preece.

In the mid-1880s, AWH led experiments into three new types of telegraphy systems without connecting wires at Town Moor, Newcastle, with many specific innovations added by AWH himself. These experiments were part of a systematic experimental programme overseen by Preece to develop systems of telegraphy without connecting wires that could be used in challenging locations such as lighthouses and islands while still connecting into the existing inland telegraphy network, then managed by the GPO [4].[1]

Throughout his 34 year career at the GPO from 1870 and 1904, AWH and OH exchanged frequent correspondence including discussions of theoretical and technical matters, and AWH regularly consulted his brother about his published papers, incorporating the ideas, suggestions

[1]For further details of this experimental programme, see Chapter 2: 'Something in the Air': The Post Office and early wireless experiments, 1882–1899 of my PhD thesis [4].

rsta.royalsocietypublishing.org Phil. Trans. R. Soc. A 376: 20170454

Figure 1. IET Archives UK0108 IMAGE 1/1/0195 Heaviside, Arthur West, head and shoulders portrait, n.d. Image courtesy of IET Archives.

and corrections of his brother.[2] Such was the closeness of their work that authorship was often collaborative and joint, and in 1880 the two brothers applied for a joint patent 'Electrical conductors' but AWH's role in the GPO meant that the patent was granted solely to OH.

This remained true throughout their lifetimes, the two brothers collaborated informally and published separately upon three key areas of experimentation: duplex telegraphy, the 'bridge system' of telephony and 'cross talk' (interference between telegraph wires and telephone lines). These separate publications in the field of telecommunications were, in part, due to their increasingly divergent personalities and working environments: AWH was a loyal company man and a practically minded engineer who worked in the northeast GPO district for most of his career, while OH had a brief period of paid employment working for the Great Northern Telegraph Company before, partially due to his increasing deafness, choosing to devote himself to independent scientific research and electrical theory. The two brothers' collaboration (and sometimes lack thereof) was also heavily impacted by senior and influential GPO electrical engineer William Preece, AWH's professional superior and an 'electrician' who strongly resisted and at times suppressed the newer theoretical and mathematical work of OH and other Maxwellians.

This paper will explore the difficult and complex relationship between Preece and the two Heaviside brothers and how these personal relationships reflect the wider reception of Maxwellian ideas and theorists in British electrical engineering as well as the engineering practice

[2]IET archives, London, holds many examples of correspondence between AWH and OH although it is clear from the large gaps and references within the correspondence that this is largely incomplete.

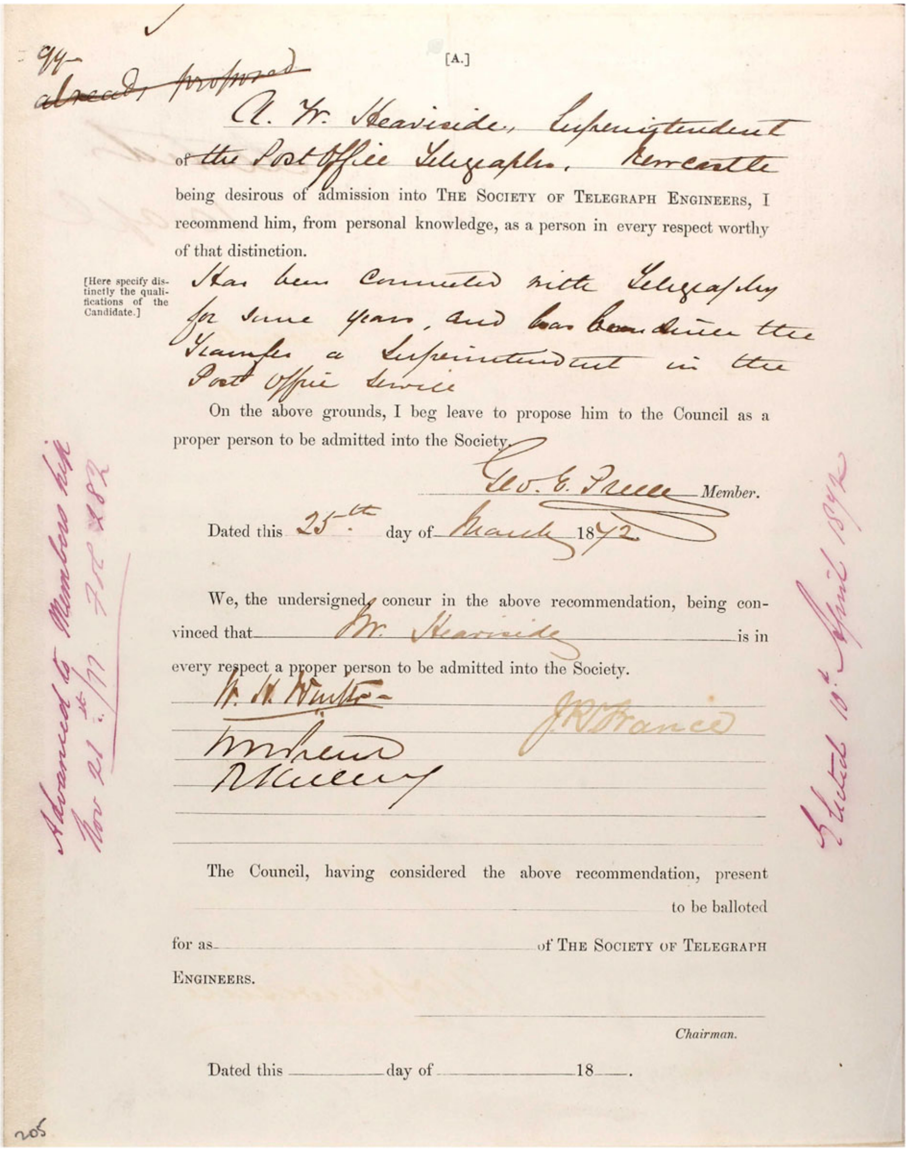
Figure 2. IET Archives UK0108 IMAGE 1/1/0196.06 OH, with family group, at the Guard House of the castle at Berry Pomeroy, Totnes, Devon, n.d. OH is at the back with a pipe while AWH is at the front right. Image courtesy of IET Archives. (Online version in colour.)

of the GPO, a state institution that could be both innovative and resistant to change in equal measure.

2. Early lives and duplex telegraphy

AWH was born in 1844 to Thomas Heaviside, a wood-engraver, and Rachel Heaviside, a former governess and then a schoolteacher. Six years later in May 1850, his youngest of four brothers, Oliver Heaviside (OH), was born at 55 King Street, Camden Town, London (figure 2). The two brothers and their siblings were raised in a home later described as 'almost literally Dickensian'—their home was just around the corner from where Charles Dickens had lived during the most miserable part of his own childhood [5]. Their one good fortune was that their aunt Emma West (their mother's sister) had married Charles Wheatstone. Emma, described as 'a lady of considerable attraction', was 11 years younger than her husband and had been described as Wheatstone's cook but was more likely, like her sister Rachel, a governess of some kind before she was married [6,7]. Emma and Charles Wheatstone's first child, Charles Pablo Wheatstone, was born just three months after they married and it was reportedly a very quiet wedding. It was through their uncle, Charles Wheatstone, that both AWH and OH made their way into the telegraphy profession.

AWH joined the Universal Private Telegraph Company at Newcastle on 1 January 1861 and began working his way up the company. When the domestic British telegraph network was transferred to the GPO in 1870, AWH joined the Post Office Engineering Department as District Superintendent for North-East (North) District based at Newcastle and was based there for the entirety of his career until his retirement in 1904. In 1872, AWH sought further professional recognition through membership of the Society of Telegraph Engineers (STE), a techno-scientific institution founded less than a year previously in May 1871 (figure 3). AWH was most likely encouraged to join by William Preece, then Assistant Chief Engineer of the GPO and a founding member of the STE. It was Preece's brother and fellow telegraph engineer George E. Preece who stated AWH's qualifications for membership of the STE, affirming AWH had 'been connected


[A.]

94

~~already proposed~~

A. W. Heaviside, Superintendent of the Post Office Telegraphs, Newcastle

being desirous of admission into THE SOCIETY OF TELEGRAPH ENGINEERS, I recommend him, from personal knowledge, as a person in every respect worthy of that distinction.

[Here specify distinctly the qualifications of the Candidate.] Has been Connected with Telegraphy for some years, and has been since the Transfer a Superintendent in the Post Office Service

On the above grounds, I beg leave to propose him to the Council as a proper person to be admitted into the Society.

Geo. E. Preece Member.

Dated this 25th day of March 1872.

We, the undersigned, concur in the above recommendation, being convinced that Mr. Heaviside is in every respect a proper person to be admitted into the Society.

W. H. Winter

J. R. France

The Council, having considered the above recommendation, present ______ to be balloted for as ______ of THE SOCIETY OF TELEGRAPH ENGINEERS.

______ Chairman.

Dated this ______ day of ______ 18__.

Advanced to Membership Nov 21st/77 F.E.R.

Elected 10th April 1872

205

Figure 3. AWH's STE membership application, submitted 25 March 1872. Image courtesy of IET Archives. (Online version in colour.)

with Telegraphy for some years, and had been given a transfer as Superintendent in the Post Office Service'.[3] His application might have been supported by William Preece, but it is difficult to tell from the signature on the form above.

By now, AWH's younger brother OH was also working in telegraphy: in 1868, OH—self-taught in Morse code and some aspects of electricity—began working for the Dansk-Norsk-Engeslske Telegraf Selskab (Danish-Norwegian-English Telegraph Company, later the Great Northern Telegraph Company) on the Anglo-Danish cable. Initially, OH worked in Denmark but transferred to Newcastle in 1870, where he lived (he lodged with AWH and his wife Isabella) and worked until resigning his post in May 1874.

[3]IET Archives Membership Register A.W. Heaviside, Superintendent of the Post Office Telegraphs, Newcastle, submitted 25 March 1872 and later accepted.

While still employed in the telegraph industry, OH was encouraged by AWH to join the STE, which AWH had joined and been accepted by 2 years earlier. However, AWH became concerned that OH would not be accepted at the STE as he thought the society would not accept 'telegraph clerks' although by then OH had moved beyond his original role as a telegraph clerk to diagnosing and repairing faults with cables and equipment.[4] OH's role in the Great Northern Telegraph Company was described by later historians of telecommunications as 'making available to them engineering talent of the same order as that available to the [Anglo-American Telegraph Company] in Lord Kelvin' with a further note adding it was 'not clear whether Heaviside's genius had yet blossomed or was ever directly beneficial to Northern' [17]. Furthermore, in relation to OH's STE membership application, by then OH and AWH were collaborating together, 'solving such electrical problems as arose' on the GPO system AWH was then in charge of as District Superintendent for the North-East of England and experimenting more generally together in telegraphy [18].

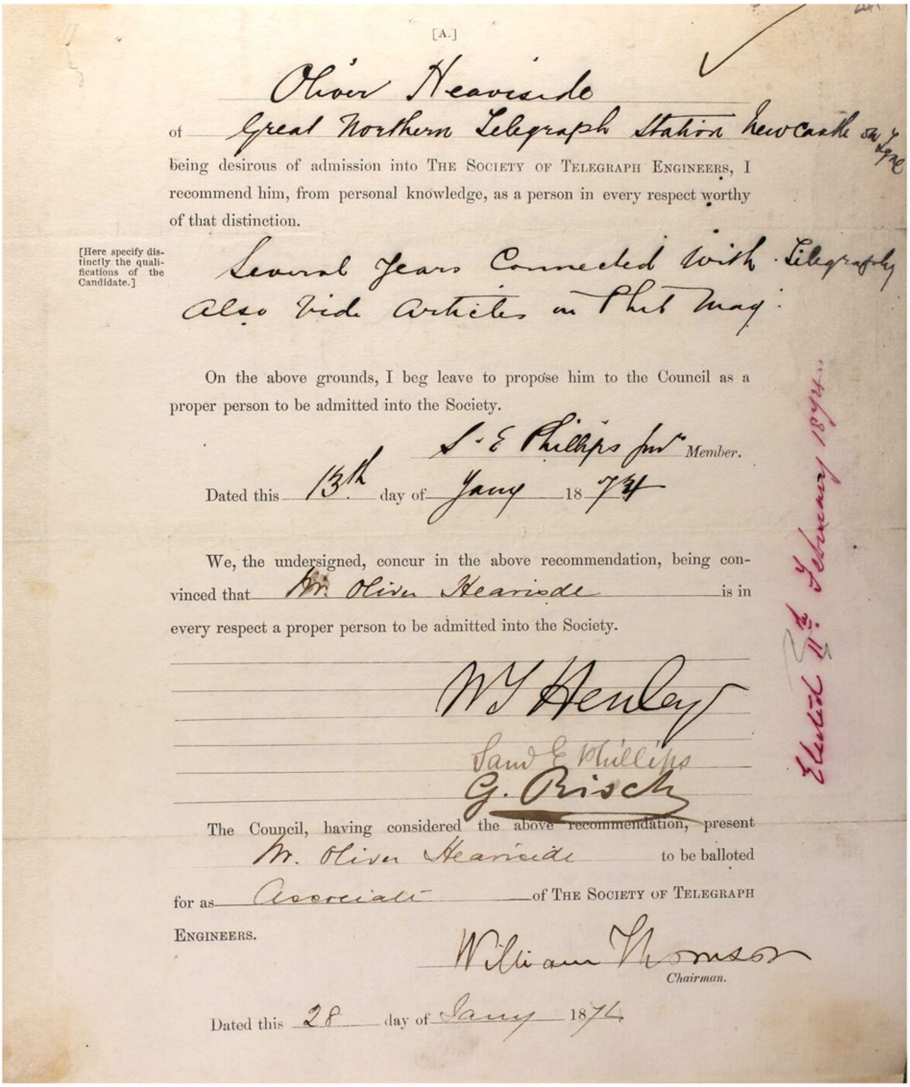
To this end, in January 1873, OH and AWH collaborated together, experimenting with a duplex telegraph system with an artificial line using single-needle instruments and later introduced the system on the telegraph line between Newcastle and Sunderland where messages were sent 'simultaneously from both stations as fast as they could be transmitted by key' and even managed quadruplex operation, sending two messages simultaneously in each direction [18]. The trials were conducted in secrecy as there were other, more senior parties in the GPO engineering department working on duplex telegraphy: William Preece held an 1855 patent for a type of duplex of telegraphy and GPO Engineer-in-Chief R.S. Cullley (then William Preece's immediate superior) was experimenting with a duplex telegraphy system [19]. Fearing a negative impact on AWH's GPO career, OH solely published their researches and results in two parts in the *Philosophical Magazine* in 1873 and 1876 and is also sometimes credited (also alone) as the inventor of quadruplex telegraphy [20,21]. A copy of the first of OH's two *Philosophical Magazine* duplex telegraphy can be found in the papers belonging to William Preece at IET Archives with handwritten notes, most probably in Preece's hand—Preece unsurprisingly not being an admirer of OH's work because OH did not acknowledge Preece's 1855 patent in duplex telegraphy and also OH ridiculed Culley's work on duplex telegraph in his 1873 article.[5] Preece wrote at that time to Culley that OH's 1873 paper was 'most pretentious and impudent' with OH claiming 'to have done everything, even Wheatstone Automatic duplex. He must be met somehow' to which Culley replied that 'He [OH] claims or is supposed to have brought out lots of other things. We will try to pot Oliver somehow' [22].

By January 1874 when OH submitted his application to join the STE, OH had moved beyond a mere telegraph clerk to experimenting and innovating in electrical telegraphy, with much of this collaboration being with his older brother and GPO engineer, AWH. Evidence of OH's stature in the field of telegraphy was such that in 1874 he approached William Thomson to support his membership application (figure 4). However, Thomson was busy and instead his membership application was proposed by electrical and telegraph engineer Samuel Edmund Phillips Junior, who stated OH's qualifications for membership of the society as being 'several years connected with telegraphy also vide. [see or refer to] articles in Phil Mag'; OH was accepted as an Associate Member in late January 1874 'in spite of the P.O. [Post Office] snobs' most probably Preece and Culley among them.[6]

[4]IET Archives UK0108 SC MSS 005/I/3/01/08 Primary collection of Heaviside papers: two letters from Heaviside to Highfield, President of the IEE, of 5 March 1922 and 14 March 1922, with regard to the IEE honouring Heaviside and securing him a pension. Letters are numbered by Heaviside OH4 and OH5. There are five major publications on OH and none on AWH; in order of publication, they are [8,9] (republished as [10]); [11–13] (recently reprinted in a slightly revised edition as [14]). See also [15,16].

[5]See BT Archives TCB 274/1 Experiments upon wireless telegraphy by methods of induction and conduction, 1886–1900.

[6]IET Archives Membership Register Oliver Heaviside, of the Great Northern Telegraph Station, Newcastle upon Tyne, submitted 13 January 1874 and accepted by William Thomson on 28 January 1874. IET Archives UK0108 SC MSS 005/I/3/01/08 Primary collection of Heaviside papers: two letters from Heaviside to Highfield, President of the IEE, of


[A.]

Oliver Heaviside

of Great Northern Telegraph Station Newcastle on Tyne

being desirous of admission into THE SOCIETY OF TELEGRAPH ENGINEERS, I recommend him, from personal knowledge, as a person in every respect worthy of that distinction.

[Here specify distinctly the qualifications of the Candidate.]

Several Years Connected with Telegraphy
Also vide articles in Phil Mag.

On the above grounds, I beg leave to propose him to the Council as a proper person to be admitted into the Society.

S. E. Phillips Junr Member.

Dated this 13th day of Jany 1874

We, the undersigned, concur in the above recommendation, being convinced that Mr. Oliver Heaviside is in every respect a proper person to be admitted into the Society.

W. T. Henley
Saml E Phillips
G. Risch

The Council, having considered the above recommendation, present Mr. Oliver Heaviside to be balloted for as Associate of THE SOCIETY OF TELEGRAPH ENGINEERS.

William Thomson
Chairman.

Dated this 28 day of Jany 1874

Elected 11th February 1874

Figure 4. OH's STE membership application, submitted 13 January 1874. Image courtesy of IET Archives. (Online version in colour.)

One of the *Philosophical Magazine* articles referred to in OH's STE application was 'On the best arrangement of Wheatstone's bridge of measuring a given resistance with a given galvanometer and battery,' an early article by OH published in 1873 and praised by William Thomson and James Clerk Maxwell with the latter citing it in the second edition of *Treatise on electricity and magnetism* [23].[7] OH's theoretical work in electrical engineering, even at this early stage, was being recognized by longstanding senior figures in the field of electrical engineering. However, not all had praise for OH's work—as mentioned previously, senior GPO engineers William Preece and R.S. Cullen among them.

Despite this professional recognition by many but not all fellow telegraph and electrical engineers, in 1881, OH was struck off the members' list of the Institution of Electrical Engineers

5 March 1922 and 14 March 1922, with regard to the IEE honouring Heaviside and securing him a pension. Letters are numbered by Heaviside OH4 and OH5.

[7]The paper was cited in [24].

(formerly the STE) having consistently failed to pay his membership fees [25].[8] By contrast, AWH remained an active member of the society for the duration of his career, contributing occasional papers and less occasional post-paper discussions until at least 1908.[9] His papers and remarks on papers display a wider contribution to electrical engineering beyond telecommunications—to electrical tramways and electrical power, the latter being a particularly strong interest—unsurprising given his active role as one of the founders of the Newcastle Electricity Supply Company [30].

In May 1874, a few months after his successful STE membership application, OH's partial deafness and general health began to decline and this may have contributed to his difficulty in the working world. OH also chose a more precarious and isolated existence, to devote himself to scientific research and electrical theory. Instead, OH lived with his parents in London for the next 15 years from 1874 and 1889 relying on a small income from articles written for *The Electrician*, abstracts for the *Journal of the Society of Telegraph Engineers* (JSTE) and possibly some small financial assistance from his brother AWH, who by the 1880s was a 'well-known and respected electrician [electrical engineer] in his own right'.[10] OH had a difficult, single-minded personality and his growing seclusion made him unsuitable for teamwork including working in telegraphy, according to his brother AWH; OH also had other priorities, for which he turned down experimental research roles in the GPO Engineering Department, albeit roles for which he would have had to report to William Preece with whom he had an antagonist relationship [32]. Because of the physical separation between the two brothers—AWH living in Newcastle with OH in London and later Devon—and possibly due to OH's deafness, much of the relationship—working and personal—between the two brothers is documented through written correspondence, some of which has survived in the IET archives in London and which are referred to throughout this paper.[11] The two did continue to meet in person—meeting regularly in London for dinner [34] (figure 5).

Despite the physical distance between the two, until 1888 OH's sole scientific collaborator was his brother AWH, with his experimental notebooks showing the various experiments and projects the brothers collaborated on between 1880 and 1887.[12] AWH used his brother as a kind of technical and scientific adviser and sounding board—in one letter promising in his next letter to write about 'what we are doing electrically in the post office and some facts about insulation which will make you stare'.[13] AWH's experiments and technical developments at this time included experiments with ABC telegraph exchange for mercantile communication in Newcastle, in particular collieries and shipping offices in 1880 and in 1882 devised a metallic circuit system of telephones known as the 'North East (Northern)' system, moving away from rented telegraphy lines [1]. The latter made AWH a pioneer of underground wires for telephones and resulted in very few open telephone wires in Newcastle from the early days of telephony. AWH's interest in the potential of telephony in the first decade of development continued and, between 1886 and 1887, the two brothers collaborated on an innovative 'Bridge telephone' system for long-distance telephone lines in Newcastle—a strong example of their collaborative relationship as well as their difficult relationship with William Preece.

[8]For a meeting report, see [26].

[9]Arthur's papers published in the *Journal of the Institution of Electrical Engineers* from the title and institution change around 1881 were: [1,27–29]. Arthur also contributed remarks on numerous papers in various sub-fields of electrical engineering throughout his career.

[10]IET Archives UK0108 SC MSS 005 Heaviside Collection letter from AWH to OH, dated 15 July 1881 and letters from William Ayrton to OH, dated 16 February [no year given] and 15 March [no year given], cited in [31]. See also [5]. For details of work for *The Electrician*, see IET Archives UK0108 SC MSS 005/I Primary collection of Heaviside papers, 1872–1921.

[11]For details of OH's personality, see [33].

[12]IET Archives Heaviside Collection UK0108 SC MSS 005 Heaviside Collection, 1872–1923 including 21 notebooks consisting of mathematical equations and calculations.

[13]IET Archives Heaviside Collection UK0108 SC MSS 005 Heaviside Collection Box 9:6:2 letter from AWH to OH dated 12 October 1881.

rsta.royalsocietypublishing.org Phil. Trans. R. Soc. A 376: 20170454

Figure 5. BT Archives TCB 475/ZA/ZA9 OH, ca. 1900 [captioned as Professor Heaviside, n.d.]. Image courtesy of BT Archives.

3. 'Bridge system' of telephony and relationship with William Preece

AWH's experimentation with early telephony originated with the work of Alexander Graham Bell and William Preece's promotion of Bell's system of telephony (figure 6). In 1876, Alexander Graham Bell introduced and patented one form of telephony and in August 1877, William Thomson (later Lord Kelvin) and William Preece introduced Bell's telephone before the annual meeting of the British Association for the Advancement of Science (BAAS) held that year in Plymouth [35]. Bell was one system of telephony introduced around this time; Edison's 'speaking telegraph' referred to in William Preece's notes from 1877 below was another (figure 6). The two systems were in competition—in and out of the courtroom and through the patent system—until priority was awarded to Bell.

A year after he introduced Bell's telephone system, Preece together with William Thomson made the first practical demonstration on the British Isles of a pair of telephones—although reportedly they had difficulty getting them to work—before the annual meeting of the BAAS held in Dublin in August 1878 [36]. Later that same year, the Post Office provided its first telephones—a pair of Bell telephones—on rental to a firm in Manchester.[14] In late 1880, the case of Attorney General v Edison Telephone Company of London Ltd came to court and in a landmark ruling Mr Baron Pollock and Mr Justice Stephen decided in favour of the state, thereby awarding responsibility for management and licensing of the domestic British telephone network to the best-placed state institution, the GPO.[15] This provided an extension of the GPO's equivalent role

[14]BT Archives POST 30/330 Rental of Professor Graham Bell's instruments by the Post Office, 1877–1881.

[15]Preece's personal annotated copy of the judgement can be found at IET Archives UK0108 NAEST 039/3—Attorney General v Edison Telephone Company of London Ltd: Arguments and Judgement in Exchequer Division, High Court of Justice (1880).


Notes on Edison's Plumbago Telegraph
[illegible] from the Journal of the Telegraph
June 1. 1877.

He (Edison) takes advantage of the property which plumbago possesses of decreasing the resistance enormously under slight pressure. Then discs of that material are placed upon the upper pole of an electro-magnet, the coils of which have several hundreds of ohms resistance. Upon the discs of plumbago is laid the armature which is provided with a binding screw for clamping the local battery wire.

The cores of the magnet, the plumbago discs & the armature are included in a local circuit which also contains an ordinary sounder & cells. Then follows a description of its action which is obvious. No adjustment is required & it is probably the only device which will allow of the transmission of signals of variable strengths from one circuit to another. The [illegible] was designed for repeating [illegible] vibrations of variable strength in his (Edison's) speaking telegraph.

Figure 6. A set of handwritten notes by Sir William Preece, dated June 1 1877, describing the mechanical and physical properties of Thomas Edison's newly invented 'speaking telegraph'.[16] Image courtesy of IET Archives. (Online version in colour.)

for the inland telegraph network, as outlined in the Telegraph Acts of 1868 and 1869, and led to GPO engineers such as Preece, AWH and others experimenting and innovating with this new technology.

As demonstrated, a key figure in the introduction of the telephone to Britain, William Preece was a Victorian electrical engineer of the practical variety with a background in telegraphy. Preece joined the Electric Telegraph Company in 1853 and transferred to the GPO's engineering department in 1870 as Southern District engineer when the private telegraph companies were taken over by the state. In 1877, he was promoted to Electrician at the Post Office and in

[16]See https://www.theiet.org/resources/library/archives/featured/edison.cfm for further details.

1880 he was President of the STE. Between 1892 and 1899, he served as Engineer-in-Chief and Electrician of the GPO before retired aged sixty five. Throughout his career, Preece emphasized practical experience over theoretical and mathematical understandings of engineering problems and solutions leading to a deep-seated mistrust of the work of Maxwellians, OH included. Between the mid-1880s and the early 1890s, Preece represented the older and more traditional class of electricians (electrical engineers) in the 'Practice versus Theory' debate and despite being on the losing side, Preece retained his authority, reputation and influence within and without the Post Office [37].

Meanwhile, in May 1877 and inspired by Graham Bell's recent telephone system, two of AWH's staff, Frank Reid and W.R. Smith, made the first telephones in the northeast using parts from a ABC telegraph [1]. In September 1877 and a month after Preece's first British demonstration of his telephone system, Graham Bell visited Britain including Newcastle in September to introduce his telephone system [38]. The introduction in the late 1870s of networks of telephone lines in various British cities led to two main problems: how to create and scale up telephone exchanges and problems of interference between telephone lines and telegraph cables, so-called cross talk. In both cases, GPO engineers such as AWH (with the assistance of his brother OH) and Preece were best placed to investigate and find solutions.

In terms of the former, between 1886 and 1887, AWH and OH collaborated to develop an innovative 'Bridge' (or parallel circuit) telephone system to improve the clarity of communication on long-distance telephone lines in Newcastle with AWH doing the practical work and OH the theoretical work. The latter was a continuation of OH's work on a 'bridge system' of duplex telegraphy, published in his 1876 *Philosophical Magazine* article [21]. As mentioned previously, this was the second part of a two-part article with the first part published in 1873 and opening with a disparagement of senior GPO engineers R.S. Culley and William Preece work on duplex telegraphy [20]. As such, OH's theoretical work on the 'bridge system' of telephony was a continuation of OH's bitter dispute with William Preece, with Preece particularly dismissive of theoretical claims made by OH about self-inductance and 'loading' of telephone circuits [5]. A joint paper by the brothers, to be published in the Journal of the Society of Telegraph Engineers, was blocked by Preece; instead, the theoretical aspect of the 'bridge system' of telephony was eventually published in OH's 1893 publication *Electromagnetic theory*, volume 1, in which OH discussed self-inductance with leaks as it related to AWH's 'bridge system' as well as an additional proposal on how to achieve a distortionless circuit [39]. AWH did not publish the details of his contribution to the 'bridge system' of telephony until his 1900 paper before the Newcastle section of the IEE with which this paper opened [1].

Ultimately, however, AWH supported Preece against his own brother when it came to improving signal transmission through the use of cable loading; this may have been as much to do with ensuring his career prospects and status within the GPO and wider electrical engineering community as it was about siding with Preece in the 'Practice versus Theory' debate (as mentioned above). In supporting Preece over his own brother in terms of self-inductance and loading coils, OH held a grudge against Preece and AWH until at least 1922 and quite possibly beyond until his death in 1925: 'AWH [Arthur Heaviside] was very afraid of P [Preece], who talked him over and convinced him that my [OH's] theory was all theoretical bosh, nothing like practical conditions and AWH went over to P's. ... The greatest mistake AWH ever made was to throw me overboard, and go in for WHP [William Henry Preece]. Spoiled all the rest of it'.[17] Reasoning from inadequate experiments, Preece declared inductance to be prejudicial to clear signalling and so took steps to block the brothers' joint publication on the 'bridge system' and actively discouraged OH's publications more generally; thereafter, OH took every opportunity to denounce Preece.[18]

[17]See IET Archives UK0108 SC MSS 068 OH correspondence with J. S. Highfield, President of the IEE, on the award of the Faraday Medal, and between Highfield and the IEE and others concerning Heaviside, 1922–1925.

[18]See IET Archives UK0108 SC MSS 005/I/1/12 [OH] Notebook, 1886–1889, which consists of a Notebook headed 'Complete MS copy of S[elf] I[nduction] of W[ires] (Right side) 163 pp Phil Mag and other papers some imperfect' with ms transcription of papers published in the *Philosophical Magazine* entitled on the Self Induction of Wires Parts I–VII published August 1886 to

It was also around this time of intense and profound technical disagreement between the two brothers in the mid-1880s that AWH began conducting a series of experiments on conductive and inductive telegraphy. The original programme of research began with William Preece's and investigations into 'cross talk' or interference between telegraph lines and telephone wires. Owing to their ongoing disagreement over cable loading and the lack of surviving correspondence between AWH and OH during this period, there is no evidence the two brothers collaborated on this project. AWH pursued these innovative experiments into what were then considered new and promising forms of electrical telegraphy without the support and theoretical work of his younger brother who instead choose to research and publish papers on electromagnetism in the *Philosophical Magazine* and *The Electrician* and which were eventually published in his three-volume series *Electromagnetic theory*, to much acclaim and public recognition in particular by Oliver Lodge [40–42].

4. 'Cross talk' and inductive telecommunications

In 1880 and a couple of years before Preece began his investigations into the topic, AWH and OH began working on a solution to the problem of 'cross talk, using electrical conductors as a solution', eventually resulting in an 1880 British patent, no. 1407 'Improvements in Electrical Conductors, and in the Arrangement and Manner of Using Conductors for Telephonic and Telegraphic Purposes.' Their goal was to obtain 'perfect protection [from inductive interference], and to render a circuit completely independent under all circumstances of external inductive influences'; the patent was granted solely to OH as AWH had complications in obtaining patents in his own right while working for the GPO (figure 7).[19] There was even some mention of an attempt to commercially exploit the patent but this came to naught, with the possible exception of an ambiguous £100 payment referred to in correspondence between the two brothers in June 1881.[20] Despite this lack of commercial success, the 1880 patent is essentially the patent for what later came to be known as a 'coaxial cable', a significant and important invention in the history of telecommunications and one still in use today [48]. Contemporaneously, the importance of OH's 1880 patent and his work on distortionless cables was recognized by telegraph engineer Charles S. Bright in his 1898 authoritative treatise on telegraph cables, *Submarine Telegraphs: Their History, Construction and Working* [49].

In 1884, Preece was directed to the use of interference as a potential form of new telegraphy by a mere accident: some telegraph messages sent through the GPO's central exchange in Bradford were being read upon the nearby circuit of a telephone company and the signals were recognized by an old telegraphist who had formerly worked for the GPO.[21] Preece investigated this matter, a serious lapse of security, thoroughly and eventually traced the cause to induction. As a result of these investigations, the GPO Engineering Department began a series of investigations between 1885 and 1886 in order to more fully understand this new form of telegraphy with various internal

July 1887, with additional notes on left hand-side pages, and notes on the fate of Part VIII and Part XI, which were declined by the *Philosophical Magazine* and blocked by Preece, and partly published in *The Electrician*.

[19]See [43,44]. I was made aware of this patent via private electronic correspondence with my then PhD supervisor, Professor Graeme Gooday at the University of Leeds, 16–18 August 2010. In relation to GPO staff and patenting, see BT Archives POST 30/2668B – Regulations concerning patents and payment for inventions by Government servants (1895–1914) and IET Archives Heaviside Collection, letter from AWH to OH, dated 29 June 1881.

[20]Appleyard makes a passing comment about the Heaviside brothers attempting to sell an invention for 'neutralizing disturbances in cables', see [45]. For discussions of the £100 payment and the letter between AWH and OH dated 29 June 1881, see; [46,47].

[21]This is described in lectures given by Preece in the lead-up to his retirement and in many articles published upon Preece's retirement in 1899 but was not mentioned in contemporaneous publications (*ca* 1884). See BT Archives TCK 89/23—[Album of press cuttings relating to Preece, mainly regarding his appointment as Chief Electrician and Engineer-in-Chief, and his retirement from the Post Office], 1890–1903; IET Archives UK0108 SC MSS 022/III/184—MS notebook on Aetheric Telegraphy [probably authored by Preece], dated 5 December 1898. The notebook contains notes on the history of telegraphy and the progress of its development prepared for talks given at the Midland Institute, Birmingham on 5 December 1898; Blackheath, Surrey on 13 March 1899 and Wimbledon Literary Society on 22 April 1899.

A.D. 1880, *6th* APRIL. N° 1407.

Electrical Conductors, &c.

LETTERS PATENT to Oliver Heaviside, of No. 3, Saint Augustine's Road, in the County of Middlesex, for an Invention of "IMPROVEMENTS IN ELECTRICAL CONDUCTORS, AND IN THE ARRANGEMENT AND MANNER OF USING CONDUCTORS FOR TELEPHONIC AND TELEGRAPHIC PURPOSES."

PROVISIONAL SPECIFICATION left by the said Oliver Heaviside at the Office of the Commissioners of Patents on the 6th April 1880.

OLIVER HEAVISIDE, of No. 3, Saint Augustine's Road, in the County of Middlesex. "IMPROVEMENTS IN ELECTRICAL CONDUCTORS, AND IN THE ARRANGEMENT AND MANNER OF USING CONDUCTORS FOR TELEPHONIC AND TELEGRAPHIC PURPOSES."

When a number of wires run parallel to one another, either suspended or otherwise, any change in the current flowing in one wire causes currents in all the rest by induction, and the effect may be so great as to seriously interfere with the working of telephonic circuits, and to a less degree of ordinary telegraphic circuits also.

It is a common practice to complete the circuit by means of a second wire instead of the earth, and it is well known that the inductive interference is thereby reduced in magnitude, the induced electro-motive forces in one wire cancelling those in the other to a certain extent. The nearer the wires are brought together the less does the inductive interference become, but it cannot be altogether eliminated in this way, because the axes of the two wires cannot be made to coincide so that they shall be both at the same distance from any disturbing wire.

My improvements have for object to obtain perfect protection, and to render a circuit completely independent under all circumstances of external inductive influences. For this purpose I use two insulated conductors for the circuit, and place one of them inside the other; thus, one conductor may be a wire, and the other a tube or sheath, thus forming a compound conductor consisting of a central wire surrounded by an insulating covering, which is in its turn surrounded by a conducting tube or sheath, which must also be insulated. When the tube and inner wire are electrically connected at both ends of the line, as through apparatus in the usual manner, the circuit as thus described is completely independent of other circuits, and any number of such circuits, each containing an insulated tube

[*Price 6d.*]

Figure 7. The cover page of OH's 1880 British Patent No. 1407 'Improvements in Electrical Conductors, and in the Arrangement and Manner of Using Conductors for Telephonic and Telegraphic Purposes.' Image available in the public domain, source unknown.

reports written by the many GPO engineers working on this research project including John Gavey, J.E. Taylor and AWH.[22]

[22]This is described in lectures given by Preece in the lead-up to his retirement and in many articles published upon Preece's retirement in 1899 but was not mentioned in contemporaneous publications (*ca* 1884). See BT Archives TCK 89/23—[Album of press cuttings relating to Preece, mainly regarding his appointment as Chief Electrician and Engineer-in-Chief, and his retirement from the Post Office], 1890–1903; IET Archives UK0108 SC MSS 022/III/184—MS notebook on Aetheric Telegraphy [probably authored by Preece], dated 5 December 1898. The notebook contains notes on the history of telegraphy and the

Between 1885 and 1886, AWH conducted experiments with inductive, conductive and reflective telegraphy systems at Town Moor, Newcastle.[23] The intriguingly titled 'reflective telegraphy' was mentioned only once in passing, was not described at all and did not feature in any other experimentation; given the title of AWH's report as 'reporting on tests done on the telegraph cable in Newcastle using a "mirror"', 'reflective telegraphy' may have involved some form of mirror, possibly some form of light telegraphy similar to Bell's Photophone.[24] These 'exhaustive' experiments were conducted alongside long-distance telegraphy and telephony experiments with the intention of determining whether the 'disturbances' were due to electromagnetic induction and were independent of Earth conduction and also how far the 'disturbances' could be detected.[25]

The experiments were based on AWH and OH's longstanding experience with interference: AWH recorded in the *JSTE* in January 1881, one had only to walk about the Newcastle Battery Room with a telephone and a coil of wire to read the Morse signalling from every battery in use [27]. The apparatus for the 1885 and 1886 experiments used telegraphic and telephonic apparatus arranged in an experimental layout to continue Preece's experiments and then extend them including the use of telephone receivers which were used as a more sensitive form of receiver rather than to receive voice signals. By the 1880s, this had become the accepted use for telephone receivers on the telegraph network, particularly in relation to the British military, and something with which GPO engineers such as AWH would have been familiar.[26] The initial experiment consisted of insulated squares of wire, each side being 440 yards long, being laid out horizontally on the ground one quarter of a mile apart, and distinct speech by telephones was transmitted between them; when removed 1000 yards apart inductive effects were still appreciable [51]. Eventually, signals were received over a distance of a quarter of a mile with 'long-distance' effects being detected just over 10 miles away.[27]

Further investigations into the inductive telegraphy system had four key objectives to discover:

(1) At what distance from a primary parallel could these effects be detected on the telephone?
(2) How much of the effect was conductive?
(3) How much of the effect was inductive?
(4) Were all bodies transparent to them?

By April 1886, AWH was able to report that inductive or conductive telegraphy could be used for 'telegraphing without wires between distant places', depending on the type of return —Earth-returns or insulated metallic returns—employed and a further round of experiments began in mid-April 1886.[28] Working on the section of Town Moor where horse racing had previously taken

progress of its development prepared for talks given at the Midland Institute, Birmingham on 5 December 1898; Blackheath, Surrey on 13 March 1899 and Wimbledon Literary Society on 22 April 1899.

[23]IET Archives UK0108 SC MSS 005/II/4/32 Letter from AWH, Superintending Engineer, reporting on tests done on the telegraph cable in Newcastle using a 'mirror'. 18–20 August 1885.

[24]See footnote 5. In 1880, Alexander Graham Bell developed and patented an original device, the 'photophone' (also referred to as a 'radiophone'), which used beams of light to transmit sound over relatively short distances. See Bell, Alexander Graham. 'US Patent No 235199 Apparatus for Signalling and Communicating Called 'Photophone',' edited by United States Patent Office, USA, 1880. https://patents.google.com/patent/US235199

[25]IET Archives UK0108 SC MSS 005/II/4/32 Letter from AWH, Superintending Engineer, reporting on tests done on the telegraph cable in Newcastle using a 'mirror'. 18–20 August 1885 and IET Archives UK0108 SC MSS 005/I/3/11/04 3 pp of rules for testing by received current with new tangent galvanometer, Newcastle upon Tyne: Corrections to Circular E19, produced for the Post Office, possibly by AWH.

[26]See [50]. I was made aware of this practice via private electronic correspondence with Michael Kay, a then fellow PhD student in history of science at the University of Leeds, on 15 October 2012.

[27]IET Archives UK0108 SC MSS 022/III/184 MS notebook on Aetheric Telegraphy [probably authored by Preece], dated 5 December 1898.

[28]BT Archives TCB 274/1 Experiments upon wireless telegraphy by methods of induction and conduction, 1886–1900.

place, initial experiments were conducted over a distance of 660 yards from a Blue House to the Grand Stand.[29]

A few months later in August 1886, the experiments continued at the estate of Hugh Andrews, Swarland Park, where AWH showed that telegraphy signal was inductive rather than conductive in nature—that is the signal was induced in the receiving circuit or line rather than being conducted by the Earth.[30] This eventually led to 'Morse-speaking' across 40 miles of country between Shap and Gretna on the west coast and Newcastle and Jedburgh on the east coast, produced by interrupting the vibrations with a Morse key, although this was hard to verify due to the large network of telegraph cables between the two places.[31] Despite much of the work having being done by AWH without recorded input from Preece, the experimental outcomes were reported by Preece at the 1886 meeting of the British Association for the Advancement of Science, held in Birmingham that year, with the figure of 40 miles being picked up and highlighted by many of the newspaper articles, omitting the original qualifiers and uncertainty in the original paper [52].[32]

These experiments continued with successful experiments on the parallel telegraph lines between Durham and Darlington, 10.25 miles apart using telephone handsets, and in 1887, Preece and AWH conducted experiments on earth conduction and inductive telephony down mines at Broomhill Colliery where communication was successfully received 360 feet below the surface [53]. Here, AWH successfully set up an inductive telephony (speech) circuit in a triangular form along the galleries about two and a quarter miles in total length, and at the surface a similar circuit of equal size over and parallel to the underground line [54]. By now, the GPO had begun experiments with these new forms of telegraphy over bodies of water, so they might be used for communication with lighthouses and lightships and these new experimental trials continued, albeit without the input of AWH after the 1886 experiments, until the early 1890s when mobile electromagnetic wave wireless telegraphy and the work of Guglielmo Marconi came to the fore. Marconi's system of wireless telegraphy was actively and publicly supported by William Preece, who provided access to GPO engineering department expertise and resources.[33]

The practical telecommunication experiments conducted by GPO engineer AWH in the northeast of England into 'disturbances' (interference) on telephone lines were devised to discover the theoretical and practical purposes of the system: first, to understand how this interference interacted with the developing telephone network and existing telegraphy network; and secondly, to discover whether they might be practically useful for 'wireless' communications. The experiments were an innovative use of the existing telephone and telegraph apparatus and systems and were underpinned by a theoretical and mathematical understanding of these experiments which is made explicit in AWH's detailed reports of the time.[34] Although there is no concrete evidence to suggest OH continued his role as informal consultant in matters of the mathematics and theory of telecommunications in relation to these experiments, they did, however, build upon the brothers' earlier collaborations in duplex telegraphy and the 'bridge system' of telephony.

[29]BT Archives TCB 274/1: Report of experiments upon telegraphing without wires between distant places by induction and conduction—commencing 14 April 1886, Newcastle-on-Tyne, signed AWH and dated 15 June 1886, 1.

[30]BT Archive TCK 89/10 Reports, with diagrams, on telephonic experiments in Newcastle, Sunderland and South Wales, 1880–1889.

[31]BT Archives TCK 89/22 Album of press cuttings relating to Preece, Marconi and others; scientific experiments to do with telegraphy, collected by 'Romeike and Curtice', Press Cutting and Information Agency, on Preece's behalf, 1892–1899.

[32]BT Archives TCK 89/22 Album of press cuttings relating to Preece, Marconi and others; scientific experiments to do with telegraphy, collected by 'Romeike and Curtice', Press Cutting and Information Agency, on Preece's behalf, 1892–1899.

[33]See BT Archives Post 30/1066C General technical report on wireless telegraphy, 1903.

[34]BT Archives TCB 274/1 Experiments upon wireless telegraphy by methods of induction and conduction, 1886–1900 and BT Archive TCK 89/10 Reports, with diagrams, on telephonic experiments in Newcastle, Sunderland and South Wales, 1880–1889.

Figure 8. BT Archives TCE 361/ARC 146 [GPO] Group Portrait: Superintending Engineers' Conference, 1898 or 1899 with AWH marked in red. Image courtesy of BT Archives. (Online version in colour.)

5. Conclusion

Although AWH did not realize this when he conducted the experiments into inductive telegraphy in the mid-1880s, by the time of his 1900 lecture, AWH had come to the realization that these new forms of telegraphy were electromagnetic: '... detection of inductive effects at long distances is marked. This is the so-called wireless telegraphy of the electromagnetic order. It was always there; most people knew of it from the days of Faraday and Henry in 1831; but what they did not know was its extent ....' [1]. AWH's work and experimental practice was built upon earlier collaborations in duplex telegraphy and the 'bridge system' of telephony and possibly guided by OH's researches into electrical theory and electromagnetism. These researches were being conducted in parallel with his brother's experiments and which were published, as mentioned above, in the *Philosophical Magazine* and *The Electrician* and which were eventually published in his three-volume series 'Electromagnetic Theory', to much acclaim and public recognition in particular by Oliver Lodge [40–42].

In 1902, OH wrote an article on telegraphy for the 10th edition of the *Encyclopaedia Britannica* in which he discussed the transmission of radio waves and suggested the possibility of a conducting layer above the Earth which, taken in conjunction with the surface of the Earth, would conduct electricity and reflect radio waves in the form of a kind of waveguide [26]. Independently and based on observations that Guglielmo Marconi's successful transatlantic radio transmission between Cornwall, England, and Newfoundland in December 1901 exceeded the distance and quality predicted by existing radio-wave theory, American electrical engineer Arthur Kennelly also proposed the existence of such a region in the ionosphere: Kennelly published in March 1902 and Heaviside in December 1902 [55].[35]

[35]Engineering and Technology History Wiki: Kennelly-Heaviside Layer. http://ethw.org/Kennelly-Heaviside_Layer.

In March 1903, 3 years after AWH's speech before the Newcastle local section of the Institution of Electrical Engineers with which this paper opened, John Ambrose Fleming—Professor of Electrical Engineering at University College, London, and consulting engineer to the Marconi Company—gave the prestigious Cantor lectures on the topic of 'Hertzian Wave Wireless Telegraphy'. Although relating to different types of wireless telegraphy compared to that experimented with by Preece, AWH and other GPO engineers, Fleming acknowledged that, in relation to the electromagnetic framework, the only reasonable proofs involving advanced mathematical analysis could be found J.J. Thomson's *Recent researches in electricity and magnetism* published in 1892 and OH's *Electromagnetic theory*, published in three volumes in 1893, 1899 and 1912.[36]

However, it was only after the Kennelly–Heaviside layer (now called the E region) was experimentally proved by British physicist Edward Appleton with a series of experiments begun in late 1924 that the GPO began to formally engage with the work of OH.[37] By the 1920s and 1930s, articles on OH's electromagnetic theory were actively used by GPO employees and by 1932 the GPO produced a report 'The Application of the Heaviside Analysis to Electrical Engineering Problems', no copies of which survive.[38] While the author of the 1932 GPO report is undocumented, by 1946 GPO Research Engineer H.J. Josephs had authored 'Heaviside's electric circuit theory', which focused on OH's electrical theory as it applied to telegraphic and telephonic problems; a second edition, which was also published internationally, was published in 1950 [56–58]. H.J. Josephs was also the author of an early biography of OH, *Oliver Heaviside: a biography*, probably self-published, around 1963 and numerous other articles on Heaviside, many based on papers of OH discovered under the floorboards of OH's former home in Paignton around 1957; very few copies of Josephs' biography have survived—the COPAC catalogue of over 100 UK and Irish academic, national and specialist libraries lists a single copy at the Science Museum and there is also a copy in the IET archives [59–63].

In many ways, the relationship between AWH and OH and hence between the GPO's engineering practice and OH's theoretical work reflected the more general and sometime antagonistic 'Practice versus Theory' debate as well as the wider reception of Maxwellian ideas and theorists in British electrical engineering in the late nineteenth century. Initial collaborative experiments with duplex telegraphy and the 'bridge system' of telephony showed how AWH's practical work and implementations benefited from OH's theoretical work but was also constrained by the more traditional engineering values of his GPO engineering department workplace and, perhaps more importantly, the longstanding authority, reputation and influence of his superior and GPO Engineer-in-Chief William Preece. Although OH's antagonistic relationship with Preece pre-dated AWH and OH's collaborative work, the inter-relationship between the Heaviside brothers and Preece in combination with the 'Practice versus Theory' debate, Preece's deep-seeded mistrust of Maxwellian ideas including OH's theoretical work, and the gradual acceptance of Maxwellian ideas led to a difficult and complex relationship.

By the 1890s and the latter stages of both Preece and AWH's longstanding careers with the GPO, the wider reception of Maxwellian ideas and theorists in British electrical engineering had shifted towards the Maxwellians and OH began to receive recognition from the wider electrical engineering community, in particular for his volumes on Electromagnetic Theory. By the time of AWH's 1900 lecture with which this paper opened, AWH had come to the realization that the inductive and conductive telegraphy systems he and the GPO had experimented with the

[36] IET Archives UK0108 NAEST 147 Hertzian Wave Wireless Telegraphy—John Ambrose Fleming.

[37] Appleton was awarded a Nobel Prize in 1947 for his research on the ionosphere; see 'Edward V. Appleton—Biographical'. Nobelprize.org. Nobel Media AB 2014. http://www.nobelprize.org/nobel_prizes/physics/laureates/1947/appleton-bio.html.

[38] See, for example, TCK 225/4 Papers of GPO engineer J.E. Taylor: The physical basis of Heaviside's radiational theory, n.d. and TCK 225/5 Papers of GPO engineer J.E. Taylor: Rough handwritten drafts of articles and jottings relating to J.E. Taylor's research. Topics covered are the theories of the electric current, theories of Maxwell, Heaviside, Einstein, Poynting, Faraday, electromagnetic theories, radiational theories, the propagation of electric waves, wave guidance and metallic conduction, 1927–1948. Also, BT Archives TCB 422/5896. The Application of the Heaviside Analysis to Electrical Engineering Problems, n.d. [1932]. Report was not issued.

1880s were, in fact, electromagnetic rather than new systems of electrical telegraphy as he had understood them at the time [1]. These experiments and the wider research programme and experimental practice of the GPO in the 1880s and 1890s demonstrated a state institution which could be both innovative in its engineering practice and deeply resistant to change. The latter was also further emphasized by the GPO's lack of engagement with OH's theoretical work, in particular the 'Heaviside Layer', until it was experimentally proved in the mid-1920s.

Data accessibility. This article has no additional data.
Competing interests. I declare I have no competing interests.
Funding. I received no funding for this study.
Acknowledgements. Many thanks go to Anne Locker, Asha Gage and Jonathan Cable at IET Archives and David Hay at BT Archives for their knowledgeable assistance and the use of images from their archives; David Rooney and Doug Millard, my line managers at the Science Museum, for their encouragement and for providing research time for this article; and the two anonymous reviewers for their many helpful comments and suggestions—all errors are mine. Last and definitely not least, thanks go to my wife Camen for her support, proofreading and many cups of tea.

References

1. Heaviside AW. 1900 Newcastle local section. Inaugural address by the Chairman. *J. Inst. Electr. Eng.* **29**, 900–924. (doi:10.1049/jiee-1.1900.0094)
2. Heaviside AW. 1901 Electrical applications in the North East of England. *St Martin's-le-Grand* **11**, 52–68.
3. Hunt B. 2011 *Heaviside, OH (1850–1925), physicist and electrical engineer*. Oxford, UK: Oxford Dictionary of National Biography.
4. Bruton EM. 2012 *Beyond Marconi: the roles of the admiralty, the post office, and the Institution of Electrical Engineers in the invention and development of wireless communication up to 1908*. Leeds, UK: University of Leeds.
5. Hunt BJ. 2012 Oliver Heaviside: a first-rate oddity. *Phys. Today* **65**, 48. (doi:10.1063/PT.3.1788)
6. Bowers B. 2001 *Sir Charles Wheatstone FRS: 1802–1875*, pp. 187–188. London: IET.
7. Thompson SP, Bowers B. 2011 *Wheatstone, Sir Charles (1802–1875), developer of telegraphy*. Oxford Dictionary of National Biography. http://www.oxforddnb.com/view/10.1093/ref:odnb/9780198614128.001.0001/odnb-9780198614128-e-29184.
8. Searle GFC. 1987 *Oliver Heaviside, the man*. St. Albans: CAM Publishing.
9. Nahin PJ. 1987 *Oliver Heaviside: Sage in Solitude*. New York: NY: IEEE Press.
10. Nahin PJ. 2001 Oliver Heaviside: the life, work, and times of an electrical genius of the Victorian age. Baltimore, MD: Johns Hopkins University.
11. Yavetz I. 1995 *From obscurity to enigma: the work of Oliver Heaviside, 1872–1889*. Basel: Birkhäuser Verlag.
12. Heather A. 2008 *Oliver Heaviside: it's my genius that keeps me warm*. Paignton, UK: Creative Media.
13. Mahon B. 2009 *Oliver Heaviside: maverick mastermind of electricity*. London, UK: IET.
14. Mahon B. 2017 *The forgotten genius of Oliver Heaviside*. Amherst, NY: Prometheus.
15. Lynch AC. 1991 The sources for a biography of Oliver Heaviside. *Hist. Technol.* **13**, 145–149.
16. Spargo CM, Yakovlev A. 2015 Oliver Heaviside FRS: Newcastle upon Tyne 1868–1874. In *IET History of Technology Network: Conference on the history of power and engineering*, Newcastle upon Tyne, UK, 6 June 2015. http://dro.dur.ac.uk/18558/1/18558.pdf.
17. Coggeshall IS, de Cogan D (eds). 1993 *An annotated history of submarine cables and overseas radiotelegraphs 1851–1934*, p. 76. Norwich, UK: School of Information Systems, UEA.
18. Appleyard R. 1930 *Pioneers of electrical communication*, p. 221. London: Macmillan and Company Limited.
19. Culley RS. Improvements in telegraphs. A new duplex system, granted on 19 November 1855. British Patent No. 2608
20. Heaviside O. 1873 On duplex telegraphy [part I]. *Philos. Mag. Ser. 4*, **45**, 426–432. (doi:10.1080/14786447308640882)
21. Heaviside O. 1876 On duplex telegraphy [part II]. *Philos. Mag. Ser. 5*, **1**, 32–43. (doi:10.1080/14786447608638999)

22. Baker EC. 1976 *Sir William Preece, F.R.S.: Victorian Engineer Extraordinary*, pp. 109–110. London: Hutchinson.
23. Heaviside O. 1873 On the best arrangement of Wheatstone's bridge of measuring a given resistance with a given galvanometer and battery. *Philos. Mag.* **45**, 3–8.
24. Maxwell JC. 1881 *A treatise on electricity and magnetism*, vol. 1, p. 442, 2nd edn. Oxford: Oxford University Press.
25. Lee G. 1950 '*Oliver Heaviside—the Man, The Heaviside Centenary Volume*, vol. 12. London: IEE.
26. Lee G. 1950 The Heaviside Centenary Meeting. *J. Inst. Electr. Eng.* **1950**, 185–186. (doi:10.1049/jiee-2.1950.0082)
27. Heaviside AW. 1881 Some experiments on induction with the telephone. *J. Soc. Telegr. Eng. Electr.* **10**, 21–28. (doi:10.1049/jste-2.1881.0002)
28. Heaviside AW, Jackson RC. 1892 Electrical distribution by the Newcastle-on-Tyne Electric Supply Co., Limited. *J. Inst. Electr. Eng.* **21**, 475–488. (doi:10.1049/jiee-1.1892.0024)
29. Heaviside AW. 1892 Reply to discussion on 'Electrical distribution by the Newcastle-on-Tyne Electric Supply Co., Limited. *J. Inst. Electr. Eng.* **21**, 501–506. (doi:10.1049/jiee-1.1892.0026)
30. AJS. 1923 Arthur West Heaviside ISO. *J. Inst. Electr. Eng.* **61**, 1154–1155.
31. Nahin PJ. 2002 *Oliver Heaviside, sage in solitude: the life, work, and times of an electrical genius of the Victorian age*, p. 107. New York: IEEE.
32. Baker EC. 1976 *Sir William Henry Preece, FRS, Victorian Engineer Extraordinary*, pp. 208–209. London: Hutchinson.
33. Searle GFC. 1987 *Oliver Heaviside, the Man*. St Albans: CAM Publishing.
34. Yavetz I. 1995 *From obscurity to enigma: the work of Oliver Heaviside, 1872–1889*, p. 15. Basel: Birkhäuser Verlag.
35. 1878 *Report of the Forty-Seventh Meeting of the British Association for the Advancement of Science: Held at Plymouth in August 1877*. London: John Murray. See https://www.biodiversitylibrary.org/item/94483.
36. 1878 *Report of the Forty-Eighth Meeting of the British Association for the Advancement of Science: Held at Dublin in August 1878*. London: John Murray. See https://www.biodiversitylibrary.org/item/94499.
37. Arapostathis S. 2013 Electrical innovations, authority and consulting expertise in late Victorian Britain. *Notes Rec. R. Soc.* **67**, 59–76. (doi:10.1098/rsnr.2012.0048)
38. Kay M. 2015 Troublesome telephony: how users and non-users shaped the development of early British exchange telephony. *Sci. Museum Group J.* **3**, (doi:10.15180/150308)
39. Heaviside OH. 1903 *Electromagnetic theory*, volume I, pp. 433–437. London: The Electrician.
40. Heaviside O. 1893 *Electromagnetic theory*, volume 1, London, UK: The Electrician Printing and Publishing Co.
41. Heaviside O. 1899 *Electromagnetic theory*, volume 2, London, UK: The Electrician Printing and Publishing Co.
42. Heaviside O. 1912 *Electromagnetic theory*, volume 3, . London, UK: The Electrician Printing and Publishing Co.
43. Kraeuter DW. 2001 *Radio patent lists and index, 1830–1980*, p 179. New York: Edwin Mellen Press.
44. Nahin PJ. 2002 *Oliver Heaviside: the life, work, and times of an electrical genius of the Victorian age*, pp. xvi–xvii. Baltimore, MD: Johns Hopkins University.
45. Appleyard R. 1930 *Pioneers of electrical communication*, p. 215. London: Macmillan and Company Limited.
46. Yavetz I. 1995 *From obscurity to enigma: the work of Oliver Heaviside, 1872–1889*, p. 13. Basel: Birkhäuser Verlag.
47. Nahin PJ. 2002 *Oliver Heaviside: the life, work, and times of an electrical genius of the Victorian age*, p. xvii. Baltimore, MD: Johns Hopkins University.
48. Nahin PJ. 2002 *Oliver Heaviside: the life, work, and times of an electrical genius of the Victorian age*, p. xvii. Baltimore, MD: Johns Hopkins University.
49. Charles BS. 1898 *Submarine telegraphs: their history, construction and working*, vol. 186, p. 575. London: Crosby Lockwood and Son.
50. Preece W, Julius M. 1881 *The telephone*, pp. 451–455. London: Whittaker & Co.
51. Fahie JJ. 1899 *A history of wireless telegraphy, 1838–1899: including some bare-wire proposals for subaqueous telegraphs*, p. 145. Edinburgh: Edward Blackwood and Sons.

52. 1887 *Report of the Fifty-Sixth Meeting of the British association for the Advancement of Science: held at Birmingham in September 1886*. London: John Murray. See https://www.biodiversitylibrary.org/item/92792.
53. 1893 *International electrical congress at Chicago*, p. 522. London, UK: The Electrician.
54. Fahie JJ. 1899 *A history of wireless telegraphy, 1838–1899: including some bare-wire proposals for subaqueous telegraphs*, p. 146. Edinburgh: Edward Blackwood and Sons.
55. Kennelly AE. 1902 *On the Elevation of the Electrically-Conducting Strata of the Earth's Atmosphere. Electrical World and Engineer*, vol. XXXIX, p. 473. http://ethw.org/w/images/2/21/Kennelly_-_on_the_elevation_of_the_electrically-conducting_strata_of_the_earth%27s_atmosphere.pdf.
56. Josephs HJ. 1946 *Heaviside's electric circuit theory*. London: Methuen.
57. Josephs HJ. 1950 *Heaviside's electric circuit theory*, 2nd edn. London, UK: Methuen.
58. Nahin PJ. 2002 *Oliver Heaviside: the life, work, and times of an electrical genius of the Victorian age*, p. 309. Baltimore, MD: Johns Hopkins University.
59. Josephs HJ. 1963 *Oliver Heaviside: a biography*. London: HJ Josephs.
60. Josephs HJ. 1954 Unpublished work of Heaviside. *Electrical Review* **155**, 9–12.
61. Josephs HJ. 1959 History under the floorboards. An account of the recent discovery of some Heaviside papers. *J. Inst. Electr. Eng.* **5**, 26–30. (doi:10.1049/jiee-3.1959.0008)
62. Josephs HJ. 1959 The Heaviside papers found at Paignton in 1957. *Proc. Inst. Electr. Eng.* **106**, 70–76.
63. Josephs HJ. 1963 Postscript to the work of Heaviside. *J. Inst. Electr. Eng.* **9**, 511–512. (doi:10.1049/jiee-3.1963.0385)

PHILOSOPHICAL TRANSACTIONS A

rsta.royalsocietypublishing.org

Review

Cite this article: Nahin PJ. 2018 Oliver Heaviside: an accidental time traveller. *Phil. Trans. R. Soc. A* **376**: 20170448. http://dx.doi.org/10.1098/rsta.2017.0448

Accepted: 7 February 2018

One contribution of 13 to a theme issue 'Celebrating 125 years of Oliver Heaviside's 'Electromagnetic Theory''.

Subject Areas:
electrical engineering

Keywords:
faster-than-light, Cherenkov radiation, electromagnetic mass, paradoxes, tachyons, time travel

Author for correspondence:
Paul J. Nahin
e-mail: paul.nahin@unh.edu

Oliver Heaviside: an accidental time traveller

Paul J. Nahin

Department of Electrical Engineering, University of New Hampshire, Durham, NH 03824, USA

PJN, 0000-0002-0486-141X

A little discussed aspect of Heaviside's work in electromagnetics concerned faster-than-light (FTL) charged particles, precursors to the hypothetical tachyon and his discovery that such motion should produce a characteristic radiation signature (now called *Cherenkov radiation*). When Heaviside wrote, the time travel implications of FTL were not known (Einstein was still a teenager), and in this paper some speculations are offered on what Heaviside would have thought of FTL time travel, and of the associated (now classic) time travel paradoxes, including the possibility (or not) of sending information into the past.

This article is part of the theme issue 'Celebrating 125 years of Oliver Heaviside's 'Electromagnetic Theory''.

1. Introduction

> 'We cannot fight the laws of Nature.'
>
> 'Nature be damned! Feed more fuel into the tubes. We must break through the speed of light . . . Give me a clear road and plenty of fuel and I'll build you up a speed of half a million miles a second, . . . What's there to stop it?'
>
> —words exchanged by the first officer and the captain of a starship on its way to Alpha Centauri, in a 1936 science fiction story in which the captain (we are told) 'Had heard, of course, of the limiting velocity of light, but it meant nothing to him' [1].

> 'Don't be afraid of infinity!'
>
> —Oliver Heaviside, in an 1898 response to critics who had expressed doubts about the possibility of faster-than-light charged particles [2].

THE ROYAL SOCIETY PUBLISHING

rsta.royalsocietypublishing.org Phil. Trans. R. Soc. A 376: 20170448

The name Oliver Heaviside conjures up different images today in the minds of different technical people, as a function of their particular scientific interest. For atmospheric scientists, the image is that of the ionosphere (the famous 'Heaviside layer' in the 1982 musical *Cats*). For electrical engineers, it is his theoretical discovery of the condition for distortionless transmission of electrical signals on communication cables. For mathematicians (and electrical engineers, again), it is his use of the operational calculus to get answers to problems that, while often correct (but sometimes not), were not infrequently obtained via brazen manipulations that made both mathematicians and engineers literally cringe in horror (but often for different reasons). For physicists, his discovery-in-parallel with John Henry Poynting (1852–1914), Professor of Physics at the University of Birmingham, of the mathematics of the flow of electromagnetic energy in space (the Poynting vector), is what springs to the forefront. For historians of science, it is his decades-long battle-of-words with William Henry Preece (1834–1913), the pompous head of engineering at the British General Post Office (the controlling agency of electrical communications in England), battles that seem like something a Hollywood script-writer with an overheated imagination might cook up.

In other words, the life of Oliver Heaviside was a pretty interesting one [3].

In the rest of this paper, I will discuss another 'adventure' of his that is not as well known as the ones I just mentioned, one that has found new life in the modern physics literature dealing with the theoretical possibility (or not) of time machines and time travel to the past. What Heaviside would have thought of such things will be pure speculation on my part, as he never wrote (as far as I know) on them, and was dead before the early science fiction magazines embraced the time travel concept with great enthusiasm.

Heaviside must, of course, have surely been aware of H. G. Wells' famous 1895 novella *The Time Machine* but, again, I do not believe he ever put his thoughts about that work down on paper (with *one* possible exception, which I will mention later). So, I will feel free to speculate, but I will also do my very best to keep one foot (or at least a toe) in the real world.

Now, before we get down to business, let me end this opening section with the comment that Heaviside has reminded some of his admirers of the eccentric, fictional scientist Haskel van Manderpootz, professor of the 'new physics' (that is, relativity), who appeared in an hilarious science fiction tale published 10 years after Heaviside's death [4]. Compared to the 'modest' Van Manderpootz (read Heaviside), all other physicists in the world are a mere 'pack of jackals, eating the crumbs of ideas that drop from [his] feast of thoughts'. This view might seem to cast Heaviside in a less-than-flattering light, but in fact I do not believe *he* held that image of himself; rather, it is a hagiographic view of the man that has gained some traction among his more enthusiastic contemporary fans. I think a proper view of Heaviside is somewhat less grand (although perhaps not by much).

2. Electric current and moving charges

We are, today, so used to thinking of electricity as coming in tiny little lumps called *electrons* that it is difficult to realize that, not much more than 125 years ago, there was no such concept. For us, electrons *are* electricity, and electrons in copper wires are the same electrons we see jumping across space as sparks. The *nature* of electric current was a Victorian mystery, however, and nobody then thought of it as charged (for that matter and by-the-way, what is *charge*?) subatomic particles in motion. In the early 1880s, when Heaviside began his electromagnetic studies, nobody had that picture of electric current in their head.

Now, like all blanket statements, if you look hard enough, exceptions do turn up. For example, Michael Faraday (1791–1867) wrote, 11 years before Heaviside was born, 'if a ball be electrified positively in the middle of a room and be moved in any direction, effects will be produced, *as if a current in the same direction had existed* [my emphasis]: or if the ball be negatively electrified, and then moved, effects as if a current in a direction contrary to that of the motion had been formed, will be produced' [5]. But this was not common knowledge among the electricians of the day.

Well surely, you object, a scientist as great as James Clerk Maxwell (1831–1879), who was Heaviside's hero (as Heaviside wrote in his *Electrical Papers*, 'It will be understood that I preach the gospel according to my interpretation of Maxwell'), must have had *some* idea of what electricity is, and you would be right. He often thought it 'to be like' an incompressible fluid, a view that lasted quite a while in Victorian times (although often mocked as the 'water in a drain pipe' view). When Maxwell's friend, William Thomson (1824–1907), later Lord Kelvin, wrote his pioneering analysis of what eventually became the trans-Atlantic electric telegraph cable, he made direct use of the heat flow analyses of the French mathematical physicist Joseph Fourier (1768–1830)—see Fourier's 1822 *The Analytical Theory of Heat*—which was equivalent to treating electricity as an incompressible fluid [6].

But Maxwell was not locked into the incompressible fluid idea. As he wrote in 1873 (Article 770 of his *Treatise on Electricity and Magnetism*), in an echo of Faraday, 'a moving electrified body is equivalent to an electric current'. That, however, is just a description of how electricity was thought to behave; it does not say anything about what it *is*. Maxwell, of course, fully appreciated the deficiency of his day's understanding of electricity and, to illustrate that, here is a wonderful little story, taken from a book by two mathematicians, that reveals Maxwell's awareness of the incomplete state of affairs concerning electricity. It may be apocryphal but, even if it is, it honestly reflects the *present* state of *our* understanding, as well: 'In Cambridge, they tell the following story about Maxwell: Maxwell was lecturing and, seeing a student dozing off, awakened him, asking "Young man, what is electricity?" "I'm terribly sorry, sir," the student replied, "I knew the answer but I have forgotten it." Maxwell's response to the class was, "Gentlemen, you have just witnessed the greatest tragedy in the history of science. The one person who knew what electricity is has forgotten it"' [7]. The ambivalence in Maxwell's mind about the nature of electricity continued until his death and he struggled to the end in a quest that ended in failure [8–12].

Even today, the words of the French mathematician Henri Poincaré (1854–1912) are reflective of our present knowledge of the nature of electricity: 'One of the French scientists who has probed Maxwell's work the most deeply said to me one day, "I understand everything in this book [Maxwells' *Treatise*] except what is meant by a charged sphere"' [13].

Serious experimental studies of the nature of electric current can be traced back to Henry Rowland (1848–1901), Professor of Physics at The Johns Hopkins University who, after first visiting with Maxwell in the summer of 1875, then showed that electrostatic charge in motion (on a disc spinning at more than 3600 r.p.m.) creates a magnetic field 'just like' a current in a wire. In 1879, Rowland's student, Edwin Hall (1855–1938), showed that a current-carrying conductor, when immersed in a magnetic field, develops a potential difference (the Hall effect) perpendicular to the current direction. These two experiments perhaps suggest to the modern mind that electric current is charged particles in motion, but Hall himself seems to have been still thinking of the fluid view ('if we regard an electric current *as a single stream flowing* [my emphasis]' [14]).

Two years later, Joseph John Thomson (1856–1940), the Cavendish Professor of Experimental Physics at Cambridge University, made the first theoretical study of the magnetic effect of moving electric charge [15]. Thomson, credited with discovering the electron, was motivated to do that by his interest in recent experiments involving electric discharges in a vacuum, and he started his paper with these words: 'Particles of matter highly charged with electricity and moving with great velocities form a prominent feature in the phenomena. ... It seems therefore to be of some interest ... to take some theory of electrical action [Thomson 'took' Maxwell's theory] and find out what, according to it, is the force existing between two moving electrified bodies, what is the magnetic force produced by such a moving body, ... '. And that is just what he did.

3. Electromagnetic mass

In his paper, Thomson deduced the form of the force experienced by a charge q as it moves with velocity $\boldsymbol{u}$ through a magnetic field $\boldsymbol{B}$ to be $(1/2)q\boldsymbol{u}x\boldsymbol{B}$. The factor of 1/2 is, as is well known to modern college freshmen, wrong (it was corrected by Heaviside in 1889 when he showed the factor is unity) [16]. But the really exciting outcome of Thomson's analysis was its initial promise of providing an *electromagnetic explanation for mass*. Thomson argued that a moving sphere carrying a uniform surface charge would, according to Maxwell, produce a magnetic field which, in turn, implies energy stored in space (because of the Poynting/Heaviside vector). This energy must come from somewhere, and the obvious place is the mechanical motion of the sphere. That, in turn, means the moving sphere must experience a 'resistance' as it moves through the ether, the mysterious dielectric medium (thought to be necessary for electromagnetic radiation to have something to 'wave' in) that both Thomson and Heaviside (along with most Victorian physicists) believed existed, despite the shocking negative results of the Michelson/Morley interferometer experiments of the mid-1880s. (That puzzling business, they imagined, could perhaps be 'explained' by the ad hoc device of the Lorentz–FitzGerald contraction hypothesis.)

It was via this 'resistance', thought Thomson, that the mechanical energy of motion is converted to electromagnetic energy. This 'resistance' could not, however, be a mere frictional drag as, after all, the orbits of the planets (also rushing through the ether) showed not the slightest decay, even after centuries of observation. Instead, as Thomson wrote, the 'resistance' 'must correspond to the resistance theoretically experienced by a solid moving through a perfect fluid. In other words, it must be equivalent to an increase in the mass of the charged sphere'. For a sphere of radius a carrying a charge q through a dielectric medium with magnetic permeability μ, Thomson calculated the mass increase (for a slowly moving charge) to be $(4/15)\mu(q^2/a)$, a result independent of the charge velocity. In 1885, Heaviside replaced the 4/15 factor with 2/3, but otherwise agreed with the functional form of Thomson's result [17].

When, however, Thomson took the sphere to be the size of the Earth, charged to the maximum possible (the point at which the electric field at the Earth's surface would be on the verge of ionizing the atmosphere), he calculated a mass increase of only 'about 650 tons, a mass which is quite insignificant when compared with the mass of the Earth'. Because of this failure to 'explain' Earth's mass, Thomson did not pursue the idea of electromagnetic mass—but Heaviside did.

In 1889, he published a paper in the *Philosophical Magazine* in which he corrected calculation errors in Thomson's 1881 paper [18]. In addition, he extended the analysis from a *slowly moving spherical* charge to that of a *point* charge moving at *any* speed u, right up to the speed of light (which he wrote as v) *and beyond*. Unlike Thomson's speed-independent result, Heaviside's solution included speed-dependent factors of the form $1 - (u/v)^2$, terms that predicted a mass increasing with speed as $u \to v$.

The possibility that at least part of the mass of a charged body could be attributed to the electrodynamic effects of its motion held out the exciting possibility that perhaps *all* of the mass (despite Thomson's failure to account for Earth's mass) could be so explained. If so, and if all matter is merely a vast number of atoms that are themselves mere constellations of fast-moving charged bodies (Thomson's electrons), then the ultimate interpretation of the world would be electromagnetic in nature, rather than the one based on the mechanical models of Lord Kelvin and Maxwell. This hope was eventually shown to be a false one [19–21].

Heaviside soon came to that negative conclusion himself. In one of his surviving, numbered notebooks, he went so far as to claim he had only briefly been seduced by the possibility, writing 'I will not go so far to say that the view which is so popular now, that "mass" is due to electromagnetic inertia, is a mere Will o' the Wisp. I will however say that the light it gives is somewhat feeble and that it eludes or evades distinct localization. The mere *idea*, that electromagnetic inertia *might* account for "mass," occurred to me in my earliest work on moving charges, but it seemed so vague and unsupported by evidence, that I set it on one side. It explains too much, and it does not explain enough' [22].

Despite the 'Einsteinian look' of Heaviside's speed-dependent terms, his analysis was greatly lacking when compared with Einstein's. Heaviside started with moving charged matter and then applied some heavy mathematics to Maxwell's electrodynamics, while Einstein used nothing but the fundamental ideas of space and time, some simple algebra, and the two relativity principles (all physical laws look the same in all inertial frames, and observers in different inertial frames will measure the same value for the speed of light). Einstein's analysis is free of any special assumptions concerning electricity in particular and the nature of matter in general: to put it bluntly, Einstein saw the whole forest, while Heaviside and Thomson were looking through a magnifying glass at the bark on a single tree.

When Heaviside tried to extend his analysis from a point charge to Thomson's spherical surface charge, he stumbled when he assumed the electric field lines would be normal to the sphere's surface. This was a fortuitous error, in one sense, however, as it prompted a letter (dated 19 August 1892) from the young Demonstrator of Experimental Physics at Thomson's Cavendish Laboratory at Cambridge, G. F. C. Searle (1864–1954), who had been working with Thomson since October 1888. That letter was the beginning of a positive relationship, on both the technical and human levels, that would last the rest of Heaviside's life.

In his letter, Searle wrote 'I quite agree with your solution for the motion of an electrified point, but I do not understand how it will apply when we deal with a *sphere*'. Searle went on to show Heaviside that as long as 'u^2/v^2 is utterly negligible' then Heaviside's field line assumption is permissible, but if 'u^2/v^2 is not neglected' then the assumption is *not* valid. Heaviside received this criticism with good spirits, and from that point on, he certainly had a high opinion of his new acquaintance. The fact that Searle, as a youngster, had been given a tour of the Cavendish laboratory by the great Maxwell himself, surely added to Searle's bona fides in Heaviside's eyes as well. In 1896, Searle published corrections to Heaviside's point charge analysis, arriving at expressions including the same famous factor of $\sqrt{1-(u^2/v^2)}$ that appeared in the Lorentz–FitzGerald contraction hypothesis, indicating that they were probing at the very edges of relativistic electrodynamics [23].

4. Faster-than-light

An immediate, impossible to overlook implication of the predicted mass variation with speed is the infinity that occurs at the speed of light. This result is generally interpreted today to mean that nothing with a non-zero rest mass can travel as fast as light (in a vacuum). Searle studied the electromagnetic effects of moving charges for decades and accepted that conclusion, but Heaviside never did. Scattered all through his books, in fact, are analyses concerning charges moving faster-than-light (FTL), with the earliest dating from 1888 [24]. This caused Searle much frustration as, for example, when he wrote in his 1896 paper 'Mr. Heaviside has stated the result when u is greater than v [the speed of light], but has not up to the present divulged the manner in which he has obtained the solution in this case'. Searle was referring to a cryptic Heaviside comment in *Electrical Papers*, where he states a result for $u > v$ and then writes 'I regret that there is no space for the mathematical investigation, which cannot be given in a few words' [25]. Just like Fermat's Last Theorem scribbled in a book page margin!

The following year (1897) Searle published a continuation of his 1896 paper, in which he worked out the energy associated with moving charged ellipsoids. His closing words were 'In all these cases it will be found that when $u = v$ the energy becomes infinite, so it would seem to be impossible to make a charged body move at a greater speed than that of light' [26]. Heaviside rejected Searle's caution, and in 1898 he wrote a reply to offer a 'proof' showing that Searle was wrong: 'The argument ... seems to be that since the calculated energy of a charged body is infinite ... at the speed of light, and since this energy must be derived from an external source, an infinite amount of work must be done. ... There is a fallacy here. One easy way of disproving the argument ... is to use not one, but two bodies, one positively and the other negatively charged to the same degree. Then the infinity disappears, and there you are, with finite energy when moving at the speed of light' [27].

With that slightly desperate analysis, Heaviside was apparently thinking of the q^2 factor in his energy-increase expressions being replaced with $\{q + (-q)\}^2 = 0$. He ignored the alternative possibility of $q^2 + (-q)^2 = 2q^2$. This was all pre-Einstein, of course, and Heaviside did not know the infinity comes from (much) deeper considerations about the natures of space and time than simply whether or not the mass carries a net electric charge.

rsta.royalsocietypublishing.org Phil. Trans. R. Soc. A 376: 20170448

Heaviside discussed the issue of FTL motion with others besides Searle; the most influential of all (perhaps more so than even Searle) was the Irish physicist George Francis FitzGerald (1851–1901), the 'FitzGerald' in the Lorentz–FitzGerald contraction hypothesis. Heaviside had enormous admiration for FitzGerald, and so asked for his friend's opinion. In an undated letter (most likely from early 1899), the response was probably not what Heaviside was looking for: 'You ask "what if the velocity be greater than that of light?" I have often asked myself that but got no satisfactory answer. The most obvious thing to ask in reply is "Is it possible?"' [28].

Heaviside clearly thought the answer is *yes*, and devoted years to thinking about charges moving at light and superluminal speeds. In 1903, he wrote of the visual image he had of charged FTL bodies: 'The photographs taken some years ago by Prof. Boys of flying bullets showed the existence of a mass of air pushed along in front of the bullet. Is there anything analogous to this in the electromagnetics of an electron? Suppose, for example, that an electron is jerked away from an atom so strongly that its speed exceeds that of light. ... So long as its speed is greater than that of light, *it is accompanied in its motion by a conical wave* [my emphasis]' [29].

Heaviside's reference was to Charles Vernon Boys (1855–1944), the inventor of the high-speed *Boys camera*, which used a steel mirror revolving at 60 000 r.p.m. to take pictures via the light of electric sparks mere microseconds in duration. In the early 1890s, Boys published the dramatic photographs that caught Heaviside's attention, showing the conical acoustical shock waves produced by bullets moving faster than the speed of sound (up to 1400 mph) [30]. (Modern aficionados of high-speed photography—almost always—think first of MIT electrical engineer Harold Edgerton (1903–1990), but Boys was taking such pictures 10 years before Edgerton was born.) As impressed as he was with Boys' photographs, they were not the inspiration behind Heaviside's FTL work, as he was writing on the effects of FTL charged particles five years before Boys' images appeared.

The shock wave, for both bullet and subatomic particle, is created when the object moves through a medium at a speed faster than the medium can propagate a disturbance. For a bullet, the crucial speed is that of sound, and for a charged particle, the crucial speed is the speed of light *in the medium*. So, in a certain physical sense, Heaviside was correct in his belief of the possibility of superluminal motion, *if* the medium is other than a vacuum. Then the speed of light is *less* than it is in a vacuum, a charged particle *can* exceed the local speed of light, and Heaviside's conical, electromagnetic shock wave *is* observed. Today, we call it *Cherenkov radiation*, after the Russian physicist Pavel Cherenkov (1904–1990) who studied it experimentally in the 1930s. (Madam Curie had, years earlier, observed the pale-blue glow from FTL electrons emitted in radium solutions, but had not appreciated the glow's significance. Submerged nuclear reactors—so-called swimming pool reactors—produce the same glow from their production of FTL electrons [31].) Along with two colleagues, Cherenkov received the 1958 Nobel physics prize. Alas, none of the three so much as mentioned Heaviside in their Nobel talks, even though he had anticipated their work (indeed, Heaviside's work anticipated the first *observation* of the radiation) by decades [32].

Of course, Heaviside did not do everything, but then neither did Cherenkov and his colleagues: it was not until 1963 that some very peculiar properties of the electric field inside and on the conical wave of an FTL charge were finally explained [33]. For example, it seemed that Gauss' law failed at FTL speeds—the integral of the electric field (which is found to point *towards* a positive FTL charge rather than away as it does at sublight speeds) over a closed surface appears to *not* equal the enclosed charge, but rather to *diverge*! Heaviside was both stunned and stumped by that result, calling it 'an impossible electrical problem' [34].

Despite this theoretical puzzle, however, Heaviside's belief in FTL particles did not waver. In early 1898, for example, he wrote (in connection with 'Cathode Rays' and 'X-Rays') 'J. J. Thomson and others have lately concluded from experiment that immense speeds of the charged particles, comparable with the speed of light, are concerned. If this be fully confirmed, we may well believe that increased voltage will produce speeds exceeding that of light, if they do not exist already, and so bring in the conical theory' [2]. And so here we see Heaviside's prediction of Cherenkov radiation *six years* before Cherenkov's birth (as well as sounding, just a bit, like the starship captain in the quotation that opens this paper). It had been known since 1850 that the speed of light in water is significantly less than it is in a vacuum, and Heaviside could have suggested that his FTL conical theory predictions be looked for—if he had, perhaps Cherenkov radiation might be known today instead as *Heaviside radiation* and his life would have taken a profoundly different path.

According to modern physics, for more than 40 years after Heaviside's death, it would have been just wrong in believing FTL particle motion *in a vacuum* to be possible. That is because, as already mentioned, of the energy infinity when a particle is accelerated up to the speed of light. Einstein's special theory of relativity places an upper limit on the speed of particles with non-zero rest mass because of that infinity; particles of *light*, however, are excepted from that infinity as there is *no* inertial frame in which a photon is at rest (*every* inertial frame observer *always* measures the *same* speed of light), and so the photon is taken to have zero rest mass (which nicely does away with the infinity problem). And in any case, photons are not *accelerated* up to the speed of light, but are *created* (by well-known physical processes) *at* the speed of light from their first instant of existence. They are therefore immune from the objection so elegantly expressed by Haskel Van Manderpootz about the difficulty of *attaining* light speed: that would require 'the use of more than an infinite number of horsepowers' [4].

And then we come to 1967.

That was the year the American physicist Gerald Feinberg (1933–1992) suggested the possibility of an FTL particle he called the *tachyon* (from the Greek word *tachys* for 'swift') [35]. He imagined that tachyons are particles created by some (as yet unknown) physical process, moving from their first instant of existence at a speed *greater* than that of light; for tachyons, the speed of light is a *lower* limit. For Heaviside, with his unshakeable Victorian belief in both the ether and in Newton's absolute, universal time, FTL simply meant that such a particle would complete a journey in less time than it would take light. With the coming of relativity, however, it was soon realized that the implications of FTL are much deeper than just 'getting there sooner'. In his 1917 book, *The Theory of the Relativity of Motion*, Richard Tolman (1881–1948), then a professor of physical chemistry at the University of Illinois, wrote 'The question naturally arises whether velocities which are greater than that of light could ever possibly be obtained'. He answered that question in a way that would have almost certainly have astonished the post-1900 Heaviside (as early as the turn of the century Heaviside was complaining of memory loss and general ill health, and two years before his death told a visitor 'he had forgotten all his writings'): in such a case, it was conceivable that an observer could see an effect *before* its cause. Tolman's statement (which came to be known as *Tolman's paradox*) is usually thought to be rather odd, of course, but he was careful to also state that such a thing was *not* 'a logical impossibility'.

This bizarre idea quickly caught on with the popular press, with the best-known example probably being the limerick by A. H. R. Buller (1874–1944) that appeared in a 1923 issue of *Punch*:

> 'There was a young lady named Bright
> Whose speed was far faster than light;
> She set out one day
> In a relative way
> And returned on the previous night.'

So, with Tolman's paradox in mind, Feinberg's suggestion caught the imaginations of many normally cautious analysts, and sparked an explosion of literature (by both physicists and philosophers) on what would be the tachyon's actual behaviour. The central objection to tachyons quickly zeroed in on such particles seeming to open the door to the possibility of sending messages (and not just Miss Bright) into the past, a possibility that nearly all physicists rejected as being simply absurd (but some philosophers were not so sure about that) [36]. Indeed, for that property alone the tachyon seemed, to many, to be science fiction for romantic juveniles, not physics for serious thinkers. (Ironically, physicist Feinberg admitted that he got his idea for the tachyon after reading a 1954 science fiction tale!)

rsta.royalsocietypublishing.org *Phil. Trans. R. Soc. A* **376**: 20170448

Well, what would *Heaviside* have made of the apparent ability of tachyons to visit the past? He never wrote a word on such things, so let me speculate just a bit.

5. And into the past

Transmitting information into the past is a poor man's version of time travel, about which the popular mind conjures up the image of H. G. Wells' Time Traveller climbing into the saddle of a machine that, at the touch of a crystal-knobbed lever, vanishes like the morning mist into either the future or the past. That image probably strikes most physicists as being a hopeless fantasy, and there *is* a lot wrong with it (as Wells fully understood) when analysed carefully. To achieve the illusion of being more than a mere childish fantasy, Wells fell back on the popular fascination in the late nineteenth century with the fourth dimension (which, in *The Time Machine*, is taken to be time). Serious analysts of the late-Victorian age took that association with a grain of salt and a wry smile, and dismissed *The Time Machine* as simply a clever artistic ruse. Wells, too, never took the idea seriously, writing in the preface of a 1934 collection of his scientific romances that the fourth dimension, for him, was simply a 'magic trick for a glimpse of the future'. Three years earlier, in a 1931 edition of the novella, he declared the fourth dimension as time to be nothing more than a whimsical fantasy.

Then, in 1974, the American physicist Frank Tipler wrote a doctoral dissertation (University of Maryland) that caused quite serious people to rethink the possibility of time machines, and of time travel to the past. What Tipler did was nothing less than describe, using Einstein's general theory of relativity, *how to make a time machine*. Indeed, the final sentence of the journal paper reporting his work is pretty blunt: 'If we construct a sufficiently large rotating cylinder, we create a time machine' [37]. Subsequent studies have focused on the clearly *stupendous* engineering difficulties of constructing Tipler's time machine (three hard-to-miss difficulties are that the cylinder is infinitely long, is made from stuff that is fantastically more dense than any known matter, and is spinning at half the speed of light). Other various space–time geometries that have been discovered since 1974 (wormholes and cosmic strings) have their own engineering problems as well, but the *physics* seems to be valid [36].

Wells died in 1946, and so we do not know how he would have reacted to Tipler's discovery. We can do a bit better with Heaviside, however, as in July 1893 he wrote the following comment, which shows how perplexed he was about gravity: 'To form any notion at all of the flux of gravitational energy, we must first localize the energy [as he and Poynting did with electromagnetic energy]. In this respect it resembles the legendary hare in the cookery book. Whether this notion will turn out to be a useful one is a matter for subsequent discovery' [38]. The warped space–times of post-1974 time machines would have left Heaviside completely dumbstruck.

And I think Heaviside would almost certainly have agreed with Wells' reaction to the fourth dimension. In one of his surviving, personal copies of *Nature*, next to the review of a book titled *Euclid's Parallel Postulate*, which quotes the author as writing 'We may some time gain experience of a new kind, presenting itself as spatial and requiring us to assume more than three dimensions in space', Heaviside scribbled in reply 'Can't see it, and never could' [39]. This, admittedly, is not quite the view of the fourth dimension presented in *The Time Machine*, but spatial or temporal, I think Heaviside's opinion of the fourth dimension would be an unhappy one.

But let us forget the mathematical physics. For most people, the most significant obstacle to taking time travel to the past seriously is, I believe, the *logical paradoxes* that the concept seems to create in abundance. There are many other serious issues with the *science* of time travel and time machines, of course, but I will limit myself here to just the paradoxes, because they (as opposed to relativistic space–time physics) would have been well within Heaviside's ability to instantly appreciate *if he had thought of them*.

And, he might well have *not* thought of them, as even Wells, himself, failed to pursue the paradoxical aspects of backwards time travel. Now Wells was not oblivious to the possibility of paradoxes, but his failure to weave them into *The Time Machine* seems to indicate he simply did not know what to make of such puzzles. In the opening of the tale, for example, during the dinner party at which the Time Traveller tries to convince his friends of the possibility of a time machine, one of them observes 'It would be remarkably convenient for the historian. One might travel back and verify the accepted account of the Battle of Hastings, for instance'. To that another guest replies 'Don't you think you would attract attention? Our ancestors had no great tolerance for anachronisms'. The Time Traveller has no reply to any of that because, I believe, Wells had no reply.

The issue of paradoxes is not mentioned again in *The Time Machine* until near the end. And once again, it seems clear that Wells was puzzled by paradoxes. When a dishevelled Time Traveller returns from AD 802, 701 and beyond, one of his incredulous friends exclaims 'What *was* this time traveling? A man couldn't cover himself with dust by rolling in a paradox, could he?'

The best known paradox is almost certainly the *grandfather paradox*: what if the Time Traveller journeys back to when granddad is a baby and kills him—from where then comes the Time Traveller? This is an old paradox, one that generations of science fiction writers have had a lot of fun with, but in fact there *is* a resolution to it that has found general agreement among most physicists and philosophers, a resolution that does *not* conclude that time travel to the past is impossible [36,40]. There are, however, plenty of other even more bizarre paradoxes that have *not* been so tamed, and I'll concentrate here on just one type, a type that lends itself in a natural way to Heaviside's FTL particles, as well as to his early life as a telegraph operator. I am referring to the possibility (or not) of creating an *information loop in time*.

6. Messages to the past

The classic science fiction example of a temporal information loop is the story of a physicist who has been trying, without success, to build a time machine. Then, when he has become sufficiently frustrated as to be just about ready to toss the whole business in the trash, a silent visitor (who bears a remarkable resemblance to our hero) appears on the physicist's doorstep, hands him a thick notebook and leaves. In the notebook are detailed instructions, written in a bold hand, in ink, on how to build a time machine. This the physicist does and, 20 years later and the winner of three consecutive Nobel prizes, he decides to journey back to give the notebook to his younger self. We seem to have a self-consistent loop (or do we?—more on this in just a moment), but the obvious, immediate question that presents itself is: *who wrote the notebook?* It certainly *was not* the physicist: the book was handed to him, he kept it in his possession for 20 years and then he handed it back to himself. Not even once did he put pen or ink to paper. So who did?

Perhaps surprisingly, information loops in time have a long literary history that pre-dates the science fiction magazines. In the 1904 novel *The Panchronicon*, for example, we learn how a Shakespeare suffering from writer's block came to write one of his plays: a time-travelling fan from 1898 whispers the magic lines she has memorized (for her literary club meetings) into his ear. Would this make Shakespeare a plagiarist? Of himself? In whose mind did the words originate? [41] This sort of loop does avoid a serious objection to the time-travelling notebook loop: when the physicist gives the notebook to his younger self, both the ink and the paper on which it is written are 20 years older than they were when he received the notebook, and so the closed loop in time is actually *not* a consistent one. There is, in contrast, no such 'ageing' concern in the Shakespeare loop. The difference in the two loops is that one contains a circulating *physical* artefact, and the other does not.

Well, what does all of this have to do with Heaviside's FTL particles? Suppose such particles exist, and further suppose we can build devices that both generate (transmit) and detect (receive) them. If we could modulate a beam of such particles (by, for example, simply turning the beam on and off to transmit Morse code, as Heaviside once did in his youth), then we would have what has been called a *tachyonic anti-telephone* [42]. (Ordinary telephones send messages to the future, and so a communication device that transmits to the past is an *anti*-telephone.) One of the authors of the anti-telephone journal paper (the American physicist Gregory Benford) later used this concept to write a well-received novel in which the world of 1998, on the verge of ecological collapse, tries to warn the world of 1963 via tachyon transmissions [43]. This then brings into play concerns about changing the past, free will and other equally bothersome issues [36].

Sending messages to the past has (and this should be no surprise) been enthusiastically embraced by science fiction writers [44–49]. But the concept has just as enthusiastically been rejected by no less an authority than Einstein, who emphatically denied the possibility of doing such a thing, specifically writing 'We cannot send wire messages into the past' [50]. That's pretty blunt!

It is my opinion that Wells and Heaviside would both have agreed with Einstein. Wells would have done so, completely and immediately, even before Einstein's relativity, with Heaviside joining him after being confronted with the logical paradoxes. Heaviside certainly would not have liked Einstein's forbidding of FTL particles, but I think the logical paradoxes would have been simply too much for him to accept. Heaviside was often a radical mathematician, yes, but at his core, he was still a conservative, Victorian-age physicist.

In the July 1945 issue of *The Cornhill Magazine*, which appeared just 13 months before his death, Wells published an essay that clearly shows what he thought of time travel, time machines and their associated paradoxes. It is an essay that could have been written by Heaviside's ghost. Appearing under the pseudonym 'Wilford B. Batterave', with the title 'A Complete Exposé of This Notorious Literary Humbug', the essay has a critic (Wells!) describing *The Time Machine* as

> a tissue of absurdities in which people are supposed to rush to and fro along the 'Time Dimension.' By a few common tricks of the story-teller's trade, Wells gets rid of his Machine before it can be subjected to a proper examination. He cheats like any common spook raiser. Otherwise it is plain commonsense that a man might multiply himself indefinitely, pop a little way into the future and then come back. There would then be two of him. Repeat *da capo* and you have four, and so on, until the whole world would be full of this Time Travelling Individual's vain repetitions of himself. The plain-thinking mind apprehends this in spite of all the Wellsian mumbo-jumbo and is naturally as revolted as I am by the insult to its intelligence.

7. A final, personal note

While I do not think we will ever be sending messages back in time, the idea *is* undeniably one of immense appeal. Who can deny the infinite seductiveness (if only it were possible) of sending a copy of this commemorative issue of the *Transactions* back in time to the 25-year-old Heaviside of 1875? Who can deny how excited they would be to knock on the door of 117 Camden Street, Camden Town, London, the home of his parents (with whom he then lived), and to hand him this tribute to his future accomplishments? Well, of course, all those pesky logical paradoxes then come into play, and there goes the idea of such a wonderful visit right out the window!

Or does it? More than 30 years ago I thought long and hard about reversing the role of Heaviside, by making *him* the time traveller. The result was, I believe, a paradox-free scenario [51]. The story may be difficult to find today, but if you write to me I will send you a pdf-scan and you can tell me what you think.

Data accessibility. This article has no additional data.
Competing interests. We declare we have no competing interests.
Funding. We received no funding for this study.

References

1. Schachner N. 1936 Reverse universe. *Astounding Stories*.
2. Heaviside O. 1899 *Electromagnetic theory*, vol. 2, p. 535. London, UK: Electrician Publishing.
3. Nahin PJ. 2002 *Oliver Heaviside: the life, work, and times of an electrical genius of the Victorian age*. Baltimore, MD: The Johns Hopkins University Press (originally published in 1987).
4. Weinbaum S. 1935 The worlds of if. *Wonder Stories*.
5. Faraday M. 1839 *Experimental researches*, vol. 1, article 1644. London, UK: Taylor and Francis.

6. Thomson W. 1855 On the theory of the electric telegraph. *Proc. R. Soc. Lond.* **7**, 382–399.
7. Doyle PG, Snell JL. 1984 *Random walks and electric networks*, p. 70. Washington, DC: Mathematical Association of America.
8. Heimann PM. 1971 Maxwell, Hertz, and the nature of electricity. *Isis* **62**, 149–157.
9. Moyer DF. 1978 Continuum mechanics and field theory: Thomson and Maxwell. *Stud. Hist. Phil. Sci.* **9**, 35–50. (doi:10.1016/0039-3681(78)90020-1)
10. Turner J. 1956 A note on Maxwell's interpretation of some attempts at dynamical explanation. *Ann. Sci.* **11**, 238–245. (doi:10.1080/00033795500200235)
11. Klein MJ. 1972 Mechanical explanation at the end of the nineteenth century. *Centaurus* **17**, 58–82.
12. Kargon R. 1969 Model and analogy in Victorian science: Maxwell's critique of the French scientists. *J. Hist. Ideas* **30**, 423–426. (doi:10.2307/2708567)
13. Bromberg J. 1968 Maxwell's electrostatics. *Am. J. Phys.* **36**, 142–151. (doi:10.1119/1.1974442)
14. Hall EH. 1879 On a new action of the magnet on electric current. *Am. J. Math.* **2**, 287–292. (doi:10.2307/2369245)
15. Thomson JJ. 1881 On the electric and magnetic effects produced by the motion of electrified bodies. *Phil. Mag.* **11**, 229–249. (doi:10.1080/14786448108627008)
16. Heaviside O. 1892 *Electrical papers*, vol. 2, p. 506. London, UK: Macmillan.
17. Heaviside O. 1892 *Electrical papers*, vol. 1, p. 446. London, UK: Macmillan.
18. Heaviside O. 1892 *Electrical papers*, vol. 2, p. 504–518. London, UK: Macmillan.
19. McCormmach RH. 1970 A. Lorentz and the electromagnetic view of nature. *Isis* **61**, 459–497. (doi:10.1086/350681)
20. Hirosige T. 1966 Electrodynamics before the theory of relativity, 1890–1905. *Jap. Stud. Hist. Sci.* **5**, 1–49.
21. Cushing JT. 1981 Electromagnetic mass, relativity, and the Kaufmann experiments. *Am. J. Phys.* **49**, 1133–1149. (doi:10.1119/1.12561)
22. *Heaviside notebook 18*, p. 326, Heaviside collection. London, UK: Institution of Engineering and Technology.
23. Searle GFC. 1896 Problems in electric convection. *Phil. Trans. R. Soc. Lond. A* **187**, 675–713. (doi:10.1098/rsta.1896.0017)
24. Heaviside O. 1892 *Electrical papers*, vol. 2, p. 494. London, UK: Macmillan.
25. Heaviside O. 1892 *Electrical papers*, vol. 2, p. 516. London, UK: Macmillan.
26. Searle GFC. 1897 On the steady motion of an electrified ellipsoid. *Phil. Mag.* **44**, 329–341. (doi:10.1080/14786449708621072)
27. Heaviside O. 1899 *Electromagnetic theory*, vol. 2, pp. 533–535. London, UK: Electrician Publishing.
28. Hackett FE. 1952/1954 FitzGerald as revealed by his letters to Heaviside. *Sci. Proc. R. Dublin Soc.* **26**, 3–7.
29. Heaviside O. 1912 *Electromagnetic theory*, vol. 3, pp. 164–169. London, UK: Electrician Publishing.
30. Boys CV. 1893 On electric spark photographs. *Nature* **47**, 415–421 and 440–446. (doi:10.1038/047415a0)
31. Jelley JV. 1958 *Cherenkov radiation*. Elmsford, NY: Pergamon Press.
32. 1964 *Nobel lectures (physics), 1942–1962*, pp. 421–483. Amsterdam, The Netherlands: Elsevier.
33. Volkoff GM. 1963 Electric field of a charge moving in a medium. *Am. J. Phys.* **31**, 601–605. (doi:10.1119/1.1969690)
34. Heaviside O. 1892 *Electrical papers*, vol. 2, p. 499. London, UK: Macmillan.
35. Feinberg G. 1967 Possibility of faster-than-light particles. *Phys. Rev.* **25**, 1089–1105. (doi:10.1103/PhysRev.159.1089)
36. Nahin PJ. 2017 *Time machine tales: the science fiction adventures and philosophical puzzles of time travel.* New York, NY: Springer.
37. Tipler FJ. 1974 Rotating cylinders and the possibility of global causality violation. *Phys. Rev. D* **9**, 2203–2206. (doi:10.1103/PhysRevD.9.2203)
38. Heaviside O. 1893 *Electromagnetic theory*, vol. 1, pp. 455–466. London, UK: Electrician Publishing.
39. 1907 *Heaviside's personal copy of nature*, pp. 220–221. London, UK: Heaviside Collection, Institution of Engineering and Technology.

40. Lewis D. 1976 The paradoxes of time travel. *Am. Phil. Q.* **13**, 145–152.
41. Mackaye HS. 1904 *The panchronicon.* New York, NY: Charles Scribner's Sons.
42. Benford G, Book DL, Newcomb WA. 1970 The tachyonic antitelephone. *Phys. Rev. D* **2**, 263–265. (doi:10.1103/PhysRevD.2.263)
43. Benford GA. 1980 *Timescape.* New York, NY: Simon and Schuster.
44. Finney J. 1981 The love letter. In *Tales out of time* (ed. B Ireson). New York, NY: Philomel.
45. Scortia TN. 1981 When you hear the tone. In *The best of Thomas N. Scortia* (ed. G Zebrowski). Garden City, NY: Doubleday.
46. Padgett L. 1953 Line to tomorrow. In *Worlds of tomorrow* (ed. A Derleth). New York, NY: Pellegrini & Cudahy.
47. Stapleton D. 1942 *How much to Thursday?*. *Thrilling Wonder Stories*.
48. Leinster M. 1976 Sam, this is you. In *The best of Murray Leinster* (ed. B Davis). London, UK: Corgi.
49. Blish J. 1976 Beep. In *Galactic empires* (ed. B Aldiss), vol. 2. New York, NY: St Martin's Press.
50. Einstein A. 1922 La Théorie de la Relativité. *Bull. Soc. Francaise Phil.* **17**, 91–113.
51. Nahin PJ. 1985 The invitation. In *The fourth Omni book of science fiction* (ed. E Datlow). New York, NY: Zebra.
rsta.royalsocietypublishing.org *Phil. Trans. R. Soc. A* **376**: 20170448

PHILOSOPHICAL TRANSACTIONS A

rsta.royalsocietypublishing.org

Review

Cite this article: Donaghy-Spargo C. 2018 On Heaviside's contributions to transmission line theory: waves, diffusion and energy flux. *Phil. Trans. R. Soc. A* **376**: 20170457. http://dx.doi.org/10.1098/rsta.2017.0457

Accepted: 11 June 2018

One contribution of 13 to a theme issue 'Celebrating 125 years of Oliver Heaviside's 'Electromagnetic Theory''.

Subject Areas:
electrical engineering

Keywords:
diffusion, electromagnetism, Oliver Heaviside, transmission lines, wave propagation

Author for correspondence:
Christopher Donaghy-Spargo
e-mail: christopher.spargo@durham.ac.uk

THE ROYAL SOCIETY PUBLISHING

On Heaviside's contributions to transmission line theory: waves, diffusion and energy flux

Christopher Donaghy-Spargo

Department of Engineering, Durham University, South Road, Durham DH1 3LE, UK

CD-S, 0000-0001-5337-4522

This paper surveys some selected contributions of Oliver Heaviside FRS (1850–1925) to classical electromagnetic theory and electrical engineering science. In particular, the paper focuses on his contributions to the development of electrical transmission line theory and his deep insights into the 'physical' nature of the phenomena relating to nineteenth century telegraphic problems. Following a brief historical introduction to the life of Heaviside to put his achievements in context, we explore his contributions to the reformulation of Maxwell's equations and the understanding of electromagnetic wave propagation along the external region of transmission lines. This leads naturally to his researches regarding the electromagnetic diffusion process inside the line conductors and his subsequent realization that the circuital parameters, usually assumed constant, are not always so. Finally, taking both these internal and external viewpoints of the conductors, his important work regarding the flow of energy described by his 'energy current' concept is presented.

This article is part of the theme issue 'Celebrating 125 years of Oliver Heaviside's 'Electromagnetic Theory''.

1. Introduction

Oliver Heaviside FRS (figure 1) was born on 18 May 1850 at 55 King Street (now known as Plender Street, Camden), London, UK, and died in Torquay on 3 February 1925, at the age of 74. Perhaps one of the most remarkable aspects of Heaviside's life was that he

Figure 1. Oliver Heaviside FRS (SC MSS 3/B/2/141—courtesy of the Institution of Engineering & Technology Archives).

left school at a young age, having received no formal education beyond the age of 14—thus, he would have left with little knowledge beyond algebra and trigonometry. Despite this lack of advanced mathematical and scientific training, he was able to make very many significant scientific achievements to rival the most revered of his university-educated peers in the great Victorian era of scientific discovery. His trajectory from schoolboy to an eminent *electromagnetician* (as prescribed by his friend G.F.C. Searle FRS) can be thought of as an embodiment and direct result of the well-known 'self-improvement' philosophy that was a staple part of the Victorian middle-class culture in which he was brought up. In Heaviside's case, this seemed to be fuelled by indefatigable intellectual curiosity and the desire for deep knowledge and understanding of the unknown—he was a self-taught electrical engineer, physicist and mathematician, finding his own way through the new electrical engineering science of the nineteenth century. Today, contemporary degree-qualified electrical engineers, physicists and mathematicians are likely to know little of Heaviside or his works and his significant influence in the field of electromagnetic theory. It can be argued that his work is vastly underappreciated. This is surprising as he made major and lasting contributions not only to classical physics, but also to both pure and applied mathematics, as well as countless contributions to the electrical engineering science that underpins most modern electrical/electronic technology. The *Heaviside Centenary Volume* published by the Institution of Electrical Engineers (IEE) provides a good account of his outstanding works in these areas, including some important findings from his unpublished notebooks [1]. H.J. Josephs, a Post Office Research Station engineer, gives valuable insight into those unpublished works held at the archives of the Institution of Engineering and Technology (IET), showing how Heaviside developed new methods to solve indefinite integrals, linked with rather tricky electrical circuit problems. The *Centenary Volume* also includes a sketch of Heaviside's personality by his friend, Cavendish Professor G.F.C. Searle FRS.

Heaviside's approach in his work was that of the 'physical'; he sought to understand the physical nature of the electrical and electromagnetic phenomena, despite his heavy use of mathematics. Much of his fundamental and most important work on electromagnetic theory, transmission lines and electric circuit theory is contained within his two volumes of *Electrical papers*, published in 1892. It is true to say that the scope of his work was wide ranging, at first focused on the great practical telegraphic and telephonic problems of the late nineteenth century, leading to the invention of inductive loading coils and a patent for the coaxial cable [2]. Later he explored (in his three volumes on electromagnetic theory, published from 1893 onwards, *125 years ago*) gravitational and electromagnetic analogies, waves from moving sources and even skirted around superconductivity with his comments on low-temperature conductors.

Heaviside was an engineer, but he was also a mathematical pioneer and an early adopter of the vector analysis and calculus, developing it to suit his engineering and physics problems. More than 50 pages of the *Philosophical Transactions of the Royal Society* are devoted to his development of these vector methods. Using these methods, he made a most substantive and very important contribution by recasting James Clerk Maxwell's 22 quaternionic equations into the four coupled partial differential equations we know today [3]. Maxwell had published his two-volume work *A treatise on electricity and magnetism* in 1873 [4]. Heaviside first came across this seminal work as part of his 'self-study' regime while he was living in Newcastle upon Tyne in late 1873, while working for the Great Northern Telegraph Company (his only paid work, which he left in 1873 at the age of 23 to concentrate on his studies). A description of this encounter is best kept in his own witty words, as he recollected in a personal letter later in life [5]:

> I remember my first look at the great treatise of Maxwell's when I was a young man ... I saw that it was great, greater and greatest, with prodigious possibilities in its power ... I was determined to master the book and set to work. I was very ignorant. I had no knowledge of mathematical analysis (having learned only school algebra and trigonometry which I had largely forgotten), and thus my work was laid out for me. It took me several years before I could understand as much as I possibly could. Then I set Maxwell aside and followed my own course. And I progressed much more quickly ... It will be understood that I preach the gospel according to my interpretation of Maxwell.

Following Maxwell, cutting his own path forward through electromagnetic theory, he explored the fringes of relativistic electrodynamics in 1888 [6,7] and studied the physical implications of hyper-light particle motion, leading him to the theoretical prediction of electromagnetic radiation under these conditions in 1888–1889 [8]. This radiation was verified experimentally by Cherenkov, for which his team received the Nobel Prize in Physics in 1958. Heaviside first studied a generalized form of the 'skin effect' in conductors in 1885 [9], and in 1902 he published a theory of global electromagnetic wave propagation using a conducting region in the atmosphere, now known as the Heaviside–Kennelly layer [10]. These remarkable achievements, among many others, led to many accolades, including election in 1891 as a Fellow of the Royal Society and in 1908 an Honorary Member of the IEE. In 1922, 3 years before his death, he was awarded the very first Faraday Medal by the IEE. These achievements are outstanding for a self-taught, lower middle-class, independent researcher and the son of a wood engraver from Stockton-on-Tees. Heaviside was a remarkable man, an original thinker with brilliant mathematical powers and physical insight.

It is beyond the scope of this paper to explore deeper into the life of Oliver Heaviside. For an excellent and in-depth account of his life, illuminating his relationships with eminent scientists of the age and with the 'establishment', the reader is directed to the excellent books on the subject [7,11,12]. Here, we are concerned mainly with his achievements in the field of electromagnetic theory in relation to transmission lines, this brief introduction serving to 'set the scene' for those not familiar with the life and work of Heaviside and to provoke the reader to explore deeper into the brief and limited descriptions of his contribution here. It is not claimed here that the contributions within are comprehensive; they merely scratch the surface of Heaviside's work until around 1900. The focus of this paper is on his studies on electromagnetic wave propagation, electromagnetic diffusion and energy flow.

2. Electromagnetic waves

Heaviside made many original contributions to the theory of electromagnetic wave propagation through his desire to develop a comprehensive theory of the transmission line and its application to nineteenth century telegraphic problems. As pointed out by Baron Jackson of Burnley FRS [1], Heaviside published over 'one and a half million words' on his researches; Burnley gives a detailed overview of some of these contributions. Presented here are those original contributions that are perhaps the most important in the development of his comprehensive theory, setting the scene for his subsequent work.

(a) Heaviside's equations: duplex forms

Most of Heaviside's step change contributions begin with his interpretation of Maxwell's theory of the electromagnetic field as he discovered it in 1873 [3,4]. The so-called *Maxwell's* equations, a set of partial differential equations that describe electromagnetic field interactions and that are a cornerstone of classical physics, developed by Maxwell from 1861 onwards, are immediately recognizable to physicists, electrical engineers and most mathematicians today [13]. In differential form and modern notation, they can be written as follows:

$$\left.\begin{aligned} \nabla\cdot\mathbf{D}(\mathbf{r},t) &= \rho, \\ \nabla\cdot\mathbf{B}(\mathbf{r},t) &= 0, \\ \nabla\times\mathbf{E}(\mathbf{r},t) &= -\frac{\partial\mathbf{B}(\mathbf{r},t)}{\partial t} \\ \text{and}\qquad \nabla\times\mathbf{H}(\mathbf{r},t) &= \boldsymbol{J}(\mathbf{r},t) + \frac{\partial\mathbf{D}(\mathbf{r},t)}{\partial t}. \end{aligned}\right\} \tag{2.1}$$

Here $\mathbf{D}$ is the electric displacement field, ρ is the electric charge density, $\mathbf{B}$ is the magnetic flux density, $\mathbf{E}$ is the electric field, $\mathbf{H}$ is the magnetic field intensity and finally $\boldsymbol{J}$ is the electric current density; all being a function of space $\mathbf{r}$ and time t. These equations, along with the constitutive relations, describe electromagnetic phenomena. However, these equations are not *really* from Maxwell. The original electrodynamic equations from 1873 were written in a different language, that of the mathematics of quaternions, an extension of complex numbers developed by Sir William Hamilton FRSE. This language obscured the physics according to Heaviside, and he viewed the mathematical presentation as 'hard to read and lengthy in execution' [14]. Much preferring his vector methods (alongside J.W. Gibbs), Heaviside was first to recast the 22 Maxwell equations into this vector algebra form, as is common today in almost all textbooks on the subject. This pioneering work made the theory of electromagnetism more accessible and readily able to be applied to solve practical problems such as those Heaviside sought to solve, as a practical 'telegrapher'. It must be noted that Heaviside had not changed the content of Maxwell's work, but was 'preach[ing] the gospel according to [his] interpretation of Maxwell', albeit in a different language. However, this achievement is not trivial [3], it was a long road to the form we are now familiar with.

As if this recasting was not significant enough, Heaviside then turned his attention to the 'symmetry' of the equations when electric charges and conduction currents are present. The curl equation for the $\mathbf{E}$ field is not symmetric with the curl equation for the $\mathbf{H}$ field. Heaviside presented his vectorial 'duplex form' of the equations in 1884 [9] when he introduced hypothetical magnetic conduction currents $\boldsymbol{G}$ and magnetic charge ρ_{m} to preserve this symmetry as found with the 'free space' Maxwell equations. In modern notation:

$$\left.\begin{aligned} \nabla\cdot\mathbf{D}(\mathbf{r},t) &= \rho_{\mathrm{e}}, \\ \nabla\cdot\mathbf{B}(\mathbf{r},t) &= \rho_{\mathrm{m}}, \\ -\nabla\times\mathbf{E}(\mathbf{r},t) &= \boldsymbol{G}(\mathbf{r},t) + \frac{\partial\mathbf{B}(\mathbf{r},t)}{\partial t} \\ \text{and}\qquad \nabla\times\mathbf{H}(\mathbf{r},t) &= \boldsymbol{J}(\mathbf{r},t) + \frac{\partial\mathbf{D}(\mathbf{r},t)}{\partial t}. \end{aligned}\right\} \tag{2.2}$$

He realized that these artefacts had no experimental evidence and said there is 'probably no such thing' [9], but maintained that the duplex equations (equation (2.2)) should be presented for completeness, symmetry and some *analytical advantage*. While, in reality, the magnetic charge and its magnetic conduction current go to zero, if one considered both an electric and a magnetic conduction current and associated heat losses when an electromagnetic wave propagates and that they could be somehow balanced, the wave would travel distortion free [15]. This was a major result and leap forward into the theoretical understanding of how to improve communications systems which was to have a profound impact on Heaviside's later work regarding transmission lines; later work was based on the application of his reformulation of Maxwell's original equations. The work led him to explore various aspects of electromagnetic wave propagation, including spherical waves and harmonics, electromagnetic radiation from faster-than-light (in the medium) particles and attenuation of electromagnetic waves in seawater throughout his volumes on electromagnetic theory (published between 1893 and 1912). His works are notoriously difficult to read due to their volume and difficulty and Heaviside's own notation used in his mathematical work—this can easily be seen with a glance through any volume of his *Electrical papers* or those on electromagnetic theory. It will also be noted that there are very few diagrams among the mathematics.

(b) The telegrapher's equations

With his new duplex equations, Heaviside turned to solving some practical problems. Long-distance telegraphy with 'good' signal rates was a significant technical challenge; many companies involved in the commercialization of this technology struggled for faster communication over longer distances. The signal was found to be 'smeared out' at the receiving end, reducing the possible signal rate. Eminent scientists such as Lord Kelvin had made some headway in obtaining a theoretical understanding [16], which was in the form of a 'heat diffusion equation' (written here for voltage V only),

$$\frac{\partial^2 V}{\partial x^2} = RC\frac{\partial V}{\partial t}, \tag{2.3}$$

where x is the space coordinate, t is time and R and C are the electrical resistance and capacitance, respectively. Heaviside revolutionized the theory of signalling along wires using 'electromagnetic waves' in his early theoretical work on the subject [17]. Lord Kelvin had missed a crucial component in the physical model, his 'telegraph equation'—*electromagnetic inductance*. Limiting the model to electrostatic capacitance and electrical resistance missed this element of Maxwell's theory and hence had fallen short of the 'wave' description. Heaviside, in his article on 'The extra current' [18] and later more explicitly [19], derived the now familiar differential equations that describe either the voltage V and current I on an electrical transmission line as a function of distance x and time t—which have the form of a damped wave equation,

$$\frac{\partial^2 V}{\partial x^2} = RC\frac{\partial V}{\partial t} + LC\frac{\partial^2 V}{\partial t^2}. \tag{2.4}$$

Here, L is the inductance; the form of the equation is a 'damped wave equation'. Heaviside realized the nature of this equation and readily determined that the speed and characteristics of the signalling depend on the nature of the circuit parameters (now known as the propagation/attenuation constants determined by line parameters R, L and C). Subsequently, he used these now familiarly termed telegrapher's equations to derived closed-form solutions for numerous electric circuit configurations in his volumes of *Electrical papers*, including producing results relating to faults in cables and signalling through telegraph circuits, undoubtedly inspired by his time as a paid telegraph clerk at the Great Northern Telegraph Company [11,12] in Denmark and Newcastle upon Tyne. This deep insight falls out of Heaviside's interpretation of Maxwell's theory and led to his comprehensive understanding of the practical engineering problems of the time, reconceptualizing the fundamentals of telegraphic propagation of signals.

(c) Distributed elements and the Heaviside condition

The signal on a transmission line can become 'distorted', as Victorian telegraphs had found out at an early stage. Heaviside's telegraphy equations gave insights into their behaviour; here, he used the concept of 'distributed circuit elements' for the first time, effectively developing the 'distributed element transmission line model' in the mid-1880s. He investigated the effect of 'continuously distributed resistance' along with a line on wave propagation [20] and moved on to consider uniformly distributed leakage conductance and the other circuit parameters. This work was of great importance, and it ultimately led him to two major advances in the field—one theoretical and one practical.

The first discovery is the mathematical condition necessary for 'distortionless propagation' of a telegraphic signal (but not *attenuation-less*), obtained in 1887 [21] in his work on electromagnetic induction and its propagation. Then, second, using this mathematical condition in conjunction with his 'distributed element model' and operational calculus, he proposed how one could achieve this condition with a practical telegraph line [22]. The condition that Heaviside derived for distortionless transmission was that the line parameters must obey the following equality:

$$\frac{L}{C} = \frac{R}{G}. \tag{2.5}$$

Here R, C and L are the electrical resistance, capacitance and inductance, respectively; now G is the electrical leakage conductance between the 'go' and 'return' conductors of the transmission line. Heaviside showed that, if the equality was satisfied, this would lead to a condition of a 'distortionless circuit' as the 'compensation' of each of the circuit elements' respective effects creates the necessary conditions. This equality is the origin of the *loading coil* as he realized that, in real telegraph lines, typically $R/G \gg L/C$, and, as such, this insight prompted him to suggest an increase in L to fulfil the equality [23] within the constraints of the technical and engineering economic factors known to him. His practical method (1893) was to use discrete inductors (distributed inductive elements) at intervals along the line [22], but he did not patent this idea; consequently, others did and probably made a substantial monetary gain in the process [11,12].

From this volume of work, a large cross-section of modern electrical engineering science can be said to be derived. It forms the foundation for analysing DC circuits through microwave and THz wave propagation. Coupled with his recasting of electromagnetic theory into the vector equations and his development of the operational calculus (not discussed here), Heaviside's work on transmission line modelling and distortionless conditions presents a comprehensive toolset. This toolset was developed for solving both hypothetical (insightful) and practical engineering and physics problems that has 'propagated' into the modern day with great success, having widespread significance and impact on the profession.

3. Electromagnetic diffusion

Heaviside's studies of electromagnetic wave propagation and the transmission line brought him to *electromagnetic diffusion*. It is apparent that he sought an answer to the following question: 'How does electric current or magnetic flux penetrate and flow in media?' He addressed this

rsta.royalsocietypublishing.org Phil. Trans. R. Soc. A 376: 20170457

question using his expert command of Maxwell's theory and some very challenging mathematics from around 1883. Heaviside used his vector form of Maxwell's equations to derive the following partial differential equation for the magnetic field **H** [24]:

$$\nabla^2 \mathbf{H}(\mathbf{r}, t) = \sigma\mu \overbrace{\frac{\partial \mathbf{H}(\mathbf{r}, t)}{\partial t}}^{\text{diffusion}} + \varepsilon\mu \underbrace{\frac{\partial^2 \mathbf{H}(\mathbf{r}, t)}{\partial t^2}}_{\text{wave}}, \tag{3.1}$$

where σ is the electrical conductivity, μ is the magnetic permeability and ε is the electric permittivity. From this equation, it is clear that two different processes are involved in the distribution of a magnetic field, one wave like and one diffusion like. He wrote, in his long-running article on electromagnetic induction and its propagation [24], 'discard the last term when the wire ... is in question and discard the previous when the dielectric is in question'. Heaviside generalized the physical picture of the transmission line to include both wave and diffusion terms, reconceptualizing telegraphic propagation in the process. From here, he set up problems to investigate the phenomena, mostly using his 'coaxial cable' configuration when analysing electric conductors, but his interest was also in the penetration of magnetic flux into highly permeable cores. Perhaps this interest was influenced by the use of *iron* wires, rather than copper, in some nineteenth century telegraph circuits. However, this was to lead him towards the important results regarding the 'skin effect' and guide his research towards considering the 'energy flux' in the dielectric, which is the focus of the next section.

(a) The 'heat-like' behaviour of magnetic cores and electric conductors

The wave effects in the dielectric are ignored in electrically conducting media and thus the second term on the RHS of equation (3.1) vanishes. The resulting equation is then the same equation used originally by Lord Kelvin [16] to describe the 'heat-like' diffusion of electric current in telegraph and submarine cables, neglecting the inductance of such lines. Heaviside changed the viewpoint to that of 'electromagnetic waves'; however, of concern was the induction of currents in the originating circuit as well as those induced in external media.

The first works in which Heaviside made a major attempt at understanding electromagnetic field penetration were in 'The induction of currents in cores' [25]. The penetration of magnetic flux into a magnetic core due to a change in external driving current was investigated in depth. He realized that periodic currents in an exciting coil cause the magnitude of the magnetic flux density to diminish progressively into the core, and he termed this phenomenon 'magnetic waves' but recognized the 'diffusion' behaviour [26] and attributed eddy currents as a driving factor. Heaviside went further and calculated the 'core-heat per second' [26], due to the alternating magnetic field in the core. Subsequently, this approach was extended by many authors, with particular attention on the penetration of magnetic flux into laminated steel sheets (e.g. [27]), as used in electric motors and generators to reduce eddy current losses.

Heaviside's work goes on to describe in great mathematical detail the effects of variable 'inductivity' (magnetic permeability) [26], fringing around the topic of magnetic saturation within the core and its effect on the rate and character of magnetic diffusion. This, as with his other special cases, often resulted in daunting looking series equations scattered with Bessel functions of various orders and argument—his mathematical powers being of great utility with these difficult problems.

Following his work on the characteristics of magnetic cores, in [24], Heaviside changed direction on considered longitudinal currents and the circular magnetic field surrounding straight electrical conductors. Based on his previous work, he recognized at once [25] that

> The propagation of magnetic force and of electric current (a function of the former) in conductors takes place according to the mathematical laws of diffusion, as of by heat conduction.

He set out a complete mathematical treatment of the subject in [26]. In this work, he moved straight from the descriptive differential equation into the solution of the current density distribution as a function of space and time involving *Fourier's cylinder functions* (Bessel functions), immediately obtaining a solution for the magnetic field and the magnetic vector potential of a straight conductor. His solution is, in general, based on an arbitrary impressed force driving the change—he stated that the central part of the wire is the 'last to get its current' on impression of a driving force, and on removal of that force that same central part is the 'last to get rid of its currents'. The current begins at the conductor surface and diffuses inwards; almost as if the current 'enters' from the surface when the circuit is switched on. He then made an astute observation that 'If the steady state is not fully set up before the impressed force is removed, the central part of the wire is less useful as a conductor ... [with] sufficient rapid reversals ... the central part of the wire is practically inoperative'.

His study of electromagnetic diffusion, the complementary electromagnetic process to wave propagation, when analysing transmission lines led him to discuss in very general terms the now well-known 'skin effect'. His analysis is valid for arbitrarily shaped conductors with any transient impressed force, not only persistent, steady-state AC currents. He presented a complete theory of such electromagnetic phenomena, including associated energies and forces [26]. Some discussion is given to the 'proximity' of the return conductor, and the effect of its distance from the main conductor, translatable to the electromagnetic proximity effect that engineers know today.

(b) The skin effect and circuit parameters

The effect of current tending to flow on the surface of a conductor was first described by Horace Lamb FRS [28] in 1883; his analysis was limited to the special case of spherical conductors. A full-generalized analysis was first presented by Heaviside in 1885, following his work on electromagnetic diffusion described above. Following the insights into the electromagnetic field distributions, Heaviside worked on translating the field theory to circuit theory. Using his transmission line wave theory, he found modified expressions for both the resistance and inductance of the lines and discussed the effect of increasing frequency on these parameters [29] due to the change in magnetic and current density field distributions. Heaviside correctly recognized that the resistance is found to be increased and the internal inductance is found to be decreased under these transient circumstances, and he developed complex series solutions for these circuital parameters and formed the circuital descriptive differential equations with the modified parameters to take into account skin and proximity effects in his transmission line model.

As Nahin points out [12], the papers by Lamb and Heaviside firmly established a mathematical description of the skin effect. This statement is extended by the present author to suggest that Heaviside's contribution laid out the proximity effect and expressions for the resistance and inductance of the conductors that experience these transient and persistent electromagnetic effects.

4. The energy flux

Heaviside had made substantial contributions to both the wave phenomenon (the second term on the RHS of equation (3.1)) external to the guiding conductors and the slower diffusive electromagnetic phenomena (the first term on the RHS of equation (3.1)) internal to those conductors. He thought that the sum of those studies does not form a complete physical picture of the transmission line—once again, he set to work. The wave and diffusion viewpoints he considered are concerned with the electric and magnetic fields only; Maxwell's theory of electromagnetism dictates that those fields must have associated energy densities and thus it follows that transfer of energy must take place somewhere—it is this that was missing.

The Russian physicist and mathematician Nikolay Umov first postulated a vector in 1874 to describe a generalized view of energy flux in liquid and elastic media [30], and this energy flux is the rate of transfer of energy—energy movement within media. In essence, this permits new physical insights into how and where the energy flows spatially and temporally in the fluid.

(a) Lumped element 'confinement'

The distributed element model of the transmission line is concerned with propagating waves of voltage and currents (equation (2.4)). The progress of these waves is governed by some forcing function (usually an impressed voltage, V), the load and the line parameters (R, G, C and L), and nothing else. Thus, this approach confines all interested activity to the internals of the conductors themselves and gives no explicit insight into the processes outside the domain of the wire. The nature of the governing electromagnetic fields is that of distributed quantities ($\mathbf{D}$, $\mathbf{B}$, $\mathbf{E}$, $\boldsymbol{J}$ and $\mathbf{H}$) in all space; the nature of circuit theory is, at best, a two-dimensional drawing with all processes confined to within the drawn lines. This is usually fine in the mathematical analysis of a transmission line problem, but does not give the great 'physical' insight that Heaviside sought, especially when it came to the 'flow' of energy. He sought to escape this confinement and pursued a line of research concerned with the electromagnetic energy densities in the free space surrounding the current-carrying conductors.

(b) The Poynting theorem and the 'energy current'

In 1884, Poynting first published his paper [31] that independently formulated an electromagnetic variant of the Umov energy flux vector using Maxwell's original quaternionic equations alongside Oliver Heaviside in his vector notation in mid-1884—this is sometimes termed the 'energy current' [32,33]. Hertz [34] attributed both Poynting and Heaviside as the originators of this concept; however, Heaviside stated that his 'transfer-of-energy' formula was a 'most general form' [35], but credited Poynting with the discovery of the fundamental electromagnetic concept. To get there, Maxwell's theory of electromagnetism 'localizes' the electromagnetic field energy densities. As this energy is localized, it must be able to 'flow' from one location to another. Before Poynting and Heaviside, this was not really understood, but its discovery was to provide a different physical viewpoint, opposed to the normal electric current (flow of charge), in which the process of energy transfer in situations where the transfer of electromagnetic energy can usefully be viewed in terms of a 'flow'. The so-called Heaviside–Poynting theorem is found in almost all textbooks on electromagnetism, expressed in modern notation as follows:

$$\iiint \underbrace{\mathbf{E}\cdot\boldsymbol{J}\,\mathrm{d}V}_{\text{electrical power dissipated}} + \overbrace{\iiint \left[\mathbf{E}\cdot\frac{\partial\mathbf{D}}{\partial t} + \mathbf{H}\cdot\frac{\partial\mathbf{B}}{\partial t}\right]\mathrm{d}V}^{\text{change in electromagnetic energy}} = \underbrace{-\oiint (\mathbf{E}\times\mathbf{H})\cdot\mathrm{d}\mathbf{S}}_{\text{power flowing out of closed surface}}. \tag{4.1}$$

Equation (4.1) is an energy-conservation law derived directly from Maxwell's equations. The new feature here is the vector on the RHS, inside the surface integral $\oiint (\mathbf{E}\times\mathbf{H})\cdot\mathrm{d}\mathbf{S}$ containing the cross-product of $\mathbf{E}\times\mathbf{H}$, the electrical and magnetic fields. The resultant vector must therefore be perpendicular to both of these fields—it describes physically the energy transfer due to time-varying electric and magnetic fields, directed perpendicular to the fields; in other words, the directional energy flux (the energy transfer per unit area per unit time) of an electromagnetic field. The derivation and physical interpretation is a stroke of genius, which we must attribute to Poynting and Heaviside. The terms on the LHS of equation (4.1) represent the power dissipated and the change in electromagnetic energy. Heaviside used the symbol $\mathbf{W}=\mathbf{E}\times\mathbf{H}$ in his work and he describes it with the now informal term 'energy current'. It must be noted that the units of $\mathbf{W}$ are watt per square metre, $\mathrm{W\,m^{-2}}$, and by dividing through by c^2 the units transform to momentum density. This is an 'electromagnetic momentum', associated with the electromagnetic fields.

It is well known that this approach is useful in antenna theory and microwave circuits. However, Heaviside extended the use of the theorem to DC electric circuits. In doing so, he

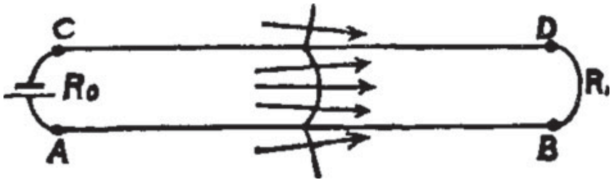


Figure 2. Energy flow in a simple circuit, ca. 1892 [33].

reversed the contemporary view of electric current, proposing that the electric and magnetic fields due to the current are the primitives, rather than being a result of the motion of the electronic charge in the conductor. This is a controversial viewpoint and, in his *Electrical papers*, the phrase 'we reverse this' [36], referring to the 'current in the wire being set up by the energy transmitted through the medium around it', reverberates even to this day. This view is supported by his work on electromagnetic diffusion (previous section) and the nature of the electromagnetic field and current density penetration of an electrical conductor subjected to a step current. This is discussed in a modern context by Feynman [37], who showed that the electromagnetic momentum is 'required' in order to conserve angular mechanical momentum associated with the energy flux vector **W**, and a detailed historical discussion is presented by Nahin [12]. The 'uniqueness' of the vector **W** and the physical existence of mysterious and counterintuitive circulating 'energy currents' emerging from static fields (e.g. a point electric charge with a superimposed magnetic dipole) has been the subject of some debate among many scientists over a long period of time. Heaviside was the first to consider these issues in 1893 [38]. The reader is directed to [12,37] for a detailed analysis and discussion on these matters.

The concepts presented here, as derived by Heaviside, complete his physical picture. From wave to diffusion phenomena, they are processes involving energy, and, as such, the Heaviside–Poynting theorem can be used to develop a complete understanding of an electric circuit to escape the 'lumped element confinement'.

(c) A simple circuit?

One problem that Heaviside used to illustrate the energy current concept was that of a battery connected by a simple circuit to a resistor load in his 1886–1887 work 'The transfer of energy and its application to wires. Energy current' [33]—effectively an analysis of a 'twin wire transmission line'. The standard approach is to assume that electrical energy flows inside metal wires (confinement), but Heaviside's energy current approach dictates otherwise. Poynting had first published on this arrangement [31], which was criticized by Heaviside due to Poynting's misconception of the nature of the external electric field surrounding the wires [32,34]. Poynting considered only a tangential electric field in the axis of the wire, combined with the circumferential magnetic field, resulting in an inwardly radial component of the energy flux density, **W**, only. Thus, given no energy flux in the direction of the electric current—how does energy get from the battery to the load? This is not answered by conventional circuit theory. Heaviside argued that this 'Poynting component' is simply the heat lost due to Joule heating in the conductor; however, a more prominent component exists outside the wire due to a radial electric field—the surface charges on the conductors that set up the field and maintain the electric current are responsible for the energy transfer external to the conductors. This complete field picture then presents a 'map' of the energy flow. Heaviside showed for the first time that a radial electric field and a circumferential magnetic field produce an 'energy current', a flow of electromagnetic energy in the space surrounding the electric conductors, directed from the battery along the axis of the conductor towards and entering the load. Figure 2 shows a diagram from Heaviside's work, illustrating the energy flow from a battery (at A–C) to a resistor (at B–D) guided by conductors C–D and A–B.

rsta.royalsocietypublishing.org Phil. Trans. R. Soc. A 376: 20170457

This remarkable result is one that is rarely presented, but that Heaviside gives in great detail in his *Electrical papers*, vol. II, which have recently been the subject of a rigorous modern mathematical treatment [39], confirming Heaviside's results. It is readily shown that the energy current approach is compatible with the circuit theory approach by applying the integral formulation of the Poynting theorem to the problem, determining the power dissipated in the load resistor as VI, as dictated by conventional 'confined' circuit theory. This 'energy current approach' gives physical insight, but requires detailed knowledge of the electric and magnetic fields surrounding the analysed circuits. It complements his work on electromagnetic wave propagation by analysing the energy associated with the wave and also his work on electromagnetic diffusion in which he found that the current in the wire penetrates from the outside surface inwards.

5. Conclusion

This paper has presented some selected contributions of Oliver Heaviside to electromagnetic theory, specifically in relation to transmission line analysis from the wave propagation, electromagnetic diffusion and energy flow perspectives. The scope of Heaviside's work is vast, and is contained within his two volumes of *Electrical papers* and his three volumes on electromagnetic theory. Heaviside radically changed how engineers and physicists treated transmission line problems and gave practical insights into how the operation of high-frequency landline and subsea telegraphic cables should be analysed, designed and operated. He went a step further, moving from the circuit theory domain into the full electromagnetic field theory description of such lines, providing advances in the understanding of the granular electromagnetic detail including the theoretical establishment of transient and persistent skin effects, as well as proximity-related phenomena. His contributions of the physical description of energy flow were a major achievement, extending physical insights into not only high-frequency radiofrequency and microwave applications but also into understanding DC circuits and the 'transference of energy', localized by Maxwell's theory. A great deal of Heaviside's electrical engineering science underpins modern advances, which continue to this day. Modern applications and theories for THz wave propagation and electric power transient analysis, to name but a few, owe much to Heaviside's nineteenth century contributions to electromagnetic theory. However, Heaviside's work is as relevant today as it was in the Victorian period. It has underpinned a vast expanse of modern theory and has application well beyond what he considered at the time or that he could have predicted into the future. Presented here, very briefly, are a few modern examples that can be argued to have evolved from Heaviside's work in electromagnetic theory.

(a) THz wave propagation

THz waves (1–10 THz) are electromagnetic waves of recent research interest. Many modern applications present themselves, such as medical and scientific imaging, process monitoring and quality control in manufacturing. Heaviside's distributed element model and his mathematical analysis of high-frequency lines have paved the way for analysis in this frequency range [40]. The analytical tools used are fundamentally based on his work intended for solving nineteenth century telegraphic problems, with appropriate modifications to include additional physics at such frequencies.

(b) Transient skin effects

In modern electrical power systems, much research is underway regarding DC systems, rather than the traditional AC approach. In the AC systems, the persistent skin effect is taken into account in the design of cables to operate at power frequencies. With modern power systems being ever more 'power electronic controlled', fast switching devices inherently introduce transient electromagnetic effects, and, as such, a transient skin effect can be present. This follows

Heaviside's work on electromagnetic diffusion and his generalized viewpoint on the 'skin effect'. Fault conditions in both AC and DC power systems also are concerned with transient electromagnetic phenomena—the fundamental work having been put forth by Heaviside.

(c) Power flow in electrical machines

The 'energy current' or the Poynting vector approach to antenna design is well known. It has been shown by Donaghy-Spargo [41], and others, that this approach complements the traditional approaches (Faraday and Lorentz) in understanding electrical machines (motors and generators). Lorentz forces are usually calculated to determine the torque produced in a rotary electrical machine; these forces are located at either side of an air gap separating the stationary component from the rotor. These are confined to the respective mechanical components—what happens in the air gap between and how does power flow from one member to another? Application of Heaviside's approach illuminates a radial flux of energy across the air gap when the motor is loaded; a circulating component wholly tangential to the cylindrical rotor surface exists when the motor is unloaded. This is a unique physical insight into the processes involved in the electromechanical conversion process.

Data accessibility. This article has no additional data.

Competing interests. I declare I have no competing interests.

Funding. I received no funding for this study.

Acknowledgements. CD-S acknowledges the very many fruitful conversations with Prof. Alex Yakovlev of the School of Engineering at Newcastle University regarding the work and life of Oliver Heaviside. The 'Introduction' section of this paper is adapted from the following presentation and unpublished paper: C.M. Spargo and A. Yakovlev, 'Oliver Heaviside FRS: Newcastle upon Tyne 1868–1874', prepared for the IET History of Technology Network Conference: The History of Power and Engineering, in Newcastle upon Tyne, 2015.

References

1. Institution of Electrical Engineers. 1950 *The Heaviside Centenary Volume*. London, UK: Institution of Electrical Engineers.
2. Heaviside O. 1880 *Electrical conductors &c.* UK Patent no. 1407.
3. Arthur JW. 2008 The fundamentals of electromagnetic theory revisited. *IEEE Antennas Propag. Mag.* **50**, 19–65. (doi:10.1109/MAP.2008.4494503)
4. Maxwell JC. 1973 *A treatise on electricity and magnetism*, vols. I & II. London, UK: Clarendon Press.
5. Bethenode J. 1925 Oliver Heaviside [obituary]. *Annales des Posies Telegraphs* **14**, 521–538.
6. Heaviside O. 1889 On the electromagnetic effects due to the motion of electrification through a dielectric. *Phil. Mag. J. Sci.* **27**, 324–339. (doi:10.1080/14786448908628362)
7. Searle GFC. 1987 *Oliver Heaviside: the man* (ed. I Catt). St Albans, UK: Westfields Press.
8. Heaviside O. 1888 *Electromagnetic waves, the propagation of potential, and the electromagnetic effects of a moving charge*. London, UK: The Electrician.
9. Heaviside O. 1892 *Electrical papers*, vol. I, pp. 440–441. London, UK: Macmillan and Co.
10. Lee G. 1947 *Oliver Heaviside and the mathematical theory of electrical communications*. London, UK: Longmans, Green and Co.
11. Mahon B. 2009 *Oliver Heaviside: Maverick mastermind of electricity*. London, UK: IET.
12. Nahin PJ. 1988 *Oliver Heaviside: sage in solitude: the life, work, and times of an electrical genius of the Victorian age*. New York, NY: IEEE Press.
13. Griffiths DJ. 2014 *Introduction to classical electrodynamics*, 4th edn. Cambridge, UK: Cambridge University Press.
14. Heaviside O. 1893 Vectors versus quaternions. *Nature* **47**, 533–534. (doi:10.1038/047533c0)
15. Heaviside O. 1892 *Electrical papers*, vol. II, p. 379. London: Macmillan and Co.
16. Thomson W. 1854 On the electro-statical capacity of a Leyden phial and of a telegraph wire insulated in the axis of a cylindrical conducting sheath. *Philos. Mag.* **9**, 531–555. (doi:10.1080/14786445508641914)

17. Heaviside O. 1892 *Electrical papers*, vol. II, p. 81. London, UK: Macmillan and Co.
18. Heaviside O. 1892 *Electrical papers*, vol. II, p. 53. London, UK: Macmillan and Co.
19. Heaviside O. 1892 *Electrical papers*, vol. II, p. 82. London, UK: Macmillan and Co.
20. Heaviside O. 1892 *Electrical papers*, vol. II, pp. 137–141. London, UK: Macmillan and Co.
21. Heaviside O. 1892 *Electrical papers*, vol. II, p. 39. London, UK: Macmillan and Co.
22. Heaviside O. 1893 Electromagnetic theory. *The Electrician* **I**, 112.
23. Heaviside O. 1892 *Electrical papers*, vol. I, pp. 139–140. London, UK: Macmillan and Co.
24. Heaviside O. 1892 *Electrical papers*, vol. II, pp. 53–54. London, UK: Macmillan and Co.
25. Heaviside O. 1892 *Electrical papers*, vol. II, p. 44. London, UK: Macmillan and Co.
26. Heaviside O. 1892 *Electrical papers*, vol. I, pp. 361–365. London, UK: Macmillan and Co.
27. Warren AG. 1946 *Mathematics applied to electrical engineering*. London, UK: Chapman & Hall.
28. Lamb H. 1883 On electrical motions in a spherical conductor. *Phil. Trans. R. Soc. Lond.* **174**, 519–549. (doi:10.1098/rstl.1883.0013)
29. Heaviside O. 1892 *Electrical papers*, vol. II, pp. 39–155. London, UK: Macmillan and Co.
30. Umov NA. 1874 Ein Theorem über die Wechselwirkungen in Endlichen Entfernungen. *Z. Math. Phys.* **19**, 97–114.
31. Poynting JH. 1884 On the transfer of energy in the electromagnetic field. *Phil. Trans. R. Soc. Lond.* **175**, 343–361. (doi:10.1098/rstl.1884.0016)
32. Heaviside O. 1892 *Electrical papers*, vol. I, p. 449. London, UK: Macmillan and Co.
33. Heaviside O. 1892 *Electrical papers*, vol. II, pp. 91–95. London, UK: Macmillan and Co.
34. Hertz H. 1893 *Electric waves: being researches on the propagation of electric action with finite velocity through space*. New York, NY: Dover Publications.
35. Heaviside O. 1886 On the self-induction of wires. *Philos. Mag.* **22**, 118–138. (doi:10.1080/14786448608627908)
36. Heaviside O. 1892 *Electrical papers*, vol. I, p. 434. London, UK: Macmillan and Co.
37. Feynman RP. 1964 *Lectures on physics*, vol. II. Reading, MA: Addison–Wesley.
38. Heaviside O. 1893 Electromagnetic theory. *The Electrician* **I**, 247.
39. Matar M, Welti R. 2017 Surface charges and J H Poynting's disquisitions on energy transfer in electrical circuits. *Eur. J. Phys.* **38**, 065201. (doi:10.1088/1361-6404/aa7d0f)
40. Mousavi SH, Williamson IA, Wang Z. 2016 Kinetic inductance driven nanoscale 2D and 3D THz transmission lines. *Sci. Rep.* **6**, 25303. (doi:10.1038/srep25303)
41. Donaghy-Spargo CM. 2017 Rotating electrical machines: Poynting flow. *Eur. J. Phys.* **38**, 055204. (doi:10.1088/1361-6404/aa7dcc)

rsta.royalsocietypublishing.org *Phil. Trans. R. Soc. A* **376**: 20170457

PHILOSOPHICAL TRANSACTIONS A

rsta.royalsocietypublishing.org

Energy velocity and reactive fields

Hans G. Schantz

Q-Track Corporation, 2223 Drake Avenue SW 1st Floor, Huntsville, AL 35805, USA

HGS, 0000-0002-7586-7341

Research

Check for updates

Cite this article: Schantz HG. 2018 Energy velocity and reactive fields. *Phil. Trans. R. Soc. A* **376**: 20170453.
http://dx.doi.org/10.1098/rsta.2017.0453

Accepted: 17 July 2018

One contribution of 13 to a theme issue 'Celebrating 125 years of Oliver Heaviside's 'Electromagnetic Theory''.

Subject Areas:
electrical engineering

Keywords:
energy velocity, reactive field, Lagrangian, Heaviside, near field

Author for correspondence:
Hans G. Schantz
e-mail: h.schantz@q-track.com

Conventional definitions of 'near fields' set bounds that describe where near fields may be found. These definitions tell us nothing about what near fields are, why they exist or how they work. In 1893, Heaviside derived the electromagnetic energy velocity for plane waves. Subsequent work demonstrated that although energy moves in synchronicity with radiated electromagnetic fields at the speed of light, in reactive fields the energy velocity slows down, converging to zero in the case of static fields. Combining Heaviside's energy velocity relation with the field Lagrangian yields a simple parametrization for the reactivity of electromagnetic fields that provides profound insights to the behaviour of electromagnetic systems. Fields guide energy. As waves interfere, they guide energy along paths that may be substantially different from the trajectories of the waves themselves. The results of this paper not only resolve the long-standing paradox of runaway acceleration from radiation reaction, but also make clear that pilot wave theory is the natural and logical consequence of the need for quantum mechanics correspond to the macroscopic results of the classical electromagnetic theory.

This article is part of the theme issue 'Celebrating 125 years of Oliver Heaviside's 'Electromagnetic Theory''.

1. Introduction

Ask a half dozen electromagnetic practitioners to define 'near field' and you will likely obtain a dozen answers [1]. For an electrically small dipole source, electrical engineer Harold Wheeler (1903–1996) identified that the reactive field terms dominate within a radius $r = \lambda/2\pi$; the radiation terms dominate beyond. Wheeler

THE ROYAL SOCIETY PUBLISHING

dubbed this the 'radiansphere' [2,3]. Electromagnetic compatibility engineers take five times this distance ($r = 5\lambda/2\pi$) as the limit at which the field impedance converges to 376.7 Ω, the impedance of free space. For larger sources, such as an antenna of aperture size 'D', the radiated fields approximate far-field plane waves at $r = D^2/\lambda$ to eighth of a wavelength precision and at $r = 2D^2/\lambda$ to a precision of a 16th of a wavelength. Traditional definitions of 'near fields' tell where to find them, not what they are or how they work.

In 1884–1885, John Henry Poynting (1852–1914) [4] and Oliver Heaviside (1850–1925) [5] independently discovered the relation governing how electromagnetic energy moves from place to place. The 'Poynting vector', or directed power flow per unit area is given by $\mathbf{S} = \mathbf{E} \times \mathbf{H}$, the cross product of the electric ($\mathbf{E}$) and magnetic ($\mathbf{H}$) fields. Just a few years later, Heinrich Hertz (1857–1894) caught electromagnetic energy in the act of moving from place to place in a series of simple and striking experiments [6]. Hertz demonstrated that electromagnetic waves propagate at the speed of light, proving the suspected link between electromagnetics and optics.

Heaviside derived a 'wave velocity' under the assumption that the energy propagates along with the wave [7]. His wave velocity or energy velocity, $\mathbf{v}_u$, is the ratio of the Poynting vector to the energy density, u:

$$\mathbf{v}_u = \frac{\mathbf{S}}{u} = \frac{\mathbf{E} \times \mathbf{H}}{\frac{1}{2}\varepsilon_0 |\mathbf{E}|^2 + \frac{1}{2}\mu_0 |\mathbf{H}|^2}, \tag{1.1}$$

where $\varepsilon_0 = 8.85 \times 10^{-12}\ \mathrm{F\,m^{-1}}$ is the permittivity, and $\mu_0 = 4\pi \times 10^{-7}\ \mathrm{H\,m^{-1}}$ is the permeability of free space. Kirk McDonald provides an excellent discussion of electromagnetic momentum and velocity in a DC circuit [8].

Hertz's experiments dealt with standing waves in which the fields propagate back and forth at the speed of light, but the average energy velocity is zero. The fact that electromagnetic energy velocity may, in general, be less than c appears to have first been explicitly noted by Harry Bateman (1882–1946) in 1915 [9]. Recently, Gerald Kaiser re-examined this question, interpreting energy velocity as a local time-dependent characteristic of electromagnetic fields [10].

Clearly, the ratio of electric to magnetic energy relates to energy velocity. The Poynting vector goes to zero if either field goes to zero. A concentration of electric energy with no magnetic energy will be static, as will a concentration of magnetic energy with no electric energy. One form of energy must transform, at least in part, to the other in order for energy to move. Only when both forms of energy coexist at the same location can there be a progression or motion of energy. This observation, due to Oliver Lodge (1851–1940), aids in understanding electromagnetic energy velocity [11]. The remarkable properties of the electric to magnetic field intensity ratio were not appreciated at the time, however.

Only in the 1930s did Sergei Schelkunoff (1897–1992) identify the significance of the 'field impedance' in applied electromagnetics [12]. Schelkunoff extended Heaviside's impedance concept from AC electronics to electromagnetic fields. The impedance of the fields, Z, is the ratio of the electric to magnetic field intensity, $Z = |\mathbf{E}|/|\mathbf{H}|$. Schelkunoff's ground-breaking work identified the distinction between the characteristic impedance of a medium—the natural ratio assumed by propagating fields, and the actual, instantaneous impedance which may be radically different under interference between multiple waves.

The final ingredient is the difference between the electric and magnetic energy densities at a particular place and time, the Lagrangian, $L = 1/2\ \varepsilon_0\ |\mathbf{E}|^2 - 1/2\ \mu_0\ |\mathbf{H}|^2$. Heaviside was sceptical of the Lagrangian or Least Action approach [13]:

> Now at Cambridge, or somewhere else, there is a golden or brazen idol called the Principle of Least Action. Its locality is kept secret, but numerous copies have been made and distributed amongst the mathematical tutors and lecturers at Cambridge, who make the young men fall down and worship the idol.

James Clerk Maxwell (1831–1879) also expressed concern that the abstract nature of Lagrangian formalism severed the link to the physical system under consideration [14]:

> . . . Lagrange and most of his followers, to whom we are indebted for these methods, have in general confined themselves to a demonstration of them, and, in order to devote their attention to the symbols before them, they have endeavoured to banish all ideas except those of pure quantity, so as not only to dispense with diagrams, but even to get rid of the ideas for velocity, momentum, and energy, after they have been once for all supplanted by symbols in the original equations.

This paper offers a definition for 'near' or 'reactive' fields by building upon Heaviside's definition of energy velocity using impedance and employing the electromagnetic Lagrangian.

2. Theory

An isolated electromagnetic wave in free space exhibits equal amounts of electric and magnetic energy. Thus, the Lagrangian is zero ($L = 1/2\,\varepsilon_0\,|\mathbf{E}|^2 - 1/2\,\mu_0\,|\mathbf{H}|^2 = 0$). Equivalently, the ratio of electric to magnetic intensity is given by the free space impedance, $Z_s = \sqrt{\mu_0/\varepsilon_0} = 376.7\Omega$. Under these circumstances, the wave and its energy propagate in synchronicity at the speed of light. If we disrupt the balance of electric to magnetic energy, some of the energy slows down.

This phenomenon is implicit in Heaviside's theory of distortionless propagation along a transmission line. With a balance of inductance (L) and capacitance (C), a wave propagates along a transmission line without distortion. In Heaviside's theory, distortionless propagation requires the impedance of a transmission line must be

$$Z = \sqrt{\frac{L}{C}} = \sqrt{\frac{R}{G}}, \tag{2.1}$$

where R is the resistance and G is the conductance, and all quantities are taken per unit length of the transmission line [15]. Alas, the number of physical quantities exceeds that of the number of symbols to express them, so the reader will have to distinguish between Lagrangian and inductance and between resistance and radial distance by context.

If an imbalance exists between electric and magnetic energy, some of the energy slows down and the wave becomes distorted. William Suddards Franklin (1863–1930) explained the distinction between 'pure' and 'impure' waves in 1909 [16]. Franklin noted that a pure wave propagates without distortion. When waves superimpose or interfere, the resulting 'impure' wave exhibits distortion—a change in shape from the original waves due to their interaction.

First, consider the field impedance, normalized to the impedance of free space so as to yield a dimensionless quantity

$$z = \frac{Z}{Z_S} = \frac{E/H}{\sqrt{\mu_0/\varepsilon_0}} = \frac{\sqrt{\varepsilon_0}E}{\sqrt{\mu_0}H}. \tag{2.2}$$

This normalized impedance comes in handy for analysing energy velocity. When $z = 1$, the electric and magnetic energy are in balance, and the energy propagates at the speed of light ($c = 2.9979 \times 10^{+8}\ \mathrm{m\,s^{-1}}$). We may normalize the energy velocity with respect to the speed of light to obtain a similarly dimensionless vector [17]:

$$\begin{aligned}\boldsymbol{\gamma} = \frac{\mathbf{v}}{c} = \frac{\mathbf{S}}{cu} &= \frac{\mathbf{E} \times \mathbf{H}}{c\left(\frac{1}{2}\varepsilon_0|\mathbf{E}|^2 + \frac{1}{2}\mu_0|\mathbf{H}|^2\right)} = \frac{2(E\hat{\mathbf{e}} \times H\hat{\mathbf{h}})}{\sqrt{\varepsilon_0/\mu_0}|E|^2 + \sqrt{\mu_0/\varepsilon_0}|H|^2} \\ &= \frac{2E/H(\hat{\mathbf{e}} \times \hat{\mathbf{h}})}{(1/Z_S)(E/H)^2 + Z_S} = \frac{2z}{(1+z^2)}(\hat{\mathbf{e}} \times \hat{\mathbf{h}}).\end{aligned} \tag{2.3}$$

This result is similar to the transmission coefficient for a discontinuity in the impedance of a medium and describes how electromagnetic energy slows down when the electric and magnetic fields are no longer balanced, i.e. when $z \neq 1$.

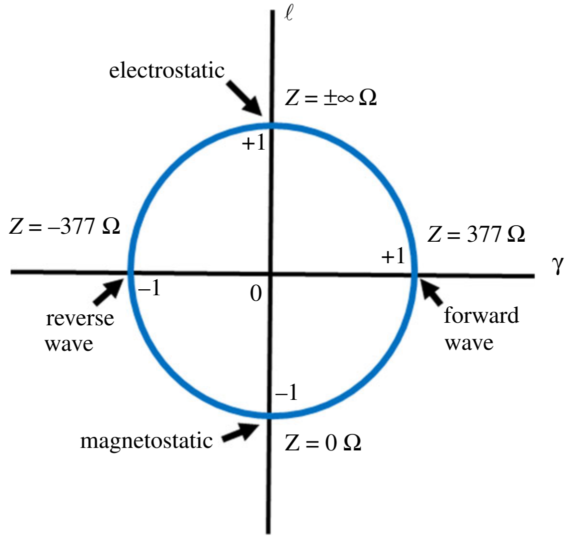


Figure 1. The ebb and flow of electromagnetic energy in a given direction is captured by the Great Electromagnetic Circle comprising the relationship between energy velocity and Lagrangian density. All one-dimensional electromagnetic states lie somewhere on the locus of the circle. (Online version in colour.)

Second, consider the Lagrangian, as normalized by the Hamiltonian (H, or equivalently the total energy) [18]:

$$\ell = \frac{L}{H} = \frac{\frac{1}{2}\varepsilon_0|\mathbf{E}|^2 - \frac{1}{2}\mu_0|\mathbf{H}|^2}{\frac{1}{2}\varepsilon_0|\mathbf{E}|^2 + \frac{1}{2}\mu_0|\mathbf{H}|^2} = \frac{z^2-1}{z^2+1}. \tag{2.4}$$

Schelkunoff encouraged us to simplify electromagnetic analysis by considering one-dimensional propagation, as with transmission lines or plane waves [12]. This allows us to capture much of the essential physics of electromagnetic phenomena while avoiding the complexity of a full three-dimensional treatment. We may drop the vector dependence on the normalized energy velocity and identify the circular relationship between the normalized Lagrangian and the normalized energy velocity:

$$\ell^2 + \gamma^2 = \frac{(z^2-1)^2}{(z^2+1)^2} + \frac{(2z)^2}{(z^2+1)^2} = \frac{(z^2+1)^2}{(z^2+1)^2} = 1. \tag{2.5}$$

The normalized Lagrangian allows for a simple characterization of near or reactive fields. In the limit $\ell \to 0$, we have a pure, far-field electromagnetic wave. In the limit $\ell \to +1$, the fields are electrostatic, and in the limit $\ell \to -1$, the fields are magnetostatic. Figure 1 shows what has been dubbed the Great Electromagnetic Circle of normalized Lagrangian versus energy velocity [19]. Figure 1 intuitively relates the energy velocity to the normalized Lagrangian, allowing a precise parametrization of intermediate 'near-field' states. No longer must we characterize near fields by proximity to a source. Using the normalized Lagrangian, we can parametrize near fields along the continuum from electrostatic fields, through radiation fields, to magnetostatic fields and through reverse propagating radiation fields, back again to electrostatic fields. This result makes clear the close connexion between the reactive fields found near a radiating source, and the reactive fields found when one electromagnetic wave interferes with another.

For simplicity, consider two waves of equal, unit amplitude interfering with each other as in figure 2. We define interference with respect to the electric field. In a destructive interference, the electric field cancels out, leaving only a magnetic field. In a constructive interference, the electric field doubles and the magnetic field cancels out. Energy is proportional to the square of the field intensity. In the constructive interference case, the initial unit electric energy on each side combines to yield four times the electric energy; in the destructive interference case, the initial

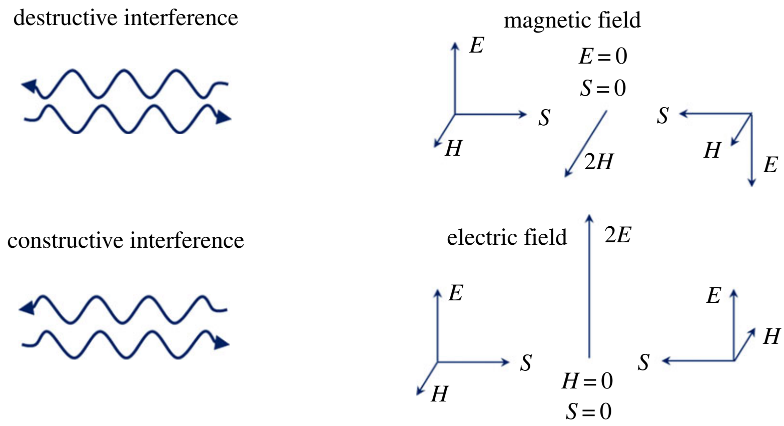


Figure 2. In a destructive interference (top left), the electric field cancels, and the magnetic field reinforces (top right). In a constructive interference (bottom left), the electric field reinforces and the magnetic field cancels (bottom right). (Online version in colour.)

unit electric energy on each side combines to cancel out entirely. This apparently counterintuitive result follows because conservation applies to the total energy, not the electric or magnetic energy individually. The excess electric energy in constructive interference arises because the magnetic field and associated energy are zero. All the magnetic energy in the original balance has become electric energy. The 'missing' electric energy in destructive interference has become magnetic energy due to the concomitant interference of the magnetic field component [20].

Total energy is conserved. When we diminish one component of the electromagnetic field, we increase the other in offset to maintain a fixed amount of total energy. Further, when either component of the field goes to zero, there is no power flow and the remaining field is momentarily static. In an isolated wave, the energy moves at a constant velocity at the speed of light in spatially distributed impulses. For a sine wave or harmonic wave, the impulses occur every half wavelength. When two waves interfere, however, the energy actually stops and changes direction. The following section provides a variety of examples.

3. Applications

The previous section demonstrated how we may combine Heaviside's energy velocity and the field Lagrangian to yield a simple, elegant framework within which to understand near or reactive fields and energy flow. As electromagnetic waves interact and interfere, the energy they convey slows down and may even come to a rest. Counter-propagating waves may exchange energy with each other, temporarily storing energy in the resulting quasi-static fields. There is a fundamental correspondence between the near fields present around a small electromagnetic source and the reactive fields that form as two waves interfere in free space. This section examines that similarity by considering three examples: the fields and energy flow around an exponentially decaying dipole, around an accelerating charge, and the fields and energy flow around two interfering waves.

(a) The exponentially decaying dipole

The usual analysis of a small or 'Hertzian' dipole with moment p_0 assumes harmonic oscillation. Each successive oscillation of the dipole results in an outward propagating radiation wave that strips off some of the energy presents in the near fields and conveys that energy away in radiation. The process—first identified by Hertz—is complicated, and has been described by the author elsewhere [21]. This section will examine a simpler case of radiation—that of an exponentially decaying dipole, to illustrate some of the fundamental physics underlying electromagnetic waves and energy.

An exponential decay time dependence of the form $T(t) = \exp[-t/(c\tau)]$ arises in the discharge of a small capacitor. The time constant, $\tau = RC$, depends on the product of associated resistance (R) and capacitance (C). The fascinating details of the fields and energy flow were originally noted by Mandel [22] and examined elsewhere in more detail [23]. Analytically simple cases like these serve as good test cases for validating finite difference time domain algorithms. Assuming the dipole moment aligns with the z-axis, the fields are

$$\mathbf{H} = \frac{p_o}{4\pi r} e^{-(t-rc)/\tau} \left(-\frac{1}{\tau r} + \frac{1}{\tau^2 c} \right) \sin\theta \hat{\varphi} \tag{3.1}$$

and

$$\mathbf{E} = \frac{p_o}{4\pi\varepsilon_0} e^{-(t-rc)/\tau} \left(\left(\frac{1}{r^3} - \frac{1}{\tau c r^2} \right) (2\cos\theta \hat{\mathbf{r}} + \sin\theta \hat{\theta}) + \frac{\sin\theta}{\tau^2 c^2 r} \hat{\theta} \right), \tag{3.2}$$

where r is the radial distance from the dipole in spherical coordinates. Unlike the harmonic dipole case, all field terms in an exponential dipole decay share the same time dependence. Thus, the impedance, the normalized energy velocity, and the Lagrangian are constant with respect to time; they depend only upon radial distance and angle. The radial impedance for an exponentially decaying electric dipole evaluated along the equatorial ($\theta = \pi/2$) plane is

$$Z_E = Z_S \frac{1 - r/\tau c + (r/\tau c)^2}{(r/\tau c)^2 - r/\tau c}. \tag{3.3}$$

Figure 3*a* plots the magnitude of the field impedance in units of free space impedance (Z_s) for an electric dipole as a function of range in units of $c\ \tau$. The impedance of the corresponding exponentially decaying magnetic dipole is the inverse of the electric case, and is offered for comparison. The electric dipole impedance exhibits a pole at $r = c\ \tau$ because the magnetic field goes to zero. Inside this spherical shell, the impedance is negative corresponding to the inward flow of energy. Outside this shell, the impedance is positive, corresponding to the outward flow of radiation energy. Figure 3*b* presents a qualitative diagram of this process. Figure 3*c* plots the normalized Lagrangian and energy velocity as a function of the radial distance. Figure 3*d* shows the trajectory of the process on the Great Electromagnetic Circle of normalized Lagrangian versus energy velocity.

Within the spherical region of radius $r = c\ \tau$, the radiation fields from an exponentially decaying dipole ripple outward through the inflowing energy absorbed by the dipole at the origin. Only after the radiation fields propagate through the $r = c\ \tau$ boundary do they associate themselves with the energy originally stored in the electrostatic fields around the dipole but outside the boundary. Inside the boundary, inductive field components dominate the radiation fields and the net energy flow is inward, despite the outward propagating radiation fields. Outside the boundary, the radiation fields dominate and energy propagate outward. Fields guide energy, but the net energy flows with the dominant fields. One must consider the superposition of all fields to understand the energy flow, not look at a field component in isolation.

(b) Accelerating charge: is radiation 'kinky'?

Heaviside also turned his attention to understanding how fields and energy radiate from an accelerating charge (q), pioneering what might be termed the 'kinked field' model of radiation [24]. He imagined a charge suddenly jerked into motion at the speed of light. Defining the axis of a globe as the direction of the displacement, the resulting field lines lie along the longitudes and propagate outward at the speed of light, as in figure 4*a*. Heaviside's approach—as later refined and extended by Richard Feynman (1918–1988)—can be applied to yield the electric field of a

rsta.royalsocietypublishing.org Phil. Trans. R. Soc. A 376: 20170453

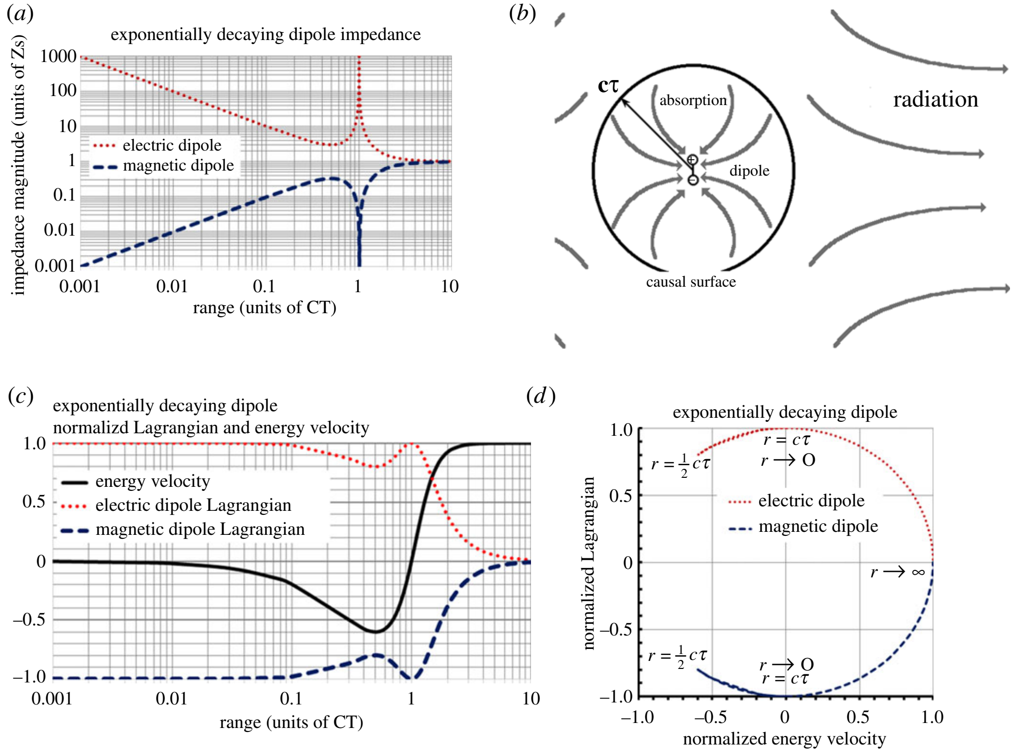


Figure 3. (*a*) Impedance of an exponentially decaying dipole as a function of radial range. (*b*) Diagram of energy flow around exponentially decaying dipole. (*c*) Normalized Lagrangian and energy velocity versus radial range. (*d*) Normalized Lagrangian versus normalized energy velocity. (Online version in colour.)

non-relativistically accelerated charge in a co-moving reference frame where velocity goes to zero ($v \to 0$) [25]:

$$\mathbf{E} = \frac{q}{4\pi\varepsilon_0}\left(\frac{\hat{\mathbf{r}}}{r^2} + \frac{\dot{v}\sin\theta}{c^2 r}\hat{\theta}\right). \tag{3.4}$$

Assume the acceleration is due to an applied external field:

$$\mathbf{E}_{\text{ext}} = E_0\hat{\mathbf{z}} = E_0(\cos\theta\hat{\mathbf{r}} - \sin\theta\hat{\theta}). \tag{3.5}$$

Note that the magnitude of the force on the charge of mass m is $F = ma = qE_0$. If we assume the accelerating charge is an electron, then the tangential fields go to zero on a spherical shell whose radius is the classical electron radius

$$r_e = \frac{q^2}{4\pi\varepsilon_0 mc^2}. \tag{3.6}$$

The corresponding magnetic field due to the accelerating charge is

$$H_\phi = \frac{q}{4\pi}\frac{\dot{v}\sin\theta}{cr}. \tag{3.7}$$

The radial impedance becomes

$$Z = \frac{E_\theta}{H_\phi} = \frac{q\dot{v}/4\pi\varepsilon_0 c^2 r - m\dot{v}/q}{\frac{q}{4\pi}\frac{\dot{v}}{cr}} = \frac{1}{\varepsilon_0 c} - \frac{4\pi crm}{q^2} = Z_S\left(1 - \frac{r}{r_\text{e}}\right). \tag{3.8}$$

Certain qualitative features of the energy flow, depicted in figure 4*b*, become obvious upon plotting the normalized Lagrangian and energy velocities as in figure 4*c*,*d*.

In the limit as r approaches zero, radial energy flow is directed out. At the classical electron radius, r_e, the transverse fields are magnetostatic and there is no radial energy flow. The radial

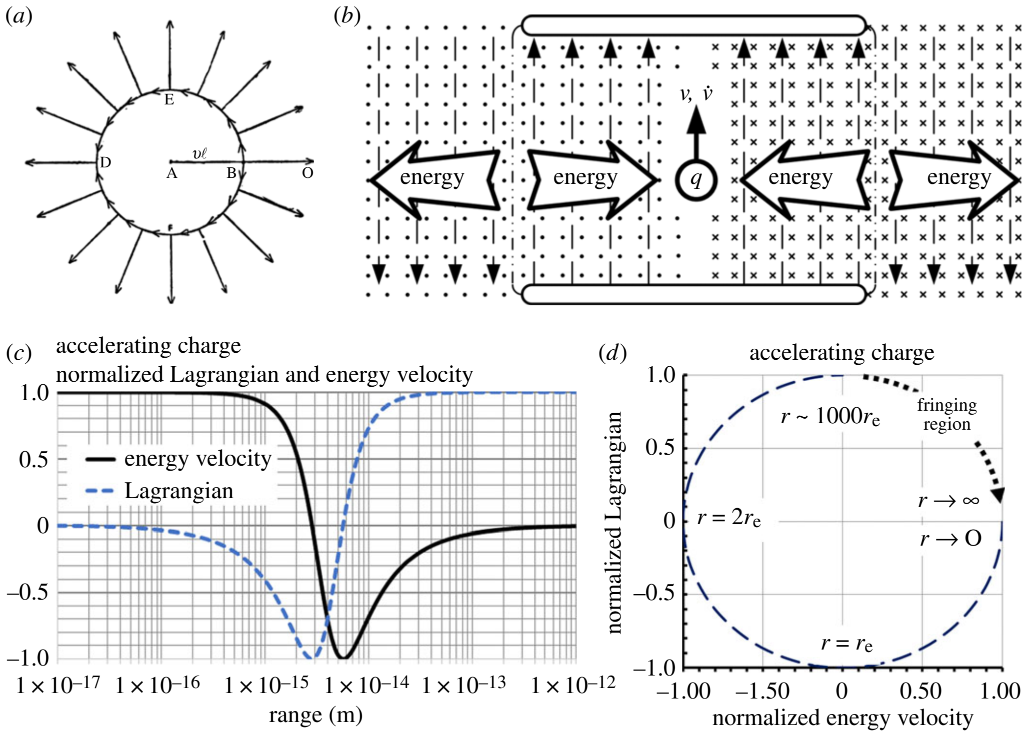


Figure 4. (*a*) Heaviside's diagram of transverse radiation fields near a suddenly accelerated charge. (*b*) Qualitative diagram of energy flow from an applied field to an accelerating charge. (*c*) Normalized energy velocity and Lagrangian for an accelerating charge (electron) versus range (r). (*d*) Normalized Lagrangian versus normalized energy velocity for an accelerating charge. (Online version in colour.)

energy flows in at the speed of light at twice the classical electron radius. The energy flow remains inward yet converges to zero as the electrostatic applied field becomes dominant at 'great' distances, which in this case is by about a thousand times the classical electron radius. Interestingly, an accelerating charge does not emit energy—it absorbs energy. The source of the energy radiated by an accelerating charge is the fringing fields at the boundary of the applied field. As the applied field decreases, the radiation field from the accelerating charge overwhelms the fringing field, causing the inward radial energy flow to reverse direction. In the fringing region, the Great Circle of the Lagrangian versus energy velocity closes. The normalized Lagrangian goes to zero and the normalized energy velocity converges to the speed of light.

The problem of how radiation energy reacts against accelerating charges has long defied the imagination of physicists [26]. Conventional theory appears to require paradoxical results like a non-physical 'runaway' or exponentially increasing acceleration. The present analysis suggests that the proper question is not 'how' radiation reacts against accelerating charges, but 'whether' it does. Radiated energy reacts against the applied field as it pushes off in the fringing region—not against the accelerating charge itself. The fringing region may be arbitrarily far away from the accelerating charge. There appears to be no reason to expect an action–reaction force balance between accelerating charges and their radiation, except in the overall context including the applied field and its fringing region.

The traditional view espoused by Heaviside that radiation is a transverse component or kink in an otherwise radial field leads to confusion because it presents only a partial picture. Here again, the lesson is that energy flow depends upon the big picture—the broader context including all the relevant fields present in a given context. The radiation fields from an accelerating charge do not convey energy from the accelerating charge to the far field. Rather they slightly decrease the

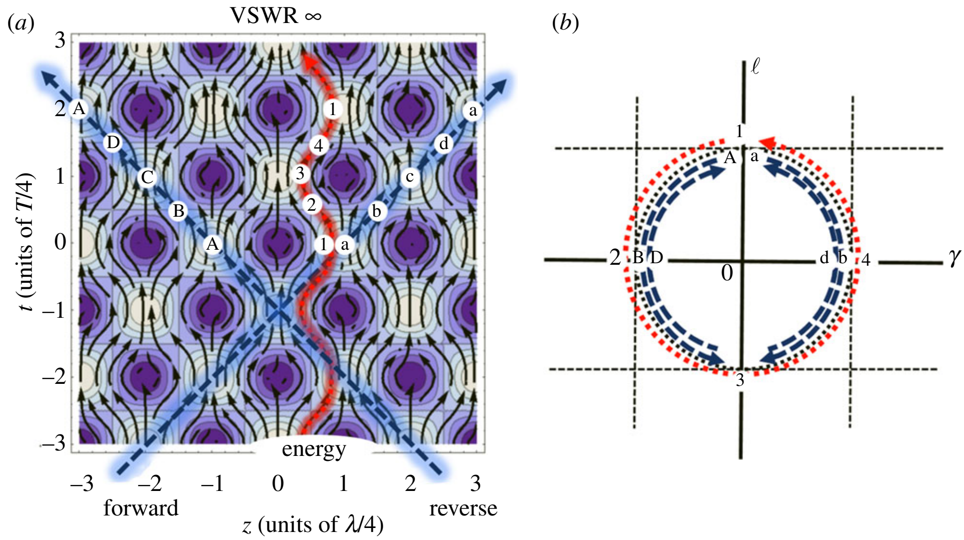


Figure 5. (*a*) Space–time diagram of the interference between identical counter-propagating harmonic waves. Shading shows energy density—light shows maximum energy, dark minimum. Arrows show space–time energy trajectories. (*b*) Behaviour in the Lagrangian-velocity space described by the great electromagnetic circle. (Online version in colour.)

dominant inward flow of energy from the applied field, and only collect energy from the fringing field region and convey it to the far field.

(c) Interacting waves

Consider two identical harmonic electromagnetic waves, one propagating forward and the other in the reverse direction along the z-axis. We may plot representative space–time trajectories for the waves and for their energy flow in the space–time energy flow diagram of figure 5*a*. The diagram is scaled in units of quarter wavelength for distance (z) and units of quarter period (T) for time (t), so the space–time trajectories of the waves, propagating at the speed of light, have unit slope. Shading shows the energy density as the two waves interfere constructively and destructively. The energy bounces back and forth between constructive interference electrostatic and destructive interference magnetostatic nodes spaced a quarter wavelength in distance and a quarter period in time. Gerald Kaiser appears to have been the first to examine the physics of this situation in detail [10].

We can also plot this behaviour in the Lagrangian-velocity space of the Great Circle of figure 5*b*. The Forward wave propagates through a region of electrostatic energy (a), conveys forward propagating electromagnetic energy (b), propagates through a region of magnetostatic energy (c), conveys forward propagating electromagnetic energy (d) and propagates through another region of electrostatic energy (a) as the Forward wave repeats the cycle. Similarly, the Reverse wave propagates through a region of electrostatic energy (a), conveys reverse propagating electromagnetic energy (b), propagates through a region of magnetostatic energy (c), conveys reverse propagating electromagnetic energy (d), and propagates through another region of electrostatic energy (a) as the Reverse wave repeats the cycle.

From the energy's space–time perspective, a trajectory may begin in an electrostatic node (1), move in the reverse direction under the influence of the reverse propagating wave (2), change direction in a magnetostatic node (3), move in the forward direction under the influence of the forward propagating wave (4) and change direction again in an electrostatic node (1) repeating the cycle. More general cases of interactions between waves have been examined elsewhere [27].

Note how decoupled the wave motion is with respect to the energy flow. Half the time, the waves convey energy in their respective directions of propagation while unopposed by the counter-propagating wave. The other half the time, the counter-propagating wave opposes the energy flow resulting in an electrostatic node for the case of constructive interference, or a magnetostatic node for the case of destructive interference. The waves propagate through each other at the speed of light, but every half wavelength or so, an identifiably different lump of energy is associated with a particular segment of the wave.

4. Conclusion

This paper applies ideas pioneered by Oliver Heaviside—energy flow and energy velocity—along with an idea Heaviside regarded and dismissed—the electromagnetic Lagrangian—to offer a novel way of looking at and understanding near or reactive fields. There is a 'Great Electromagnetic Circle' in a Lagrangian-energy velocity space that allows for an intuitive characterization of electromagnetic phenomena from electrostatic fields, through radiation or propagating fields, to magnetostatic fields, and back again. A deep kinship exists between the 'near' fields close to an electromagnetic source and the reactive fields present when two waves interfere with each other in free space. Fields and waves propagate through each other and in so doing they exchange energy. Fields and waves serve to guide the flow of energy along space–time trajectories that may be entirely different from the trajectory of any particular wave. These ideas are entirely consistent with and follow from the ideas of Maxwell and Heaviside, and offer a different perspective in considering practical problems from applied electromagnetics.

Conventional thinking assumes that energy remains tightly coupled to fields—that a particular electromagnetic wave and its energy are linked together as a wave propagates from place to place. To the contrary, the lesson of this paper is that fields and energy are only loosely coupled. As fields and waves interact, they exchange energy. Electromagnetic waves propagate at the speed of light—energy propagates at the speed of light only in a pure and isolated wave. In the presence of other electromagnetic fields or waves, the energy propagates at speeds less than that of light—even slows to a stop and changes direction in many near-field and interference situations.

This paper examines well-established principles in electromagnetic physics and offers a novel reinterpretation of their meaning with the potential to make clear how electromagnetics works. The concepts of this paper may resolve the long-standing paradox of radiation reaction, for instance. There is one final important consideration worthy of examination: the implications of these electromagnetic energy flow concepts for our understanding of quantum mechanics.

Paul A. M. Dirac (1902–1984) was one of the founding fathers of quantum mechanics (as Richard Feynman quipped, physicists may read Schiff, but they quote Dirac). In his classic quantum mechanics text, Dirac argued that photons do not interfere with each other [28]:

> Suppose we have a beam of light consisting of a large number of photons split up into two components of equal intensity. On the assumption that the intensity of this beam is connected with the probable number of photons in it, we should have half the total number of photons going into each component. If the two components are now made to interfere, we should require a photon in one component to be able to interfere with one in the other. Sometimes these two photons would have to annihilate one another, and other times they would have to produce four photons. This would contradict the law of conservation of energy. The new theory, which connects the wave function with probabilities for one photon, gets over the difficulty by making each photon go partly into each of the two components. Each photon then interferes only with itself. Interference between two different photons never occurs.

We know on the macroscopic level that waves interfere with each other, giving rise to regions in which the propagating electromagnetic energy transforms to electrostatic energy, becomes magnetostatic energy, or shifts elsewhere on the continuum, somewhere in between. The conventional wisdom, then, as stated by Dirac, is that although this energy is quantized, there can be no interactions or interference between the photons associated with the forward propagating wave and the photons associated with the reverse propagating wave. We get something—the macroscopically observed shifts in the energy balance—from nothing: the supposed non-interaction of the photons themselves. To quote Newton's comments on action-at-a-distance, this is 'so great an absurdity that I believe no man, who has in philosophical matters a competent faculty of thinking, can ever fall into it' [29]. Arguing from the quantum perspective, Roy J. Glauber calls Dirac's 'a highly simplistic remark which has sown confusion among physicists ever since' [30].

In the conventional wisdom, fields and energy appear tightly coupled. This paper demonstrates they are not. Fields exchange energy all the time. Electromagnetic waves and fields guide energy. Fields go one way, other fields go other ways, and the energy they convey goes in a completely different trajectory under the influence of the interacting fields. Waves and fields serve to guide electromagnetic energy along trajectories governed by the interactions and interference of the relevant fields. How would this classical theory naturally segue into understanding reality at the quantum level?

In the classical electromagnetic picture, an electromagnetic wave is an incident on the two slits. Each slit becomes a source for secondary waves that interfere with each other, generating an interference pattern. The flow of energy is not governed by the fields or wave from either slot, but rather by the interference pattern generated by the superposition of the two wave sources. A continuous flow of electromagnetic energy propagates along the streamlines set up by the interference pattern guided by the interaction of the fields. Now instead of a continuous flow of energy, imagine that the energy is quantized and rarefied to the point that only a single photon traverses the system at a time. By chance or happenstance, the photon will follow a particular energy flow streamline set up by the superposition of the fields. As in the classical case, no particular slot or particular wave source dictates the trajectory of the photon—only the interference resulting from the superposition of the fields.

Physicists confuse themselves speculating which slit the photon went through, assuming it had to go through both slits in order to generate the interference pattern and had to be in both places at the same time. Feynman argued the phenomenon of two-slit interference is 'impossible, absolutely impossible, to explain in any classical way, and which has in it the heart of quantum mechanics' [31]. He noted that all the mystery of quantum reality is wrapped up in this simple experiment. The mystery vanishes upon understanding the simple classical picture of this paper.

Fields behave like waves. They generate interference patterns between the two slits, throughout the region, 'non-locally', even when there is only enough energy for one photon to be moving through the slits. Energy behaves like particles, following a particular trajectory under the influence of the interacting waves. Each does its own thing. The quantum weirdness and confusion are gone. There is an interpretation of quantum mechanics that adopts a qualitatively similar perspective: the pilot wave theory of Louis de Broglie (1892–1987) and David Bohm (1917–1992). As de Broglie observed, '[a] photon coming from one laser or the other and arriving in the interference zone is guided – and this seems to us to be physically certain – by the superposition of the waves emitted by the two lasers . . . ' [32]. The results of this paper make clear that pilot wave theory is the natural and logical consequence of the need for quantum mechanics correspond to the macroscopic results of the classical electromagnetic theory.

Heaviside's and Poynting's work on energy flow led to a revolution in the understanding of how electric circuits convey energy. Rather than a continuous flow of energy through wires like water through a pipe, Heaviside and Poynting taught that wires sculpt adjacent fields which in turn convey energy through the space around the wires depositing it in the surface of the wire. The charge carriers themselves move slowly with a relatively short mean free path, while the fields propagate unimpeded at the speed of light along the wire.

An analogous worldview emerges from the present work. There is no need for photons to couple exclusively with any particular electromagnetic wave. Rather, they engage in a complex dance, pausing to form a momentary electrostatic or magnetostatic pirouette, before shifting direction and waltzing off with a different wave. In most realistic, complicated electromagnetic environments, photons have relatively short mean free paths and a drift velocity notably less than the speed of light.

In 1889, Heaviside observed, 'There are many more things in Maxwell which are not yet discovered' [33]. The same may be said of Heaviside's own work. When we apply Heaviside's insights to contemporary problems, we 'make patent what was merely latent' in the work and thought of the great man upon whose shoulders we stand.

Data accessibility. This article has no additional data.

Competing interests. The author declares that he has no competing interests.

Funding. The author received no funding for the present work.

Acknowledgements. The author dedicates this paper to the memory of his doctoral advisor, E.C.G. 'George' Sudarshan (1931–2018). In addition, the author would like to thank Kirk McDonald, Travis Norsen, Tasos Papaioannou, Greg Merchán, Jack Gardner and the peer reviewers for useful suggestions and feedback.

References

1. Capps C. 2001 Near field or far field?' EDN, pp. 95–101. See: https://www.edn.com/design/communications-networking/4340588/Near-field-or-far-field.-
2. Wheeler HA. 1947 Fundamental limitations of small antennas. *Proc. IRE* **35**, 1479–1484. (doi:10.1109/JRPROC.1947.226199)
3. Wheeler HA. 1959 The radiansphere around a small antenna. *Proc. IRE* **47**, 1325–1331. (doi:10.1109/JRPROC.1959.287198)
4. Poynting JH. 1884 On the transfer of energy in the electromagnetic field. *Phil. Trans. R. Soc.* **175**, 343–361. Collected in Collected Scientific Papers by John Henry Poynting, Cambridge at the University Press, 1920, pp. 175–193. (doi:10.1098/rstl.1884.0016)
5. Heaviside O. 1892 *Electrical papers*, vol. 1, pp. 429–556. London: Macmillan and Co. (Originally published in The Electrician, 1885–1886).
6. Hertz H. 1893 *Electric waves: being researches on the propagation of electric action with finite velocity through space*, (D.E. Jones, Trans.), New York: NY: Macmillan and Co.
7. Heaviside O. 1893 *Electromagnetic theory*, vol. 1, pp. 78–80. London, UK: The Electrician' Printing and Publishing Company. See http://bit.ly/uCSzEw.
8. McDonald KT. 2006 'Momentum in a DC Circuit,' May 2006, updated May 13, 2014. See: http://puhep1.princeton.edu/~mcdonald/examples/loop.pdf.
9. Bateman H. 1915 *The mathematical analysis of electrical and optical wave-motion On the basis of Maxwell's equations*, p. 6. Cambridge University Press. (http://bit.ly/1ykUfeE)
10. Kaiser G. 2011 Electromagnetic inertia, reactive energy and energy flow velocity. *J. Phys. A: Math. Theor.* **44**, 345206. (doi:10.1088/1751-8113/44/34/345206)
11. Lodge OJ. 1885 On the identity of energy: in connection with Mr. Poynting's paper on the transfer of energy in an electromagnetic field; and on the two fundamental forms of energy. *Philos. Mag.* **19**, 482–487. (doi:10.1080/14786448508627704)
12. Schelkunoff SA. 1938 The impedance concept and its application to problems of reflection, refraction, shielding and power absorption. *Bell System Tech. J.* **17**, 17–48. (doi:10.1002/j.1538-7305.1938.tb00774.x)
13. Heaviside O. 1912 *Electromagnetic theory*, vol. 3, p. 175. London, UK: The Electrician Printing and Publishing Company.
14. Maxwell JC. 1991 *Matter and motion*, p. 136. Mineola, NY: Dover Publications, Inc. (Originally published 1920. Revised edition with notes and appendices by Sir Joseph Larmor).
15. Heaviside O. 1892 *Electrical papers*, vol. 2, pp. 124–128. London, UK: Macmillan.
16. Franklin WS. 1990 *Electric waves: an advanced treatise on alternating-current theory*, pp. 88–93, ch. 4. New York, NY: The Macmillan Company. (http://bit.ly/1rerFKe)
17. Schantz HG. 2014 On the superposition and elastic recoil of electromagnetic waves. *FERMAT* 4, [ART-2014-Vol4-Jul_Aug-002]. See http://arxiv.org/abs/1407.1800.
18. Schantz HG. 2017 Electromagnetic models and their application. In *Proc. of the 2017 Antenna Applications Symp., Allerton Park, Monticello, IL, 19–21 September 2017*, pp. 239–268. Urbana, IL: University of Illinois.
19. Schantz HG. 2017 *A rambling wreck*, pp. 222–226. Seattle, WA: Createspace.

20. Heaviside O. 1894 *Electrical papers*, vol. 2, pp. 126–127. London, UK: Macmillan.
21. Schantz HG. 2001 Electromagnetic energy around Hertzian dipoles. *IEEE Antennas Propag. Mag.* **43**, 50–62. (doi:10.1109/74.924604)
22. Mandel L. 1972 Energy flow from an atomic dipole in classical electrodynamics. *J. Opt. Soc. Am.* **62**, 1011–1012. (doi:10.1364/JOSA.62.001011)
23. Schantz HG. 1995 The flow of electromagnetic energy in the decay of an electric dipole. *Am. J. Phys.* **63**, 513–520. (doi:10.1119/1.17862)
24. Heaviside O. 1893 *Electromagnetic theory*, vol. 1, p. 55. London, UK: The Electrician Printing and Publishing Company, Limited.
25. Feynman RP, Leighton RB, Sands M. 1963 *The Feynman lectures on physics*, vol. 1, pp. 28. Reading, MA: Addison-Wesley Publishing Company.
26. McDonald KT. 2018 On the history of the radiation reaction. 6 May 2017; updated 7 March 2018. See http://puhep1.princeton.edu/~mcdonald/examples/selfforce.pdf.
27. Schantz HG. 2014 On energy flow in standing waves. In *Antenna Applications Symp., Allerton Park, IL, 25 September 2014*. Urbana, IL: University of Illinois.
28. Dirac PAM. 1967 *The principles of quantum mechanics*, 4th edn. (revised), p. 9. New York, NY: Oxford University Press.
29. Newton I. 1981 *Letter to Richard Bently* (*January 1693*)*, as quoted by A. Rupert Hall, From Galileo to Newton*, pp. 314–315. New York, NY: Dover Publications.
30. Glauber RJ. 1995 Dirac's famous dictum on interference: one photon or two? *Am. J. Phys.* 63, 1995, 12. (doi:10.1119/1.17790)
31. Feynman RP, Leighton RB, Sands M. 1963 *The Feynman lectures on physics*, vol. 1, pp. 37. Reading, MA: Addison-Wesley Publishing Company.
32. de Broglie L, Silva JAE. 1968 Interpretation of a recent experiment on interference of photon beams. *Phys. Rev.* **172**, 1284–1285. (doi:10.1103/PhysRev.172.1284)
33. Heaviside O. 1889 *Electromagnetic waves*, p. 24. London, UK: Taylor and Francis.

rsta.royalsocietypublishing.org *Phil. Trans. R. Soc. A* **376**: 20170453

PHILOSOPHICAL
TRANSACTIONS A

rsta.royalsocietypublishing.org

Research

Cite this article: Yakovlev A. 2018 Energy current and computing. *Phil. Trans. R. Soc. A* **376**: 20170449.
http://dx.doi.org/10.1098/rsta.2017.0449

Accepted: 9 June 2018

One contribution of 13 to a theme issue 'Celebrating 125 years of Oliver Heaviside's 'Electromagnetic Theory''.

Subject Areas:
electrical engineering, microsystems, electromagnetism, theory of computing

Keywords:
electromagnetism, energy current, energy-modulated computing, Oliver Heaviside, Petri nets, transmission lines

Author for correspondence:
Alex Yakovlev
e-mail: alex.yakovlev@ncl.ac.uk

†Present address: School of Engineering, Newcastle University, Merz Court, Newcastle upon Tyne NE1 7RU, UK.

THE ROYAL SOCIETY
PUBLISHING

Energy current and computing

Alex Yakovlev†

School of Engineering, Newcastle University, Newcastle upon Tyne, UK

AY, 0000-0003-0826-9330

In his seminal *Electrical papers*, Oliver Heaviside stated 'We reverse this . . . ' referring to the relationship between energy current and state changes in electrical networks. We explore implications of Heaviside's view upon the state changes in electronic circuits, effectively constituting computational processes. Our vision about energy-modulated computing that can be applicable for electronic systems with energy harvesting is introduced. Examples of analysis of computational circuits as loads on power sources are presented. We also draw inspiration from Heaviside's way of using and advancing mathematical methods from the needs of natural physical phenomena. A vivid example of Heavisidian approach to the use of mathematics is in employing series where they emerge out of the spatio-temporal view upon energy flows. Using *series* expressions, and types of natural discretization in space and time, we explain the processes of discharging a capacitive transmission line, first, through a constant resistor and, second, through a voltage controlled digital circuit. We show that event-based models, such as Petri nets with an explicit notion of causality inherent in them, can be instrumental in creating bridges between electromagnetics and computing.

This article is part of the theme issue 'Celebrating 125 years of Oliver Heaviside's 'Electromagnetic Theory''.

1. Preface

This year we are celebrating the 125th anniversary of Heaviside's electromagnetic theory. For me personally, the turning point in terms of rediscovering Heaviside was the year 2013. Nearly 40 years had passed then since I started my Master's degree in Computer Engineering at what was the oldest Electrical Engineering school in the USSR, now called St Petersburg Electrotechnical University. The 5 year curriculum of the Computer Engineering degree included two substantive courses on Engineering Maths and on Electrical Circuit Theory. In those lectures, I had first heard of Heaviside in connection with the step function and operational calculus. For some reason, our curriculum did not contain a course on Electromagnetic Fields, though many of my friends doing other degrees such as Automatic Control or Electrical Measurements, let alone Radio Engineering, did Fields. Curiously, the national academic board of Computer Science degree studies felt that we, as future computer engineers, would not need to go as deep as the Fields. The Board probably thought that knowing Circuit Theory would be sufficient for all our practical needs, in order to understand how analogue and digital circuits work. Today, in hindsight, I realize that was an omission. I metaphorically talk about this 'smart separation of concerns' as building a 'circuit theory wall', beyond which computer engineers should not go! Perhaps, indeed, most of the contemporary computer engineers and academics never need to look over this wall. Some knowledge of electromagnetics passes to them through a standard undergraduate Physics course, and everybody accepts that this should be enough for having a bigger picture of the electrical and electronic world.

Yet, in the year 2013, I came across Oliver Heaviside's work in full. How? From around 2010 I was involved in a large research project on the new generation of energy-harvesting electronics. The project embraced several research teams from different universities. The full set of expertise of these teams included power generation, power regulation and energy-harvesting-aware computational circuits. My team at Newcastle was responsible for the latter. It was during that project that I started to build the vision of computational electronics that was driven by potentially unpredictable and variable power sources rather than what had usually been seen as the recipient of unlimited power or energy supply. I called this type of computing 'energy-modulated' [1], to emphasize the fact that it was energy flow that defined the computational action. In the following section on energy-modulated computing, I shall briefly explain its main idea. To underpin this concept, a patent on a self-powered voltage sensor was filed in 2011 [2]. In this method and apparatus, we used a capacitor to store the energy of the measured signal as a charge. This charge then was used to drive a self-timed frequency divider (or counter), which effectively converted the charge in the capacitor into binary code. At some point, the voltage across the capacitor became too low to drive the increment action in the counter. The resulting code could then be used as a digital measure of the initial voltage across the capacitor. An interesting challenge was to determine the strict relationship between this voltage and the code. Furthermore, another interesting question came up. What is the law according to which the charge, and hence the voltage across the capacitor, follows in time when the capacitor is discharged through a digital load formed by such a self-timed circuit? This answer was found and it appeared to be a hyperbolic (not exponential!) decay law—we presented in the following paper [3] in 2013. This derivation is reproduced in this paper.

Next, in the same 2013, by sheer coincidence I exchanged emails with Mr Ivor Catt about the late Professor David Kinniment, my colleague and mentor of many years, who studied an interesting and challenging phenomenon called metastability (connected to the philosophical problem of choice and the story of Buridan's ass) [4,5] in digital circuits during his 45-year academic career. From David Kinniment I had known that Ivor Catt was one of the early discoverers of this phenomenon, which he called The Glitch [6]. To my amazement, in my conversation with Ivor Catt, he told me about his other passion. That other work, which had absorbed him for nearly 40 years, was on developing and promoting his own version of electromagnetic theory (called Catt-theory or Theory C) [7]. Ivor Catt sent me his book and several articles in IEEE journals and in the *Wireless World* magazine. They showed how this theory advanced Heaviside's theory (Theory H) of transverse electromagnetic (TEM) waves and the concept of energy current. I managed to organize a seminar on Electromagnetism at Newcastle on 9 October 2013 to which I invited Mr Ivor Catt and Dr David Walton, who worked with Ivor Catt on various parts of his theory, particularly on demonstrating that a capacitor is a transmission line (TL) [8]. Coincidentally, David Walton obtained both of his Physics degrees from Newcastle

rsta.royalsocietypublishing.org Phil. Trans. R. Soc. A 376: 20170449

University, and on the same day of 9 October 2013 there was a historical 50th anniversary reunion of Electrical Engineering graduates of 1963, some of whom had known David Walton (moreover some again, by coincidence had known Ivor Catt), so the date was truly momentous. Ivor Catt himself gave a 2 h lecture [9] which was followed by an hour-long lecture by David Walton [10]. These lectures showed a demonstration of the physics of some phenomena, ordinarily known to engineers, such as charging a capacitor, in an unconventional form—namely by applying a step voltage to a TL. The well-known exponential charging was the result of an approximated series of discrete steps caused by the cyclic process of the travelling TEM wave. This theory was supported by an experiment, known as Wakefield experiment [11], which led to the conclusion that there is no such a thing as a static electric field in a capacitor. In other words, a capacitor is a form of TL in which a TEM wave moves with a single fixed velocity, which is the speed of light in the medium. Below we reproduce both the derivation of the TL-based capacitor discharge and the description of the Wakefield experiment.

Those lectures triggered my deep interest in studying Oliver Heaviside's work and, even more, his whole life. And this very interest drew me to (then PhD student but now Dr) Christopher Donaghy-Spargo, with whom we founded NEMIG—northeast Electromagnetics Interest Group, which since 2013 has enjoyed a formidable series of seminars given by scientists, engineers, historians and entrepreneurs, driven by the ideas and lives of Maxwell, Heaviside and generally by the exciting field of electromagnetism.

Coming back to the main object of this paper, which is the relationship between energy current and computing, I must admit that I had drawn most of inspiration from my familiarization with Heaviside's work, his legacy in the work of others, and to a great extent by the fact that both Ivor Catt and David Walton came to studying electromagnetic theory from the point of view of energy current through their experiences in dealing with high-speed digital electronics. This electronics does not deal with sine waves. It deals with digital pulses, which are physical enough to be dealt with in a 'more physical way' rather than expressing them as an algebraic sum of sine wave harmonics stretching in the time domain from $-\infty$ to $+\infty$. Such pulses have a clear starting point in time and endpoint in time. They naturally lend themselves to causality between actions, such as a rising edge of one pulse causes a falling edge of another pulse, for example, as the signal passes through a logic NOT element (inverter). As I spent most of my own 40 working years exploring asynchronous self-timed digital circuits, and such circuits could work directly when the power is applied to their vdd lines, I was firmly attracted by the natural beauty of the ideas of the electromagnetic theory approach relying basically only on energy current, Poynting vector ($S = E \times H$, vector product of the electrical field vector and magnetic field vector, representing the directional energy flux, measured in Watt per square metre; note that it is sometimes referred to as Umov–Poynting vector) and TEM wave—particularly by its compactness and parsimony of Occam's Razor.

Another important aspect of my fascination of the energy-current approach to computational electronics is associated with the fundamental role that mathematical series play there. Series, so much loved and revered (to the poetic level!) by Heaviside, are at the core of the vision of all electromagnetic phenomena because they relate all state changes in the electromagnetic field with the geometry of the space and medium. Likewise, series form the foundation to the definition of main measurements of the dynamic fields and hence for the subsequent discretization and quantization important for the computational procedures. This paper pays particular attention to the use of mathematical series, their constructions, summations and other analytical operations. By doing so, this paper would like to pay homage to Heaviside as a magician of the series mathematics.

Setting the scene, I would like to finish this preface with a quote from David Walton's lecture abstract [10]:

> It is normally recognised that the postulation of Displacement current by James Clerk Maxwell was a vital step which led to the understanding that light was an electromagnetic wave. I will examine the origins of displacement current by consideration of the behaviour

of the dielectric in a lumped capacitor and will show that it has no physical reality. In the absence of an ether there is no rationale for displacement current. We are then left with a theory which works mathematically but has no basis in physical reality. I will discuss the remarkable property of empty space in that it has the ability to accommodate energy. I will then show that Faraday's law and conservation of charge lead to the existence of electromagnetic energy which travels at a single fixed velocity and has a determined relationship between the electric and magnetic fields. Because this mathematics is reversible it follows that these two physical laws can be considered to be consequences of the nature of electromagnetic energy rather than the reverse.

2. Energy-modulated computing

The question we pose here is: *How does energy drive computations?*

The traditional way of making computers perform calculations is to connect a processor, memory and all the interfaces to the outside world to a power supply. The power supply is a source of energy, which must always be THERE and ON and deliver as much power (energy per time unit) as the computer needs to run its programs and interact with the environment. Typically, these needs are determined by the rate of activity of the processor, or more specifically by the frequency at which the clock generator runs the processor. In traditional systems, the 'motive force' for the processor comes from the clock generator under whose tick-events the hardware of the computer updates its state. This way of building computer systems considers power supply and timing mechanisms, both needed to the processor, separately, and such a separation requires significant effort (including power and time costs) to make sure that the power level always matches the frequency, otherwise the computer will produce erroneous results. With this approach, which has been dominating computer engineering for many years, any considerations of energy usage are usually in context with energy consumption, which is seen as an effect of the performed computational actions rather than cause thereof.

Now, instead of separating power and timing sources, we can think of the processor's own timing mechanism that can be powered by the same power supply as the rest of the system. It will then turn the whole system into an energy-driven machine, which will be driven by the energy flow, pretty much like any biological organism functions in nature. Traditionally, the relationships such as Power $= f(F, V)$, where F is the switching frequency and V is the supply voltage, have been studied and the figure of merit has been, for example, Watt/MegaHertz. In energy-driven approach we are more interested in the relations such as Delay $= f(V)$ or Delay $= f(P)$, or Rate $= f(V)$, hence the figure of merit is MegaHertz/Watt or MegaHertz/Volt.

Figure 1 illustrates the above distinction, by showing two designs of a functionally same system, which switches between different modes, idle, data processing and communication. Clearly, the traditional way is significantly less robust to possible variations of supply voltage vdd, which either makes the system error-prone or inefficient with lots of interruptions. This view upon the computing systems being an exemplar of physical systems, rather than purely mathematical objects, i.e. products of human imagination, brings us to a most inspiring quote from Heaviside's work [12]:

> Now, in Maxwell's theory .. the potential energy . . . and . . . the kinetic or magnetic energy are supposed to be set up by the current in the wire. We reverse this; the current in the wire is set up by the energy transmitted through the medium around it. The sum of the electric and magnetic energies is the energy of the electric machinery which is transmitting energy from the battery to the wire. It is definite in amount, and the rate of transmission of energy (total) is also definite in amount.

Although this reversal of the viewpoint mostly concerned the analytical aspect of Electrical Science, it can be used today as a system design principle or paradigm. Real-power computing [13] describes hardware and software systems subject to a 'real-power constraint', for example, the availability of energy from a power source, or restriction on power dissipation. Real-power

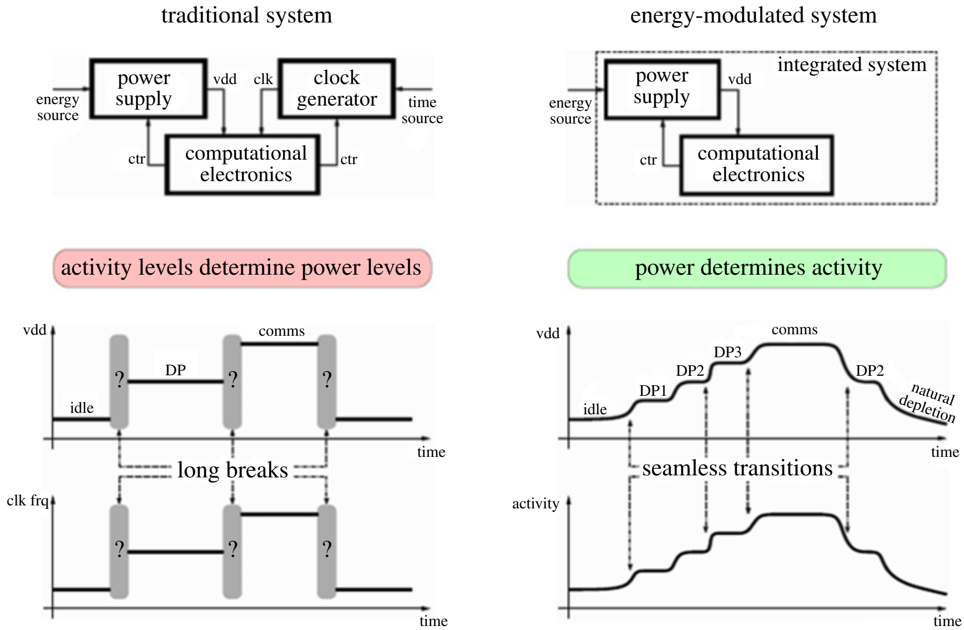


Figure 1. Traditional versus energy-modulated view upon system design. (Online version in colour.)

applications must guarantee performance within specified definite power constraints, referred to as 'power bands'. Systems of this type are linked to the notion of survivability, which depends on their power aspects as well as their ability to morph and retain functional aspects to ensure continued computation. Real-power systems go well beyond conventional low-power systems that are optimized for minimum power consumption only.

The energy-modulated approach to system design permits building systems in which information and energy flows to the system are not structured as fundamentally different thereby, again, bringing our computing engines closer to natural conditions. Traditionally, they are supplied separately, which makes them scale poorly with modern technology. Namely, either there is much energy unused where there is no need for processing information, or information is lost when energy is not available. Contrary to this, we can actually supply information with energy in bundles, hence leading to *computing with energy*.

In this spirit, we can structure information and energy in the form of tokens (cf. Heavisidian pulse-based view). An information token must consist of Data that is Powered (Energy Validity) and Ready (Timing Validity). Here we can associate such a token with a signal in the form of an electric current which can be integrated as a charge in a capacitor (figure 2). In our subsequent sections, we will show how this charge can be turned into a computable object or datum, i.e. for example measured and digitized. It should be noted that a signal which contains timing and power validity is both self-timed and self-powered (STSP). A practical way to process such signals in electronics is to use asynchronous circuits [14], where researchers have built a significant body of knowledge since the late 1950s.

This idea is schematically shown in figure 3. As we show further in this paper, computing can be involved at the level of energy current or at the level of the accumulated energy, such as electric charge. What one can see here is an STSP signal, or several such signals, that trigger processing in the functional block. The triggering action can be tuned to very small scale energy levels, and instead of using traditional amplification, which requires energy-costly separate power regulator, one can use integration of charge above a certain threshold. (Fundamentally, integration is also far more noise-immune than amplification!) With modern technology scaling, this approach can help us avoid two main shortcomings of the conventional computing, namely (i) unused energy when there is no need for processing, and (ii) lost information when energy is not available.

rsta.royalsocietypublishing.org Phil. Trans. R. Soc. A 376: 20170449

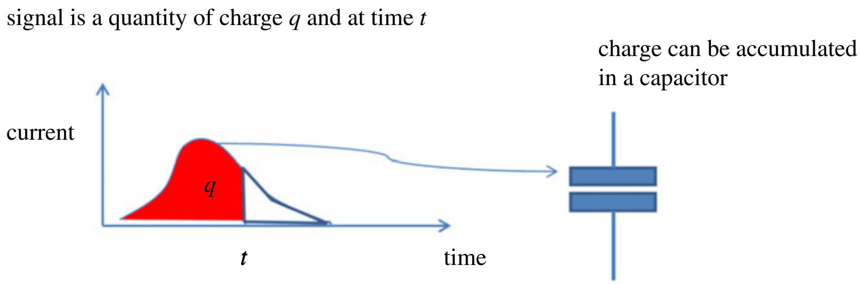


Figure 2. Catching the energy of a signal in a capacitor. (Online version in colour.)

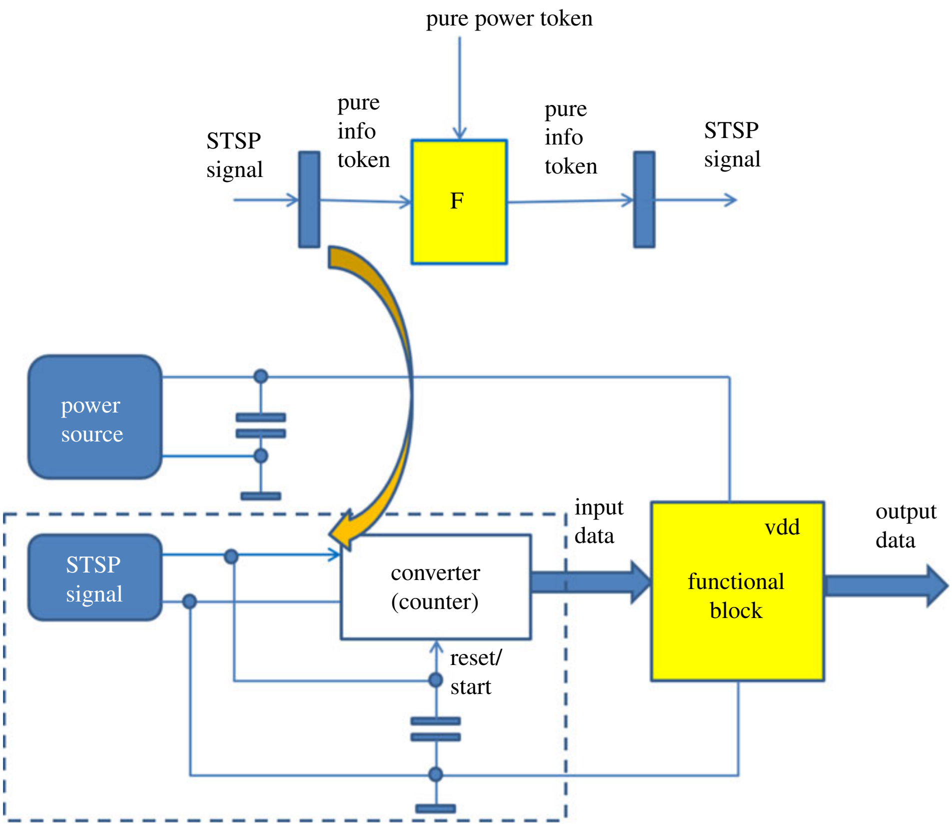


Figure 3. Processing self-timed self-powered signals. (Online version in colour.)

What's crucial for energy-modulated computing is the notion of causality, discussed below.

(a) Energy current and causality

Interestingly, the view in which energy flow is seen as a cause to performing action goes back to Galileo and Newton, who emphasized the role of geometry in representing causality and dynamics in nature. Dr Ed Dellian, who translated and edited the key works of Galileo and Newton, wrote to me (E Dellian 2018, personal communication):

> I have learned that the matter requires to distinguish carefully between arithmetic (algebra) and (Euclidean) geometry. Galileo's saying in the 'Saggiatore' (1623) refers to geometry as the 'art of measuring' what there 'really is'. Arithmetic (algebra) on the other hand is 'the art of calculating' what there 'should be'. In order to understand the importance of the

> art of measuring in natural philosophy, that is, geometry as the 'language of nature', one must consider what Newton says in his foreword of 1686 to the 'Principia'. This is at least as relevant as Galileo's dictum in the 'Saggiatore': For Newton, geometry is the general basis of all of 'mechanics'. He calls geometry 'that part of mechanics in general which demonstrates and teaches the art of exact measuring'.

To us, pursuing the above energy-modulated viewpoint, measuring and computing are synonymous. It is therefore important to see the distinction between geometric (physical) and algebraic (metaphysical) views on nature, with the former building on and expressing causality and causal proportionality while the latter being non-causal, and what is sometimes called tautological proportionality. To this end, Dellian further wrote:

> You know already that this 'E', the symbol of 'cause', is not equivalent to the classical energy 'E'. Rather it is the same thing as Galileo's and Newton's 'vis impressa', the cause of (change of) motion (cf. Newton's def. 4). Note that Newton explicitly says to have taken his first and second law from Galileo!

In other words, we have two 'kinds' of energy—one is *creative and causal*, and the other is purely *characterizational*. In the following, a series of statements summarize my understanding of Dellian's analysis and its relation to electromagnetic theory:

1. Energy current (**E-vector**) causes **momentum *p***.
2. **Causality** is made via the proportionality coefficient c (**speed of energy current, i.e. the speed of light**)
3. Momentum p is what mediates between E-vector and changes in the **matter**.
4. Momentum p is preserved as energy current hits the matter.
5. Momentum in the matter presents another form of energy (**E-scalar**).
6. E-scalar characterizes the elements of the matter as they move with **a (material) velocity**.
7. As elements of the matter move, they cause changes in Energy current (E-vector) and this forms a *fundamental feedback mechanism* (which is recursive/fractal ...).

Telling this in terms of **electromagnetic theory and electricity**:

1. E-vector (Poynting vector/Heaviside signal) causes E-scalar (electric current in the matter).
2. This causality between E-vector and E-scalar is mediated by momentum p causing the motion of charges.
3. The motion of charges with material velocity causes changes in E-vector, i.e. the feedback effect mentioned above (e.g. self-induction).

We can now conclude that the key aspects for energy-modulated computing, inspired by Heaviside's notion of energy current, are the following:

— Capturing energy current that moves in space with speed of light into material form (possibly electric charge in a capacitor) to enable measurement and hence information processing.
— Detecting the completion of such process in the form of validity information.
— Discretizing information in time in the form of steps and cause–effect relationships between signals.

Thus computation can take place at different levels, at both energy current level as well as at the material level. As we show in the following sections the important role in defining computation at both these levels mathematically is in the use of series. Such series characterize the process of energy accumulation and division in space and in time.

3. Computing by accumulating and dividing energy

(a) On the creative role of *series*

Use of functional series was essential for Heaviside's mathematical toolbox. He wrote in [15]:

> The subject of the decomposition of an arbitrary function into the sum of functions of special types has many fascinations. No student of mathematical physics, if he possesses any soul at all, can fail to recognise the poetry that pervades this branch of mathematics.

This is an inspiring statement and the energy current approach to the analysis of electronic systems is exactly what encourages one to use series. Why? This is because the energy current carrying any change in the electromagnetic field always moves with a speed of light (which is a constant for a given medium), and all processes of changing the state of the field can be seen in the form of spatial division, or discretization, so that exact time intervals between them can be determined from the spatial values and the speed of light. It is then very handy to use simply the indexation of steps in iterative processes associated with the changes of the states. One of the best illustrations of this methodology comes with the analysis of TLs. In the following section, we will illustrate this with the use of the energy current approach in describing the process of discharging a capacitor through a constant resistance. The capacitor will be modelled by a lossless TL. I find this analysis, which is reproduced from Catt *et al.* [16], highly illuminating. In this analysis we avoid modelling a TL as a set of distributed RC parameters, and only rely on the notion of a TEM wave travelling in the medium between two metal wires (or, equivalently, plates of a capacitor) with a speed of light, as postulated by Catt's electromagnetic theory [7].

(b) Capacitor as transmission line

The configuration that we want to consider here is shown in figure 4.

Assume, first, that the capacitor was charged via resistor R to the voltage V (via switch S1). Then we disconnect S1 and connect S2. The capacitor is a (e.g. coaxial cable) TL with a characteristic impedance Z0. Let us assume that $R \gg Z0$, and we assume that R is constant. The reflection coefficient at the right-hand side terminals of the open-ended TL: $\rho = (R - Z0)/(R + Z0)$. So, at time $t = 0$, we have S2 in the ON state, while S1 is in OFF state, and $V_0^c = V$. At this moment a down-step of magnitude $(Z0/(R + Z0))V$ starts to propagate down the TL. As it is open on the right-hand side it is reflected with a coefficient 1 and is seen on the left-hand side as $V_1^c = V - 2(Z0/(R + Z0))V$. After n two-way passes of the TEM wave:

$$V_n^c = V_{n-1}^c - 2\frac{Z0}{R + Z0} V\rho^n = V\left(1 - 2\frac{Z0}{R + Z0}\sum_{i=0}^{n}\rho^i\right). \tag{3.1}$$

Denote $a = 2\,(Z0/(R + Z0))$ and, because $\rho < 1$ we can use the sum of geometric progression for $\sum_{i=0}^{n-1}\rho^i = ((1 - \rho^n)/(1 - \rho))$. Therefore, we have $V_n^c = V(1 - a((1 - \rho^n)/(1 - \rho)))$. However, $a/(1 - \rho) = 1$, hence we have

$$V_n^c = V\rho^n. \tag{3.2}$$

It is now clear that the discharge process is exponentially decaying. In the long run, $V_n^c \to 0$ as $n \to \infty$.

Now we can show that this exponential process can be expressed in the form $e^{-(t/\tau)}$, which is customary to standard lumped capacitor discharge analysis. Consider the term

$$\rho^n = \left[\frac{R - Z0}{R + Z0}\right]^n = \left[\frac{1 - (Z0/R)}{1 + (Z0/R)}\right]^n. \tag{3.3}$$

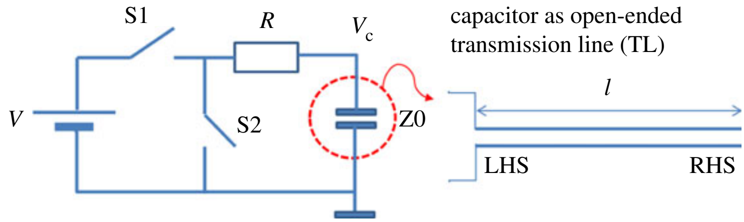


Figure 4. Circuit for charging and discharging a capacitor seen as a transmission line. (Online version in colour.)

If, as we assumed, $Z0/R \ll 1$, then $\rho^{\mathrm{n}} \approx [1\text{–}2(Z0/R)]^{\mathrm{n}}$. Define $k = 2(Z0/R)n$, and hence $\rho^{\mathrm{n}} \approx [1-(k/n)]^{\mathrm{n}}$. So,

$$\lim_{n\to\infty}\left(1-\frac{k}{n}\right)^{n} = \mathrm{e}^{-k} = \mathrm{e}^{-(2Z0n/R)}. \tag{3.4}$$

Now let us connect the index of step n with time t. At time t, $n = vt/2l$, where v is the velocity of propagation, i.e. the speed of light in the medium, and l is the length of TL. Hence, $V^{c}_{n+1} = V^{c}(t) = V\mathrm{e}^{-(2Z0n/R)} = V\mathrm{e}^{-(v/l)(Z0/R)t}$. Now, for any TL, we know that $Z0 = f\sqrt{\mu/\varepsilon}$ and $v = 1/\sqrt{\mu\varepsilon}$, where μ is the permeability and ϵ is the permittivity of the dielectric medium between the plates of the TL and f is the geometric factor. This geometric factor relates a linear unit capacitance c with ε, as follows: $\varepsilon = cf$. So, for the TL of length l, the total capacitance is $C = cl$. Substituting all these elements into $(v/l)(Z0/R) = 1/\mathrm{RC}$, and finally, we have $V^{\mathrm{c}}(t) = \mathrm{e}^{-(t/\mathrm{RC})}$. In other words, thanks to the two important series that we summed, a geometric series and the power series representing a natural exponential, we have been able to derive the approximation of the step-wise process driven by energy current captured in the space defined by the TL. This process is equivalent to the process of discharging a lumped capacitor of capacitance C through a lumped resistor of resistance R.

Interestingly, the appropriateness of modelling capacitors with fast switching as TLs has been demonstrated experimentally by Sullivan & Kern [17]. Without resorting to fundamental arguments, such as Heaviside's energy current or Catt theory, they observed the following:

> The ideal transmission line model exaggerates the definition of the steps in the voltage waveform, particularly after more than one round-trip time. If one simply evaluates rms error between the simulated waveform and the actual waveform, one might think that the ideal transmission line is a poor model, and that more accurate transmission line models, as begun in section V-D, are necessary before transmission line models can be useful to the designer. We intend to work on such improved models, but we believe that the simple ideal transmission line model will often be more useful to the designer. It provides a conceptual model of what behavioural features to expect, and relates those features directly to geometry. In particular, the height and timing of the first and second steps are probably the most important features. The height of the first step is given by $Z_0\Delta I$ for the height of the first step, and twice that for the second step.

(c) The Wakefield experiment

An experimental evidence of the stepwise discharge process for a capacitor modelled by a co-axial cable has been presented by Ivor Catt in *Electronics World* in April 2013 [11]. Here is only a brief recap of this description. The experiment bears the name of Mr Tony Wakefield of Melbourne, who actually built the configuration and performed all the measurements. Catt wrote:

> We now have experimental proof that the so-called steady charged capacitor is not steady at all. Half the energy in a charged capacitor is always travelling from right to left at the speed of light, and the other half from left to right [see figure 5].

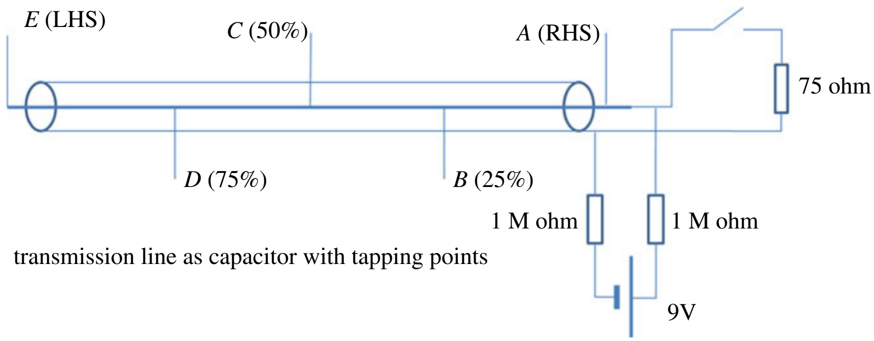


Figure 5. Wakefield experiment set-up: coaxial cable as a cap with tapping points. (Online version in colour.)

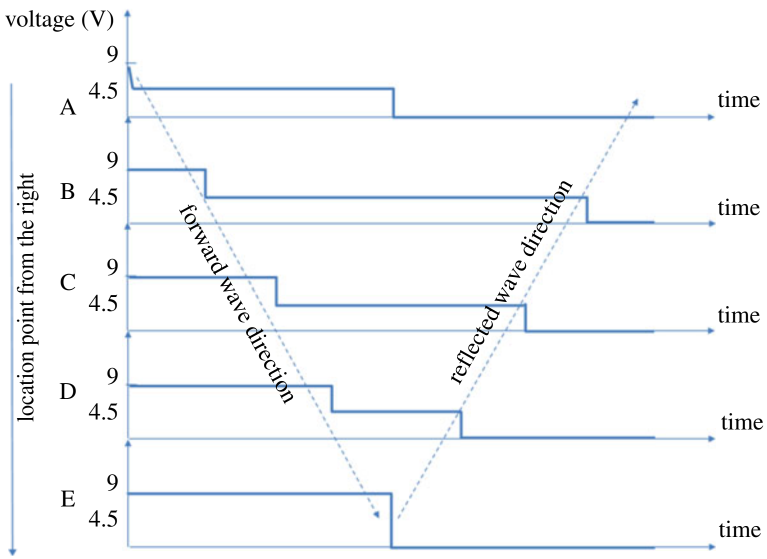


Figure 6. Signal plots for the Wakefield experiment, in five different locations. (Online version in colour.)

The Wakefield experiment uses a 75-ohm coax 18 meters long. The left-hand end is an open circuit. The right-hand end is connected to a small, 1 cm long, normally open reed switch. On the far side of the reed switch is a 75-ohm termination resistor simulating an infinitely long coaxial cable. A handheld magnet is used to operate the switch.

The coax is charged from a 9 V battery via 2×1 megohm resistors, close-coupled at the switch to centre and ground. The two resistors are used to isolate the relatively long battery wires from the coax. High value resistors are used to minimize any trickle charge after the switch is closed.

A 2-channel HP 54510B digital sampling scope set to $2\,\mathrm{V\,div^{-1}}$ vertical and $20\,\mathrm{ns\,div^{-1}}$ horizontal is used to capture five images.

For the reasons of copyright, I cannot copy these images from Catt's paper. But, they were taken in the following points: (A) across the terminator 75-ohm resistor, (B) 25% to the left of the reed switch (4.5 m), (C) 50% to the left of the reed switch (9 m), (D) 75% to the left of the reed switch (13.5 m), (E) at the extreme left of the open end of the cable (figure 6).

rsta.royalsocietypublishing.org Phil. Trans. R. Soc. A 376: 20170449

(d) What if the load is a self-timed digital circuit?

When I asked myself a question of what the law of discharging a capacitor via a self-timed switching circuit is, I had already suspected that it was not a normal exponential process that we see in an RC circuit with a constant R. Clearly the fact that the switching circuit forms a voltage-controlled impedance would affect the discharging process. Theoretically, one can try to produce a model of resistor R and apply it at every step of the discharging process. But then the question arises, what if the dynamics of change of R is not related to the dynamics of the capacitor discharge? While it would still be interesting to study the discharging process for a TL in place of a lumped C, here we can consider the model of discharging a lumped capacitor via a digital switching circuit as the first step in this direction. Here, we will focus on the step-wise process associated with the changes in the resistive load due to switching from one destination capacitor to another capacitor and modelling the dynamics of the delay in such switching. As a self-timed digital switching circuit, we consider a ring oscillator. The original details of this derivation can be found in papers with my colleagues Dr Reza Ramezani, Dr Alex Kushnerov and Dr Andrey Mokhov [14,18].

Let us consider a capacitor and a ring oscillator built out of an odd number of inverters, as shown in figure 7*a*. With two switches we can first charge the capacitor from a voltage source (switch S1 On and Switch S2 Off), and then discharge it by powering up the ring oscillator (switch S1 Off and Switch S2 On). A switching process in any digital circuit, built from CMOS gates, is a process of charging and discharging its parasitic capacitances through PMOS and NMOS transistors, respectively. For a ring oscillator, this behaviour proceeds in series, between each pair of adjacent inverters, wherein the preceding inverter, after charging its parasitic capacitance from the power line, activates the input of the next inverter, which pulls down its output, thereby discharging its parasitic capacitor, and so on. This process is illustrated in more detail in figure 8*a*, where it is shown that the pull up (charging) is done via the PMOS transistor in the ON state and the pull-down (discharging) via the NMOS transistor in the ON state. The charging of each such parasitic capacitance C_p is by taking part of the charge from the main capacitance C.

Because there is always only one inverter active, we can describe the operation of the ring oscillator and the main capacitor in three states, as shown in figure 8*b*. Configuration ① represents the circuit state when no charging or discharging action takes place. Half of the inverters (say, odd-numbered) are at logical '1', i.e. their corresponding capacitors C_p are charged to the voltage level equal to that of C. The remaining inverters are at logical '0' and their capacitors C_p are empty. This state also describes the circuit status between two successive switching events. The system is in state ② when a switching occurs. One of the charged capacitors C_p (of the inverter which is supposed to switch from 1 to 0) is discharged via the corresponding NMOS transistor. In state ③ the discharged capacitor associated with the next inverter in the ring (which is supposed to switch from 0 to 1) receives charge from the source capacitor. This is the only state which draws energy from the source capacitor. Note that, in real operation an overlap exists between states ② and ③, however, to simplify the analysis we consider them as distinct states.

At the level of abstraction where we consider configurations shown in figure 8, we would normally deal with the exponentials involved at each state of charging and discharging the current parasitic capacitor. As discussed in the previous section if we look at the main capacitor as a TL, such exponentials are actually step-wise processes. In this section we shall abstract away from those TL-related processes, assuming that their dynamics is much faster than the delays associated with the states shown in figure 8. Instead, we will go one level of abstraction higher and look at the steps associated with the transitions between such states. Thus we will concentrate only on the charge division at each step, which we call the V drop. The timing aspect, i.e. the propagation delay involved in the charging and discharging state, can be considered separately. The V drop in this circuit can be found by the law of conservation of charge, which redistributes between the capacitors $C_T = C + C_t$ and C_p according to their capacitances. The voltage across the main capacitor can be written as a recurrence relation that is unfolded then into an iterative process with the initial condition of V_0.

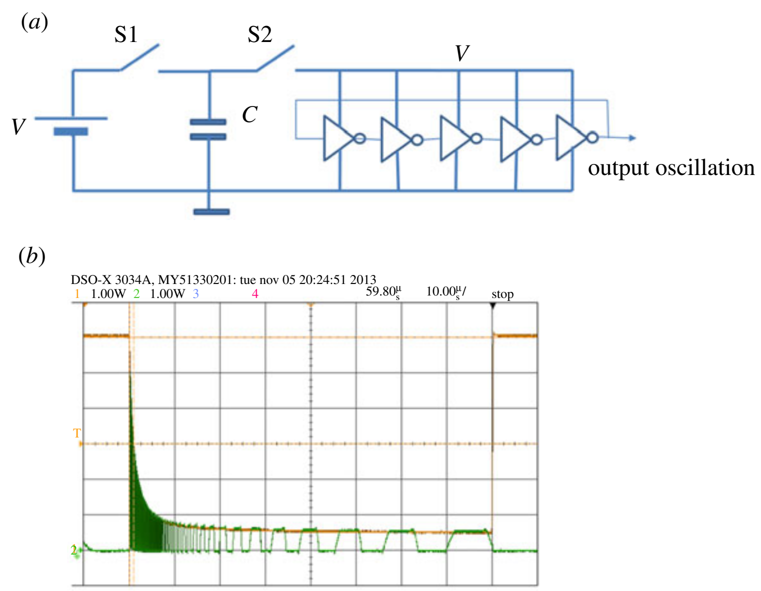


Figure 7. Ring oscillator powered by a capacitor (*a*) and its behaviour inverters (*b*), showing both the voltage drop and the output oscillation. (Online version in colour.)

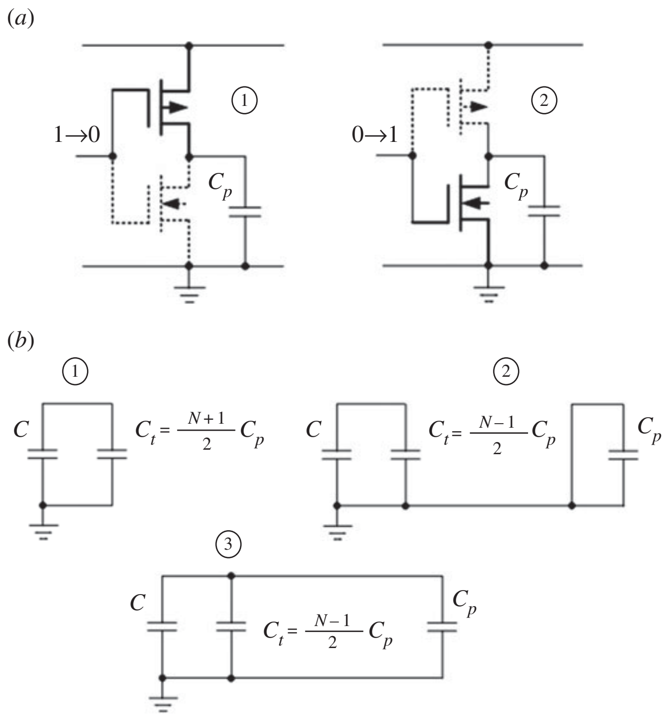


Figure 8. Charge switching: transistor switching (*a*), and the circuit state at dynamic switching (*b*), where *N* is the number of inverters.

Thus, at the n-th step, charge equilibrium occurs at

$$V_{n+1} = KV_n = K^n V_0, \tag{3.5}$$

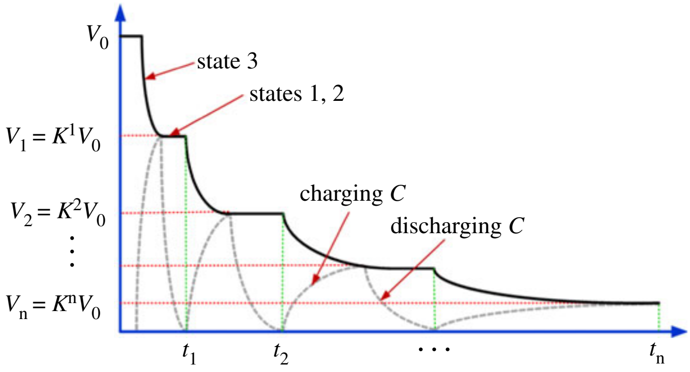


Figure 9. The V drop over time, with charging and discharging of parasitic capacitors in the inverters. (Online version in colour.)

where

$$K = \frac{C_T}{C_T + C_p} \tag{3.6}$$

is the coefficient of voltage decay at each switching action (or 'charge division coefficient'). Since $C_T \gg C_p$, K is only slightly less than one. Thus, the switching index n determines the voltage at each step. This process is shown in figure 9, where for the sake of better illustrating the effect of charge division we exaggerate the ratio so that K is much smaller than one. Let us now consider the delay of such steps. The propagation delay of an inverter is a function of supply voltage, which determines the behaviour of its NMOS and PMOS transistors. For operation above the transistor threshold (super-threshold region), the time of switching (propagation delay) is approximately reciprocal to V.

However, below threshold (sub-threshold region), the propagation delay becomes a more complex function of V. The following is a model for the propagation delay of a single inverter proposed in [19]:

$$t_d = \begin{cases} t_{\text{sup-th}} = \dfrac{pC_pV}{(V-V_t)^{\alpha}}, & V > V_t, \\[2ex] t_{\text{sub-th}} = \dfrac{pC_pV}{I_0 e^{(V-V_t)/N_s}}, & 0 < V \le V_t, \end{cases} \tag{3.7}$$

where p is the fitting parameter, α is the velocity saturation index (which we consider here for simplicity to be equal to 2), I_0 is the drain current at $V_{gs} = V_t$ and $N_s = mkT/q$ (m is the sub-threshold slope factor, $1/N_s \approx 28$) [20]. In the following, we will consider, for simplicity, the operation in the super-threshold region and approximate the propagation delay as $t_d = A/V$, where $A = pC_p$. The overall elapsed time at step i is conveniently represented by the sum of the individual propagation delays of the steps. Thus the physical time t is also determined by the increments of the switching index:

$$t = \frac{1}{V_0}\sum_{i=0}^{n-1}\frac{A}{K^i} = \frac{A}{V_0}\left(\frac{K^{n-1}}{K^{n-1}} + \frac{K^{n-2}}{K^{n-1}} + \cdots + \frac{K^0}{K^{n-1}}\right) = \frac{A}{V_0K^{n-1}}\sum_{i=0}^{n-1}K^i. \tag{3.8}$$

This adds up all the time intervals t_d that precede n and represents the total time spent on the first n switching events. We know that $K < 1$, so we have finally the sum of a geometric progression:

$$t = \frac{A}{V_0K^{n-1}} \cdot \frac{1-K^n}{1-K} = \frac{AK}{V_0K^n} \cdot \frac{1-K^n}{1-K}. \tag{3.9}$$

And if we consider that after the n-th switching event $V_n = V_0 K^{\mathrm{n}}$ we can substitute it into (3.9) and obtain

$$t = \frac{AK(1 - V_n/V_0)}{V_n(1 - K)}. \tag{3.10}$$

Now, solving this equation with respect to V_n, we obtain it as a function of time

$$V_n = \frac{V_0 AK}{V_0(1 - K)t + AK} = V_0 \frac{1}{wt + 1}, \tag{3.11}$$

where

$$A = 2pC_p \quad \text{and} \quad w = \frac{(1 - K)V_0}{AK}. \tag{3.12}$$

This characteristic is a hyperbola rather than exponential as it would have been if the capacitor was discharged through a constant resistance. Note again that the hyperbola of the kind of (3.11) is the result of considering the behaviour of transistors in the inverters of the circuit in the super-threshold region. In near-threshold and sub-threshold regions, the character of the discharging process is fairly complicated due to the exponential increase of the delay as voltage drops, and this significantly reduces the rate of the voltage drop due to logic switching. Eventually, the switching process stops. The remaining charge decay is mainly determined by the leakage current.

In this analysis, performed in a Heaviside way, an intermediate factor, called a switching index n, was introduced to simplify the process of deriving the important relationship between the voltage on the capacitor V and time. Notably, both time and voltage of the power supply of the circuit are thus quantized according to the switching index n, rather than using a fixed time discretization step as often happens in numerical analyses of circuits. Voltage is dropping exponentially with n, and time is stretched as a function of the supply voltage, or in other words, of the energy available in the main capacitor. In consequence, the overall time is obtained by accumulating the intervals of switching in the form of generating functions. Solving these functions provides analytical solutions which express voltage as a function of (global) time. Using the discrete parameter n as an intermediate step appears to be a useful 'trick' in deriving the relationship between the two characteristics, voltage and time, that are themselves continuous.

On the basis of the idea of discharging a capacitor via a self-timed circuit, we were able to build charge-to-digital converters (CDC) which could be used to measure capacitance or voltage. For this, we simply had to replace the ring oscillator with a counter of switching activity [21]. In those paper, we showed that, either we could use the self-timed counter directly as an oscillator-counter, or have a delay chain and a separate counter.

Figures 10 and 11 show how to build a CDC using an iterative delay chain discharge mechanism. An unknown capacitance C is charged to an initial voltage level V_H and then discharged to a pre-defined reference voltage level V_L. The number of iterations of discharging is then accumulated in a counter and it is related to the initial voltage level V_H and the value of C.

Let $k = (1 - K) = C_p/(C_T + C_p) \approx C_p/C_T$, because $C_T \gg C_p$, where $C_T = C + C_t$ and K was defined by equation (3.6). The step-wise capacitor discharge process was described above and led to the equation:

$$V_L = V_H K^n = V_H (1 - k)^n. \tag{3.13}$$

Based on Taylor series, $(1 - k)^{\mathrm{n}} = 1 - nk + n(n - 1)k^2/2! - n(n - 1)(n - 2)k^3/3! + \cdots$ and, if $nk \ll 1$, the above formula can be approximated as $(1 - k)^{\mathrm{n}} = 1 - nk$. From the fact that we have levels V_H and V_L fixed (i.e. *const*) by the measurement method, we must have $(1 - k)^{\mathrm{n}} = 1 - nk = const$. Thus, under $nk \ll 1$, $nk = const$, and hence

$$n \frac{C_p}{C + C_t} = const. \tag{3.14}$$

So, if $C \gg C_t$, then $C + C_t \approx C$. Thus, $C_p/C = const$, which means n must be linearly proportional to C. The value of n, if accumulated in the counter, can be used to obtain the value of a measured capacitance C.

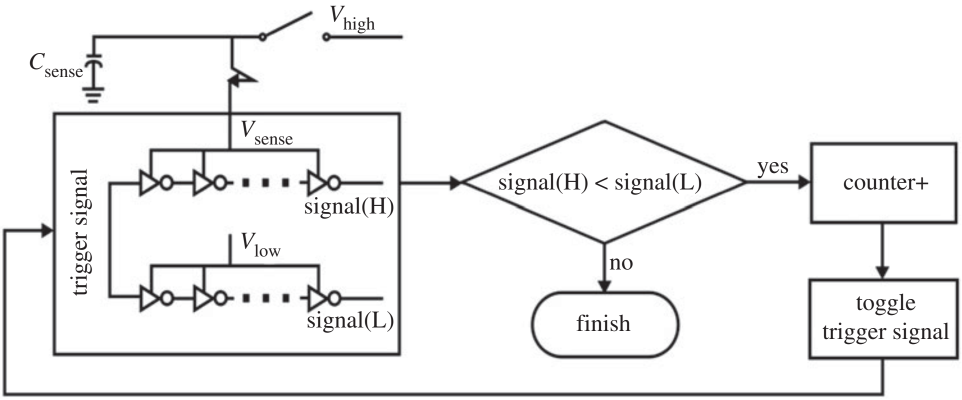


Figure 10. The idea of building a sensor using a charge to digital converter.

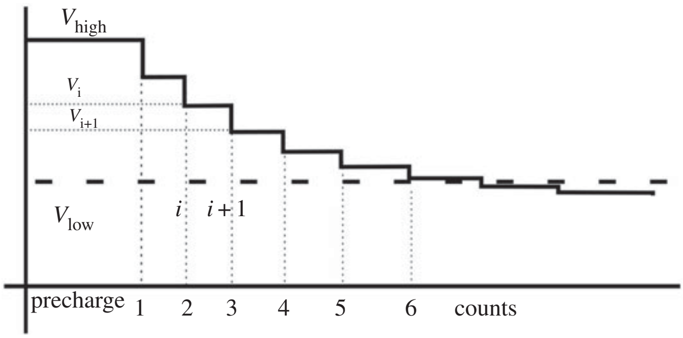


Figure 11. Discharging and counting process.

Using similar construction, we can also measure voltage V_H if we know the capacitance value C. Suppose we also know V_L. Therefore, since $V_L = V_H K^n$, after the number of discharging steps n, the latter can be determined as

$$n = \log_K \frac{V_L}{V_H}. \tag{3.15}$$

Thus, n is logarithmic with the measured voltage V_H.

(e) On quantization and discretization: hypotheses

In this section, I will consider some rather interesting, and possibly controversial, implications of the transients that we visited above thanks to Catt, Davidson and Walton's derivations. The artefact that those transients had envelopes that were exponential or sine/cosine curves was the result of having them been sums of series of steps in the first place. Furthermore, they originated as series of steps from one, rather simple but fundamental, postulate—that of the existence of energy current that is never stationary but always moves with the speed of light (Catt's Theory of [7]).

Understanding this postulate and the various analyses of transients in electrical systems is important. It is crucial for settling with the idea of the world being quantized by virtue of energy currents being trapped between some reflection points, and the continuous pictures of the transients are just the results of some step-wise processes.

I deliberately use word 'quantized' in the above because I tend to think that 'quantization' and 'discretization' are practically (in the physical sense; mathematicians may argue of course because they may add some abstract notion to these terms) synonyms. So I will try to explain my point below.

Let us see what happens with the TEM as it works in a TL with reflection. We have a series of steps in voltage which eventually form an exponential envelope (with a linear time-invariant resistor). If we examine these steps, they show discrete sections in time and amplitude. The values of time sections between these steps are determined by the finite and specific characteristics of the geometry of the TL and the properties of the (dielectric) medium. The value of the amplitude levels between these steps is determined by the electrical properties of the line and the power level of the source. So, basically, these discrete values associated with the energy entrapment in the TL are determined by the inherent characteristics of the matter and the energetic stimulus. If we stimulated the TL with periodic changes in the energy current, we could observe the periodic process with discretized values in those steps—the envelope of which could be a sequence of charging and discharging exponentials. Likewise, if we set up a TL (which is largely capacitive in the above) with an inductance, so we will have an LC oscillator; this would produce a periodic, similarly step-wise, discretized process whose envelope will be a sine wave (see [7]).

Now, if we analyse such a system in its discretized (rather than enveloped) form, we could produce some sort of histogram showing the distribution of how much time the object in which we trap energy current, spends in what level of amplitude (we could even assign specific energy levels). Now, we can call such an object a 'Quantum Object'. Why not? I guess the only difference between our 'quantum object' and ones that Quantum Physicists are talking about would be purely mathematical. We know the object well and our characterization of the discretized process is deterministic while in Quantum Physics the exact evolution of an object between discretized states may be unknown, hence probabilities are employed (interesting polemic issues may arise here as I seem to touch the territory of the EPR paradox [22]).

On the basis of these arguments, I would like to make some hypotheses.

We live in the world that has finite size objects of matter, whatever large or small they are. These objects have boundaries. The boundaries act as reflection points on the way of the energy current. Hence associated with these objects and boundaries we have entrapments of energy. These entrapments, due to reflections give rise to discretization in time and level. The grains of our (discretized) matter can be quite small so the entrapments can be very small and we cannot easily measure these steps in their sequences, but rather characterize by some integrative measurements (accumulate and average them—like in luminescence), hence at some point we end up being histogrammatic or probabilistic.

One more thing that may still bother us is the verticality of steps and their slopes.

Let us look at the moment when we change the state of a reed-switch or pull up the line to vdd or down to Ground. The time with which this transition takes place is also non-zero. I.e. even if the propagation of the change is with the speed of light, modulo the 'epsilon' and 'mu' of the medium, i.e. with finite time to destination, the transition of the voltage level must also be associated with some propagation of the field, or forces, inside the reed-switch or in the transistor, respectively, that pulls the line up or down. Clearly that time-frame is much smaller than the time frame of propagating the energy current in the medium along the TL, but still, it is not zero. I presume that, recursively, we can look at the finer granularity of this state change and see that it is itself a step-wise process of some reflections of the energy current in that small object, the switch, and what we see as a continuous slope is actually an envelope of the step-wise process.

So ultimately, we live in the recursive or fractal world of quantized space and the current limit of what we can measure as a step is something defined by, say, electron tunnelling microscope.

These ideas are pretty much in line with the state of the art knowledge in chemical physics and modern positions of some leading physicists on 'classical' Quantum Mechanics [23]. For example, from my discussions with Prof Werner Hofer of Newcastle University, I came to the understanding that electron is a portion of space, surrounding the nucleus of an atom, which has trapped energy current, pretty much analogous to a capacitor!

4. Mathematical models for energy-modulated computing

(a) Modelling Wakefield experiment in Petri-nets

In this section, I would like to make a brief excurse to a computational model called Petri nets. The name that this model bears is attributed to Prof. Carl Adam Petri, whose work [24] on causal models of information processing and communication using a graphical formalism of condition-event nets in the 50s and early 60s had led to a significant amount of research and applications in computer science and beyond. Petri nets have a very clear distinction between locality and nonlocality. In fact, in Petri nets, one can express cause–effect relationship between discrete events in a very succinct form, even if one has many events occurring concurrently and independently in space. I spent nearly 40 years studying different types of Petri nets and their semantics and applying Petri nets to the analysis and design of electronic systems.

What is a Petri net? It is a bipartite directed graph [25] with nodes being either places (denoted by circles) or transitions (denoted by bars). Sometimes places are also called conditions and transitions are called events. The arcs of the graph, called a flow relation, can only go from places to transitions and transitions to places. These arcs define the causality relation between transitions, i.e. events, via places. Places, transitions and arcs define the structure of the net. A Petri net can produce a dynamic behaviour, which is defined by the initial position of token in some places (called initial marking or initial state) and some rules of a 'token game', i.e. the method according to which the current marking can change into the following marking and thus produce sequences of markings. The token game has two main rules. Rule 1 (Enabling Rule) defines the condition under which a transition is enabled. Namely, a transition is enabled if all its input places (places that are connected to the transition by arcs going from the places to the transition) contain tokens. Every transition that is enabled under a given marking can fire. Rule 2 (Firing Rule) says that when a transition fires, we have to remove a token from each input place and add a token to each output place (such places are connected by arcs going from the transition to places). On the basis of these rules, one can perform exploration of the reachable markings (called reachability analysis). This exploration, done systematically and exhaustively, requires performing algorithmic search techniques, such as depth-first search or breadth-first search. The size of the reachable state space may grow exponentially if the Petri net has transitions that are enabled *concurrently* in the same marking, because all possible interleavings of firing such transitions need to be explored. The important advantage of Petri nets themselves as a model for representing distributed concurrent systems is their compactness. On the contrary, the reachable state space represented by the so-called reachability graph would have all concurrency captured in the form of interleaving sequences, and hence suffers from complexity burst.

We shall avoid discussing formal definitions of the Petri net structure and behaviour (there are several semantics of concurrency and choice that can be associated with this behaviour). Instead, we shall simply proceed to the use of Petri nets for the modelling of the distributed system associated with the TL and energy current moving in it taking place in the Wakefield experiment [11]. The system previously shown in figure 5 is now modelled by the Petri net in figure 12. This net captures the energy current by tokens in the upper and lower threads—they are called 'energy tokens'. These tokens move in the upper subnet right-to-left and in the lower subnet left-to-right to demonstrate the effect of the direct (from the switch to the far end) and reflected TEM waves (the wave that moves to the right will again be reflected back to form another direct wave after some loss of energy into the subnet modelling the terminating resistor). Specific transitions can be associated with signal events in particular locations of the cable. Namely, transitions 1 and 9 stand for events in point A, 2 and 8 in B, 3 and 7 in C, 4 and 6 in D and 5 in E. The Petri net of such kind is usually called a 'ring pipeline net'. Ring pipeline conveniently shows the cyclic rotation of the energy tokens in the cable medium.

Note that this net can operate cyclically without any outgoing tokens to the resistor while the 'control token' initially sits in the place labelled 'a'. This place enables transition 1, which helps to sustain cyclicity in the process. This cyclic behaviour does not change the potential of the cable as

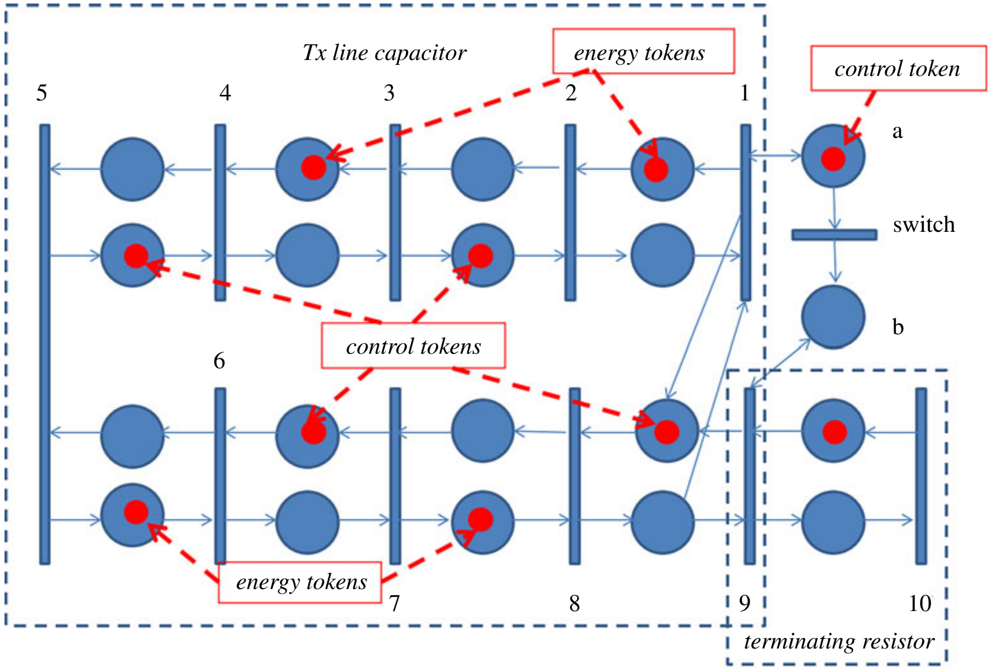


Figure 12. Petri net model of a transmission line with a switch and terminating resistor (Wakefield experiment).

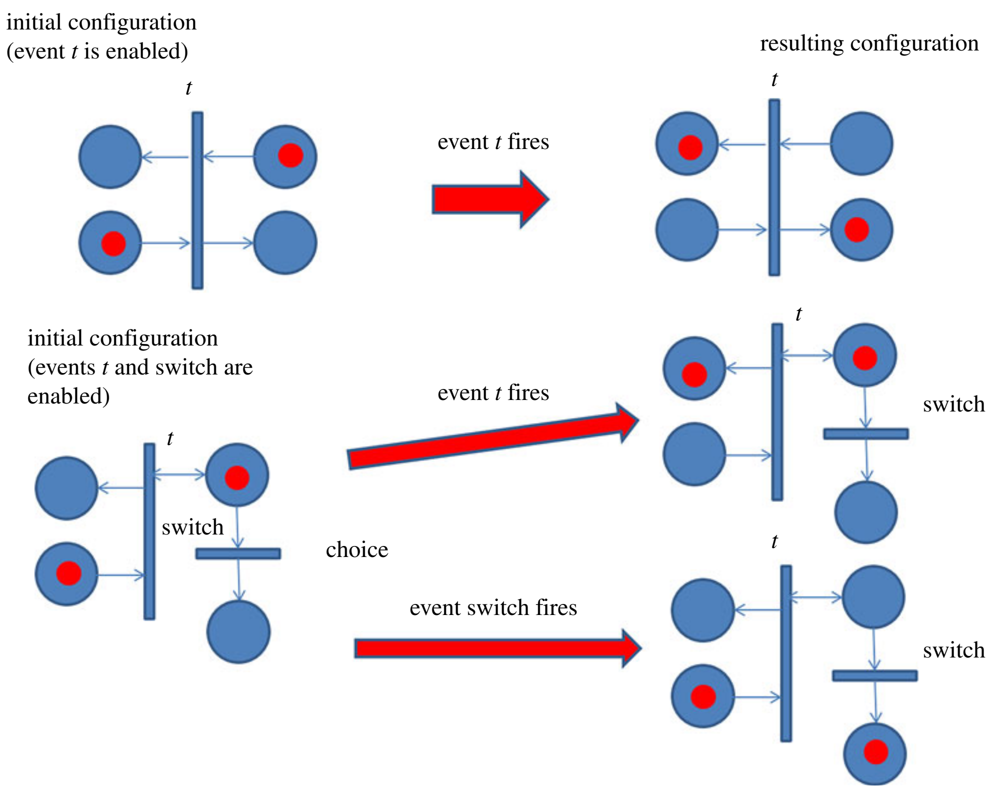


Figure 13. Illustration of Firing Rule for transitions in the Petri net of figure 12.

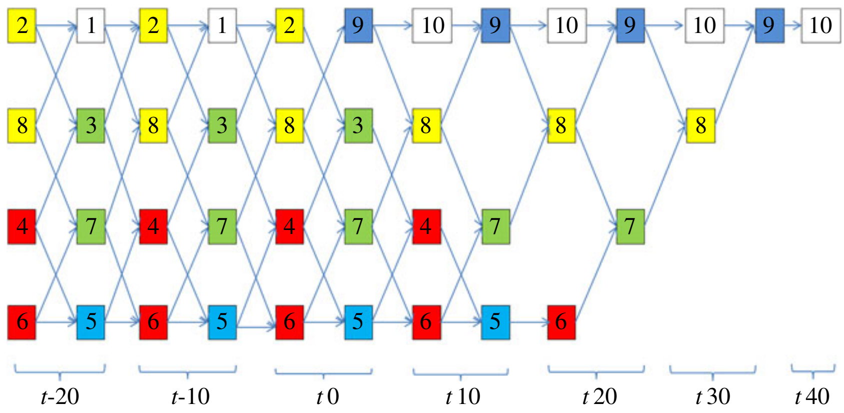


Figure 14. Petri net behaviour unfolding.

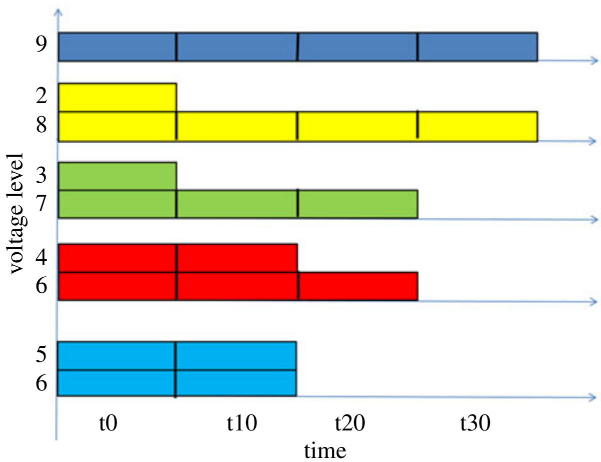


Figure 15. Waveform interpretation of the Petri net behaviour (cf. figure 6).

it stays charged at 9 V. As soon as the switch is operated, the 'control token' moves from place to 'a' to place 'b', and this activates the transition 9. When the control token is in place 'b' transition 1 cannot be enabled anymore because there is no more token in place 'a'. The firing rules for the main transition types used in this net are illustrated in figure 13.

The behaviour of the Petri net model and its interpretation in terms of the states of the signal in the five main points of the cable can be analysed using an acyclic unfolding of the Petri net, whose fragment is shown in figure 14.

The initial segment of the unfolding is labelled with time stamps *t*-20 and *t*-10, which correspond to the time before the switch is ON. This shows the repetitive operation of pairs of transitions associated with points A (1), B (2 and 8), C (3 and 7), D (4 and 6) and E(5). Not all of these transitions are concurrent (this is because of the way we represented the pipeline—we alternate transitions passing the energy and control tokens—in principle we could show a denser pipeline but this would have required using some dummy events), but all of them fire in two adjacent batches at t-20 and t-10, which models the passage of the two cycles of rotation of energy current in the cable. At time stamp t0, we now have the first occurrence of event 9 (no longer 1) to represent action in point A. We can now trace the section of the unfolding which is marked by the trace of events: 2, 3, 4, 5 (direct wave), 6, 7, 8, 9 (reflected wave) taking place in the steps t0, t10, t20 and t30. The end of the signal propagation is at time t40, where transition 10 fires indicating that the initial energy level has left the cable through the terminating resistor. The Petri net unfolding can be interpreted as a waveform (with the colours of the corresponding transitions in the net), shown in figure 15. There is a distinct similarity with the waveforms from the scope in the Wakefield experiment shown in figure 6. Some discrepancy is caused by a bit coarse level of granularity with which we modelled the pipeline of the signals in the cable. Increasing granularity of modelling, i.e. discretizing the space and using more transitions would have improved the matching effect. But this would have cluttered the diagrams, so we avoided that here.

Petri nets are generally a good model to model repetitive group behaviour which was explored throughout many years by Carl Adam Petri. In the recent paper by Prof. Rudiger Valk [26], Petri's cycloids are introduced with a concept of slowness. As the paper claims, this approach can be applied to Organization, to Work Flow (Just-in-time Production), and to Physical Systems. It would be interesting to investigate how Petri nets and cycloids can capture such behaviour for example as propagations of slivers of pulses in groups of TLs.

5. Conclusion

More than 125 years ago Oliver Heaviside stated that energy current was the primal standpoint. In this paper, we looked at the potential impact of the idea of energy current on the connection between electromagnetic theory and computing. This connection is manifold. It permeates through the notion of energy-modulated computing. It also drives the research into computing which is based on physical phenomena such as causality and encourages the engineers to develop or use the 'right kind' of mathematics to build the bridge between the behaviour of signals in physics and exploiting this behaviour in computations. The bridge between the physics of electromagnetism and computing fundamentally lies in Time domain analysis and appropriate forms of discretization of processes in space and time (cf. geometric approach of Galileo and Newton [27]). Immediate switching to Frequency domain analysis for pulse-based signals (and this is what we deal with in computers!) would bring a 'wrong type' of mathematics on the way of physics and reality. This sounds controversial but this is what we could and should learn from Heaviside.

What about more specific methodological innovation of this paper? We have now explored two types of step-wise physical processes that we can link with computing. One is associated with energy-current—this is a fast computing paradigm associated with the speed of light. An example is the energy-division in TLs—here we can form oscillations at super-Gigahertz frequencies on a chip. Another form is associated with the switching of logic gates, where we rely on mass effects such as movement of charge, and division of electrical energy associated with it. This is illustrated by the capacitor discharge via digital switching logic. Here our typical speeds are sub-Gigahertz. These two forms are orthogonal but can work together, for example in a nested way, like the second and minute hands of the clock. We could combine the TL discharge (step-wise discretization of an exponential—inner loop) with a logic circuit switching (step-wise discretization of hyperbolic discharge—outer loop).

This is a conjecture with which I conclude this paper. It is based on the stepwise process of TL models of capacitors by Ivor Catt and his associates and our stepwise processes with a ring oscillator discharging a capacitor, even a lumped one. These are two orthogonal discretization operators. The study of their superposition is a subject of our future work. This will open up some new dimensions for energy-modulated computing!

Besides, a potentially useful result of this paper in terms of modelling is the fact that Petri net unfolding can be interpreted as a waveform of signals whose states are associated with some places in the net.

Data accessibility. This article has no additional data.
Competing interests. I declare I have no competing interests.
Funding. I received no funding for this study.

Acknowledgements. I am grateful to Dr Christopher Donaghy-Spargo of Durham University for numerous stimulating interactions on the life and work of Oliver Heaviside and our regular chats about electromagnetism. I would also like to thank Mr Ivor Catt and Dr David Walton for opening the world of electromagnetics to me in 2013 through the prism of Catt Theory. I am also indebted to my research group at Newcastle, with whom we are exploring the arcades of energy-modulated computing. Last but not least, many thanks to the two anonymous reviewers for their thorough reading of the paper and producing invaluable comments and corrections.

References

1. Yakovlev A. 2011 Energy-modulated computing. In *Proc. Design Automation & Test in Europe (DATE), Grenoble*, pp. 1–6. Piscataway, NJ: IEEE. (doi:10.1109/DATE.2011.5763216)
2. Yakovlev A, Ramezani R, Mak T. 2011 *Apparatus and method for voltage sensing*. US Patent application US13638330, 28.02.2011, granted US9121871B2 01.09.
3. Ramezani R, Yakovlev A. 2013 Capacitor discharging through asynchronous circuit switching. In *Proc. IEEE 19th Int. Symp. on Asynchronous Circuits and Systems, Santa Monica, CA*, pp. 16–22. Piscataway, NJ: IEEE. (doi:10.1109/ASYNC.2013.11)
4. Kinniment DJ. 2008 *Synchronization and arbitration in digital systems*. Chichester, UK: Wiley Publishing.
5. Kinniment DJ. 2011 *He who hesitates is lost*. Newcastle upon Tyne, UK: Newcastle University, School of Engineering.
6. Catt I. 1966 Time loss through gating of asynchronous logic signal pulses. *IEEE Trans. Electron. Comput.* **EC-15**, 108–111. (doi:10.1109/PGEC.1966.264407)
7. Catt I. 1995 *Electromagnetics 1*. St Albans, UK: Westfields Press.
8. Catt I, Davidson MF, Walton DS. 1978 Displacement current – and how to get rid of it. In *Wireless World*, pp. 51–52.
9. Catt I. 2013 Advances in electromagnetic theory. In *Video recording of the Electromagnetism Seminar at Newcastle University, 9 October 2013*. See http://async.org.uk/IvorCatt+DavidWalton.html.
10. Walton D. 2013 Maxwell's electromagnetic theory. In *Video recording of the Electromagnetism Seminar at Newcastle University, 9 October 2013*. See http://async.org.uk/IvorCatt+DavidWalton.html.
11. Catt I. 2013 The end of the road. *Electronics World, The Centenary Issue* **119**, 72–74. See https://www.electronicsworld.co.uk/magazine/centenary-issue.
12. Heaviside O. 1892 *Electrical papers*, vol. I. London: Macmillan and Co., p. 438.
13. Shafik R, Yakovlev A, Das S. 2018 Real-power computing. *IEEE Trans. Comput.* **67**, 1445–1461. Early Access. (doi:10.1109/TC.2018.2822697)
14. Yakovlev A, Vivet P, Renaudin M. 2013 Advances in asynchronous logic: from principles to GALS & NoC, recent industry applications, and commercial CAD tools. In *Proc. Design Automation & Test in Europe (DATE), Grenoble, France*, pp. 1715–1724. (doi:10.7873/DATE.2013.346)
15. Heaviside O. 1892. *Electrical papers*, vol. II. London: Macmillan and Co., p. 201.
16. Catt I, Davidson MF, Walton DS. 1979 *Digital hardware design*. London, UK: The Macmillan Press Ltd.
17. Sullivan CR, Kern AM. 2001 Capacitors with fast current require distributed models. In *Proc. IEEE 32nd Annu. Power Electronics Specialists Conf., Vancouver, BC*, pp. 1497–1503. Piscataway, NJ: IEEE. (doi:10.1109/PESC.2001.954331)
18. Yakovlev A, Kushnerov A, Mokhov A, Ramezani R. 2014 On hyperbolic laws of capacitor discharge through self-timed digital loads. *Int. J. Circ. Theor. Appl.* **43**, 1243–1262. (doi:10.1002/cta.2010)
19. Wang A, Calhoun BH, Chandrakasan AP. 2006 *Subthreshold design for ultra-low power systems*. Berlin, Germany: Springer Verlag.
20. Nose K, Sakurai T. 2000 Optimization of VDD and VTH for low-power and high-speed applications. In *Proc. Asia and South Pacific Design Automation Conf. (ASP-DAC), Yokohama, Japan*, pp. 469–474. New York, NY: ACM (doi:10.1145/368434.368755)
21. Xu Y, Shang D, Xia F, Yakovlev A. 2016 A smart all-digital charge to digital converter. In *Proc. IEEE Int. Conf. on Electronics, Circuits and Systems (ICECS), Monte Carlo*, pp. 668–671. Piscataway, NJ: IEEE. (doi:10.1109/ICECS.2016.7841290)

22. Einstein A, Podolsky B, Rosen N. 1935 Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.* **47**, 777–780. (doi:10.1103/PhysRev.47.777)
23. Hofer W. 2018 *Is quantum mechanics creationism, and not science?* ArXiv e-prints, February 2018. See http://adsabs.harvard.edu/abs/2018arXiv180200227H
24. Petri CA. 1966 Communication with automata. Dissertation, University of Hamburg. See http://edoc.sub.uni-hamburg.de/informatik/volltexte/2010/155/
25. Murata T. 1989 Petri nets: properties, analysis and applications. *Proc. IEEE* **77**, 541–580. (doi:10.1109/5.24143)
26. Falk R. 2018 On the structure of cycloids introduced by Carl Adam Petri. In *Proc. 39th Int. Conf. on Applications and Theory of Petri nets and Concurrency*, *Bratislava, Slovakia*.
27. Dellian E. 1989 Newton, die Traegheitskraft und die absolute Bewegung. *Philosophia Naturalisz* **26**, 192–201. See http://www.neutonus-reformatus.de/frameset.html

PHILOSOPHICAL TRANSACTIONS A

royalsocietypublishing.org/journal/rsta

Correction

Cite this article: Yakovlev A. 2019 Correction to 'Energy current and computing'. *Phil. Trans. R. Soc. A* **377**: 20180351.
http://dx.doi.org/10.1098/rsta.2018.0351

Correction to 'Energy current and computing'

Alex Yakovlev

Phil. Trans. R. Soc. A **376**, 20170449. (Published online 29 October 2018). (doi:10.1098/rsta.2017.0449)

AY, 0000-0003-0826-9330

In figure 6, there is an error with one of the waveforms. The correct version of figure 6 is below.

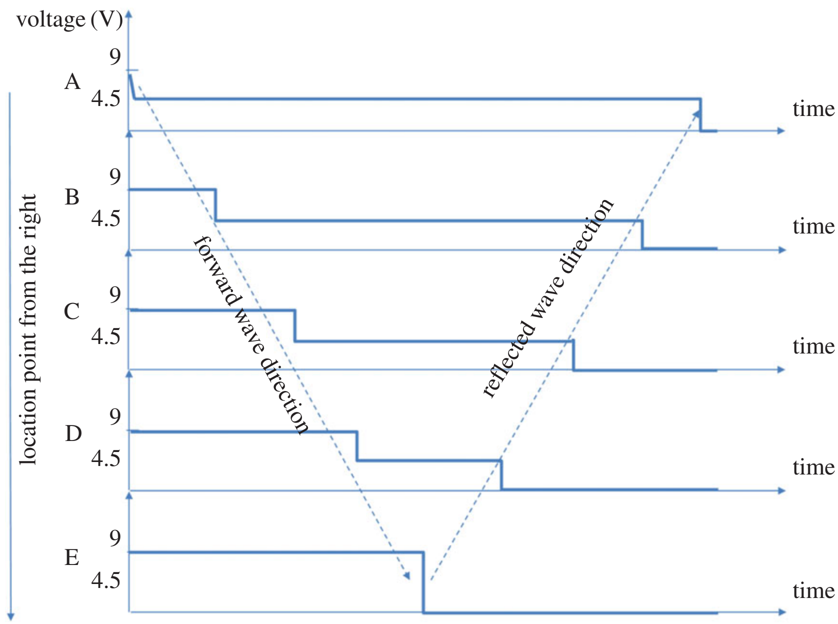


Figure 6. Signal plots for the Wakefield experiment, in five different locations.

THE ROYAL SOCIETY
PUBLISHING

PHILOSOPHICAL TRANSACTIONS A

rsta.royalsocietypublishing.org

Research

Cite this article: Gradoni G, Russer J, Baharuddin MH, Haider M, Russer P, Smartt C, Creagh SC, Tanner G, Thomas DWP. 2018 Stochastic electromagnetic field propagation— measurement and modelling. *Phil. Trans. R. Soc. A* **376**: 20170455. http://dx.doi.org/10.1098/rsta.2017.0455

Accepted: 27 July 2018

One contribution of 13 to a theme issue 'Celebrating 125 years of Oliver Heaviside's 'Electromagnetic Theory''.

Subject Areas:
electrical engineering, electromagnetism

Keywords:
electromagnetic fields, stochastic fields, propagation, Heaviside

Author for correspondence:
Gabriele Gradoni
e-mail: gabriele.gradoni@nottingham.ac.uk

Stochastic electromagnetic field propagation— measurement and modelling

Gabriele Gradoni[1,2], Johannes Russer[3], Mohd Hafiz Baharuddin[2,4], Michael Haider[3], Peter Russer[3], Christopher Smartt[2], Stephen C. Creagh[1], Gregor Tanner[1] and David W. P. Thomas[2]

[1]School of Mathematical Sciences, and [2]George Green Institute for Electromagnetics Research, University of Nottingham, University Park, Nottingham NG7 2RD, UK

[3]Institute of Nanoelectronics, Technical University Munich, Munich, Germany

[4]Centre of Advanced Electronic and Communication Engineering, Universiti Kebangsaan Malaysia, Bangi 43600, Malaysia

GG, 0000-0001-8321-6883; DWPT, 0000-0002-1815-8477

This paper reviews recent progress in the measurement and modelling of stochastic electromagnetic fields, focusing on propagation approaches based on Wigner functions and the method of moments technique. The respective propagation methods are exemplified by application to measurements of electromagnetic emissions from a stirred, cavity-backed aperture. We discuss early elements of statistical electromagnetics in Heaviside's papers, driven mainly by an analogy of electromagnetic wave propagation with heat transfer. These ideas include concepts of momentum and directionality in the realm of propagation through confined media with irregular boundaries. We then review and extend concepts using Wigner functions to propagate the statistical properties of electromagnetic fields. We discuss in particular how to include polarization in this formalism leading to a Wigner tensor formulation and a relation to an averaged Poynting vector.

THE ROYAL SOCIETY PUBLISHING

This article is part of the theme issue 'Celebrating 125 years of Oliver Heaviside's 'Electromagnetic Theory''.

1. Introduction

Oliver Heaviside was a gifted engineer and a tenacious supporter of the *treatise* of J. C. Maxwell [1]. Along with George Francis FitzGerald and Oliver Lodge, he was part of a group of scientists later named 'The Maxwellians', who worked to develop an electromagnetic (EM) theory from Maxwell's explanation of Hertz' experiments. Heaviside became known as the idiosyncratic genius [2] and maverick mind [3] of electricity for his practical intuition in understanding the wave guidance for wireless telegraphy, predicting a conducting layer in the atmosphere, later named after him and Arthur Edwin Kennelly, who formulated this idea independently at the same time [4]. The feasibility of long-distance wireless telegraphy had earlier been doubted by many scientists. Guglielmo Marconi pioneered wireless telegraphy and accomplished the first transatlantic signal transmission in 1902 using coupled resonant circuits as developed by Ferdinand Braun [5,6]. In 1909, Ferdinand Braun and Guglielmo Marconi jointly received the Nobel Prize for their groundbreaking contributions to wireless telegraphy. Besides his forward-looking activity in EM phenomenology, Heaviside gained a reputation as an important mathematician when he argued in favour of the use of vectors over quaternions, emerging in mathematical physics through the Cambridge fashion for Lagrangians [7]. Heaviside invented the vector notation to bring Maxwell's views closer to the more practical attitude of telegraphers as well as to relate the underlying symbols to intuitive geometrical operations [8, 5.5.3]. Heaviside's vector notation for Maxwell's equations has become standard, providing a balance in complexity and intuitiveness which makes it appealing for engineering. Further frameworks have been developed and applied to EM theory in the meantime which, unlike Maxwell's original formulation, provide a more intuitive insight such as the framework based on exterior differential forms [9,10]. Besides Heaviside's work on the generation and motion of wavefronts in open media and telegraphic lines, he soon focused on the definition of energy flux in electromagnetism. Thanks to his—geometry-based—vectorial notation, he derived an expression for $\mathbf{E} \times \mathbf{H}$ that made it clear that this quantity is related to the transport of energy. He achieved this result independently of Poynting, and even generalized it to include displacement currents [8, 5.5.3 and 5.5.4, pp. 196–199].

There is therefore a tradition in Maxwellian electromagnetics to reconcile theories with concepts involving the definition of EM energy and its transport vector. We shall take up this tradition, using the established Heaviside vector notation to extend the characterization of random scalar wave-fields to stochastic vector EM fields exploring the connection between correlation functions (CFs) and the Poynting vector. Furthermore, the use of Wigner functions is introduced to preserve the concept of directionality thus building on Heaviside's intuition finding regularities in the presence of complex field distributions. The Wigner function (WF) has originated in quantum mechanics [11,12] and has since found widespread applications in fields such as optics [13–15] and radio science [16].

Today, with the principles and models of electromagnetism well established, it is easy to forget the tribulations and controversy that the Maxwellians went through in the early days. One of the first episodes of this kind concerned MacCullagh's geometrical theory of optics and his formulation in terms of a quadratic Lagrangian describing the EM waves within the aether [17, p. 9], a theory dissmissed by Thomson and Stokes. Elastic theories of an aether based on Green's dynamical models were better regarded at that time in England. Stokes stated that MacCullagh's theory violated the conservation of angular momentum and so it was dynamically inadmissible [17, p. 10]. FitzGerald, who studied under MacCullagh at Trinity College Dublin, was reluctant to accept Stokes' demolition of his alternative theory and put a serious effort into understanding Maxwell's *treatise*. Just when Stokes proved that Maxwell's theory was inadmissible as an elastic theory, FitzGerald concluded that Maxwell's theory was identical to MacCullagh's theory. But the acceptance of Maxwell's theory in the scientific community was still all but certain; this changed only when a young, self-taught, British telegrapher named Oliver Heaviside crossed FitzGerald's path to reshape the entire subject of electromagnetics. Heaviside, who never held an academic position and left school at 16 to work for a telegraphic company, dedicated his early efforts to the motion of EM waves in cables. Later, the generation and motion of EM waves in aether became a central theme in his studies. Inherently, Heaviside considered the concepts of electric and magnetic forces and fluxes central to the theory and preferred them over potentials. This was a key intuition in reducing the long list of Maxwell's original equations to a few equations constituting the *Maxwell's equations* of the EM theory as we know them today. Heaviside's interests were far-reaching to the point that he considered more complex motion scenarios besides guided and free waves, including the presence of nearby objects as well as wave confinement in enclosures. We will in the following analyse a small part of volume 1 of the *Electromagnetic Theory* series [7], where Heaviside mentions the *formation of a new regularity in the mean from complex fields in irregular geometries*. This is a clear precursor of the field of statistical electromagnetics, which is today pervasive in practical electromagnetic compatibility (EMC) and wireless communications scenarios. Importantly, Heaviside mentions the possibility of complex field distributions resulting from irregular sources, again, of paramount importance in modern EMC studies concerning the characterization of stochastic emissions from multifunctional digital electronics.

Modelling stochastic EM fields is necessary for improving the design of electronic devices, printed circuit boards (PCBs) and electronic systems taking into consideration their susceptibility to electromagnetic interference (EMI). Radiated EMI is caused by fast transients due to switching and information transfer processes within the electronic device. From an EMC perspective, these processes can be considered as being quasi-random, i.e. noisy. Power levels of the radiated field are in general low and emissions are spatially distributed over the PCB, while various hot spots may be identified. A widely used technique for characterizing emissions is near-field scanning (NFS) because of its high measurement accuracy and reliability [18–20]. This allows for using a canonical method in EM theory: by considering a source distribution enclosed within a virtual surface, one can compute the EM field in a source-free region outside this surface from amplitude and phase of the tangential field components on this surface by using Huygens' principle. For stochastic fields, however, numerical values of noise amplitudes cannot be specified by this method. On the circuit level, noisy signals can be described by energy and power spectra [21] and noisy linear circuits can be modelled numerically by correlation matrix-based methods [22–24]. In addition, network methods can be adapted for numerical modelling of EM fields [25] efficiently. Correlation matrix-based methods have been expanded in [26,27] to model noisy EM fields.

Another approach to modelling the propagation of field CFs is obtained by using a representation of the field in the combined space of position and wavevectors. The process is carried out through a representation known as the WF method [15], from which a ray-tracing-based approximation of the propagated CF can be derived [28,29]. Furthermore, in the propagation of CFs calculated from NFS of noisy fields, the connection between CFs and phase-space distributions to extract explicit directional information from the measurement process can be exploited. Interestingly, the method is able to include the transport of evanescent wave CFs as a leading order approximation of phase-space diffusion [30].

The concept of directionality was stressed by Heaviside when he was considering reflections from irregular geometries, envisaging new regularity for mean values for highly irregular fields. Representing EM fields in phase space goes beyond Heaviside's insight and leads to new insight in the form of finding the universal behaviour of completely randomized fields and including partial positional and directional information for geometries that are only partially irregular [31]. So far, we have treated the propagation of CFs of scalar field components in free space [29] and confined space [32]. In this paper, we discuss extensions of this formalism to include vector fields, which naturally leads to a Wigner tensor (WT). In previous work, predictions based on this WF approach have been verified in both the far field [29] and the near field [33].

NFS for field correlations in stochastic EM fields requires at least two field probes and has been addressed in [34–37]. The method has been validated experimentally in the context of cavity-backed apertures [29], Arduino printed circuit boards (PCBs) [38,39], and equipment-level enclosures [30].

2. From deterministic to statistical electromagnetic theory

Heaviside's masterpiece *Electromagnetic Theory* [7,40,41] contains a thorough analysis of two important properties of the electromagnetic wave (EMW):

— The electric and magnetic vector forces are subject to an effect of self-induction. Stress is made on this effect in pure (plane) EM waves, where *electric and magnetic forces have a constant ratio, electric and magnetic energies are equal, they have the property of being perpendicular to one another [...] and their plane is in the wavefront, or the direction of motion of waves is perpendicular to* ***E*** *and to* ***H****. It is the direction of the flux of energy*.

— The EMW is constituted by force fields best described by *vectors*. Of particular importance is the introduction of the vector product as a tool for unveiling the principle of the *continuity of the energy flux*. It is remarked that *the principle of the continuity of energy is a special form of that of its conservation*, or Newton's *principle of the conservation or persistence of energy*. However, *in the ordinary understanding of the conservation principle, it is the integral amount of energy that is conserved, and nothing is said about its distribution or its motion. This involves continuity of existence in time, but not necessarily in space also. But if we can localise energy definitely in space, then we are bound to ask how energy gets from place to place.* As we will see later in the paper, the concept of lines guiding the energy flux was introduced by Heaviside independently from J. H. Poynting, and using the vector EMW notation.

Heaviside's perspective allowed for reducing the numerous equations introduced by Maxwell into four compact vector equations, as they are known today. An energy-based interpretation of EMW motion emerges by picturing its continuous flux that traverses a region of space from source to receiver. This interpretation is fully supported by mathematical tools from vector calculus, as opposed to quaternions put forward by Hamilton [7], and finds a direct connection to the directional energy flux analysis performed by Poynting. Inherently, it is the physics-based use of electric and magnetic field products connected to the energy flow that introduces the concept of directionality when waves transition through an unbounded medium. In this unique and imaginative reading of a new form of motion, inspired by cable studies of EMW, and consolidated by the tight interaction between the Maxwellians, modern EM theory finds its origins.

Heaviside was also keen to understand the diffusive nature of EMW energy across diverse media. Volume I of [7] contains a detailed discussion on the nature of EM wave transmission in analogy with diffusion processes in thermodynamics and mechanics. Special attention is devoted to the self-interaction mechanism between electric and magnetic fields, which makes up the propagation mechanism and determines its faster occurrence as compared to fluids.

It is probably the marriage between self-interaction and diffusion that lead Heaviside to think about the mechanism of EMW energy confinement within enclosed reflecting surfaces. The description of such a scenario is peculiar and deserves special attention as it is most likely related to a less known appetite of Heaviside for complex EMW distribution in irregular geometries underpinned by statistically isotropic waves giving rise to new forms of regularities, the *regularity of the mean*.

(a) Characterization of stochastic sources

Electric current densities (or *Gaussage*) flowing through surfaces or volumes become a source of EMW that propagates in the surrounding space. Many of the ideas used below to predict such propagation in the context of complex or stochastic fields find precursors already in the work of Heaviside.

In Heaviside's idea of source operation, a central goal is *to find the form of the wavefront due to any collection of point-sources and trace the changes of its shape and position as it progresses* [7]. More specifically, he states that *when point sources are spread continuously over a surface to form a surface-source, we have a continuous wavefront from the first moment; or rather, two wavefronts, one on each side of the surface*. He continues by pointing out that *when we have once got a wavefront we may ignore the sources which produced it and make the wavefront itself tell us what its subsequent history will be*. Unilateral approximation and wavefront reconstruction is what can be achieved with the NFS method that will be explained later. The distinction between using the source currents explicitly, or ignoring them in favour of working directly with field measurements, is reflected in the two approaches to stochastic field propagation described in §2b.

Heaviside did not exclude the possibility of having, similarly to thermal diffusion, a diffusion of wavefronts originating from arbitrary sources in confined geometries and leading to complex EMW distributions. He realized that any localized perturbation will cause a distortion in both the space and time motion of waves. This perturbation can be described and predicted mathematically for elementary (point-)sources. When the source is extended and supports a complicated spatial current pattern, it can be described statistically, as depicted later in this section.

A natural step to take from the concept of energy defined by Heaviside is to consider a mean energy flux and Poynting vector through products of local, simultaneous fields. It can be shown that joint positional and directional information can be extracted from products of non-local and non-simultaneous fields. Discretely or continuously spread sources that are spatially extended, complex/irregular in their geometry, and stochastic in time, radiate wavefronts that have an irregular and stochastic structure. From any resultant (vector) field $\mathbf{F}(\mathbf{x}, z; t)$, taking the product of fields and averaging in time to extract the typical field behaviour, leads to the *field correlation tensor (dyadic)*, that is,

$$\begin{aligned}\mathcal{C}_z(\mathbf{x}_a, \mathbf{x}_b; \tau) &= \langle \mathbf{F}(\mathbf{x}_a, z; t+\tau)\, \mathbf{F}(\mathbf{x}_b, z; t)\rangle \\ &= \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} \mathbf{F}(\mathbf{x}_a, z; t+\tau)\, \mathbf{F}(\mathbf{x}_b, z; t)\, \mathrm{d}t,\end{aligned} \tag{2.1}$$

where the dyadic product is denoted $\mathbf{FF} = \mathbf{FF}^H = \mathbf{F} \otimes \mathbf{F}$, with superscript H meaning conjugate transpose, while $\mathbf{x}_a$ and $\mathbf{x}_b$ are the (two-dimensional) transverse coordinates within planes parallel to the source and perpendicular to z, the outward normal to the source. Then using space–time stochastic magnetic fields yields $\mathbf{F} \equiv \mathbf{H}$, while starting from the electric field yields $\mathbf{F} \equiv \mathbf{E}$. In a Cartesian frame of reference $\mathbf{F} = [F_x, F_y, F_z]$ the structure in (2.1) reads

$$\mathcal{C}_z(\mathbf{x}_a, \mathbf{x}_b; \tau) = \begin{bmatrix} C_z^{xx} & C_z^{xy} & C_z^{xz} \\ C_z^{yx} & C_z^{yy} & C_z^{yz} \\ C_z^{zx} & C_z^{zy} & C_z^{zz} \end{bmatrix}, \tag{2.2}$$

where the dependence on $\mathbf{x}_a$ and $\mathbf{x}_b$ and on τ in the entries has been omitted for compactness. Individual components of the correlation tensor (CT) (2.1) have been used in previous work on radiated emission in EMC, particularly using the experimental access to magnetic field components that is viable with commercial loop probes [29]. In the next section, we will describe how magnetic fields can be extracted accurately through scanning measurements in close proximity to the source. The experimental layout here gives direct access to the magnetic field components parallel to the source, H_x and H_y, while the normal component H_z can be obtained both from the Gauss law for magnetism and from Faraday's law of induction [7]. An application of Love's equivalence principle has been proposed in [42], where the concept arises of a Huygens box embracing a radiative emission source for EMC tests. Only tangential magnetic fields on the box surface are needed to find the radiated emissions from the source by defining the equivalent surface current density $\mathbf{J}_s = \mathbf{n} \times \mathbf{H}$, with $\mathbf{n}$ the outward normal to the surface of the Huygens box. Then curl equations with the forcing term $\mathbf{J}_s$ are used to calculate the tangential components of the electric field. Alternatively to measuring magnetic fields, experimental set-ups may also measure electric fields. Notable examples are the sleeve dipole used at UMIST [42] and the atomic field probe recently conceived at NIST [43]. Therefore, defining the coherence tensor in terms of either electric or magnetic fields allows one to create a model of the source based on laboratory measurements. Combining electric and magnetic fields, it is also possible, in principle, to form hybrid *magneto-electric* CFs. In fact, any combination for which there is an explicit relation between the resulting coherence tensor and the Poynting vector, which describes the local direction of the energy flow [44], allows propagation of emissions from the source. We illustrate this explicitly later in this section, by showing how the coherence tensor can be used to devise a wave-dynamical phase-space representation from which both the energy flow vector and the local energy density can be extracted self-consistently.

In the frequency domain, this field–field CT is represented by the Fourier transform

$$\Gamma_z\left(\mathbf{x}_a, \mathbf{x}_b; \omega\right) = \int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}\omega\tau} \mathcal{C}_z\left(\mathbf{x}_a, \mathbf{x}_b; \tau\right)\, \mathrm{d}\tau. \tag{2.3}$$

The averaging in (2.1) produces a function that is independent of the time reference in the case of stationary stochastic fields. Hidden in the CF is significant information on the physical structure of the random emissions. However, the spatial representation (2.3) does not give simple access to this in terms of a clean physical interpretation. Following arguments similar to those employed in semiclassical quantum mechanics for the single-particle density function [13], it can be argued that both positional and directional properties can be extracted from the CT through the WT [28].

The WT has a direct connection with the CT and transforms its entries to WFs that represent vector wave components' functions on phase space, combining both position and direction of propagation. More precisely, the WT is defined as

$$\mathcal{W}_z\left(\mathbf{x}, \mathbf{p}\right) = \int \mathrm{e}^{-\mathrm{i}k\mathbf{p}\cdot\mathbf{s}} \Gamma_z\left(\mathbf{x} + \frac{\mathbf{s}}{2}, \mathbf{x} - \frac{\mathbf{s}}{2}\right) \mathrm{d}\mathbf{s}, \tag{2.4}$$

where k is the constant wavenumber coordinates $(\mathbf{x}, \mathbf{s})$ related to $(\mathbf{x}_a, \mathbf{x}_b)$ by the transformation

$$\left.\begin{aligned} \mathbf{x} &= (\mathbf{x}_a + \mathbf{x}_b)/2, \\ \mathbf{s} &= \mathbf{x}_a - \mathbf{x}_b, \end{aligned}\right\} \tag{2.5}$$

so that $\mathbf{x}$ is the average position and $\mathbf{s}$ is the difference in positions of a pair of measured fields. More explicitly, $\mathbf{s} = (s_x, s_y)$ represents, in the NFS of planar sources, the in-plane displacement (for fixed z) between measurement positions. The conjugate momentum vector $\mathbf{p} = (p_x, p_y)$ in (2.4) takes the geometrical meaning of the components of the wavevector parallel to the source plane, normalized so that

$$p_x = \sin\theta\cos\phi, \tag{2.6}$$

$$p_y = \sin\theta\sin\phi, \tag{2.7}$$

and $|\mathbf{p}| = \sin\theta$, where θ is the angle of the ray with respect to the outward normal.

A useful property of the WF is that it allows for treating position and momentum variables symmetrically, so that (2.4) can also be obtained from

$$\mathcal{W}_z\left(\mathbf{x}, \mathbf{p}\right) = \left(\frac{k}{2\pi}\right)^2 \int \mathrm{e}^{\mathrm{i}k\mathbf{x}\cdot\mathbf{q}} \tilde{\Gamma}_z\left(\mathbf{p} + \frac{\mathbf{q}}{2}, \mathbf{p} - \frac{\mathbf{q}}{2}\right) \mathrm{d}\mathbf{q}, \tag{2.8}$$

where coordinates $(\mathbf{p}, \mathbf{q})$ are similarly obtained from $(\mathbf{p}_a, \mathbf{p}_b)$ through the rotation

$$\left.\begin{aligned} \mathbf{p} &= (\mathbf{p}_a + \mathbf{p}_b)/2, \\ \mathbf{q} &= \mathbf{p}_a - \mathbf{p}_b. \end{aligned}\right\} \tag{2.9}$$

The momentum representation of the CF (2.3) is given by the double partial Fourier transform, defined in the frequency domain by

$$\tilde{\Gamma}_z(\mathbf{p}_a, \mathbf{p}_b) = \iint \mathrm{e}^{-\mathrm{i}k(\mathbf{p}_a\cdot\mathbf{x}_a - \mathbf{p}_b\cdot\mathbf{x}_b)}\Gamma_z\left(\mathbf{x}_a, \mathbf{x}_b\right)\,\mathrm{d}\mathbf{x}_a\,\mathrm{d}\mathbf{x}_b, \tag{2.10}$$

with $\mathbf{p}_a$ and $\mathbf{p}_b$, respectively, denoting the variables conjugate to $\mathbf{x}_a$ and $\mathbf{x}_b$: one similarly finds that $(\mathbf{p}, \mathbf{q})$ are respectively conjugate to $(\mathbf{s}, \mathbf{x})$. It is evident from (2.10) that phase-space methods are analogous to plane wave-based near-to-far-field transformations.

From the WT, an inverse Fourier transform retrieves the CT, in either a position or in a direction basis through

$$\tilde{\Gamma}_z\left(\mathbf{p} + \frac{\mathbf{q}}{2}, \mathbf{p} - \frac{\mathbf{q}}{2}\right) = \int \mathrm{e}^{-\mathrm{i}k\mathbf{x}\cdot\mathbf{q}}\mathcal{W}_z\left(\mathbf{x}, \mathbf{p}\right)\,\mathrm{d}\mathbf{x}, \tag{2.11}$$

$$\Gamma_z\left(\mathbf{x} + \frac{\mathbf{s}}{2}, \mathbf{x} - \frac{\mathbf{s}}{2}\right) = \left(\frac{k}{2\pi}\right)^2 \int \mathrm{e}^{\mathrm{i}k\mathbf{p}\cdot\mathbf{s}}\mathcal{W}_z\left(\mathbf{x}, \mathbf{p}\right)\mathrm{d}\mathbf{p}. \tag{2.12}$$

We now show that a direct connection between the Poynting vector, expressing the directional flow of energy across a predefined surface, and phase-space tensors, can be found through the CT of solely electric (magnetic) fields; see the Appendix for details. The starting point is the partial inverse Fourier transform introduced in (2.10) of both the electric and magnetic fields

$$\mathbf{E}(\mathbf{x}, z; \omega) = \left(\frac{k}{2\pi}\right)^2 \int \mathrm{e}^{\mathrm{i}k\mathbf{p}\cdot\mathbf{x}}\tilde{\mathbf{E}}(\mathbf{p}, z; \omega)\,\mathrm{d}\mathbf{p}, \tag{2.13}$$

$$\mathbf{H}(\mathbf{x}, z; \omega) = \left(\frac{k}{2\pi}\right)^2 \int \mathrm{e}^{\mathrm{i}k\mathbf{p}\cdot\mathbf{x}}\tilde{\mathbf{H}}(\mathbf{p}, z; \omega)\,\mathrm{d}\mathbf{p}. \tag{2.14}$$

We now write the Poynting vector for a partially coherent field as [14]

$$\mathbf{S}(\mathbf{x}, z; \omega) = \frac{1}{8\pi}\left\langle \mathbf{E}(\mathbf{x}, z; \omega) \times \mathbf{H}^*(\mathbf{x}, z; \omega) - \mathbf{H}(\mathbf{x}, z; \omega) \times \mathbf{E}^*(\mathbf{x}, z; \omega)\right\rangle, \tag{2.15}$$

where $\langle\cdot\rangle$ denotes an ensemble average over field realizations. Next, using Maxwell's equations in momentum space, along with the definitions of the CT in (2.3) and the WT in (2.4), we get (see Appendix)

$$\mathbf{S}(\mathbf{x}, z; \omega) = \frac{1}{4\pi\eta_0}\left(\frac{k}{2\pi}\right)^2 \int \left[\hat{\mathbf{P}}(\mathbf{p})\,\mathrm{Tr}\left(\mathcal{W}_z\left(\mathbf{x}, \mathbf{p}\right)\right) - 2\mathrm{Re}\left(\hat{\mathbf{P}}(\mathbf{p})^\dagger \cdot \mathcal{W}_z\left(\mathbf{x}, \mathbf{p}\right)\right)\right]\mathrm{d}\mathbf{p}, \tag{2.16}$$

where $\hat{\mathbf{P}}$ is a vector operator defined in (A 9) and $\eta_0 = \sqrt{\mu/\epsilon}$ is the free-space impedance. This gives a direct relation between the Poynting vector and the WT and a similar expression holds, with the substitution $\eta_0 \to \eta_0^{-1}$, for the corresponding magnetic CT. Note that this relation holds in the non-paraxial regime for fully vectorial EM fields and can therefore be used to characterize stochastic near fields. Inherently, the first term in (2.16) has been found by Keller *et al.* in the context of waves propagating in weakly random media [45], while the additional terms have been found in [14] using a different formalism. The second term (see Appendix) is believed to be important in the near field of stochastic sources as it is non-zero only for non-plane wavefronts, i.e. it vanishes for plane wave-like EM fields, whence we expect it to be less important in the far-field region beyond the stochastic source. We remark also that the Poynting theorem, which establishes a relation between the rate of change of the energy density and the Poynting vector, can be used to derive a continuity equation for the Wigner representation of EM waves [45].

To conclude this section, we point out that a NFS of two rectangular magnetic field components tangential to the source is enough to have a full reconstruction of the average Poynting vector from the WT. Other forms of the Poynting vector can be used in connection with the WT of partially coherent fields, which offers an alternative representation to directional fields used to find a paraxial equation for the field intensity [46].

(b) Transport of stochastic fields

rsta.royalsocietypublishing.org Phil. Trans. R. Soc. A 376: 20170455

We now review two approaches to propagate the source CT, a first one constructed through the intermediate representation based on the WT from NFS measurements [29], and a second one based on the method of moments (MoM) [27]. Once partially coherent stochastic fields radiated by a complex spatio-temporal source are characterized in terms of the WT, a phase-space transport equation is used to advance wave densities along their positional and directional characteristics. According to Liouville's theorem, the flow of scalar wave densities such as radiance is conserved along phase-space paths. It has been shown in a different context that this translates into a particularly simple law to advance WFs across finite regions in homogeneous media. More precisely, the WF of in-plane fields near a source can be translated to WFs beyond the source by a simple linear mapping that represents a sheared motion in phase space [13,15]. This has been obtained as a leading order approximation from the solution of both Helmholtz [28] and vector wave equations [29]. Briefly, the exact solution is obtained by representing the second Green identity in $\mathbf{p}$ space and applying the measured in-plane field from NFS as boundary data. In the context of vector EM fields, the Stratton–Chu boundary integral equation is solved in the z-plane [47] in the momentum $\mathbf{p}$ basis, and the free-space propagator for individual field components can be derived exactly. The propagated CF at $z > 0$ is thus obtained by using the fields measured at $z = 0$ as boundary conditions. An exact propagator follows for the field–field CF, from which the corresponding WF can be derived. Recasting the source CF obtained from boundary data into a source WF, an exact transport operator is found in integral form, which can be approximated at leading order with a Frobenius-Perron transport equation. This procedure has been applied to scalar fields, i.e. to a single component of the CT (2.3) and for the WT (2.4) [29, eq. (26)]. More specifically, individual tangent vector components of the in-plane magnetic field have been measured and used to guide and verify the approximate transport equations [29, fig. 9]. Starting with the source correlation in momentum space, calculated from boundary data, propagation along the normal direction to the source is described by the propagator

$$\tilde{\Gamma}_z(\mathbf{p}_a, \mathbf{p}_b) = \mathrm{e}^{ikz[T(\mathbf{p}_a) - T^*(\mathbf{p}_b)]}\tilde{\Gamma}_0(\mathbf{p}_a, \mathbf{p}_b), \tag{2.17}$$

using the notation of (2.9) and where

$$T(\mathbf{p}) = \begin{cases} \sqrt{1-p^2} & \text{for } p^2 \le 1 \\ i\sqrt{p^2-1} & \text{for } p^2 > 1 \end{cases}, \tag{2.18}$$

and $p = |\mathbf{p}|$. The transport equation for the WT can then found by inserting (2.17) into (2.8)

$$\mathcal{W}_z(\mathbf{x}, \mathbf{p}) = \iint \mathcal{G}_z(\mathbf{x}, \mathbf{p}, \mathbf{x}', \mathbf{p}') \cdot \mathcal{W}_0(\mathbf{x}', \mathbf{p}')\, \mathrm{d}\mathbf{x}'\, \mathrm{d}\mathbf{p}' \tag{2.19}$$

with a kernel given by the dyadic operator

$$\mathcal{G}_z(\mathbf{x}, \mathbf{p}, \mathbf{x}', \mathbf{p}') = \left(\frac{k}{2\pi}\right)^2 \delta(\mathbf{p} - \mathbf{p}')\,\mathbf{1} \int \mathrm{e}^{ik(\mathbf{x}-\mathbf{x}')\cdot\mathbf{q} + ikz(T(\mathbf{p}+\mathbf{q}/2) - T^*(\mathbf{p}-\mathbf{q}/2))}\, \mathrm{d}\mathbf{q}, \tag{2.20}$$

where $\mathbf{1}$ is the unit dyad, from which (2.21) can be written as [33]

$$\mathcal{W}_z(\mathbf{x}, \mathbf{p}) = \mathcal{G}_z(\mathbf{x}, \mathbf{p}) *_{\mathbf{x}} \cdot \mathcal{W}_0(\mathbf{x}, \mathbf{p}), \tag{2.21}$$

where $*_{\mathbf{x}}$ denotes the convolution operation acting only on the spatial variable $\mathbf{x}$. It has been shown in [28,29] that $\mathcal{G}_z$ can be simplified through a ray-based approximation for spatial variations in the source correlation that are on a scale that is larger than the wavelength.

The exponent in (2.20) can be approximated by expanding in $\mathbf{q}$, to yield the representation of a Dirac delta function for propagating waves,

$$\mathcal{G}_z\left(\mathbf{x},\mathbf{x}';\mathbf{p},\mathbf{p}'\right)\approx\delta\left(\mathbf{x}-\mathbf{x}'-\frac{z\mathbf{p}}{T(\mathbf{p})}\right)\delta\left(\mathbf{p}-\mathbf{p}'\right)\mathbf{1} \tag{2.22}$$

and an exponential damping for evanescent waves

$$\mathcal{G}_z\left(\mathbf{x},\mathbf{x}';\mathbf{p},\mathbf{p}'\right)\approx \mathrm{e}^{-2kz\sqrt{p^2-1}}\delta\left(\mathbf{p}-\mathbf{p}'\right)\mathbf{1}. \tag{2.23}$$

Finally, using (2.22) and (2.23) in (2.21) yields

$$\mathcal{W}_z\left(\mathbf{x},\mathbf{p}\right)\approx\begin{cases}\mathcal{W}_0\left(\mathbf{x}-\dfrac{z\mathbf{p}}{T\left(\mathbf{p}\right)},\mathbf{p}\right) & \left|\mathbf{p}\right|<1\\ \mathcal{W}_0\left(\mathbf{x},\mathbf{p}\right)\,\mathrm{e}^{-2kz\sqrt{p^2-1}} & \left|\mathbf{p}\right|>1\end{cases}, \tag{2.24}$$

thus leading to a ray-tracing approach for propagating waves and to a $\mathbf{p}$-dependent damping rate for evanescent waves. In the cases of proximity to the source or of low-frequency emissions, the evanescent component may be significant, if not dominant. Then the significance of large momenta going well beyond the leading order in (2.24) motivates an asymptotic calculation for high $\mathbf{p}$ leading to

$$T(\mathbf{p})\approx\mathrm{i}\left|\mathbf{p}\right|,\quad\left|\mathbf{p}\right|\gg1. \tag{2.25}$$

It has been found that the effect of the convolution on evanescent waves is twofold: there is a $\mathbf{p}$-dependent decay combined with a diffusion in $\mathbf{x}$. It is therefore found that the propagator can be derived conveniently by using the WT as an intermediate representation [29, Eq. (22)]. This approximation of the transport rule for WTs is obtained explicitly by retaining the leading order of the series expansion of the exponent of $\mathcal{G}_z$ in (2.21) [28,48]. The propagated CT can be retrieved in a configuration space by an inverse Fourier transform as depicted in [29]. Interestingly, the average scalar EM field intensity can be obtained from the transported $\mathcal{W}_z(\mathbf{x},\mathbf{p})$ in the far field, which has the physical meaning of a (local) average radiation pattern from the statistical source. Used in the Poynting vector, the energy flow takes a particularly simple form in this free-space approximation, propagating field intensities along straight lines, with a tangent vector given by $\mathbf{p}$ from the source plane to the observation plane at z. The propagation rule for an in-plane CF offers an analogue to NF transformation of fields radiated by deterministic sources such as antennas [49]. We now describe an alternative propagation method using the MoM.

Heaviside pioneered the use of vector electric and magnetic potentials. Their use has become standard in EM theory and offers a convenient mathematical framework on which to calculate the radiated fields from a source electric $\mathbf{J}_e$ and/or magnetic $\mathbf{J}_m$ surface current density. The Poynting vector allows for creating equivalent electric currents from in-plane tangential magnetic field components, $\mathbf{J}_e=\hat{n}\times\mathbf{H}$. In the presence of deterministic sources, an equivalent distribution of electric dipoles can be reconstructed from an equivalent surface current density by back-propagating near fields. Having access only to field–field CFs, the same philosophy can be used with stochastic sources to derive propagated CF by forming a stochastic current–current CF [50].

It is shown in [26,27] that the source-field dyadic Green functions can be introduced through the vector potentials to obtain the field dyadic

$$\mathit{\Gamma}_{\mathbf{F}}\left(\mathbf{x}_a,\mathbf{x}_b\right)=\iint G_{\mathbf{FJ}}\left(\mathbf{x}_a-\mathbf{x}'_a\right)\mathit{\Gamma}_{\mathbf{J}}\left(\mathbf{x}'_a,\mathbf{x}'_b\right)G^{\dagger}_{\mathbf{FJ}}\left(\mathbf{x}_b-\mathbf{x}'_b\right)\,\mathrm{d}\mathbf{x}'_b\,\mathrm{d}\mathbf{x}'_b, \tag{2.26}$$

rsta.royalsocietypublishing.org Phil. Trans. R. Soc. A 376: 20170455

describing the correlation of the field $\mathbf{F}$ at points $\mathbf{x}_a$ and $\mathbf{x}_b$. The field dyadic $\Gamma_{\mathbf{F}}(\mathbf{x}_a, \mathbf{x}_b)$ is defined in (2.3) and the current dyadic, related to the correlation of the the surface current density, is defined as

$$\Gamma_{\mathbf{J}}\left(\mathbf{x}_a, \mathbf{x}_b\right) = \left\langle \mathbf{J}_{e,m}\left(\mathbf{x}_a, \omega\right) \mathbf{J}_{e,m}^{\dagger}\left(\mathbf{x}_b, \omega\right)\right\rangle. \tag{2.27}$$

The twofold integral (2.26) can be treated analytically or the MoM [51] can be used to convert the field problem into a network problem and can be solved by subsequently applying network correlation matrix methods [22–24] as shown in [26,27]. By expanding the field and current densities in a vectorial basis

$$\mathbf{F}(\mathbf{x}) = \sum_{n=1}^{N} V_n \mathbf{u}_n(\mathbf{x}), \tag{2.28}$$

$$\mathbf{J}(\mathbf{x}) = \sum_{n=1}^{N} I_n \mathbf{u}_n(\mathbf{x}), \tag{2.29}$$

and substituting (2.28) and (2.29) in (2.26) and (2.27), we obtain a set of algebraic equations

$$\boldsymbol{C}^{V}\left(\omega\right) = \mathbf{Z}\left(\omega\right) \boldsymbol{C}^{I}\left(\omega\right) \mathbf{Z}^{\dagger}\left(\omega\right), \tag{2.30}$$

where we define *impedance* matrix elements through

$$Z_{mn}\left(\omega\right) = \iint \mathbf{u}_m^{\dagger}(\mathbf{x}) G_{\mathbf{FJ}}\left(\mathbf{x}, \mathbf{x}', \omega\right) \mathbf{u}_n(\mathbf{x}')\,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{x}'. \tag{2.31}$$

Having introduced the generalized voltage and current vectors, in (2.30)

$$\mathbf{V}(\omega) = [V_1(\omega), V_2(\omega), \ldots, V_N(\omega)]^{\mathrm{T}}, \tag{2.32}$$

$$\mathbf{I}(\omega) = [I_1(\omega), I_2(\omega), \ldots, I_N(\omega)]^{\mathrm{T}}, \tag{2.33}$$

we have defined voltage–voltage

$$\boldsymbol{C}^{V}\left(\omega\right) = \left\langle \mathbf{V}(\omega)\mathbf{V}^{\dagger}(\omega)\right\rangle, \tag{2.34}$$

and current–current

$$\boldsymbol{C}^{I}\left(\omega\right) = \left\langle \mathbf{I}(\omega)\mathbf{I}^{\dagger}(\omega)\right\rangle, \tag{2.35}$$

correlation matrices. Finally, the propagated field–field correlation dyadic (2.12) is retrieved in the position basis function as

$$\Gamma_{\mathbf{F}}\left(\mathbf{x}_a, \mathbf{x}_b\right) = \sum_{n=1}^{N} \sum_{m=1}^{N} C_{nm}^{V}\left(\omega\right) \mathbf{u}_n(\mathbf{x}_a)\mathbf{u}_m^{*}(\mathbf{x}_b), \tag{2.36}$$

which can be compared with the propagated correlation obtained with the WT intermediate representation.

When detected by probes, EM fields are perturbed by the probe structure itself. As explained in [50], a more subtle effect in NFS concerns the coupling of probes with the surface current density distribution flowing at the source surface. A correction factor can be evaluated by either dedicated experiments or simulations, which can be used to predict the transfer impedance between the field incident on the probe and the voltage read from its port [52]. For statistical sources, it is therefore natural to consider correcting entries of the CT involving the correction factor. Nevertheless, it is shown through plane-wave expansion [53] and operator theory [50] that the propagation of probe voltage–voltage CFs is equivalent to the propagation of field–field CFs. A proposal for standardization of near-field measurement of stochastic EM fields led by the European COST Action 1407 has been initiated in [54]. This initiative has resulted in an IEEE standard proposal being specified for single-probe, dual-probe and multi-probe scanning systems. The measurement methods are compared in terms of their RF, accessible resolution, reliability (including mechanical stress) performances and test time for industrial deployment. The amount of data recorded in two-point measurements required for the characterization of stochastic EM near fields can be reduced considerably by principal component analysis [55–57].

The Green function and MoM methods also have been extended for application to cyclostationary stochastic fields [58–60]. Areas of application are the modelling of the EM interference radiated by digital circuitry inside the system and also into the environment, where the period of the cyclostationary EMI is given by the clock frequency of the digital circuits.

(c) Statistical description of confined fields

The characterization of stochastic fields operating *in vacuo* is important in having an energy-based model of complex sources. It is then natural to include environmental effects in the propagation of CTs. Proximity of objects and/or source confinement, such as for a PCB inside a shielding enclosure, generates reflected waves, which interfere with free-space waves to create complicated distributions of the EM field. This phenomenon is enhanced inside resonant cavities. In irregular cavities, energy is diffused and localization is reduced at high frequencies. Heaviside's analysis of equilibrium radiation has strong precursory elements at the basis of modern statistical electromagnetics. The discussion in [7, sec. 186] starts from the realization that ... *a perfectly conducting screen enclosing a dielectric region supporting electromagnetic disturbances, keeps in their energy, which remains in the electric and magnetic forms, and if there be no source of energy present, the total energy remains constant*. Then it continues towards the definition of the concept of resonance, which develops *the very rudimentary case of a plane wave running to and fro between parallel plane reflecting boundaries, without the slightest tendency to change the type of the vibrations*, for which *there is no necessary tendency for the initial state [...] to break up and fritter down into irregular vibrations*. The subsequent sentence intentionally spoils this intuitive picture through acutely observing that *there does appear to be a general tendency to this [irregular] effect, when the initial states are not so artfully selected as to prevent it happening. Even when we start with some quite simple type of electromagnetic disturbance, the general effect of the repeated reflections from the boundary (especially when of irregular form) and the crossing of waves is to convert the initial simplicity into a highly complex and irregular state of vibration throughout the whole region*. More importantly for the case of statistical sources, Heaviside continues by saying that this *irregularity* occurs *if the initial state be itself of an irregular type, when it is tolerably clear that the irregularity will persist, and become more complete*. Heaviside's reasoning does not stop here and furthers the investigation by assuming a fully developed, *extreme*, irregularity and arguing that *the very irregularity gives rise to a regularity of a new kind, the regularity of averages. The total [average] energy, which is a constant quantity, will be half electric and half magnetic and will be uniformly spread throughout the enclosure, so that the energy density (or energy per unit volume) is constant. As regards the [electrical] displacement and the [magnetic] induction, they take all directions in turn at any one spot, quite irregularly, but so that their time-averages show no directional preference*. Invoking the variability in amount and direction of the flux of energy, expressed by the Poynting vector, Heaviside anticipates an interesting calculation to perform *[...] in virtue of the constancy of the mean density and the preservation of the normal state by constant exchanges of energy, there is a definite mean energy flux to be obtained by averaging results. This mean flux expresses the flux of energy per second across a unit area anywhere situated within the enclosure*. Letting *[...] the mean density of the energy be U [...]*, and fixing *[...] attention upon a unit area, A, [...]*, the flux of energy through A is considered *under different circumstances*. The first one regards the overall energy moving at the same speed v, *[...] as in simple plane progressive waves, and the direction of its motion were perpendicular to the fixed unit area A, then the energy passing through it would belong to a ray (or bundle of rays) of unit section, and the energy flux would be* Uv *[...]*. However, Heaviside points out that *[...] this is impossible, because energy would accumulate on one side of A at the expense of the other. The next approximation, to prevent the accumulation, is to let half the energy go one way and half the other; still, however, in the same line. This brings us down to* $\frac{1}{2}Uv$. *To go further, we must take all possible directions of motion into account.* In order to introduce an additional approximation based on multiple directions of wave motion, we need to imagine the ray *[...] to make an angle* θ *with the normal to A [...]*. Considering this reduction of the ray energy, *the true flux through the area A is* *therefore the mean value of $Uv\cos\theta$ for all directions in space assumed by the ray. Now the mean value of $\cos\theta$ for a complete sphere is zero, and therefore the mean flux through A is zero. This is right, as it asserts that as much goes through one way as the other. To obtain the amount going either way we must average over the hemisphere only. The mean value of $\cos\theta$ is then $\frac{1}{2}$. But we are only concerned with half the total energy, or $\frac{1}{2}U$, when we are confined to one hemisphere. Consequently, we have*

$$W = \tfrac{1}{4}Uv, \tag{2.37}$$

to express the flux of energy W per second each way through any unit area in the enclosure. This result is extremely important as it anticipates early studies on random EM fields in mode-stirred enclosures and reverberation chambers, which are at the basis of modern statistical electromagnetics [61]. Interestingly, Heaviside remarks that the result in (2.37) can be obtained following an alternative procedure, which is very similar to the earlier argument, but whose starting point is *to divide the original ray of unit section along which the flux is Uv into a very great number n of equal rays of unit section, each conveying $1/n$ part of the same, and placed at such inclinations to the normal to A that no direction in space is favoured. This amounts to dividing the surface of a sphere whose centre is that of the area A into n equal parts, the centre of every one of which defines the position of one of the n rays. Any ray now sends $(Uv/n)\cos\theta$ through A per second. Now sum this up over the whole hemisphere and the result is W [...]* in (2.37). More explicitly, *in the limit, when n is infinitely great, we have*

$$W = \int_0^{2\pi}\int_0^{\pi/2} \frac{Uv\cos\theta}{4\pi}\sin\theta\,\mathrm{d}\phi\,\mathrm{d}\theta = \frac{1}{4}Uv, \tag{2.38}$$

as before. The division of energy into partial rays has a remarkable analogy with the random plane wave hypothesis used to explain field fluctuations in reverberation chambers [62]. Having obtained this fundamental result, Heaviside goes on to point out analogies with previous investigators in thermodynamics, where it would also appear *[...] that the result is general, and is independent of sources of heat, and of the emissivity and temperatures*. Nevertheless, it is implicit that the fraction of radiation absorbed inside the cavity would be compensated by the same amount of emission, thus implying the presence of a source maintaining the extreme EM field state. Further considerations are put forward by Heaviside in [7, Sec. 187 and Sec. 188] concerning the mean pressure of radiation and the analogy between emissivity and temperature. Once again, results offer a precursor to developments in statistical electromagnetics, where a strong analogy with thermodynamics concepts is used to obtain average quantities, as summarized in [63]. Although Heaviside's intuition suggested that the conversion of simple sources into a *highly complex and irregular state [...] cannot happen universally*, advancement of statistics as well as modern wave chaos—the study of wave systems whose classical, high-frequency asymptotics, analogue supports a chaotic dynamics – achieved a deeper understanding of the distribution of spectral eigenvalues of irregular systems, thus defining universal laws through random matrix theory (RMT) [64,65]. The random plane wave hypothesis is also supported by phase-space studies of chaotic systems, from which we now understand that the *extreme state* defined by Heaviside can be achieved by considering the field amplitude statistics governed by a universal Gaussian probability distribution: this offers one further example of *a regularity of a new kind*. The WF of a maximally entropic EM field state is uniform in phase space [31]. More modern statistical theories have been formulated to characterize average energy and probability density functions of irregular cavity fields, some inspired by the reverberation chamber [62,66,67], some inspired by wave chaos in confined microwave billiards [64,65]. However, the transition between regularity and irregularity is not sharp and in practical propagation scenarios, especially in wireless communication studies, confining geometries are not completely irregular as they can present flat walls facing each other, thus leading to mixed regular-irregular phase-space structures. The WT offers a valuable way to characterize the development of such a structure from partially coherent and partially polarized sources radiating inside partially irregular environments. A recent approach called the dynamical energy analysis (DEA) has been proposed to transport phase-space densities on triangular meshes [68,69]—similar to those employed in the finite-element method (FEM)—through the Frobenius-Perron operator. DEA is obtained as an asymptotic limit of a wave transfer operator for field–field CFs [70,71]. Taking the WT entries of (2.4) in the short-wavelength limit upon asymptotic expansion, we obtain at leading order that the transported average WT takes the form

$$\mathcal{W}_z(\mathbf{x},\mathbf{p}) \approx \sum_{n=1}^{N} \mathcal{W}_{z'}\left[\phi^n\left(\mathbf{x},\mathbf{p};z,z'\right)\right], \tag{2.39}$$

whose elements are the solution of the stationary Liouville equation used in the DEA method [69]. In (2.39), the linear phase-space flow ϕ is dependent on the chord connecting the point z to the point z' at the boundary. This is the underlying ray-tracing scheme on which the WF develops wave effects. Specific scenarios have been studied to test the propagators that we have derived, including a fully chaotic quantum map [32], which serves as a prototype of the propagation in a diffuse environment such as reverberation chambers.

3. Experiments on stochastic electromagnetic fields

We validate the propagators of §3b through laboratory experiments conducted at the George Green Institute of Electromagnetics Research (GGIEMR), University of Nottingham. A source CF is obtained from measured magnetic fields of stochastic EM fields, see [29].

(a) Experimental set-up

The WF-based approximate propagator in (2.24) will be compared to the exact MoM-based propagator (2.36) and experiments. A one-probe 3-D scanning system is used to perform measurements of a single magnetic field component radiated from the DUT in figure 1. The DUT consists of a metallic brass cavity with a 0.8 m × 0.8 m aperture on the lids shown in figure 1. A metallic rotating stirrer which is driven by a stepper motor is placed inside the cavity to mix and randomize the EM field radiated from the aperture. The source of the radiation is a monopole inside the cavity, a metallic rectangular enclosure of dimensions 1 m × 1 m × 0.5 m. The monopole is a loop antenna, Langer EMV-Technik RF R50-1 magnetic field probe, connected to an Agilent E5062A vector network analyser (VNA). Both magnitude and phase of the coupling between the monopole and the probe, S_{21} measurement, at a frequency of 3 GHz, are captured at each scanning position. The phase reference provided by the VNA makes it possible to calculate a CT by performing a single-probe scan over the source plane.

The scanner system is then placed inside an anechoic chamber to minimize external interference while being controlled using a LabVIEW software on a PC from outside. We perform a near-field scan close to the aperture: the magnetic field is recorded across a dense grid of spatial points by moving a loop probe over the scanning plane. The experiment is done at a fixed frequency $f = 3$ GHz ($\lambda = 0.1$ m).

The scanning plane size is 0.3 m × 0.3 m with 0.005 m scanning steps yielding 60 × 60 scan points per plane. The measurements were carried out for two scanning planes at heights of $z = 0.01$ m and $z = 0.10$ m above the source plane. The measurements were repeated for 36 different paddle positions with a 10-degree rotation step to create an ensemble of fields. Measurement for one paddle position on both scanning planes will capture the VNA transmission parameter S_{21} for a total of 2 heights ×60 points ×60 points = 7200 measurement points. For 36 paddle positions, 259 200 data points are captured and this will provide some difficulty in processing of the data.

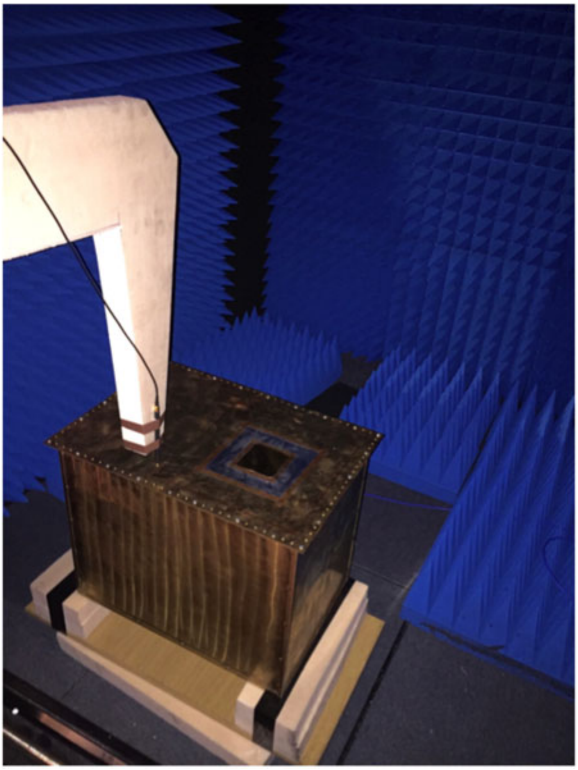


Figure 1. Loop probe moved above the cavity-backed aperture by the scanner system.

(b) Results

Since we only consider magnetic fields oriented in the y-direction, the equivalent surface currents are oriented in x-direction. Therefore, we only need to consider the following component of the Green dyadic

$$G_{HJ}\left(\mathbf{x}-\mathbf{x}', \omega\right) = -\frac{1}{4\pi}\left(z-z'\right)\frac{1-\mathrm{i}k|\mathbf{x}-\mathbf{x}'|}{|\mathbf{x}-\mathbf{x}'|^3}\mathrm{e}^{\mathrm{i}k|\mathbf{x}-\mathbf{x}'|}. \tag{3.1}$$

Unit pulse functions were applied as expansion functions in the discretization scheme based on the MoM

$$u_n(\mathbf{x}) = \begin{cases} 1 & \text{for } \mathbf{x} \in U_n \\ 0 & \text{otherwise} \end{cases}, \tag{3.2}$$

where U_n is the solution domain for the n-th basis function and Dirac delta distributions $\delta(\mathbf{x}) = \delta(x)\delta(y)$ used as weighting functions. In the post-processing of the data, a four-dimensional (4D) correlation dataset $\gamma_z(x_1, y_1, x_2, y_2)$ has been calculated for each scan plane. The 4D dataset for $z = 0.01$ m is used as a source for the propagation calculation. The propagated version of the 4D dataset at $z = 0.10$ m is obtained using both methods. The computational cost for the numerical propagation of the WT in (2.21) is low, since it only involves a convolution. The MoM-based approach requires matrix multiplications for determining field correlation propagation, which is computationally very efficient. However, the computational cost of determining the impedance matrix varies depending on the chosen Green function and requires numerical integration when all near-field contributions are included. The mosaic representation described in [29] is used and the comparison between the two propagated CFs is performed on a selected square of the mosaic. This is shown in figure 2 along with resulting energy densities. There is a good agreement between the two correlation patterns both in terms of height and spreading of the correlation length.

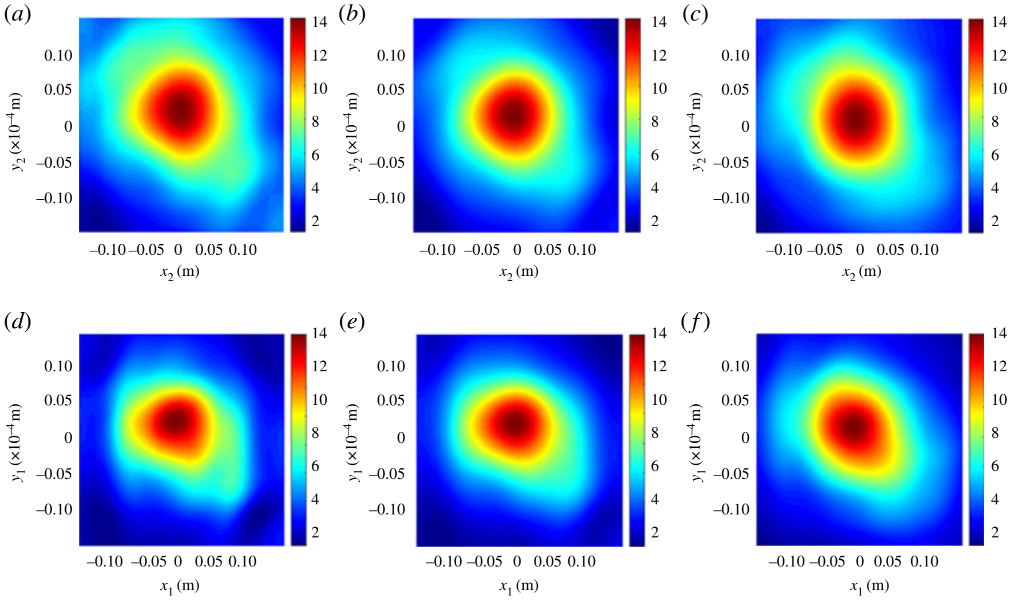


Figure 2. Upper row: comparison between field CFs obtained from (*a*) WF-based and (*b*) MoM-based propagators and (*c*) with measured data. Lower row: energy densities obtained from (*d*) WF-based and (*e*) MoM-based propagators and (*f*) with measured data. The observation plane correlation is taken with the reference point at (0 mm; 0 mm), and results are shown for the plane at $z = 10$ cm beyond the source. (Field correlation and energy density are in units of (A/m)^2.)

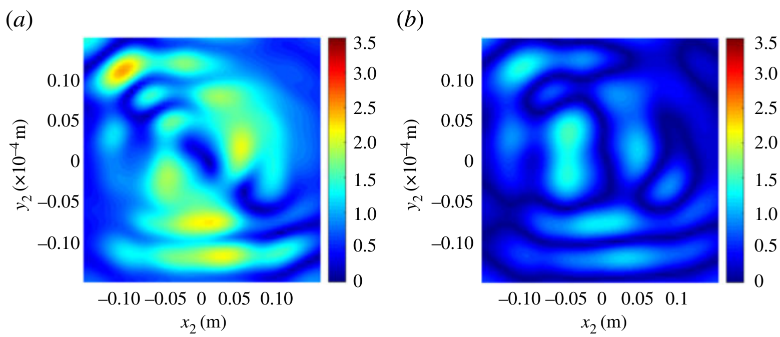


Figure 3. Difference between the intensity from (*a*) WF-based propagator and measured data and (*b*) MoM-based propagator and measured data. (Energy density in units of (A/m)^2.)

Even better agreement is found for the energy densities reported in figure 2*d*,*e*, both accurately reproducing the measured intensity in figure 2*f*. To have a better insight into the accuracy of the *approximated* WF method, we use the *exact* MoM method as a reference (where an exact WF propagator has been used in previous work [28,33]). In particular, the differences between both the propagated energy densities and the measured data have been calculated for the example in figure 2, in order to quantify the relative errors of the two methods with respect to measurements. The error plots are shown in figure 3. Both the propagated energy densities have small differences compared to the measured data with the (numerically exact) MoM method being more accurate than the approximated WF method, as expected. However, the MoM is computationally more intensive than the approximated WF and does not transparently provide information on space-angular properties of emissions as the WF does. We stress that *exact* formulations of the WF method, based on the full kernel defined in (2.20), are available [28,33] as an alternative to MoM, but these would be similarly computationally more intensive. The WF method propagates input data with the level of efficiency of a double Fourier transform. In this example, the computation time of the MoM-based propagation, with full calculation of impedances including near-field contributions, takes hours, while the computation time of the WF-based propagation takes minutes in a standard desktop computer. Since the MoM corresponds to an exact propagator, propagation inaccuracies are only related to discretization and discrepencies are likely due to measurement uncertainty. Theoretical methods and measurement techniques apply to other field components and therefore the same level of agreement is expected for other components of the CT, whence we envisage a successful reconstruction of the average Poynting vector for partially coherent stochastic vector fields.

4. Conclusion

This paper has presented a comparison of approaches to stochastic field measurements and modelling using WFs and MoM. It is shown that by quantifying the coherency tensor of stochastic fields, a wave-dynamical phase-space representation can be devised to extract both the energy flow vector and the local energy density. The spatial and directional properties can be extracted from the CT through the WT and the relation between the WT and Poynting vector is derived. This is then used to develop a propagation rule for the CT. A comparative study based on experimentally measured stochastic fields and propagated fields using the WF and the MoM technique is provided showing good agreement. It is shown that Heaviside anticipated many of these advanced ideas leading to the new field of statistical electromagnetics.

Data accessibility. Datasets and codes used from both UoN and TUM to process data have been deposited: http://ggiemr-repository.engineering.nottingham.ac.uk/UoN/Wigner_MoM.
Authors' contributions. G.G. and J.R. conceived and drafted the manuscript. G.G., S.C.C. and G.T. developed the Wigner function theory. J.R. and P.R. developed the method of moments theory. M.H.B., C.S. and D.W.P.T. conceived and performed the validation measurements. M.H.B. and M.H. did the numerical work and compared with the experimental data. All authors analysed the comparison results. All authors gave final approval for publication.
Competing interests. We declare we have no competing interests.
Funding. Financial support by the COST Action IC1407 'ACCREDIT', the Horizon 2020 FET network NEMF21, grant no. 664828, and the ONR Global, grant no. N62909-16-1-2115, is gratefully acknowledged.
Acknowledgements. We acknowledge helpful discussions with Steve Greedy (Nottingham), Sendy Phang (Nottingham), Timo Hartmann (Nottingham), Mohamed Ismaeel Maricar (Nottingham) and Deepthee Madenhoor Ramaprija (Nottingham).

Appendix

Using Maxwell's equations for a source- and charge-free medium in the $\mathbf{p}$ domain,

$$\mathbf{P} \times \mathbf{E}(\mathbf{p}, z; \omega) = \eta_0 \mathbf{H}(\mathbf{p}, z; \omega), \tag{A 1}$$

$$\mathbf{P} \times \mathbf{H}(\mathbf{p}, z; \omega) = -\eta_0^{-1} \mathbf{E}(\mathbf{p}, z; \omega), \tag{A 2}$$

where $\eta_0 = \sqrt{\mu/\epsilon}$ and $\mathbf{P} = [\mathbf{p}, p_n]$, and with p_n standing for the momentum along the normal direction to the surface in which in-plane components of $\mathbf{p}$ are defined, the first term in (2.15) can be written

$$\mathbf{E}(\mathbf{x}, z; \omega) \times \mathbf{H}^*(\mathbf{x}, z; \omega) = \eta_0^{-1} \left(\frac{k}{2\pi}\right)^4 \iint e^{ik(\mathbf{p}_a - \mathbf{p}_b)\cdot \mathbf{x}} \left[\tilde{\mathbf{E}}(\mathbf{p}_a, z; \omega) \times (\mathbf{P}_b^* \times \tilde{\mathbf{E}}^*(\mathbf{p}_b, z; \omega))\right] \mathrm{d}\mathbf{p}_a \, \mathrm{d}\mathbf{p}_b, \tag{A 3}$$

while the second term in (2.15) is

$$\mathbf{H}(\mathbf{x}, z; \omega) \times \mathbf{E}^*(\mathbf{x}, z; \omega) = \eta_0^{-1} \left(\frac{k}{2\pi}\right)^4 \iint e^{ik(\mathbf{p}_a - \mathbf{p}_b)\cdot \mathbf{x}} \left[(\mathbf{P}_a \times \tilde{\mathbf{E}}(\mathbf{p}_a, z; \omega)) \times \tilde{\mathbf{E}}^*(\mathbf{p}_b, z; \omega)\right] \mathrm{d}\mathbf{p}_a \, \mathrm{d}\mathbf{p}_b. \tag{A 4}$$

Combining the two expressions yields

$$\mathbf{S}(\mathbf{x},z;\omega) = \frac{1}{8\pi\eta_0}\left(\frac{k}{2\pi}\right)^4 \left\langle \iint e^{ik(\mathbf{p}_a-\mathbf{p}_b)\cdot\mathbf{x}}[\tilde{\mathbf{E}}(\mathbf{p}_a,z;\omega)\times(\mathbf{P}_b^*\times\tilde{\mathbf{E}}^*(\mathbf{p}_b,z;\omega)) - (\mathbf{P}_a\times\tilde{\mathbf{E}}(\mathbf{p}_a,z;\omega))\times\tilde{\mathbf{E}}^*(\mathbf{p}_b,z;\omega)]\,\mathrm{d}\mathbf{p}_a\,\mathrm{d}\mathbf{p}_b\right\rangle, \tag{A 5}$$

from which, unfolding the kernel inside the square brackets yields

$$\mathbf{S}(\mathbf{x},z;\omega) = \frac{1}{8\pi\eta_0}\left(\frac{k}{2\pi}\right)^4 \left\langle \iint e^{ik(\mathbf{p}_a-\mathbf{p}_b)\cdot\mathbf{x}}[(\mathbf{P}_a+\mathbf{P}_b^*)(\tilde{\mathbf{E}}(\mathbf{p}_a,z;\omega)\cdot\tilde{\mathbf{E}}^*(\mathbf{p}_b,z;\omega)) - (\tilde{\mathbf{E}}(\mathbf{p}_a,z;\omega)\cdot\mathbf{P}_b^*)\tilde{\mathbf{E}}^*(\mathbf{p}_b,z;\omega) - \tilde{\mathbf{E}}(\mathbf{p}_a,z;\omega)(\mathbf{P}_a\cdot\tilde{\mathbf{E}}^*(\mathbf{p}_b,z;\omega))]\,\mathrm{d}\mathbf{p}_a\mathrm{d}\mathbf{p}_b\right\rangle. \tag{A 6}$$

Now, since averaging and integration commute, and using the property of the electric CT, $\mathbf{P}_a\cdot\tilde{\Gamma}_z = \tilde{\Gamma}_z\cdot\mathbf{P}_b^* = 0$, along with the Hermiticity of $\tilde{\Gamma}_z$, we may write

$$\mathbf{S}(\mathbf{x},z;\omega) = \frac{1}{8\pi\eta_0}\left(\frac{k}{2\pi}\right)^4 \iint e^{ik(\mathbf{p}_a-\mathbf{p}_b)\cdot\mathbf{x}}\Big[(\mathbf{P}_a+\mathbf{P}_b^*)\,\mathrm{Tr}\left(\tilde{\Gamma}_z(\mathbf{p}_a,\mathbf{p}_b)\right) - \tilde{\Gamma}_z(\mathbf{p}_a,\mathbf{p}_b)\cdot(\mathbf{P}_a+\mathbf{P}_b^*) - \tilde{\Gamma}_z^*(\mathbf{p}_a,\mathbf{p}_b)\cdot(\mathbf{P}_a+\mathbf{P}_b^*)\Big]\,\mathrm{d}\mathbf{p}_a\,\mathrm{d}\mathbf{p}_b, \tag{A 7}$$

which is consistent with the expression derived in [14, eq. (2.13)]. We can exploit the calculation further by involving the phase-space representation (2.11)

$$\mathbf{S}(\mathbf{x},z;\omega) = \frac{1}{4\pi\eta_0}\left(\frac{k}{2\pi}\right)^4 \iiint e^{ik\mathbf{q}\cdot(\mathbf{x}-\mathbf{x}')}\int[\mathbf{P}\,\mathrm{Tr}\,(\mathcal{W}_z(\mathbf{x}',\mathbf{p})) - 2\mathrm{Re}\,(\mathcal{W}_z(\mathbf{x}',\mathbf{p})\cdot\mathbf{P})]\,\mathrm{d}\mathbf{p}\,\mathrm{d}\mathbf{q}\,\mathrm{d}\mathbf{x}', \tag{A 8}$$

where

$$\mathbf{P} = \frac{\mathbf{P}_a+\mathbf{P}_b^*}{2} = \mathbf{P}(\mathbf{p},\mathbf{q}) \tag{A 9}$$

is implicitly a function of $\mathbf{p}$ and $\mathbf{q}$, but not of $\mathbf{x}$ or $\mathbf{x}'$. We may formally write this in the form (2.16), where the vector operator $\hat{\mathbf{P}}(\mathbf{p})$ acts by convolution on functions of $\mathbf{x}$ as

$$\hat{\mathbf{P}}(\mathbf{p})g(\mathbf{x}) = \int \mathbf{K}(\mathbf{x}-\mathbf{x}',\mathbf{p})g(\mathbf{x}')\,\mathrm{d}\mathbf{x}',$$

where

$$\mathbf{K}(\mathbf{x},\mathbf{p}) = \left(\frac{k}{2\pi}\right)^2\int e^{ik\mathbf{q}\cdot\mathbf{x}}\mathbf{P}(\mathbf{p},\mathbf{q})\,\mathrm{d}\mathbf{q}.$$

Note that the components of this operator parallel to the source plane act simply through the delta kernel

$$P_x(\mathbf{p},\mathbf{q}) = p_x \quad\Rightarrow\quad K_x(\mathbf{x},\mathbf{p}) = p_x\delta(\mathbf{x})$$

but the normal component is non-trivial.

References

1. Maxwell J. 1954 *A treatise on electricity and magnetism*, vol. 1. Dover Publications.
2. Harman PM. 1993 Bruce J. Hunt, *The Maxwellians*. Ithaca and London: Cornell University Press, 1991. pp. xiii 266. *Br. J. Hist. Sci.* **26**, 117–118. (doi:10.1017/S0007087400030569)
3. Mahon B. 2009 *Oliver Heaviside: maverick mastermind of electricity*. History and Management of Technology Series. London, UK: Institution of Engineering and Technology.
4. Nahin P. 2002 *Oliver Heaviside: the life, work, and times of an electrical genius of the Victorian age*. Baltimore, MD and London, UK: Johns Hopkins University Press.
5. Hong S. 2001 *Wireless: from Marconi's black-box to the audion*. Transformations. Cambridge, MA: MIT Press.

6. Russer P. 2012 Ferdinand Braun – a pioneer in wireless technology and electronics. In *A wireless world* (eds K Grandin, P Mazzinghi, N Olander, G Pelosi). Contributions to the history of the Royal Swedish Academy of Sciences, vol. 42, pp. 228–247. Firenze, Italy: Firenze University Press.
7. Heaviside O. 2011 *Electromagnetic theory, vol 1*. Cambridge Library Collection. Cambridge, UK: Cambridge University Press.
8. Darrigol O. 2003 *Electrodynamics from ampère to Einstein*. Oxford, UK: Oxford University Press.
9. Warnick KF, Russer PH. 2014 Differential forms and electromagnetic field theory (invited paper). *PIER Progr. Electromagn. Res.* **148**, 83–112. (doi:10.2528/PIER14063009)
10. Haider M, Russer JA. 2017 Differential form representation of stochastic electromagnetic fields. *Adv. Radio Sci.* **15**, 21–28. (doi:10.5194/ars-15-21-2017)
11. Hillery M, O'Connell R, Scully M, Wigner E. 1984 Distribution functions in physics: fundamentals. *Phys. Rep.* **106**, 121–167. (doi:10.1016/0370-1573(84)90160-1)
12. Bolte J, Glaser R, Keppeler S. 2001 Quantum and classical ergodicity of spinning particles. (http://arxiv.org/abs/nlin/0101022)
13. Bastiaans M. 1979 The Wigner distribution function and Hamilton's characteristics of a geometric-optical system. *Opt. Commun.* **30**, 321–326. (doi:10.1016/0030-4018(79)90361-4)
14. Alonso MA. 2004 Wigner functions for nonparaxial, arbitrarily polarized electromagnetic wave fields in free space. *J. Opt. Soc. Am. A* **21**, 2233–2243. (doi:10.1364/JOSAA.21.002233)
15. Torre A. 2005 *Linear ray and wave optics in phase space: bridging ray and wave optics via the Wigner phase-space picture*. Amsterdam, The Netherlands: Elsevier Science.
16. Marcuvitz N. 1991 The quasiparticle view of wave propagation. *Proc. IEEE* **79**, 1350–1358. (doi:10.1109/5.104209)
17. Hunt B. 2005 *The Maxwellians*. Cornell History of Science Series. Ithaca, NY: Cornell University Press.
18. Yaghjian A. 1986 An overview of near-field antenna measurements. *IEEE. Trans. Antennas Propag.* **34**, 30–45. (doi:10.1109/TAP.1986.1143727)
19. Tong X, Thomas D, Nothofer A, Sewell P, Christopoulos C. 2010 Modeling electromagnetic emissions from printed circuit boards in closed environments using equivalent dipoles. *IEEE Trans. Electromagn. Compat.* **52**, 462–470. (doi:10.1109/TEMC.2010.2044181)
20. Tong X, Thomas D, Nothofer A, Sewell P, Christopoulos C. 2011 Reduction of sensitivity to measurement errors in the derivation of equivalent models of emission in numerical computation. *ACES J.* **26**, 603–610.
21. Davenport WB, Root WL. 1958 *An introduction to the theory of random signals and noise*. New York, NY: McGraw-Hill.
22. Hillbrand H, Russer P. 1976 An efficient method for computer aided noise analysis of linear amplifier networks. *IEEE Trans. Circuits Syst.* **23**, 235–238. (doi:10.1109/TCS.1976.1084200)
23. Russer P, Müller S. 1990 Noise analysis of linear microwave circuits. *Int. J. Num. Model. Electr. Netw. Dev. Fields* **3**, 287–316. (doi:10.1002/jnm.1660030408)
24. Russer P. 1996 Noise analysis of linear microwave circuits with general topology. In: *The review of radio science 1993–1996* (ed. W Ross Stone), pp. 887–890. Oxford, UK: Oxford University Press.
25. Felsen LB, Mongiardo M, Russer P. 2009 *Electromagnetic field computation by network methods*. Berlin: Springer.
26. Russer JA, Russer P. 2011 An efficient method for computer aided analysis of noisy electromagnetic fields. In *Microwave Symposium Digest (MTT), 2011 IEEE MTT-S Int., June 2011*, pp. 1–4. New York, NY: IEEE.
27. Russer JA, Russer P. 2015 Modeling of noisy EM field propagation using correlation information. *IEEE. Trans. Microw. Theory. Tech.* **63**, 76–89. (doi:10.1109/TMTT.2014.2376962)
28. Gradoni G, Creagh SC, Tanner G, Smartt C, Thomas DWP. 2015 A phase-space approach for propagating field-field correlation functions. *New. J. Phys.* **17**, 093027. (doi:10.1088/1367-2630/17/9/093027)
29. Gradoni G, Arnaut LR, Creagh SC, Tanner G, Baharuddin MH, Smartt C, Thomas DWP. 2017 Wigner-function-based propagation of stochastic field emissions from planar electromagnetic sources. *IEEE Trans. Electromagn Compat.* **60**, 580–588. (doi:10.1109/temc.2017.2738329)
30. Baharuddin MH, Nasser H, Smartt C, Thomas DWP, Gradoni G, Creagh SC, Tanner G. 2016 Measurement and Wigner function analysis of field-field correlation for complex PCBs in

near field. In *2016 Int. Symp. Electromagnetic Compatibility – EMC Europe, September 2016*, pp. 7–11.
31. Berry MV. 1977 Regular and irregular semiclassical wavefunctions. *J. Phys. A Math. Gen.* **10**, 2083–2091. (doi:10.1088/0305-4470/10/12/016)
32. Creagh SC, Gradoni G, Hartmann T, Tanner G. 2017 Propagating wave correlations in complex systems. *J. Phys. A Math. Theor.* **50**, 045101. (doi:10.1088/1751-8121/50/4/045101)
33. Gradoni G, Ramapriya DM, Creagh SC, Tanner G, Baharuddin MH, Nasser H, Smartt C, Thomas DWP. 2017 Near-field scanning and propagation of correlated low-frequency radiated emissions. *IEEE Trans. Electromagn. Compat.* **60**, 2045–2048. (doi:10.1109/temc.2017.2778046)
34. Yaghjian A. 1986 An overview of near-field antenna measurements. *IEEE Trans. Antennas Propag.* **34**, 30–45. (doi:10.1109/TAP.1986.1143727)
35. Sugawara K, Chen CP, Ma Z, Anada T, Thomas DWP. 2007 Non-contacting electric and magnetic field probe for measuring em fields on microwave planar circuits. In *2007 Asia-Pacific Microwave Conf., Bangkok, Thailand, December 2007*, pp. 1–4.
36. Thomas DWP, Biwojno K, Xin T, Nothofer A, Sewell P, Christopoulos C. 2008 Measurement and simulation of near field emissions from microstrip lines. In *2008 Int. Symp. on Electromagn. Compat. – EMC Europe, Hamburg, Germany, September 2008*, pp. 1–6.
37. Russer JA, Uddin N, Awny AS, Thiede A, Russer P. 2015 Near-field measurement of stochastic electromagnetic fields. *IEEE Electromagn. Compat. Mag.* **4**, 79–85. (doi:10.1109/memc.2015.7336761)
38. Smartt C, Thomas DWP, Nasser H, Baharuddin M, Gradoni G, Creagh SC, Tanner G. 2015 Challenges of time domain measurement of field-field correlation for complex PCBs. In *2015 IEEE Int. Symp. Electromagnetic Compatibility (EMC), Dresden, Germany, Dresden, Germany, August 2015*, pp. 953–958.
39. Thomas DWP, Smartt C, Nasser H, Baharuddin M, Greedy S, Gradoni G, Creagh SC, Tanner G. 2016 Time domain measurement of near field emissions from complex PCBs. In *2016 46th European Microwave Conf. (EuMC), London, UK, October 2016*, pp. 707–710.
40. Heaviside O. 2011 *Electromagnetic theory: volume 2*. Cambridge Library Collection. Cambridge, UK: Cambridge University Press.
41. Heaviside O. 2011 *Electromagnetic theory: volume 3*. Cambridge Library Collection. Cambridge, UK: Cambridge University Press.
42. Gao X, Fan J, Zhang Y, Kajbaf H, Pommerenke D. 2014 Far-field prediction using only magnetic near-field scanning for EMI test. *IEEE Trans. Electromagn. Compat.* **56**, 1335–1343. (doi:10.1109/TEMC.2014.2322081)
43. Holloway CL, Gordon JA, Jefferts S, Schwarzkopf A, Anderson DA, Miller SA, Thaicharoen N, Raithel G. 2014 Broadband Rydberg atom-based electric-field probe for SI-traceable, self-calibrated measurements. *IEEE. Trans. Antennas Propag.* **62**, 6169–6182. (doi:10.1109/TAP.2014.2360208)
44. Jackson J. 2012 *Classical electrodynamics*. Hoboken, NJ: Wiley.
45. Ryzhik L, Papanicolaou G, Keller JB. 1996 Transport equations for elastic and other waves in random media. *Wave Motion* **24**, 327–370. (doi:10.1016/S0165-2125(96)00021-2)
46. Kinsler P, Favaro A, McCall MW. 2009 Four Poynting theorems. *Eur. J. Phys.* **30**, 983–993. (doi:10.1088/0143-0807/30/5/007)
47. Tai C 1996. *Dyadic Green functions in electromagnetic theory*. IEEE/OUP Series on Electromagnetic Wave Theory. Piscataway, NJ: IEEE Press.
48. Gradoni G, Creagh SC, Tanner G. 2014 A Wigner function approach for describing the radiation of complex sources. In *Proc. IEEE EMC Conf., Raleigh, NC, 2014*.
49. Sarkar TK, Taaghol A. 1999 Near-field to near/far-field transformation for arbitrary near-field geometry utilizing an equivalent electric current and MoM. *IEEE. Trans. Antennas Propag.* **47**, 566–573. (doi:10.1109/8.768793)
50. Gradoni G, Ramapriya DM, Creagh SC, Tanner G, Baharuddin MH, Smartt C, Thomas DWP. 2017 Propagation methods for stochastic field emissions and source reconstruction. In *2017 Int. Symp. Electromagnetic Compatibility – EMC Europe, Angers, France, September 2017*, pp. 1–6.
51. Harrington RF. 1968 *Field computation by moment methods*. San Francisco, CA: IEEE Press.

52. Tzoulis A, Eibert T. 2005 Efficient electromagnetic near-field computation by the multilevel fast multipole method employing mixed near-field/far-field translations. *IEEE. Antennas Wirel. Propag. Lett.* **4**, 449–452. (doi:10.1109/LAWP.2005.860195)
53. Cozza A, Masciovecchio F, Dorgan C, Serhir M, Monsef F, Lecointe D. 2017 A dielectric low-perturbation field scanner for multi-path environments. *IEEE. Trans. Antennas Propag.* **65**, 1978–1987. (doi:10.1109/TAP.2017.2670563)
54. Haider M, Russer J. 2018 Computer aided characterization of radiated electromagnetic interference of printed circuit boards by stochastic electromagnetic fields. In *Proc. of the 2nd URSI AT-RASC, Gran Canaria, 28 May–1 June 2018*.
55. Haider M, Russer JA. 2017 Principal component analysis for efficient characterization of stochastic electromagnetic fields. *Int. J. Num. Modell. Electr. Netw. Dev. Fields* **31**, e2246. (doi:10.1002/jnm.2246)
56. Haider M, Russer JA, Baev A, Kuznetsov Y, Russer P. 2017 Principal component analysis applied in modeling of stochastic electromagnetic field propagation. In *Proc. European Microwave Conf. (EuMC), Nuremberg, Germany, 8–13 October 2017*.
57. Russer JA, Haider M, Baharuddin MH, Smartt C, Baev A, Wane S, Bajon D, Kuznetsov Y, Thomas D, Russer P. 2016 Correlation measurement and evaluation of stochastic electromagnetic fields. In *Proc. Int. Symp. Electromagnetic Compatibility, EMC Europe, Wroclaw, Poland, 5–9 September 2016*.
58. Russer JA, Russer P, Konovalyuk M, Gorbunova A, Baev A, Kuznetsov Y. 2015 Near-field propagation of cyclostationary stochastic electromagnetic fields. In *Int. Conf. Electromagnetics in Advanced Applications (ICEAA), 2015, Torino, Italy, 7–11 September 2015*, pp. 1456–1459.
59. Russer JA, Russer P. 2016 Modeling and analysis of stationary and cyclostationary noisy EM fields. In *Proc. Int. Symp. Electromagnetic Compatibility, EMC Europe, Workshop WS1A-WS1B, Wroclaw, Poland, 5–9 September 2016*.
60. Kuznetsov Y, Baev A, Haider M, Russer JA. 2018 Cyclostationary source separation in the near-field of electronic devices. In *Proc. 2nd URSI AT-RASC, Gran Canaria, Spain, 28 May–1 June 2018*.
61. Corona P, Ladbury J, Latmiral G. 2002 Reverberation-chamber research-then and now: a review of early work and comparison with current understanding. *IEEE Trans. Electromagn. Compat.* **44**, 87–94. (doi:10.1109/15.990714)
62. Hill D. 1998 Plane wave integral representation for fields in reverberation chambers. *IEEE Trans. Electromagn. Compat.* **40**, 209–217. (doi:10.1109/15.709418)
63. Migliaccio M, Gradoni G, Arnaut LR. 2016 Electromagnetic reverberation: the legacy of Paolo Corona. *IEEE Trans. Electromagn. Compat.* **58**, 643–652. (doi:10.1109/TEMC.2016.2546183)
64. Gradoni G, Yeh J-H, Xiao B, Antonsen TM, Anlage SM, Ott E. 2014 Predicting the statistics of wave transport through chaotic cavities by the random coupling model: a review and recent progress. *Wave Motion* **51**, 606–621. (doi:10.1016/j.wavemoti.2014.02.003)
65. Selemani K, Gros JB, Richalot E, Legrand O, Picon O, Mortessagne F. 2015 Comparison of reverberation chamber shapes inspired from chaotic cavities. *IEEE Trans. Electromagn. Compat.* **57**, 3–11. (doi:10.1109/TEMC.2014.2313355)
66. Arnaut LR. 2003 Limit distributions for imperfect electromagnetic reverberation. *IEEE Trans. Electromagn. Compat.* **45**, 357–377. (doi:10.1109/TEMC.2003.811301)
67. Cozza A. 2011 The role of losses in the definition of the overmoded condition for reverberation chambers and their statistics. *IEEE Trans. Electromagn. Compat.* **53**, 296–307. (doi:10.1109/TEMC.2010.2081993)
68. Tanner G. 2009 Dynamical energy analysis – determining wave energy distributions in vibro-acoustical structures in the high-frequency regime. *J. Sound. Vib.* **320**, 1023–1038. (doi:10.1016/j.jsv.2008.08.032)
69. Chappell D, Tanner G. 2013 Solving the stationary Liouville equation via a boundary element method. *J. Comput. Phys.* **234**, 487–498. (doi:10.1016/j.jcp.2012.10.002)
70. Gradoni G, Creagh SC, Tanner G. 2015 Propagation of correlation functions in cavities. In *2015 1st URSI Atlantic Radio Science Conf. (URSI AT-RASC), Gran Canaria, Spain, May 2015*, p. 1.
71. Gradoni G, Creagh SC, Tanner G. 2016 Transfer operator approach for cavities with apertures. In *2016 URSI Int. Symp. Electromagnetic Theory (EMTS), Espoo, Finland, August 2016*, pp. 682–685.

PHILOSOPHICAL TRANSACTIONS A

rsta.royalsocietypublishing.org

Research

Check for updates

Cite this article: Duffy AP, Arihilam E, Sasse H, Parker S, Armstrong R, Wainwright N, Steggles S, Thomas T. 2018 Measuring the shielding properties of flexible or rigid enclosures for portable electronics. *Phil. Trans. R. Soc. A* **376**: 20170456.
http://dx.doi.org/10.1098/rsta.2017.0456

Accepted: 10 August 2018

One contribution of 13 to a theme issue 'Celebrating 125 years of Oliver Heaviside's 'Electromagnetic Theory''.

Subject Areas:
electrical engineering

Keywords:
Heaviside, shielding effectiveness, cybersecurity, reverberation chamber, electromagnetics, electromagnetic compatibility

Author for correspondence:
A. P. Duffy
e-mail: apd@dmu.ac.uk

Measuring the shielding properties of flexible or rigid enclosures for portable electronics

A. P. Duffy[1], E. Arihilam[1], Hugh Sasse[1], Sarah Parker[2], R. Armstrong[2], N. Wainwright[2], S. Steggles[3], and T. Thomas[4]

[1]Faculty of Technology, De Montfort University, The Gateway, Leicester LE1 9BH, UK
[2]Eurofins (previously York EMC Services), Market Square, University of York, York YO10 5DD, UK
[3]Disklabs Ltd, Galena Close, Tamworth, Staffordshire B77 4AS, UK
[4]Privoro Inc, 3100 West Ray Road, Chandler, AZ 85226, USA

APD, 0000-0002-5074-4273

Heaviside, in volume 1 of *Electromagnetic theory*, considered shielding of conducting materials in the form of attenuation. This treatment is still significant in the understanding of shielding effectiveness. He also considered propagation of electromagnetic waves in free-space. What Heaviside (1850–1925) could never have imagined is that 125 years later, there would be devices we know as mobile phones (or cell phones, handies, etc.) with capabilities beyond the dreams of the great science fiction writers of the day like H. G. Wells (1866–1949) or Jules Verne (1828–1905). More than this, that there would be a need for law enforcement agencies, among others, to use electromagnetically shielded enclosures to protect electronic equipment from communicating with the 'outside world'. Nevertheless, Heaviside's work is still fundamental to the developments discussed here. This paper provides a review of Heaviside's view of shielding and propagation provided in volume 1 of *Electromagnetic theory* and develops that to the design of new experiments to test the shielding of these portable enclosures in a mode-stirred reverberation chamber, a test environment that relies entirely on reflections from conducting surfaces for its operation.

THE ROYAL SOCIETY PUBLISHING

This article is part of the theme issue 'Celebrating 125 years of Oliver Heaviside's 'Electromagnetic Theory''.

1. Introduction

(a) Contribution from Oliver Heaviside's 'electromagnetic theory'

Electromagnetic theory, volume 1, was originally published in 1893. The commentary here is based on Heaviside [1]. The preface of this edition gives an insight into his approach to communicating and quotes the *Heaviside centenary volume* [2] by saying 'As the reader can see by opening the present volume almost at random, Heaviside seldom writes more than a page or two without introducing a pleasant touch of informality or humour'. It also quotes 'Oliver Heaviside' by E. T. Whittaker where it describes what would be identified today as a disadvantaged childhood. This led to the desire to leave home. Prof. Charles Wheatstone of Kings College London was his uncle by marriage and, through that connection, Heaviside became a telegraph operator in Newcastle. This employment also saw his agile and probing mind turn to problems in telegraphy where he would publish several papers in inter alia the *Philosophical Magazine*. He left this job when only 24, his last formal employment. From then, his life was dedicated to the study of electromagnetics and related phenomena.

His approach to intellectual investigation resulted in his summarizing and condensing Maxwell's equations into the four equations that we recognize and describe today as 'Maxwell's Equations'. (Perhaps they should be referred to commonly as the Heaviside–Maxwell equations.)

His work on transmission lines is still widely used and, with the general increase in the electrical length of circuits and resulting prevalence in signal integrity, perhaps more so than previously. However, he did not endear himself to the establishment by doing things like pointing out the conditions for distortionless transmission against the commonly held understanding of the day. This non-conformist approach to research extended to mathematics with the development of operational calculus which, again, was poorly received at the time by the establishment. Now, much of his work is regarded as, at least, ground-breaking and his three volumes on electromagnetic theory as seminal.

This paper considers the measurement of shielding of portable enclosures for electronics. This is a problem that encompasses electromagnetic propagation, shielding and reflections in a reverberation chamber. Heaviside described the foundations of these in [1] in the following sections:

§180 Describes the relationship between the electric and magnetic fields in a plane wave
§181 Presents wavefronts
§184 Talks about the behaviour of a perfect conductor in a wave as an obstruction but noting it does not absorb
§189 Introduces internal obstructions and conduction on the surface
§190 Discusses reflections
§191 Continues the discussion of conductors
§193 Describes thin plates
§195 Discusses attenuation
§206 Presents thoughts on the guiding of waves

While Heaviside produced mathematical analyses of clarity and vision, his use of mathematics was as a support to his more discursive approach to communication.

The discussions in this paper are clearly grounded in both the scientific foundations and style of communication laid down by Oliver Heaviside.

rsta.royalsocietypublishing.org Phil. Trans. R. Soc. A 376: 20170456

Figure 1. Examples of shielded enclosures. (Online version in colour.)

(b) Shielding for portable electronics

Portable electronics, such as mobile telephones, laptops and tablets, all have myriad connectivity options with receiving and/or transmitting functions available for WiFi, cellular telephone connectivity, GPS, Bluetooth and near-field communications (NFC), etc. There are situations where these devices need to have their connectivity limited in a way that turning on ‘airplane mode’ is simply not appropriate. Some examples of this are to avoid distractions while driving, to ensure there can be no surreptitious transmission of conversations during sensitive meetings, or to allow law enforcement agencies to render the equipment isolated without turning it off (which means it is impossible for the call log or messages to be deleted remotely, potentially removing evidence of nefarious activities).

With the growing sophistication in electronic crime, especially involving the use of computer systems, it has become important to have the opportunity to secure such items of electronics with protective shielding. Such portable items of electronics may contain personal information and data which if not protected could be maliciously hacked (such as NFC hacking). A shielded enclosure will use a conductive material or a mesh to attenuate the propagation of electronic fields into the enclosure. Sufficient shielding (attenuation) of the enclosure is important to make it difficult to connect to the devices (wirelessly) and alter, delete data or add corrupt digital materials/data onto the devices [3], or to allow the devices to act remotely to perform such functions such as timer-based dialling of a detonator on an improvised explosive device.

These enclosures can be in the form of flexible pouches, with or without inspection windows, or solid rigid boxes. Figure 1 shows some examples of commercial products designed to achieve these goals. The construction of the enclosures is not standardized and so various approaches and materials are used by different vendors and for different applications. The authors have seen clearly spurious claims for the shielding of similar devices based on measurements of materials alone, and not accounting for any seams, folds or fasteners. These factors indicate the importance of producing a test method that is reliable, repeatable and robust.

The quality of this shielding is vital in being able to achieve necessary isolation. A typical measure of shielding effectiveness (SE) is as given as follows:

$$\mathrm{SE} = -20\log_{10}\left(\frac{E_\mathrm{S}}{E_\mathrm{NS}}\right). \tag{1.1}$$

where E_S is the electric field strength with the shield in place and E_NS is the electric field strength with no shield in place, but all other elements in the configuration remain the same. The minus sign is there to allow a positive level of shielding to be reported.

An obvious question is how much shielding is needed for a given application? This, of course, depends on factors such as the separation between the electronics and the transmitter. By the

rsta.royalsocietypublishing.org Phil. Trans. R. Soc. A 376: 20170456

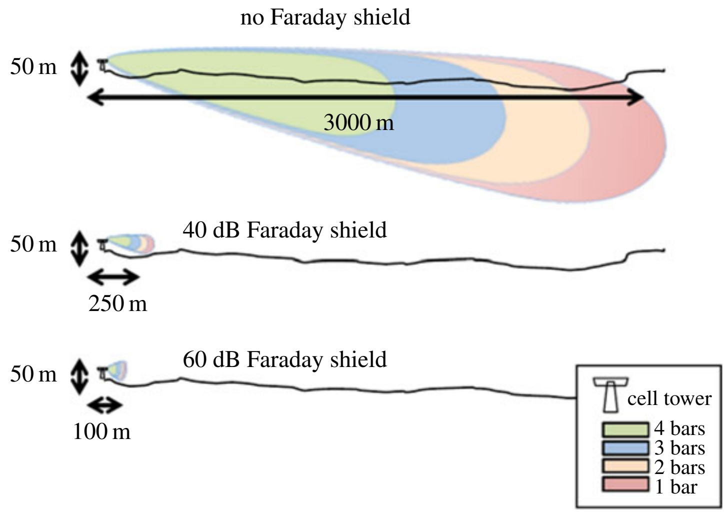


Figure 2. Indication of the effect of shielding on a cellular telephone as a function of distance based on the Hata–Okamura model. (Online version in colour.)

way of example, consider electromagnetic propagation according to the classic empirical Hata–Okamura model [4]. The following equation gives the model

$$P_L(\mathrm{dB}) = 26.16 \log f_r - 13.82 \log h_{\mathrm{bs}} + (44.9 - 6.55 \log h_{\mathrm{bs}})R - H + 69.55, \tag{1.2}$$

where f_r is the frequency of operation, h_{bs} is the height of the transmitter, R is the distance between the transmitter and receiver and H is a factor based on the height of the receiver and frequency and is dependent on the classification of the environment. Figure 2 shows how the level of shielding relates to the distance from the transmitter for different 'bar levels' of a mobile telephone. It can be seen that the effective usable distance reduces substantially as the shielding increases.

Clearly, knowledge of the value of the shielding performance of the enclosures is of utmost importance, particularly where the performance of those enclosures is critical. This is a particularly timely question given the rise in shielding products coming onto the market and the need for *inter alia* law enforcement agencies being able to specify the performance they need or to be able to compare the cost-effectiveness for various competitor products. This is a piece of work that has been picked up by the IEEE EMC Society in their standards project P2710 [5] which looks to define an approach to measuring this SE that can be used by manufacturers, vendors, and purchasers to ensure consistency of measurements and their comparison. This paper is a foundational contribution to that study and shows how a mode-stirred reverberation chamber in conjunction with a broadband source can provide a robust measurement of the shielding.

The rest of this paper looks at mode-stirred reverberation chamber measurements. It defines the proposed test method and looks at some typical results. The conclusion is drawn that the approach is both suitable and robust.

2. Reverberation chamber tests

The mode-stirred reverberation chamber is a large, over-moded, resonant cavity with moving elements internally that modify the boundary conditions which give rise to modal patterns that vary as the moving elements change position. Typically, an internal paddle or similar stirrer is used, centred on an axis, with sufficient steps through a complete revolution to ensure good stirring and statistical independence of fields from one step to the next. This change in modal pattern over a full revolution of the stirrer gives rise to any test object in the chamber, excited

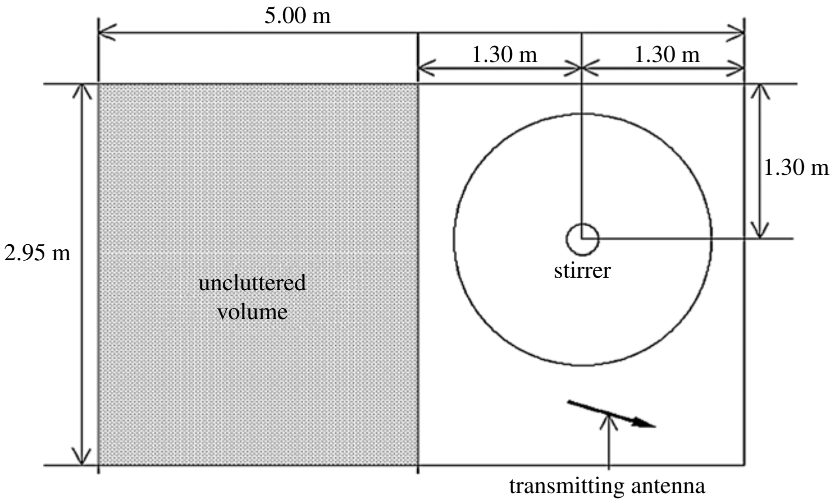


Figure 3. De Montfort University mode-stirred reverberation chamber configuration [8].

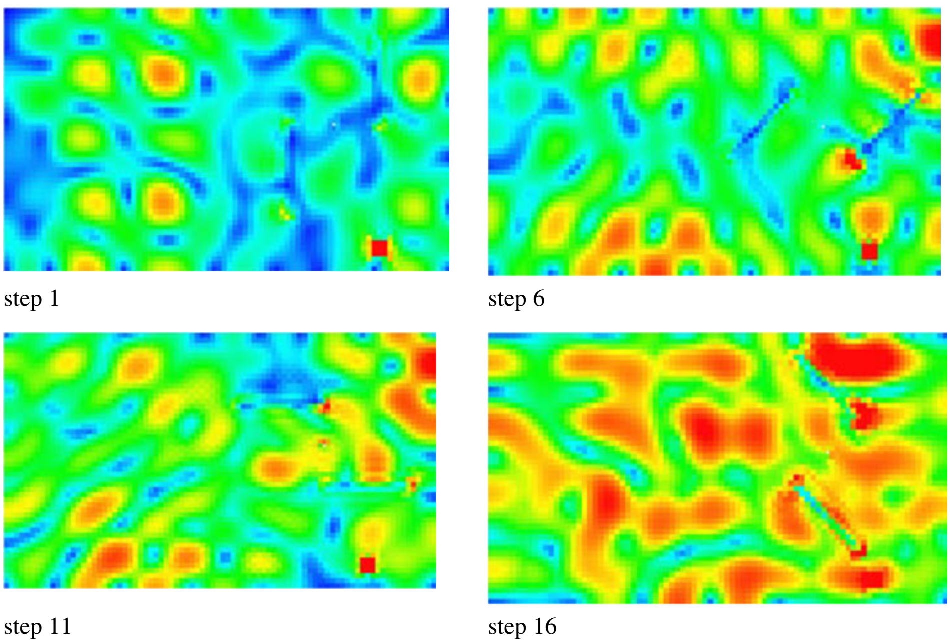


Figure 4. Field patterns in a simulation of the cross section of the reverberation chamber [8].

by an antenna (also in the chamber), receiving a statistically uniform field illumination. Any radiator in the chamber is, as a consequence, statistically isotropic. This means that, through one revolution of the stirrer, a maximum field coupling should occur between the source and receiver, irrespective of the polarity or orientation of either. That has the benefit of allowing tests to be undertaken in the chamber with the position of both the transmitter and receiver having little influence on the results obtained providing that the test object is placed within a 'working volume' [6].

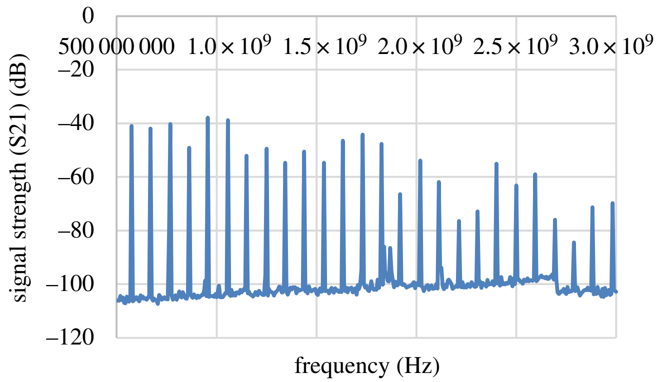


Figure 5. De Montfort comb generator performance when measured in a reverberation chamber. (Online version in colour.)

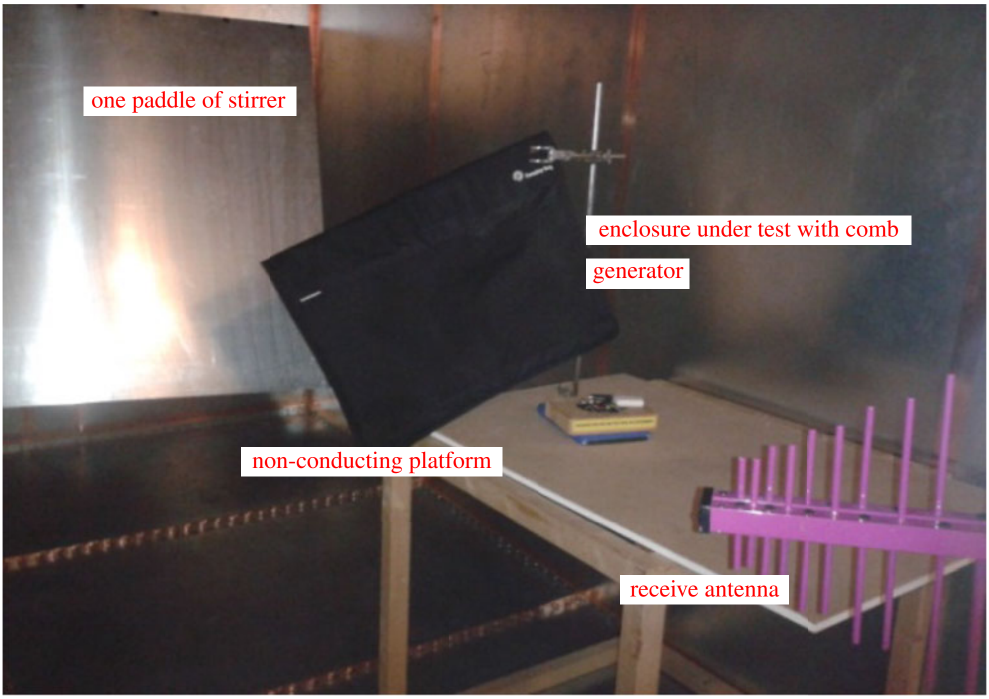


Figure 6. Equipment used for shielding measurements.

The reverberation chamber is a resonant cavity with a notional modal structure as given in equation (2.1). Notional is used because the presence of the internal stirrer will cause some variation from this.

$$f_{m,n,p} = \frac{c_0}{2}\sqrt{\left(\frac{m}{l}\right)^2 + \left(\frac{n}{w}\right)^2 + \left(\frac{p}{h}\right)^2}, \tag{2.1}$$

where c_0 is the speed of light; m, n, p are mode number integers and l, w, h are the length, width and height of the cavity, respectively.

According to BSi [7] the minimum number of cumulative modes that should be present for the tests to be valid is 60. A useful rule of thumb is also three times the fundamental frequency. These are generally relatively closely aligned. So, for example, the reverberation chamber at De Montfort University is illustrated in figure 3. The cumulative 60 modes occur at 174 MHz. The fundamental frequency is 58.6 MHz, meaning the 'three times' frequency is 178 MHz. Typically, for convenience, 200 MHz is used as the lowest frequency of record.

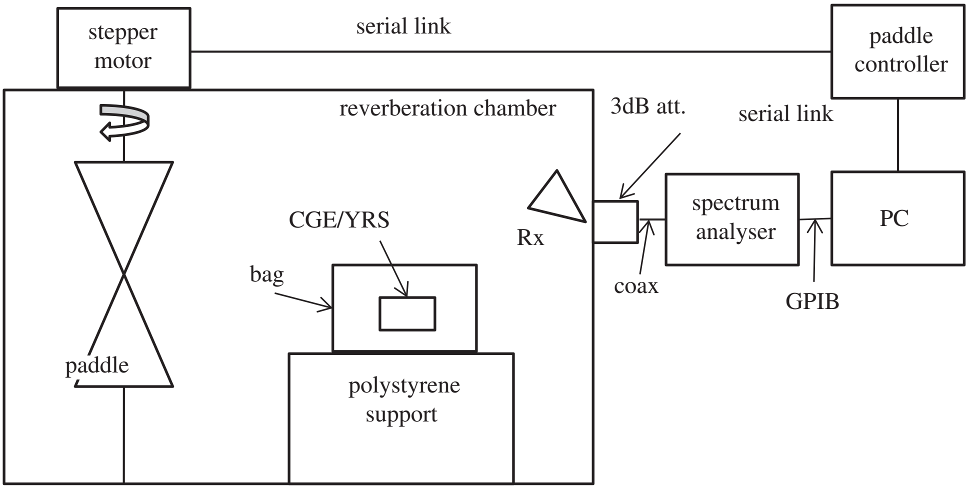


Figure 7. EY measurement configuration.

Figure 8. EY test configuration.

The effect of the stirrer can be seen in the sequence of simulations shown in figure 4, which is a simulation of the field strength of a horizontal cross section of the reverberation chamber as a function of stirrer position. Here, for the purpose of generating the illustration, 20 steps were used to complete a full revolution (as opposed to the 200 normally used for formal testing). The stirrer is two 1 m^2 vanes either side of the axis—they can be seen as parallel thin blue lines in the top right

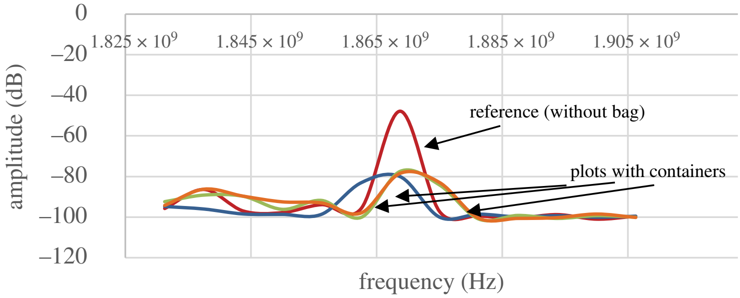


Figure 9. Identifying possible frequency drift between tests. (Online version in colour.)

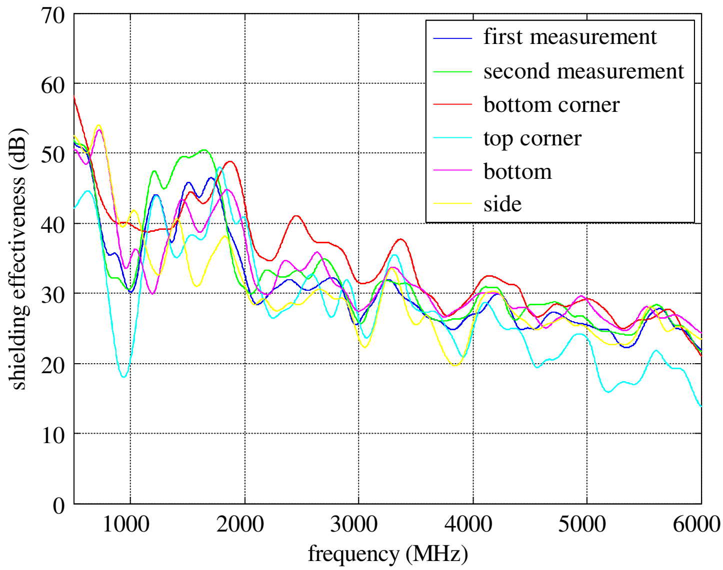


Figure 10. Comparison of measurements with signal source placed in different locations of a large flexible bag. (Online version in colour.)

corner of each pane. The red dot in the bottom right of each pane is the source. A temperature colour scale has been used with blue being low field strength and red high field strength.

(a) Test configuration

There are a number of specific requirements required for undertaking the shielding tests. The first is that any enclosed transmitter or receiver should allow the enclosure to be operated as designed. Any enclosed transmitter or receiver should be broadband, allowing early generations of cellular telephony to be tested alongside newer generations and emerging WiFi applications. Given, to a first approximation, the shields are reciprocal, a well-stirred chamber renders a source (and receiver) as statistically isotropic and the fields are statistically uniform, it was decided that placing a transmitter in the enclosure was the most sensible of approaches. In this case, the best approach for a broadband source would be a comb generator. A comb generator is a frequency source based on the generation of harmonics of a fundamental. The discrete frequencies produced, when viewed in the frequency domain, resemble a comb. Figure 5 shows the characteristic performance of the comb generator, when tested in the reverberation chamber of the comb generator used at De Montfort University. This was a simple battery powered

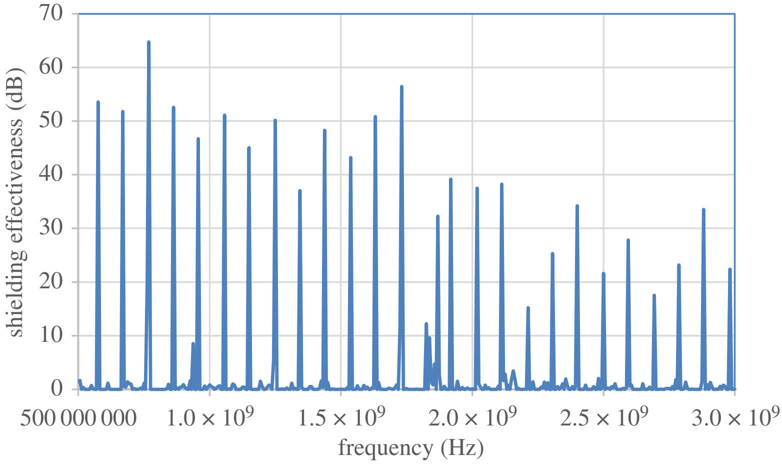


Figure 11. SE results in the DMU facility. (Online version in colour.)

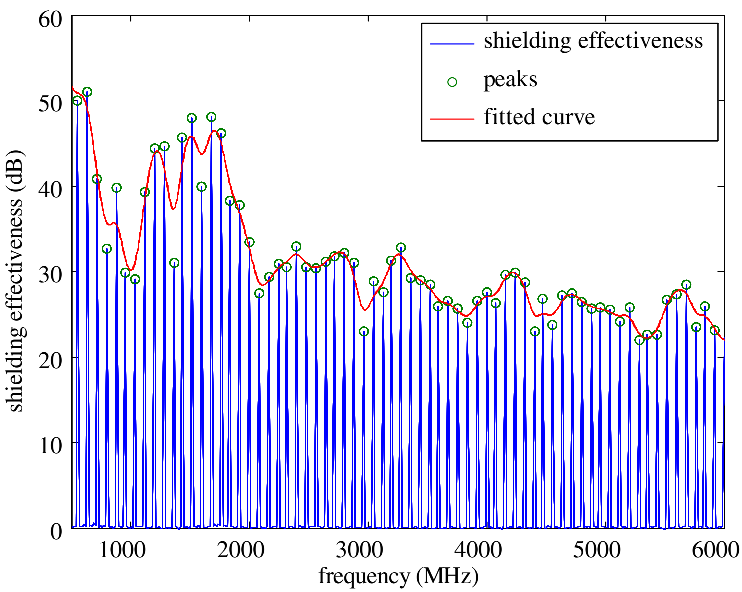


Figure 12. SE results in the EY facility. (Online version in colour.)

comb generator kit bought for a few pounds off the Internet coupled with a WiFi antenna. This approach allowed a 'compare and contrast' approach to be used against the professional-level comb generator used by Eurofins York, EY (formerly York EMC Services). This diagram also shows that there is an approximately 60 dB of dynamic range at approximately 1 GHz reducing to about 30 dB of dynamic range at approximately 3 GHz.

Following on from the definition in equation (1.1), the approach used to determine the shielding of any enclosure is to perform coupling measurements in the reverberation chamber with and without the enclosure shielding the comb generator.

One question that needed to be addressed is whether the presence of the shield would have any material effect on the location of the resonances due to the loading of the transmitter. Any such effect would make a comparison between the two measurements difficult to undertake. A simple experiment to consider loading was undertaken where a foil sheet was placed close to the comb

rsta.royalsocietypublishing.org *Phil. Trans. R. Soc. A* **376**: 20170456

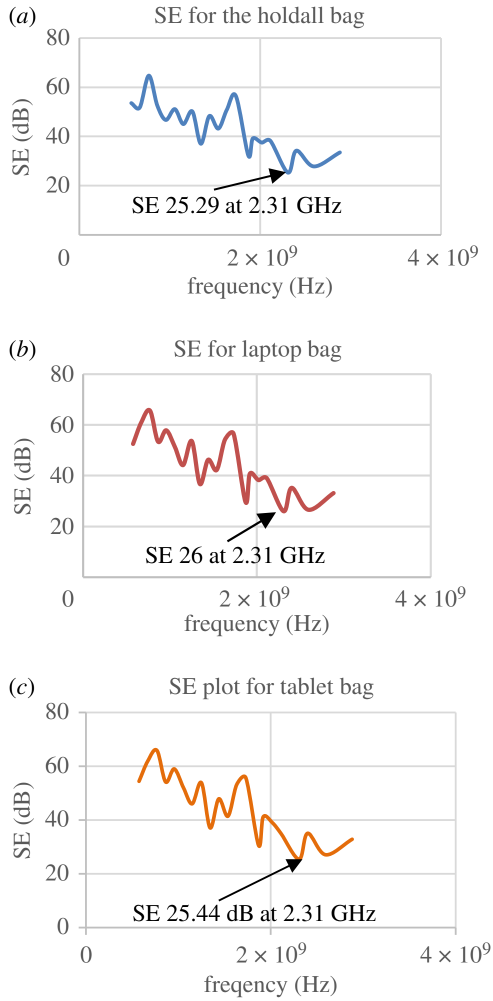


Figure 13. SE results for different enclosure sizes. (Online version in colour.)

generator. This showed that there was no change in the location of the resonances, only their amplitudes, as expected.

An illustration of the component elements of the tests in the De Montfort University reverberation chamber is given in figure 6, which shows the stirrer, an enclosure and the measurement antenna.

The comparator measurements were undertaken by Eurofins York (formerly York EMC Services), using a similar approach of having a broadband comb generator in a large reverberation chamber (similar in size to the De Montfort University Chamber). In this case, the comb generator was a professional-level noise source, the design of the paddle and number of steps were different (nearly full height, narrower and half the number of steps), the receive antenna was a wall mounted blade antenna rather than a bi-log antenna. Figure 7 shows the schematic and figure 8 shows an internal view of the chamber.

3. Results

The above experiments were undertaken to obtain the shielding measurements of a number of container designs.

The effect of any possible loading on the circuit was investigated to ensure that the peaks were consistent in frequency between reference tests and tests with the signal generator placed in the containers. Figure 9 shows representative results of one of those tests, located around one frequency point. It can be seen that the coupling value has changed but the location of the peaks is reasonably consistent. Note that this was with the low-cost comb generator at the De Montfort University facility, so some inherent drifting between experiments might be expected.

A further investigation about the effect on the placement of the signal generator within a large enclosure was undertaken. This was to identify the contribution that placement might make on the results. The comb generator was placed at various positions inside the container (in this case a bag designed to hold something like a large laptop) undertaken at the EY facility. The 'top' was close to the opening, which was folded over and fastened with hook-and-loop fasteners. Figure 10 shows the effect of placement, near to or away from various seams and fasteners, in the spirit of IEEE Standards 299 and 299.1 [9,10]. It can be seen that there is some variation between the various tests but, given the variability over the frequency range, it is difficult to argue that this is significant. One factor to note is that there is a greater reduction in SE at the higher frequencies when the signal generator is placed near the folded opening of the bag.

Tests were undertaken of the SE in the two different facilities. The DMU results are shown in figure 11. Note the upper frequency of 3 GHz. It can be seen that the SE is approximately 50 dB up to 1.75 GHz, reducing to 20–30 dB at 3 GHz. The comparative results from the EY facility are shown in figure 12. A very similar structure is seen, with values close to 50 dB being seen in the lower frequencies reducing to 20–30 dB at the middle to higher frequencies.

A further interesting investigation is the identification of the minimum SE of different sized enclosures. Figure 13 shows three different sized constructions, a holdall size, a laptop size and a tablet size. All three use the same materials and construction approach. It is interesting to note that the lowest SE occurs at the same frequencies and is of a very similar level, indicating that the limitation may very well be the material itself (although this is the subject of additional investigations).

4. Conclusion

This paper has investigated the shielding of containers used to provide radio-security for electronic devices with inherent connectivity. The purpose being to develop a measurement technique that is reliable between facilities in order to allow designers, manufacturers and specifiers to have confidence in the reported shielding of these products.

The foundations of much of the research involved in this paper can trace its origins to [1].

The results show that a reverberation chamber based test method, using a comb generator as a broadband noise source produces consistent results and is a strong candidate technique for [5].

An area for further study is the effect of the enclosure on the antenna used for the measurements. It is likely that there will be some performance change in the antenna with it either inside or outside the enclosure and should that be different for different types of antennas used for the measurement and in the devices the enclosure is designed for, the accuracy of the results would be enhanced by better understanding this.

Data accessibility. This article has no additional data.
Authors' contributions. All authors made approximately equal contribution to the work described in this paper.
Competing interests. We declare we have no competing interests.
Funding. We received no funding for this study.

References

1. Heaviside O. 1971 *Electromagnetic theory*, 3rd edn. New York: NY: Chelsea Publishing Co Int.
2. IEE. 1950 *Heavide centenary volume*. London, UK: IEE.
3. Doherty E. 2012 *Digital forensics for handheld devices*. Boca Raton, NJ: CRC Press.

4. Anusha VS, Nithya GK, Rao S. 2017 A comprehensive survey of electromnagetic propagation models. In *Int. Conf. on Communication and Signal Processing (ICCSP)*, Chennai, India. Piscataway, NJ: IEEE.
5. I. S. Association. 2016 P2710 Electromagnetic shielding performance of enclosures for portable electronic devices (5 February 2016). See https://development.standards.ieee.org/get-file/P2710.pdf?t=88569200003 (accessed 14 April 2018).
6. Coates A, Duffy A. 2007 Maximum working volume and minimum working frequency tradeoff in a reverberation chamber. *IEEE Trans. EMC* **49**, 719–722.
7. BSI. 2011 BS EN 61000-4-21:2011 Electromagnetic Compatibility (EMC) Part 4-21: Testing and measurement techniques—Reverberation chamber test methods.
8. Coates AR. 2004 Electromagnetic compatibility testing of communication cables. PhD thesis, De Montfort University, Leicester, UK.
9. 2006 *IEEE Std 299-2006, Method for Measuring the Effectiveness of Electromagnetic Shielding Enclosures*. Piscataway, NJ: IEEE.
10. IEEE. 2013 *IEEE Std 299.1-2013. Method for Measuring the Shielding Effectiveness of Enclosures and Boxes Having all Dimensions between 0.1 m and 2 m*. Piscataway, NJ: IEEE.

PHILOSOPHICAL TRANSACTIONS A

rsta.royalsocietypublishing.org

Research

Cite this article: Hampshire DP. 2018 A derivation of Maxwell's equations using the Heaviside notation. *Phil. Trans. R. Soc. A* **376**: 20170447.
http://dx.doi.org/10.1098/rsta.2017.0447

Accepted: 4 June 2018

One contribution of 13 to a theme issue 'Celebrating 125 years of Oliver Heaviside's 'Electromagnetic Theory''.

Subject Areas:
electromagnetism

Keywords:
Heaviside, Maxwell's equations, Coulomb's Law, Ampere's Law, electromagnetism, charge conservation

Author for correspondence:
Damian P. Hampshire
e-mail: d.p.hampshire@durham.ac.uk

A derivation of Maxwell's equations using the Heaviside notation

Damian P. Hampshire

Superconductivity Group, Centre for Materials Physics, Department of Physics, University of Durham, South Road, Durham DH1 3LE, UK

DPH, 0000-0001-8552-8514

Maxwell's four differential equations describing electromagnetism are among the most famous equations in science. Feynman said that they provide four of the seven fundamental laws of classical physics. In this paper, we derive Maxwell's equations using a well-established approach for deriving time-dependent differential equations from static laws. The derivation uses the standard Heaviside notation. It assumes conservation of charge and that Coulomb's law of electrostatics and Ampere's law of magnetostatics are both correct as a function of time when they are limited to describing a local system. It is analogous to deriving the differential equation of motion for sound, assuming conservation of mass, Newton's second law of motion and that Hooke's static law of elasticity holds for a system in local equilibrium. This work demonstrates that it is the conservation of charge that couples time-varying $\boldsymbol{E}$-fields and $\boldsymbol{B}$-fields and that Faraday's Law can be derived without any relativistic assumptions about Lorentz invariance. It also widens the choice of axioms, or starting points, for understanding electromagnetism.

This article is part of the theme issue 'Celebrating 125 years of Oliver Heaviside's 'Electromagnetic Theory''.

1. Introduction

This research paper is written in the celebration of 125 years of Oliver Heaviside's work *Electromagnetic*

THE ROYAL SOCIETY PUBLISHING

rsta.royalsocietypublishing.org *Phil. Trans. R. Soc. A* **376**: 20170447

theory [1]. Heaviside was broadly self-taught, an eccentric and a fabulous electrical engineer. He very probably first read Maxwell's great treatise on electricity and magnetism [2] while he was in the library of the Literary and Philosophical Society of Newcastle upon Tyne, just up the road from Durham [3]. He called Maxwell 'heaven-sent' and Faraday 'the prince of experimentalists' [1]. Heaviside restructured Maxwell's original 20 equations to be the four equations that we now recognize as Maxwell's equations. In every high school, good physics students can write down Newton's laws. In every university, they can write down Maxwell's equations in the mathematical form developed by Heaviside.

Axioms in mathematics play a central pedagogical role in learning and understanding this discipline. In 300 BC, Euclid wrote *Elements*, his seminal text about geometrical mathematics [4]. It included his 10 axioms and the proofs of more than 400 propositions or theorems. It has provided the template for the logical approach that students have used over the following two-and-a-half millennia. Students demonstrate their knowledge and skill by using the axioms as starting points and derive all the consequences that follow. Axioms in science are usually taken to be generally true, or at least very widely true, and are distinguished from those more limited statements or equations that can be derived from the axioms and then used to describe a particular system or to provide results for an examination. Hence, we expect the professional mathematical and scientific communities to specify the axioms of their disciplines clearly and mark the development of new knowledge by changes in axioms. There is also the expectation that the most useful axioms, as the Greek word axioma (self-evident truth or starting point) suggests, cannot be derived from other equations or laws.

Probably, the most famous physics textbook of modern times is the three-volume textbook *The Feynman lectures on physics*. In it, Feynman says 'we can understand the complete realm of classical physics' from just seven equations [5]. The first three equations describe forces: Newton's law of motion, Newton's law of gravity and the force law for a charged particle moving in a magnetic and electric field. The remaining four are Maxwell's differential equations. Students of electromagnetism are introduced to Maxwell's equations and taught that they are generally true, not least because of the overwhelming body of experimental data that validates them. Not only do they describe the $\boldsymbol{E}$-fields and $\boldsymbol{B}$-fields from charges and currents in vacuum but by considering the charges and currents produced in materials, they describe the fields produced by all the important technologically useful materials and an enormous range of physical phenomena in the world around us. They also include a prohibition on the creation of net charge that is consistent with all experimentation to date.

Maxwell's original work used a heuristic approach to derive 20 scalar equations that describe electromagnetism and was first to demonstrate that light is a transverse electromagnetic wave. The equations have a form that follows Newton and emphasize the electromotive force produced by electric and magnetic fields, as shown in table 1. Heaviside, who was first employed as an engineer in the British Post Office telegraph system in Newcastle upon Tyne [3], took the equations, eliminated the vector and scalar potentials and developed the differential vector calculus notation necessary to write them down in the form that we currently use [8]. Heaviside's form gives the $\boldsymbol{E}$- and $\boldsymbol{B}$-fields an importance beyond the forces they can produce and opens the way to describe wave and energy propagation more directly.

The historical development of electromagnetism has influenced its modern-day teaching. Undergraduate textbooks derive the electrostatic and magnetostatic differential equations mathematically from Coulomb's Law and Ampère's Law. However, to arrive at Maxwell's time-dependent equations, students follow the heuristic approach. Most science students are then taught relativity without understanding properly the axioms of classical electromagnetism. This is pedagogically unsound because if we do not make explicit the axioms of classical electromagnetism, in the (albeit unlikely) event that there are new experiments that are not consistent with current understanding, we undermine our students' ability to identify which axioms can be retained and which ones should be discarded. For example, many students think that Faraday's Law is axiomatic or that the postulates of relativity are required to derive Faraday's Law. In this paper, we show that Faraday's Law can be derived without using any of the

Table 1. Maxwell's 20 (scalar) equations in modern form, labelled with his original lettering notation (A)–(H) [6]. The first 18 of his equations, (A)–(F), are given here as six vector equations using Heaviside's curl notation. There are also two scalar equations, (G) and (H). We have avoided Maxwell's use of 'electromotive force' and 'actual electromotive force' and taken $\boldsymbol{\xi}$ as the electromotive force per unit charge. Also, ρ_{N} is the resistivity, ε_{r} is relative permittivity and μ_{r} is relative permeability. Standard symbols have their usual meanings [7]. Equations (A)–(D) and (G) include what are now known as Maxwell's four equations together with the expression for the Lorentz force. The subscript 'free' that is used now for charge densities and current densities that can travel over macroscopic distances, was called 'true conduction' by Maxwell. He also considered the magnetic vector potential $\boldsymbol{A}$ in terms of the electromagnetic momentum per unit charge.

$\boldsymbol{J}_{\mathrm{Total}} = \boldsymbol{J}_{\mathrm{free}} + \dfrac{\partial \boldsymbol{D}}{\partial t}$	(A)	$\boldsymbol{\xi} = \dfrac{\boldsymbol{D}}{\varepsilon_{\mathrm{r}}\varepsilon_0}$	(E)
$\mu_{\mathrm{r}}\mu_0\boldsymbol{H} = \nabla \times \boldsymbol{A}$	(B)	$\boldsymbol{\xi} = \rho_{\mathrm{N}}\boldsymbol{J}_{\mathrm{free}}$	(F)
$\nabla \times \boldsymbol{H} = \boldsymbol{J}_{\mathrm{Total}}$	(C)	$\nabla \cdot \boldsymbol{D} = \rho_{\mathrm{free}}$	(G)
$\boldsymbol{\xi} = -\nabla V - \dfrac{\partial \boldsymbol{A}}{\partial t} + \boldsymbol{v} \times \mu_{\mathrm{r}}\mu_0\boldsymbol{H}$	(D)	$\dfrac{\partial \rho_{\mathrm{free}}}{\partial t} = -\nabla \cdot \boldsymbol{J}_{\mathrm{free}}$	(H)

assumptions from Einstein's theory of relativity. Indeed, the derivations here beg the question as to whether Coulomb's Law, Ampère's Law and Faraday's Law should all have the status of laws at all, given that we can derive Faraday's Law from the other two.

The next section of this paper discusses the process by which static laws can be used to derive time-dependent differential equations. As an exemplar, it considers the textbook use of Hooke's static law of elasticity to derive the time-dependent differential equation that describes the propagation of sound. Section 3 uses a similar approach to derive Maxwell's equations. We apply the vector calculus approach developed by Heaviside [9] to derive all four of Maxwell's equations. Finally, we speculate about possible sources of experimental evidence for the breakdown of Maxwell's equations.

2. Deriving time-dependent differential equations from static laws

Scientists are well versed in using static laws to derive time-dependent partial differential equations. To derive the time-dependent differential equation for the propagation of sound, we start with Hooke's static law of elasticity, which when used to describe static equilibrium in a gas, can be written

$$p - p(0) = \frac{B}{\rho_{\mathrm{D}}(0)}(\rho_{\mathrm{D}} - \rho_{\mathrm{D}}(0)), \tag{2.1}$$

where B is the bulk modulus, $p(0)$ and $\rho_{\mathrm{D}}(0)$ are the initial pressure and density of the gas under test, p is the applied pressure and ρ_{D} the resultant density. Hooke's static law is then rewritten as

$$\left(\frac{\partial p}{\partial \rho_{\mathrm{D}}}\right)_t = \frac{B}{\rho_{\mathrm{D}}(0)}. \tag{2.2}$$

Equations (2.1) and (2.2) are quite different types of equations. Equation (2.1) relates how a change in the external pressure applied to a uniform and static gas changes the density throughout the entire gaseous system. Equation (2.2) is a differential equation that describes how a differential pressure across an infinitesimal volume causes a differential change in density. We note that equation (2.2) is derived by considering an element in which the cause (∂p) and the effect ($\partial \rho_{\mathrm{D}}$) are infinitesimally close together (i.e. local). We describe the gas as being in local equilibrium, so that even though the pressure and density can vary as a function of space and time, every point throughout the system has local values of ρ and p related by Hooke's Law. Local equilibrium also ensures that the differential of pressure with respect to space or time is related to an equivalent differential for density, for example, by an equation of the form

$$\left(\frac{\partial p}{\partial t}\right)_x = \frac{B}{\rho_D(0)}\left(\frac{\partial \rho_D}{\partial t}\right)_x. \tag{2.3}$$

However, it is important to note that strictly, it is *not possible* to derive either equation (2.2) or (2.3) from (2.1) using mathematics alone. In (2.1), ρ and p do not include the variables x and t (i.e. space and time), whereas in (2.2) and (2.3), they are functions of x and t. Textbooks do not usually emphasize that we have used our physical intuition and followed Occam's razor [10], so that among the many possible dependencies that include the time dependence for p and ρ_D, we have selected the one with the fewest additional assumptions.

To derive the equation that describes the propagation of sound, we then use Newton's second law of motion in the form

$$\left(\frac{\partial p}{\partial x}\right)_t = -\rho_D(0)\left(\frac{\partial u}{\partial t}\right)_x, \tag{2.4}$$

where u is the velocity and t is the time. Using the identity $(\partial p/\partial x)_t = (\partial p/\partial \rho_D)_t(\partial \rho_D/\partial x)_t$ and equation (2.2), Newton's Law gives

$$\left(\frac{\partial \rho_D}{\partial x}\right)_t = -\frac{\rho_D^2(0)}{B}\left(\frac{\partial u}{\partial t}\right)_x. \tag{2.5}$$

We then use the conservation of mass:

$$\left(\frac{\partial \rho_D}{\partial t}\right)_x = -\rho_D(0)\left(\frac{\partial u}{\partial x}\right)_t, \tag{2.6}$$

and partially differentiating (2.5) with respect to t and partially differentiating (2.6) with respect to x and allowing changes in the order of differentiation, we find

$$\left(\frac{\partial^2 u}{\partial t^2}\right)_x = \frac{B}{\rho_D(0)}\left(\frac{\partial^2 u}{\partial x^2}\right)_t \tag{2.7}$$

From (2.7), the velocity of sound v is given by $v^2 = B/\rho_D(0)$. The extension of Hooke's static law to the time domain allows us to describe a whole new range of phenomena associated with pressure waves (e.g. sound). However, this derivation also serves as a useful reminder of the limitations with this approach. In practice, the propagation of sound in a gas does not strictly obey (2.7), because propagation depends on how the temperature changes while the pressure is changing. Experimental results show that new physics, not found in the static measurements, is relevant in time-dependent systems (i.e. the rate of heat flow). There are other examples of systems in physics, where we start with a static law and can derive time-dependent differential equations some of which to first order do not require additional terms (e.g. in deriving dispersion relations such as the classical derivation for magnons using the Heisenberg spin Hamiltonian) and other examples where additional frictional terms are added (e.g. in describing energy loss such as in dispersive resonant polarisation in dielectrics). Hence, we emphasize that the validation of any time-dependent equations is ultimately an issue for experimentation. In this paper, we follow the simple approach described in equations (2.1)–(2.7). We postulate that if Coulomb's law of electrostatics and Ampère's law of magnetostatics are limited to describe what could be called 'local equilibrium'—a local point of observation with local charges and current densities (i.e. local cause and effect), and written in the most simple time-dependent form (invoking Occam's razor), then the derivatives of these laws, the differential equations with respect to time and space, hold throughout the whole system. The philosophy of this approach looks to make the system sufficiently general that it includes all the components necessary to provide general differential equations.

3. Derivations of Maxwell's four equations

(a) The divergence of $\boldsymbol{E}$

The most simple generalization of Coulomb's law of electrostatics, to a time-dependent form where the point of observation and the charges present are local, is

$$\boldsymbol{E}(\boldsymbol{r},t)=\lim_{\eta\to 0}\frac{1}{4\pi\varepsilon_0}\underset{\text{Primed spatial variables}}{\int}\frac{\hat{\boldsymbol{\eta}}}{\eta^2}\rho'(\boldsymbol{r}',t)\mathrm{d}\tau', \tag{3.1}$$

where as shown in figure 1, the electric field, $\boldsymbol{E}$, at a point of observation P located at a point $\boldsymbol{r}$ (x, y, z) and time t, is produced by the charge densities $\rho'(\boldsymbol{r}',t)$ located at the primed points $\boldsymbol{r}'$ (x', y', z') at the same time t. By definition, $\boldsymbol{\eta}=\boldsymbol{r}-\boldsymbol{r}'$and $\mathrm{d}\tau'$ denotes integrating over the primed spatial variables of the charge densities while the unprimed spatial variables remain constant. Neither the spatial variables x, y and z, nor x', y' and z' are functions of time. $\rho'(\boldsymbol{r}',t)$ is the charge density at position $\boldsymbol{r}'$ and time t. $\boldsymbol{E}$ is a function of the unprimed spatial variables x, y and z as well as time t. We assume that all the charge densities are local—very close to the point of observation. Hence, the partial time derivative of the $\boldsymbol{E}$-field at the point of observation is

$$\frac{\partial \boldsymbol{E}}{\partial t}=\lim_{\eta\to 0}\frac{1}{4\pi\varepsilon_0}\underset{\text{primed spatial variables}}{\int}\frac{\hat{\boldsymbol{\eta}}}{\eta^2}\frac{\partial\rho'}{\partial t}\mathrm{d}\tau', \tag{3.2}$$

where $\partial\rho'/\partial t$ is calculated at time t. As is standard convention, partial derivatives with respect to time are calculated assuming all spatial variables (i.e. primed and unprimed) are held constant. We state the standard definition of the del operator ∇:

$$\nabla=\hat{\mathbf{i}}\left.\frac{\partial}{\partial x}\right)_{y,z}+\hat{\mathbf{j}}\left.\frac{\partial}{\partial y}\right)_{x,z}+\hat{\mathbf{k}}\left.\frac{\partial}{\partial z}\right)_{x,y}, \tag{3.3}$$

and note that for this operator, in addition to the unprimed spatial variables that are explicitly shown to be held constant, for each of the partial derivatives, the variable t and the primed variables x', y' and z' are also held constant. Equations (3.1) and (3.3) lead to

$$\nabla\cdot\boldsymbol{E}=\frac{1}{4\pi\varepsilon_0}\lim_{\eta\to 0}\int\nabla\cdot\left(\frac{\hat{\boldsymbol{\eta}}}{\eta^2}\rho'(\boldsymbol{r}',t)\right)\mathrm{d}\tau'. \tag{3.4}$$

Using the identities $\nabla\cdot((\hat{\eta}/\eta^2)\rho')=\rho'\nabla\cdot(\hat{\boldsymbol{\eta}}/\eta^2)+(\hat{\boldsymbol{\eta}}/\eta^2)\cdot\nabla\rho'$and $\nabla\cdot(\hat{\boldsymbol{\eta}}/\eta^2)=4\pi\delta^3(\boldsymbol{\eta})$ and noting that ρ' only depends on primed variables and the time t, we obtain one of Maxwell's equations,

$$\nabla.\boldsymbol{E}=\frac{\rho}{\varepsilon_0}. \tag{3.5}$$

Equation (3.5) has the same form and uses similar mathematical identities to those used to derive the standard result from electrostatics. However, in this paper, we derive it from local time-dependent equations and then assume it is one of the underlying or fundamental differential equations that is correct at all points in space and time. Maxwell's equations have no agreed order. We call it Maxwell's first equation.

(b) The divergence of $\boldsymbol{B}$

We use Ampère's law of magnetostatics and again invoke Occam's razor to postulate that the local time-dependent $\boldsymbol{B}$-field at time t is

$$\boldsymbol{B}(\boldsymbol{r},t)=\lim_{\eta\to 0}\frac{\mu_0}{4\pi}\underset{\text{primed spatial variables}}{\int}\boldsymbol{J}'(\boldsymbol{r}',t)\times\frac{\hat{\boldsymbol{\eta}}}{\eta^2}\mathrm{d}\tau', \tag{3.6}$$

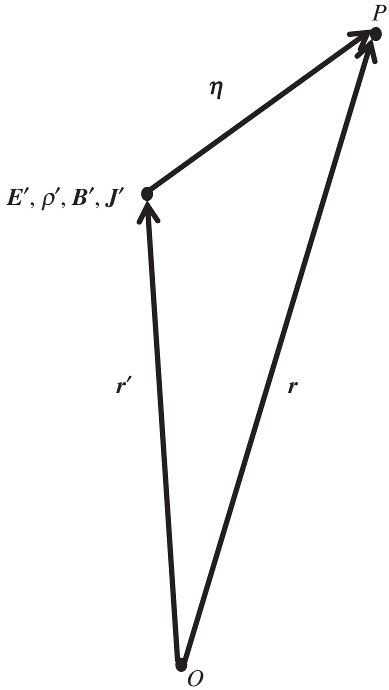


Figure 1. A coordinate system in which charge densities and current densities are observed. O is the origin. P is the point of observation. Charge densities and current densities are displaced from the origin at points $\boldsymbol{r}'$. The vector separation between the charge density $\rho'(\boldsymbol{r}')$ or current density $\boldsymbol{J}'(\boldsymbol{r}', t)$ and the observation point is given by $\boldsymbol{\eta} = \boldsymbol{r} - \boldsymbol{r}'$. The electric and magnetic fields at the primed locations are $\boldsymbol{E}'$ and $\boldsymbol{B}'$, respectively. For the special case of the electric field, charge density, magnetic field and current density at the point of observation, we use unprimed values $\boldsymbol{E}$, ρ, $\boldsymbol{B}$ and $\boldsymbol{J}$, respectively.

where $\boldsymbol{J}'$ is only a function of the primed spatial variables and the time is t. Again we assume that equation (3.6) is only valid for a system where all the current densities are local to the point of observation. We can also write

$$\frac{\partial \boldsymbol{B}}{\partial t} = \lim_{\eta \to 0} \frac{\mu_0}{4\pi} \int_{\text{primed spatial variables}} \frac{\partial \boldsymbol{J}}{\partial t} \times \frac{\hat{\boldsymbol{\eta}}}{\eta^2} \mathrm{d}\tau', \tag{3.7}$$

where $\partial \boldsymbol{J}'/\partial t$ is calculated at t. To improve brevity, we will omit including $\lim_{\eta \to 0}$ in the integral equations and the limits of the integration in this paper hereafter. Using a general vector field identity written in the form $\nabla \cdot (\boldsymbol{J}' \times \hat{\boldsymbol{\eta}}/\eta^2) = (\hat{\boldsymbol{\eta}}/\eta^2) \cdot (\nabla \times \boldsymbol{J}') - \boldsymbol{J}' \cdot (\nabla \times \hat{\boldsymbol{\eta}}/\eta^2)$ [7], given $\nabla \times \boldsymbol{J}' = 0$ (because ∇ is not primed but $\boldsymbol{J}'$ is primed), the divergence of (3.6) leads to

$$\nabla . \boldsymbol{B} = \frac{\mu_0}{4\pi} \int -\boldsymbol{J}' . \nabla \times \frac{\hat{\boldsymbol{\eta}}}{\eta^2} \mathrm{d}\tau'. \tag{3.8}$$

Using the vector field identity $\nabla \times (\hat{\boldsymbol{\eta}}/\eta^2) = 0$ leads to the second of Maxwell's equations:

$$\nabla . \boldsymbol{B} = 0. \tag{3.9}$$

This equation also has the same form and uses similar mathematical identities to those used to derive the standard result from magnetostatics. However, as noted above, we have derived it from local time-dependent equations and assume it is correct at all points in space and time.

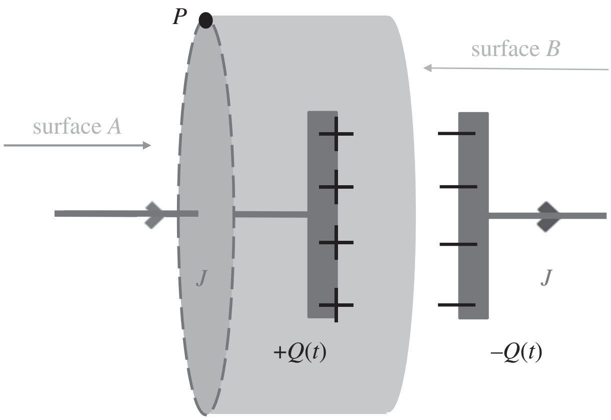


Figure 2. Straight wires carrying a constant current density $\boldsymbol{J}$ and charging two capacitor plates. Surface A is the flat circular surface bounded by the dotted ring path on which the point P is located and through which the current passes. Surface B is bounded by the same dotted ring path but passes through the capacitor plates so no current passes through it. The magnitude of the charge on each plate increases with time t and has magnitude $Q\,(t)$.

(c) The curl of B

(i) Maxwell's displacement current density

Textbooks [11] show, by taking the curl of both sides of Ampère's magnetostatic law, that

$$\nabla \times \boldsymbol{B} = \mu_0 \boldsymbol{J}. \tag{3.10}$$

Maxwell realized that equation (3.10) cannot be generally true as a function of time, given the vector field identity$\nabla \,.\, (\nabla \times \boldsymbol{B}) = 0$. By invoking the continuity of charge equation given by $\nabla \,.\, \boldsymbol{J} = -\partial\rho/\partial t$ and considering the partial time derivative of (3.5), he added his famous displacement current density term $\varepsilon_0 \partial \boldsymbol{E}/\partial t$ to equation (3.10) to give the third of his equations:

$$\nabla \times \boldsymbol{B} = \mu_0 \boldsymbol{J} + \mu_0 \varepsilon_0 \frac{\partial \boldsymbol{E}}{\partial t}. \tag{3.11}$$

Another approach used to justify the generalization from the magnetostatic equation (3.10) to the time-dependent equation (3.11) is found by considering figure 2. A current density flows in wires to charge capacitor plates and produces a changing $\boldsymbol{E}$-field between the plates. Using Stoke's theorem, one can rewrite (3.10) in terms of the line integral of the magnetic field around the path that bounds surface A and the surface integral across surface A where

$$\int \boldsymbol{B} \,.\, \mathrm{d}\boldsymbol{l} = \mu_0 \int \boldsymbol{J} \,.\, \mathrm{d}\boldsymbol{S}, \tag{3.12}$$

$\boldsymbol{J}$ is the current density in the wire and $\boldsymbol{S}$ is the cross-sectional area of the wire. However, for (3.12) to describe the $\boldsymbol{B}$-field produced in figure 2 correctly, the line integral $\int \boldsymbol{B}.\mathrm{d}\boldsymbol{l}$ must not depend on whether we choose surface A or surface B over which to complete the surface integral. The right-hand side of (3.12) is $\mu_0 \int \boldsymbol{J} \,.\, \mathrm{d}\boldsymbol{S}$ for surface A and zero for surface B because no current passes through surface B. To ensure that the line integral of $\boldsymbol{B}$ does not depend on whether surface A or surface B is chosen, and noting that the current density in the wire is given by $\boldsymbol{J} = \varepsilon_0 \partial \boldsymbol{E}/\partial \mathrm{t}$, where $\boldsymbol{E}$ is the field between the plates, one can add Maxwell's displacement current density term to (3.10) to produce (3.11). Maxwell's brilliant addition led to the unification of electricity and magnetism.

rsta.royalsocietypublishing.org Phil. Trans. R. Soc. A 376: 20170447

(ii) Maxwell's third equation

In deriving Maxwell's third (and fourth) equation, we assume the system is constrained by the conservation of charge. The constraint implies

$$\frac{\partial \rho'}{\partial t} = -\nabla' . \boldsymbol{J}', \tag{3.13}$$

where $\boldsymbol{J}'$ and ρ' are the current density and charge density at the point $\boldsymbol{r}'$. We have used the standard definition (cf equation (3.3))

$$\nabla' = \hat{\mathbf{i}}\left.\frac{\partial}{\partial x'}\right)_{y',z'} + \hat{\mathbf{j}}\left.\frac{\partial}{\partial y'}\right)_{x',z'} + \hat{\mathbf{k}}\left.\frac{\partial}{\partial z'}\right)_{x',y'}. \tag{3.14}$$

For this operator, similarly to equation (3.3), in addition to the primed spatial variables explicitly shown, the variable t and the non-primed variables x, y and z, are also held constant. Substituting (3.13) into (3.2) and then using standard vector field manipulations that include changing the order of partial derivatives and the vector field identity $\nabla(1/\eta) = -\hat{\boldsymbol{\eta}}/\eta^2$, we find that

$$\frac{\partial \boldsymbol{E}}{\partial t} = \frac{1}{4\pi\varepsilon_0}\nabla\int\frac{1}{\eta}\nabla'.\boldsymbol{J}'\mathrm{d}\tau'. \tag{3.15}$$

Textbook vector field algebraic techniques used in magnetostatics for functions of just three spatial variables can be used to rearrange the right-hand side of (3.15). Using the identity

$$\nabla'.\left[\frac{1}{\eta}\boldsymbol{J}'\right] = \frac{1}{\eta}\nabla'.\boldsymbol{J}' + \nabla'\left(\frac{1}{\eta}\right).\boldsymbol{J}'. \tag{3.16}$$

and then integrating gives

$$\int\frac{1}{\eta}\nabla'.\boldsymbol{J}'\mathrm{d}\tau' = \int\nabla'.\left[\frac{1}{\eta}\boldsymbol{J}'\right]\mathrm{d}\tau' - \int\nabla'\left(\frac{1}{\eta}\right).\boldsymbol{J}'\mathrm{d}\tau'. \tag{3.17}$$

The second of the three integrals in (3.17) can be written as a surface integral and then set to zero, because without loss of generality, we can assume $\boldsymbol{J}' = 0$ over the surface that defines τ'. Using (3.17), $\nabla'(1/\eta) = -\nabla(1/\eta)$, (3.16) with the del operator unprimed and $\nabla.\boldsymbol{J}' = 0$ (because ∇ is not primed but $\boldsymbol{J}'$ is primed) (3.15) then becomes:

$$\frac{\partial \boldsymbol{E}}{\partial t} = -\frac{1}{4\pi\varepsilon_0}\nabla\int\nabla'\left(\frac{1}{\eta}\right).\boldsymbol{J}'\mathrm{d}\tau' = \frac{1}{4\pi\varepsilon_0}\nabla\int\nabla\left(\frac{1}{\eta}\right).\boldsymbol{J}'\mathrm{d}\tau' = \frac{1}{4\pi\varepsilon_0}\nabla\int\nabla.\left(\frac{\boldsymbol{J}'}{\eta}\right)\mathrm{d}\tau'. \tag{3.18}$$

Using the vector field identity for the curl of the curl of a vector field, $\nabla^2(1/\eta) = -4\pi\delta^3(\boldsymbol{\eta})$, the vector field identity for the curl of the product of a vector field and a scalar, $\nabla\times\boldsymbol{J}' = 0$, and $\nabla(1/\eta) = -\hat{\boldsymbol{\eta}}/\eta^2$ gives

$$\begin{aligned}\frac{\partial \boldsymbol{E}}{\partial t} &= \frac{1}{4\pi\varepsilon_0}\int\nabla^2\left(\frac{\boldsymbol{J}'}{\eta}\right)\mathrm{d}\tau' + \frac{1}{4\pi\varepsilon_0}\int\nabla\times\left(\nabla\times\frac{\boldsymbol{J}'}{\eta}\right)\mathrm{d}\tau' \\ &= \frac{1}{4\pi\varepsilon_0}\int\boldsymbol{J}'\nabla^2\left(\frac{1}{\eta}\right)\mathrm{d}\tau' - \frac{1}{4\pi\varepsilon_0}\int\nabla\times\left(\frac{\hat{\boldsymbol{\eta}}}{\eta^2}\times\boldsymbol{J}'\right)\mathrm{d}\tau' = -\frac{\boldsymbol{J}}{\varepsilon_0} + \frac{1}{\varepsilon_0\mu_0}\nabla\times\boldsymbol{B}.\end{aligned} \tag{3.19}$$

Hence, we find the third of Maxwell's equations:

$$\nabla\times\boldsymbol{B} = \mu_0\boldsymbol{J} + \mu_0\varepsilon_0\frac{\partial \boldsymbol{E}}{\partial t}. \tag{3.20}$$

We can compare the mathematical approach we have used to derive (3.20) to Maxwell's heuristic approach. Maxwell considered a steady-state system, whereas this paper considers a

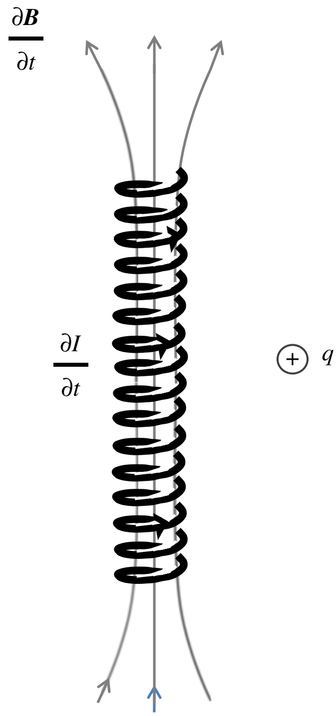


Figure 3. A stationary positive charge (q) outside a long solenoid. The current flowing in the solenoid is being increased (i.e. $\partial I/\partial t > 0$)as is the magnetic field inside the solenoid (i.e. $\partial \boldsymbol{B}/\partial t > 0$). (Online version in colour.)

local system. Both have invoked Occam's razor to generalize Coulomb's law of electrostatics to find an expression for $\partial \boldsymbol{E}/\partial t$ and are sufficiently general to find the same underlying differential equation (3.20).

(d) The curl of *E*

(i) Faraday's Law

Textbooks show that Coulomb's law for electrostatics [7] leads to

$$\nabla \times \boldsymbol{E} = 0. \tag{3.21}$$

Equations (3.5), (3.9), (3.10) and (3.21) in time-independent form are known as the equations of electrostatics and magnetostatics. The Helmholtz theorem tells us that a vector field is completely specified by knowing its divergence and curl [12]. To generalize (3.21) to include time dependence, Maxwell used Faraday's experimental results [13]. Faraday found that if the magnetic field is steadily increased inside a long solenoid, there is a force on a stationary charge outside the solenoid (cf. figure 3). He measured the force on such stationary charges using loops of metallic wires (carrying unbound stationary charges) attached to voltmeters. Faraday's experiments (together with Lenz's experiments [14]) can be described mathematically as

$$\nabla \times \boldsymbol{E} = -k\frac{\partial \boldsymbol{B}}{\partial t}. \tag{3.22}$$

where k is a constant of proportionality approximately equal to unity. In order to constrain k to be exactly unity, some textbooks then assume that (3.22) is invariant under a Galilean transformation and then further assume this result remains true even for systems where relativistic effects are important. Such assumptions are not employed in the derivation of Faraday's Law below.

(ii) Maxwell's fourth equation

We first consider the primed partial time derivative of equation (3.20) which is

$$\nabla' \times \frac{\partial \boldsymbol{B}'}{\partial t} = \mu_0 \frac{\partial \boldsymbol{J}'}{\partial t} + \mu_0\varepsilon_0 \frac{\partial^2 \boldsymbol{E}'}{\partial t^2}, \tag{3.23}$$

where $\boldsymbol{B}'$, $\boldsymbol{J}'$ and $\boldsymbol{E}'$ are the magnetic field, the current density and the $\boldsymbol{E}$-field at the position $\boldsymbol{r}'$. Substituting for $\partial \boldsymbol{J}'/\partial t$ in (3.7) and using the vector field identity

$$\nabla' \times (\nabla' \times \boldsymbol{E}') - \nabla'(\nabla' \cdot \boldsymbol{E}') + \nabla'^2 \boldsymbol{E}' = 0, \tag{3.24}$$

gives

$$\frac{\partial \boldsymbol{B}}{\partial t} = \frac{1}{4\pi} \int \left\{ \left[\nabla' \times \left(\frac{\partial \boldsymbol{B}'}{\partial t} + \nabla' \times \boldsymbol{E}' \right) \right] + \left[\nabla'^2 \boldsymbol{E}' - \nabla'(\nabla' \cdot \boldsymbol{E}') - \mu_0\varepsilon_0 \frac{\partial^2 \boldsymbol{E}'}{\partial t^2} \right] \right\} \times \frac{\hat{\boldsymbol{\eta}}}{\eta^2} \mathrm{d}\tau'. \tag{3.25}$$

Using the vector field identity

$$\nabla' \times \left(\frac{\rho' \hat{\boldsymbol{\eta}}}{\eta^2} \right) = \rho' \nabla' \times \left(\frac{\hat{\boldsymbol{\eta}}}{\eta^2} \right) + \nabla'(\rho') \times \left(\frac{\hat{\boldsymbol{\eta}}}{\eta^2} \right), \tag{3.26}$$

and given the curl of a radial function is zero, the second term in (3.26) is zero. Hence, using the vector field identity $\nabla(1/\eta) = -\hat{\boldsymbol{\eta}}/\eta^2$ and (3.26), we can rewrite the second term in the second square bracket of (3.25) as

$$\begin{aligned} \int (\nabla'(\nabla' \cdot \boldsymbol{E}')) \times \frac{\hat{\boldsymbol{\eta}}}{\eta^2} \mathrm{d}\tau' &= \frac{1}{\varepsilon_0} \int (\nabla'(\rho')) \times \frac{\hat{\boldsymbol{\eta}}}{\eta^2} \mathrm{d}\tau' = \frac{1}{\varepsilon_0} \int \nabla' \times \left(\frac{\rho'}{\eta^2} \hat{\boldsymbol{\eta}} \right) \mathrm{d}\tau' = -\frac{1}{\varepsilon_0} \int \nabla' \times \nabla \left(\frac{\rho'}{\eta} \right) \mathrm{d}\tau' \\ &= \frac{1}{\varepsilon_0} \int \nabla \times \nabla' \left(\frac{\rho'}{\eta} \right) \mathrm{d}\tau' \end{aligned} \tag{3.27}$$

Using the vector field identities $\nabla'(\rho'/\eta) = (1/\eta)\nabla'(\rho') + \rho'\nabla'(1/\eta)$ and $\nabla'(1/\eta) = \hat{\boldsymbol{\eta}}/\eta^2$ gives

$$\int \nabla'(\nabla' \cdot \boldsymbol{E}') \times \frac{\hat{\boldsymbol{\eta}}}{\eta^2} \mathrm{d}\tau' = \frac{1}{\varepsilon_0} \nabla \times \int \frac{1}{\eta} \nabla'(\rho') \mathrm{d}\tau' + \frac{1}{\varepsilon_0} \nabla \times \int \rho' \frac{\hat{\boldsymbol{\eta}}}{\eta^2} \mathrm{d}\tau'. \tag{3.28}$$

In using Coulomb's Law and Ampère's Law, we have ignored the internal structure of any element of charge density and current density. Hence without loss of generality, we assume that the volume occupied by every element of charge density and current density can be considered negligible and set the second integral in (3.28) to be zero. Therefore

$$\int \nabla'(\nabla' \cdot \boldsymbol{E}') \times \frac{\hat{\boldsymbol{\eta}}}{\eta^2} \mathrm{d}\tau' = 4\pi \nabla \times \boldsymbol{E}. \tag{3.29}$$

Substituting (3.29) into (3.25) then gives

$$\frac{\partial \boldsymbol{B}}{\partial t} = \frac{1}{4\pi} \int \left\{ \left[\nabla' \times \left(\frac{\partial \boldsymbol{B}'}{\partial t} + \nabla' \times \boldsymbol{E}' \right) \right] + \left[\nabla'^2 \boldsymbol{E}' - \mu_0\varepsilon_0 \frac{\partial^2 \boldsymbol{E}'}{\partial t^2} \right] \right\} \times \frac{\hat{\boldsymbol{\eta}}}{\eta^2} \mathrm{d}\tau' - \nabla \times \boldsymbol{E}. \tag{3.30}$$

Satisfying (3.30) (together with the other Maxwell equations (3.5) and (3.24) already derived) is equivalent to requiring

$$\nabla \times \boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\partial t}. \tag{3.31}$$

Equation (3.31) equates the left-hand side of (3.30) with the last term on the right-hand side. The primed version of (3.31) sets the first square bracket to zero. Given we already have Maxwell's equations (3.5) and (3.24), we can take them in primed form together with the primed version of (3.31) and the mathematical identity (3.24) to derive the wave-equation for $\boldsymbol{E}'$ propagating through vacuum (i.e. $\rho' = 0$ and $J' = 0$) which is of the form $\nabla'^2 \boldsymbol{E}' - \mu_0\varepsilon_0 \partial^2 \boldsymbol{E}'/\partial t^2 = 0$. This ensures the second square bracket in (3.30) is zero. Hence, given Maxwell's first three equations, (3.31) is a solution to (3.30). We note that an alternative solution to (3.30) is of the form of (3.21). In the context of this paper, Maxwell's first three equations together with equation (3.21) provide an alternative set of four time-dependent differential equations for electromagnetism. We put this set of equations aside as non-physical, because they imply that any change in charge density or current density would instantaneously change the $\boldsymbol{E}$-fields and $\boldsymbol{B}$-fields throughout the entire Universe. Equation (3.31) is Faraday's Law. It is the fourth of Maxwell's equations. We have shown that Coulomb's Law, Ampère's Law and the conservation of charge are sufficient to expect Faraday's Law and that the value for the constant k in equation (3.22) is not a matter for experimentation but is fixed to be unity. Faraday's Law can be derived without any relativistic assumptions about Lorentz invariance.

4. Are Maxwell's equations universally true?

The Jefimenko equations [15] are the general solutions to Maxwell's four equations where

$$\boldsymbol{E}(\boldsymbol{r},t) = \frac{1}{4\pi\varepsilon_0}\left\{\int \frac{\hat{\boldsymbol{\eta}}}{\eta^2}\rho'_r \mathrm{d}\tau' + \int \frac{\hat{\boldsymbol{\eta}}}{\eta c}\frac{\partial \rho'_r}{\partial t}\mathrm{d}\tau' - \int \frac{1}{\eta c^2}\frac{\partial \boldsymbol{J}'_r}{\partial t}\mathrm{d}\tau'\right\} \tag{4.1}$$

and

$$\boldsymbol{B}(\boldsymbol{r},t) = \frac{\mu_0}{4\pi}\left\{\int \boldsymbol{J}'_r \times \frac{\hat{\boldsymbol{\eta}}}{\eta^2}\mathrm{d}\tau' + \int \frac{\partial \boldsymbol{J}'_r}{\partial t} \times \frac{\hat{\boldsymbol{\eta}}}{\eta c}\mathrm{d}\tau'\right\}, \tag{4.2}$$

given that ρ'_r and $\boldsymbol{J}'_r$ are subject to the constraint of the conservation of charge (cf. equation (3.13)). In the Jefimenko equations, the charge density, ρ'_r, and current density, $\boldsymbol{J}'_r$, are calculated at the retarded time t_r where $t_\mathrm{r} = t - \eta/c$. Jefimenko pointed out that his equations show that the fields are caused by the charge densities and current densities in the system and that when Maxwell added the displacement current density to his fourth equation, he coupled $\nabla \times \boldsymbol{B}$ and $\partial \boldsymbol{E}/\partial t$ but did not introduce 'a cause and effect relationship' [15]. Similarly, this paper shows that Faraday's Law (in the differential form given by equation (3.31)) couples $\nabla \times \boldsymbol{E}$ and $\partial \boldsymbol{B}/\partial t$ as a result of the conservation of charge, but they also should not be considered to be in 'a cause and effect relationship'. It is interesting to identify those terms in the Jefimenko general solutions that were used by Maxwell and those used in this paper to help identify Maxwell's four underlying differential equations. Maxwell used Coulomb's Law and Faraday's Law associated with just two of the three terms for $\boldsymbol{E}(\boldsymbol{r},t)$ in equation (4.1). By also turning to Ampère's Law and the conservation of charge, he identified both terms for $\boldsymbol{B}(\boldsymbol{r},t)$ in equation (4.2). He avoided the complexities of retarded time by considering a steady-state system. In this paper, the central assumptions that Coulomb's electrostatic static law and Ampère's magnetostatic law are both true in the extreme local limit as a function of time is confirmed by Jefimenko's equations (i.e. the leading terms in equations (4.1) and (4.2) are equations (3.1) and (3.6), respectively). The complexity of retarded time is avoided by considering a local system where $t_\mathrm{r} \approx t$. The $\boldsymbol{E}$-fields and $\boldsymbol{B}$-fields are coupled by the conservation of charge.

The experimental evidence for Maxwell's equations is overwhelming. Furthermore, as the gateway to Einstein's theory of relativity [16], which in itself also brings its own compelling experimental evidence, any speculation about charge creation or the breakdown of Maxwell's equations, is very probably destined to be fruitless. However, because we have derived Maxwell's equations using the conservation of charge as a constraint, we complete the paper by considering what would happen if this constraint did not always apply, or more precisely, where might we look for the breakdown of Maxwell's equations. We suggest that the essence of an entity that has been *created* is that there should be no experimental methods that can determine the properties of the created entity prior to creation. The probability of the entity's existence can be considered to increase from zero to one. Looking for events described by this language of probability naturally points us towards quantum mechanics. Given that quantum mechanics has been tested to exquisite accuracy and that all known interactions conserve charge, it becomes a remote possibility at best, that we can find charge creation. Alternative tests of Maxwell's equations include looking for the creation of current density, or electric and magnetic waves that do not obey Maxwell's equations. Our best chances are to seek out events that are so difficult to produce that they have not been extensively interrogated experimentally, and hence may offer something completely unexpected. We suggest investigating an Einstein–Podolsky–Rosen experiment [17]. Typically, an entangled electron–positron pair is mixed and prepared as a superposition of states with equal and opposite magnetic moments (or spins). The charges are separated and the magnetic moment or spin of the electron is measured. The well-known instantaneous collapse of the wavefunction occurs, so that the positron ends up with the opposite magnetic moment (or spin) to the electron. The appearance of the moment of the positron is triggered by entirely quantum mechanical effects—no direct electromagnetic communication occurs between the electron and positron. Indeed, one can think about the two charges as a single entity. However, one can argue that we do not really know how the information that leads to the positron producing a magnetic moment of opposite sign is instantaneously received—beyond asserting it is part of the fabric of quantum mechanics, or part of the nature of a macroscopic wavefunction. We suggest that while the moment of the positron is being *created* (rather than excited), the production of the $\boldsymbol{B}$-field associated with its magnetic moment may not be coupled to the production of any $\boldsymbol{E}$-field at all. So, one could measure $(\partial \boldsymbol{E}/\partial t)_{\boldsymbol{r}}$ and $(\partial \boldsymbol{B}/\partial t)_{\boldsymbol{r}}$ in the wavefront of the positron, hoping to find $\boldsymbol{B}$-fields with $\boldsymbol{E}$-fields that are inconsistent with Maxwell's equations.

5. Conclusion and final comments

Although Euclid's choices of starting points or axioms for geometrical mathematics seem obvious, Feynman has reminded us that they are not a unique set. We can use Euclid's axioms to derive Pythagoras' theorem or we can take Pythagoras' theorem as an axiom and drop one of Euclid's axioms [18]. Hence, we have the paradox that although Euclid's choices have remained generally accepted for centuries because they are closest to being self-evident truths, none of them is indispensable. The choice of axioms ultimately includes some of the 'beauty is truth, truth beauty' [19] sentiment. At the moment, most of the scientific community uses Feynman's seven equations of classical physics, including Maxwell's equations, as axioms. If we discover charge creation, or electric and magnetic waves that do not obey Maxwell's equations, then treating Maxwell's equations as axioms would become untenable. In this paper, we have shown that Maxwell's equations can be justified using a mathematical derivation that follows from Coulomb's Law, Ampère's Law and the conservation of charge. Therefore, as with other differential equations in physics, in the unlikely event that Maxwell's equations are not true under all circumstances, we can discuss how the equations are derived and make other choices of axioms.

Data accessibility. The data and figures in the paper are available at: http://dx.doi.org/10.15128/gh93gz492 and associated materials are on the Durham Research Online website: http://dro.dur.ac.uk/26207/.

Competing interests. The author has no competing interests.

Funding. This work is funded by EPSRC grant nos EP/K502832/1 for the Durham Energy Institute and grant EP/L01663X/1 for the Fusion Energy Doctoral Training Network.

Acknowledgements. I very warmly thank those with whom I have discussed the nature of axioms and truth over many years. Most obviously with my wonderful wife and children, Amanda, Emily, Peter, Alexander and Michael, as well as with Andrew Davis, Douglas Halliday, Ifan Hughes, Martin McGovern, Kozo Osamura and Mark J Raine. Of course, none of them are responsible for any non-truths that may have slipped into this paper.

References

1. Heaviside O. 2005 *Electromagnetic theory*. London, UK: Elibron Classics.
2. Maxwell JC. 1873 *Treatise on electricity and magnetism*. Oxford, UK: Clarendon Press.
3. Spargo CM, Yakovlev A. 2015 Oliver Heaviside FRS: Newcastle-upon-Tyne 1868–1874. See http://dro.dur.ac.uk/18558/.
4. Euclid, Densmore D. 2007 *Euclid's elements*. Santa Fe, NM: Green Lion Press.
5. Feynman RP, Leighton RB, Sands M. 1977 *The Feynman lectures on physics*. Reading, MA: Addison-Wesley Publishing Company.

6. Maxwell JC. 1865 A dynamical theory of the electromagnetic field. *Phil. Trans. R. Soc. Lond.* **155**, 459–512. (doi:10.1098/rstl.1865.0008)
7. Woan G. 2003 *The Cambridge handbook of physics formulas*. Cambridge, UK: Cambridge University Press.
8. Yavetz I. 2005 *Landmark writings in western mathematics 1640–1940*. Amsterdam, The Netherlands: Elsevier B.V.
9. Nahin PJ. 2002 *Oliver Heaviside: the life, work, and times of an electrical genius of the Victorian age*. Baltimore, MD: JHU Press.
10. Poorna S. 2011 *Oxford English dictionary*. Oxford, UK: Oxford University Press.
11. Jackson JD. 1999 *Classical electrodynamics*. Hoboken, NJ: John Wiley and Sons.
12. Bladel J. 1958 *On Helmholtz's theorem in finite regions*. Wisconsin, NM: Midwestern Universities Research Association.
13. Day P. 1999 *The philosopher's tree: Michael Faraday's life and work in his own words*. Boca Raton, FL: CRC Press.
14. Lenz E. 1834 Ueber die Bestimmung der Richtung der durch elektodynamische Vertheilung erregten galvanischen Ströme. *Ann. Phys.* **107**, 483–494. (doi:10.1002/andp.18341073103)
15. Jefimenko OD. 1966 *Electricity and magnetism—an introduction to the theory of electric and magnetic fields*. Star City, WV: Meredith Publishing Company.
16. Einstein A. 1905 Zur Elektrodynamik bewegter Körper (On the electrodynamics of moving bodies). *Ann. Phys.* **17**, 891. (doi:10.1002/andp.19053221004)
17. Einstein A, Podolsky B, Rosen N. 1935 Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.* **47**, 777. (doi:10.1103/PhysRev.47.777)
18. Feynman RP. 1964 *The relation of mathematics and physics—online TUVA lecture*. Redmond, WA: Microsoft. See http://research.microsoft.com/apps/tools/tuva/index.html#data=4%7Cd71e62e2-0b19-4d82-978b-9c0ea0cbc45f%7C%7C.
19. Keats J. 1994 *The complete poems of John Keats*. Ware, UK: Wordsworth Editions Limited.

PHILOSOPHICAL TRANSACTIONS A

rsta.royalsocietypublishing.org

A catalogue of hidden momenta

David J. Griffiths

Department of Physics, Reed College, Portland, OR 97202, USA

DJG, 0000-0002-4406-4781

Research

Check for updates

Cite this article: Griffiths DJ. 2018 A catalogue of hidden momenta. *Phil. Trans. R. Soc. A* **376**: 20180043.
http://dx.doi.org/10.1098/rsta.2018.0043

Accepted: 14 May 2018

One contribution of 13 to a theme issue 'Celebrating 125 years of Oliver Heaviside's 'Electromagnetic Theory''.

Subject Areas:
electromagnetism

Keywords:
classical electrodynamics, electromagnetic momentum, hidden momentum, electric dipoles, magnetic dipoles, magnetic monopoles

Author for correspondence:
David J. Griffiths
e-mail: griffith@reed.edu

Electromagnetic fields carry momentum: $\mathbf{P}_{\text{em}} = \epsilon_0 \int (\mathbf{E} \times \mathbf{B})\, d\tau$. But if the centre of energy of a (localized) system is at rest, its *total* momentum must be zero. The compensating term has come to be called 'hidden' momentum: $\mathbf{P}_{\text{h}} = -\mathbf{P}_{\text{em}}$. It is (typically) ordinary *mechanical* momentum, relativistic in nature, and is 'hidden' only in the sense that it is not associated with motion of the system as a whole—only with that of its constituent parts. This article develops a catalogue of field momenta and hidden momenta for ideal electric and magnetic dipoles—both the 'standard' variety made from electric charges and currents and the 'anomalous' variety made from hypothetical magnetic monopoles and their currents—in the presence of electric and magnetic fields (which themselves may be produced by 'standard' or 'anomalous' sources).

This article is part of the theme issue 'Celebrating 125 years of Oliver Heaviside's 'Electromagnetic Theory''.

1. Electric and magnetic dipoles

In the static case, Maxwell's equations read

$$\left.\begin{aligned} &(a)\ \nabla \cdot \mathbf{E} = \left(\frac{1}{\epsilon_0}\right)\rho, && (c)\ \nabla \times \mathbf{E} = \mathbf{0}, \\ &(b)\ \nabla \cdot \mathbf{B} = 0, && (d)\ \nabla \times \mathbf{B} = \mu_0 \mathbf{J}, \end{aligned}\right\} \tag{1.1}$$

where ρ is the electric charge density and $\mathbf{J}$ is the electric current density. In a world with magnetic monopoles, there would also exist electromagnetic fields sourced by magnetic charges ($\tilde{\rho}$) and their currents ($\tilde{\mathbf{J}}$):

$$\left.\begin{aligned} &(a)\ \nabla \cdot \tilde{\mathbf{E}} = 0, && (c)\ \nabla \times \tilde{\mathbf{E}} = -\mu_0 \tilde{\mathbf{J}}, \\ &(b)\ \nabla \cdot \tilde{\mathbf{B}} = \mu_0 \tilde{\rho}, && (d)\ \nabla \times \tilde{\mathbf{B}} = \mathbf{0}. \end{aligned}\right\} \tag{1.2}$$

(I will use a tilde for these 'anomalous' sources and fields, to distinguish them from the 'standard'

THE ROYAL SOCIETY PUBLISHING

rsta.royalsocietypublishing.org Phil. Trans. R. Soc. A 376: 20180043

variety associated with ordinary electric charge.) It follows (by applying the divergence to equations (1.1(*d*)) and (1.2(*c*)) that the electric and magnetic currents are divergenceless:

$$\nabla\cdot\mathbf{J}=0,\quad \nabla\cdot\tilde{\mathbf{J}}=0, \tag{1.3}$$

(the associated charges are locally conserved).

The force on an electric charge (q) moving with velocity $\mathbf{v}$ is given by the Lorentz force law:

$$\mathbf{F}=q[\mathbf{E}'+(\mathbf{v}\times\mathbf{B}')], \tag{1.4}$$

where $\mathbf{E}'=\mathbf{E}+\tilde{\mathbf{E}}$ is the *total* electric field (standard plus anomalous), and $\mathbf{B}'=\mathbf{B}+\tilde{\mathbf{B}}$. Likewise, the force on a magnetic monopole ($\tilde{q}$) is

$$\mathbf{F}=\tilde{q}[\mathbf{B}'-\epsilon_0\mu_0(\mathbf{v}\times\mathbf{E}')]. \tag{1.5}$$

(The slight asymmetry in all these formulae is an unfortunate artefact of the SI system, and would not appear in Gaussian units.)

The fields can be expressed in terms of scalar and vector potentials:[1]

$$\left.\begin{aligned}&\mathbf{E}=-\nabla V, &&\mathbf{B}=\nabla\times\mathbf{A},\\ \text{and}\quad &\tilde{\mathbf{E}}=-\nabla\times\tilde{\mathbf{A}}, &&\tilde{\mathbf{B}}=-\nabla\tilde{V}.\end{aligned}\right\} \tag{1.6}$$

Adopting the gauge condition

$$\nabla\cdot\mathbf{A}=0,\quad \nabla\cdot\tilde{\mathbf{A}}=0, \tag{1.7}$$

Maxwell's equations become

$$\left.\begin{aligned}&\nabla^2V=-\frac{1}{\epsilon_0}\rho, &&\nabla^2\mathbf{A}=-\mu_0\mathbf{J},\\ \text{and}\quad &\nabla^2\tilde{V}=-\mu_0\tilde{\rho}, &&\nabla^2\tilde{\mathbf{A}}=-\mu_0\tilde{\mathbf{J}}.\end{aligned}\right\} \tag{1.8}$$

For localized charge and current configurations, it follows that

$$V(\mathbf{r})=\frac{1}{4\pi\epsilon_0}\int\frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,\mathrm{d}\tau',\quad \mathbf{A}(\mathbf{r})=\frac{\mu_0}{4\pi}\int\frac{\mathbf{J}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,\mathrm{d}\tau', \tag{1.9}$$

and

$$\tilde{V}(\mathbf{r})=\frac{\mu_0}{4\pi}\int\frac{\tilde{\rho}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,\mathrm{d}\tau',\quad \tilde{\mathbf{A}}(\mathbf{r})=\frac{\mu_0}{4\pi}\int\frac{\tilde{\mathbf{J}}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,\mathrm{d}\tau'. \tag{1.10}$$

The electric dipole moment of an electric charge distribution (ρ) is defined by

$$\mathbf{p}\equiv\int\mathbf{r}\rho\,\mathrm{d}\tau, \tag{1.11}$$

where $\mathbf{r}$ is the vector from the origin to the volume element $\mathrm{d}\tau$. The magnetic dipole moment of an electric current configuration ($\mathbf{J}$) is

$$\mathbf{m}\equiv\frac{1}{2}\int(\mathbf{r}\times\mathbf{J})\,\mathrm{d}\tau. \tag{1.12}$$

I shall call these 'standard' dipoles, to distinguish them from the hypothetical 'anomalous' variety associated with monopole charges and currents:

$$\tilde{\mathbf{m}}\equiv\int\mathbf{r}\tilde{\rho}\,\mathrm{d}\tau, \tag{1.13}$$

and

$$\tilde{\mathbf{p}}\equiv-\frac{\epsilon_0\mu_0}{2}\int(\mathbf{r}\times\tilde{\mathbf{J}})\,\mathrm{d}\tau. \tag{1.14}$$

[1]Many of the equations in §1 are taken from [1] and are recapitulated here in order to make the paper self-contained. However, I have reversed an unfortunate sign convention in the definition of $\tilde{\mathbf{A}}$.

rsta.royalsocietypublishing.org Phil. Trans. R. Soc. A 376: 20180043

If (as we shall always assume) the dipoles are *neutral*,

$$\int \rho \, \mathrm{d}\tau = 0, \quad \int \tilde{\rho} \, \mathrm{d}\tau = 0, \tag{1.15}$$

then $\mathbf{p}$ and $\tilde{\mathbf{m}}$ are independent of the choice of origin. It follows from equation (1.3) that

$$\begin{aligned} 0 = \int r_i(\boldsymbol{\nabla} \cdot \mathbf{J}) \, \mathrm{d}\tau = \int r_i(\nabla_j J_j) \, \mathrm{d}\tau = \int \nabla_j(r_i J_j) \, \mathrm{d}\tau - \int (\nabla_j r_i) J_j \, \mathrm{d}\tau \\ = -\int \delta_{ji} J_j \, \mathrm{d}\tau = -\int J_j \, \mathrm{d}\tau \end{aligned} \tag{1.16}$$

(repeated indices are to be summed from 1 to 3; we assume that charge and current distributions are *localized*, so all boundary terms coming from integration by parts vanish). Thus,

$$\int \mathbf{J} \, \mathrm{d}\tau = \mathbf{0}, \tag{1.17}$$

and in this case, the standard magnetic dipole moment ($\mathbf{m}$) is independent of origin. By the same reasoning,

$$\int \tilde{\mathbf{J}} \, \mathrm{d}\tau = \mathbf{0}, \tag{1.18}$$

and $\tilde{\mathbf{p}}$ is independent of origin.

Similarly,

$$\begin{aligned} 0 = \int r_i r_j (\nabla_k J_k) \, \mathrm{d}\tau = -\int [\nabla_k(r_i r_j)] J_k \, \mathrm{d}\tau = -\int (r_i \delta_{jk} + r_j \delta_{ik}) J_k \, \mathrm{d}\tau \\ = -\int (r_i J_j + r_j J_i) \, \mathrm{d}\tau. \end{aligned} \tag{1.19}$$

On the other hand, from equation (1.12),

$$\begin{aligned} \epsilon_{ijk} m_k = \frac{1}{2}(\epsilon_{ijk}\epsilon_{klm}) \int r_l J_m \, \mathrm{d}\tau = \frac{1}{2}\left(\delta_{il}\delta_{jm} - \delta_{im}\delta_{jl}\right) \int r_l J_m \, \mathrm{d}\tau \\ = \frac{1}{2}\int (r_i J_j - r_j J_i) \, \mathrm{d}\tau = \int r_i J_j \, \mathrm{d}\tau. \end{aligned} \tag{1.20}$$

By the same token,

$$\epsilon_{ijk} \tilde{p}_k = -\mu_0 \epsilon_0 \int r_i \tilde{J}_j \, \mathrm{d}\tau. \tag{1.21}$$

In general, the charge and current configurations constituting a physical dipole will be distributed over some finite region of space. However, we shall from now on confine our attention to 'ideal' dipoles, localized at a single point. More precisely, we are interested in the *limiting case*, in which the size of the dipole shrinks to zero.

To compute the potential of such a dipole, we note that

$$|\mathbf{r} - \mathbf{r}'| = \sqrt{r^2 + r'^2 - 2\mathbf{r} \cdot \mathbf{r}'}. \tag{1.22}$$

For an *ideal* dipole at the origin, the charge (or current) distribution at $\mathbf{r}'$ vanishes except at $\mathbf{r}' \to 0$, so (in equations (1.9) and (1.10)) we may safely confine our attention to the region $r' \ll r$, for which the binomial expansion gives

$$\frac{1}{|\mathbf{r} - \mathbf{r}'|} \approx \frac{1}{r}\left(1 + \frac{\mathbf{r} \cdot \mathbf{r}'}{r^2}\right). \tag{1.23}$$

In the case of a standard electric dipole,

$$V(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\left\{\frac{1}{r}\int \rho(\mathbf{r}') \, \mathrm{d}\tau' + \frac{\mathbf{r}}{r^3} \cdot \int \mathbf{r}' \rho(\mathbf{r}') \, \mathrm{d}\tau'\right\}, \tag{1.24}$$

or (using equations (1.11) and (1.15)),

$$V(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \frac{\mathbf{p} \cdot \mathbf{r}}{r^3}. \tag{1.25}$$

rsta.royalsocietypublishing.org Phil. Trans. R. Soc. A 376: 20180043

Similarly, the potential of a non-standard magnetic dipole is

$$\tilde{V}(\mathbf{r}) = \frac{\mu_0}{4\pi} \frac{\tilde{\mathbf{m}} \cdot \mathbf{r}}{r^3}. \tag{1.26}$$

For a standard magnetic dipole,

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi} \left\{ \frac{1}{r} \int \mathbf{J}(\mathbf{r}')\, \mathrm{d}\tau' + \frac{1}{r^3} \int (\mathbf{r} \cdot \mathbf{r}')\mathbf{J}(\mathbf{r}')\, \mathrm{d}\tau' \right\}, \tag{1.27}$$

or (using equations (1.17) and (1.20))

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi} \frac{\mathbf{m} \times \mathbf{r}}{r^3}. \tag{1.28}$$

Likewise, for an anomalous electric dipole:

$$\tilde{\mathbf{A}}(\mathbf{r}) = -\frac{1}{4\pi\epsilon_0} \frac{\tilde{\mathbf{p}} \times \mathbf{r}}{r^3}. \tag{1.29}$$

To determine the field of a dipole, we take the gradient or curl of the pertinent potential (equation (1.6)). This requires some care, however, because the dipole potentials are very singular at the origin. In general [2, eqn 6],

$$\nabla_i \left(\frac{r_j}{r^3} \right) = \frac{1}{r^3} \left(\delta_{ij} - 3\frac{r_i r_j}{r^2} \right) + \frac{4\pi}{3} \delta_{ij} \delta^3(\mathbf{r}), \tag{1.30}$$

so for any constant vector $\mathbf{a}$,

$$\nabla \left(\frac{\mathbf{a} \cdot \mathbf{r}}{r^3} \right) = \frac{1}{r^3} \left(\mathbf{a} - 3\frac{\mathbf{r}(\mathbf{a} \cdot \mathbf{r})}{r^2} \right) + \frac{4\pi}{3} \mathbf{a} \delta^3(\mathbf{r}) \tag{1.31}$$

and

$$\nabla \times \left(\frac{\mathbf{a} \times \mathbf{r}}{r^3} \right) = -\frac{1}{r^3} \left(\mathbf{a} - 3\frac{\mathbf{r}(\mathbf{a} \cdot \mathbf{r})}{r^2} \right) + \frac{8\pi}{3} \mathbf{a} \delta^3(\mathbf{r}). \tag{1.32}$$

Using these two identities, we find

$$\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \frac{1}{r^3} \left(3\frac{\mathbf{r}(\mathbf{r} \cdot \mathbf{p})}{r^2} - \mathbf{p} \right) - \frac{\mathbf{p}}{3\epsilon_0} \delta^3(\mathbf{r}), \tag{1.33}$$

$$\mathbf{B}(\mathbf{r}) = \frac{\mu_0}{4\pi} \frac{1}{r^3} \left(3\frac{\mathbf{r}(\mathbf{r} \cdot \mathbf{m})}{r^2} - \mathbf{m} \right) + \frac{2\mu_0 \mathbf{m}}{3} \delta^3(\mathbf{r}), \tag{1.34}$$

$$\tilde{\mathbf{E}}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \frac{1}{r^3} \left(3\frac{\mathbf{r}(\mathbf{r} \cdot \tilde{\mathbf{p}})}{r^2} - \tilde{\mathbf{p}} \right) + \frac{2\tilde{\mathbf{p}}}{3\epsilon_0} \delta^3(\mathbf{r}) \tag{1.35}$$

and

$$\tilde{\mathbf{B}}(\mathbf{r}) = \frac{\mu_0}{4\pi} \frac{1}{r^3} \left(3\frac{\mathbf{r}(\mathbf{r} \cdot \tilde{\mathbf{m}})}{r^2} - \tilde{\mathbf{m}} \right) - \frac{\mu_0 \tilde{\mathbf{m}}}{3} \delta^3(\mathbf{r}). \tag{1.36}$$

The delta-function terms are often left out, because one is usually interested in the field at some remove from the dipole; what remains has the same form in all four cases. This 'universal' part holds outside a sphere of vanishingly small radius; the delta-function describes the field inside this sphere. Although the latter contributes only at one point, it is essential for the internal consistency of the theory.

rsta.royalsocietypublishing.org Phil. Trans. R. Soc. A 376: 20180043

It is useful to note that this entire theory is invariant under the following duality transformation:

$\rho \to \frac{1}{c}\tilde{\rho}$	$\mathbf{J} \to \frac{1}{c}\tilde{\mathbf{J}}$	$\tilde{\rho} \to -c\rho$	$\tilde{\mathbf{J}} \to -c\mathbf{J}$
$\mathbf{E} \to c\tilde{\mathbf{B}}$	$\mathbf{B} \to -\frac{1}{c}\tilde{\mathbf{E}}$	$\tilde{\mathbf{E}} \to c\mathbf{B}$	$\tilde{\mathbf{B}} \to -\frac{1}{c}\mathbf{E}$
$V \to c\tilde{V}$	$\mathbf{A} \to \frac{1}{c}\tilde{\mathbf{A}}$	$\tilde{V} \to -\frac{1}{c}V$	$\tilde{\mathbf{A}} \to -c\mathbf{A}$
$\mathbf{p} \to \frac{1}{c}\tilde{\mathbf{m}}$	$\mathbf{m} \to -c\tilde{\mathbf{p}}$	$\tilde{\mathbf{p}} \to \frac{1}{c}\mathbf{m}$	$\tilde{\mathbf{m}} \to -c\mathbf{p}$

2. Field momentum

The (linear) momentum in electromagnetic fields is

$$\mathbf{P}_{\text{em}} = \epsilon_0 \int (\mathbf{E}' \times \mathbf{B}')\, d\tau \tag{2.1}$$

(the fields could be standard or anomalous, or—in principle—some of each). It is often more convenient to express this equation in terms of potentials; the resulting formula depends on the nature of the sources:

1. *Standard electric and magnetic fields*: $\mathbf{E} = -\nabla V$, so

$$\begin{aligned}\mathbf{P}_{\text{em}} &= -\epsilon_0 \int [(\nabla V) \times \mathbf{B}]\, d\tau = -\epsilon_0 \left[\int \nabla \times (V\mathbf{B})\, d\tau - \int V(\nabla \times \mathbf{B})\, d\tau\right] \\ &= \mu_0\epsilon_0 \int V(\mathbf{r})\mathbf{J}(\mathbf{r})\, d\tau.\end{aligned} \tag{2.2}$$

On the other hand, since[2] (equation (1.9))

$$\begin{aligned}V(\mathbf{r}) &= \frac{1}{4\pi\epsilon_0}\int \frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\, d\tau' \quad \text{and} \quad \mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi}\int \frac{\mathbf{J}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\, d\tau', \\ \mathbf{P}_{\text{em}} &= \mu_0\epsilon_0 \frac{1}{4\pi\epsilon_0} \iint \frac{\rho(\mathbf{r}')\mathbf{J}(\mathbf{r})}{|\mathbf{r}-\mathbf{r}'|}\, d\tau'\, d\tau = \int \left\{\frac{\mu_0}{4\pi}\int \frac{\mathbf{J}(\mathbf{r})}{|\mathbf{r}'-\mathbf{r}|}\, d\tau\right\} \rho(\mathbf{r}')\, d\tau' \\ &= \int \rho(\mathbf{r})\mathbf{A}(\mathbf{r})\, d\tau.\end{aligned} \tag{2.3}$$

2. *Standard electric field and anomalous magnetic field*: The first line of equation (2.2) still holds, but since $\nabla \times \tilde{\mathbf{B}} = \mathbf{0}$, the field momentum is zero:

$$\mathbf{P}_{\text{em}} = \mathbf{0}. \tag{2.4}$$

[2] Of course, you could go back to equation (2.1) and insert $\mathbf{B} = \nabla \times \mathbf{A}$, but it's a little easier to work from equation (2.2).

3. *Anomalous electric and magnetic fields*:

$$\mathbf{P}_{\text{em}} = \epsilon_0 \int (\tilde{\mathbf{E}} \times \tilde{\mathbf{B}})\,\mathrm{d}\tau = -\epsilon_0 \int (\tilde{\mathbf{E}} \times \boldsymbol{\nabla}\tilde{V})\,\mathrm{d}\tau = -\epsilon_0 \int \tilde{V}(\boldsymbol{\nabla} \times \tilde{\mathbf{E}})\,\mathrm{d}\tau$$
$$= \mu_0\epsilon_0 \int \tilde{V}(\mathbf{r})\tilde{\mathbf{J}}(\mathbf{r})\,\mathrm{d}\tau, \tag{2.5}$$

and since (equation (1.10))

$$\left.\begin{aligned} &\tilde{V}(\mathbf{r}) = \frac{\mu_0}{4\pi}\int \frac{\tilde{\rho}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,\mathrm{d}\tau' \quad \text{and} \quad \tilde{\mathbf{A}}(\mathbf{r}) = \frac{\mu_0}{4\pi}\int \frac{\tilde{\mathbf{J}}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,\mathrm{d}\tau', \\ \text{and hence} \qquad &\mathbf{P}_{\text{em}} = \mu_0\epsilon_0 \int \tilde{\rho}(\mathbf{r})\tilde{\mathbf{A}}(\mathbf{r})\,\mathrm{d}\tau. \end{aligned}\right\} \tag{2.6}$$

4. *Anomalous electric field and standard magnetic field*:

$$\mathbf{P}_{\text{em}} = \epsilon_0 \int \tilde{\mathbf{E}} \times (\boldsymbol{\nabla} \times \mathbf{A})\,\mathrm{d}\tau$$
$$= -\epsilon_0 \int \left[\mathbf{A} \times (\boldsymbol{\nabla} \times \tilde{\mathbf{E}}) + (\mathbf{A}\cdot\boldsymbol{\nabla})\tilde{\mathbf{E}} + (\tilde{\mathbf{E}}\cdot\boldsymbol{\nabla})\mathbf{A}\right]\mathrm{d}\tau$$
$$= \mu_0\epsilon_0 \int (\mathbf{A} \times \tilde{\mathbf{J}})\,\mathrm{d}\tau. \tag{2.7}$$

I used the fact that the ith component of $\int (\mathbf{A}\cdot\boldsymbol{\nabla})\tilde{\mathbf{E}}\,\mathrm{d}\tau$ is

$$\int \mathbf{A}\cdot\boldsymbol{\nabla}(\tilde{E}_i)\,\mathrm{d}\tau = -\int (\boldsymbol{\nabla}\cdot\mathbf{A})\tilde{E}_i\,\mathrm{d}\tau = 0, \tag{2.8}$$

and the same goes for $\int (\tilde{\mathbf{E}}\cdot\boldsymbol{\nabla})\mathbf{A}\,\mathrm{d}\tau$. Finally, using the now-familiar trick,

$$\mathbf{P}_{\text{em}} = \mu_0\epsilon_0 \int (\mathbf{A} \times \tilde{\mathbf{J}})\,\mathrm{d}\tau = \mu_0\epsilon_0 \iint \frac{\mu_0}{4\pi}\frac{\mathbf{J}(\mathbf{r}') \times \tilde{\mathbf{J}}(\mathbf{r})}{|\mathbf{r}-\mathbf{r}'|}\,\mathrm{d}\tau'\,\mathrm{d}\tau$$
$$= \mu_0\epsilon_0 \int (\mathbf{J} \times \tilde{\mathbf{A}})\,\mathrm{d}\tau. \tag{2.9}$$

I summarize these results as follows:

$\mathbf{E}$ and $\mathbf{B}$	$\mathbf{P}_{\text{em}} = \mu_0\epsilon_0 \int V\,\mathbf{J}\,d\tau = \int \rho\,\mathbf{A}\,d\tau$
$\mathbf{E}$ and $\tilde{\mathbf{B}}$	$\mathbf{P}_{\text{em}} = \mathbf{0}$
$\tilde{\mathbf{E}}$ and $\mathbf{B}$	$\mathbf{P}_{\text{em}} = \mu_0\epsilon_0 \int (\mathbf{A} \times \tilde{\mathbf{J}})\,d\tau = -\mu_0\epsilon_0 \int (\tilde{\mathbf{A}} \times \mathbf{J})\,d\tau$
$\tilde{\mathbf{E}}$ and $\tilde{\mathbf{B}}$	$\mathbf{P}_{\text{em}} = \mu_0\epsilon_0 \int \tilde{V}\,\tilde{\mathbf{J}}\,d\tau = \mu_0\epsilon_0 \int \tilde{\rho}\,\tilde{\mathbf{A}}\,d\tau$

Now we will use these formulae to determine the field momenta of electric and magnetic dipoles (both standard and anomalous) in external electric and magnetic fields (both standard

rsta.royalsocietypublishing.org Phil. Trans. R. Soc. A 376: 20180043

and anomalous). The dipoles are at rest (we might as well put them at the origin), and since they occupy an infinitesimal volume we can expand the external potentials:

$$V(\mathbf{r}) = V(\mathbf{0}) + \mathbf{r}\cdot(\boldsymbol{\nabla}_0 V) = V(\mathbf{0}) - \mathbf{r}\cdot\mathbf{E}(\mathbf{0}); \quad \mathbf{A}(\mathbf{r}) = \mathbf{A}(\mathbf{0}) + (\mathbf{r}\cdot\boldsymbol{\nabla}_0)\mathbf{A}; \tag{2.10}$$

and

$$\tilde{V}(\mathbf{r}) = \tilde{V}(\mathbf{0}) + \mathbf{r}\cdot(\boldsymbol{\nabla}_0 \tilde{V}) = \tilde{V}(\mathbf{0}) - \mathbf{r}\cdot\tilde{\mathbf{B}}(\mathbf{0}); \quad \tilde{\mathbf{A}}(\mathbf{r}) = \tilde{\mathbf{A}}(\mathbf{0}) + (\mathbf{r}\cdot\boldsymbol{\nabla}_0)\tilde{\mathbf{A}}. \tag{2.11}$$

(Here $\boldsymbol{\nabla}_0$ means 'evaluate the derivatives at $\mathbf{r}=\mathbf{0}$,' and we do not need any higher-order terms.)

1. *Standard electric dipole in standard magnetic field.* Use equation (2.3):

$$\begin{aligned}\mathbf{P}_{\text{em}} &= \int \rho(\mathbf{r})[\mathbf{A}(\mathbf{0}) + (\mathbf{r}\cdot\boldsymbol{\nabla}_0)\mathbf{A}]\,\mathrm{d}\tau \\ &= \left\{\int \rho(\mathbf{r})\,\mathrm{d}\tau\right\}\mathbf{A}(\mathbf{0}) + \left(\left\{\int \mathbf{r}\rho(\mathbf{r})\,\mathrm{d}\tau\right\}\cdot\boldsymbol{\nabla}_0\right)\mathbf{A} \\ &= (\mathbf{p}\cdot\boldsymbol{\nabla})\mathbf{A}.\end{aligned} \tag{2.12}$$

(I dropped the subscript on $\boldsymbol{\nabla}$; it is to be evaluated at the location of the dipole.)

2. *Standard magnetic dipole in standard electric field.* Use equation (2.2):

$$\mathbf{P}_{\text{em}} = \mu_0\epsilon_0\left\{V(0)\int \mathbf{J}(\mathbf{r})\,\mathrm{d}\tau - E_j(\mathbf{0})\int \mathbf{r}_j\mathbf{J}\,\mathrm{d}\tau\right\}. \tag{2.13}$$

But $\int \mathbf{J}\,\mathrm{d}\tau = \mathbf{0}$ and $\int r_i J_j\,\mathrm{d}\tau = \epsilon_{ijk}m_k$, so the jth component is

$$P_{\text{em}j} = -\mu_0\epsilon_0 E_i(\mathbf{0})\int \mathbf{r}_i J_j\,\mathrm{d}\tau = -\mu_0\epsilon_0 E_i(\mathbf{0})\epsilon_{ijk}m_k. \tag{2.14}$$

(I used equation (1.20)) and therefore

$$\mathbf{P}_{\text{em}} = \mu_0\epsilon_0(\mathbf{E}\times\mathbf{m}). \tag{2.15}$$

3. *Standard electric dipole in anomalous magnetic field, or anomalous magnetic dipole in standard electric field.* Equation (2.4) says

$$\mathbf{P}_{\text{em}} = \mathbf{0}. \tag{2.16}$$

4. *Anomalous electric dipole in anomalous magnetic field.* Use equation (2.5):

$$\left.\begin{aligned} & \mathbf{P}_{\text{em}} = \mu_0\epsilon_0\int[\tilde{V}(\mathbf{0}) - \mathbf{r}\cdot\tilde{\mathbf{B}}(\mathbf{0})]\tilde{\mathbf{J}}\,\mathrm{d}\tau. \\ & P_{\text{em}i} = -\mu_0\epsilon_0\tilde{B}_j(\mathbf{0})\int r_j\tilde{J}_i\,\mathrm{d}\tau = -\mu_0\epsilon_0\tilde{B}_j(\mathbf{0})\left(-\frac{1}{\mu_0\epsilon_0}\right)\epsilon_{jik}\tilde{p}_k \\ & \quad = -\epsilon_{ijk}\tilde{B}_j(\mathbf{0})\tilde{p}_k, \\ \text{so}\quad & \mathbf{P}_{\text{em}} = -\tilde{\mathbf{B}}\times\tilde{\mathbf{p}}. \end{aligned}\right\} \tag{2.17}$$

5. *Anomalous magnetic dipole in anomalous electric field.* Use equation (2.6):

$$\left.\begin{aligned} & \mathbf{P}_{\text{em}} = \mu_0\epsilon_0\int[\tilde{\mathbf{A}}(\mathbf{0}) + (\mathbf{r}\cdot\boldsymbol{\nabla}_0)\tilde{\mathbf{A}}]\tilde{\rho}(\mathbf{r})\,\mathrm{d}\tau. \\ & P_{\text{em}i} = \mu_0\epsilon_0(\nabla_{0j}\tilde{A}_i)\int r_j\tilde{\rho}(\mathbf{r})\,\mathrm{d}\tau = \mu_0\epsilon_0(\nabla_{0j}\tilde{A}_i)\tilde{m}_j \\ \text{so}\quad & \mathbf{P}_{\text{em}} = \mu_0\epsilon_0(\tilde{\mathbf{m}}\cdot\boldsymbol{\nabla})\tilde{\mathbf{A}}. \end{aligned}\right\} \tag{2.18}$$

6. *Anomalous electric dipole in standard magnetic field.* Use equation (2.7):

$$\left.\begin{aligned}
\mathbf{P}_{\text{em}} &= \mu_0\epsilon_0 \int [\mathbf{A}(\mathbf{0}) + (\mathbf{r}\cdot\boldsymbol{\nabla}_0)\mathbf{A}] \times \tilde{\mathbf{J}}(\mathbf{r})\,\mathrm{d}\tau.\\
P_{\text{em}i} &= \mu_0\epsilon_0\epsilon_{ijk}(\nabla_{0l}A_j)\int r_l\tilde{J}_k\,\mathrm{d}\tau = \mu_0\epsilon_0\epsilon_{ijk}(\nabla_{0l}A_j)\left(-\frac{1}{\mu_0\epsilon_0}\right)\epsilon_{lkm}\tilde{p}_m\\
&= -(\delta_{im}\delta_{jl}-\delta_{il}\delta_{jm})(\nabla_{0l}A_j)\tilde{p}_m = -(\nabla_0\cdot\mathbf{A})\tilde{p}_i + (\nabla_{0i}A_j)\tilde{p}_j,\\
\text{so}\qquad \mathbf{P}_{\text{em}} &= \tilde{p}_j\boldsymbol{\nabla}A_j = \boldsymbol{\nabla}(\tilde{\mathbf{p}}\cdot\mathbf{A}) = \tilde{\mathbf{p}}\times(\boldsymbol{\nabla}\times\mathbf{A}) + (\tilde{\mathbf{p}}\cdot\boldsymbol{\nabla})\mathbf{A}\\
&= (\tilde{\mathbf{p}}\times\mathbf{B}) + (\tilde{\mathbf{p}}\cdot\boldsymbol{\nabla})\mathbf{A}.
\end{aligned}\right\}\qquad (2.19)$$

(The form $\boldsymbol{\nabla}(\tilde{\mathbf{p}}\cdot\mathbf{A})$ is tidy but dangerous: the derivative does *not* act on $\tilde{p}$, only on $\mathbf{A}$.)

7. *Standard magnetic dipole in anomalous electric field.* Use equation (2.9):

$$\left.\begin{aligned}
\mathbf{P}_{\text{em}} &= -\mu_0\epsilon_0 \int [\tilde{\mathbf{A}}(\mathbf{0}) + (\mathbf{r}\cdot\boldsymbol{\nabla}_0)\tilde{\mathbf{A}}] \times \mathbf{J}(\mathbf{r})\,\mathrm{d}\tau.\\
P_{\text{em}i} &= -\mu_0\epsilon_0\epsilon_{ijk}(\nabla_{0l}\tilde{A}_j)\int r_l J_k\,\mathrm{d}\tau = -\mu_0\epsilon_0\epsilon_{ijk}(\nabla_{0l}\tilde{A}_j)\epsilon_{lkm}m_m\\
&= -\mu_0\epsilon_0(\delta_{im}\delta_{jl}-\delta_{il}\delta_{jm})(\nabla_{0l}\tilde{A}_j)m_m = \mu_0\epsilon_0(\nabla_{0i}\tilde{A}_j)m_j\\
\text{so}\qquad \mathbf{P}_{\text{em}} &= \mu_0\epsilon_0 m_j\boldsymbol{\nabla}\tilde{A}_j = \mu_0\epsilon_0\boldsymbol{\nabla}(\mathbf{m}\cdot\tilde{\mathbf{A}}) = \mu_0\epsilon_0[\mathbf{m}\times(\boldsymbol{\nabla}\times\tilde{\mathbf{A}}) + (\mathbf{m}\cdot\boldsymbol{\nabla})\tilde{\mathbf{A}}]\\
&= \mu_0\epsilon_0[-\mathbf{m}\times\tilde{\mathbf{E}} + (\mathbf{m}\cdot\boldsymbol{\nabla})\tilde{\mathbf{A}}].
\end{aligned}\right\}\qquad (2.20)$$

I summarize the results as follows:

$\mathbf{p}$ in $\mathbf{B}$	$\mathbf{P}_{\text{em}} = (\mathbf{p}\cdot\boldsymbol{\nabla})\mathbf{A}$	$\tilde{\mathbf{p}}$ in $\mathbf{B}$	$\mathbf{P}_{\text{em}} = (\tilde{\mathbf{p}}\times\mathbf{B}) + (\tilde{\mathbf{p}}\cdot\boldsymbol{\nabla})\mathbf{A}$
$\mathbf{m}$ in $\mathbf{E}$	$\mathbf{P}_{\text{em}} = -\mu_0\epsilon_0(\mathbf{m}\times\mathbf{E})$	$\tilde{\mathbf{m}}$ in $\mathbf{E}$	$\mathbf{P}_{\text{em}} = \mathbf{0}$
$\mathbf{p}$ in $\tilde{\mathbf{B}}$	$\mathbf{P}_{\text{em}} = \mathbf{0}$	$\tilde{\mathbf{p}}$ in $\tilde{\mathbf{B}}$	$\mathbf{P}_{\text{em}} = \tilde{\mathbf{p}}\times\tilde{\mathbf{B}}$
$\mathbf{m}$ in $\tilde{\mathbf{E}}$	$\mathbf{P}_{\text{em}} = -\dfrac{1}{c^2}[\mathbf{m}\times\tilde{\mathbf{E}} - (\mathbf{m}\cdot\boldsymbol{\nabla})\tilde{\mathbf{A}}]$	$\tilde{\mathbf{m}}$ in $\tilde{\mathbf{E}}$	$\mathbf{P}_{\text{em}} = \mu_0\epsilon_0(\tilde{\mathbf{m}}\cdot\boldsymbol{\nabla})\tilde{\mathbf{A}}$

3. Hidden momentum

Now, there is a general theorem [3,4] in special relativity that says 'if the centre of energy of a localized system is at rest, then the total momentum is zero.' In the cases, we are considering (stationary dipoles in static fields) the centre of energy is certainly not moving, and yet the *field* momentum is *not* zero, as we have seen. Evidently, there must be some *other* momentum, equal and opposite to $\mathbf{P}_{\text{em}}$. This other momentum has come to be called 'hidden' momentum [5,6],[3] though there is nothing secret about it—in the present context, it is perfectly ordinary mechanical momentum, relativistic in nature, and 'hidden' only in the sense that it is not associated with motion of the object (here, the dipole) as a whole, but rather with its internally moving parts.

The most illuminating example of hidden momentum goes back to Penfield & Haus in the mid-1960s [8]. Imagine a rectangular loop of wire, carrying a steady current I in the presence of a uniform electrostatic field $\mathbf{E}$ (figure 1). Picture the current as a resistanceless flow of free positive

[3] For a history, and comprehensive references, see [7].

rsta.royalsocietypublishing.org Phil. Trans. R. Soc. A 376: 20180043

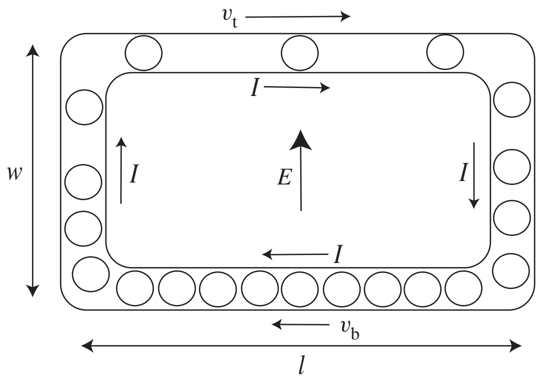


Figure 1. The Penfield–Haus model.

charges,[4] each with charge q and mass m. The electric field accelerates them as they ascend the left side, and slows them down as they descend the right side. Accordingly, their speed is greater along the top segment than at the bottom: $v_t > v_b$. On the other hand, they are further apart in the top segment, so there are more of them at the bottom: $N_b > N_t$. The current (which, remember, is constant around the loop) is

$$I = \frac{N_t q}{l} v_t = \frac{N_b q}{l} v_b \;\Rightarrow\; N_t v_t = N_b v_b = \frac{Il}{q}. \tag{3.1}$$

The net (relativistic) momentum of the charges—to the right—is

$$P_h = \gamma_t N_t m v_t - \gamma_b N_b m v_b = \frac{Ilm}{q}(\gamma_t - \gamma_b). \tag{3.2}$$

Now, the kinetic energy gained in ascending the left leg is equal to the work done by the electric force:

$$\gamma_t mc^2 - \gamma_b mc^2 = qEw \;\Rightarrow\; \gamma_t - \gamma_b = \frac{qEw}{mc^2}, \tag{3.3}$$

so

$$P_h = \left(\frac{Ilm}{q}\right)\left(\frac{qEw}{mc^2}\right) = \frac{1}{c^2}(Ilw)E. \tag{3.4}$$

But Ilw is the magnetic dipole moment of the loop; it points into the page. So

$$\mathbf{P}_h = \frac{1}{c^2}(\mathbf{m} \times \mathbf{E}) \quad [\mathbf{m} \text{ in } \mathbf{E}]. \tag{3.5}$$

This is the hidden momentum of the configuration—it is nothing but the net mechanical momentum of the charges constituting the current. It is independent of the *size* of the dipole (as long as $\mathbf{E}$ is constant over its area), so it applies in particular to an ideal (point) dipole. And it is just right to cancel the field momentum (equation (2.15))

$$\mathbf{P}_{em} = -\mu_0 \epsilon_0 (\mathbf{m} \times \mathbf{E}),$$

as required by the centre of energy theorem.

Notice that the Penfield–Haus mechanism applies to particles *in motion*. If, for example, this were an *anomalous* magnetic dipole, made from monopoles at rest, there would be *no* hidden momentum (in a standard electric field):

$$\mathbf{P}_h = \mathbf{0} \quad [\tilde{\mathbf{m}} \text{ in } \mathbf{E}]. \tag{3.6}$$

But, in that case, there is no *field* momentum either (equation (2.16)), and the total is again zero.

[4] This is not, of course, a realistic model of an actual current-carrying wire, but it makes the essential point with a minimum of extraneous baggage. For other models, see [9,10].

What if $\mathbf{E}$ is *not* uniform over the current loop? Consider a segment $d\mathbf{l}$; its momentum is

$$d\mathbf{P} = \gamma(\lambda_m\, dl)\mathbf{v} = \gamma\frac{\lambda_m}{\lambda_e}\lambda_e v\, d\mathbf{l} = \gamma(\alpha I)\, d\mathbf{l}, \tag{3.7}$$

where λ_m is the mass (of the moving charges) per unit length, λ_e is their charge per unit length and α is the mass-to-charge ratio). For the whole loop, then,

$$\mathbf{P}_\mathrm{h} = \alpha I \oint \gamma\, d\mathbf{l}. \tag{3.8}$$

Picking as the reference point for potential some convenient spot $\mathcal{O}$ on the loop,

$$\gamma(\mathbf{r}) = \gamma_0 + \int_{\mathcal{O}}^{\mathbf{r}} \frac{dW}{mc^2} = \gamma_0 + \frac{q}{mc^2}\int_{\mathcal{O}}^{\mathbf{r}} \mathbf{E}\cdot d\mathbf{l} = \gamma_0 - \frac{1}{\alpha c^2}V(\mathbf{r}), \tag{3.9}$$

where γ_0 is the value at $\mathcal{O}$, and dW is the work done on a charge as it advances by $d\mathbf{l}$ along the loop. But

$$\oint \gamma_0\, d\mathbf{l} = \gamma_0 \oint d\mathbf{l} = \mathbf{0}, \tag{3.10}$$

and hence[5]

$$\mathbf{P}_\mathrm{h} = -\frac{I}{c^2}\oint V(\mathbf{r})\, d\mathbf{l}. \tag{3.11}$$

For example, if the source of the electric field is an ordinary electric dipole, $\mathbf{p}$, (I'm changing the reference point, but the closed line integral is independent of any added constant)

$$\mathbf{P}_\mathrm{h} = -\frac{I}{c^2}\frac{1}{4\pi\epsilon_0}\oint \frac{[\mathbf{p}\cdot(\mathbf{r}-\mathbf{r}')]}{|\mathbf{r}-\mathbf{r}'|^3}\, d\mathbf{l} = -\frac{\mu_0}{4\pi}\oint \mathbf{I}\frac{[\mathbf{p}\cdot(\mathbf{r}-\mathbf{r}')]}{|\mathbf{r}-\mathbf{r}'|^3}\, dl, \tag{3.12}$$

where $(\mathbf{r}-\mathbf{r}')$ is the vector from the dipole (at $\mathbf{r}'$) to the point $\mathbf{r}$. (It does not matter whether you associate the vector with I or with dl, since they are in the same direction: $I\, d\mathbf{l} = \mathbf{I}\, dl$.) We can express this result in terms of the vector potential (due to the current loop) at the location of the electric dipole:

$$\mathbf{A}(\mathbf{r}') = \frac{\mu_0}{4\pi}\oint \frac{\mathbf{I}(\mathbf{r})}{|\mathbf{r}'-\mathbf{r}|}\, dl. \tag{3.13}$$

Thus

$$[(\mathbf{p}\cdot\boldsymbol{\nabla}')\mathbf{A}(\mathbf{r}')]_i = \frac{\mu_0}{4\pi}\oint I_i p_j \nabla'_j\left(\frac{1}{|\mathbf{r}'-\mathbf{r}|}\right) dl = -\frac{\mu_0}{4\pi}\oint I_i p_j \frac{(\mathbf{r}'-\mathbf{r})_j}{|\mathbf{r}'-\mathbf{r}|^3}\, dl, \tag{3.14}$$

or

$$[(\mathbf{p}\cdot\boldsymbol{\nabla}')\mathbf{A}(\mathbf{r}')] = -\frac{\mu_0}{4\pi}\oint \mathbf{I}\frac{[\mathbf{p}\cdot(\mathbf{r}'-\mathbf{r})]}{|\mathbf{r}'-\mathbf{r}|^3}\, dl. \tag{3.15}$$

So the hidden momentum is

$$\mathbf{P}_\mathrm{h} = -(\mathbf{p}\cdot\boldsymbol{\nabla})\mathbf{A}, \quad [\mathbf{p} \text{ in } \mathbf{B}], \tag{3.16}$$

which once again is just right to cancel the field momentum (equation (2.12)).

The same argument applies to an anomalous *magnetic* dipole in an anomalous *electric* field, except that what does the work is now the magnetic force ($\mathbf{F} = \tilde{q}\tilde{\mathbf{B}}$) acting

[5]For volume currents equation (3.11) becomes $\mathbf{P}_\mathrm{h} = -(I/c^2)\int V\mathbf{J}\, d\tau$, and we see immediately that it cancels the field momentum (equation (2.2)). If $\mathbf{E}$ is *uniform* over the current region, then $V(\mathbf{r}) = V(\mathbf{0}) - \mathbf{E}\cdot\mathbf{r}$, and (using equation (1.20)) $\mathbf{P}_{\mathrm{h}_j} = \mu_0\epsilon_0 E_i \int r_i J_j\, d\tau = \mu_0\epsilon_0 E_i \epsilon_{ijk} m_k = \mu_0\epsilon_0(\mathbf{m}\times\mathbf{E})_j$, so we recover equation (3.5).

on the particles in the monopole current loop:

$$\mathbf{P}_\mathrm{h} = -\frac{1}{c^2}(\tilde{\mathbf{m}}\cdot\boldsymbol{\nabla})\tilde{\mathbf{A}} \quad [\tilde{\mathbf{m}} \text{ in } \tilde{\mathbf{E}}]. \tag{3.17}$$

Alternatively, you can get equation (3.17) by applying the duality transformation to equation (3.16). Likewise, from equations (3.5) and (3.6),

$$\mathbf{P}_\mathrm{h} = -(\tilde{\mathbf{p}}\times\tilde{\mathbf{B}}) \quad [\tilde{\mathbf{p}} \text{ in } \tilde{\mathbf{B}}], \tag{3.18}$$

and

$$\mathbf{P}_\mathrm{h} = \mathbf{0} \quad [\mathbf{p} \text{ in } \tilde{\mathbf{B}}]. \tag{3.19}$$

The original Penfield–Haus model made no reference to the *source* of the electrostatic field—they presumably took it to be some collection of stationary electric charges (perhaps in the form of a surrounding parallel-plate capacitor). The hidden momentum in the magnetic dipole itself would be the same (equation (3.5)) if the electric field were due to a current of magnetic monopoles ($(1/c^2)(\mathbf{m}\times\tilde{\mathbf{E}})$). However, in that case there would *also* be hidden momentum residing in the monopole current (the monopoles accelerating and decelerating in response to the magnetic field of the electric current loop). The latter is given by equation (3.17) (for the momentum *in the monopole current loop* it does not matter whether the source of the magnetic field is standard or anomalous). Combining the two we get

$$\mathbf{P}_\mathrm{h} = \frac{1}{c^2}[(\mathbf{m}\times\tilde{\mathbf{E}}) - (\mathbf{m}\cdot\boldsymbol{\nabla})\tilde{\mathbf{A}}] \quad (\mathbf{m} \text{ in } \tilde{\mathbf{E}}). \tag{3.20}$$

By the same argument (or by invoking the duality transformation)

$$\mathbf{P}_\mathrm{h} = -[(\tilde{\mathbf{p}}\times\mathbf{B}) + (\tilde{\mathbf{p}}\cdot\boldsymbol{\nabla})\mathbf{A}] \quad (\tilde{\mathbf{p}} \text{ in } \mathbf{B}). \tag{3.21}$$

The following catalogue summarizes these results:

$\mathbf{p}$ in $\mathbf{B}$	$\mathbf{P}_\mathrm{h} = -(\mathbf{p}\cdot\boldsymbol{\nabla})\mathbf{A}$	resides in source of $\mathbf{B}$
$\mathbf{m}$ in $\mathbf{E}$	$\mathbf{P}_\mathrm{h} = \frac{1}{c^2}(\mathbf{m}\times\mathbf{E})$	resides in $\mathbf{m}$
$\mathbf{p}$ in $\tilde{\mathbf{B}}$	$\mathbf{P}_\mathrm{h} = \mathbf{0}$	(nothing moving)
$\mathbf{m}$ in $\tilde{\mathbf{E}}$	$\mathbf{P}_\mathrm{h} = \frac{1}{c^2}[\mathbf{m}\times\tilde{\mathbf{E}} - (\mathbf{m}\cdot\boldsymbol{\nabla})\tilde{\mathbf{A}}]$	resides in $\mathbf{m}$ and source of $\tilde{\mathbf{E}}$
$\tilde{\mathbf{p}}$ in $\mathbf{B}$	$\mathbf{P}_\mathrm{h} = -[(\tilde{\mathbf{p}}\times\mathbf{B}) + (\tilde{\mathbf{p}}\cdot\boldsymbol{\nabla})\mathbf{A}]$	resides in $\tilde{\mathbf{p}}$ and source of $\mathbf{B}$
$\tilde{\mathbf{m}}$ in $\mathbf{E}$	$\mathbf{P}_\mathrm{h} = \mathbf{0}$	(nothing moving)
$\tilde{\mathbf{p}}$ in $\tilde{\mathbf{B}}$	$\mathbf{P}_\mathrm{h} = -(\tilde{\mathbf{p}}\times\tilde{\mathbf{B}})$	resides in $\tilde{\mathbf{p}}$
$\tilde{\mathbf{m}}$ in $\tilde{\mathbf{E}}$	$\mathbf{P}_\mathrm{h} = -\frac{1}{c^2}(\tilde{\mathbf{m}}\cdot\boldsymbol{\nabla})\tilde{\mathbf{A}}$	resides in source of $\mathbf{E}$

rsta.royalsocietypublishing.org Phil. Trans. R. Soc. A 376: 20180043

In each case, the hidden momentum is just right to cancel the field momentum.

4. Interacting dipoles

As an application, suppose that the field is itself due to another dipole. There are four possibilities: (1) $\mathbf{p}$ and $\mathbf{m}$, (2) $\mathbf{p}$ and $\tilde{\mathbf{m}}$, (3) $\tilde{\mathbf{p}}$ and $\mathbf{m}$ and (4) $\tilde{\mathbf{p}}$ and $\tilde{\mathbf{m}}$. We could regard the first as a standard magnetic dipole in the electric field of a standard electric dipole:

$$\begin{aligned}\mathbf{P}_{\mathrm{h}} &= \frac{1}{c^2}(\mathbf{m}\times\mathbf{E}) = \mu_0\epsilon_0\mathbf{m}\times\left\{\frac{1}{4\pi\epsilon_0}\frac{[3(\mathbf{p}\cdot\hat{\mathbf{r}})\hat{\mathbf{r}}-\mathbf{p}]}{r^3} - \frac{1}{3\epsilon_0}\mathbf{p}\delta^3(\mathbf{r})\right\}\\ &= \frac{\mu_0}{4\pi r^3}[3(\mathbf{p}\cdot\hat{\mathbf{r}})(\mathbf{m}\times\hat{\mathbf{r}}) - (\mathbf{m}\times\mathbf{p})] - \frac{\mu_0}{3}(\mathbf{m}\times\mathbf{p})\,\delta^3(\mathbf{r}),\end{aligned} \tag{4.1}$$

(where $\mathbf{r}$ is the vector from one dipole to the other), or as a standard electric dipole in the magnetic field of a standard magnetic dipole:

$$\left.\begin{aligned}&\mathbf{P}_{\mathrm{h}} = -(\mathbf{p}\cdot\boldsymbol{\nabla})\mathbf{A} = -(\mathbf{p}\cdot\boldsymbol{\nabla})\frac{\mu_0}{4\pi}\frac{(\mathbf{m}\times\mathbf{r})}{r^3},\\ \text{or}\quad &P_{\mathrm{h}i} = -\frac{\mu_0}{4\pi}\epsilon_{ijk}p_l m_j\nabla_l\left(\frac{r_k}{r^3}\right) = -\frac{\mu_0}{4\pi}\epsilon_{ijk}p_l m_j\left[\frac{1}{r^3}\left(\delta_{lk} - 3\frac{r_l r_k}{r^2}\right) + \frac{4\pi}{3}\delta_{lk}\,\delta^3(\mathbf{r})\right]\\ &\quad = -\frac{\mu_0}{4\pi}\left\{\frac{[-3(\mathbf{p}\cdot\hat{\mathbf{r}})(\mathbf{m}\times\hat{\mathbf{r}})_i + (\mathbf{m}\times\mathbf{p})_i]}{r^3} + \frac{4\pi}{3}(\mathbf{m}\times\mathbf{p})_i\delta^3(\mathbf{r})\right\},\end{aligned}\right\} \tag{4.2}$$

and we recover equation (4.1).

In the same way, we obtain the hidden momentum in the other three cases. These results are summarized below:

$\mathbf{p},\mathbf{m}$	$\mathbf{P}_{\mathrm{h}} = \frac{\mu_0}{4\pi r^3}[3(\mathbf{p}\cdot\hat{\mathbf{r}})(\mathbf{m}\times\hat{\mathbf{r}}) - (\mathbf{m}\times\mathbf{p})] + \frac{\mu_0}{3}(\mathbf{p}\times\mathbf{m})\,\delta^3(\mathbf{r})$	resides in $\mathbf{m}$
$\mathbf{p},\tilde{\mathbf{m}}$	$\mathbf{P}_{\mathrm{h}} = \mathbf{0}$	nothing moving
$\tilde{\mathbf{p}},\mathbf{m}$	$\mathbf{P}_{\mathrm{h}} = \frac{\mu_0}{4\pi r^3}\{3[(\tilde{\mathbf{p}}\times\mathbf{m})\cdot\hat{\mathbf{r}}]\hat{\mathbf{r}} - (\tilde{\mathbf{p}}\times\mathbf{m})\} - \frac{\mu_0}{3}(\tilde{\mathbf{p}}\times\mathbf{m})\,\delta^3(\mathbf{r})$	resides in $\tilde{\mathbf{p}}$ and $\mathbf{m}$
$\tilde{\mathbf{p}},\tilde{\mathbf{m}}$	$\mathbf{P}_{\mathrm{h}} = -\frac{\mu_0}{4\pi r^3}[3(\tilde{\mathbf{m}}\cdot\hat{\mathbf{r}})(\tilde{\mathbf{p}}\times\hat{\mathbf{r}}) - (\tilde{\mathbf{p}}\times\tilde{\mathbf{m}})] + \frac{\mu_0}{3}(\tilde{\mathbf{p}}\times\tilde{\mathbf{m}})\,\delta^3(\mathbf{r})$	resides in $\tilde{\mathbf{p}}$

5. Spherical shell models

In this paper, I have treated ideal (point) dipoles, whose fields include the subtle delta function terms. It is embarrassingly easy to get these 'contact' contributions wrong, and wise to check one's results using a finite model. What if we picture the dipoles as spherical shells, of radius R, carrying appropriate surface charges (σ) or currents ($\mathbf{K}$)?[6] Letting $v \equiv \frac{4}{3}\pi R^3$ be the volume of the sphere:

[6]If you prefer, think of them as uniformly polarized or uniformly magnetized solid spheres, but this raises diverting questions about the correct formula for the field momentum inside a material medium (Abraham versus Minkowski), which I would like to avoid.

rsta.royalsocietypublishing.org Phil. Trans. R. Soc. A 376: 20180043

1. *Standard electric dipole*, $\mathbf{p}$: $\sigma = (\mathbf{p}\cdot\hat{\mathbf{r}})/v$,

$$V(\mathbf{r}) = \begin{cases} \dfrac{1}{4\pi\epsilon_0}\dfrac{(\mathbf{p}\cdot\hat{\mathbf{r}})}{r^2}, & (r > R), \\ \dfrac{(\mathbf{p}\cdot\mathbf{r})}{3\epsilon_0 v}, & (r < R) \end{cases} \tag{5.1}$$

and therefore

$$\mathbf{E}(\mathbf{r}) = \begin{cases} \dfrac{1}{4\pi\epsilon_0}\dfrac{[3(\mathbf{p}\cdot\hat{\mathbf{r}})\hat{\mathbf{r}} - \mathbf{p}]}{r^3}, & (r > R), \\ -\dfrac{\mathbf{p}}{3\epsilon_0 v}, & (r < R). \end{cases} \tag{5.2}$$

2. *Standard magnetic dipole*, $\mathbf{m}$: $\mathbf{K} = (\mathbf{m}\times\hat{\mathbf{r}})/v$,

$$\mathbf{A}(\mathbf{r}) = \begin{cases} \dfrac{\mu_0}{4\pi}\dfrac{(\mathbf{m}\times\hat{\mathbf{r}})}{r^2}, & (r > R), \\ \dfrac{\mu_0(\mathbf{m}\times\mathbf{r})}{3v}, & (r < R) \end{cases} \tag{5.3}$$

and therefore

$$\mathbf{B}(\mathbf{r}) = \begin{cases} \dfrac{\mu_0}{4\pi}\dfrac{[3(\mathbf{m}\cdot\hat{\mathbf{r}})\hat{\mathbf{r}} - \mathbf{m}]}{r^3}, & (r > R), \\ \dfrac{2\mu_0\mathbf{m}}{3v}, & (r < R). \end{cases} \tag{5.4}$$

3. *Anomalous magnetic dipole*, $\tilde{\mathbf{m}}$: $\tilde{\sigma} = (\tilde{\mathbf{m}}\cdot\hat{\mathbf{r}})/v$,

$$\tilde{V}(\mathbf{r}) = \begin{cases} \dfrac{\mu_0}{4\pi}\dfrac{(\tilde{\mathbf{m}}\cdot\hat{\mathbf{r}})}{r^2}, & (r > R), \\ \dfrac{\mu_0(\tilde{\mathbf{m}}\cdot\mathbf{r})}{3v}, & (r < R) \end{cases} \tag{5.5}$$

and therefore

$$\tilde{\mathbf{B}}(\mathbf{r}) = \begin{cases} \dfrac{\mu_0}{4\pi}\dfrac{[3(\tilde{\mathbf{m}}\cdot\hat{\mathbf{r}})\hat{\mathbf{r}} - \tilde{\mathbf{m}}]}{r^3}, & (r > R), \\ -\dfrac{\mu_0\tilde{\mathbf{m}}}{3v}, & (r < R). \end{cases} \tag{5.6}$$

4. *Anomalous electric dipole*, $\tilde{\mathbf{p}}$: $\tilde{\mathbf{K}} = -c^2[(\tilde{\mathbf{p}}\times\hat{\mathbf{r}})/v]$,

$$\tilde{\mathbf{A}}(\mathbf{r}) = \begin{cases} -\dfrac{1}{4\pi\epsilon_0}\dfrac{(\tilde{\mathbf{p}}\times\hat{\mathbf{r}})}{r^2}, & (r > R), \\ -\dfrac{(\tilde{\mathbf{p}}\times\mathbf{r})}{3\epsilon_0 v}, & (r < R) \end{cases} \tag{5.7}$$

and therefore

$$\tilde{\mathbf{E}}(\mathbf{r}) = \begin{cases} \dfrac{1}{4\pi\epsilon_0}\dfrac{[3(\tilde{\mathbf{p}}\cdot\hat{\mathbf{r}})\hat{\mathbf{r}} - \tilde{\mathbf{p}}]}{r^3}, & (r > R), \\ \dfrac{2\tilde{\mathbf{p}}}{3\epsilon_0 v}, & (r < R). \end{cases} \tag{5.8}$$

As $R \to 0$, $1/v \to \delta^3(\mathbf{r})$, and we recover the ideal dipole fields (equations (1.33)–(1.36)).

Now suppose the sphere is *both* an electric dipole (either kind) *and* a magnetic dipole (either type).[7] Let us first calculate the field momentum for each combination. The external contribution ($r > R$) is the same for all of them; letting $\mathbf{a} \equiv (\mathbf{p} \times \mathbf{m})$:

$$\mathbf{P}^{\text{out}}_{\text{em}} = \epsilon_0 \frac{1}{4\pi\epsilon_0}\frac{\mu_0}{4\pi}\int \frac{[3(\mathbf{p}\cdot\hat{\mathbf{r}})\hat{\mathbf{r}} - \mathbf{p}] \times [3(\mathbf{m}\cdot\hat{\mathbf{r}})\hat{\mathbf{r}} - \mathbf{m}]}{r^6}\,\mathrm{d}\tau$$
$$= \frac{\mu_0}{(4\pi)^2}\int \frac{[3(\mathbf{a}\cdot\hat{\mathbf{r}})\hat{\mathbf{r}} - 2\mathbf{a}]}{r^6}\,\mathrm{d}\tau. \tag{5.9}$$

Setting the z axis along $\mathbf{a}$, so that $\mathbf{a} = a\hat{\mathbf{z}}$, $\mathbf{a}\cdot\hat{\mathbf{r}} = a\cos\theta$ and $\hat{\mathbf{r}} = \sin\theta\cos\phi\hat{\mathbf{x}} + \sin\theta\sin\phi\hat{\mathbf{y}} + \cos\theta\hat{\mathbf{z}}$, the ϕ integral kills the $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ components, leaving

$$\mathbf{P}^{\text{out}}_{\text{em}} = \frac{\mu_0}{(4\pi)^2}(a\hat{\mathbf{z}})(2\pi)\int_R^\infty \frac{1}{r^4}\,\mathrm{d}r \int_0^\pi (3\cos^2\theta - 2)\sin\theta\,\mathrm{d}\theta \tag{5.10}$$
$$= -\frac{\mu_0}{4\pi}\frac{(\mathbf{p}\times\mathbf{m})}{3R^3}. \tag{5.11}$$

The internal contribution ($r < R$) is trivial (since the fields are uniform), but different for the (four) different configurations:

1. <u>$\mathbf{p}$ and $\mathbf{m}$</u>: $\mathbf{P}^{\text{in}}_{\text{em}} = \epsilon_0(-\mathbf{p}/3\epsilon_0 v) \times (2\mu_0\mathbf{m}/3v)v = -2\mu_0(\mathbf{p}\times\mathbf{m})/9v$.
$$\text{So } \mathbf{P}_{\text{em}} = -\frac{\mu_0}{4\pi}\frac{(\mathbf{p}\times\mathbf{m})}{R^3}. \tag{5.12}$$
2. <u>$\mathbf{p}$ and $\tilde{\mathbf{m}}$</u>: $\mathbf{P}^{\text{in}}_{\text{em}} = \epsilon_0(-\mathbf{p}/3\epsilon_0 v) \times (-\mu_0\tilde{\mathbf{m}}/3v)v = (\mu_0/9v)(\mathbf{p}\times\tilde{\mathbf{m}})$.
$$\text{So } \mathbf{P}_{\text{em}} = \mathbf{0}. \tag{5.13}$$
3. <u>$\tilde{\mathbf{p}}$ and $\mathbf{m}$</u>: $\mathbf{P}^{\text{in}}_{\text{em}} = \epsilon_0(2\tilde{\mathbf{p}}/3\epsilon_0 v) \times (2\mu_0\mathbf{m}/3v)v = 4\mu_0(\tilde{\mathbf{p}}\times\mathbf{m})/9v$.
$$\text{So } \mathbf{P}_{\text{em}} = \frac{\mu_0}{4\pi}\frac{(\tilde{\mathbf{p}}\times\mathbf{m})}{R^3}. \tag{5.14}$$
4. <u>$\tilde{\mathbf{p}}$ and $\tilde{\mathbf{m}}$</u>: $\mathbf{P}^{\text{in}}_{\text{em}} = \epsilon_0(2\tilde{\mathbf{p}}/3\epsilon_0 v) \times (-\mu_0\tilde{\mathbf{m}}/3v)v = -(2\mu_0/9v)(\tilde{\mathbf{p}}\times\tilde{\mathbf{m}})$.
$$\text{So } \mathbf{P}_{\text{em}} = -\frac{\mu_0}{4\pi}\frac{(\tilde{\mathbf{p}}\times\tilde{\mathbf{m}})}{R^3}. \tag{5.15}$$

Now let's calculate the hidden momentum in each configuration:

1. <u>$\mathbf{p}$ and $\mathbf{m}$</u>: I would *like* to use the Penfield–Haus formula (equation (3.5)), but that assumes the electric field is uniform over the current region. In this case, the field *is* uniform *inside* the sphere, but right at the surface (where the current is located) $\mathbf{E}$ is discontinuous, and the external field is *not* uniform. You can finesse this problem by a trick: make the magnetic sphere ever-so-slightly *smaller* than the electric sphere; then the electric field really *is* uniform over the current, and we get[8]
$$\mathbf{P}_{\text{h}} = \frac{1}{c^2}(\mathbf{m}\times\mathbf{E}) = \mu_0\epsilon_0\left[\mathbf{m}\times\left(\frac{-\mathbf{p}}{3\epsilon_0 v}\right)\right] = \frac{\mu_0}{4\pi}\frac{(\mathbf{p}\times\mathbf{m})}{R^3}. \tag{5.16}$$
2. <u>$\tilde{\mathbf{p}}$ and $\tilde{\mathbf{m}}$</u>: using the duality transformation and equation (5.19)
$$\mathbf{P}_{\text{h}} = \frac{\mu_0}{4\pi}\frac{(\tilde{\mathbf{p}}\times\tilde{\mathbf{m}})}{R^3}. \tag{5.17}$$
3. <u>$\mathbf{p}$ and $\tilde{\mathbf{m}}$</u>: nothing is moving, so
$$\mathbf{P}_{\text{h}} = \mathbf{0}. \tag{5.18}$$

[7]The 'standard' case ($\mathbf{p}$ and $\mathbf{m}$) was introduced by Romer [11].

[8]If this bothers you, go back to equation (3.11) (or rather, its analogue for surface currents), $\mathbf{P}_{\text{h}} = -(1/c^2)\int V\mathbf{K}\,\mathrm{d}a$; you get the same answer either way.

4. $\tilde{\mathbf{p}}$ and $\mathbf{m}$: In this case, there is a hidden momentum in *both* spheres. Using the trick (making the magnetic sphere slightly smaller than the electric sphere—you can do it the other way, of course, but you get the same answer), the hidden momentum in the electric current is

$$\mathbf{P}_{\mathrm{h}}^{e} = \frac{1}{c^2}(\mathbf{m}\times\tilde{\mathbf{E}}) = \mu_0\epsilon_0\left[\mathbf{m}\times\left(\frac{2\tilde{\mathbf{p}}}{3\epsilon_0 v}\right)\right] = -2\,\frac{\mu_0}{4\pi}\frac{(\tilde{\mathbf{p}}\times\mathbf{m})}{R^3}. \tag{5.19}$$

But the magnetic field in the vicinity of the monopole current is *not* uniform, and we must use equation (3.8) (or rather, its analogue for a monopole current ($\tilde{I}$) in a standard magnetic field $\mathbf{B}$):

$$\mathbf{P} = \tilde{\alpha}\tilde{I}\oint \gamma\,\mathrm{d}\mathbf{l}. \tag{5.20}$$

In this case, equation (3.9) becomes

$$\gamma(\mathbf{r}) = \gamma_0 + \frac{1}{\tilde{\alpha}c^2}\int_{\mathcal{O}}^{\mathbf{r}} \mathbf{B}\cdot\mathrm{d}\mathbf{l}. \tag{5.21}$$

We might as well choose our axes so that $\tilde{\mathbf{p}}$ points in the z direction. I will first calculate the hidden momentum in a single ring of monopole current, $\tilde{I}$, at $z = R\cos\theta$, with radius $R\sin\theta$.

$$\mathbf{B} = \frac{\mu_0}{4\pi}\frac{[3(\mathbf{m}\cdot\hat{\mathbf{r}})\hat{\mathbf{r}}-\mathbf{m}]}{R^3},\quad \mathrm{d}\mathbf{l} = R\sin\theta\,\mathrm{d}\phi\hat{\boldsymbol{\phi}}, \tag{5.22}$$

and (setting the reference point directly above the x axis),

$$\int_{\mathcal{O}}^{\mathbf{r}} \mathbf{B}\cdot\mathrm{d}\mathbf{l} = \frac{\mu_0}{4\pi}\frac{R\sin\theta}{R^3}\int_0^{\phi}[3(\mathbf{m}\cdot\hat{\mathbf{r}})\hat{\mathbf{r}}-\mathbf{m}]\cdot\hat{\boldsymbol{\phi}}\,\mathrm{d}\phi = -\frac{\mu_0\sin\theta}{4\pi R^2}\int_0^{\phi}(\mathbf{m}\cdot\hat{\boldsymbol{\phi}})\,\mathrm{d}\phi.$$

Now

$$\hat{\boldsymbol{\phi}} = -\sin\phi\hat{\mathbf{x}} + \cos\phi\hat{\mathbf{y}},\quad \text{so } \mathbf{m}\cdot\hat{\boldsymbol{\phi}} = -m_x\sin\phi + m_y\cos\phi,$$

and therefore

$$\begin{aligned}\int_{\mathcal{O}}^{\mathbf{r}} \mathbf{B}\cdot\mathrm{d}\mathbf{l} &= -\frac{\mu_0\sin\theta}{4\pi R^2}\int_0^{\phi}(-m_x\sin\phi + m_y\cos\phi)\,\mathrm{d}\phi\\ &= -\frac{\mu_0\sin\theta}{4\pi R^2}\left[m_x(\cos\phi - 1) + m_y\sin\phi\right].\end{aligned} \tag{5.23}$$

The hidden momentum in this ring is

$$\begin{aligned}\mathbf{P}_{\mathrm{h}}^{\mathrm{ring}} &= \frac{\tilde{I}}{c^2}\left(-\frac{\mu_0\sin\theta}{4\pi R^2}\right)\oint\left[m_x(\cos\phi-1)+m_y\sin\phi\right]R\sin\theta\,\mathrm{d}\phi\hat{\boldsymbol{\phi}}\\ &= -\frac{\tilde{I}\mu_0\sin^2\theta}{4\pi c^2 R}\int_0^{2\pi}\left[m_x(\cos\phi-1)+m_y\sin\phi\right](-\sin\phi\hat{\mathbf{x}}+\cos\phi\hat{\mathbf{y}})\,\mathrm{d}\phi\\ &= -\frac{\tilde{I}\mu_0\sin^2\theta}{4\pi c^2 R}\left(-\pi m_y\hat{\mathbf{x}} + \pi m_x\hat{\mathbf{y}}\right).\end{aligned} \tag{5.24}$$

Now we integrate over all the rings that cover the monopole current sphere, using $\tilde{I}\to -|\tilde{\mathbf{K}}|R\,\mathrm{d}\theta$ and $\tilde{\mathbf{K}} = -c^2(\tilde{\mathbf{p}}\times\hat{\mathbf{r}})/v$

$$\begin{aligned}\mathbf{P}_{\mathrm{h}}^{\mathrm{m}} &= \frac{\mu_0}{4c^2}(-m_y\hat{\mathbf{x}}+m_x\hat{\mathbf{y}})\int_0^{\pi}\sin^2\theta\,\frac{c^2|\tilde{\mathbf{p}}\times\hat{\mathbf{r}}|}{v}\,\mathrm{d}\theta\\ &= \frac{\mu_0}{4v}\left(-m_y\hat{\mathbf{x}}+m_x\hat{\mathbf{y}}\right)\tilde{p}\int_0^{\pi}\sin^3\theta\,\mathrm{d}\theta = \frac{\mu_0}{4v}\left(-m_y\hat{\mathbf{x}}+m_x\hat{\mathbf{y}}\right)\tilde{p}\left(\frac{4}{3}\right)\\ &= \frac{\mu_0}{3v}(\tilde{\mathbf{p}}\times\mathbf{m}) = \frac{\mu_0}{4\pi}\frac{(\tilde{\mathbf{p}}\times\mathbf{m})}{R^3}.\end{aligned} \tag{5.25}$$

Finally, combining equations (5.19) and (5.25),

$$\mathbf{P}_{\mathrm{h}} = \mathbf{P}_{\mathrm{h}}^{e} + \mathbf{P}_{\mathrm{h}}^{m} = -\frac{\mu_0}{4\pi}\frac{(\tilde{\mathbf{p}}\times\mathbf{m})}{R^3}. \tag{5.26}$$

These results confirm the contact terms in the table in §4.[9] As always, the hidden momentum is equal and opposite to the field momentum. There is nothing *surprising* in any of this, but it is gratifying to see it work out in explicit detail.

Data accessibility. This article has no additional data.

Competing interests. I declare I have no competing interests.

Funding. No funding has been received for this article.

Acknowledgements. I thank Vladimir Hnizdo, Kirk McDonald and Pablo Saldanha for illuminating correspondence.

References

1. Griffiths DJ. 1992 Dipoles at rest. *Am. J. Phys.* **60**, 979–987. (doi:10.1119/1.17001)
2. Frahm CP. 1983 Some novel delta-function identities. *Am. J. Phys.* **51**, 826–829. (doi:10.1119/1.13127)
3. Coleman S, Van Vleck JH. 1968 Origin of 'hidden momentum forces' on magnets. *Phys. Rev.* **171**, 1370–1375. (doi:10.1103/PhysRev.171.1370)
4. Calkin MG. 1971 Linear momentum of the source of a static electromagnetic field. *Am. J. Phys.* **39**, 513–516. (doi:10.1119/1.1986204)
5. Shockley W, James RP. 1967 'Try simplest cases' discovery of 'hidden momentum' forces on 'magnetic currents'. *Phys. Rev. Lett.* **18**, 876–879. (doi:10.1103/PhysRevLett.18.876)
6. Costa de Beauregard O. 1967 A new law in electrodynamics. *Phys. Lett.* **A24**, 177–178. (doi:10.1016/0375-9601(67)90752-9)
7. McDonald KT. 2018 See http://physics.princeton.edu/~mcdonald/examples/hiddendef.pdf.
8. Penfield Jr P, Haus HA. 1967 *Electrodynamics of moving media*, pp. 214–216. Cambridge, MA: M.I.T. Press.
9. Vaidman L. 1990 Torque and force on a magnetic dipole. *Am. J. Phys.* **58**, 978–983. (doi:10.1119/1.16260)
10. Hnizdo V. 1997 Hidden momentum of a relativistic fluid carrying current in an external electric field. *Am. J. Phys.* **65**, 92–94. (doi:10.1119/1.18500)
11. Romer RH. 1995 Question #26. Electromagnetic field momentum. *Am. J. Phys.* **63**, 777–779. (doi:10.1119/1.18075)

[9]Since the two spheres coincide we are *only* checking the contact term. On the other hand, if we *separate* the spheres (by a distance greater than the sum of their radii) there is really nothing to check, since the fields are precisely those of an ideal dipole.

PHILOSOPHICAL TRANSACTIONS A

rsta.royalsocietypublishing.org

Research

Check for updates

The Noether current in Maxwell's equations and radiation under symmetry breaking

Dhiraj Sinha[1] and Gehan Amaratunga[2]

[1]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

[2]Division of Electrical Engineering, Department of Engineering, University of Cambridge, 9 J.J. Thomson Avenue Cambridge CB3 0FA, UK

DS, 0000-0003-3589-1862

Cite this article: Sinha D, Amaratunga G. 2018 The Noether current in Maxwell's equations and radiation under symmetry breaking. *Phil. Trans. R. Soc. A* **376**: 20170452. http://dx.doi.org/10.1098/rsta.2017.0452

Accepted: 6 September 2018

One contribution of 13 to a theme issue 'Celebrating 125 years of Oliver Heaviside's 'Electromagnetic Theory''.

Subject Areas:
electromagnetism, gauge theory

Keywords:
radiation, Noether current, symmetry breaking

Author for correspondence:
Dhiraj Sinha
e-mail: dhiraj@mit.edu

Symmetries in physical systems are defined in terms of conserved Noether Currents of the associated Lagrangian. In electrodynamic systems, global symmetry is defined through conservation of charges, which is reflected in gauge symmetry; however, loss of charges from a radiating system can be interpreted as localized loss of the Noether current which implies that electrodynamic symmetry has been locally broken. Thus, we propose that global symmetries and localized symmetry breaking are interwoven into the framework of Maxwell's equations which appear as globally conserved and locally non-conserved charges in an electrodynamic system and define the geometric topology of the electromagnetic field. We apply the ideas in the context of explaining radiation from dielectric materials with low physical dimensions. We also briefly look at the nature of reversibility in electromagnetic wave generation which was initially proposed by Planck, but opposed by Einstein and in recent years by Zoh.

This article is part of the theme issue 'Celebrating 125 years of Oliver Heaviside's 'Electromagnetic Theory''.

THE ROYAL SOCIETY PUBLISHING

1. Introduction

Spontaneous symmetry breaking where the ground-state breaks the system's symmetry plays the main role in generation of cooper pairs in superconductivity [1–3], Goldstone Bosons [4,5], acquisition of mass through Higgs mechanism in the Standard Model [6–9] and generation of phonons in crystals [10]. In a related phenomenon called explicit symmetry breaking, the defining equations of motion represented by the Lagrangian break the symmetry [10]. Initial work in the field of symmetry in physical systems was done by Curie who proposed that in any physical phenomenon, the element of symmetry of its cause is also present in the resulting effect [11]. He further argued that the asymmetry associated with an effect is also present in the causes. Symmetries and the associated conserved physical quantities defined by the Noether Theorem play a central role in physics [12]. For example, temporal symmetry of the Lagrangian of a system is associated with energy conservation [13]. Similarly, symmetry of the electromagnetic field is expressed through gauge invariance [14].

In the current article, we propose that transformations of force fields from one form to the other are linked with loss of invariance of at least one physical quantity resulting in a localized symmetry breaking of the associated field in the localized region of space and time which generates a Noether current which is globally conserved. In that context, we argue that loss of invariance in the temporal domain of a force field causes loss of invariance in the spatial domain of a corresponding physical quantity and vice-versa. For example, potential energy of a body is transformed into change in momentum when the symmetry of the potential energy surface is locally broken by some external agent. These observations converge towards our novel discovery of the fact that in Maxwell's equations on radiation, the globally conserved Noether current is generated as a consequence of localized symmetry breaking of an electromagnetic field. At a global level, it plays a critical role in defining the global symmetries expressed through gauge invariance and generation of electromagnetic radiation.

According to Maxwell, radiation is generated due to acceleration of electrons [15,16]. This aspect was further highlighted and put on a simpler mathematical framework by Heaviside [17,18]. An important contribution of Heaviside was reformulation of the initial Maxwell's equations in quaternion form into a vector form comprising of four equations in terms of electric and magnetic fields and charges (static and dynamic) [19].

There are some physical instances where electron acceleration does not result in radiation, in apparent violation of the Maxwell's theoretical prediction. For example, an infinitely long wire placed at an infinitesimal distance, from a conductive ground plane, excited by a time varying voltage source, is a classical example where electrons are accelerated but the overall radiation field is cancelled due to opposing currents induced in the ground plane. A closely spaced transmission line, excited by a time varying voltage source, is another such example, where there is no measurable level of radiation despite electron acceleration as the charges flowing in the wires are in opposite directions which cancels the radiation field.

An important consequence of Maxwell's work is that charges at uniform velocities do not emit electromagnetic radiation. However, there are physical conditions, where electromagnetic radiation is emitted from charges moving at constant velocities. For example, a charged particle moving through a dielectric material at a uniform velocity leads to polarization and radiation. Similarly, in Cherenkov radiation, electrons moving at superluminal velocities in a medium generate electromagnetic radiation [20–22].

Maxwell and Heaviside's work on radiation was put on a stronger physical foundation by J.J. Thomson, whose perspective on radiation was related to the topological changes in the nature of electric lines of field [23]. He argued that acceleration changes the geometric pattern of field lines. The prior theoretical work on radiation were validated through the experimental work by Hertz [24,25], who proved the existence of electromagnetic waves. The other pioneers in the field of early work on wireless communication were Bose [26], Tesla [27] and Marconi [28], who worked on the empirical aspects of generation, transmission and reception of electromagnetic waves. One of the key findings by Marconi was the fact that radiation efficiency significantly increases when

rsta.royalsocietypublishing.org Phil. Trans. R. Soc. A 376: 20170452

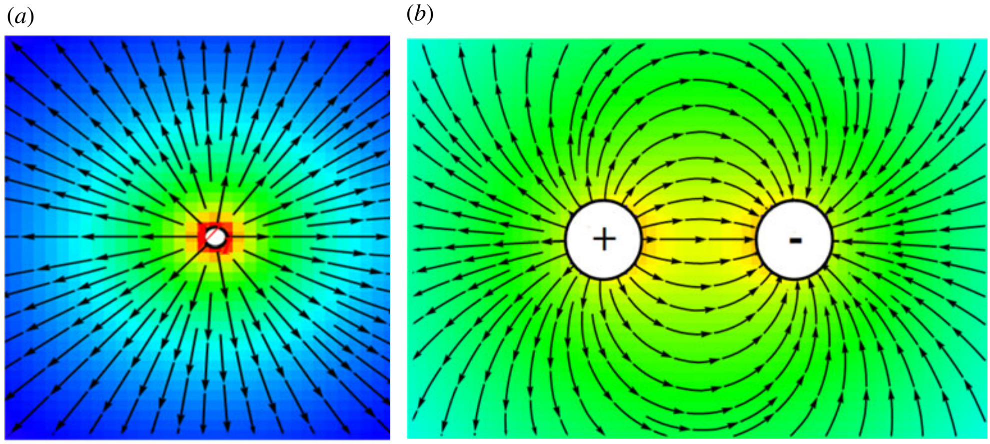


Figure 1. Symmetry in static charges. (*a*) The electric lines of force are oriented radially from the centre in a symmetric manner which illustrates rotational symmetry of the electric potential. (*b*) The spatial symmetry of the electric field lines appear to be broken in an electric dipole, however, the mathematical symmetry defined by Gauss's Law is maintained which is reflected in temporal symmetry of the associated Lagrangian. (Online version in colour.)

a metallic pole, being used as a transmitting antenna, is held vertically over a horizontal plane [29]. The current work aims to unify the diverse aspects of radiation explored along theoretical and empirical dimensions through a common thread based on broken symmetries of force fields.

2. Symmetry in electrostatics

The Lagrangian of an electrodynamic system is represented by electric and magnetic fields [30], hence, the geometric patterns of lines of electromagnetic fields can also be used to study the associated symmetries in the localized region of the given electrodynamic system. For example, in electrostatics, for a static point charge, symmetry is observed in electric field lines which are directed away from the centre of the charge and possess rotational symmetry (figure 1*a*) [31,32]. The Lagrangian density of a static point charge Q within a sphere of radius R is expressed by

$$L_d = \frac{1}{2}\epsilon E^2 = \frac{Q^2}{32\pi^2 \epsilon R^4}, \tag{2.1}$$

where ϵ is the permittivity of the medium, and E is the electric field. Here, $\mathrm{d}L_d/\mathrm{d}t$ and $\mathrm{d}L_d/\mathrm{d}\theta$ are equal to zero where, t is time and θ is angular displacement from the centre of the charge, which indicates symmetry of the Lagrangain. The symmetry of a dynamic system is associated with a conserved quantity, which is charge in this case. In electrostatics, the net value of charge does not change with time. The symmetry is also expressed through Gauss's Law

$$\boldsymbol{\nabla} \cdot \boldsymbol{D} = \rho, \tag{2.2}$$

where $\boldsymbol{D}$ is electric flux density and ρ is the total charge density enclosed by the given Gaussian surface. It could be argued that for a system of more than one charge, the geometric symmetry of the electric lines of fields is broken as shown in figure 1*b*, however, the symmetry defined by equation (2.2) is maintained around a Gaussian surface S, enclosing the charges, though $\boldsymbol{D}$ and ρ vary over the surface such that $\int \boldsymbol{D} \cdot d\boldsymbol{S} = Q$, where Q is the total charge enclosed by the surface. In addition to this, the temporal variation of the Lagrangian is zero, which also indicates symmetry.

Symmetries in electromagnetism are integral parts of the current scientific literature, however, an interpretation of the physics of the system in terms of geometric symmetry of the electric lines of field and conservation of the associated charges at local and global level can establish the condition of generation of radiation or its absence from an electrodynamic system. This offers a novel perspective which supplements the theoretical prediction of Maxwell, according to which charges at rest and at uniform velocities do not radiate and charges under acceleration result in radiation. For example, in the context of charging of a parallel plate capacitor, with infinitely large plates with an infinitesimal separation, the conservation of charges defined by equation (2.2) remains valid at a local as well as a global level besides the symmetry of the electric field, which confirms the absence of radiation. A charge moving at a uniform velocity in a dielectric material cannot be expressed in terms of the symmetry of the electric field as defined by equation (2.2); besides this, its Lagrangian density does not remain invariant along temporal and spatial dimensions, so, we can associate a radiation field with it. Thus, our symmetry based interpretation offers a novel dimension to our current understanding of electromagnetic phenomenon while complementing the existing understanding based on the foundations of Maxwell, Heaviside and Thomson.

3. Localized symmetry breaking in electromagnetism

The relationship between symmetries and conservation was discovered by Emmy Noether [12], which is widely referred to as the Noether theorem. It defines a relationship between global symmetries and globally conserved currents in a physical system. However, the Noether theorem is largely silent about localized symmetry breaking and the role of non-conserved currents, within the localized region of space and time, which play an important role in defining global symmetries of a physical system. Its connection to Maxwell's equations and role in radiation can lead to a deeper and refined understanding of electrodynamic systems.

A mass kept at the highest point of a gravitational potential surface as shown in figure 2*a* has a finite value of potential energy. At the highest point, the potential energy surface has symmetry, i.e. the Lagrangian $\mathcal{L}$ has spatial symmetry, $\mathrm{d}\mathcal{L}/\mathrm{d}r = 0$, along a spatial dimension r [30]. The Noether current in this context is momentum, which is generated only when $\mathrm{d}\mathcal{L}/\mathrm{d}r \neq 0$. In the context of the ball shown in figure 2*b*, there is a net force due to non-zero value of the potential gradient. In other words, the localized symmetry breaking of the Lagrangian generates a change in momentum of force expressed through the Euler Lagrange equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}\dot{r}}\right) = \frac{\mathrm{d}\mathcal{L}}{\mathrm{d}r}, \tag{3.1}$$

where $\dot{r}$ is rate of change of position of the particle along the dimension r. The change in momentum with respect to time gives a sense of symmetry breaking of momentum $\boldsymbol{p}$ along temporal dimension expressed in terms of non-zero value of $\mathrm{d}\boldsymbol{p}/\mathrm{d}t$. The discussion converges towards the point that localized symmetry breaking of one physical quantity along spatial dimension breaks the symmetry of a related physical quantity along the temporal dimension. We extend the concept to electromagnetic force fields, where it leads to the condition of generation of radiation fields under broken symmetries.

An important contribution in the case of electric lines of force of a momentarily accelerated electron was done by J.J. Thomson, who argued that the electric lines of field of an accelerated charge undergo a distortion as illustrated in figure 3*a* [33]. Although Thomson's picture was highly intuitive and led to a better understanding of the nature of geometric lines of force associated with charges under radiation, for a long time there was no effort to develop a mathematical model corresponding to distortions in the electric lines of field from the framework of broken symmetries. Interpretation of the distortions in electric lines of force associated with an accelerated charge as a consequence of breaking of the translational symmetry of the electric field in the radial direction within a localized region of space and time due to rotation of the electric field, which results in generation of electromagnetic waves as shown in figure 3*b*, was published recently [34,35]. Thus, although a periodic acceleration of the dipole configuration of charged particles results in electromagnetic radiation which is associated with generation of rotating electric field in space is well known, its correlation to loss of temporal and spatial invariance of the associated Lagrangian has remained unnoticed, for long. In fact, it is an instance of explicit

rsta.royalsocietypublishing.org Phil. Trans. R. Soc. A 376: 20170452

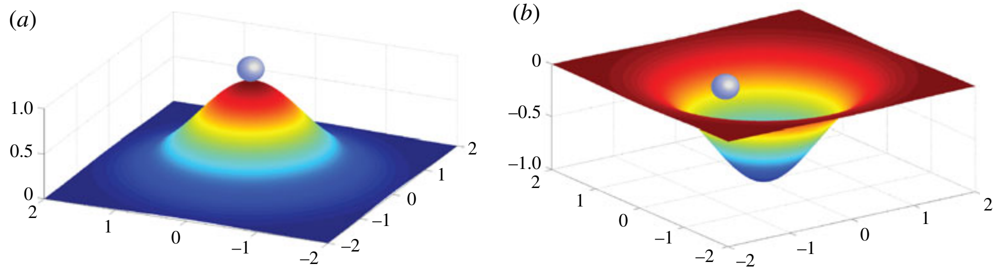


Figure 2. Explicit symmetry breaking in potential surfaces. (*a*) A body (light blue ball) placed at the symmetrical peak of the potential energy surface is not subjected to a net force due to symmetry. Its symmetry can be explicitly broken within a localized region of space, by displacing it towards the edge of the potential surface when it is under some finite force field which can result in change in momentum. (*b*) The black ball in the figure in a potential well is under a net force due to explicit symmetry breaking of the potential energy surface within the localized region of the body. Despite global conservation of momentum, its momentum is not conserved within the localized region of space with broken symmetry of the potential energy surface. (Online version in colour.)

symmetry breaking, where the defining equations of motion represented by the Lagrangian break the symmetry [36].

The fact that translational symmetry of the electric field is explicitly broken, resulting in the generation of time varying magnetic field and radiation, is expressed by Maxwell's equation

$$\nabla \times \boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\mathrm{d}t}, \tag{3.2}$$

where $\boldsymbol{E}$ is electric field intensity and $\boldsymbol{B}$ is magnetic flux density.

The above arguments also apply in the case of radiation from a dipole antenna connected to a parallel two wire transmission line excited by a time varying voltage source (figure 3*c*). If the parallel section of the transmission line is assumed to be infinitely long with an infinitesimally small separation, the electric field has translational symmetry along the transmission line which is broken along the flared end where rotational electric field is generated, resulting in radiation. Here, it is assumed that the currents not aligned to the axis of the transmission line at the supply end are shielded. This, in essence, assumes the complete absence of common mode current components in the parallel section of the wire. These aspects have also been highlighted by Ianconescu & Vulfin, who show that in the case of an infinite transmission line, there is no far field radiation, but there is a non-radiative field which decays as inverse square of distance from the axis of the wire as in the case of an electrostatic field [37], rather than inverse of distance as associated with radiation) [35]. However, a two wire transmission line radiates when one takes into account the currents which do not lie in the axis of the transmission line. For example, current flowing at the supply end and low end of the transmission line leads to a loop antenna like effect [37].

The concept of explicit symmetry breaking has also been applied in experiments on generation of tunable broadband radiation from a superconducting ring where a pulse of laser light breaks the symmetry of cooper pairs, resulting in electromagnetic radiation [38]. Under explicit symmetry breaking, the momentum current associated with the test charge loses its conserved value within a localized region of space and time as the corresponding Noether current is the driver of electromagnetic radiation.

An important goal of the current article is to underscore the fact that localized symmetry breaking is expressed within the framework of Maxwell's equation on radiation, and it denotes a close relationship between dependence of the force fields along temporal and spatial dimensions. The rotation of the electric field expressed by curl of the vector $\boldsymbol{E}$ indicates that the symmetry of the electric field along a set of dimensions in the physical space has been broken. Thus,

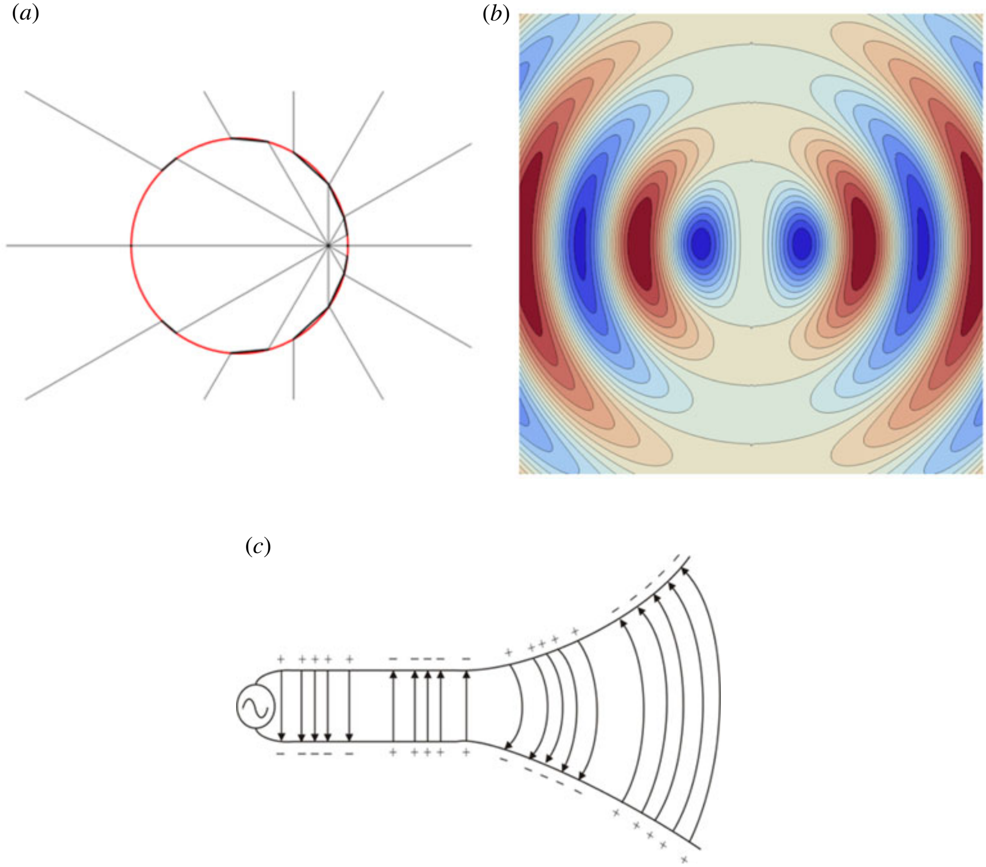


Figure 3. Explicit symmetry breaking of electric lines of force. (*a*) The changes in rotational symmetry in the electric field distribution of a static positive charge under acceleration as discovered by Thomson [33], can be analysed from their perspective of explicit symmetry breaking of the electric field. (*b*) When the dipole configuration of charges is subjected to time varying acceleration, the symmetry of the electric field is explicitly broken periodically resulting in electromagnetic radiation which is associated with rotation of the electric field resulting in completely broken translational symmetries. (*c*) The electric field lines are straight in an infinitely long, parallel transmission line, placed close to each other at an infinitesimally small distance, indicating translational symmetry which is broken along the flared end in an explicit manner due to geometric asymmetry of the electrodynamic system. (Online version in colour.)

the equation essentially implies that broken symmetry of magnetic flux lines in the temporal dimension results in symmetry breaking along a set of spatial dimensions. In a curved space around a mass, the rotation of the electric lines of force would show an additional curvature. For a uniform plane wave propagating along the z-axis, with the electric and magnetic field lines aligned along x- and y-axes respectively, equation (3.2) takes the form [31]

$$\frac{\partial E_x}{\partial z} = \frac{\partial B_y}{\partial t}. \tag{3.3}$$

Here, the finite value of rate of change of magnetic flux density can be interpreted as broken symmetry of the magnetic flux lines along the temporal dimension. Here $\mathrm{d}E_x/\mathrm{d}B_y$ is also a measure of the degree of curvature of space around a mass. If the space is flat, the fields are orthogonal to each other in the context of a plane polarized electromagnetic wave. However, when the space is curved, the orthogonality is lost. In metamaterials, by artificially engineering a given material, the fields are changed in order to change the existing symmetry [39].

Equation (3.2) can be transformed into a line integral using Stoke's theorem resulting in the expression $\int \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{r} = -\partial \int \boldsymbol{B}.\mathrm{d}\boldsymbol{S}/\partial t = \mathrm{d}\phi/\mathrm{d}t$, where ϕ is the flux density. The finite value of $\mathrm{d}\phi/\mathrm{d}t$ as the magnetic flux lines vary can be interpreted as localized symmetry breaking of the magnetic flux lines along the temporal dimension. Similarly, the electric field intensity $\boldsymbol{E}$ due to a charge q in free space can also be interpreted as localized symmetry breaking of the electric potential V along the spatial dimension.

The time varying electric flux density under the rotation of magnetic field intensity is expressed by

$$\nabla \times \boldsymbol{H} = \frac{\partial \boldsymbol{D}}{\partial t}, \tag{3.4}$$

where $\boldsymbol{H}$ is magnetic field intensity and $\boldsymbol{D}$ is electric flux density. Equation (3.4) takes the form of

$$-\frac{\partial H_y}{\partial z} = \frac{\partial D_x}{\partial t}, \tag{3.5}$$

for the uniform plane wave, and it indicates its symmetry breaking of electric flux density along the dimension of time which, in turn, drives symmetry breaking of magnetic field intensity along the spatial dimension. The temporal and spatial dependence of force fields during their transformation and generation of locally non-conserved, but globally conserved currents, under symmetry breaking, along with their role in radiation offers a new degree of freedom in understanding electrodynamic systems.

4. Charge in an electromagnetic field

According to the symmetry principle of electromagnetism, an electron in a simple harmonic motion generates a plane polarized electromagnetic wave; similarly, a plane polarized electromagnetic wave generates a simple harmonic motion in an electron [32]. Here, we argue that this aspect of symmetry in electromagnetism partially breaks down as the rotational component of electric field of electromagnetic wave excites the angular momentum of the electron which generates circularly polarized radiation under acceleration. As force $\boldsymbol{F}$ on a charged particle with a charge q in an electric field $\boldsymbol{E}$ is given by $q\boldsymbol{E}$, we can multiply both sides of equation (3.2) by q to get the following expression:

$$\nabla \times \boldsymbol{F} = -\frac{\partial q\boldsymbol{B}}{\partial t}. \tag{4.1}$$

The left-hand side of the equation (4.1) can be transformed into a line integral using Stoke's theorem $\int \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r} = -q\partial \int \boldsymbol{B} \cdot \mathrm{d}\boldsymbol{S}/\partial t = U$, where U is the electric potential energy. Its derivative along the dimension r leads to, $\boldsymbol{F} = \partial U/\partial r$ or $\partial V/\partial r$, where V is the electric potential. The force exerted by the electric field $\boldsymbol{E}$ on another charge q generates a momentum $\boldsymbol{p}$ which can be expressed as $\boldsymbol{F} = q\boldsymbol{E} = \mathrm{d}\boldsymbol{P}/\mathrm{d}t$. The expression defined by equation (4.1) expresses the fact that a charged particle in an electromagnetic field shows rotational motion and the radiation generated by it is not linearly but circularly polarized.

The equations discussed earlier can also be expressed using the electromagnetic field tensor [32], $\boldsymbol{F}_{\mu\nu} = \partial \boldsymbol{A}_\nu/\partial x_\mu - \partial \boldsymbol{A}_\mu/\partial x_\nu$ where $\boldsymbol{A}$ is magnetic vector potential, x is a dimension in space, μ and ν are subscripts denoting spatial components. The symmetry breaking of the electromagnetic field tensor can be expressed as $\partial \boldsymbol{F}_{\mu\nu}/\partial x = \boldsymbol{J}_\mu$, where $\boldsymbol{J}_\mu$ is the four current vector. The finite value of spatial variation of $\partial \boldsymbol{F}_{\mu\nu}$ indicates its symmetry breaking along the spatial dimension which results in corresponding breaking in symmetry along the temporal domain as expressed by the finite value of four current vector $\boldsymbol{J}_\mu$.

The explicit symmetry breaking of the electric and magnetic field during electromagnetic wave propagation is also reflected in the non-conserved values of its Lagrangian within localized

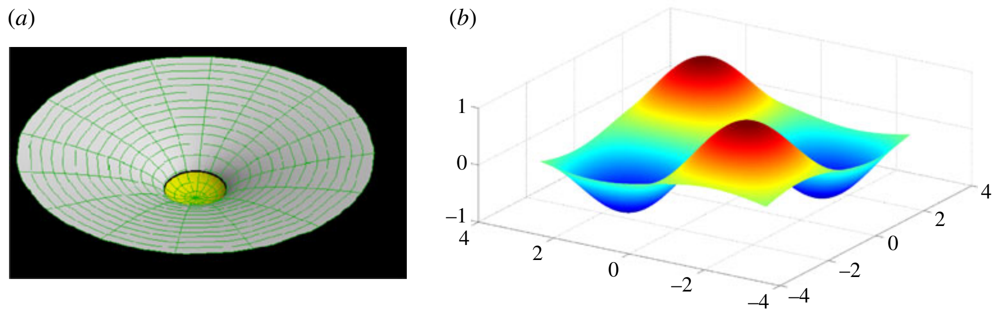


Figure 4. Explicit symmetry breaking of space–time surface under the presence of mass. (*a*) Mass breaks the symmetry of space–time surface resulting in change in curvature. (*b*) Elevation and depression on the space–time surface appears around mass. (Online version in colour.)

regions of space and time. The Euler–Lagrange equation for the electromagnetic Lagrangian density can be stated as follows:

$$\partial_\beta \left[\frac{\partial \mathcal{L}}{\partial(\partial_\beta A_\alpha)}\right] - \frac{\partial \mathcal{L}}{\partial A_\alpha} = 0. \tag{4.2}$$

Here, the Lagrangian $\mathcal{L}$ has subscripts, α and β, denoting components in space and $\partial\mathcal{L}/\partial A_\alpha = -J^\alpha$, which indicates that symmetry breaking in the spatial domain results in generation of Noether current in the localized region of space and time [32]. As the electromagnetic wave moves into a certain localized region of space and time, the Lagrangian of the electromagnetic field loses its temporal symmetry within that region. It acquires a positive value as the wave moves in and slowly starts getting reduced and eventually reduces to zero as the wave moves out of the region.

Symmetry is never absolute and is represented in terms of the invariance of a system with a well-defined transformation group called the symmetry group [40]. The phenomenon of Symmetry breaking expresses relationships of transformation between a group and its sub-groups. It indicates that the initial symmetry is broken to one of its sub-groups [40]. In the context of static charges, the translational or rotational symmetry of the electric lines of field can also be interpreted as a broken symmetry of a higher order. For example, a stack of ferroelectric material having randomly oriented dipole moments has rotational symmetry as the net dipole moment is invariant with respect to rotation. When it is inserted within the plates of a charged parallel plate capacitor where the electric field lines have translational symmetry, the net polarization induced in it is $\boldsymbol{P}_e = \chi_e \epsilon \boldsymbol{E}$ where χ_e is electric susceptibility and ϵ is permittivity [41]. This can also be interpreted as explicit breaking of rotational symmetry of the polarization field resulting in generation of translational symmetry. Similarly, induced magnetism in ferromagnetic material under a strong magnetic field can be considered as explicit breaking of rotational symmetry and induction of translational symmetry [42].

The general symmetry breaking of a force field in either space or time which results in symmetry breaking of a corresponding quantity in the other dimension is unified in the general theory of relativity where both the symmetries are simultaneously broken. The presence of mass explicitly breaks the symmetry of space–time fabric (figure 4*a*). This is expressed in the basic equation [43,44]

$$\boldsymbol{G}_{\mu\nu} = \boldsymbol{R}_{\mu\nu} - \frac{1}{2}\boldsymbol{R}\boldsymbol{g}_{\mu\nu} = \frac{8\pi}{c^4} G\boldsymbol{T}_{\mu\nu}, \tag{4.3}$$

where $\boldsymbol{G}_{\mu\nu}$ indicates the curvature of space–time surface which is an indication of the total potential energy gradient of the system, $\boldsymbol{T}_{\mu\nu}$ indicates the energy momentum tensor and can be assumed to quantify the total kinetic energy components. Although space and time occur implicitly in both the sides of the equation, space has a dominant role in the potential energy component and time has the key role in the kinetic energy component. When there is no mass present, $\boldsymbol{R}_{\mu\nu} = 0$ represents zero curvature of space indicating translational symmetry and a zero value of energy-momentum tensor. Figure 4*b* further illustrates the explicit symmetry breaking of the space–time curve in the form of elevations and depressions around a given mass.

5. Broken symmetry in electrodynamic systems, gauge symmetry and radiation

Symmetries in electromagnetism are defined in terms of gauge invariance [14]. Here, we attempt to link the temporal and spatial symmetry breaking of electric and magnetic field through the magnetic vector potential and establish its correlation to gauge invariance. We argue that such interactions involve localized symmetry breaking of the magnetic vector potential which is essentially linked to gauge invariance and global symmetries. The magnetic vector potential $\boldsymbol{A}$ is expressed by [15]

$$\boldsymbol{E} = -j\frac{\nabla\boldsymbol{\nabla}\cdot\boldsymbol{A}}{\omega\mu\epsilon} - \frac{\partial\boldsymbol{A}}{\partial t}, \tag{5.1}$$

where μ is permeability, ω is frequency and j indicates the imaginary component in the complex plane. It can be interpreted as comprising two symmetry breaking terms–the broken symmetry of the vector magnetic potential along spatial and temporal dimension. In equation (5.1), the first term denotes near-field effects and the second term denotes the radiation term.

An important goal of this discussion is to establish that the localized symmetry breaking of the electromagnetic field is closely associated with gauge symmetry at a global level. The magnetic flux density $\boldsymbol{B}$ is related to magnetic vector potential by the relation [31,32]

$$\boldsymbol{B} = \boldsymbol{\nabla} \times \boldsymbol{A}. \tag{5.2}$$

The expression of equation (5.2) can be interpreted from a novel perspective by considering that the translational symmetry breaking of magnetic vector potential along spatial dimensions results in its rotation, leading to the generation of magnetic flux density. The rotational symmetry breaking of the magnetic vector potential is associated with its invariance under the transformation $\boldsymbol{A} = \boldsymbol{A} + \boldsymbol{\nabla}\psi$, where ψ is a scalar field. This leads to the invariance of $\boldsymbol{B}$, expressed as, $\boldsymbol{B} = \boldsymbol{\nabla} \times (\boldsymbol{A} + \boldsymbol{\nabla}\psi) = \nabla \times \boldsymbol{A}$. In physical terms, as the magnetic flux density is generated as a consequence of translational symmetry breaking of the magnetic vector potential, it is not influenced by any additional changes in its symmetry along a specific spatial dimension.

In electrodynamics, an increase in the magnetic vector potential $\boldsymbol{A}$, by $\boldsymbol{\nabla}\psi$ is also associated with a corresponding decrease in the corresponding electric potential V, expressed as $V - \partial\psi/\partial t$, which is the basic condition of gauge invariance [31,32]. For a uniform plane wave propagating in free space, the electric field is expressed by $\boldsymbol{E} = -\partial\boldsymbol{A}/\partial\boldsymbol{t}$. It establishes the relationship between temporal symmetry breaking of the magnetic vector potential and generation of an electric field, which remains invariant to any temporal changes in the value of electric potential V, expressed by $V - \partial\psi/\partial t$. Thus, the localized symmetry breaking of the magnetic vector potential, $\boldsymbol{A}$, along the spatial and temporal dimensions is an essential condition of the invariance of $\boldsymbol{E}$ and $\boldsymbol{B}$ which implies the invariance of the associated Lagrangian defining an electromagnetic wave and generation of global symmetries.

The ideas outlined above provide new insights in understanding the phenomenon of radiation from a diverse set of systems including piezoelectric and dielectric materials, which are insulators and do not have free charge carriers. Our current understanding of dielectric antennas is governed by the empirical and theoretical work by Long, McAllister and Shen who used a cylindrical structure of length L and radius R mounted on a ground plane whose radiation pattern was used to validate the theoretical assumption that the antenna has a boundary comprising a perfect magnetic conductor [45]. The electromagnetic field patterns were expressed using Bessel functions while arguing that the dielectric material behaved like a magnetic dipole antenna [45]. The magnetic wall model fails in offering a theoretical framework towards calculation of input impedance of dielectric antennas which play a key role in defining the radiation efficiency [46]. Besides this, it assumes the presence of an infinite ground plane which does not satisfy the operation of such devices at an empirical level [47]. Under empirical conditions, dielectric antennas have been found to have radiation patterns similar to electric dipole, magnetic dipole and quadrupole antennas, depending on shapes and mounting patterns and feeding geometries [48].

Earlier, it was pointed out by Sinha & Amaratunga that the assumptions by Long *et al.* violate the foundational aspects of radiating systems as a time varying excitation fed to a dielectric material generates oscillations of bound charges which is essentially oscillation of electrical dipoles which are inversely proportional to the cube of distance [35,38]. The electric field $\boldsymbol{E}$, associated with a Hertzian dipole in spherical coordinates along the dimensions r, θ and ϕ, is expressed by [49]

$$E_r = \eta \frac{IL\cos\theta}{2\pi r^2}\left[1+\frac{1}{jkr}\right]\mathrm{e}^{-jkr} \tag{5.3}$$

and

$$E_\theta = j\eta \frac{kIL\sin\theta}{2\pi r^2}\left[1+\frac{1}{jkr}-\frac{1}{(kr)^3}\right]\mathrm{e}^{-jkr}. \tag{5.4}$$

Here, I is the current in the antenna, L is its length, η is the intrinsic impedance of space, k is the wavenumber associated with the wave and $j=\sqrt{-1}$. The other component of electric field is $E_\phi = 0$. Oscillating dipoles generate an electric field which corresponds to the $1/r^3$ term in the antenna field equation [49]. Considering this, the power generated by a dielectric antenna should hold a vanishingly small value in the far field region, which defies empirical observations.

A dielectric antenna has a capacitive reactance, which acts as a reservoir of electrons and is mounted on a finite ground plane, whose inductive reactance provides a framework for motion of charges between the capacitive and inductive elements. Hence, at a fundamental level, dielectric antennas work like simple capacitor–inductor antennas where electrons swing between the two elements resulting in radiation [34,35]. Thus, the physical operation of dielectric antennas can be understood by considering the role of non-conserved Noether currents in the system generated as a consequence of electrodynamic asymmetry which facilitates transformation of current into electromagnetic wave. The magnetic vector potential which determines the radiation pattern can be found by solving the inhomogeneous Helmholtz equation, $\boldsymbol{\nabla}^2\boldsymbol{A} + \omega^2\mu\epsilon\boldsymbol{A} = \boldsymbol{J}$, where the current density $\boldsymbol{J}$ is the non-conserved Noether current, driving radiation.

The vector magnetic potential $\boldsymbol{A}$, at a point r and time t in a medium of permeability μ, is expressed as [32]

$$\boldsymbol{A}(r,t) = \frac{\mu}{4\pi}\int \frac{\boldsymbol{J}(r',t')\mathrm{d}^3r'}{|r-r'|}, \tag{5.5}$$

where $\boldsymbol{J}(r',t')$ is current at time t' and a position r' flowing along a length dr'. The time t' is called retarded time defined by $t' = t - |r-r'|/c$, where c is the speed of light in the given medium. For a wire of length L oriented along the z-direction, carrying a current I, the magnetic vector potential can be expressed as, $\boldsymbol{A}(x,y,z) = \boldsymbol{a}_z\mu IL\mathrm{e}^{-jkr}/(4\pi r)$ [32]. An important point which we wish to make in the current work, which has not been addressed in the literature on electromagnetism until now, is the fact that the factor IL plays a key role in determining the radiation field, which has a much higher value in a metallic antenna as the electrons flow along the length L determined by the antenna's spatial dimension. In a dielectric antenna, in the absence of a ground plane, under electric excitation, the capacitive current can be made high, which is equivalent to the current in a metallic antenna in terms of flow of charge per unit time, however, in such an antenna, due to the fact that the charges are bound and the net displacement around the mean position, L, has atomic dimensions, which is miniscule in comparison to a metallic antenna of a similar size, the charges can move freely along the spatial dimensions of the antenna. The result is that the overall value of IL is much lower in a dielectric antenna resulting in low radiation efficiency. This point further strengthens the arguments on the role of ground plane in dielectric antennas which facilitate radiation by providing a longer path of current flow.

Our analysis of radiating systems in terms of non-conserved Noether current and broken symmetries can help in analysing empirical results on radiating structures with low degrees of freedom. For example, in a recent work published on the problem of antenna miniaturization in wireless communication by Nan *et al.* [50], they reported that bulk acoustic waves in a magneto electric (ME) nano plate resonator (NPR) structure induce oscillation of the magnetization of ferromagnetic thin film leading to electromagnetic radiation which can realize antenna size miniaturization by 1–2 orders of magnitude. According to the Chu–Harrington (CH) limit which determines the antenna size for a given bandwidth, the antenna size reduction reduces the bandwidth and its efficiency [51,52]. Hence, a technology on antenna miniaturization must demonstrate efficiency measurements over a given bandwidth under standardized test conditions. Surpassing the CH limit over a given bandwidth is the only indicator of antenna miniaturization and Nan *et al*. in the given article [50] mention explicitly that their antennas do not surpass the CH limit, so the question of antenna miniaturization as claimed by the authors in the first part of the paper is refuted by themselves in the latter part. An analysis in terms of the Noether current associated with the system can shed novel insights on radiation from such a structure.

The theoretical foundation of the operation of ME antennas operating on electroacoustic effect addressed by Nan *et al.* [50] remains unclear as they point out that far-field radiation is a consequence of oscillating magnetic dipole moments. In a magnetic loop antenna of radius a carrying a current I, the magnetic flux density is defined as [49]

$$B_r = j\mu \frac{ka^2 I \cos\theta}{2r^2}\left(1 + \frac{1}{jkr}\right) \mathrm{e}^{-jkr} \tag{5.6}$$

and

$$B_\theta = \mu \frac{(ka)^2 I \sin\theta}{4r}\left(1 + \frac{1}{jkr} - \frac{1}{(kr)^2}\right) \mathrm{e}^{-jkr}. \tag{5.7}$$

The other component of magnetic field, in the ϕ direction, is $H_\phi = 0$. In equations (5.6) and (5.7), the $1/r^3$ term is the magnetic flux density generated by the oscillating magnetic dipoles which falls off inversely as cube root of distance and can induce current in test antennas during mid-field measurements giving the appearance of a lossy antenna. It can account for low radiation efficiency of 0.28% measured by the ME antenna reported by Nan *et al.* The radiation efficiency matches the value of radiation efficiency of 2%, in terms of orders of magnitude in piezoelectric films under symmetric excitation explored by Sinha & Amaratunga which effectively feeds excitation into the oscillating dipoles [35], creating a system where the corresponding Noether current is fed back to the ground plane, instead of being radiated out in space.

In the context of piezoelectric antennas, asymmetric excitation where the piezoelectric structure is used as a capacitive element on an inductive ground plane, the antenna efficiency increases as an electrodynamic asymmetry is created which provides a framework for loss of the corresponding Noether current, but this leads to miniaturization by a factor of 3–5 [34,35] and not 100 as Nan *et al*. claim. Thus, the results on antenna miniaturization as claimed by Nan *et al.* [50] need further validation as the antenna efficiency is in accordance with Harrington Chu limit. However, there is some near-field coupling between the RF field and ME NPR system which induces time varying oscillation of the magnetic moment, leading to generation of magnetic field in the near-field region which induces a finite current in the test antenna leading to a lossy antenna like effect.

Thus, an analysis of radiating structures by considering the flow of Noether current within the system can help in a more refined understanding of the mechanism of radiation from systems of low spatial dimensions. It becomes more important in electrodynamic systems which are much below the operational wavelength and parasitic elements start behaving as effective radiators.

6. The nature of irreversibility

In this section, we establish a relationship between the role of Noether current and broken symmetries in physical systems in defining a mathematical framework towards understanding irreversibility. One of the mysteries of physics is that the laws of physics have time reversal symmetry, however, irreversibility is intertwined with the fabric of real physical systems. Since

rsta.royalsocietypublishing.org *Phil. Trans. R. Soc. A* **376**: 20170452

rsta.royalsocietypublishing.org Phil. Trans. R. Soc. A 376: 20170452

the days of Boltzmann, scientist have looked at irreversibility from a statistical perspective, which considers irreversibility as an evolution of the number of microstates leading to an increase in entropy as the initial states become less probable. Thus, a closed system in a highly ordered state evolves towards a state of disorder as its entropy increases. Brian Greene in his book, The fabric of the cosmos: space, time and the structure of reality, writes, 'No one has ever discovered any fundamental law which might be called the Law of the Spilled Milk or the Law of the Splattered Egg' [53]. Lebowitz writes on Scholaropedia, 'It is only secondary laws, which describe the behavior of macroscopic objects containing many, many atoms, such as the second law of thermodynamics, which explicitly contain this time asymmetry. The obvious question then is; how does one go from a time symmetric description of the dynamics of atoms to a time asymmetric description of the evolution of macroscopic systems made up of atoms' [54]. According to the second law of thermodynamics, entropy of an isolated system tends to increase leading to an increase in its phase space volume [55], but Liouville's theorem states that the phase space volume is a conserved physical quantity [56]. This leads to an unresolved puzzle on the reconciliation of the entropy increase with the evolution of microstates and the onset of irreversibility in thermodynamics with time reversal symmetry and reversibility integrated with Newtonian mechanics which is referred to as the Loschmidt paradox [57]. The explanation of the Loschmidt paradox is statistical rather than physical and is usually defined in terms of the coarse graining of the phase space according to which the overall volume of phase space is conserved, but it evolves under Markov criterion into coarse-grained cells of phase space [58]. Currently, a mathematical framework which could define irreversibility in a macroscopic system does not exist. Here, we will see that a geometric analysis of trajectories of the Noether current in a system can lead to new insights in defining macroscopic irreversibility.

The first-order relationship between space and time in transformation of force fields is a characteristic of a reversible process. For example, the mechanical potential change $\mathrm{d}U_m$ along a given spatial dimension, r, leads to a gain of momentum $\mathrm{d}p$ over a time $\mathrm{d}t$, which is expressed as, $\mathrm{d}U_m/\mathrm{d}r = \mathrm{d}p/\mathrm{d}t$. It leads to a momentum gain, as a particle is dropped from a given height. In a thermodynamic system, at equilibrium, the net momentum change is zero, i.e. $\sum_i^n \boldsymbol{p}_i = 0$ over a given closed volume V. We cannot define a conserved current over a given volume which also means that spatial trajectories of particles in the phase space diagram are not uniquely defined.

At microscopic dimensions, we can define a surface, S, through which a definite momentum traverses at a certain instant of time, so it can be considered to be reversible during those specific instants of time. However, if the system is in equilibrium, the sum of momentum components vanish over a period of time $\mathrm{d}t$ and we cannot define a definite conserved current. Thus, in a process like diffusion or heat transfer, we cannot define reversibility. For example, diffusion is given as [55]

$$\nabla^2 C = K_\mathrm{D}\frac{\partial C}{\partial t}, \tag{6.1}$$

where C is the concentration of particles and K_D is the diffusion constant. Here the first term $\nabla^2 C$ leads to the second term, $K_\mathrm{D}\partial C/\partial t$, but the second term does not lead to the first term. A microscopic model of each of the particles follows the Euler–Lagrange equation and is reversible. However, irreversibility is only a macroscopic model of a set of multiple microscopic processes which are fundamentally reversible. Similarly, heat transfer at a given temperature T, is expressed by [55]

$$\nabla^2 T = K_\mathrm{T}\frac{\partial T}{\partial t}, \tag{6.2}$$

where K_T is thermal conductivity. Such systems do not have global conserved currents, neither do they have locally non-conserved physical quantities associated with transformation of force fields. Hence, the relationship between changes in space and time corresponding to the changes in a set of physical quantities, which can be considered to be an independent marker of reversibility of a process, do not exist in such processes.

The points mentioned above are of interest with regard to a historical debate on the nature of reversibility of generation of a photon. While Planck argued that emission of a photon is a reversible process [59], Einstein pointed out that it is an irreversible process, as the reversible process of a spontaneous absorption of a photon is not observed [60]. The debate was further expanded by Zeh, who supported Einstein's viewpoint [61]. However, an analysis in terms of the relationship between transformation of force fields and their spatial and temporal dependance implies that generation of an electromagnetic wave is a reversible process as the space and time are correlated by interdependent fields and the Noether current is lost locally but is conserved globally. In an irreversible process, the Noether current appears to be lost as a consequence of the statistical interactions of the given macroscopic system. For example, a hot furnace generates photons and slowly cools down. The process of heat transfer can be modelled according to equation (6.2) and it can appear to be an irreversible process as the globally conserved current loses the character of a vector field. In other words, the statistical nature of the system annihilates the vector fields at a macroscopic level, where the corresponding Noether current is annihilated at a macroscopic level.

7. Conclusion

We have shown that explicit symmetry breaking of a force field along the spatial domain is associated with explicit symmetry breaking along the spatial dimension within a localized region of space and time and vice versa. This is particularly reflected in electromagnetism where broken symmetries of electric and magnetic field are associated with temporal and spatial changes of corresponding force fields. In electrodynamic systems, radiation is generated as a consequence of loss of the corresponding Noether current, in the localized region of space and time. However, at a global level, the symmetries are maintained through conservation of charged and gauge invariance. The logic has been used to define the fundamental nature of a reversible process while arguing that in an irreversible process the globally conserved Noether current appears to be lost due to statistical addition of the corresponding vector field.

Data accessibility. This article has no additional data.
Authors' contributions. All authors have read and approved the manuscript.
Competing interests. We declare we have no competing interests.
Funding. DS acknowledges support from SUTD-MIT Fellowship.

References

1. Nambu Y. 1960 Quasi-particles and gauge invariance in the theory of superconductivity. *Phys. Rev.* **117**, 648–663. (doi:10.1103/PhysRev.117.648)
2. Nambu Y, Giovanni JL. 1961 Dynamical model of elementary particles based on an analogy with superconductivity. I. *Phys. Rev.* **122**, 345–358. (doi:10.1103/PhysRev.122.345)
3. Anderson P. 1963 Plasmons, gauge invariance, and mass. *Phys. Rev.* **130**, 439–442. (doi:10.1103/PhysRev.130.439)
4. Goldstone J. 1955–1965 Field theories with Superconductor solutions. *Nuovo Cimento* **19**, 154–164. (doi:10.1007/bf02812722)
5. Salam J, Weinberg A, Goldstone S. 1962 Broken symmetries. *Phys. Rev.* **127**, 331–334. (doi:10.1103/PhysRev.127.331)
6. Englert F, Brout R. 1964 Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.* **13**, 321–323. (doi:10.1103/PhysRevLett.13.321)
7. Higgs PW. 1964 Broken symmetries, massless particles and gauge fields. *Phys. Lett.* **12**, 132–133. (doi:10.1016/0031-9163(64)91136-9)
8. Higgs PW. 1966 Spontaneous symmetry breakdown without massless bosons. *Phys. Rev.* **145**, 1156–1163. (doi:10.1103/PhysRev.145.1156)
9. Susskind L. 1979 Dynamics of spontaneous symmetry breaking in the Weinberg-Salam theory. *Phys. Rev. D* **20**, 2619–2625. (doi:10.1103/physrevd.20.2619)
10. Wagner H. 1966 Long-wavelength excitations and the Goldstone theorem in many-particle systems with 'broken symmetries'. *Z. Phys.* **195**, 273–299. (doi:10.1007/BF01325630)

11. Curie P. 1894 Sur la symétrie dans les phènomènes physiques. Symétrie d'un champ électrique et d'un champ magnétique. *J. Phys.* **3**, 393–415. (doi:10.1051/jphystap:018940030039300)
12. Noether EA. 1918 Nachr. D. König. Gesellsch. D. Wiss. Zu Göttingen. *Math-Phys. Klasse* **325**, 235–257.
13. Trautman A. 1967 Noether equations and conservation laws. *Comm. Math. Phys.* **6**, 248–261. (doi:10.1007/BF01646018)
14. Quigg C. 2013 *Gauge theories of the strong, weak, and electromagnetic interactions*. Princeton, NJ: Princeton University Press.
15. Maxwell JC. 1865 VIII. A dynamical theory of the electromagnetic field. *Phil. Trans. R. Soc. Lond.* **155**, 459–512. (doi:10.1098/rstl.1865.0008)
16. Maxwell JC. 1881 *A treatise on electricity and magnetism*, vol. 1. Oxford, UK: Clarendon Press.
17. Heaviside O. 1889 The electromagnetic effects of a moving charge—III. *Electrician* **23**, 458–459.
18. Heaviside O. 1902 The waste of energy from a moving electron. *Nature* **67**, 6–8. (doi:10.1038/067006d0)
19. Nahin PJ. 2002 *Oliver Heaviside: the life, work, and times of an electrical genius of the Victorian age*. Baltimore, US: JHU Press.
20. Cherenkov PA. 1934 Visible emission of clean liquids by action of γ radiation. *Dokl. Akad. Nauk SSSR* **2**, 451–454.
21. Vavilov SI. 1934 On Possible Reasons for the Blue γ Radiation in Fluids. *Dokl. Akad. Nauk.* **2**(8), 457–459.
22. Cerenkov PA. 1937 Visible radiation produced by electrons moving in a medium with velocities exceeding that of light. *Phys. Rev.* **52**, 378–379. (doi:10.1103/PhysRev.52.378)
23. Thomson JJ. 1904 *Elements of the mathematical theory of electricity and magnetism*, p. 539. Cambridge, UK: Cambridge University Press.
24. Hertz HR. 1887 Ueber sehr schnelle electrische Schwingungen. *Annalen der Physik* **267**, 421–448. (doi:10.1002/andp.18872670707)
25. Hertz H, Kelvin WTB. 1893 *Electric waves*, p. 177. London, UK: Macmillan.
26. Geddes P. 1920 *An Indian pioneer of science: the life and work of Sir Jagadis C. Bose*. London, UK: Longmans, Green, and Company.
27. Tesla N. 1901 Apparatus for the utilization of radiant energy. U.S. Patent 685,957, issued 5 November 1901.
28. Marconi G. 1900 Improvements in apparatus for wireless telegraphy. British Patent No. 7777 (1900).
29. Marconi G. 1906 On methods whereby the radiation of electric waves may be mainly confined to certain directions, and whereby the receptivity of a receiver may be restricted to electric waves emanating from certain directions. *Proc. R. Soc. Lond. A* **77**, 413–421. (doi:10.1098/rspa.1906.0036)
30. Goldstein H. 1957 *Classical mechanics*. Boston, MA: Pearson.
31. Hayt WH. 1981 *Engineering electromagnetics*. New York, NY: McGraw-Hill.
32. Jackson JD. 1988 *Classical electrodynamics*. Amsterdam, The Netherlands: Wiley.
33. Thomson JJ. 1910 On a theory of the structure of the electric field and its application to Röntgen radiation and to light. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **19**, 301–313. (doi:10.1080/14786440208636803)
34. Sinha D, Amaratunga GAJ. 2015 *Explicit symmetry breaking in electrodynamic systems and electromagnetic radiation*. San Rafael: Morgan & Claypool Publishers.
35. Sinha D, Amaratunga GAJ. 2015 Electromagnetic radiation under explicit symmetry breaking. *Phys. Rev. Lett.* **114**, 147701. (doi:10.1103/PhysRevLett.114.147701)
36. Brading K, Castellani E. 2003 *Symmetries in physics: new reflections*. Cambridge, UK: Cambridge University Press.
37. Ianconescu R, Vulfin V. 2017 Free space TEM transmission lines radiation losses model. (http://arxiv.org/abs/1701.04878v2)
38. Bulmer J, Bullard T, Dolasinski B, Murphy J, Sparkes M, Pangovski K, Powers P, Haugan T. 2015 Tunable broadband radiation generated via ultrafast laser illumination of an inductively charged Superconducting ring. *Sci. Rep.* **5**, 18151. (doi:10.1038/srep18151)
39. Werner DH, Kwon DH. 2015 *Transformation electromagnetics and metamaterials*. Amsterdam, The Netherlands: Springer.

40. Hamermesh M. 1989 *Group theory and its application to physical problems*. N. Chelmsfod, CA: Courier Corporation.
41. Resta R. 1992 Theory of the electric polarization in crystals. *Ferroelectrics* **136**, 51–55. (doi:10.1080/00150199208016065)
42. Birss RR. 1964 *Symmetry and magnetism*. Amsterdam, The Netherlands: North-Holland.
43. Carroll SM. 2004 *Spacetime and geometry: an introduction to general relativity*. Boston, MA: Addison Wesley.
44. Hartle JB. 2003 *Gravity: an introduction to Einstein's general relativity*. Boston, MA: Addison Wesley.
45. Long SA, McAllister MW, Shen LC. 1983 The resonant cylindrical dielectric cavity antenna. *IEEE Trans. Ant. Prop.* **31**, 406–412. (doi:10.1109/TAP.1983.1143080)
46. Petosa A. 2007 *Dielectric resonator antenna handbook*, p. 79. London, UK: Artech House.
47. Luk KM, Leung KW. 2003 *Dielectric resonator antennas*, p. 366. London, UK: Research Studies Press.
48. Mongia RK, Bhartia P. 1994 Dielectric resonator antennas—a review and general design relations for resonant frequency and bandwidth. *Int. J. Micr. Mill. Wave Comp.-Aid. Engg.* **4**, 230–247. (doi:10.1002/mmce.4570040304)
49. Balanis CA. 1982 *Antenna theory, analysis and design*. New York, NY: Harper & Row.
50. Nan T *et al.* 2017 Acoustically actuated ultra-compact NEMS magnetoelectric antennas. *Nat. Comm.* **8**, 57. (doi:10.1038/s41467-017-00103-8)
51. Harrington RF. 1960 Effect of antenna size on gain, bandwidth, and efficiency. *J. Res. Nat. Bur. Stand.* **64**, 1–12. (doi:10.6028/jres.064d.003)
52. McLean JS. 1996 A re-examination of the fundamental limits on the radiation Q of electrically small antennas. *IEEE Trans. Antenn. Prop.* **44**, 672. (doi:10.1109/8.496253)
53. Greene B. 2007 *The fabric of the cosmos: space, time, and the texture of reality*. New York, NY: Alfred Knopf.
54. Lebowitz JL. 2008 Time's arrow and Boltzmann's entropy. *Scholarpedia* **3**, 3448. (doi:10.4249/scholarpedia.3448)
55. Reif F. 1965 *Fundamentals of statistical and thermal physics*. New York, NY: McGraw Hill.
56. Gibbs JW. 1884 On the fundamental formula of statistical mechanics, with applications to astronomy and thermodynamics. In *Proc. of the American Association for the Advancement of Science*, vol. 33, p.57–58. US, s.n.: American Association for the Advancement of Science.
57. Loschmidt J. 1876 Über den Zustand des Wärmegleichgewichtes eines Systems von Körporen mit Rücksicht auf die Schwerkraft, 1. Teil. *Sitzungsber. Kais. Akad. Wiss. Wien, Math. Naturwiss. Classe* **73**, 128–142.
58. Vladislav C, Daniel P, Sheehan DP. 2005 *Challenges to the second law of thermodynamics*. Dordrecht, The Netherlands: Springer.
59. Planck M. 1991 *The theory of heat radiation*. New York, NY: Dover Publications.
60. Einstein A. 1905 Generation and conversion of light with regard to a heuristic point of view. *Ann. Phys.* **17**, 132–148.
61. Zeh HD. 2010 *The physical direction of time*. Berlin, Germany: Springer.

PHILOSOPHICAL TRANSACTIONS A

rsta.royalsocietypublishing.org

Research

Check for updates

Cite this article: Thöle F, Spaldin NA. 2018 Magnetoelectric multipoles in metals. *Phil. Trans. R. Soc. A* **376**: 20170450. http://dx.doi.org/10.1098/rsta.2017.0450

Accepted: 11 June 2018

One contribution of 13 to a theme issue 'Celebrating 125 years of Oliver Heaviside's 'Electromagnetic Theory''.

Subject Areas:
solid state physics

Keywords:
magnetoelectric effect, density functional theory, non-centrosymmetric materials

Author for correspondence:
Nicola Spaldin
e-mail: nicola.spaldin@mat.ethz.ch

Magnetoelectric multipoles in metals

Florian Thöle and Nicola A. Spaldin

Materials Theory, ETH Zürich, Wolfgang-Pauli-Strasse 27, Zürich 8093, Switzerland

FT, 0000-0001-6006-9876

In this paper, we demonstrate computationally the existence of magnetoelectric multipoles, arising from the second-order term in the multipole expansion of a magnetization density in a magnetic field, in non-centrosymmetric magnetic metals. While magnetoelectric multipoles have long been discussed in the context of the magnetoelectric effect in non-centrosymmetric magnetic *insulators*, they have not previously been identified in metallic systems, in which the mobile carriers screen any electrical polarization. Using first-principles density functional calculations, we explore three specific systems: first, a conventional centrosymmetric magnetic metal, Fe, in which we break inversion symmetry by introducing a surface, which both generates magnetoelectric monopoles and allows a perpendicular magnetoelectric response. Next, the hypothetical cation-ordered perovskite, $SrCaRu_2O_6$, in which we study the interplay between the magnitude of the polar symmetry breaking and that of the magnetic dipoles and multipoles, finding that both scale proportionally to the structural distortion. Finally, we identify a hidden *antiferromultipolar* order in the non-centrosymmetric, antiferromagnetic metal $Ca_3Ru_2O_7$, and show that, while its competing magnetic phases have similar magnetic dipolar structures, their magnetoelectric multipolar structures are distinctly different, reflecting the strong differences in transport properties.

This article is part of the theme issue 'Celebrating 125 years of Oliver Heaviside's 'Electromagnetic Theory''.

THE ROYAL SOCIETY PUBLISHING

1. Introduction

The interaction energy, H_{int} of a magnetization density, $\boldsymbol{\mu}(\mathbf{r})$ with a magnetic field $\mathbf{H}(\mathbf{r})$ is given in general by the integral of their vector product over all space,

$$H_{\text{int}} = -\int \boldsymbol{\mu}(\mathbf{r}) \cdot \mathbf{H}(\mathbf{r})\, \mathrm{d}^3\mathbf{r}. \tag{1.1}$$

The response of many magnetic materials, however, is well described by the approximate interaction

$$H_{\text{int}} = -\mathbf{m} \cdot \mathbf{H}(0)\,, \tag{1.2}$$

where

$$\mathbf{m} = \int \boldsymbol{\mu}(\mathbf{r})\, \mathrm{d}^3\mathbf{r} \tag{1.3}$$

is the magnetization and $\mathbf{H}(0)$ is a *uniform* magnetic field. This description captures, for example, the usual Zeeman effect in which the magnetism in a ferro- or ferri-magnetic tends to align parallel to the field, as well as the well-known susceptibility of antiferromagnets, in which an applied field induces a net magnetization from the compensating magnetic sublattices.

In a particular class of magnetic materials—those which are insulating and lack a centre of inversion symmetry—this level of treatment is now known to miss important physics, even in the case when the applied field is uniform. The simultaneous breaking of space-inversion and time-reversal symmetry in such materials allows them to exhibit the linear magnetoelectric effect, in which an electric field induces a magnetization with magnetoelectric susceptibility α and vice versa [1]. This phenomenon is not readily captured by a description of the magnetization at the dipole level, but is revealed transparently in analyses of the next-highest-order multipoles in a multipole expansion of equation (1.1), since these depend on the product of $\mathbf{r}$ and $\boldsymbol{\mu}(\mathbf{r})$ and so have the appropriate symmetry [2]. Specifically, in the second order of the multipole expansion,

$$H^{\text{ME}}_{\text{int}} = -\int r_i \mu_j(\mathbf{r}) \partial_i H_j(0)\, \mathrm{d}^3\mathbf{r}, \tag{1.4}$$

the $\int r_i \mu_j(\mathbf{r})\, \mathrm{d}^3\mathbf{r}$ component can be decomposed into a sum of three terms,

$$a = \frac{1}{3}\int \mathbf{r} \cdot \boldsymbol{\mu}(\mathbf{r})\, \mathrm{d}^3\mathbf{r}, \tag{1.5}$$

$$\mathbf{t} = \frac{1}{2}\int \mathbf{r} \times \boldsymbol{\mu}(\mathbf{r})\, \mathrm{d}^3\mathbf{r} \tag{1.6}$$

and

$$q = \frac{1}{2}\int \left[r_i\mu_j + r_j\mu_i - \frac{2}{3}\delta_{ij}\mathbf{r} \cdot \boldsymbol{\mu}(\mathbf{r}) \right] \mathrm{d}^3\mathbf{r}. \tag{1.7}$$

These are referred to as the magnetoelectric monopole, toroidal moment and quadrupole, respectively, and correspond to the diagonal, antisymmetric and symmetric and traceless components of the magnetoelectric tensor [2–4]. In addition to their connection to the magnetoelectric effect, the magnetoelectric multipoles have proved useful in identifying hidden *antimagnetoelectric* ordering [2], as well as providing an unambiguous route to defining the size of the local magnetic moment in certain magnetoelectric antiferromagnets [5].

Since metallic materials cannot sustain an electric polarization due to screening of the electric field by free charge carriers, the conventional linear magnetoelectric effect can only manifest in insulating materials. Magnetoelectric multipoles, on the other hand, should still be non-zero by symmetry in non-centrosymmetric magnetic metals. While higher-order multipoles in metals have recently been considered in the context of the anomalous Hall [6,7] and magnetopiezoelectric effects [8], second-order magnetoelectric multipoles have never, to our knowledge, been observed or discussed in the context of metallic systems. The purpose of this paper, therefore, is twofold. First, to compute the properties of some representative magnetic metals with broken inversion symmetry, in order to establish whether magnetoelectric multipoles exist and to determine their

rsta.royalsocietypublishing.org Phil. Trans. R. Soc. A 376: 20170450

magnitudes. And second, to discuss possible properties that might manifest as a result of such a hidden multipolar order.

We proceed by computing the structural and electronic ground states of three model magnetic metals using density functional theory (DFT), then extract the atomic-site magnetoelectric multipoles around each ion by transforming the atomic-site density matrix into its irreducible spherical tensor moments [2]. We begin by studying a conventional magnetic metal (iron, Fe) which is centrosymmetric in its bulk form, and we break the inversion symmetry by creating a surface. Provided that the vacuum is sufficiently insulating, the system then becomes insulating perpendicular to the surface, and so can exhibit a linear magnetoelectric effect in the surface normal direction [9,10]. To study the influence of the insulator, we also calculate the behaviour when the vacuum is replaced by a representative dielectric, MgO. In this example, we expect magnetoelectric multipoles to occur, and to be substantial only in the vicinity of the interface.

Next, we investigate the hypothetical non-centrosymmetric magnetic metal, cation-ordered $SrCaRu_2O_6$, which has been shown theoretically to have strong coupling between the spins and the polar distortions of the lattice [11–13]. This system allows us to explore the relationship between the magnitude of the symmetry breaking and the magnitudes of the resulting multipoles by manually modifying the amplitude of the polar distortion.

Finally, we study an established non-centrosymmetric magnetic metal, $Ca_3Ru_2O_7$, in which we anticipate hidden magnetoelectric multipoles in the bulk material because of its combined magnetic order and polar crystallographic structure. Several different magnetic dipole orders are known to exist as a function of temperature [14–18], allowing us to search for a relationship between dipolar and multipolar magnetic arrangements. Additionally, because of its strong spin-charge coupling, $Ca_3Ru_2O_7$ shows highly anisotropic magnetoresistance [19] as well as two-dimensional conductivity at low temperatures; we explore whether these properties can be related to the behaviour of the magnetoelectric multipoles.

The remainder of this paper is organized as follows: in §2, we describe the computational methods and approximations used. In §3, we describe in turn our results for Fe surfaces, for $SrCaRu_2O_6$ and for $Ca_3Ru_2O_7$. We conclude by discussing the relevance of magnetoelectric multipoles for the properties and description of non-centrosymmetric magnetic metals, as well as giving suggestions for further work.

2. Methods

We perform DFT calculations within the local density approximation (LDA) augmented where appropriate by a Hubbard U correction. Two plane-wave basis codes are employed: for calculations of magnetoelectric multipoles, we use the VASP software package [20,21] with projector-augmented wave (PAW) potentials [22], while for the calculation of field responses, we use the Quantum Espresso package [23] with ultrasoft pseudopotentials [24].

We model the Fe and Fe/MgO slabs as periodic Fe/vacuum and Fe/MgO/vacuum superlattices, and use an energy cut-off of 540 eV and a $8 \times 8 \times 1$ Monkhorst-Pack k-point grid [25]. For Mg and O, the 3*s* orbitals and 2*s*2*p* orbitals, respectively, are treated as valence states in the PAW potentials and ultrasoft pseudopotentials. For Fe, the 3*p*4*s*3*d* orbitals are treated as valence states in the PAW potentials, and the 4*s*3*d* orbitals are treated as valence states in the ultrasoft pseudopotentials. Electric fields are treated by adding a sawtooth potential with a compensating dipole layer in the vacuum region [26,27]. Magnetic fields are introduced as a Zeeman term in the potential; note that this does not capture contributions from the orbital magnetic response [28].

For $SrCaRu_2O_6$ and $Ca_3Ru_2O_7$, we use energy cut-offs of 550 eV and 500 eV, and k-point grids of $7 \times 7 \times 5$ and $10 \times 10 \times 2$, respectively. We treat the following orbitals as valence states in the PAW potentials: 3*p*4*s* for Ca, 4*s*4*p*5*s* for Sr, 4*p*5*s*4*d* for Ru and 2*s*2*p* for O. For $Ca_3Ru_2O_7$, we add a Hubbard U correction [29], $U = 2$ eV, on the Ru sites, which is the value used in [18]. We emphasize that we do not expect DFT to give an accurate description of the detailed electronic structure of these correlated ruthenates, and our emphasis is to reproduce the gross features of $Ca_3Ru_2O_7$ as a model compound for studying magnetoelectric multipoles.

The atomic-site magnetoelectric multipoles are calculated through the decomposition of the density matrix into irreducible spherical tensor moments, as described in [2]. The dielectric susceptibility was calculated following Giustino & Pasquarello [30]. All atomically smoothed quantities were calculated by convolution with a trapezoidal kernel whose width was chosen to minimize fluctuations in the bulk-like regions of the slabs [30]. Crystal structure visualizations were produced with VESTA [31].

3. Results and discussion

(a) Fe surfaces and interfaces

We begin with an analysis of a conventional, centrosymmetric ferromagnetic metal, Fe, in which we break the inversion symmetry by introducing a surface. Bulk Fe occurs in the bcc structure at low temperatures, with a net magnetization from the ferromagnetic ordering of the local moments on the Fe atoms. The magnetocrystalline anisotropy orients the magnetic moments along a ⟨001⟩ direction, which lowers the space group symmetry to $I4/mm'm'$; the resulting site symmetry of Fe, which is on Wyckoff position $2a$, is $4/mm'm'$. The presence of inversion symmetry at the atomic site forbids the presence of magnetoelectric multipoles. Our first-principles calculations in this setting confirm this fact.

On introduction of a surface, or an interface to a dielectric material, the inversion symmetry is broken and the system becomes insulating in the normal direction. These two properties combine to allow the linear magnetoelectric effect, which has been demonstrated for slabs of other ferromagnetic metals both computationally [9,10] and experimentally [32,33].

Our first model system is a ferromagnetic Fe slab, represented by a superlattice containing 16 (001)-oriented layers of bcc-structure Fe separated by 102 Å-thick vacuum layers (figure 1*b*, upper panel). We set the in-plane lattice constant of the bcc Fe unit cell to 2.98 Å, corresponding to an in-plane Fe–Fe distance of 2.43 Å, so that it forms a coherent interface with MgO in the Fe/MgO heterostructures that we study next ($a_{\mathrm{MgO}}/\sqrt{2} = 2.98$ Å). We then relax the out-of-plane Fe–Fe distances, and obtain a Fe–Fe distance of 2.41 Å in the outer layers, while the Fe–Fe distance converges to 2.43 Å in the bulk-like interior of the slabs. For this strain state, the magnetization orients in the uniaxial out-of-plane direction (note that we did not include a demagnetizing field in our calculations); in our presented results it is along the positive direction of figure 1*b*,*c*. In figure 1*a*, we show the magnetic dipole moment on each Fe atom in the slab as a function of the layer. We obtain a magnetic moment of 2.7 μ_{B} on the Fe atoms at the surfaces and an interior value of 1.5 μ_{B}, differing slightly from the LDA bulk value of 2.2 μ_{B} due to the epitaxial constraint on the in-plane lattice parameters of our slab.

A symmetry analysis reveals that in the slab geometry we are considering—a tetragonal structure with purely out-of-plane magnetization—only the magnetoelectric monopole and the q_z component of the magnetoelectric quadrupole are non-zero. The calculated magnetoelectric monopoles are shown in figure 1*b* as a function of their layer number in the slab. (Since the q_z quadrupole component is proportional to the magnetoelectric monopole in our calculation, we do not list it explicitly.) The monopoles are zero towards the centre of the slab, where the local atomic environment is close to the bulk structure and the influence of the inversion-symmetry breaking at the surface becomes negligible, but are non-zero at the surfaces, with opposite signs at opposite surfaces. The opposite signs can be understood either in terms of the opposite position of the vacuum relative to the magnetization orientation at the two surfaces, or by the fact that the surfaces are related to each other by a glide plane, which reverses the monopole sign.

Next, we repeat our calculations for a heterostructure of the same 16 layers of Fe, this time adjacent to nine layers of (001)-oriented MgO with vacuum on each side. In our DFT calculations, the most stable configuration has Fe atoms situated on top of O atoms, as found previously by Butler *et al.* [34], with a Fe-O distance of 2.12 Å. (Note that, while our set-up is similar to that of

rsta.royalsocietypublishing.org Phil. Trans. R. Soc. A 376: 20170450

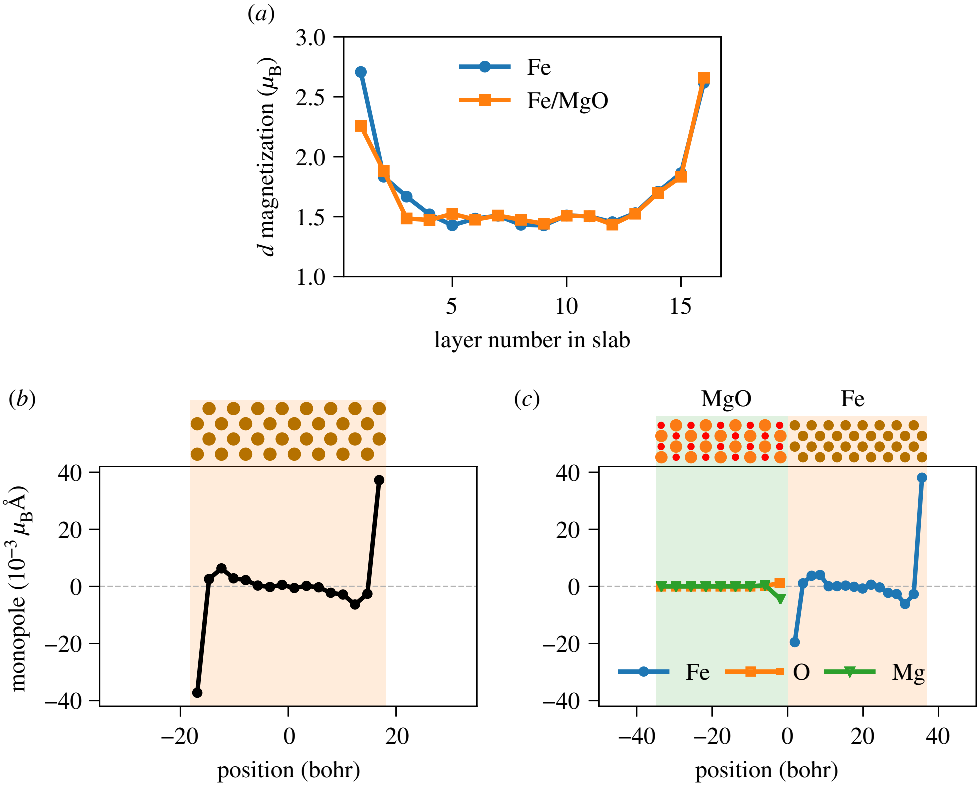


Figure 1. (*a*) Magnetic dipole moment per Fe atom in the Fe slab (blue circles) and the Fe/MgO slab (orange squares). The MgO is at the left side of the figure. (*b*) Magnetoelectric monopoles on each Fe atom in the Fe slab. (*c*) Atomic magnetoelectric monopoles in the Fe/MgO slab. (Online version in colour.)

[34], we assume a fixed MgO lattice constant instead of a fixed Fe lattice constant.) As for the Fe slab, we find that the ferromagnetic magnetization in the Fe slab is oriented in the out-of-plane direction. The magnetic moment on the Fe atom at the Fe/MgO interface is reduced to $2.3\,\mu_B$ compared with $2.7\,\mu_B$ at the Fe/vacuum interface (figure 1*a*).

Our calculated magnetoelectric monopoles are shown in figure 1*c*. Again, we find that the monopoles are zero in the interior of the Fe slab, and that the surface layers have the largest, and oppositely signed, values. The monopole size is reduced to $20 \times 10^{-3}\,\mu_B$ Å at the Fe/MgO interface, around half of its magnitude at the Fe/vacuum interface, reflecting (although larger than) the decrease of the interfacial Fe magnetic moment from $2.7\,\mu_B$ to $2.3\,\mu_B$. A small monopole is also visible on the Mg and O atoms in the interface layer, indicating a spillover of the magnetic polarization from Fe into the MgO; our calculations yield small magnetic dipole moments of $0.01\,\mu_B$ on the Mg atom and $0.06\,\mu_B$ on the O atom adjacent to the interface.

The patterns of monopoles that we obtain for both the Fe and Fe/MgO slabs reflect the pattern of electric field-induced magnetization presented for $SrRuO_3/SrTiO_3$ heterostructures in [9], in that they are largest at the surfaces and of opposite sign at either surface of the slab. Therefore, as a next step, we calculate the changes in magnetization induced by electric fields applied perpendicular to the surface.

We apply an electric field perpendicular to the slabs by applying a sawtooth potential with a discontinuity in the vacuum region, and set the average field in the supercell to $514\,\mathrm{V\,\mu m^{-1}}$ for both Fe and Fe/MgO cases. Since the field is screened in the metal, the field in the vacuum depends on the size of the vacuum and dielectric regions in the slab supercell. From our DFT calculations, we find that the electric field in the vacuum region is $626\,\mathrm{V\,\mu m^{-1}}$ in the Fe slab, and

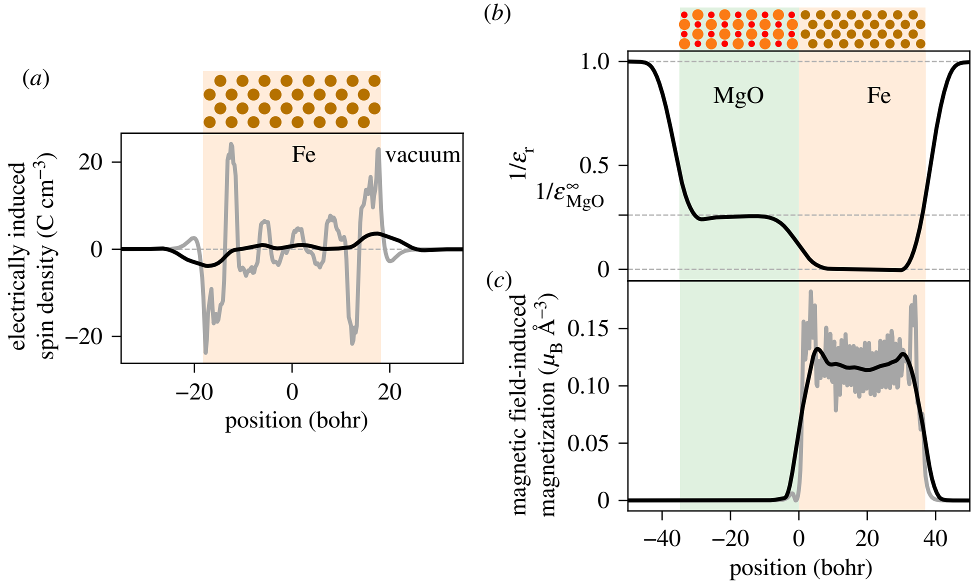


Figure 2. (*a*) Macroscopically and planar-averaged magnetoelectric response in slab of Fe along the slab axis, i.e. induced magnetization under an electric field. Black line denotes atomically smoothed response; grey line, unsmoothed response. (*b*) Inverse dielectric constant in the Fe/MgO slab with vacuum on both sides. (*c*) Magnetic field-induced magnetization, that is the magnetic susceptibility, of the Fe/MgO slab. (Online version in colour.)

$735\,\mathrm{V\,\mu m^{-1}}$ in the Fe/MgO slab, where the field is reduced in the dielectric. Note that we extract here the electronic contributions to the magnetoelectric response by performing all calculations at fixed ionic structure.

The macroscopic and planar-averaged electric field-induced magnetization in the Fe slab is shown in figure 2*a*. We see that the pattern of magnetoelectric response follows closely the pattern of the magnetoelectric monopoles presented above. First, it is non-zero only in the surface regions, where the inversion symmetry is broken. Second (and as seen previously for SRO/STO), it is largest on the surface atoms with a small contribution of opposite sign on the next-nearest atoms. Third, at opposite surfaces, the response has opposite signs. Specifically, the change in magnetic dipole moment on the leftmost Fe atom is $-1.18 \times 10^{-3}\,\mu_B$, while the change in magnetic dipole moment on the rightmost Fe atom is $1.21 \times 10^{-3}\,\mu_B$; in addition, a small overall ferromagnetic component $0.33 \times 10^{-3}\,\mu_B$ is induced. The magnetoelectric response in the Fe/MgO slab follows the same pattern, but with a slight reduction in induced magnetization at the Fe/MgO interface compared with the Fe/vacuum interface, $1.33 \times 10^{-3}\,\mu_B$. Note, however, that due to our choice of electrostatic boundary conditions, the voltage at each interface is different, which forbids a direct comparison of the magnitude of the induced response. Again there is a net induced ferromagnetic component of $0.38 \times 10^{-3}\,\mu_B$.

Since the magnetoelectric susceptibility, α, is a bulk property, it is not the relevant quantity for describing surface electric field-induced magnetism. Instead, Rondinelli *et al.* introduced the concept of spin capacitance, $C^s = \sigma^s/V$, which, by analogy to the usual charge capacitance, C, is the spin polarization per unit area induced by the voltage V [9]. They then suggested a magnetoelectric 'figure of merit' given by $\eta = C^s/C$ [9]. For our Fe/vacuum slabs, we obtain $\eta = 0.28$, similar to that found earlier for the SRO/STO interface [9], and for the Fe/MgO interface in our Fe/MgO slabs, we obtain $\eta = 1.27$. The large and somewhat unintuitive $\eta > 1$ is a result of spin transfer from the majority to minority channel at the Fe/MgO interface, in addition to the capacitive charge and spin accumulation. Whether this huge magnetoelectric figure of merit is relevant for the favourable tunnelling magnetoresistance in Fe/MgO heterostructures [34] is an interesting question for future exploration.

Lastly, we reflect on the relation between susceptibility and magnetoelectric response. In bulk materials, the diagonal magnetoelectric response α is known to be related to the size of the magnetoelectric monopole per unit volume by

$$\alpha = c(\epsilon_r - 1)\chi_m A, \tag{3.1}$$

where ϵ_r is the relative permittivity, χ_m is the magnetic susceptibility, A is the magnetoelectric monopole per unit volume and c is a proportionality constant [2,5].

In figure 2*b*, we show our calculated inverse relative permittivity and in figure 2*c* the magnetization induced by a Zeeman field of 1 mT for the Fe/MgO slab. (The interface in the Fe/vacuum slab behaves similarly to the Fe/vacuum interface shown here.) As expected, the relative permittivity is unity in the vacuum and diverges in the metallic region. It has a finite value in the dielectric and at the interfaces of the metal. The magnetic susceptibility, on the other hand, is only non-zero in the metallic (Fe) region. We see that the product of both susceptibilities is non-zero and finite only in the region in which we observe a magnetoelectric response.

In summary, our first-principles calculations confirm that magnetoelectric monopoles and quadrupoles can be generated in nominally centrosymmetric Fe by introducing interfaces that break the inversion symmetry. These magnetoelectric multipoles are large only at the interfaces and vanish rapidly towards the bulk region. At the same time, since such a system is insulating in the direction normal to the surface, it exhibits a magnetoelectric response, which, like the multipoles, is large only at the interfaces. We find that the sizes of the magnetic dipole moment and the magnetoelectric monopole depend on the detailed nature of the interface, with those at the Fe/MgO interface being smaller than at the Fe/vacuum interface. Finally, we show that, while the magnetoelectric response coincides with the region in which the product of electric and magnetic susceptibilities is finite and non-zero, there is no obvious connection between their magnitudes.

(b) $SrCaRu_2O_6$

We now turn our attention to the hypothetical magnetic polar metal, A-site ordered $SrCaRu_2O_6$, in which it has been shown computationally that the magnetism can be modified by modulation of the non-centrosymmetric structural distortion. The disordered solid solution series $Sr_{1-x}Ca_xRuO_3$ of the isostructural perovskites $SrRuO_3$ and $CaRuO_3$ exists experimentally, and is metallic and non-polar at all compositions, with the centrosymmetric *Pnma* space group (octahedral tilt pattern $a^+b^-c^-$) [35]. From $x = 0$ up to $x \approx 0.7$, it is an itinerant ferromagnet with $T_C \approx 57$ K [13]. Above this value of x, the ferromagnetic order is suppressed.

Puggioni & Rondinelli showed using DFT calculations that inversion symmetry is broken when the A-site cations are ordered in layers along the c-direction [11], while the magnetic and metallic behaviour persist. Constraining the a and b lattice constants to be equal to mimic coherent epitaxial growth, they obtained a $Pmc2_1$ space group, with the same tilt pattern as the *Pnma* of the disordered alloy and an additional distortion corresponding to the polar Γ_5^- mode (shown in figure 3*a*) of the parent $P4/mmm$ space group. Interestingly, suppression of the polar Γ_5^- mode caused a collapse of the ferromagnetic order [11]; we will use this fact to investigate the interplay between polarization, magnetization and magnetoelectric multipoles.

Using the ground-state structure of the layered A-site ordered compound from [11], we calculate the electronic and magnetic structure with spin-orbit coupling included. Within the LDA, we obtain a magnetic moment on the Ru atoms of $0.1\,\mu_B$; this increases to $0.7\,\mu_B$ with even a small $U = 0.5$ eV. Since the experimentally measured value for the magnetic moment per Ru atom in disordered $Sr_xCa_{1-x}RuO_3$ with $x = 0.53$ is below $0.2\,\mu_B$ [13], we do not apply a Hubbard U correction in the following. We find that the magnetic dipole moments are oriented in the orthorhombic b-direction, leading to the magnetic space group $Pm'c2_1'$. Since the site symmetry of the Ru atoms on Wyckoff position 4c in this space group is 1, all magnetoelectric

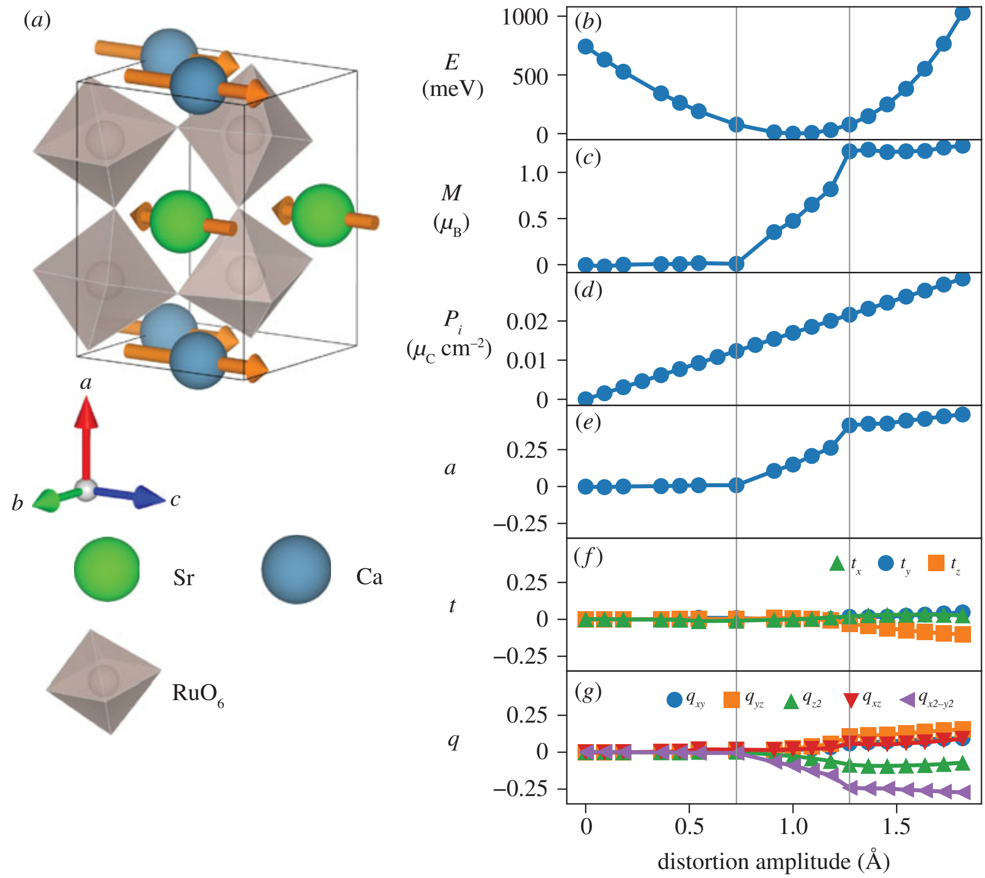


Figure 3. (*a*) Structure of A-site cation-ordered $SrCaRu_2O_6$. Orange arrows denote the atomic displacements in the polar Γ_5^- mode of $SrCaRu_2O_6$. (*b–g*) Properties of $SrCaRu_2O_6$ as a function of polar mode amplitude: (*b*) energy, *E*, (*c*) ferromagnetic magnetization, *M*, (*d*) ionic polarization of the lattice, P_i, (*e*) magnetoelectric monopoles, *a*, (*f*) all components of the toroidal moment, *t*, and (*g*) all quadrupoles, *q*. Note that the normalization of the distortion mode in this work is such that the relaxed structure has an amplitude of 1 Å, which is different from the normalization used in [11]. (Online version in colour.)

multipoles are allowed on each site. Symmetry analysis of the allowed arrangements of the magnetoelectric multipole orders in this space group yields the results summarized in table 1, where + and − indicate the sign of the allowed multipoles. For the t_x toroidal moment and the q_{yz} quadrupole, a ferro-type order is allowed, while the remaining multipoles order in different antiferro-type patterns. If the system were insulating, the corresponding bulk magnetoelectric effect in an insulator would have two non-zero components, α_{23} and α_{32}, with the symmetric part, $(\alpha_{23} + \alpha_{32})/2$, determined by the magnetoelectric quadrupole q_{yz}, and the antisymmetric part, $(\alpha_{23} - \alpha_{32})/2$, determined by the toroidal moment t_x.

Next, we analyse the behaviour of the magnetic order when the polar mode amplitude is changed. Keeping all modes except for the polar mode at their bulk amplitudes, we calculate the energy and magnetic structure for several different amplitudes of the polar mode between zero and 1.8 Å, where we normalize the mode amplitude such that the ground state structure corresponds to an amplitude of 1 Å. Figure 3*b*,*c* shows our calculated energy and magnetization as a function of amplitude. At zero and small amplitude, we obtain a non-magnetic solution, with the onset to the ferromagnetic state occurring at a mode amplitude of 0.72 Å. The ferromagnetic moment then increases up to 1.2 μ_B at a mode amplitude of 1.27 Å where it saturates. The ground-state structure has its mode amplitude in the intermediate region in which the magnetization has

Table 1. Symmetry analysis of the dipolar and multipolar order on the Ru atoms, which occupy the Wyckoff position 4*c* in the $Pmc2_1$ space group. The Ru magnetic dipole moments order ferromagnetically and are oriented along *y*, leading to the magnetic space group $Pm'c2_1'$. This results in a ferro ordering of the t_x and q_{xy} multipoles, and the antiferro orderings shown for the other multipoles.

atom	m_y	$a, q_{z^2/x^2-y^2}$	t_x, q_{yz}	t_y, q_{xz}	t_z, q_{xy}
Ru1	+	+	+	+	−
Ru2	+	−	+	−	−
Ru3	+	−	+	+	+
Ru4	+	+	+	−	+

Table 2. Calculated size of the magnetoelectric multipoles in $SrCaRu_2O_6$ in the equilibrium structure.

($\times 10^{-3}\ \mu_B$ Å)								
a	t_x	t_y	t_z	q_{xy}	q_{yz}	q_{z^2}	q_{xz}	$q_{x^2-y^2}$
0.15	−0.00	0.00	0.01	0.01	0.03	−0.03	0.02	−0.09

not reached its saturation value. Note that at all amplitudes, antiferromagnetic orders with the same magnetic unit cell are higher in energy than the calculated non-magnetic and ferromagnetic orders.

Next, we compute the magnetoelectric multipoles as a function of the polar mode amplitude and show the size of the monopole, toroidal moment and quadrupole components as a function of mode amplitude in figure 3*e*–*g*. As expected, at zero and small distortion amplitude there are no multipoles, since there is no time-reversal symmetry breaking magnetic dipole order. In the intermediate region, the size of all multipole components increases roughly proportionally to the size of the ferromagnetic dipole moment. The sizes of the multipoles at the ground state amplitude are listed in table 2. At this amplitude, the toroidal moments are essentially zero, and the magnetoelectric monopole is the largest component. In the region in which the magnetization saturates, the magnetoelectric multipoles saturate also, showing negligible change with further increase of the polar mode amplitude.

In summary, we find a region of polar distortion amplitude in which both the usual dipolar magnetization as well as the magnetoelectric multipoles scale proportionally to the strength of the inversion-symmetry breaking. While the polar mode amplitude cannot be modified using an electric field due to metallic screening, the reciprocal effect—modification of the magnetization with a magnetic field to tune the amplitude of the polar structural distortion—should be accessible. This hidden magnetoelectric response might be observable using second harmonic generation, which is sensitive to the inversion-symmetry breaking. Since the magnetoelectric multipoles saturate together with the magnetization even when the polarization amplitude continues to increase, we conclude that they scale with the magnitude of the dipolar magnetization rather than being explicitly sensitive to the magnitude of the inversion-symmetry breaking.

(c) $Ca_3Ru_2O_7$

$Ca_3Ru_2O_7$ is a member of the $A_{n+1}B_nX_{3n+1}$ Ruddelsden–Popper series with $n = 2$; its structure is shown in figure 4*a*. Much previous work has focused on understanding its electrical and magnetic phase diagram, sketched in figure 4*b* [14,15,17]. Above $T_{MIT} = 48$ K, the system is metallic in-plane and insulating out-of-plane, in the sense that the in-plane resistivity ρ_a decreases with decreasing temperature, while the out-of-plane resistivity ρ_c shows increasing resistivity with decreasing temperature [14]. Below T_{MIT}, it shows insulating behaviour in all directions, until,

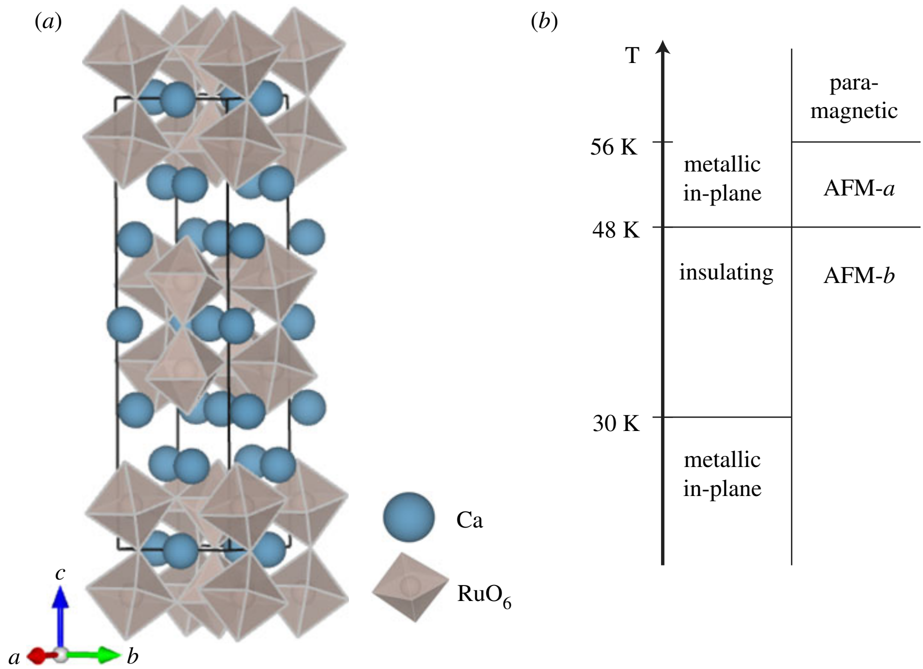


Figure 4. (*a*) Crystal structure of $Ca_3Ru_2O_7$. (*b*) Schematic of the electronic and magnetic phase diagram. (Online version in colour.)

below 30 K, it regains metallic conductivity in the a-b plane. At $T_N = 56$ K, within the metallic phase, it undergoes a phase transition from paramagnetic to antiferromagnetic with the so-called AFM-*a* arrangement, in which each double layer is ordered ferromagnetically with AFM coupling between the double layers. In this phase, the anisotropy causes the spins to lie along the *a* axis [16]. At T_{MIT}, the magnetic moments reorient to align parallel to the *b* axis while keeping the overall magnetic ordering, forming the so-called AFM-*b* state. The *c*-axis magnetoresistance is different between the two orientations of the AFM order and exhibits a pronounced temperature dependence around the transition temperatures [19]. This has been attributed to the strong spin-charge coupling in the material [17,36].

In the following, we will show that, while the point group symmetry of the material contains time reversal, magnetoelectric multipoles are realized in both antiferromagnetic phases because the *site symmetry* is not time-reversal symmetric. We will further show that magnetoelectric multipoles provide a sensitive indicator of the differences in hybridization between the two different magnetic orientations, and so can be helpful in revealing the coupling between charge and spin degrees of freedom. As a consequence, we will argue that the magnetoelectric multipoles are a sensitive tool for characterizing the nature of the microscopic magnetic anisotropies, even when the magnetization densities are essentially identical for different choices of easy axis.

We start by analysing the magnetic symmetry of the system. The *b*-axis anisotropy of the $Bb2_1m$ space group [15,16] leads to a magnetic space group of $B_Pb'2_1m'$ (in OG setting), while the AFM-*a* phase has the magnetic space group $B_Pb2_1'm'$. In both cases, the magnetic point group is $mm21'$ and so contains time-reversal symmetry. This prohibits a ferro-type order of magnetoelectric multipoles, so if $Ca_3Ru_2O_7$ were insulating a bulk magnetoelectric effect would not be allowed. Antiferro-type multipolar orders are allowed, however, because the time-reversal symmetry in this magnetic space group occurs in combination with a translation through the *B* centring of the unit cell. This is consistent with the 1 site symmetry of the Ru atoms on Wyckoff position 8*b*, which does not contain time-reversal (or any other) symmetry. As a result, magnetoelectric multipoles occur on the individual Ru sites, arranged in the antiferro-type orders shown in table 3.

Table 3. Allowed multipole orders on Wyckoff position 8*b* in the $Bb2_1m$ space group, for antiferromagnetic arrangements of magnetic moments aligned along the *a*-axis (AFM-*a*) and along the *b* axis (AFM-*b*).

	AFM-*a*				AFM-*b*			
Wyckoff position 8*b*	$a, q_{z^2/x^2-y^2}$	t_x, q_{yz}	t_y, q_{xz}	t_z, q_{xy}	$a, q_{z^2/x^2-y^2}$	t_x, q_{yz}	t_y, q_{xz}	t_z, q_{xy}
(x, y, z)	+	−	+	−	−	−	−	−
$(-x, y+\frac{1}{2}, -z)$	−	−	−	−	−	+	−	+
$(-x, y+\frac{1}{2}, z)$	−	+	+	−	−	−	+	+
$(x, y, -z)$	+	+	−	−	−	+	+	−
$(x+\frac{1}{2}, y, z+\frac{1}{2})$	−	+	−	+	+	+	+	+
$(-x+\frac{1}{2}, y+\frac{1}{2}, -z+\frac{1}{2})$	+	+	+	+	+	−	+	−
$(-x+\frac{1}{2}, y+\frac{1}{2}, z+\frac{1}{2})$	+	−	−	+	+	+	−	−
$(x+\frac{1}{2}, y, -z+\frac{1}{2})$	−	−	+	+	+	−	−	+

The structural and magnetic properties of $Ca_3Ru_2O_7$ have been investigated previously using density functional calculations [18,37–40], and the calculated properties were found to depend strongly on the choice of Hubbard *U*. In all calculations, the magnetic ground state was found to have the magnetic ordering of the AFM-*b* phase. Without including a Hubbard *U* or spin-orbit interactions, calculations using LDA or GGA obtained a metallic system [37,38]. Including a moderate $U = 2$ eV on the Ru atoms led to a gap opening in higher energy AFM phases, but the AFM-*b* phase retained its metallic character [39]. Further increase of *U* to 3.5 eV and inclusion of spin-orbit interactions opened a gap in the AFM-*b* phase [40]. We emphasize that the most appropriate description of this correlated oxide is a difficult and ongoing question [41] which we do not address here. Rather, since our motivation is to use $Ca_3Ru_2O_7$ as a model system to establish the existence of magnetoelectric multipoles, we take the simplest method that gives qualitatively correct behaviour, that is the LDA method with $U = 2$ eV.

Using the experimental crystal structure from [15], we calculate the electronic structure as described in §2, imposing the AFM-*b* and AFM-*a* magnetic order in turn. In both cases, we obtain a metallic system in which the density of states at the Fermi level stems mainly from the Ru *d* orbitals, with the AFM-*a* phase being 1.5 meV Ru^{-1} atom higher in energy. We find that a small antiferromagnetic tilting of the magnetic moments of the Ru atoms away from the *a* (for AFM-*a*) or *b* (for AFM-*b*) easy axis is energetically favourable; this is allowed by symmetry for the Wyckoff position 8*b* occupied by the Ru atoms. In the following, we neglect this small rotation and constrain the moments to lie along the *a* axis for the AFM-*a* and along the *b* axis for the AFM-*b* structure, respectively, to allow for a more straightforward comparison.

The magnetization density along the *a*-direction for the AFM-*a* phase is shown in figure 5*a*, while the magnetization density along the *b*-direction for the AFM-*b* phase is shown in figure 5*b*. In spite of the striking difference in properties measured for the two phases—AFM-*b* has a higher out-of-plane resistivity with a different magnetic-field dependence at temperatures around the metal-insulator transition—we see that the shape of the magnetization densities is indistinguishable on this scale. The difference between the magnetization densities, shown in figure 5*c*, is tiny, although one can resolve small changes in the regions close to the Ru atoms. This suggests a small rehybridization of the Ru *d* orbitals. The calculated density of states of the Ru *d* bands around the Fermi level, shown in figure 6, barely shows this difference between the two magnetization directions.

These small differences in magnetization are revealed much more strikingly in the magnetoelectric multipoles, which we report in table 4. As a result of the low symmetry, all components are non-zero in both magnetic phases. We see that the monopole term is strongest

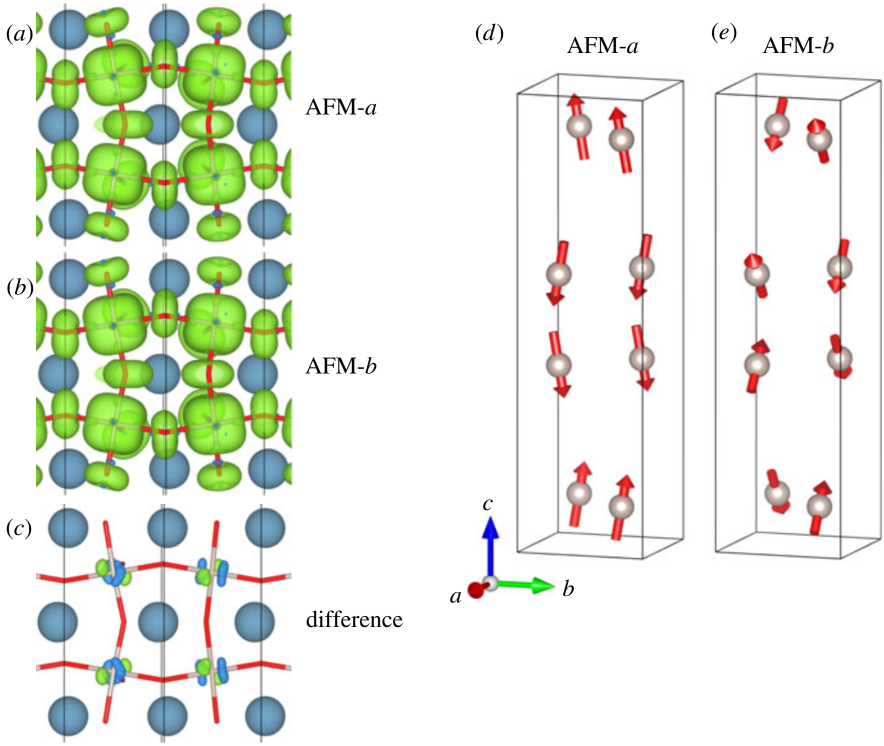


Figure 5. (*a*) Magnetization density for AFM-*a*-ordered $Ca_3Ru_2O_7$. The isosurface level is 2×10^{-3} μ_B Å^{-3}. (*b*) Magnetization density for AFM-*b*-ordered $Ca_3Ru_2O_7$. The isosurface level is 2×10^{-3} μ_B Å^{-3}. (*c*) Difference between magnetization densities of AFM-*a* and AFM-*b* orderings. The isosurface level is 1×10^{-3} μ_B Å^{-3}. (*d*), (*e*) Toroidal moments on Ru atoms in $Ca_3Ru_2O_7$ when the magnetic moments are aligned along *a* (AFM-*a*) or *b* (AFM-*b*). For the AFM-*a* structure, the toroidal moments lie in the *b*–*c* plane and are approximately oriented along *c*, while for AFM-*b*, the toroidal moments lie in the *a*–*c* plane and are approximately oriented along *c*. The Ca and O atoms are not shown. (Online version in colour.)

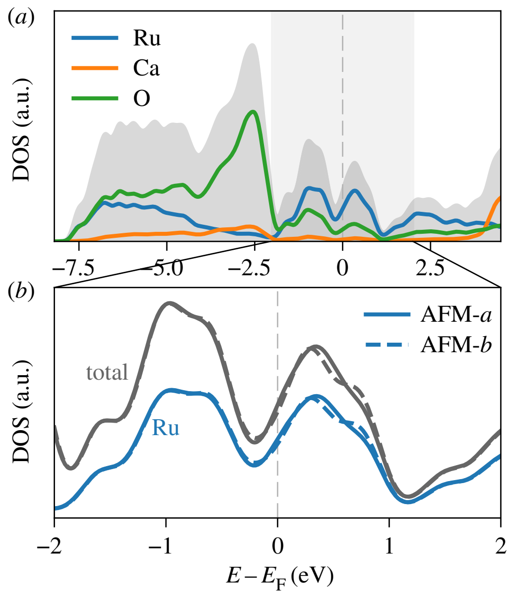


Figure 6. (*a*) Atomic-orbital projected density of states (DOS) for $Ca_3Ru_2O_7$ with AFM-*a* order. (*b*) Comparison of the DOS around the Fermi energy for AFM-*a* and AFM-*b* orders. (Online version in colour.)

Table 4. Sizes of the Ru-atom magnetoelectric multipoles in $Ca_3Ru_2O_7$, for both AFM-*a* and AFM-*b* phases.

	($10^{-3}\mu_B$Å)								
	a	t_x	t_y	t_z	q_{xy}	q_{yz}	q_{z^2}	q_{xz}	$q_{x^2-y^2}$
AFM-*a*	0.29	−0.01	0.06	−0.35	−0.36	0.01	−0.42	−0.19	0.43
AFM-*b*	−0.60	−0.19	0.00	−0.25	0.41	−0.06	0.16	0.04	0.36

in the AFM-*b* phase, while the average toroidal and quadrupole moments are larger in the AFM-*a* phase.

We focus on the toroidal moments, shown in figure 5*d*,*e*, to analyse the differences in multipole behaviour between the two magnetic orderings. Since there is no component of toroidal moment parallel to the magnetic moment, the toroidal moments are aligned perpendicular to the magnetic moments, that is in the *b*–*c* plane for AFM-*a* and the *a*–*c* plane for AFM-*b*. They are tilted away from the *c* axis by 10° in AFM-*a* and 36° in AFM-*b*. It is clear that the crystallographic differences in the orthorhombic *a* and *b* directions, which cause the tiny changes in magnetizations from the different hybridization of the Ru *d*–O *p* orbitals, manifest as distinctly different toroidal moments with different magnitudes and relative orientations.

In summary, while the symmetry of $Ca_3Ru_2O_7$ does not allow macroscopic multipole order, we find a hidden antiferromultipolar order in the magnetoelectric monopole, toroidal moment and quadrupole. The distinctly different magnetoelectric multipoles displayed by different magnetic orientations provide a useful handle for quantifying subtle rearrangements of magnetization density that are not readily revealed from an analysis of the magnetic dipole contribution alone. While there is no obvious connection between the differences in magnetoelectric multipoles and the differences in transport properties between the two magnetic orderings, this could be an interesting consideration for future work.

4. Conclusion

In conclusion, we have shown for the first time that second-order magnetoelectric multipoles exist as a 'hidden order' in non-centrosymmetric magnetic metallic systems. We investigated their behaviour in three different systems: the surface of ferromagnetic Fe, which is centrosymmetric in its bulk form, the hypothetical polar magnetic metal $SrCaRu_2O_6$, in which the polar mode strongly modifies the magnetization, and the known non-centrosymmetric antiferromagnetic metal, $Ca_3Ru_2O_7$.

We identified magnetoelectric monopoles at surfaces and interfaces of centrosymmetric Fe, and showed that they are consistent with the carrier-mediated electric field-induced magnetism previously reported in related systems. We showed that the magnetoelectric multipoles can be controlled by modulating the amplitude of the polar Γ_5^- mode in $SrCaRu_2O_6$, with their size corresponding closely to that of the corresponding dipolar magnetization. However, since the amplitude of the polar mode cannot be modified by an applied electric field, there is no accompanying magnetoelectric effect. Finally, we identified hidden antiferro-ordered magnetoelectric multipoles in both magnetic phases of $Ca_3Ru_2O_7$, and showed that they depend strongly on the orientation of the antiferromagnetic dipole moments. Consequently, they provide a sensitive indicator of the changes in magnetization density associated with the spin reorientation.

We hope that the identification of magnetoelectric multipoles in magnetic metals achieved in this work motivates future studies of the relationship between the magnetoelectric multipoles and properties such as magnetoresistance and spin-dependent transport in polar magnetic metals.

Data accessibility. All input data and results from DFT calculations are available on request.

Authors' contributions. F.T. and N.A.S. conceived of the study and discussed the research. F.T. carried out the first-principles calculations. Both authors contributed to the writing of the manuscript and gave final approval for publication.

Competing interests. The authors declare that they have no competing interests.

Funding. This work was supported financially by the ETH Zürich, by the Max Rössler Price of the ETH Zürich, the Körber Foundation and by the Sinergia program of the Swiss National Science Foundation grant no. CRSII2_147606/1. This work was supported by a grant from the Swiss National Supercomputing Center (CSCS) under project IDs s624 and p504.

Acknowledgements. We thank Massimiliano Stengel, Maxim Mostovoy and Sang-Wook Cheong for insightful discussions.

References

1. Landau LD, Lifshitz EM, Pitaevskii LP. 1984 *Electrodynamics of continuous media*, 2nd edn. London, UK: Butterworth-Heinemann.
2. Spaldin NA, Fechner M, Bousquet E, Balatsky A, Nordström L. 2013 Monopole-based formalism for the diagonal magnetoelectric response. *Phys. Rev. B* **88**, 094429. (doi:10.1103/PhysRevB.88.094429)
3. Ederer C, Spaldin NA. 2007 Towards a microscopic theory of toroidal moments in bulk periodic crystals. *Phys. Rev. B* **76**, 214404. (doi:10.1103/PhysRevB.76.214404)
4. Spaldin NA, Fiebig M, Mostovoy M. 2008 The toroidal moment in condensed-matter physics and its relation to the magnetoelectric effect. *J. Phys. Condens. Matter* **20**, 434203. (doi:10.1088/0953-8984/20/43/434203)
5. Thöle F, Fechner M, Spaldin NA. 2016 First-principles calculation of the bulk magnetoelectric monopole density: Berry phase and Wannier function approaches. *Phys. Rev. B* **93**, 195167. (doi:10.1103/PhysRevB.93.195167)
6. Suzuki MT, Koretsune T, Ochi M, Arita R. 2017 Cluster multipole theory for anomalous Hall effect in antiferromagnets. *Phys. Rev. B* **95**, 094406. (doi:10.1103/PhysRevB.95.094406)
7. Suzuki MT, Ikeda H, Oppeneer PM. 2018 First-principles theory of magnetic multipoles in condensed matter systems. *J. Phys. Soc. Jpn* **87**, 041008. (doi:10.7566/JPSJ.87.041008)
8. Watanabe H, Yanase Y. 2017 Magnetic hexadecapole order and magnetopiezoelectric metal state in $Ba_{1-x}K_xMn_2As_2$. *Phys. Rev. B* **96**, 064432. (doi:10.1103/PhysRevB.96.064432)
9. Rondinelli JM, Stengel M, Spaldin NA. 2008 Carrier-mediated magnetoelectricity in complex oxide heterostructures. *Nat. Nanotechnol.* **3**, 46–50. (doi:10.1038/nnano.2007.412)
10. Duan CG, Velev JP, Sabirianov RF, Zhu Z, Chu J, Jaswal SS, Tsymbal EY. 2008 Surface magnetoelectric effect in ferromagnetic metal films. *Phys. Rev. Lett.* **101**, 137201. (doi:10.1103/PhysRevLett.101.137201)
11. Puggioni D, Rondinelli JM. 2014 Designing a robustly metallic noncenstrosymmetric ruthenate oxide with large thermopower anisotropy. *Nat. Commun.* **5**, 3432. (doi:10.1038/ncomms4432)
12. Puggioni D, Rondinelli JM. 2014 Linear optical and electronic properties of the polar metallic ruthenate (Sr,Ca)Ru_2O_6. *J. Phys. Condens. Matter* **26**, 265501. (doi:10.1088/0953-8984/26/26/265501)
13. Cao G, McCall S, Shepard M, Crow J, Guertin R. 1997 Thermal, magnetic, and transport properties of single-crystal $Sr_{1-x}Ca_xRuO_3$ ($0 \leq x \leq 1.0$). *Phys. Rev. B* **56**, 321–329. (doi:10.1103/PhysRevB.56.321)
14. Yoshida Y, Nagai I, Ikeda SI, Shirakawa N, Kosaka M, Môri N. 2004 Quasi-two-dimensional metallic ground state of $Ca_3Ru_2O_7$. *Phys. Rev. B* **69**, 220411. (doi:10.1103/PhysRevB.69.220411)
15. Yoshida Y, Ikeda SI, Matsuhata H, Shirakawa N, Lee CH, Katano S. 2005 Crystal and magnetic structure of $Ca_3Ru_2O_7$. *Phys. Rev. B* **72**, 054412. (doi:10.1103/PhysRevB.72.054412)
16. Bao W, Mao ZQ, Qu Z, Lynn JW. 2008 Spin valve effect and magnetoresistivity in single crystalline $Ca_3Ru_2O_7$. *Phys. Rev. Lett.* **100**, 247203. (doi:10.1103/PhysRevLett.100.247203)
17. Baumberger F, Ingle NJC, Kikugawa N, Hossain MA, Meevasana W, Perry RS, Shen KM, Lu DH, Damascelli A, Rost A, Mackenzie AP, Hussain Z, Shen Z-X. 2006 Nested Fermi surface and electronic instability in $Ca_3Ru_2O_7$. *Phys. Rev. Lett.* **96**, 107601. (doi:10.1103/PhysRevLett.96.107601)

18. Zhu M, Peng J, Zou T, Prokes K, Mahanti SD, Hong T, Mao ZQ, Liu GQ, Ke X. 2016 Colossal magnetoresistance in a Mott insulator via magnetic field-driven insulator-metal transition. *Phys. Rev. Lett.* **116**, 216401. (doi:10.1103/PhysRevLett.116.216401)
19. Fobes D, Peng J, Qu Z, Liu TJ, Mao ZQ. 2011 Magnetic phase transitions and bulk spin-valve effect tuned by in-plane field orientation in $Ca_3Ru_2O_7$. *Phys. Rev. B* **84**, 014406. (doi:10.1103/PhysRevB.84.014406)
20. Kresse G, Furthmüller J. 1996 Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11 169–11 186. (doi:10.1103/PhysRevB.54.11169)
21. Kresse G, Furthmüller J. 1996 Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **6**, 15–50. (doi:10.1016/0927-0256(96)00008-0)
22. Kresse G. 1999 From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758–1775. (doi:10.1103/PhysRevB.59.1758)
23. Giannozzi P, Baroni S, Bonini N, Calandra M, Car R, Cavazzoni C, Ceresoli D, Chiarotti GL, Cococcioni M, Dabo I, Dal Corso A, de Gironcoli S, Fabris S, Fratesi G, Gebauer R, Gerstmann U, Gougoussis C, Kokalj A, Lazzeri M, Martin-Samos L, Marzari N, Mauri F, Mazzarello R, Paolini S, Pasquarello A, Paulatto L, Sbraccia C, Scandolo S, Sclauzero G, Seitsonen AP, Smogunov A, Umari P, Wentzcovitch RM. 2009 QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *J. Phys. Condens. Matter* **21**, 395502. (doi:10.1088/0953-8984/21/39/395502)
24. Dal Corso A. 2014 Pseudopotentials periodic table: from H to Pu. *Comput. Mater. Sci.* **95**, 337–350. (doi:10.1016/j.commatsci.2014.07.043)
25. Monkhorst HJ, Pack JD. 1976 Special points for Brillouin-zone integrations. *Phys. Rev. B* **13**, 5188–5192. (doi:10.1103/PhysRevB.13.5188)
26. Neugebauer J, Scheffler M. 1993 Theory of adsorption and desorption in high electric fields. *Surf. Sci.* **287–288**, 572–576. (doi:10.1016/0039-6028(93)91030-S)
27. Bengtsson L. 1999 Dipole correction for surface supercell calculations. *Phys. Rev. B* **59**, 12 301–12 304. (doi:10.1103/PhysRevB.59.12301)
28. Bousquet E, Spaldin NA, Delaney KT. 2011 Unexpectedly large electronic contribution to linear magnetoelectricity. *Phys. Rev. Lett.* **106**, 107202. (doi:10.1103/PhysRevLett.106.107202)
29. Liechtenstein AI, Anisimov VI, Zaanen J. 1995 Density-functional theory and strong interactions: orbital ordering in Mott-Hubbard insulators. *Phys. Rev. B* **52**, R5467–R5470. (doi:10.1103/PhysRevB.52.R5467)
30. Giustino F, Pasquarello A. 2005 Theory of atomic-scale dielectric permittivity at insulator interfaces. *Phys. Rev. B* **71**, 144104. (doi:10.1103/PhysRevB.71.144104)
31. Momma K, Izumi F. 2011 VESTA 3 for three-dimensional visualization of crystal, volumetric and morphology data. *J. Appl. Crystallogr.* **44**, 1272–1276. (doi:10.1107/S0021889811038970)
32. Weisheit M, Fahler S, Marty A, Souche Y, Poinsignon C, Givord D. 2007 Electric field-induced modification of magnetism in thin-film ferromagnets. *Science* **315**, 349–351. (doi:10.1126/science.1136629)
33. Zhernenkov M, Fitzsimmons MR, Chlistunoff J, Majewski J, Tudosa I, Fullerton EE. 2010 Electric-field modification of magnetism in a thin CoPd film. *Phys. Rev. B.* **82**, 024420. (doi:10.1103/PhysRevB.82.024420)
34. Butler WH, Zhang XG, Schulthess TC, MacLaren JM. 2001 Spin-dependent tunneling conductance of Fe–MgO–Fe sandwiches. *Phys. Rev. B* **63**, 054416. (doi:10.1103/PhysRevB.63.054416)
35. Eom CB, Cava RJ, Fleming RM, van Dover RB, Phillips JM, Marshall JH, Hsu JWP, Krajewski JJ, Peck WF. 1992 Single-crystal epitaxial thin films of the isotropic metallic oxides $Sr_{1-x}Ca_xRuO_3$ ($0 \leq x \leq 1.0$). *Science* **258**, 1766–1769. (doi:10.1126/science.258.5089.1766)
36. Peng J, Hu J, Gu XM, Zhou GT, Liu JY, Zhang FM, Wu XS, Mao ZQ. 2016 Normal and inverse bulk spin valve effects in single-crystal ruthenates. *Appl. Phys. Lett.* **108**, 162402. (doi:10.1063/1.4947489)
37. Singh DJ, Auluck S. 2006 Electronic structure and bulk spin-valve behavior in $Ca_3Ru_2O_7$. *Phys. Rev. Lett.* **96**, 097203. (doi:10.1103/PhysRevLett.96.097203)
38. Ke X, Peng J, Singh DJ, Hong T, Tian W, Dela Cruz CR, Mao ZQ. 2011 Emergent electronic and magnetic state in $Ca_3Ru_2O_7$ induced by Ti doping. *Phys. Rev. B* **84**, 201102. (doi:10.1103/PhysRevB.84.201102)

39. Zou T, Cao HB, Liu GQ, Peng J, Gottschalk M, Zhu M, Zhao Y, Leão JB, Tian W, Mao ZQ, Ke X. 2016 Pressure-induced electronic and magnetic phase transitions in a Mott insulator: Ti-doped $Ca_3Ru_2O_7$ bilayer ruthenate. *Phys. Rev. B* **94**, 041115. (doi:10.1103/PhysRevB.94.041115)
40. Liu GQ. 2011 Mott transition and magnetic anisotropy in $Ca_3Ru_2O_7$. *Phys. Rev. B* **84**, 235137. (doi:10.1103/PhysRevB.84.235137)
41. Deng X, Haule K, Kotliar G. 2016 Transport properties of metallic ruthenates: a DFT+DMFT investigation. *Phys. Rev. Lett.* **116**, 256401. (doi:10.1103/PhysRevLett.116.256401)

Cover image
Courtesy of Institution of Engineering and Technology Archives.

Typeset by Nova Techset Private Limited, Bengaluru & Chennai, India. Printed in the UK by Cambrian Printers.

This paper meets the requirements of ISO 9706:1994(E) and ANSI/NISO Z39.48-1992 (Permanence of Paper) effective with volume 338, issue 1649, 1992.

Phil. Trans. R. Soc. A (ISSN 1364-503X) is published 26 times a year for US$6531 per year by the Royal Society, and is distributed in the USA by Agent named Air Business, C/O Worldnet Shipping USA Inc., 149-35 177th Street, Jamaica, New York, NY11434, USA. US Postmaster: Send address changes to *Phil. Trans. R. Soc. A*, C/O Air Business Ltd, C/O Worldnet Shipping USA Inc, 149-35 177th Street Jamaica, New York, NY11414.

The Royal Society is a self-governing Fellowship of many of the world's most distinguished scientists drawn from all areas of science, engineering, and medicine. The Society's fundamental purpose, as it has been since its foundation in 1660, is to recognise, promote, and support excellence in science and to encourage the development and use of science for the benefit of humanity.

The Society's strategic priorities emphasise its commitment to the highest quality science, to curiosity-driven research, and to the development and use of science for the benefit of society. These priorities are:

- Promoting science and its benefits
- Recognising excellence in science
- Supporting outstanding science
- Providing scientific advice for policy
- Fostering international and global cooperation
- Education and public engagement

For further information on the Royal Society
The Royal Society
6 – 9 Carlton House Terrace
London SW1Y 5AG

T +44 20 7451 2500
W royalsociety.org

For further information on *Philosophical Transactions of the Royal Society A*
T +44 20 7451 2214
E philtransa@royalsociety.org
W rsta.royalsocietypublishing.org

1364-503X(20181213)376:2134

ISBN: 978-1-78252-358-1
ISSN 1364-503X
The Royal Society Registered Charity No 207043

Phil. Trans. R. Soc. A | Volume 376 | Issue 2134 | 13 December 2018

125 years of Oliver Heaviside's 'Electromagnetic Theory'