Arch. Hist. Exact Sci.
DOI 10.1007/s00407-016-0176-1

Kirchhoff's theory for optical diffraction, its predecessor and subsequent development: the resilience of an inconsistent theory

Jed Z. Buchwald[1] · **Chen-Pang Yeang**[2]

Received: 13 January 2016

Abstract Kirchhoff's 1882 theory of optical diffraction forms the centerpiece in the long-term development of wave optics, one that commenced in the 1820s when Fresnel produced an empirically successful theory based on a reinterpretation of Huygens' principle, but without working from a wave equation. Then, in 1856, Stokes demonstrated that the principle was derivable from such an equation albeit without consideration of boundary conditions. Kirchhoff's work a quarter century later marked a crucial, and widely influential, point for he produced Fresnel's results by means of Green's theorem and function under specific boundary conditions. In the late 1880s, Poincaré uncovered an inconsistency between Kirchhoff's conditions and his solution, one that seemed to imply that waves should not exist at all. Researchers nevertheless continued to use Kirchhoff's theory—even though Rayleigh, and much later Sommerfeld, developed a different and mathematically consistent formulation that, however, did not match experimental data better than Kirchhoff's theory. After all, Kirchhoff's formula worked quite well in a specific approximation regime. Finally, in 1964, Marchand and Wolf employed the transformation of Kirchhoff's surface integral that had been developed by Maggi and Rubinowicz for other purposes. The result yielded a consistent boundary condition that, while introducing a species of discontinuity, nevertheless rescued the essential structure of Kirchhoff's original formulation from Poincaré's paradox.

Communicated by: Jed Buchwald.

✉ Jed Z. Buchwald
buchwald@caltech.edu

Chen-Pang Yeang
chenpang.yeang@utoronto.ca

1 California Institute of Technology, Pasadena, CA, USA

2 Institute for the History and Philosophy of Science and Technology, University of Toronto, Toronto, Canada

Published online: 23 February 2016

Contents

1 Introduction

On 22 June 1882, the University of Berlin's professor of theoretical physics, Gustav Robert Kirchhoff (1824–1887), read an influential paper titled "Zur Theorie der Lichtstrahlen" ("On the theory of light rays") to a meeting of the Prussian Academy of Sciences in Berlin. The purpose of the paper was to deduce from the wave equation the expression governing the diffraction of light by an aperture on an otherwise opaque screen. To do so Kirchhoff assumed a particular set of boundary conditions: namely, that both the amplitude of the disturbance as well as its spatial gradient vanished on the screen, but that they remained unaltered over the aperture itself.[1] He was in this way able to generate a solution for scalar diffraction that could yield the empirically-successful (so far as was then known) expression that Augustin Fresnel (1788–1827) had produced six decades before using an altogether different line of argument (on which more below). Despite its frequent presence in physicists' and engineers' publications, Kirchhoff's theory of diffraction has not until recently attracted the attention of physics (or mathematics) historians for understandable reasons: historical focus has principally aimed at episodes in nineteenth century physics that brought either fundamental changes—e.g. the wave theory of light, electromagnetic field theory, kinetic theory and statistical mechanics—or influential technological breakthroughs—e.g. Hertz's production of electric waves and the subsequent invention of wireless telegraphy. Kirchhoff's theory fits neither criterion. It did not introduce any novel physical entities or mechanisms beyond what wave optics had stipulated; nor did it lead to technological innovation. It nicely fits, one might say, Thomas Kuhn's conception of "normal science," in which practitioners solve problems that arise within a given system without violating its fundamental boundaries.[2]

Although Kirchhoff's theory did not have significant ontological or technological implications, it nonetheless raised important questions concerning the use of mathematics in theoretical physics. What made the theory interesting in subsequent years

[1] Kirchhoff (1882), also printed as Kirchhoff (1883). In this article, we use an English translation by Hentschel and Zhu (forthcoming).

[2] Kuhn (1962), pp. 10–34.

is the mathematical inconsistency of the boundary conditions that were used. So far as was known at the time, however, his solution worked quite well empirically, and Kirchhoff himself never remarked the inconsistency. Decades after the French mathematician Henri Poincaré (1854–1912) published a deleterious consequence of the inconsistency in 1892,[3] the theory nevertheless continued to appear in major textbooks and research periodicals in optics and electromagnetism. Physicists and engineers treated it not as an antiquated and inconsistent effort to derive empirically workable results, but as a good enough working model, for Kirchhoff's solution nicely fit both optical and microwave experimental data under particular, but commonly applicable, conditions. Indeed, physicists' and engineers' interest in, and use of, Kirchhoff's theory has certainly not waned over the decades.[4]

The persistent deployment of an apparently inconsistent theory even after recognition of its flaws is not unique in the history of physics. The infamous divergence of the quantum field integrals in self-energy calculations in quantum electrodynamics (QED) and physicists' various ad-hoc manipulations to bypass the problem before the introduction of renormalization provides one noteworthy example.[5] Another concerns early twentieth century Cambridge mathematicians' continued use of circulatory theory to explain airfoil lifting despite a salient contradiction—namely, d'Alembert's paradox, according to which there should be no lift at all in a perfect fluid—that had been well known for centuries.[6] In these cases, a major reason for the tenacity of an admittedly problematic theory was essentially pragmatic: the physical-mathematical problem was simply too complex, and no comparable alternative was available at the time.

Despite the similarities, Kirchhoff's account of diffraction differed in one essential respect from these two examples that makes this situation particularly compelling. In the first decades of the twentieth century, Lord Rayleigh (1842–1919) and Arnold Sommerfeld (1868–1951) derived different solutions to the same diffraction problem under a set of consistent boundary conditions. Unlike the cases of QED before renormalization and the circulatory theory of airfoil lifting in the early twentieth century, therefore, a mathematically consistent alternative was in fact available. Yet the existence of a seemingly more appealing and logical alternative did not eliminate or marginalize Kirchhoff's theory. This mathematically inconsistent (and physically untenable) solution has continued to appear and thrive in textbooks and periodicals as the standard approach to the problem of diffraction. Why, one may ask, did scientists continue to stay with Kirchhoff's theory despite the presence of a consistent alternative?

Recently, several historians and philosophers have begun to pay closer attention to this curious episode.[7] These studies deepen our understanding of conceptual and

---

[3] Poincaré (1892b), pp. 187–188.

[4] According to Werner Marx's bibliometric analysis, Kirchhoff's 1882 work from 1900 to 2010 has been specifically cited 70 times in research articles; moreover, the citation count has increased over the years: Marx (2016).

[5] Schweber (1994)

[6] Bloor (2011).

[7] Klaus Hentschel, Ning Yan Zhu, Ann Hentschel, and Werner Marx have provided a comprehensive historical study of Kirchhoff's 1882 paper by translating it into English, presenting a commentary, situating it within Kirchhoff's intellectual biography, and conducting a scientometric analysis of citations to it: Hentschel and Zhu (forthcoming). In light of the realism debate in philosophy of science, Juha Saatsi and

technical aspects of Kirchhoff's theory and help us clarify its philosophical implications, but in what follows we are concerned with the details of the theory's persistence and with the consistent alternatives to it. The theory's empirical success within the experimental regime of nineteenth and early twentieth century optics certainly goes reasonably far in accounting for its persistence. More, however, was involved than the results of experiment. The theory's tenacity also reflects the hold of a long-standing intellectual tradition in optics. Rooted in a principle introduced by Christiaan Huygens (1629–1695) in 1678 and first deployed for diffraction by Fresnel in 1818, this approach calculated wave intensity by means of a single integral over regions that are not blocked by a diffracting object. Such an integral could be interpreted as comprised of waves, or wavelets, that emanate from each of the points on the object's open regions and so on the incident wave front proper. Even after the scope of the inconsistency became altogether clear, physicists continued to seek (and to find) a single integral that could give physical meaning at least to a form of Huygens' principle, if not to the original, a principle that Kirchhoff's theory so nicely exemplified. In what follows we explore the structure of Kirchhoff's theory and the alternatives to it in order to bring out the several ways in which mathematical structures that exemplify principles akin to Huygens' generated theories of diffraction.

2 Fresnel applies Huygens' principle to diffraction

When Fresnel first tackled the problem of diffraction neither he nor Thomas Young (1773–1829) before him had deployed Huygens' principle. Instead, both Young and Fresnel initially considered interference to take place between rays of light emanating directly from the source and from emission stimulated by the source at the edges of diffracting objects. That way of thinking essentially conformed to a long-standing tradition that only Huygens himself had challenged, one in which the central physical entity in optics was the ray of light.[8] Fresnel broke with that tradition and introduced Huygens' principle when confronted with an empirical conflict that arose in a particular situation. But to deploy the principle Fresnel had to develop a method for decomposing waves from a multiplicity of different loci into ones with common phases but different amplitudes. Such a method had emerged previously in his exploration of the chromatic effects generated by the passage of polarized light through thin, birefringent crystals.

To achieve that result Fresnel assumed a wave of the form $a \sin[2\pi(ft - x/\lambda)]$, with λ the wavelength and f the frequency of the disturbance. If the wave emanating from the source reaches a point *via* two different paths, one of which traverses

Footnote 7 continued
Peter Vickers have used Kirchhoff's theory as a counterexample to primitive realism—what they have termed "naïve optimism"—which contends that any significant novel predictive success can be explained by the truth content of the assumptions that play an essential role in the derivation. Kirchhoff's boundary conditions, Saatsi and Vickers noted, are both mathematically inconsistent and physically untenable. Yet they are essential assumptions for the derivation of Kirchhoff's empirically-successful (within certain regimes) formula (Saatsi and Vickers 2011). We thank Ning Yan Zhu for catching a number of misprints and missing or incorrect references in a previous version of the present article.

[8] On the history of Huygens' optics and its background see Shapiro (1973). On Young see Kipnis (1991), and on Fresnel see Buchwald (1989).

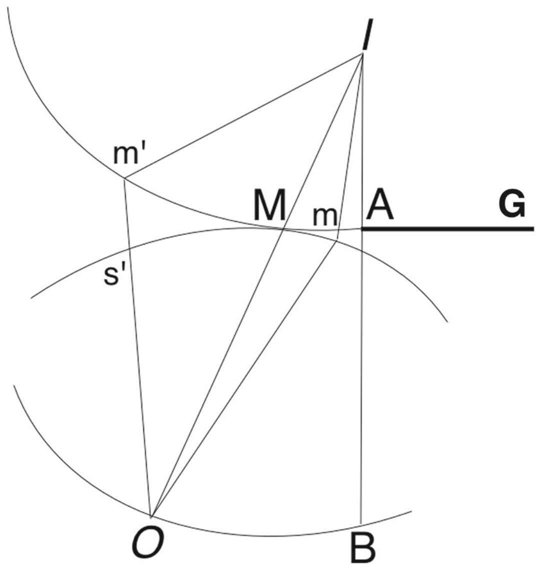

Fig. 1 Fresnel's configuration for diffraction

a distance x while the other requires an additional distance d, then the latter will be $a \sin [2\pi(ft - x/\lambda) - i]$ where i is $2\pi d/\lambda$. Algebraically decomposing this expression, Fresnel could split the second wave into two parts:

$$a \sin [2\pi(ft - x/\lambda) - i] = a \cos(i) \sin [2\pi(ft - x/\lambda)] \\ - a \sin(i) \cos [2\pi(ft - x/\lambda)]$$

That is, a single wave with arbitrary phase i can always be considered to arise from two other waves with amplitudes $a\cos(i)$, $a\sin(i)$ that differ in phase by 90°. This involves the same kind of process, Fresnel noted, as the composition of two mutually-perpendicular forces of magnitudes equal to the component amplitudes. Whence, he concluded, the magnitude of the resultant of two waves with the same frequency that otherwise differ in phase by 90° is just the square root of the sum of the squares of the component amplitudes. This decomposition enabled the application of Huygens' principle to diffraction.

Consider with Fresnel a point light source I, a semi-infinite plane obstacle AG whose edge A is directly beneath I, and a screen parallel to AG. His objective was to calculate the resultant wave at a point O on the screen but outside the shadow cast by AG (Fig. 1). According to Huygens' principle the wave intensity at O should be the sum—or integral for an effectively continuous distribution—of all the wavelets emanating from the front $m'MmA$, calculated as it passes the obstacle AG.

Fresnel assumed that the remainder of the wave was completely blocked by the obstacle, that the obstacle neither modified the wave's spherical shape, nor affected the phase of the vibrations at any unblocked point of the front. Note that in so assuming Fresnel essentially ignored the finite dimensions of such an obstacle in that he did not separately consider what the wave front might be on the side of the obstacle facing the luminous point and on the side facing the screen point. He simply annulled the wave altogether at the obstacle's locus. We shall see below that the introduction of a solution to the scalar wave equation by Kirchhoff mandated explicit consideration of both surfaces.

The resultant at O was, then, the integral of all the wavelets over the semi-infinite front $m'MmA$ whose phases were respectively proportional to the path lengths AO, mO, MO, $m'O$, etc. To carry out this integral, Fresnel had to choose a convenient origin. For various reason he selected M (subsequently termed the 'pole'), the intersection between the direct line IO and the wave front $m'MmA$.[9] Fresnel easily calculated that the distances m′s′ (s′ is the intersection between the line $m'O$ and the arc around the center O) by which the wavelets differ in their distances to the point of observation are to a very good approximation for points reasonably near the pole as $z^2(b+c)/(2bc)$ with b representing IA and c representing AB, where B is the screen point, while z is the distance Mm'. Consequently the phase difference i in Fresnel's original decomposition becomes $\pi z^2(b+c)/bc\lambda$. Each of the wavelets decomposes accordingly, with the consequence that the square of the resultant amplitude (and so the optical intensity) at the observation point must be:

$$\left[\int dz \sin(\pi z^2(b+c)/bc\lambda)\right]^2 + \left[\int dz \cos(\pi z^2(b+c)/bc\lambda)\right]^2$$

Fresnel's successful demonstration that his formulae worked well empirically won him a Paris Academy prize in 1818 despite the presence on the judging committee of no less than Pierre-Simon Laplace (1749–1827) and Siméon Denis Poisson (1781–1840), neither of whom was sympathetic to wave optics. Today, his expression is a particular approximation to wave propagation under a specific condition: it is the second-order expansion of the propagating phase that takes into account the spherical wave front when the distance of interest lies between the quasi-static near-field zone and the plane-wave-like far-field zone.[10] At the time, and for decades thereafter, his expression had a broader and more profound meaning, for it demonstrated that, far from being a curious conception which could at best be used to generate known results in a hitherto-unacceptable theory, Huygens' principle had proven empirical consequences. The result was to entrench the principle as a physically-meaningful foundation over the following decades as the wave theory of light was gradually assimilated, understood and produced altogether novel results. Most of these novelties did not depend on the principle itself since they were for the most part concerned with polarization phenomena. Nevertheless, as a unifying physical conception with a specific mathematical expression Huygens' principle provided a powerful foundation for this new, complex and difficult theory. John Herschel (1792–1871), who published the first comprehensive article on wave optics in 1827, placed particular emphasis on the principle, which he expressed in the following way:

[9] For full details see Buchwald (1989), Chap. 6.

[10] See, for example, Kong (1986), pp. 671–695

> ... conceive the surface of any wave A B C to consist of vibratory molecules, *all in the same phase of their vibrations* (*sic*). Then will the motion of any point X ... be the same, whether it be regarded as arising from the original motion of S [the source], or as the resultant of all the motions propagated to it from all the points of the surface.[11]

The essence of this way of treating the problem can be concisely expressed. The diffracted wave intensity $u(\vec{r}, t)$ at a given location $\vec{r}$ and time t is taken to be

$$u(\vec{r}, t) = \int_{\Omega} \mathrm{d}r' A(\vec{r}, \vec{r}') u_{in}(\vec{r}', t - |\vec{r} - \vec{r}'| / c)$$

The integration takes place over the open parts of a blocking screen, while $u_{in}(\vec{r}', t)$ is the intensity of the wave directly from the light source at a point $\vec{r}'$ on Ω at time t, $A(\vec{r}, \vec{r}')$ is the inclination and any other necessary factors for propagation from $\vec{r}'$ to $\vec{r}$, c is the wave speed, and $|\vec{r} - \vec{r}'| / c$ is the time delay (or, equivalently, phase factor) from $\vec{r}'$ to $\vec{r}$. In this manner of working, which remained common for decades after Fresnel's original work, the solution to the wave equation is presumptively given (because of the known u_{in}) under the assumption that the screen has no other effect than to stop propagation except at its open parts. That is, the source wave is assumed to be completely unaltered at the screen's openings and completely extinguished over its surface.

3 Stokes' "Dynamical Theory of Diffraction"

Neither Fresnel nor those who worked in wave optics for decades afterwards began their analysis of diffraction with a differential equation. In other areas of optics, such as dispersion and birefringence, analysis did proceed in the 1830s by generating solutions to a wave equation, but in those cases boundary issues were irrelevant, posing a very different type of problem. Nevertheless, the physical foundation of wave optics then required the existence of an *ether*, a substance that pervades space and material bodies and that must be governed by the laws of mechanics. In France and Germany for decades the ether was presumed to be constituted of particles governed by Coulomb-like forces of various presumptive intensities, whereas in Britain after *circa* 1830 it was widely treated as an effective continuum governed by variously-assumed constitutive stipulations.[12]

Fresnel himself never deployed a differential equation for optics to any major extent, though he certainly did consider what sorts of forces must obtain between the elements of the ether in order to justify his assumptions concerning the relationships required between wave speed and direction of oscillation in birefringent media. That is, the only circumstances at the time that raised questions of the forces that act on ether ele-

[11] Herschel (1827), Sec. 623.

[12] For the example of dispersion and Augustin Cauchy's (1789–1857) elaborate mathematics see Buchwald (2012). For a broad overview of the period see Buchwald (2013) and Darrigol (2012).

ments were ones in which the wave speed varies with the direction of propagation and of oscillation. Diffraction required nothing of the sort because the empirical circumstances were limited at the time to propagation in isotropic media. There was however one important question that did demand a consideration of the wave equation, namely the form of the inclination factor governing the amplitude of the Huygens wavelets as a function of the angle between the line to the observation point and the normal to the front which the wavelets comprise. Fresnel had argued on rather weak physical grounds that, in the circumstances he was considering, the amplitudes are independent of direction for the hemisphere tangent to the incident front in the direction of propagation, which meant that in his diffraction formulae the inclination factor was simply ignored. The first effort to determine that factor required a consideration of the wave equation, and it was accomplished in 1849 by George Gabriel Stokes (1819–1903).

A senior wrangler and Smith's prizeman at Cambridge University and eventually Lucasian Professor of Mathematics there, Stokes had worked on optics (including the aberration of light and spectroscopy) as well as fluid dynamics, a subject whose mathematical structure resembled that of optics precisely because the optical ether was presumptively governed by mechanical relations.[13] Stokes introduced his "Dynamical Theory" with the following words:

> When light is incident on a small aperture in a screen, the illumination at any point in front of the screen is determined, on the undulatory theory, in the following manner. The incident waves are conceived to be broken up on arriving at the aperture; each element of the aperture is considered as the centre of an elementary disturbance, which diverges spherically in all directions, with an intensity which does not vary rapidly from one direction to another in the neighborhood of the normal to the primary wave; and the disturbance at any point is found by taking the aggregate of the disturbances due to all the secondary waves, the phase of vibration of each being retarded by a quantity corresponding to the distance from its centre to the point where the disturbance is sought.[14]

This much is entirely similar to the standard assumption since Fresnel, and Stokes did not go beyond it. Nevertheless, he penetrated very far into the mathematical core of contemporary wave theory—farther than anyone had since Fresnel—by deriving an entirely new result involving the polarization of diffracted light that he immediately sought to confirm in the laboratory. Stokes' primary purpose however was to uncover the inclination or amplitude factor. To do so he began at once with the general differential equation of motion for an isotropic, inviscid elastic solid that he had himself developed, and that he now applied to the optical ether:[15]

$$\partial^2\vec{u}/\partial t^2 = b^2\nabla^2\vec{u} + (a^2 - b^2)\nabla(\nabla \cdot \vec{u})$$

[13] Stokes (1856); reprinted in Stokes (1883), pp. 243–328.

[14] *Ibid*., 243–244.

[15] Stokes (1845c). Of course neither Stokes nor anyone else, with the partial exception of James Clerk Maxwell (1831–1879), used vector notation until the 1890s. Nevertheless physicists and mathematicians of the period were able almost at once to produce the component equivalents of even complex vector operations, so that modern notation does not unduly alter their original understanding.

where $\vec{u}$ is the displacement, and a, b are elastic constants. He then separated the equation by defining 'for shortness' δ as the negative compression $\nabla \cdot \vec{u}$ (or 'dilatation' as he called it), and $\vec{\omega}$ as the rotation (or, again in Stokes' terminology, the 'distortion') $(1/2)\nabla \times \vec{u}$:

$$\partial^2\delta/\partial t^2 = a^2\nabla^2\delta$$
$$\partial^2\vec{\omega}/\partial t^2 = b^2\nabla^2\vec{\omega}$$

The single equation for the compression, and the three for the components of the rotation, all have precisely the same form, and Stokes could at once write down the following solution, which he obtained from Poisson[16]:

$$U = \frac{t}{4\pi}\int F(at)\mathrm{d}\sigma + \frac{1}{4\pi}\frac{\mathrm{d}}{\mathrm{d}t}\left\{t\int f(at)\mathrm{d}\sigma\right\} \tag{1}$$

Here U is the solution at some point P, t is the time, and $f(at)$, $F(at)$ are respectively the initial values of the function δ(or a component of $\vec{\omega}$) and of its time derivative at all positions whose distance from P is *at* (or *bt*). The integrals, which correspond to the mean values of the functions, are taken over a spherical surface of radius *at* (or *bt)* that surrounds the field point P. Poisson's solution has the peculiarity of representing the effect at P in terms of a time- dependent radius that is drawn from P. Instead, that is, of following a pulse as it expands outwards, with this solution we start at a given point and cut space with surfaces drawn about it until we find surfaces that pass through the regions which contain the initial disturbance. Poisson, as it were, held fixed the initial disturbance and went looking for it from the field point, and this solution (which is difficult to formulate in a rigorous manner[17]) applies to any disturbance that begins at some moment. In contrast to the manner in which the wave equation would be treated decades later, wherein the time-varying part, subjected to a Fourier expansion, is separated from the location-varying part, this solution lumps together space and time, treating the whole as an initial-value problem. Stokes then justified its extension from, in effect, a pulse to infinitely long wave trains in the following way:

> In the investigation it has been supposed that the force [disturbance] began to act at the time 0, before which the fluid was at rest, so that $f(t) = 0$ when t is negative. But it is evident that exactly the same reasoning would have applied had the force begun to act at any past epoch, so that we are not obliged to suppose $f(t)$ equal to zero when t is negative, and we may even suppose $f(t)$ periodic, so as to have finite values from $t = -\infty$ to $t = +\infty$.[18]

16 It is particularly ironic that Stokes took this from Poisson, because he at once used it to argue that the inclination factor varies in a fashion that Poisson himself would probably not have (Buchwald (1989), p. 192).

17 On which see Baker and Copson (1939), pp. 12–15.

18 Stokes was a bit disingenuous here, since not only his investigation, but Poisson's solution, requires the limitation. Stokes' quick attempt to extend the class of allowable functions to cover those which are not temporally delimited requires a great deal more justification than this Stokes (1883), pp. 278.

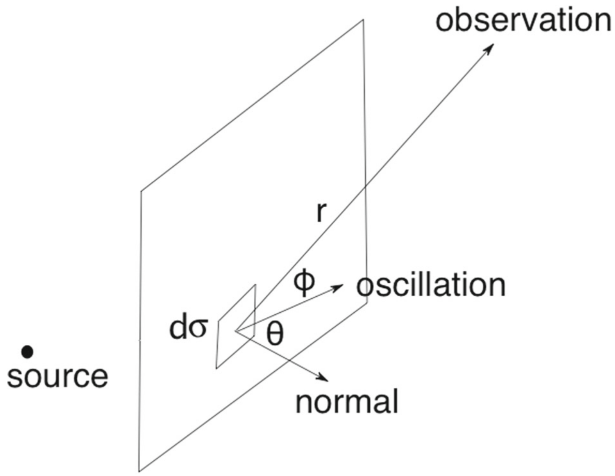

Fig. 2 Stokes' configuration for diffraction

To generate a formula that could be applied to diffraction, Stokes simply assumed the disturbance to be sinusoidal in form. The final result for the wave of distortion produced the following expression for its value at a field point located at a distance r from the disturbance on a surface element $d\sigma$ that forms part of an unblocked region:

$$\frac{d\sigma}{2\lambda r}(1+\cos\theta)\sin\phi\cos\left[\frac{2\pi}{\lambda}(bt-r)\right] \tag{2}$$

Here θ is the angle between the normal at the surface element and the line from there to the field point, while ϕ represents the angle between the direction of the oscillation and that same line: it was introduced to take account of polarization (Fig. 2). This expression constitutes the first of an inclination factor, namely $1+\cos\theta$.[19] Stokes' main interest here was to show that, were the expression integrated over a completely unblocked surface, then the original wave would be regenerated, demonstrating thereby the consistency of an analysis based on Huygens' principle. He did not however go further to amend Fresnel's original formulae by incorporating the inclination factor – which would in any case be pointless given that the configuration of diffraction experiments at the time made the factor inconsequential.

Unlike Fresnel, Stokes did not have to simply assume Huygens' principle because its wavelets now appeared as the integrands of a solution. Still, Stokes had not gone any further than had Fresnel in respect to the conditions that should obtain over a blocking

[19] Stokes set the compression wave to the side. However, the mechanically-necessary existence of both compression and distortion generally posed a problem for such investigations because they cannot easily be divorced from one another, particularly if the model involves, like Cauchy's, forces between particles [on which see Buchwald (1980, 1981)]. Cauchy assumed the compression wave to be invisible (which raised energy issues that might in principle be detectable), while Stokes took the compression constant to be so large that the corresponding wave speed was infinitely larger than the speed of the distortional wave, implying that the former would not be visible or otherwise affect the latter. This amounted to assuming that the ether is incompressible, requiring the rate of displacement $\partial\vec{u}/\partial t$ of any element to satisfy $\nabla\cdot\partial\vec{u}/\partial t=0$.

surface and within its open regions. He had also simply assumed that the wave function must vanish on the surface and remain completely unaffected over its open regions. The consistency that he proved concerned only the question of whether the expression he had generated for the wavelets could reproduce the original disturbance, not whether the presumed values of the wave within the bounded regions and over the blocking surface could themselves be justified.

Questions concerning the boundary conditions and uniqueness requirements that arise when working with partial differential equations had previously been dealt with in respect to the Laplace equation for the potential function ($\nabla^2\varphi = 0$), most notably in England by the mathematician George Green (1793–1841) in 1828.[20] Green developed, among other results, what became known as "Green's theorem" (of which more below), and he did argue that, given the value of the potential over a closed surface, there is one and only one unique solution to the Laplace or Poisson equation within or without the closed surface, provided that the potential vanishes at an infinite distance from it, or what would eventually be termed the Dirichlet principle by Bernhard Riemann (1826–1866).[21]

Now in 1845 and 1846 Stokes published two influential papers concerning the aberration of light that invoked a condition for the ether's velocity with respect to the earth.[22] He was almost certainly not familiar with the work in which Green had deduced his theorem since it had not been published in a journal, and since in any case its title seemed to concern itself with electricity. Only in 1845, when William Thomson (1824–1907, later Lord Kelvin) rediscovered Green's *Essay*, did it become broadly known. Even if Stokes had known Green's theorem, he would not have thought it relevant to the problem of aberration, despite the fact that both Green's *Essay* and Stokes' aberration problem concerned the question of solutions to the Laplace equation under specific boundary conditions.

The question Stokes addressed was to find an appropriate condition on the motion of the ether near the earth's surface that could yield the constant of aberration, i.e. the expression for the deviation of a star's apparent position as a function of the speed of light through otherwise stationary ether and the earth's velocity through the medium. Stokes demonstrated that the appropriate expression would follow provided that two conditions were satisfied: first, that the ether's velocity at the earth had to be the same as the earth's with respect to the ether at a distance, i.e. that the earth must fully drag the ether along, and second that the part of the ether's velocity due to the earth's motion must be irrotational, i.e. that $\nabla \times \vec{v}$ must vanish, in which case that part of the velocity must be the gradient of a scalar function.[23] Since Stokes had also required the divergence of the velocity to vanish in his diffraction theory in order to eliminate the compression wave, this would have meant that the ether's velocity had to satisfy a Laplace equation were it not that Stokes had limited the requirement to a 'part' of

20 Green (1828).

21 After Peter Lejeune Dirichlet, 1805–1859; *cf* Riemann (1857), p. 17.

22 Stokes (1845a, b).

23 Of course optical polarization required the existence of transverse oscillations, i.e. of $\nabla \times \vec{v}$, and so Stokes emphasized that irrotationality could hold only for "that part of the motion of the ether which is due to the motion of translation of the earth and planets" (*ibid*., p. 137).

the velocity, which accordingly essentially avoided having to deal with a boundary-value problem. He was able in this way to ignore altogether the earth's surface by simply assuming in his calculation of the aberration constant that the small additional 'part' of the ether's velocity due to the ether's being entangled with the moving earth was that of the earth itself relative to an ether undisturbed except by the ($\nabla \times \vec{v}$ bearing) motions of light. Nearly half a century later, in 1887, the Dutch physicist H. A. Lorentz (1853–1928) returned to the question and decided that Stokes' analysis had to be treated as a boundary-value problem. Whatever the mechanism of earth-ether entanglement might entail, Lorentz effectively assumed that, since the ether is acknowledged to be incompressible, the Laplace equation must be satisfied, in which case the ether's motion is specified altogether by the scalar function's normal gradient at the earth's surface, i.e. by the velocity, assumed to be that of the earth itself. Lorentz then demonstrated that under these circumstances, the component of the gradient of the requisite scalar function tangent to the surface will differ from the corresponding component of the earth's velocity, so that the ether will slip over the surface, ruining the general applicability of Stokes' calculation.[24]

Stokes in the mid-1840s had paid no attention at all to boundary-value conditions, while Lorentz in 1887 insisted on invoking one immediately by jettisoning Stokes' admittedly-vague separation of earth-plus-ether velocity for that of the ether. Lorentz in other words constrained the problem to be one that simply had to satisfy a specific boundary requirement on the Laplace equation, namely the so-called Neumann condition, according to which specification of the scalar function's normal gradient over a closed surface completely determines the function's value.[25] However, as late as the 1870s physicists at least were not generally paying close attention to the mathematical conditions imposed on harmonic functions by boundary-value requirements, as we shall now see in detail in the case of Kirchhoff.

4 Kirchhoff renovates Huygens' principle

Gustav Robert Kirchhoff was well-known when he left his two-decade-long academic base in Heidelberg and took the first chair for an *Ordinarius* in theoretical physics at the University of Berlin in 1875. His Berlin years from 1875 to his death in 1887 were capstones to a career, since his major contributions—electric circuit theory, spectroscopy, and thermal radiation—had all been made well before, while at Berlin he devoted himself principally to teaching.[26] Precisely because of the need to prepare lectures Kirchhoff looked into the issues in optical diffraction that Stokes had only partially resolved. In 1882, Kirchhoff read a seminal paper on optics at the Royal Prussian Academy of Sciences in Berlin. It was published in the Academy's *Sitzungsberichte* and the following year in the *Annalen der Physik und Chemie*.

[24] Lorentz (1887).

[25] Named after the German mathematician Carl Neumann (1832–1925). For histories of the boundary-value criteria for harmonic functions see Cross (1985); Cheng and Cheng (2005). See Kline (1972), Chap. 28 for an overview of partial differential equations in the period.

[26] Jungnickel and McCormmach (1986), pp. 30–32, Vol. 2.

The aim of Kirchhoff's paper was similar to Stokes' three-decades earlier work but more general: Kirchhoff aimed to develop a "fully satisfactory theory" of light, and not solely of diffraction *per se*, by starting with the wave equation itself.[27] Like Stokes, Kirchhoff assumed that the ether displacement $\vec{u}$ corresponding to light was altogether transverse, setting aside the possible existence of a compression wave (on the widely-accepted premise that the ether is incompressible), yielding thereby the following equation:

$$\frac{\partial^2 \vec{u}}{\partial t^2} = a^2 \nabla^2 \vec{u}$$

Also like Stokes, Kirchhoff used a scalar variable φ to express any of $\vec{u}$'s components, but unlike Stokes he altogether ignored polarization [which Stokes had partially taken into account through the introduction of the variable φ (*cf* 2)]:

$$\frac{\partial^2 \varphi}{\partial t^2} = a^2 \nabla^2 \varphi$$

Kirchhoff's method of solving the equation however differed in fundamental ways from Stokes'. Stokes had treated the situation entirely as an *initial-value problem*, and he had accordingly used Poisson's formulation to express the solution in terms of the subsequent evolution of the distributed sources at time zero (*cf* 1). That solution tacitly assumed that the disturbance vanishes on a blocking surface and is otherwise unaffected, but Stokes did not directly incorporate the conditions into the wave equation's integral solution. By contrast, Kirchhoff explicitly involved the *boundaries* of the illuminated region.

He began with a point source in unbounded space. Assuming a trigonometric oscillation with period T for an outward-propagating spherical wave, Kirchhoff first assumed the following expression for the wave function:

$$\varphi = \frac{D}{r_i} \cos\left[2\pi\left(\frac{r_i}{\lambda} - \frac{t}{T}\right)\right] + \frac{D'}{r_i} \sin\left[2\pi\left(\frac{r_i}{\lambda} - \frac{t}{T}\right)\right]$$

or (following the decomposition that Fresnel had originally developed)

$$\varphi = \sqrt{\left(\frac{D}{r_i}\right)^2 + \left(\frac{D'}{r_i}\right)^2} \sin\left\{\left[2\pi\left(\frac{r_i}{\lambda} - \frac{t}{T}\right)\right] - i\right\}$$

in which r_i is the distance between the source or, in Kirchhoff's word, 'luminous' point and the locus of observation, while D and D' are amplitude coefficients whose ratio determines the wave's phase.[28]

[27] Kirchhoff (1883), *cf* pp. 663. The paper was included in the posthumous publication of Kirchhoff's works: Kirchhoff (1891a), pp. 22–54.

[28] *Ibid*., 665.

Kirchhoff next set his “luminous point” to the side, together with his presumptive expressions for the wave, and turned instead to “Green’s theorem” specifically in order to obtain an expression that “specifies and generalizes the so-called Huygens theorem”—that, in other words, can be interpreted as an expression for the effect of Huygens’ wavelets in open regions of an illuminated space. Kirchhoff’s friend and now colleague at Berlin, Hermann von Helmholtz (1821–1894), had years before shown how to deploy Green’s theorem for the vibrations of sound in open-ended tubes. Moreover, Kirchhoff had already developed the basis of what follows in 1876 in his published lecture on propagation in a compressible fluid.[29]

Consider two scalar functions U and G of x, y, and z *within* a bounded space V whose surface is denoted by s, and whose first and second spatial derivatives are defined and continuous. Let $\mathrm{d}v$ be a differential volume element within the space V, $\mathrm{d}s$ a differential element of the surface s, and N the normal to $\mathrm{d}s$ directed toward the interior of V. Then in Kirchhoff’s formulation:[30]

$$\oiint_{s} \mathrm{d}s \left(U \frac{\partial G}{\partial N} - G \frac{\partial U}{\partial N} \right) = \iiint_{V} \mathrm{d}v \left(G \nabla^2 U - U \nabla^2 G \right)$$

Kirchhoff assumed ‘initially’ that the function G, like φ, satisfied the wave equation. Accordingly, he set U to φ and inserted both it and G into his formulation, replacing the spatial derivatives $\nabla^2 U$ and $\nabla^2 G$ with the time derivatives $\partial^2 U/\partial t^2$ and $\partial^2 G/\partial t^2$ *via* the wave equation. He then integrated over a time interval, requiring however that the lower limit be negative and the upper limit positive:

$$\int_{-t'}^{t''} d\mathrm{t} \oiint_{s} \left(\varphi \frac{\partial G}{\partial N} - G \frac{\partial \varphi}{\partial N} \right) \mathrm{d}s = \frac{1}{a^2} \left[\iiint_{V} \mathrm{d}v \left(G \frac{\partial \varphi}{\partial t} - \varphi \frac{\partial G}{\partial t} \right) \right]_{-t'}^{t''} \tag{3}$$

Kirchhoff then set G to an expression whose numerator bears a limited resemblance to what would decades later become known as Dirac’s delta function:

$$G = \frac{F(r_0 + at)}{r_0}$$

Here r_o is the distance from any point to an arbitrary but fixed point O, with both points lying within the volume of integration in (3). The unusual function F satisfies $F(x) = 0$ for $x \neq 0$, while $\int_d^e F(x) = 1$ where the lower limit d is a *finite* negative number, and the upper limit e is *finite* positive.

This distinguishes Kirchhoff’s F from Dirac’s delta since the latter’s defining integral runs from negative to positive infinity. In his later optical lectures Kirchhoff

[29] Kirchhoff (1876), pp. 314–317. There Kirchhoff had used a different specification for the limits in his time integral with attendant changes in the argument which was however less detailed than it later became, perhaps because Kirchhoff wanted to ensure the analysis would work for an infinite train of disturbances (see below).

[30] Kirchhoff (1883), p. 666.

justified the existence of such a function as a limiting case by way of an example. Suppose $F(x)$ to be $(\mu/\sqrt{\pi})e^{-\mu^2x^2}$ and let the constant μ be so large that $F(x)$ is "vanishingly small for every finite value of x" and so as infinite as μ itself when x is zero. Setting $z = \mu x$, then *even integrating over infinity* the function would equal unity because $(1/\sqrt{\pi}) \int_{-\infty}^{+\infty} e^{-z^2} dz = 1$, in which case Kirchhoff's F converges asymptotically to Dirac's delta.[31]

Kirchhoff's purpose in introducing the strange 'function' F was to generate surface expressions that could be applied to the physical situation he had in mind, namely the propagation of light within a region that is illuminated by one or more point sources that are *external* to the region. Sections of the bounding surfaces can then be assimilated to light-blocking or reflecting obstacles, allowing as well for different propagatory speeds within different parts of the region in order to accommodate refraction. In doing so, Kirchhoff assimilated the point at which the wave function's value is to be calculated to the point at which the distance r_o in the definition of function G vanishes.

This required two steps. First the volume integral on the right-hand side of (3) had to be removed. To do so Kirchhoff imposed a second condition on the lower limit of the time integration, namely that r_o-at' must be "negative and finite." This implies that the distances r_o from O to any point in the region, thereby including points infinitesimally near the bounding surface, are always less than at'. What might that mean in terms of the physics of the situation? Likely this, though Kirchhoff made no comments: the initial optical disturbance must have begun at a point in time sufficiently distant that it had long passed point O. This presumably licensed Kirchhoff to work thereafter with an effectively continuous train of disturbances of arbitrary form.

From the requirement that r_o-at' must be less than zero (and finite) it followed that the 'function' F must be zero at both limits of the integral, namely -t' and t'', since $t''$is positive, and so the entire right-hand side of (3) will indeed vanish, leaving only the surface integral on the left. This was not enough for the equation to be made physically applicable. Kirchhoff's unusual 'function' F did wipe out the volume integral, but G still appeared, and it did not have any apparent physical meaning, having been chosen for the express purpose of using Green's equation, thereby permitting the consideration of surfaces and hence the general behavior of light in the presence of physical boundaries. Moreover, the inclusion of point O involves a singularity since r_o vanishes there. Which is why Kirchhoff decided to cut O altogether out of the volume over whose surfaces the integration takes place.

To that end Kirchhoff surrounded O with an "infinitesimal sphere" which he subtracted from his region. That produced two bounding surfaces: the original one as external boundary, and the spherical surface surrounding O as internal boundary (Fig. 3). The result was to split the surface integral into two parts, one over the external, the other over the internal boundary. With lowercase s denoting the external, and

[31] Kirchhoff (1891b), pp. 24–25. Note that the requirement that the integration limits of F must be "finite positive and negative" is maintained. He had first developed an argument for the existence of F in his 1876 derivation for propagation in a compressible fluid. This addenda to the original specification of the function F was added by the editor, Kurt Hensel (1861–1941), of the *Optik* (Kirchhoff 1891b, p. 267), indicating the existence of disquiet concerning the function. We thank Ning Yan Zhu for noting Hensel's intervention.

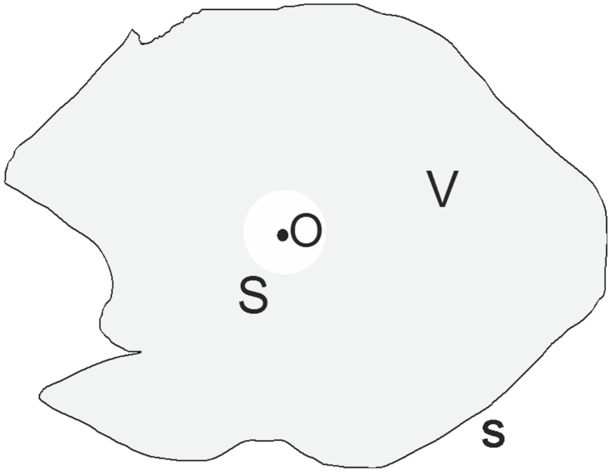

Fig. 3 Kirchhoff's configuration for Green's theorem

uppercase S the internal boundary, Kirchhoff's equation thereby becomes:[32]

$$\int_{-t'}^{t''} \mathrm{d}t \oiint_{s+S} \left(\varphi \frac{\partial G}{\partial N} - G\frac{\partial \varphi}{\partial N}\right) \mathrm{d}(s+S)$$
$$= \int_{-t'}^{t''} \mathrm{d}t \oiint_{s} \left(\varphi \frac{\partial G}{\partial N} - G\frac{\partial \varphi}{\partial N}\right) \mathrm{d}s + \int_{-t'}^{t''} \mathrm{d}t \oiint_{S} \left(\varphi \frac{\partial G}{\partial N} - G\frac{\partial \varphi}{\partial N}\right) \mathrm{d}S = 0 \quad (4)$$

Kirchhoff next moved to eliminate G. First of all, r_o now becomes the radius R of the "infinitesimal sphere," while its surface element dS has the factor R^2. Consequently Gds is proportional to R which, since the sphere is infinitesimal by assumption, means that the second term in the integral over the sphere's surface S can be discarded. Taking however the normal gradient of G at the sphere's surface yields—$(1/R^2)F(R + at)$, which is—$(1/R^2)F(at)$ since Rcan be neglected in the function's argument. Multiplication by dS accordingly leaves just $F(at)$, and so the integral over S simply reduces to $-4\pi\varphi_0$ where φ_0 is the value of the wave function at the locus of the fixed point O at time zero.[33] To eliminate F altogether Kirchhoff recurred to his requirement that its integral over any finite interval must be equal to one, in which case integrating $F(at)$ over time simply produces $1/a$, with (4) thereby reduced to the following:

$$\int_{-t'}^{t''} \mathrm{d}t \oiint_{s} \left(\varphi \frac{\partial G}{\partial N} - G\frac{\partial \phi}{\partial N}\right) \mathrm{d}s = \frac{-4\pi}{a}\varphi_0(0) \quad (5)$$

[32] *Ibid*., pp. 666–668.

[33] At time zero because $F(at)$ is itself non-zero only when its argument vanishes.

Note that the locus of whatever luminous point gives rise to the source wave is not as yet under direct consideration in Kirchhoff's analysis beyond requiring that it must lie outside the bounded region in order to avoid a further singularity.

Both G and a time integral still appear here and had to be removed in order to reach a physically useful result—in order, that is, to calculate the amplitude at a given point in the region bounded by the surface of integration. The region in question must not at this stage contain a source. Again using the integral requirement over the finite interval $\left[-t', t''\right]$ and now shifting the time origin to t, Kirchhoff reduced (5) to the following form, which we shall call his fundamental result:[34]

$$4\pi\varphi_0(t) = \oiint_s \mathrm{d}s \cdot \Omega$$

where

$$\Omega = \frac{\partial}{\partial N}\frac{\varphi(\vec{r}_0, t - r_0/a)}{r_0} - \frac{f(\vec{r}_0, t - r_0/a)}{r_0} \quad \text{and } f = \partial\varphi/\partial N \tag{6}$$

Helmholtz, as Kirchhoff knew, had derived a similar formula in 1859 to characterize the acoustic vibration within a pipe with one open end, albeit in a considerably less general manner. Unlike Kirchhoff, Helmholtz assumed temporal harmonic variation, thereby separating the time- from the space-dependent part of the wave function. Kirchhoff, recall, had not required such a limitation, which is why he had introduced his function F.[35] Note that in Kirchhoff's formulation the loci of whatever luminous points are responsible for the radiation do not appear explicitly. The waves that they engender will appear only by supposition in the boundary conditions that Kirchhoff adopted. Kirchhoff did not, in other words, separately develop Green integrals for the incident and resultant waves and then relate them through his boundary conditions.

Kirchhoff now interpreted his fundamental result explicitly in terms of Huygens' principle: "the motion of the ether [at any point] in the space enclosed in surface s can be regarded as caused by a layer of luminous points on surface s, because each one of the two terms of which Ω is composed may be described as corresponding to a luminous point situated at the location of ds."[36] However, where Huygens had based his principle on a consideration of the physical tendency to expand at each point of

[34] *Ibid*., 668–669.

[35] Helmholtz, "Theorie der Luftschwingungen in Röhren mit offenen Enden," *Journal für die reine und angewandte Mathematik*, 62:1 (1859), 23. Suppose with Helmholtz that $\varphi(\vec{r}, t) = \varphi_c(\vec{r})\cos(\omega t) + \varphi_s(\vec{r})\sin(\omega t)$ so that the time and space dependencies could be separated. Helmholtz's equation may then be written as follows:

$$4\pi\varphi_{c,s}(r_0 = 0) = \oiint_S \mathrm{d}s \cdot \left\{\frac{\partial\varphi_{c,s}}{\partial N} G_{c,s}(r_o) - \varphi_{c,s}\frac{\partial G_{c,s}(r_o)}{\partial N}\right\}$$

with

$$G_c(r) = \frac{\cos(kr)}{r}, \quad G_s(r_0) = \frac{\sin(kr)}{r}$$

[36] *Ibid*., pp. 669–670. The translation is by Anne Hentschel in Hentschel and Zhu (forthcoming).

a wave front, Kirchhoff's version emerged as a result of his application of Green's theorem without any direct consideration of the physics beyond the wave equation proper. Of primary significance, his expression (6) indicated that the contribution from each point on an arbitrary surface s to the ether vibration at a point O that is enclosed by s is determined by both the ether vibration φ and its normal gradient on s. Both values had consequently to be specified.

Kirchhoff recognized that his result could be extended to a case in which the point O is *exterior to* the bounding surface s while the originating luminous points (which are thus far exterior to s) lie *within* the enclosed region, amounting thereby to an inversion of the existing situation. Key to the first step in the derivation of this extension was to redefine the space exterior to s as a volume bounded internally by s and externally by the infinitely large surface S_∞ so that (6) becomes:

$$4\pi\varphi_0(t) = \oiint_s \mathrm{d}s \cdot \Omega + \oiint_{S_\infty} \mathrm{d}S_\infty \cdot \Omega$$

Here the normal to s is turned from inward to outward. Supposing a "particular condition" to be satisfied—namely, that before some finite past moment the values of both functions (φ and f) were everywhere zero—then, provided that the point O is not itself at infinity and choosing a subsequent but otherwise arbitrary moment, both functions will at that time also vanish on the infinitely large surface S_∞.[37] This licensed a principle of inversion, according to which whatever holds for a O point applies as well to a luminous point by exchanging their locations. In such an exchange the locus of evaluation for the wave remains near the Green function singularity.

To further extend the situation, Kirchhoff envisioned two closed surfaces that enclose volumes which have an intersection within which point O lies, while any luminous points I lie outside both regions (Fig. 4). Each of these surfaces then yields φ_0 near O. Subtracting the intersection from the union of the volumes creates a region with a new surface that bounds a region from which both O and any luminous points are excluded (because O now lies within the extracted intersection of the two original volumes). Under theses circumstances the integral in (6) always vanishes. By the principle of inversion the same follows if the original two surfaces enclosed luminous points, with O lying outside both. A similar procedure, but this time forming the composite union of two intersecting surfaces, implies that the integral in (6) also vanishes if a surface *includes* both O and I. In summary Kirchhoff has demonstrated:

1. a region that *includes O* but *excludes I* can produce a non-zero value for the integral in (6). The same holds *vice versa* by inversion.
2. a region that *excludes* both O and I produces a zero value for the integral in (6).
3. a region that *includes* both O and I also produces a zero value for the integral in (6).

[37] Kirchhoff's requirement was in later years subsumed under what became known as the Sommerfeld radiation condition, according to which the limit at infinity of the difference $\partial\varphi/\partial N - 2\pi i\varphi/\lambda$ must vanish: *cf* Goodman (1988), p. 44.

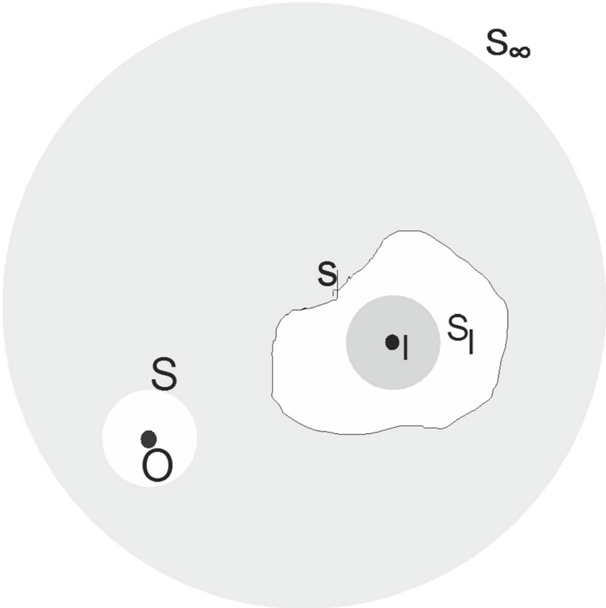

Fig. 4 Kirchhoff's further configuration for Green's theorem

These three correspond respectively to cases in which (1) I and O are separated by a surface, (2) the radiation from I simply passes through the region, and (3) I's radiation reaches O unimpeded.

Before any specific results could be obtained, Kirchhoff had to produce a useful expression for the general integral, $\oiint_s \mathrm{d}s \cdot \Omega$, given the locus of O and the waves engendered by whatever luminous points are present. To do so he assumed a single radiating point I and limited the form of the wave that it produces to a time harmonic expression, namely to $\varphi = (1/r_i)\cos[(r_i/\lambda - t/T)2\pi]$. The distances r_o, r_i that now both appear in the general integral are respectively drawn from the loci O, I to points on the surface of integration. This produced the following form for the integrand Ω:

$$\Omega = \frac{1}{r_i r_o}\left(\frac{1}{r_i}\frac{\partial r_i}{\partial N} - \frac{1}{r_o}\frac{\partial r_o}{\partial N}\right)\cos\left[\left(\frac{r_i + r_o}{\lambda} - \frac{t}{T}\right)2\pi\right] + \frac{2\pi}{r_i r_o \lambda}\left(\frac{\partial r_i}{\partial N} - \frac{\partial r_o}{\partial N}\right)\sin\left[\left(\frac{r_i + r_o}{\lambda} - \frac{t}{T}\right)2\pi\right]$$

Kirchhoff next introduced a specific coordinate system and an approximation linked to it. The origin of the new coordinate system was placed on the surface of integration at a very specific point: namely, where the sum $r_o + r_i$ is a minimum and so is stationary. This will occur if the line joining I to O passes through the surface of integration, in which case the origin will lie at the intersection of the joining line with the integration surface. Kirchhoff's new origin is effectively the same as the one that

Fresnel had introduced decades before, namely the pole. Using this coordinate system, he expanded r_o, r_i as functions of ρ_0, ρ_1, these being the distances from O, I to the pole, but only up to second order. That limitation amounted to the requirement that the wavelength is vanishingly small in comparison to the distances r_o, r_i.[38] This would in contemporary parlance be referred to as the "stationary phase approximation."

For a completely unobstructed wave Kirchhoff could then show that integrating the resultant expression simply reproduces the presumptive wave from the luminous point because the normal gradients that appear in the sine term simply cancel one another, obliterating it, leaving only the cosine term. He was now ready to grapple with the general problem of optics, including retrieving the situation that would reproduce a facsimile of an optics based on bundles of rays. To do so required a crucial physical assumption, known later as "Kirchhoff's *Ansatz*" or as his "physical optics approximation:" namely, that the light wave φ_i (and its gradient $\partial\varphi_i/\partial N$) at any point on the surface of integration is approximately equal to the source wave from I (and its gradient) plus any reflected wave φ_r (and its gradient $\partial\varphi_r/\partial N$), meaning that the presence of any obstructing body does not alter the wave that is incident upon it other than by affecting the values of the wave function and its normal gradient over the body's actual surface:[39]

$$\varphi \cong \varphi_i + \varphi_r, \quad \frac{\partial\varphi}{\partial N} \cong \frac{\partial\varphi_i}{\partial N} + \frac{\partial\varphi_r}{\partial N}.$$

Kirchhoff considered three specific situations: light scattering from a "black body," light diffracted from a black-body screen with an aperture, and diffraction by a grating. For our purposes here, we will consider only his treatment of diffraction. A light source at point i is enclosed by an opaque screen whose surface S is punctured by an aperture A. To simplify the calculation, Kirchhoff transformed the screen from an enclosed surface to an infinite plane, so that the light source i and the point of observation O were on opposite sides of the infinite, flat screen. Here we encounter Kirchhoff's requirements for *both* the wave function at the screen's two surfaces and its normal gradients there. Divide S into a part S_i that faces the light source and a part S_o that is shielded from it. Since the screen extinguishes all light striking it, the wave function must vanish altogether over S_o In addition, the wave over the aperture A should have the same form as the source wave φ_i according to Kirchhoff's *Ansatz*, which leaves the wave near S_i (and so over the aperture) unaltered if the screen is black (there being no reflection). Note Kirchhoff's explicit consideration of both sides of the screen.

However, neither the *Ansatz* nor Kirchhoff's definition of an opaque screen in themselves placed any limitation on the wave's normal gradient over S_o. Kirchhoff

[38] Kirchhoff expressed this as the wavelength being 'infinitesimal' and the sum r_o, r_i being effectively constant throughout the integration since the limitation to a second-order expansion restricts the integration's accuracy to loci within the vicinity of the origin, where the sum is a minimum (Kirchhoff 1883, p. 672).

[39] *Ibid*., pp.683–685.

nevertheless assumed that it too would vanish over the shadowed surface of the screen, producing the following set of boundary conditions:[40]

$$\varphi = \varphi_i, \quad \frac{\partial \varphi}{\partial N} = \frac{\partial \varphi_i}{\partial N} \quad \text{on } A$$
$$\varphi = 0, \quad \frac{\partial \varphi}{\partial N} = 0 \quad \text{on } S_o$$

Using these boundary conditions with a source wave set to $(1/r_i)\cos[(r_i/\lambda - t/T)\,2\pi]$, the integrand Ω produces the following result for the wave at an observation point O under the approximation that the wavelength is vanishingly small in comparison with the distances of both the source and observation points from the aperture (Fig. 5):[41]

$$\varphi_o = \iint\limits_A \mathrm{d}s\Omega = \frac{1}{2\lambda}\iint\limits_A \frac{\mathrm{d}s}{r_i r_o}\left(\frac{\partial r_i}{\partial N} - \frac{\partial r_o}{\partial N}\right)\sin\left[2\pi\left(\frac{r_i + r_o}{\lambda} - \frac{t}{T}\right)\right]$$

The term containing the difference between the normal gradients of r_o, r_i required two further approximations: namely that these distances are much larger than the dimensions of the aperture and so can be considered constant across it outside of the sine term in the integrand, and that the line from I to a point in the aperture forms an "infinitesimal angle" with the line from O to that point. In that case the normal gradients in Ω become equal and opposite since they are simply the cosines of the angles that the lines from I and O make with the normal, while the product $r_o r_i$ is effectively constant, yielding the following expression for the wave at O:

$$\varphi_o = \frac{1}{\lambda r_i r_o}\frac{\partial r_i}{\partial N}\iint\limits_A \mathrm{d}s \sin\left[2\pi\left(\frac{r_i + r_o}{\lambda} - \frac{t}{T}\right)\right]$$

40 Given these boundary conditions, and a special condition, Kirchhoff could also retrieve geometric optics. Require that the sum of the distances from I and O to the surface must not be constant for any finite part of it—meaning in effect that both the observation and luminous points are very far away. If, under this condition, the line joining I to O does not anywhere intersect a body that does not reflect light—a "black body"—then the wave at O remains unaltered. If, on the other hand, that line passes through the black body at least once, then "darkness occurs at the location of O," a true shadow is formed and in consequence "the light from the luminous point propagates rectilinearly in rays that can be regarded as independent of one another" (*ibid.*, pp. 686–687). This retrieves geometric optics for a black-body obstacle and an effectively infinitesimal wavelength. The limitation to constancy over a finite area of the surface for the sum $r_o + r_i$ is dropped for diffraction.

41 *Ibid.*, p. 688. Note that $\partial r_i/\partial N = \cos(\vec{r}_i, \vec{n}_i)$ and $-\partial r_o/\partial N = \cos(\vec{r}_o, \vec{n}_o)$, where $\vec{r}_i$ is the vector from an arbitrary point on A to the point of illumination I, $\vec{r}_o$ is the vector from the same point on A to the point of observation O, $\vec{n}_i$ is the vector normal to A and pointing toward the side of I, $\vec{n}_o$ is the vector normal to A and pointing toward the side of O, and $\cos(\vec{a}, \vec{b})$ is the cosine of the angle between the two vectors $\vec{a}$ and $\vec{b}$. Kirchhoff's diffraction integral may thereby be written as
$\varphi_o = \frac{1}{2\lambda}\iint\limits_A \frac{\mathrm{d}s}{r_i r_o}\left[\cos(\vec{r}_i, \vec{n}_i) + \cos(\vec{r}_o, \vec{n}_o)\right]\sin\left[2\pi\left(\frac{r_i + r_o}{\lambda} - \frac{t}{T}\right)\right]$.

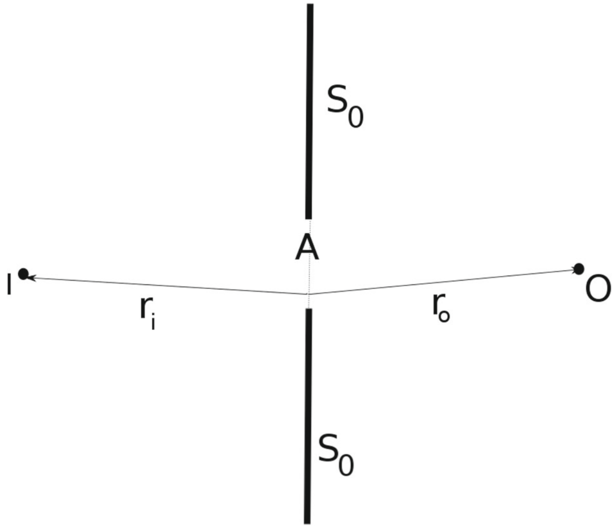

Fig. 5 Kirchhoff's diffraction configuration

Note immediately one difference from Fresnel's original expression: the Huygens wavelets are shifted by a quarter-wavelength in phase from the source wave, which generated discussion in subsequent years.[42]

Kirchhoff added an incident time-harmonic sine wave with a different amplitude for generality and could at once write down an expression for the optical intensity at O. For simplicity we will assume a unit total incident intensity in which case Kirchhoff's diffraction integral yields:[43]

$$\text{intensity at } O = \frac{1}{2\lambda^2 r_i^2 r_o^2}\left(\frac{\partial r_i}{\partial N}\right)^2 (c^2 + s^2)$$
$$\text{where } c = \iint \mathrm{d}s \cos\left(\frac{r_i + r_o}{\lambda}\right) 2\pi \text{ and } s = \iint \mathrm{d}s \sin\left(\frac{r_i + r_o}{\lambda}\right) 2\pi$$

The functions c, s are of course the Fresnel integrals, while the factor in the square of the normal gradient of r_i is Kirchhoff's inclination factor, here obtained for the first time directly from the solution to the wave equation *via* Green's theorem and suitable approximations. Except for this factor multiplying the integrals, the result has the same form as the one that Fresnel had produced decades before for the same situation. Kirchhoff's is more general in two ways: first it includes a factor that depends directly on an inclination factor and inversely as the product of the wavelength by the distances

[42] Fresnel took the wave in Huygens' principle to differ from the source wave solely by virtue of distance to the surface of integration, so that, e.g., a cosine wave remained a cosine wave plus a phase addition to its argument. However, the integrands in Kirchhoff's expression are shifted by a quarter wavelength from the source wave in addition to the phase addition. The shift is a direct consequence of applying Green's theorem to the wave equation.

[43] Kirchhoff (1883), p. 689.

of I, O to the aperture loci, and second Kirchhoff left his result in terms of r_o, r_i instead of further approximating in terms of the vertical distances of I, O to the screen.

5 The Poincaré paradox

Trained in engineering at the École Polytechnique and the École des Mines, Poincaré had always kept an active line of research in mathematical physics and applied mathematics, in spite of his better known works in more abstract areas. In 1886 he assumed the chair of mathematical physics at the University of Paris, where he lectured on optics in the winter semester of 1887–1888, which he again taught in 1891–1892.[44] Both sets were published, and in the 1892 lectures Poincaré reviewed what he then termed "Kirchhoff's hypotheses" for diffraction.[45] If, he began, two functions G, φ are continuous and finite within a given region, then Green's theorem requires:

$$\oiint_{\Sigma} \left(\varphi \frac{\partial G(r)}{\partial N} - G(r) \frac{\partial \varphi}{\partial N} \right) \mathrm{d}\sigma = \iiint_{V} dv \left(G(r)\nabla^2 \varphi - \varphi \nabla^2 G(r) \right)$$

where the surface of integration is denoted by Σ. Suppose next that G has a singularity at the origin O of coordinates but such that the product Gr becomes unity there, with r denoting distance from O. Surround O with an infinitesimal sphere and calculate the integral over its surface, producing (as with Kirchhoff) the value $-4\pi\varphi_o$, which gives the value for φ at the locus O of the singularity for G. Assuming further that both φ and G satisfy the wave equation, then the volume integral vanishes, leaving Poincaré' with (7) for the value of φ_O at a point that lies within an infinitesimal sphere which is itself surrounded by a closed surface Σ:[46]

$$\iint_{\Sigma} \left(\varphi \frac{\partial G(r_P)}{\partial N} - G(r_P) \frac{\partial \varphi}{\partial N} \right) \mathrm{d}\sigma = -\,4\pi\varphi_o \tag{7}$$

If point O is not surrounded by Σ then the integral vanishes.

To apply the theorem to diffraction, Poincaré considered a situation that was similar to Kirchhoff's but that was configured in a different manner. He divided space into four distinct surfaces as follows. Marking an arbitrary point that we shall designate as M, Poincaré denoted the surface of the screen facing M as B and the screen's opposite surface as C. He then described a closed surface S, part of which coincides with B and within which point M lies; A denotes the part of S that excludes B. Surface S is consequently the union $A + B$. Finally, he surrounded M with an infinitesimal sphere s_M that lies within the region surrounded by S. We have added a point P that lies

[44] For a thorough account of Poincaré's life and career see Gray (2013). Also Dieudonné (2008); Charpentier et al. (2010). On anomalous dispersion and its theoretical consequences see Buchwald (1985), chap. 27.

[45] Poincaré had derived formulae akin to Kirchhoff's in the 1887–1888 lectures, though at that time he was unaware of Kirchhoff's theory; he had then also noted that the boundary conditions cannot strictly hold at the same time: Poincaré (1889), pp. 115–116. On Poincaré and light see Darrigol (2015).

[46] Poincaré (1892b), pp. 141–143.

entirely outside the union $C + A$ and have surrounded it, like M, with an infinitesimal sphere s_p. Poincaré's configuration placed the locus of his luminous point at point M within the region surrounded by the surface, which is where Kirchhoff had placed his observation point. The observation point, though Poincaré did not mark it, here lies at a point P, outside $C + A$.

Poincaré introduced a function φ^i to represent the wave at any point emitted by the luminous point M in the absence of a physical screen. When B and C are the surfaces of a physical object, then the wave φ^i is altered by their presence to φ^R according to Kirchhoff's boundary conditions. Poincaré employed the distinction to establish relations between the two functions using the boundary conditions for an opaque body. Kirchhoff had not done anything similar since he had in the end directly inserted the value of the wave from the luminous point into his integral and had applied boundary conditions without having reached the result by means of relations between distinct Green's integrals for φ^i,φ^R.

To clarify Poincaré's somewhat terse discussion we introduce the following convention: the value φ_P at a point P of the integral $\frac{1}{4\pi}\iint\limits_{\sigma_k}\left[G_P\frac{\partial\varphi}{\partial N}-\varphi\frac{\partial G_P}{\partial N}\right]\mathrm{d}\sigma$ of a function φ and a Green's function G_P, the latter of which has a singularity at P,over a set of surfaces σ_k that bound the region enclosing P will be represented by $\iint\limits_{\sigma_k}[\varphi]$.[47] In order to employ Green's theorem for the value of the wave function at P we surround P with an infinitesimal sphere in Kirchhoff's fashion so that we can use Green's theorem to replace the integral over the surface of that sphere with the value of the wave function at P. Then, with reference to Fig. 6, we can with Poincaré build expressions for both φ^i, φ^R by considering different sets of the surfaces. Consider first a surface formed by the union of s_M with the surfaces B, C. A source wave φ_P^i from M that reaches the point Pwithin the region bounded by these surfaces will accordingly be represented by the following expression:

$$\varphi_P^i = \iint\limits_{s_M+B+C}\left[\varphi^i\right] \tag{8}$$

The resultant wave φ_P^R at P can be similarly represented by integrals over these same three surfaces:

$$\varphi_P^R = \iint\limits_{s_M+B+C}\left[\varphi^R\right] \tag{9}$$

However, the latter is affected by Kirchhoff's boundary conditions, for which Poincaré assumed the region surrounded by B and C to be opaque. Accordingly, and assuming Kirchhoff's *Ansatz*, the source wave at surface B, which faces the luminous point at M, is unaffected by the presence of the body, while at surface C Kirchhoff's boundary

[47] Poincaré always presumed an outer surface S_∞ at infinity at which the wave function φ and its normal derivative vanish, with the surfaces σ_k forming inner boundaries. The Green's function at point U is $G_P = \exp(-ikr_P)/r_P$, where r_P is the distance between U and P, $k = 2\pi/\lambda$ and the time dependence has been removed due to the monochromatic assumption.

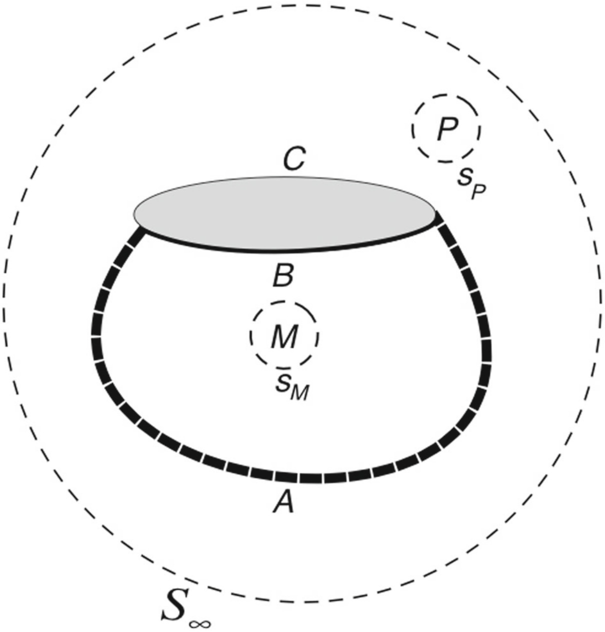

Fig. 6 Poincaré's first surfaces

conditions required the resultant wave and its normal derivatives to vanish. Consequently φ_P^R can be expressed in terms of the source wave and just the two surfaces B, s_M:

$$\varphi_P^R = \iint\limits_{s_M+B} \left[\varphi^i\right] = \iint\limits_{s_M+B+C} \left[\varphi^i\right] - \iint\limits_{C} \left[\varphi^i\right] \tag{10}$$

We next follow Poincaré by introducing a new set of boundaries using the surfaces in Fig. 6, specifically the space bounded internally by the union of A with C (and externally by the surface at infinity). The source wave at any point in this region consequently has the expression:

$$\varphi_P^i = \iint\limits_{A+C} \left[\varphi^i\right] \tag{11}$$

Equating the expressions (8) and (11) for φ_P^i produces (12):

$$\varphi_P^i = \iint\limits_{s_M+B+C} \left[\varphi^i\right] = \iint\limits_{A+C} \left[\varphi^i\right] \tag{12}$$

As a result expression (10) for the diffracted wave φ_P^R becomes (13):

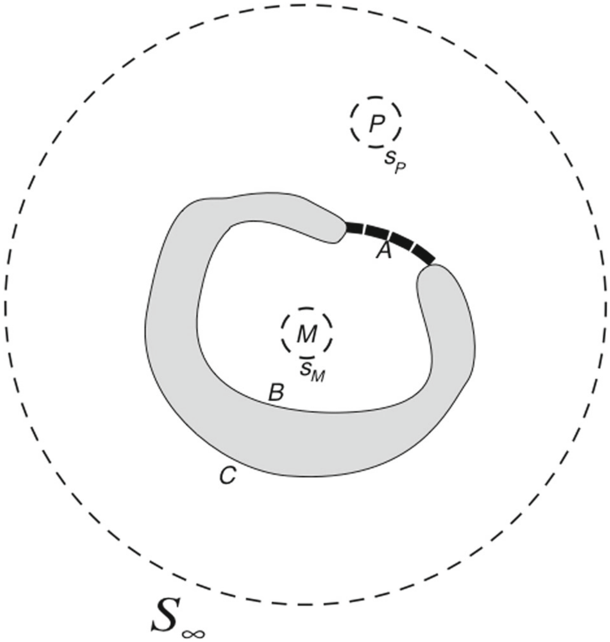

Fig. 7 Poincaré's surfaces adapted to an opaque screen ($B+C$), with an aperture (A), enclosing a luminous point (M) and an observation point (P)

$$\varphi_P^R = \iint_{s_M+B+C} \left[\varphi^i\right] - \iint_C \left[\varphi^i\right] = \iint_{A+C} \left[\varphi^i\right] - \iint_C \left[\varphi^i\right] = \iint_A \left[\varphi^i\right] \tag{13}$$

That is, the diffracted wave at a point P can be found by integrating the wave from the luminous point M over an open surface whose unclosed portion is bridged by the obstacle, with M lying within the region enclosed by the bridged surface and P lying outside it, precisely as Kirchhoff had found.

For clarity, we can redraw as follows to represent the typical case of a screen with an aperture. If we imagine the obstacle bounded by B and C to be an infinite plane screen and the open surface A to be an aperture on the screen, then Poincaré's configuration in Fig. 7 is nothing but the configuration for Kirchhoff's diffraction problem (surface C in that case becomes the screen surface facing away from Poincaré's luminous point M, i.e. C becomes the shadowed surface).

Poincaré's explicit introduction of integrals for the source wave revealed a problem that Kirchhoff would not have seen precisely because he had considered the source wave only towards the end of his calculations—until then his luminous point, though necessarily present, had not directly entered. With Poincaré we now consider a point L located *within* the region $B + C$ bounded by the surfaces of the screen itself. This point is the locus of a Green's function G_L. In doing so we first consider B and C merely as surfaces and not as the boundaries of an opaque screen in order to produce

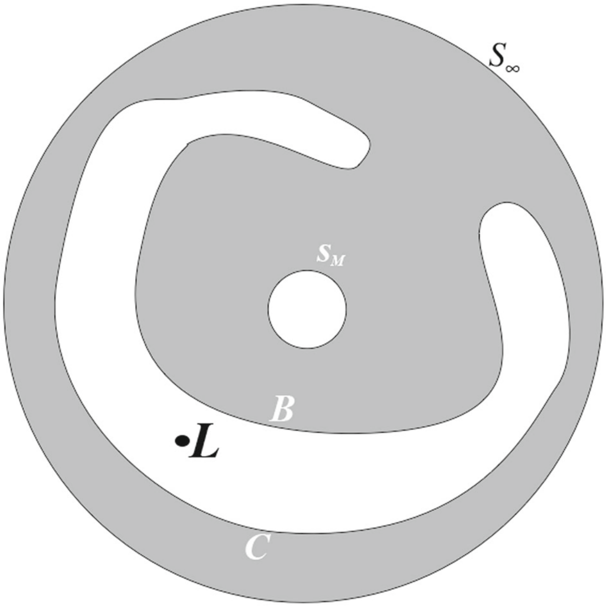

Fig. 8 Poincaré's point L within the region bounded by $B + C$

an expression for the source at L that will lead us to the difficulty that Poincaré discovered.[48]

Recall first that in these calculations a surface S_∞ at infinity is presumed where the wave function and its normal derivatives vanish. We now develop expressions for the values of the source wave and of the total wave at L by integrating *within the region that does not enclose it*—the darkened area in Fig. 8, which is bounded by s_M, S_∞, and the surfaces B, C of the screen itself.

Since point L does not lie within the region of integration, and since there are no singularities for the source wave, the total wave, or the Green's function G_L within that region, then Green's theorem requires that the integral for either the source wave or for the total wave over the boundaries and using the Green's function G_L must vanish (since the values also vanish at infinity by assumption we may as always neglect that boundary):

$$\iint\limits_{s_M+B+C} \left[G_L \frac{\partial \varphi^i}{\partial N} - \varphi^i \frac{\partial G_L}{\partial N} \right] \mathrm{d}\sigma = 0 \tag{14}$$

$$\iint\limits_{s_M+B+C} \left[G_L \frac{\partial \varphi^R}{\partial N} - \varphi^R \frac{\partial G_L}{\partial N} \right] \mathrm{d}\sigma = 0 \tag{15}$$

[48] Poincaré's deduction of the difficulty is extremely terse. What follows draws out the several relations that he developed in somewhat greater detail. Also see Baker and Copson (1939), pp. 70–72.

However—the key element in Poincaré's proof—Kirchhoff's boundary conditions reduce (15) to (16) *since* φ^R*and its normal derivatives should both vanish over C,* while taking the values of φ^i over B and s_M.

$$\iint\limits_{s_M+B}\left[G_L\frac{\partial\varphi^i}{\partial N}-\varphi^i\frac{\partial G_L}{\partial N}\right]\mathrm{d}\sigma=0 \tag{16}$$

Subtracting (16) from (14) accordingly yields (17):

$$\iint\limits_{C}\left[G_L\frac{\partial\varphi^i}{\partial N}-\varphi^i\frac{\partial G_L}{\partial N}\right]\mathrm{d}\sigma=0 \tag{17}$$

And here we find Poincaré's problem. Since φ^i is an arbitrary source wave, and since surface C is itself not only arbitrary but also without physical constraints when considering the source wave, (17) would imply that such a wave would have to vanish "whatever the form of the screen," i.e. even in its physical absence. Waves could simply not exist at all. Something was clearly wrong with the boundary condition over the shadowed surface of the screen, since it is that condition which produces (17). Poincaré did not also deduce through an example that the wave which results from Kirchhoff's boundary conditions cannot reproduce the very conditions over the aperture proper that led to it (*cf* footnote 81 below). He had instead found that waves could simply not occur at all given Kirchhoff's conditions.

In working with Green's theorem earlier in his lectures, Poincaré had remarked that "in general, it will not be possible arbitrarily to assign to one a system of values, whether of φ or $d\varphi/dN$, because these two functions are linked by a relation expressing that the integral is null in an exterior point."[49] This is perhaps what led him to investigate what occurs when the integrations are carried out for "an exterior point" in Green's theorem, leading to the contradiction between the existence of waves and Kirchhoff's boundary conditions that he uncovered. Poincaré's remark concerning overdetermination would certainly have been clear to most mathematicians and physicists in his day. However, that fact alone does not lead to his result, which requires considerably more: the problem arises because the total wave (and its normal derivative) is presumed to be the same as the source wave over the entire visible part of the obstacle (surface B in Fig. 8. Poincaré's point L within the region bounded by B+C), for that was why φ^i appeared in Poincaré's several expressions for φ^T, but that it must vanish on the shadowed surface (C).

In 1897 Lord Rayleigh in England took an entirely different tack. He did not mention the Poincaré paradox at all, but instead developed two alternative expressions for the diffracted wave. Neither of the two solutions suffered in the manner that Poincaré had pointed out because, unlike Kirchhoff's expression, Rayleigh's alternatives did not impose simultaneous requirements on both the wave and its normal derivative—which Poincaré had himself pointed to as a possible source of the problem, though

[49] Poincaré (1889), pp. 144–145.

we have seen that it required both that *and* a discontinuity of the wave and its normal derivative in crossing from the unshielded to the shielded region.

6 The Rayleigh-Sommerfeld alternatives

By the mid-twentieth century, the inconsistency of Kirchhoff's theory seemed to many to be an obvious consequence of the well-known boundary conditions for partial differential equations. As E.W. Marchand (1914–1999) at Eastman Kodak's Research Laboratory and Emil Wolf (1922-) at the University of Rochester (the latter of whom had studied under Max Born) remarked in 1966:

> Since the time-independent wave equation for U is elliptic, the specification of U or its normal derivative $\partial U/\partial n$ in the plane of the aperture (together with the specification of the asymptotic behavior of the solution at infinity in the appropriate half-space) is sufficient to specify the solution uniquely. The inconsistency in Kirchhoff's theory arises from the specification of both U and $\partial U/\partial n$ and has the consequence that, as the point of observation P approaches the plane of the aperture, the Kirchhoff solution $U_K(P)$ does not recover the assumed boundary conditions.[50]

It had for decades been common knowledge that solutions to hyperbolic or elliptic partial differential equations are fully determined by either, but not both, of the following two conditions: (1) the Dirichlet, which specifies the value of the function on a given boundary, or (2) the Neumann, which specifies the value of the function's normal derivative there.[51] From this perspective, of which we have just seen that Poincaré was quite aware, Kirchhoff's theory must be problematic. Yet the result seemed to work quite well empirically given the approximations that Kirchhoff had made. Poincaré accordingly thought that, though the two requirements are not "rigorously compatible, they are at least so in an approximate manner, when one neglects quantities of the order of a wavelength."[52] He did not however provide any clear justification for such a claim.

In 1897 Rayleigh, the leading expert on the theory of sound (having published the first edition of his magisterial treatise on the subject in 1877) detailed a theory that avoided overspecifying the boundary conditions.[53] Rayleigh examined two situations: in one (φ^I) the normal gradient of the wave vanishes over the shadowed surface of the screen, while in the other (φ^{II}) the wave itself does. In both cases the wave and

[50] Marchand and Wolf (1966), p. 1712.

[51] Hyperbolic equations, of which the time-dependent wave equation is one, have the form $\partial^2\varphi/\partial x^2 - \left(1/c^2\right)\partial^2\varphi/\partial t^2 = f(x,t)$ while elliptic equations satisfy $\partial^2\varphi/\partial x^2 + \partial^2\varphi/\partial y^2 = f(x,y)$—Laplace's, Poisson's and the time-independent wave equation developed by Helmholtz are all elliptic.

[52] Poincaré (1892b), p. 188.

[53] Rayleigh (1897). The details of Rayleigh's analysis depended upon results he had developed in his *Theory of Sound*, whose second edition of the first volume had appeared in 1894 and, for the relevant parts of the second volume, the year before.

its normal gradient must be continuous across the aperture. This produces a duo of possible solutions as follows, wherein S denotes the surface of the screen including any aperture and A denotes an aperture:[54]

$$\varphi^{I} = \frac{-1}{4\pi}\left[\iint\limits_{S-A} \mathrm{d}s \left\{\varphi\frac{\partial G}{\partial N}\right\} + \iint\limits_{A} \mathrm{d}s \left\{\varphi\frac{\partial G}{\partial N} - G\frac{\partial \varphi}{\partial N}\right\}\right]$$

$$\varphi^{II} = \frac{1}{4\pi}\left[\iint\limits_{S-A} \mathrm{d}s \left\{G\frac{\partial \varphi}{\partial N}\right\} + \iint\limits_{A} \mathrm{d}s \left\{\varphi\frac{\partial G}{\partial N} - G\frac{\partial \varphi}{\partial N}\right\}\right]$$

This division into two situations raised the problem of how to find distinct expressions for φ in each case. The issue did not arise with Kirchhoff's integral across the aperture because he dropped the integral across the screen proper. Rayleigh solved the problem by considering two cases. For φ^{I} he envisioned a *perfectly reflecting* plane screen on which a plane wave parallel to it is incident. He then offered expressions χ_m, χ_p respectively for what such a wave would be on either side of the screen when it is *unperforated*, so that for a screen S in the yz plane $\chi_m = e^{-ikx} + e^{ikx}$ and $\chi_p = 0$. To these he added, again respectively, $\psi_m = \iint\limits_{A} \Psi_m \frac{e^{-ikr}}{r} dS$, $\psi_p = \iint\limits_{A} \Psi_p \frac{e^{-ikr}}{r} dS$, integrating solely over an aperture A, that are presumed to modify these solutions to take account of the perforation. The functions Ψ_m, Ψ_p must satisfy the reduced wave equation.

For continuity conditions in φ^{I}, again, Rayleigh needed the normal gradient of the wave over the shadowed surface of the screen to vanish, but for the wave as well as its normal gradient to be continuous across the aperture, i.e. at the screen, where $x = 0$ Rayleigh required in φ^{I} that:

$$\text{on } S - A : \frac{\partial(\chi_p + \psi_p)}{\partial N} = 0$$

$$\text{on } A : 2 + \psi_m = \psi_p \text{ and } \frac{\partial(\chi_m + \psi_m)}{\partial N} = \frac{\partial(\chi_p + \psi_p)}{\partial N}$$

Since χ_p (and so its normal gradient) is required to vanish over the entire surface S, and the normal gradient of χ_m also vanishes over S (where $x = 0$), these requirements can be satisfied provided that Ψ_m,Ψ_p are equal and opposite on A, with the wave functions ψ_m,ψ_p then being equal and opposite by reflection of the one across the screen to produce the other, with the additional stipulation that the wave function ψ_m must have the value -1 on the aperture proper. The problem then consists, as Rayleigh put it, "in so determining Ψ_m that this shall be the case." Conversely, in the case of φ^{II}, where the wave is presumed to vanish over the shadowed surface $S - A$, Rayleigh altered χ_m to the difference $e^{-kx} - e^{kx}$ instead of the sum, so that both χ_m and χ_p vanish everywhere at the screen. In this situation ψ_m, ψ_p are, respectively,

[54] Rayleigh (1896), secs. 277, 278 and 292, the last of which Rayleigh referred to in his paper on optical diffraction, Rayleigh 1897.

$\iint\limits_A \Psi_m \frac{\partial(e^{-ikr}/r)}{\partial N} \mathrm{d}S$, $\iint\limits_A \Psi_p \frac{\partial(e^{-ikr}/r)}{\partial N} \mathrm{d}S$. To satisfy continuity over the aperture then required Ψ_m, Ψ_p to be equal and opposite there, where Ψ_m takes a value such that $\partial\psi_m/\partial N$ becomes ik .

Note, then, that Rayleigh's duo each in its specific manner only partially satisfy Kirchhoff's boundary conditions over the body $S-A$ of the screen: in the case of φ^I the wave function over $S-A$ must vanish, whereas its normal gradient need not, whereas in the case of ϕ^{II} the reverse holds. The problems then require judicious choices of the wave functions and appropriate conditions at the aperture using the same Green's function $\frac{e^{-ikr}}{r}$. Rayleigh's aim had not been to avoid the inconsistency that plagued Kirchhoff's solution, but rather to obtain expressions that followed directly from the assumption of one of the two different boundary conditions on $S-A$. Kirchhoff's theory of course required that both be satisfied at once. Because Rayleigh had not fundamentally altered the physical presumptions that underpin Kirchhoff's (or Poincaré's) analysis, *viz* that the screen can annul the wave or its normal gradient, and that the wave in the aperture remains unaffected, his two possible integrals were distinguished solely by the *a priori* imposition of different conditions. That is, Rayleigh's theory, like Kirchhoff's, remained tied to imposed requirements without investigating what might underpin such stipulations on the grounds, just then becoming prevalent, of electromagnetic optics.

In the early 1890s the young German physicist Arnold Sommerfeld (1868–1951) developed a theory for diffraction that avoided Green's theorem by recurring to electromagnetic relations.[55] He was at the time assistant to the mathematician Felix Klein (1849–1925) at the University of Göttingen. Like Kirchhoff and Poincaré, Sommerfeld was a dedicated teacher, and as a professor of theoretical physics at Munich from 1906 he trained an imposing array of students, including Wolfgang Pauli, Rudolf Peierls, Alfred Landé, Linus Pauling, I. I. Rabi and Max von Laue (the latter three as post-graduates.[56] In 1894 Sommerfeld, referring to Poincaré, had remarked in an article that Kirchhoff's boundary conditions are 'inadmissible' and would entail that wave function must vanish everywhere, though the theory nevertheless yields good agreement with observation (so far as was then known).[57] Two years later Sommerfeld published an intricate alternative that avoided the inconsistency but that was limited to the two-dimensional case of an infinitely thin, semi-infinite plane barrier with infinite electric conductivity struck by a linearly-polarized, plane wave parallel to it. To do so, Sommerfeld calculated the scattering of such a wave under the usual electromagnetic condition that the electric field vector in the plane of the screen vanishes, resulting after extensive, and intricate, calculation in an expression for the scattered wave.[58]

55 Sommerfeld (1896); translated as Sommerfeld (2004).

56 For a comprehensive examination of the Sommerfeld school at Munich, and in particular its concentration on the solution of specific problems, see Seth (2010).

57 Sommerfeld (1895), pp. 341–342.

58 In 1892 Poincaré had himself considered the case of diffraction from a sharp metallic edge at large diffraction angles in an effort to account for experimental results that Gouy had obtained (Gouy (1886)). To do so he limited his analysis to infinite conductivity and examined the two-dimensional case by considering the wave from the edge produced by the scattering of a converging cylindrical disturbance whose axis parallels the edge (Poincaré (1892a); on this and Gouy's experiments see Darrigol (2015), pp. 14–16.)

The resulting expression indicated that a modified form of Young's original conviction that the illuminated edge of a diffractor could be treated as though it emitted waves. This result, however limited, constituted the foundation for a rigorous theory of diffraction, and to reach it Sommerfeld deployed a form of imaging in the complex plane in order to maintain the boundary condition on the electric field. Although the theory proved influential over the years, the intricacy of the calculations continued to limit its application.[59] The use of an image procedure did nevertheless also lead Sommerfeld in his optical lectures to deploy images using Green's theorem proper. His approach there invoked the same duo of integrals that Rayleigh had used in 1897 albeit derived, unlike Rayleigh's, which worked with the form of the wave function, from a different specification of Green's function.[60]

Sommerfeld noted that if the Green's function $G(r)$ vanishes on S, then the term containing the normal derivative of the wave function will also vanish there:

$$\varphi(P) = \frac{1}{4\pi} \iint\limits_S \mathrm{d}s \left\{ \frac{\partial \varphi}{\partial N} G(r) - \varphi \frac{\partial G(r)}{\partial N} \right\}$$

$$G(r) = 0 \text{ over } S \Rightarrow \varphi(P) = \frac{-1}{4\pi} \iint\limits_S \mathrm{d}s \left\{ \varphi \frac{\partial G(r)}{\partial N} \right\}$$

In that case only the values of the wave function and of the Green function's normal derivative need to be specified. Simple in principle, but difficult to apply in practice since there is no general Green's function that could satisfy such a requirement. Accordingly Sommerfeld had to limit his theory to a situation in which the Green's function could be so specified, and the only one that worked required the screen to be a plane.[61]

This can be done in the following way. Consider with Sommerfeld an infinite screen located at the origin and parallel to the xy plane (Fig. 9). Place point P at (P_x, P_y, P_z) and consider an arbitrary point Q whose coordinates are (Q_x, Q_y, Q_z). Now reflect P across the screen to form its image P' whose coordinates will therefore be $(P_x, P_y, \text{-}P_z)$. Then the respective distances r, r' from P, P' to Q will be:

Footnote 58 continued
Sommerfeld referred to Poincaré's results for support since he had obtained a similar final expression under the same approximation (far field close to the edge of the geometric shadow): Sommerfeld 1896, p. 374.

59 Sommerfeld produced a version of his theory in his lectures on optics: Sommerfeld (1954) present the lectures that he gave in 1934. Secs. 38–39 provide his "mathematically rigorous solution" for the infinitely-thin, semi-infinite screen. See Born and Wolf (2002); chap. 11 develops the Sommerfeld theory up to the early 1950s. For subsequent developments see Babich et al. (2007).

60 Sommerfeld (1954), pp. 195–201. Sommerfeld had likely been giving these lectures for decades in some form, so that his version of the Rayleigh alternatives probably date to his early years at Munich, so a decade or more after Rayleigh's work on the subject. Sommerfeld did not mention Rayleigh in the published lectures, perhaps because his version of the alternatives involved a considerably different Green's function, as we shall see.

61 *Ibid*., 198–200.

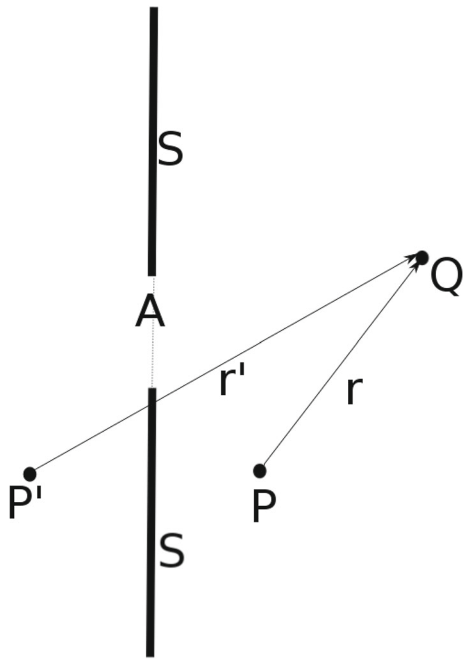

Fig. 9 Sommerfeld's configuration for his Green's function

$$r = \sqrt{(Q_x - P_x)^2 + (Q_y - P_y)^2 + (Q_z - P_z)^2}$$
$$r' = \sqrt{(Q_x - P_x)^2 + (Q_y - P_y)^2 + (Q_z + P_z)^2}$$

Now define a new Green's function as follows:

$$G(Q) \equiv \frac{\exp(ikr)}{r} - \frac{\exp(ikr')}{r'}$$

The new function is, as it were, formed by envisioning the *image* of a Green's function at point P across the screen and then summing the image with its progenitor. This new expression meets all the requirements for a Green's function, for it is a solution of Helmholtz's equation, converges to the appropriate form as r, r' reach zero, and adds nothing at infinity. We can calculate the derivative of G in the direction of the z-axis, *viz.* in the direction normal to the screen itself. Having done so, we move Q to the screen itself, where the result is that G vanishes. Over the shadowed part of the screen the wave function φ must vanish, while it must be continuous across the aperture. Sommerfeld assumed a point source and so expressed the emitted wave as $\frac{e^{ikr}}{r}$ where r is the distance from the source to a point of the aperture. For the alternative case in which the normal derivative of the Green's function must vanish, Sommerfeld merely had to form the *sum* instead of the difference, *viz.* $G(Q) \equiv \frac{\exp(ikr)}{r} + \frac{\exp(ikr')}{r'}$, since the normal is opposite in direction on either side of the aperture.

Sommerfeld's route to a duo of mathematically consistent solutions differed considerably from Rayleigh's. Sommerfeld had altered the Green's function by forming a difference or sum between the original function and its image across the screen but had retained a spherical source wave. Rayleigh had instead retained the single-term Green's function but had required a plane source wave. Both proceeded to apply their expressions to the standard configuration in which, namely, the distance from the aperture to the field point is much greater than a wavelength. Moreover, though neither noted the point, the Kirchhoff integral is easily shown to be the *mean* of the Sommerfeld duo, which are limited to a plane screen, with Rayleigh additionally requiring the incident wave to parallel the screen.[62] Kirchhoff's full integral had no such limitation, but then it suffered from the Poincaré paradox.

$$\varphi^{I}(r) = \frac{1}{2\pi}\iint\limits_{A} \varphi_{\mathrm{inc}} \frac{\partial}{\partial N}\left(\frac{\exp(-ikr)}{r}\right) \mathrm{d}s$$

$$\varphi^{II}(r) = \frac{-1}{2\pi}\iint\limits_{A} \frac{\exp(-ikr)}{r}\frac{\partial \varphi_{\mathrm{inc}}}{\partial N} \mathrm{d}s$$

$$\text{Kirchhoff: } \frac{1}{2}\left[\frac{1}{2\pi}\iint\limits_{A} \varphi_{\mathrm{inc}} \frac{\partial}{\partial N}\left(\frac{\exp(-ikr)}{r}\right) \mathrm{d}s - \frac{1}{2\pi}\iint\limits_{A} \frac{\exp(-ikr)}{r}\frac{\partial \varphi_{\mathrm{inc}}}{\partial N} \mathrm{d}s\right]$$

Neither of the two Rayleigh-Sommerfeld expressions yields the Fresnel integrals—only their sum does. Consequently evaluation of either requires novel series expansions. In 1913, for example, Rayleigh applied ϕ^{I} to the case of an infinite narrow slit, which required extensive numerical computation *via* a series for the resulting integral.[63]

In his rigorous diffraction theory of 1896 Sommerfeld had concluded with a few remarks concerning the results of experiment that are worth quoting in full in order to gain a purchase on the empirical issues of the period. Since Poincaré in 1892 had obtained the same final results under the same circumstances, he concluded as follows:

> The question of the experimental confirmation of our [1896] theory can be here settled in a word. On the one hand Kirchhoff's formulas have been often proven by observations at a small diffraction angle; on the other hand Mr. Poincaré compares his formulas with Gouy's observations under large diffraction angles and finds them to be essentially confirmed. *In the same circumstances, that is for small as for large diffraction angle our theory therefore is also confirmed by experiment* [*sic*].[64]

Consequently Kirchhoff's full integral could account for experiments at small diffraction angles, but not at the large ones in Gouy's experiments, which latter could

62 For φ^{I} the normal derivative of the Sommerfeld difference Green's function reduces to twice the value of the positive term, while for φ^{II} the Sommerfeld function is just twice the value of either term.

63 Rayleigh (1913).

64 Sommerfeld (1896), p. 374.

be handled by Poincaré's 1892 analysis based on electromagnetic relations (see note 58 above). Sommerfeld's scattering theory, which avoided Green theorem methods, could handle both.

In his optical lectures Sommerfeld did not directly address empirical issues, but he did point out that any theory involving the imposition of boundary conditions on an expression derived from Green's theorem (and so leading to an interpretation in terms of Huygens' principle) was questionable, including the Rayleigh-Sommerfeld duo:

> The question remains, are these assumptions also physically justifiable? The answer again is that *they are only approximations for sufficiently small wavelengths*. The field does not vanish completely behind the screen, nor is the field in the aperture entirely unaffected by the presence of the screen, at least not within distances of the order of magnitude of a wavelength from the edge of the screen. The introduction of the Green's function therefore involves no final justification of the method.[65]

Sommerfeld accordingly remained unconvinced that the boundary conditions associated with any such form of diffraction theory were physically reasonable, whereas the approach that he had developed in the early 1890s avoided the problem, at least to a certain extent. Nevertheless, half a decade after the appearance of Kirchhoff's paper an alternative expression based on his theory was produced that, we shall see, would lead decades later to a reformulation that could avoid the Poincaré paradox—but only at considerable mathematical cost.

7 Kirchhoff's integral transformed

A year after Kirchhoff's death in the fall of 1887, and before the appearance of the Poincaré paradox, an Italian mathematician modified Kirchhoff's theory by converting it from a surface to a line integral, a transformation that implicated a different way to work with Kirchhoff's boundary conditions. A native of Milan, Italy, Gian Antonio Maggi (1856–1937) attended Kirchhoff's lectures in Berlin. In 1886 he had become an ordinary professor in analysis at the University of Messina.[66] Two years later Maggi published a transformation of Kirchhoff's fundamental integral that avoided the use of Kirchhoff's function F with its peculiar properties. Of his time in Berlin Maggi wrote in 1914 that he "was fortunate, not long before [writing his 1888 paper] to follow Kirchhoff's lessons at the University of Berlin, and so I was in a privileged position to pay attention to that supremely important result [Kirchhoff's diffraction theory], which the present state of the theory of the electromagnetic field has recently enriched with new applications."[67]

65 Sommerfeld (1954), p. 200. The italics are Sommerfeld's.

66 Maggi (1888). On Maggi see Cisotti (1938), Anonymous, 2015. The reminiscence is from Maggi (1914). Maggi wrote the latter paper in order to compensate the difficulty of his 1888 presentation which, he wrote in 1914, "had somewhat harmed its perspicuity."

67 Maggi (1914). Since Kirchhoff died in 1887, Maggi's "not long before" means several years prior to 1888.

Kirchhoff was known for his strict adherence to precision and rigor. For example, the young Heinrich Hertz (1857–1894)—student of Kirchhoff's colleague Helmholtz—had sent him a paper on elastic collisions that Kirchhoff had extensively edited because Hertz had not been sufficiently explicit in laying out the precise conditions of the problem.[68] Perhaps Kirchhoff had expressed to Maggi directly or in his lectures some doubts concerning the propriety of his function F, or it may be that Maggi noted its presence and decided to avoid it in order to purify Kirchhoff's integral of one possible objection.[69] To do so Maggi deployed three points: two of the points are fixed, while the third is not. The wave equation's solutions V have the following usual form, with r representing the distance between a fixed point x_0, y_0, z_0 and a point x, y, z within the enclosed volume:

$$\frac{\partial^2 V}{\partial t^2} = a^2 \nabla^2 V \quad \text{with } r = \sqrt{(x_0 - x)^2 + (y_0 - y)^2 + (z_0 - z)^2}$$
$$V = \frac{\varphi(t - r/a)}{r}$$

Maggi then noted that the function can contain, separately, the variable coordinates x, y, z and still satisfy the wave equation, so that V becomes $V(t - r/a, x, y, z)$. By considering the two fixed points, one of whose coordinates (x_0, y_0, z_0) appear in the expression for r whereas the coordinates of the other does not, Maggi was able to obtain Kirchhoff's fundamental integral without using the function F while nevertheless leaving the form of the wave open. The reasoning was sufficiently obscure that he attempted a simplification in 1914 (see above, note 66).

This was not Maggi's only innovation. In addition, he effected a significant transformation of the fundamental integral. To do so he used the line joining his two fixed points to specify the edge of a plane that could be used in a complicated manner to create a surface that separated space into two contiguous regions within which the analysis took different forms. These two points represented the loci of the source and of the point at which the wave is observed past a surface that represents a screen, and the regions in question correspond respectively to the space defined by the geometric shadow and to the space outside it. Kirchhoff's boundary conditions were separately applied to these two regions. The result of the procedure yielded a complicated function (due to the choice of coordinates) for the value of the wave that involved in both regions a line integral around the edge of an aperture. In the geometrically-illuminated region the disturbance is determined by the sum of a direct wave from the source added to the line integral, while in the geometric shadow only the line integral holds. This accordingly implicated a *discontinuity* across the shadow boundary, though Maggi's use of prolate coordinates rather obscures the configuration. The transformation to a line integral was effected by Maggi's use of Stokes' theorem based on the assumption that the medium is incompressible, as was usual. If, then, $\vec{u}$ represents a disturbance in

68 On which see Buchwald (1994), chap. 8.

69 The fact that the addition by Hensel in the published lectures a few years later writes of F that such a function "actually" exists may indicate doubt at the time concerning the propriety of basing such a fundamental result upon it.

such a medium, $\nabla \cdot \vec{u}$ must vanish, in which case $\vec{u}$ can be represented by $\nabla \times \vec{v}$ with $\vec{v}$ an appropriate function. Consequently Stokes' theorem affords the transformation of a surface integral for $\vec{u}$ into a line integral for $\vec{v}$:

$$\iint_{\sigma} \vec{u} \cdot \mathrm{d}\sigma = \iint_{\sigma} (\nabla \times \vec{v}) \cdot \mathrm{d}\sigma = \oint_{\partial\sigma} \vec{v} \cdot d\vec{l}$$

In a brief consideration at the end of his article, Maggi concluded that, in the limit of vanishingly small wavelength, the disturbance occurs entirely outside the region of the geometric shadow. The signal purpose of his transformation to a line integral, then, was to yield that result, which in effect retrieved geometric optics. Yet it seems that very few people paid attention to this latter result, perhaps because it was obscurely framed in terms of prolate coordinates. Which makes it hardly surprising that the first to re-achieve such a transformation did not mention Maggi at all.

Wojciech (Adalbert) Rubinowicz (1889–1970) was born in Bukovina to Polish parents, and received his Ph.D. in physics at the University of Czernowitz in 1916. He planned to stay there as a post-graduate assistant, but the university then closed during the First World War. He obtained a temporary position at the University of Munich's institute of theoretical physics, where he began work as assistant to Arnold Sommerfeld in 1916.[70] A year later, Rubinowicz published an article in the *Annalen der Physik* that explored a line-integral expression for Kirchhoff's fundamental integral.[71] He was stimulated to do so by Sommerfeld's 1894 analysis of diffraction based on electromagnetic theory (on which see above), as Rubinowicz aimed to see whether a similar result could be obtained directly from Kirchhoff's integral by means of a line-integral transformation. Although Rubinowicz's work became much better known among European physicists and mathematicians than Maggi's, the conversion that both effected, though in considerably different ways, would later become known as the "Maggi-Rubinowicz transformation."[72]

Rubinowicz did not maintain Kirchhoff's original formulation that allowed for an arbitrary wave but instead used the reduced Helmholtz equation and so presumed that the time and space variables could be separated. Consequently, unlike Maggi he did not have to concern himself with Kirchhoff's delta-like function. Rubinowicz first introduced the general surface integral over the boundary of a region G with a Green's function e^{ikr}/r, with $\bar{u}$ representing the value of the wave function over the region's boundary:

$$\frac{1}{4\pi} \iint_{G} \mathrm{d}s \left\{ \bar{u} \frac{\partial (e^{ikr}/r)}{\partial n} - \frac{e^{ikr}}{r} \frac{\partial \bar{u}}{\partial n} \right\} \tag{18}$$

70 Eckert (2013), p. 226.

71 Rubinowicz (1917).

72 In 1923, the Austrian professor of physics Friedrich Kottler at the University of Vienna pointed out that Maggi, not Rubinowicz, had first produced the transformation. See Kottler (1923).

He then remarked that Kirchhoff had reduced (18) to an integral solely over an aperture F, producing the following expression for the wave at a point O in the diffraction region:[73]

$$u(O) \equiv \frac{1}{4\pi} \iint\limits_{F} \mathrm{d}s \left\{ \bar{u} \frac{\partial (e^{ikr}/r)}{\partial n} - \frac{e^{ikr}}{r} \frac{\partial \bar{u}}{\partial n} \right\}$$

Rubinowicz next effected a clever choice of region and geometric boundary for the otherwise arbitrary surface G, one that was effectively the same as Maggi's but considerably clearer due to the avoidance of prolate spherical coordinates. Limit G to a surface consisting of the region that would be geometrically illuminated by a source L of the form $\exp(ik\rho)/\rho$ shining through an aperture F which has a rim B.This demarcates a region with one end a surface K_∞ at infinity, capped at the other by the finite aperture F, and delimited by the conical boundary K (Fig. 10). The integral over K_∞ vanishes as usual, leaving only the region $F + K$. Since we have not introduced a physical aperture, the value of $\bar{u}$ over $F + K$ must be the same that it would be in the absence of the screen, hence just $\frac{\exp(ik\rho)}{\rho}$. Consequently the expression for the field at a point $O_E$within the $F + K$ bounded region—which is, again, completely unobstructed—must be:[74]

$$\begin{aligned} u_E(O_E) \equiv \frac{1}{4\pi} \iint\limits_{F+K} ds \left\{ \frac{\exp(ik\rho)}{\rho} \frac{\partial}{\partial n} \left(\frac{\exp(ikr)}{r} \right) \right. \\ \left. - \frac{\partial}{\partial n} \left(\frac{\exp(ik\rho)}{\rho} \right) \left(\frac{\exp(ikr)}{r} \right) \right\} = \frac{\exp(ik\rho)}{\rho} \end{aligned} \tag{19}$$

Introduce a screen such that the surface F corresponds to a physical aperture. In that case for a point O located anywhere in the diffraction region Kirchhoff's integral requires (with $\bar{u}$ on F having the same value that it would were the screen absent):[75]

$$u(O) \equiv \frac{1}{4\pi} \iint\limits_{F} \mathrm{d}s \left\{ \frac{\exp(ik\rho)}{\rho} \frac{\partial}{\partial n} \left(\frac{\exp(ikr)}{r} \right) - \frac{\partial}{\partial n} \left(\frac{\exp(ik\rho)}{\rho} \right) \left(\frac{\exp(ikr)}{r} \right) \right\} \tag{20}$$

Now place the general Kirchhoff point O at O_E within the region bounded by $F + K$. Then for such a point, since (20) integrates over F only, whereas (19) integrates over $F + K$, Rubinowicz could subtract (20) from (19) to write:

73 Rubinowicz (1917), p. 259.

74 *Ibid*., p. 260.

75 *Ibid*., p. 259.

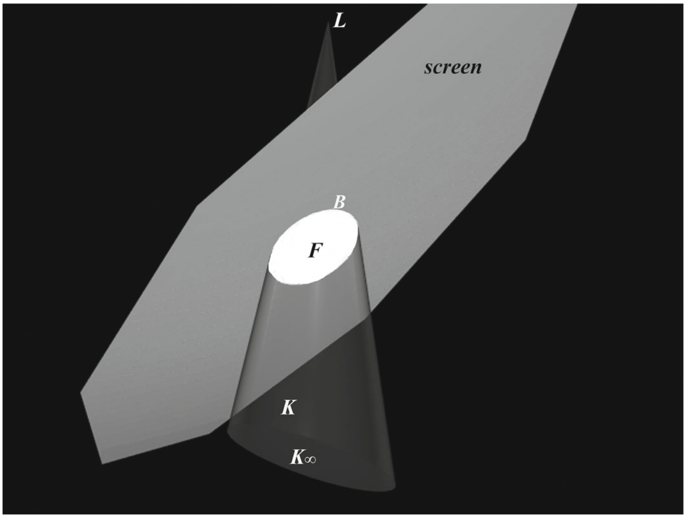

Fig. 10 Rubinowicz's surfaces

$$u(O_E) = u_E(O_E) - \frac{1}{4\pi} \iint\limits_K ds \left\{ \frac{\exp(ik\rho)}{\rho} \frac{\partial}{\partial n} \left(\frac{\exp(ikr)}{r} \right) - \frac{\partial}{\partial n} \left(\frac{\exp(ik\rho)}{\rho} \right) \left(\frac{\exp(ikr)}{r} \right) \right\} \tag{21}$$

Rubinowicz's expression (21) for $u(O_E)$ applies only to a point within the $F + K$ bounded region. Outside of this region Kirchhoff's general expression (20) for $u(O)$ would presumably apply. This might seem to be a pointless formality—we have as it were introduced surface K in (19) only to remove it again in (21). But we have in fact done something more, because we assert that (21) does properly represent the resultant wave within the region bounded by $F + K$ when a screen perforated by F is present. This is nothing more than Kirchhoff's claim for such a region but Rubinowicz's (21) has an important representational effect.

Because the value of the normal gradient $\frac{\partial \bar{u}}{\partial n}$ vanishes over K, Rubinowicz could transform the K integral into one entirely around the rim B of the cap F. Integrating from $\rho = \rho_l$ (at a point on the rim) to infinity thereby produced (recall that ρ in the K integral is the distance between L and a point on K):

$$\iint\limits_K \mathrm{d}s = \oint\limits_B \mathrm{d}l \int\limits_{\rho_l}^{\infty} \mathrm{d}\rho \tag{22}$$

Attention to the geometry yields an expression for the wave field within $F + K$ that involves an integral over the rim B added to the geometric-optics field:

$$u(O_E) \equiv u_E(O_E) + \frac{1}{4\pi} \oint\limits_B \mathrm{d}l \left\{ \frac{\exp[ik(\rho_l + r_l)]}{\rho_l r_l} \frac{\cos(\hat{n}, \vec{r}_l)}{1 + \cos(\vec{r}_l, \vec{\rho}_l)} \sin(\vec{\rho}_l, \mathrm{d}\vec{l}) \right\} \tag{23}$$

In Rubinowicz's (23), r_l is the distance between the field point O and a point l on B, $\hat{n}$ is the normal there to the screen aperture F, $\vec{\rho}_l$ is the distance from the rim point to the luminous point at L, and $\cos(\hat{n}, \vec{r}_l)$ is the cosine of the angle between the two vectors $\hat{n}$ and $\vec{r}_l$. The source wave $u_E(B)$ at the rim appears as the factor $\frac{\exp[ik\rho_l]}{\rho_l}$, so that (23) may also be written as follows:

$$u(O_E) \equiv u_E(O_E) + \frac{1}{4\pi} \oint\limits_{B} \mathrm{d}l \left\{ u_E(B) \frac{\exp[ikr_l]}{r_l} \frac{\cos(\hat{n}, \vec{r}_l)}{1 + \cos(\vec{r}_l, \vec{\rho}_l)} \sin(\vec{\rho}_l, \mathrm{d}\vec{l}) \right\} \quad (24)$$

Rubinowicz's (24) entails an inherent discontinuity at the boundary K: on it the integrand of the rim integral becomes infinite because there the two vectors $\vec{r}_l$, $\vec{\rho}_l$ point in opposite directions. In the region external to $F + K$ the untransformed Kirchhoff expression (20) presumably obtains, and so also do his conditions that both the wave function and its normal gradient vanish over the shadowed surface of the screen. Consequently, and despite the fact that Kirchhoff's $u(O)$ still holds everywhere, the Maggi-Rubinowicz introduction of the $F + K$ boundary introduces a representational discontinuity, though one that neither of the two explored since their interests lay elsewhere. Neither did they ask whether the rim integral would vanish if the observation point O_E lay in the aperture proper, where on the basis of Kirchhoff's boundary condition—which is still presumed—the incident wave $u_E(O_E)$ should be unaltered.

Although Rubinowicz's (24) was in essence similar to Maggi's transformation, Maggi's use of prolate spherical coordinates obscured a startling result upon which Rubinowicz placed considerable emphasis (and which Maggi himself likely did not perceive). Namely, that (24) amounted to a partial vindication of Thomas Young's original method of calculating diffraction by finding the interference between a direct wave from the source and waves presumptively engendered by the direct wave at the edges of the aperture—as Rubinowicz remarked at the very beginning of his article.[76] This amounted to a considerable alteration in the original sense of the Huygens-Fresnel principle from one in which the observed wave is governed entirely by wavelets that are distributed over the aperture's surface, to one in which the efficacious wavelets that alter the incident wave are located on the aperture's rim.

Rubinowicz's article appeared in 1917, over three decades after Kirchhoff's original and Maggi's own transformation, and neither he nor Maggi had concerned themselves with the problem of Kirchhoff's boundary conditions (which Maggi may not have been aware of in 1888). Each of them had a different purpose in mind, Maggi to avoid Kirchhoff's function F, which led him to his line-integral transformation, and Rubinowicz to retrieve a simulacrum of Young's original theory of diffraction. Does the *Poincaré paradox* carry over as well to the Maggi-Rubinowicz transformation? Like Maggi, but with greater geometric clarity since he had not introduced the complexity of prolate spherical coordinates, Rubinowicz had altered the way in which

[76] Partial only because Young had assumed that the wave emanating from the aperture's rim would be the same as the source wave but shifted in phase by half a wavelength. Rubinowicz's rim integral is considerably different.

Kirchhoff's boundary conditions had been applied by both Kirchhoff and by Poincaré by introducing two distinct regions across which the expression for the wave function is discontinuous. Neither of these regions had a boundary coincident with the entire surface of the diffractor proper since the one was capped by the aperture and the other by the surface outside the geometric shadow. For that reason Poincaré's proof of inconsistency, which required no discontinuities as well as the enclosing surface of the diffractor (within which Poincaré placed his point L to generate the paradox), could not be so directly applied. The question accordingly arises as to whether the new representation might capitalize on discontinuity to evade the Poincaré paradox. To do so in a way that preserved the Green theorem structure in some form would certainly require altering Kirchhoff's boundary conditions.

8 A form of consistency achieved

Work continued in the twentieth century through at least two alternatives which, though not always aimed principally at that purpose, could avoid the Poincaré paradox. In 1923 Friedrich Kottler (1886–1965), physics professor at the University of Vienna proposed formulating the integral as a solution to a "saltus" problem—one in which the field undergoes a discontinuous jump across a boundary, requiring the specification of its values on either immediate side, much as, in electrostatics, a fictitious distribution of charge at the boundary across which dielectric capacity changes produces a discontinuity in the potential.[77] Kottler proposed to treat Kirchhoff's conditions as though they were the result of a similar kind of discontinuity, thereby obviating the Poincaré paradox but at the cost of introducing an assumption without physical justification.[78] In 1933 Max Born (1882–1970), thinking in terms of his characteristic recursive method, suggested that the Kirchhoff integral might be the first-order approximation in a sequence of iterative solutions that converged to the exact solution represented by his boundary conditions.[79]

In 1964 Marchand and Wolf remarked that "the difference between the consistent solution of Rayleigh and Sommerfeld and the inconsistent solution of Kirchhoff may be regarded as due to the superposition of plane waves whose amplitude distribution has a very sharp maximum for … waves propagated along the plane of the screen. Such waves do not contribute to the far field."[80] Two years later they followed with an analysis that enabled a direct calculation of precisely such a series of waves by means of a reworking of the boundary conditions through a clever use of the Maggi-Rabinowicz transformation, thereby rescuing the essence of Kirchhoff's formulation and, with it, the meaning of an expression akin to Huygens principle.

Rubinowicz's form of the transformation, recall, could be expressed in terms of the source as follows (*cf* (23)) invoking a discontinuity for points past the screen as

[77] Kottler (1923) and, decades later, Kottler 1965.

[78] By contrast, in electro- and magneto-statics dielectric and para- or dia-magnetic substances yield the discontinuity as a result of the jump in permeability at a boundary.

[79] Born (1933), p. 152.

[80] Marchand and Wolf (1964).

one moves from within to without the geometric shadow since the denominator in the formula vanishes for points on the shadow proper:

$$u(O_E) \equiv u_E(O_E) + \frac{1}{4\pi} \oint_B \mathrm{d}l \left\{ u_E(B) \frac{\exp[ikr_l]}{r_l} \frac{\cos(\hat{n}, \vec{r}_l)}{1 + \cos(\vec{r}_l, \vec{\rho}_l)} \sin(\vec{\rho}_l, \mathrm{d}\vec{l}) \right\}$$

Rubinowciz's transformation of Kirchhoff's integral within the geometrically-illuminated region

Kirchhoff's surface-integral involves no such discontinuity of the kind for it leads to the following expression (*cf* above, note 41):

$$\varphi_o = \frac{1}{2\lambda} \iint_A \frac{\mathrm{d}s}{r_i r_o} \left[\cos(\vec{r}_i, \vec{n}_i) + \cos(\vec{r}_o, \vec{n}_o)\right] \sin\left[2\pi \left(\frac{r_i + r_o}{\lambda} - \frac{t}{T}\right)\right]$$

Marchand and Wolf considered that the inconsistency built into Kirchhoff's theory by his boundary conditions could be shown in a different manner from Poincaré by noting a related consequence. Take a point in the diffracted region (not, as with Poincaré, within the geometrically-defined body of the diffractor itself) and move it indefinitely close to the open aperture. Calculate the value of the wave function at that point using the expression that results from Kirchhoff's theory. It should by hypothesis be effectively the same as the unaltered source wave, but in fact it is not.[81] Why, one might ask, did they turn to this way to show the inconsistency? It is quite direct and perhaps more revealing than Poincaré's route, but, more to the point, it is precisely this consequence of the full set of Kirchhoff's conditions that can be avoided by exploiting the discontinuity at the geometric shadow inherent in the Maggi-Rubinowicz transformation.[82]

To that end, Marchand and Wolf began not with the transformation but with an explicit alteration of the boundary conditions, something that neither Maggi nor Rubinowicz had proposed. Consider with Marchand and Wolf a luminous point L whose distance from a given point of the aperture rim is s_0. Consider a point P' with coordinate (x,y) on the shadowed part of the screen itself, including any aperture(s), and express the presumptive value $U(P')$ of the wave at such a point of the screen by a sum of the source wave added to an integral over the aperture's rim B, of the following form, with points Q' lying on the rim:

[81] The expression given immediately above is not applicable to a point close to the aperture because of the assumption that the aperture point to observation point distance is vastly larger than a wavelength. If however we take as an example of the problem what occurs in the case of a circular disk as screen and make no approximations, then two problems arise using the full Kirchhoff integral. The solution along the axis produces two terms, the second of which is infinite and obviously unphysical. Moreover—and this is the sort of problem that Marchand and Wolf had in mind—even the first term does not vanish at the disk itself, where by hypothesis no wave at all should exist. In other words, the boundary condition presupposed cannot be recovered (see Lucke 2004, pp. 3–4). A similar situation arises for diffraction by a circular aperture, in which the incident wave is not recovered at the aperture.

[82] Marchand and Wolf (1966).

$$U(P') \equiv f_0(x, y) = \varepsilon_0(x, y)u_{\text{inc}}(x, y) + \frac{1}{4\pi}\oint_B u_{\text{inc}}(Q')\frac{\exp(ikQ'P')}{Q'P'}\vec{K}\cdot d\vec{l}$$

where

$$\vec{K} = \frac{1}{4\pi}\frac{\overrightarrow{SQ'} \times \overrightarrow{Q'P'}}{\left|\overrightarrow{SQ'}\right|\left|\overrightarrow{Q'P'}\right| - \overrightarrow{SQ'}\cdot\overrightarrow{Q'P'}}$$

and where
$\varepsilon_0 = 0$ for points on the screen but outside the aperture
$\varepsilon_0 = 1$ for points in the aperture proper
S is the location of a point light source outside the aperture

This expression for the field intensity $U(P')$ on both the screen's shadowed surface and on the aperture constitute *new boundary conditions* for Kirchhoff's problem of diffraction. Where Kirchhoff had set the value at the aperture to the incident wave, Marchand and Wolf add to that wave the rim integral. And where Kirchhoff had set the wave on the screen but outside the aperture to zero, they now add the rim integral. Neither Maggi nor Rubinowicz had suggested that the rim integral might be extended to points of the screen outside the aperture, and neither had they considered what takes place if the observation point were to lie in the aperture proper, where, again, on Kirchhoff's boundary condition the rim integral should vanish to retrieve the unaltered incident wave.

In a critical next step, Marchand and Wolf note that their expression for $\vec{K}$ is identical to the one that appears in the Maggi-Rubinowicz transformation of Kirchhoff's diffraction integral, i.e. in the factor of $u_E(B)\frac{\exp[ikr_l]}{r_l}$ in Rubinowicz's rim integral (24). From here, they claim that the diffracted wave field at a point $P = (x, y, z)$ on the side of the aperture opposite to the light source ($z > 0$) can be expressed as

$$U_K(x, y, z) = \varepsilon_0(x, y, z)u_{\text{inc}}(x, y, z) + \frac{1}{4\pi}\oint_B u_{\text{inc}}(Q')\frac{\exp(ikQ'P)}{Q'P}\vec{K}\cdot \mathrm{d}\vec{l}$$

Although the Marchand-Wolf integral appears to differ from the term that appears in Rubinowicz's expression of Kirchhoff's formula (24), they are nevertheless essentially equivalent and so both exhibit precisely the same discontinuity: at the geometric shadow proper, i.e. if P' lies on K, $\overrightarrow{SQ'}$ parallels $\overrightarrow{Q'P'}$, and consequently the expression for $\vec{K}$ is discontinuous there. Now, they continued, the rim integral alone should be extended to *any point* outside the geometrically-illuminated region. How so, when Rubinowicz's derivation apparently concerned only the latter? Though Marchand and Wolf did not explicitly provide the reason, it is not at all hard to see: if Kirchhoff's boundary condition outside the geometric region but on the screen is replaced by Marchand-Wolf's (with $\varepsilon_0 = 0$), then on the conical sides delimiting the region the normal derivative of the wave function again vanishes, while the aperture rim still delimits the region, leaving, *via* the Maggi-Rubinowicz transformation, only the rim integral over the closed part of the screen since the latter is bounded by the aperture.

Consequently the Marchand-Wolf expression for $U(P')$ now holds throughout the diffraction region (in the form of $U_K(x, y, z)$ above) provided that ε_0 is unity everywhere within the geometrically-illuminated region but vanishes outside it. Moreover, because of the new condition at the aperture, the Marchand-Wolf integral for a point in the diffraction region that is indefinitely close to the aperture now does reproduce the presumed value there (i.e., $U_K(x, y, z = 0) \rightarrow f_0(x, y)$). This is precisely what does not happen when using Kirchhoff's original with his condition that the wave at the aperture is unaltered from what it would be in the screen's absence. And this is all obtained from the transformation of Kirchhoff's original integral combined with Marchand and Wolf's alteration of the boundary conditions. The Kirchhoff inconsistency is thereby avoided altogether since the new boundary conditions involve only the value of the wave function proper and not its normal gradient, whether over the screen proper or over the aperture—at the expense of discontinuity at the geometric shadow edge.

That is not the only cost, for the transformation to a line integral also implicates a significant amendment to the deployment of Fresnel's integrals, which had (and continue to this day) to permeate practical computations of diffraction since the 1830s. Within the approximation regime that governed the Fresnel (and so Kirchhoff) expressions, namely small wavelengths and loci far from the screen but not too far from the edge of the geometric shadow, these original expressions work well despite the boundary inconsistencies necessary to their derivation. But, one might ask, does the Marchand-Wolf replacement do any better? Indeed, their 'consistent' version of the Kirchhoff integral does, they showed, produce good agreement with an experiment in which the behavior of 3.2 cm microwaves was examined across the aperture in a conducting screen. The experiment indicated that the wave pattern across it varies nearly sinusoidally, which is well predicted by Marchand-Wolf but not of course by Kirchhoff's theory, which requires the unaltered value of the source wave.[83] The Marchand-Wolf expression explains the results as due to the interference of the source wave with the wave scattered at the aperture's edge. But the sinusoidal behavior in the aperture nevertheless does imply that the Kirchhoff assumption of an unaltered incident wave at the aperture works quite well for points in the diffraction regime that are not too close. This retrieves and justifies the Kirchhoff integral within an approximation regime with its longstanding, and persuasive, interpretation in terms of Huygensian wavelets.

9 Conclusion

We have examined a diffraction theory that, for nearly two centuries (including Fresnel's original) continued in use even though, a century in, its only derivation from a wave equation had nevertheless proved to be fundamentally inconsistent with the

[83] Neither are these results accommodated by the Rayleigh-Sommerfeld alternative in which the waveform at the aperture is assumed to be the same as that of the incident disturbance. The second alternative, in which the waveform's normal derivative is specified, fares somewhat better but still misses the mark. *Ibid.*, pp. 1715–1717.

very conditions that yielded the expression that, so far as was then known, worked extremely well. One major reason for that resilience was the theory's instantiation within a mathematical framework of the principle, due to Huygens, on which Fresnel had originally grounded wave optics. Whereas Fresnel's wave optics had deployed assumptions concerning point-source radiators that presumptively constitute a wave front, Kirchhoff's theory was formulated as a solution to a partial differential equation based on Green's theorem. Critically, so far as was known the resulting formula for diffraction worked extremely well.

Empirical adequacy and pragmatic tractability certainly were—and remain—central factors for the tenacity of Kirchhoff's theory. Scientists have, and continue to this day, used the expression provided by Kirchhoff's theory—or even Young's much simpler structure involving two-ray interference—in astronomical observations and optical experiments. In for example Albert Michelson and his colleague F.G. Pease's interferometric measurements at Mt. Wilson Observatory as late as 1920, the wave-optical method they employed to estimate a star's diameter did not go beyond the Fresnel integrals proper.[84] Throughout the late nineteenth and twentieth centuries, Kirchhoff's theory continued to appear in major textbooks and monographs, continued to be used and discussed in physics and engineering periodicals, and was generally considered to be a reasonable expression for the effect of diffraction by an aperture. The problem of consistency, to the extent—and it was apparently not a very great extent—that it was recognized seemed principally to concern a defect in the manner in which the Huygens principle or some variant thereof could be represented within a consistent mathematical context.

The principal alternative to Kirchhoff's theory involved dropping one or the other of the two mutually-inconsistent boundary conditions. Rayleigh, who was not concerned with the Kirchhoff problem *per se*, had produced both in 1896 by choosing particular wave functions, while Sommerfeld elaborated the two alternatives using novel Green's functions, with the consequent result that Kirchhoff's solution proved to be the mean of the Rayleigh-Sommerfeld duo, though without any clear reason as to why that might be so. When Marchand and Wolf compared newly-produced microwave data with predictions from Kirchhoff's and the Rayleigh-Sommerfeld expressions in 1966, they found that Kirchhoff's, inconsistent mathematically though it certainly is, better matched measurements at the aperture.[85] To a certain extent such a result was not altogether surprising since Sommerfeld did not regard the boundary conditions that he had himself used to be physically reasonable, even if mathematically unexceptionable.[86] By contrast, Sommerfeld's own analytical solution in 1896, based on electromagnetic relationships, though limited to diffraction by the edge of a half-infinite plane, did not require any *a priori* assumptions concerning boundary conditions. Yet the extreme complexity and intractability of this approach significantly restricted its practical applicability, whereas Kirchhoff's integral easily produced an expression under the approximation to small wavelengths in relation to aperture size and distance from

84 Michelson and Pease (1920).

85 Marchand and Wolf (1966), pp. 1715–1717.

86 Sommerfeld (1954), p. 200.

the screen for both source and observation point that easily facilitated computations and was widely used in optical research. In that approximation Kirchhoff's theory was essentially a differential-equation-based reformulation of Fresnel's formula, a major theoretical result in early wave optics that had enjoyed significant empirical success.

The inconsistency problem was never, in one sense, truly resolved because it follows directly from the requirements of Green's theorem. And yet in the absence of Green's theorem a comparatively direct reading in terms of anything like Huygens' principle, with its persuasive physical significance, evaporated. Kottler's recommendation to reformulate the theory as a *saltus* problem amounted to introducing a fundamental *discontinuity* in applying Green's theorem in a manner that could retain the meaning of Huygens' principle. Born's suggestion to interpret the Kirchhoff expression in terms of a first-order expansion of whatever the actual solution might be was not directly aimed at preserving the physical significance of the principle. It was nevertheless also an effort to reconstrue Kirchhoff's mathematical structure in a manner that separated his solution, which certainly did grant meaning to the principle, from its grounding in a mathematical inconsistency.

In 1964 Marchand and Wolf employed the long-available Maggi-Rubinowicz transformation to replace Kirchhoff's single surface scheme with one in which the solution is applied separately to two contiguous regions across which the expression for the wave function is discontinuous. That breaks the inconsistency otherwise demanded by Dirichlet-Neumann solution requirements and accordingly breaches the proof of the Poincaré paradox. The resulting expression aligned well with Young's theory in the early nineteenth century, in which light rays from an optical source reach points within the geometrically-illuminated region directly and from the aperture's rim, and outside the region solely from the rim. This new physical meaning, gained through discontinuity and transformation to a line-integral, made the scheme particularly appealing—but not for computational purposes under the usual approximation to vanishingly small wavelength, for there the Fresnel integrals work well for reasons that, in light of Marchand-Wolf, have become clear. The price to be paid was the abandonment of Kirchhoff's original, and extremely simple, boundary condition over the surface of the screen *via* the introduction of discontinuity.

The history of Kirchhoff's solution illustrates that mathematical consistency is not inevitably a necessary condition for success in physics. While logical compatibility, mathematical rigor, and conceptual coherence are important epistemic virtues for theory choice and development, they are not inexorably requisite. Empiricist and pragmatic attitudes prompted optical scientists significantly to downplay the mathematical inconsistency of Kirchhoff's theory, a theory that exemplifies a long research program that dominated optics for two centuries. Rather than abandoning altogether such a useful scheme because of mathematical requirements, researchers instead preferred to use it and eventually to reconstruct the mathematics on a different, if doubtfully rigorous (because involving discontinuity) basis, and thereby to grant it continued life.

Acknowledgments We thank Diana Kormos Buchwald for her assistance in correcting errors.

References

Anonymous. 2015. Maggi, Gian Antonio. Treccani Enciclopedia Italiana. http://www.treccani.it/enciclopedia/gian-antonio-maggi/. Accessed 15 July 2015.

Babich, V.M., M.A. Lyalinov, and V.E. Grikurov. 2007. *Diffraction theory: The Sommerfeld–Malyuzhinets technique*. UK: Alpha Science Intl. Ltd.

Baker, B., and E.T. Copson. 1939. *The mathematical theory of Huygens' principle*. Oxford: Clarendon Press.

Bloor, D. 2011. *The enigma of the aerofoil: Rival theories in aerodynamics, 1909–1930*. Chicago: University of Chicago Press.

Born, M. 1933. *Optik*. Berlin: Julius Springer-Verlag.

Born, M., and E. Wolf. 2002. *Principles of optics*. Cambridge: Cambridge University Press.

Buchwald, J. 1980. Optics and the theory of the punctiform ether. *Archive for History of Exact Sciences* 21: 245–278.

Buchwald, J. 1981. The quantitative ether in the first half of the nineteenth century. In *Conceptions of ether: Studies in the history of ether theories, 1740–1900*, ed. G. Cantor, and M.J.S. Hodge, 215–237. Cambridge: Cambridge University Press.

Buchwald, J. 1985. *From Maxwell to microphysics: Aspects of electromagnetic theory in the last quarter of the nineteenth century*. Chicago: University of Chicago Press.

Buchwald, J. 1989. *The rise of the wave theory of light: Optical theory and experiment in the early nineteenth century*. Chicago: University of Chicago Press.

Buchwald, J. 1994. *The creation of scientific effects: Heinrich Hertz and electric waves*. Chicago: University of Chicago Press.

Buchwald, J. 2012. Cauchy's theory of dispersion anticipated by Fresnel. In *A master of science history*, ed. J. Buchwald, 399–416. Dordrecht: Springer.

Buchwald, J. 2013. Optics in the Nineteenth Century. In *The Oxford handbook of the history of physics*, ed. J. Buchwald, and R. Fox, 445–472. Oxford: Oxford University Press.

Charpentier, E., E. Ghys, and A. Lesne (eds.). 2010. *The scientific legacy of Poincaré*. Providence, RI: American Mathematical Society.

Cheng, A.H.D., and D.T. Cheng. 2005. Heritage and early history of the boundary element method. *Engineering Analysis with Boundary Elements* 29: 268–302.

Cisotti, U. 1938. Gli scritti scientifici di Gian Antonio Maggi. *Rendiconti del Seminario Matematico e Fisico di Milano* 12: 167–189.

Cross, J.J. 1985. Integral theorems in Cambridge mathematical physics, 1830–55. In *Wranglers and physicists: Studies on Cambridge mathematical physics in the nineteenth century*, ed. P.M. Harman, 112–148. Manchester: Manchester University Press.

Darrigol, O. 2012. *A history of optics from Greek antiquity to the nineteenth century*. Oxford: Oxford University Press.

Darrigol, O. 2015. Poincaré's Light. In *Henri Poincaré, 1912–2012: Poincaré Seminar 2012*, ed. B. Duplantier, and V. Rivasseau, 1–50. Basel: Springer.

Dieudonné, J. 2008. *Jules Hernri Poincaré: Complete dictionary of scientific biography*. Detroit: Charles Scribner's Sons.

Eckert, M. 2013. *Arnold Sommerfeld: Science, life and turbulent times, 1868–1951*. New York: Springer.

Goodman, J.W. 1988. *Introduction to fourier optics*. New York: McGraw Hill.

Gouy, L.G. 1886. Recherches experimentales sur la diffraction. *Annales de chimie et de physique* 52: 145–192.

Gray, J. 2013. *Henri Poincaré: A scientific biography*. Princeton: Princeton University Press.

Green, G. 1828. *An essay on the application of mathematical analysis to the theories of electricity and magnetism: Mathematical papers of the late George Green*. Cambridge: Cambridge University Press.

Hentschel, K., and N. Zhu. Gustav Robert Kirchhoff's treatise *On the theory of light rays (1882)* - English Translation, Analysis and Commentary. Stuttgarter Beiträge zur Wissenschafts- und Technikgeschichte. Berlin: Logos-Verlag (forthcoming).

Herschel, J. 1827. Light. *Encyclopedia Metropolitana*, 341–586.

Jungnickel, C., and R. McCormmach. 1986. *Intellectual mastery of nature: Theoretical physics from Ohm to Einstein*. Chicago: University of Chicago Press.

Kipnis, N. 1991. *History of the principle of interference of light*. Basel: Birkhäuser.

Kirchhoff, G. 1876. *Vorlesungen über mathematische Physik, Bd 1: Mechanik*. Leipzig: B. G. Teubner.

Kirchhoff, G. 1882. Zur Theorie der Lichtstrahlen. *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften zu Berlin part* 2: 641–669.
Kirchhoff, G. 1883. Zur Theorie der Lichtstrahlen. *Annalen der Physik* 255: 663–695.
Kirchhoff, G. 1891a. *Gesammelte Abhandlungen von G. Kirchhoff*. Leipzig: J. A. Barth.
Kirchhoff, G. 1891b. *Vorlesungen über mathematische Physik, Optik*. Leipzig: B. G. Teubner.
Kline, M. 1972. *Mathematical thought from ancient to modern times*. New York: Oxford University Press.
Kong, J.A. 1986. *Electromagnetic wave theory*. New York: Wiley.
Kottler, F. 1923. Zur Theorie der Beugung an schwarzen Schirmen. *Annalen der Physik* 375: 405–456.
Kottler, F. 1965. Diffraction at a black screen: Part I: Kirchhoff's Theory. *Progress in Optics* 4: 283–314.
Kuhn, T. 1962. *The structure of scientific revolutions*. Chicago: University of Chicago Press.
Lorentz, H.A. 1887. De l'influence du mouvement de la terre sur les phénomènes lumineux. *Archives Néerlandaises* 21: 103–176.
Lucke, R.L. 2004. *Rayleigh-Sommerfeld diffraction vs Fresnel-Kirchhoff, Fourier propagation, and Poisson's spot. NRL/FR/7218–04–10,101*. Washington, DC: Naval Research Laboratory.
Maggi, G.A. 1888. Sulla propagazione libera e perturbata delle onde luminose in un mezzo isotropo. *Annali di Matematica Pura ed Applicata* 16: 21–48.
Maggi, G.A. 1914. Sul teorema di Kirchhoff traducente il principio di Huygens. *Annali di Matematica Pura ed Applicata* 22: 171–177.
Marchand, E.W., and E. Wolf. 1964. Comparison of the Kirchhoff and the Rayleigh-Sommerfeld theories of diffraction at an aperture. *Journal of the Optical Society of America* 54: 587–594.
Marchand, E.W., and E. Wolf. 1966. Consistent formulation of Kirchhoff's diffraction theory. *Journal of the Optical Society of America* 56: 1712–1722.
Marx, W. 2016. Bibliometric analysis of Kirchhoff's paper. *Gustav Robert Kirchhoff's Treatise On the Theory of Light Rays* (1882) [forthcoming]
Michelson, A., and F.G. Pease. 1920. Measurement of the diameter of α Orionis with the interferometer. *Contributions from the Mount Wilson Observatory* 203: 249–260.
Poincaré, H. 1889. *Leçons sur la Théorie Mathématique de la Lumière, professés pendant le premier semestre 1887–1888*. Paris: Georges Carré.
Poincaré, H. 1892a. Sur la polarisation par diffraction. *Acta Mathematica* 16: 1–50.
Poincaré, H. 1892b. *Théorie mathématique de la lumière II: Nouvelles études sur la Diffraction – Théorie de la dispersion de Helmholtz: Leçons professés pendant le premier semestre 1891–1892*. Paris: Gauthier-Villars.
Rayleigh, 1896. *Theory of sound*. London: Macmillan and Co.
Rayleigh, 1897. On the passage of waves through apertures in plane screens, and allied problems. *Philosophical Magazine* 43: 259–272.
Rayleigh, 1913. On the passage of waves through fine slits in thin opaque screens. *Proceedings of the Royal Society of London* 89: 194–219.
Riemann, B. 1857. *Theorie der Abel'schen Functionen*. Berlin: Georg Reimer.
Rubinowicz, A. 1917. Die Beugungswelle in der Kirchhoffschen Theorie der Beugungserscheinungen. *Annalen der Physik* 53: 257–278.
Saatsi, J., and P. Vickers. 2011. Miraculous success? Inconsistency and untruth in Kirchhoff's diffraction theory. *British Journal for the Philosophy of Science* 10: 1–18.
Schweber, S.S. 1994. *QED and the Men Who Made It: Dyson, Feynman, Schwinger, and Tomonaga*. Princeton University Press: Princeton.
Seth, S. 2010. *Crafting the quantum: Arnold Sommerfeld and the practice of theory, 1890–1926*. Cambridge: MIT.
Shapiro, A. 1973. Kinematic optics: A study of the wave theory of light in the seventeenth century. *Archive for History of Exact Sciences* 11: 134–266.
Sommerfeld, A. 1895. Zur mathematischen Theorie der Beugungserscheinungen. *Nachrichten von der König. Gesellschaft der Wissenschaften zu Göttingen* 1: 338–342.
Sommerfeld, A. 1896. Mathematische Theorie der Diffraction. *Mathematische Annalen* 47: 317–374.
Sommerfeld, A. 1954. *Optics: Lectures on theoretical physics*, vol. IV. New York: Academic Press.
Sommerfeld, A. (ed.). 2004. *Mathematical theory of diffraction*. Basel: Birkhäuser.
Stokes, G. 1845a. On the aberration of light. In *Mathematical and physical papers*, ed. G. Stokes, 134–140. Cambridge: Cambridge University Press.

Stokes, G. 1845b. On the constitution of luminiferous ether, viewed with reference to the phenomenon of the aberration of light. In *Mathematical and physical papers*, ed. G. Stokes, 153–156. Cambridge: Cambridge University Press.

Stokes, G. 1845c. On the theories of the internal friction of fluid in motion, and of the equilibrium and motion of elastic solids. In *Mathematical and physical papers*, ed. G. Stokes, 243–328. Cambridge: Cambridge University Press.

Stokes, G. 1856. On the dynamical theory of diffraction. *Transactions of the Cambridge Philosophical Society* 9: 1–62.

Stokes, G. 1883. *Mathematical and physical papers*. Cambridge: Cambridge University Press.