Diffraction at a Straight Edge

A Gem from Sommerfeld's Work in Classical Physics

Rajaram Nityananda

The simplest problem in diffraction – light passing a straight edge – did not receive a rigorous solution till Sommerfeld's work of 1896. The earlier theories, their successes and their limitations are recounted. They led upto Sommerfeld's final solution which had eluded many great contemporaries. This solution was a landmark of rigorous diffraction theory.

Rajaram Nityananda works at the School of Liberal Studies, Azim Premji University, Bengaluru. Earlier, he spent a decade at the National Centre for Radio Astrophysics in Pune, and more than two decades at the Raman Research Institute in Bengaluru. He has taught physics and astronomy at the Indian Institute of Science in Bengaluru and the IISERs in Pune and Mohali. He is keenly interested in optics in a broad sense, and more generally, problems, puzzles, paradoxes, analogies, and their role in teaching and understanding physics.

Introduction

Diffraction refers to light not traveling in a straight line and bending into the shadow. It was observed by Grimaldi in Italy in the 17th century. Newton also studied diffraction and tried, unsuccessfully, to explain it in terms of his 'corpuscles' (particles) of light. Young and Fresnel contributed greatly to the experimental side in the early 19th century and gave wave theories to explain their results. Fresnel's theory was based on Huygens' idea of secondary waves, and needed to be put on a proper mathematical basis. This task was taken up by Kirchoff in the late 19th century. However, his theory had its own unjustified assumptions. It was Sommerfeld who, in 1896, put the whole subject on a sound physical and mathematical footing, using electromagnetic theory and the proper treatment of the screen as an electrical conductor. He introduced the 'Sommerfeld radiation condition' which is needed to obtain a unique solution to the problem. This condition has been used ever since in such problems. He also had to invent new mathematical methods, which are rarely mentioned even in advanced courses. This article introduces this chapter of optics,

Keywords
Diffraction, straight edge, Sommerfeld.

using the simplest possible example – diffraction by a straight edge.

Fresnel–Huygens Theory

Old fashioned textbooks of optics have separate chapters for ‘interference’ and ‘diffraction’. But this distinction is artificial. We are first taught superposition of two waves, each coming from a single slit. We think of each slit as a point source. However, a slit is really finite in size – otherwise no light would come out! So logically, one should first understand the wave coming out of each slit, before superposing two such waves! Typically, the single slit appears in a later chapter. Mathematically speaking, the single-slit diffraction pattern is a superposition of an infinite number of waves, one for each point on the slit, evaluated using integral calculus. In this article, we will use the word diffraction to cover all experiments when light waves from a source encounter an obstacle in the form of a screen with one or more holes or slits (apertures). This includes what is usually called ‘interference’. It is also a particular case of a more general situation, called ‘scattering’, when the obstacle need not be a thin screen with holes, but can be more general, e.g., a raindrop.

The simplest problem is diffraction by a single straight edge, regarded as infinite (*Figure* 1).

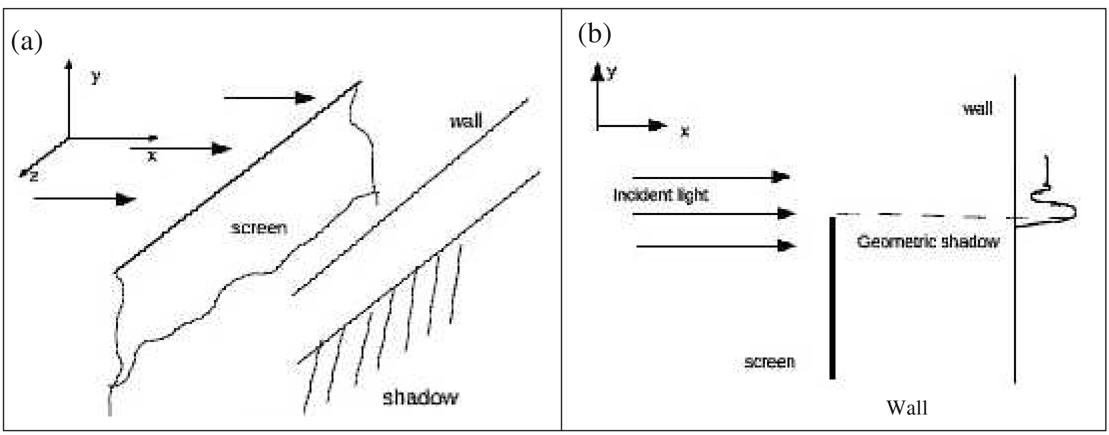

Figure 1. Geometry of diffraction by a straight edge. **(a)** The screen is in the y–z plane at $x = 0$, the incident light comes along the x-axis from the negative x-direction, and the observations are made beyond the screen at positive x, along the y-axis. The irregular lines indicate that the screen continues beyond what is shown. The shadow is sketched even though it is on the invisible side of the wall. **(b)** View of the same geometry along the z-axis, showing intensity variation along the wall schematically.

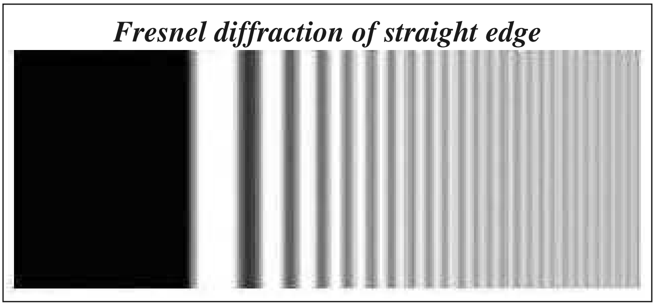

Figure 2. Intensity variation near the geometric shadow of a straight edge, as we move along the y-axis. Note that there is light even in the geometric shadow, and the intensity shows oscillations near the edge of the shadow.
Courtesy: http://commons.wikimedia.org/wiki/File:Fresnel_diffraction_of_straight_edge_density_plot.jpg

This was studied experimentally by the French physicist A Fresnel, around 1810. He was able to give a theory for the phenomenon. The resulting intensity distribution is shown in *Figure* 2. The pattern of intensity on the wall is similar at different distances, but the size of the pattern increases as the square root of the distance from the screen (*Box* 1).

Box 1. The Fresnel Theory of Diffraction

Huygens stated his principle in terms of secondary waves traveling in all directions. But he finally used only the common tangent to those secondary waves, to get a new wavefront. Nowhere in his book '*Traite de la Lumiere*' is there any hint of wavelength, or of interference, or of phase. His great achievement was to explain the double refraction of calcite using two wavefronts, one spherical and the other a spheroid (ellipse rotated about one of its axes, major or minor).

The application of the secondary wave idea to calculate diffraction patterns is due to Fresnel. In the case of the single straight edge, the geometry is shown in *Figure* 1.

According to the Fresnel theory, we think of the incident wave existing with its original strength, say 1, over the unobstructed part of the $y-z$ plane, $y > 0$. The phase of the wave is assumed to be constant over this region $y > 0$. The field at a point (x, y) is given by a sum over contributions from all points on the wavefront (*Figure* 3a). The path length from $(0, y')$ to (x, y) is given by $(x^2+(y-y')^2)^{1/2}$ which can be approximated by $x+(y-y')^2/2x$ in cases where the second term is smaller than the first. This path difference has to be multiplied by $2\pi/\lambda \equiv k$ to obtain the phase difference, $kx + k(y - y')^2/2x$. We therefore have to integrate $\exp(ikx) \times \exp(ik(y - y')^2/2x)$ with respect to y', from zero to infinity. It is natural to change variables using $u^2 = k(y - y')^2/2x$. u has the physical meaning of distance along the wavefront, measured in units of $\sqrt{(2x/k)} = \sqrt{(x\lambda/\pi)} \equiv d_{\mathrm{F}}$ The distance d_{F} – sometimes something proportional to it – is called the Fresnel scale. Note that d_{F} depends on $\sqrt{x}$, where x measures how far we are from the plane of the straight screen. To get a feel for numbers, the early diffraction experiments by Grimaldi,

Box 1. Continued...

Box 1. Continued...

take x about 10 metres, and $\lambda = 0.5 \times 10^{-4}$ cm. Note that we are not implying that Grimaldi knew he was dealing with waves!. We find $d_{\mathrm{F}} = 1.26$mm, so it is possible for a good observer to see the oscillations in intensity (fringes) on a wall in a dark room.

Figure 2 shows the intensity distribution in the straight-edge diffraction pattern. Ray optics would tell us that we would see a uniformly bright region separated by a sharp line from a fully dark shadow. The real situation is that light bends into the geometric shadow, and even before the shadowed region, the intensity is not constant but oscillates.

Our complex amplitude is proportional to $U_{\mathrm{F}} = \int_{-y/d_{\mathrm{F}}}^{\infty} \exp(iu^2)\mathrm{d}u \equiv Fr_{\mathrm{F}}(-y/d_{\mathrm{F}})$. This integral is a complex number which is a function of the lower limit. We have named it Fr in honour of Fresnel.

We can fix the constant of proportionality by taking y large and positive, far above the edge. In this case, the lower limit would be minus infinity. The answer should be our incident wave amplitude of unity. The constant of proportionality is therefore given by the reciprocal of the same integral over the full range – we will not need to use it but the integral equals $\sqrt{i\pi}$. The behaviour of the intensity – the square of this amplitude, is shown as a function of y, the position of the point where we are observing the diffraction pattern (*Figure* A).

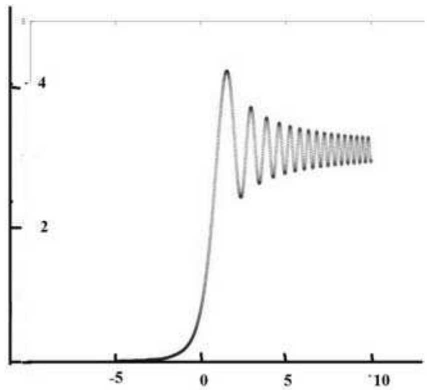

Figure A. The intensity of the diffraction pattern on a plane situated at a distance *x* from the straight edge, according to the Fresnel theory. The horizontal axis is in units of the Fresnel scale d_{F}. The vertical axis shows $|(Fr(y/d_{\mathrm{F}}))^2|$ which is proportional to the intensity of the light. (Note that the value is tending to π for large values of *y*.)

Fresnel's theory is based on Huygens' idea that every point on a wavefront acts as a source of 'secondary waves'. It is sketched in *Figure* 3a and outlined in *Box* 1. This theory was very successful in explaining the early experiments. However, the idea of secondary waves raises many questions. Why should a wave itself act as a source of secondary waves? Why do we not include the secondary wave traveling in the backward direction from a given wavefront?

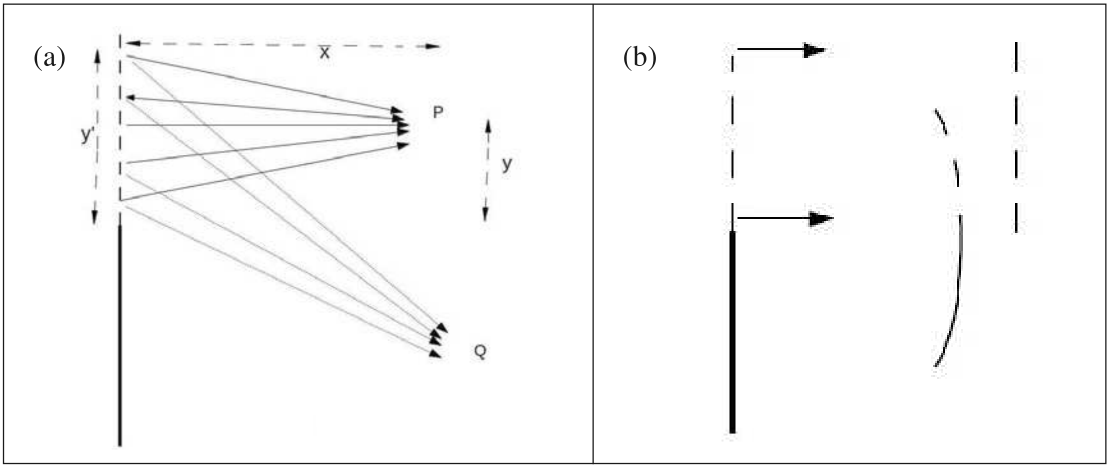

Figure 3. (**a**) The physical picture of straight edge diffraction according to Fresnel. All points on the wavefront contribute secondary waves, with equal amplitude and phases depending on the distance. For a point P in the illuminated region, there is a point where the distance is a minimum, around which the phase varies slowly, and we get a significant sum. For a point like Q well into the shadow region, the phases do not vary slowly and there is srong cancellation, leading to a greatly reduced intensity.
(b) Straight edge diffraction according to Young: In this figure, the plane wave from the source simply continues with the lower part cut off by the screen. The circular arc represents a wave originating from the edge of the screen. The intensity falls strongly away from the geometric shadow boundary. There is a discontinuity in the amplitude at the shadow, which exactly compensates for the discontinuity in the plane wavefront, giving a continuously varying sum of the two waves. This discontinuity is represented by the sudden changeover from a continuous line to a dashed line on the circular arc.

Around the same time, Young in England gave a different formulation in which the original wave falling on the screen travels unaltered into the region accessible to rays, An additional wave originates from the edge of the aperture, and enters the geometric shadow (*Figure* 3b). It might appear impossible to reconcile these two points of view, as different as the English and the French nations! Young did not give a mathematical formulation. Maggi (1890) and Sommerfeld's student Rubinowicz (1912) were able to show the equivalence of these two very different looking pictures. When the expression given by the Fresnel theory was transformed using integration by parts, it precisely gave rise to the Young edge wave!

It might appear difficult to reconcile the Young and Fresnel points of view, as different as the English and French nations.

Given its successes, Huygens' principle needed a deeper justification. This is what Kirchoff sought in 1882. His starting point was the wave equation – also called the d'Alembert equation after its discoverer. To simplify matters, he created a theory which really applies to waves like sound, not light. The difference is that sound waves are described by a single scalar quantity, the pressure. However. from the phenomenon of polarization, it was known that light waves are transverse. This means that there is a vector quantity describing the light wave, with two components, perpendicular to the direction in which the wave moves. Maxwell's electromagnetic theory of light tells us that this vector can be chosen as the electric field of the wave. In the straight edge problem there are two cases. We can take the electric – or the magnetic – vector along z, the direction of the edge. Symmetry then guarantees it will be along z everywhere. The scalar theory can be used for this field component.

The basic result of Kirchoff's theory is described in *Figure* 4. Kirchoff's theory is not really a solution to the problem but all solutions of the wave equation. The (scalar) field at any point P is expressed in terms of the same field and its first derivative, taken over a surface surrounding the point. By choosing the surface suitably, one may be able to make a good guess about the field on it, and hence use the identity to calculate the field at P. The assumptions made are as follows.

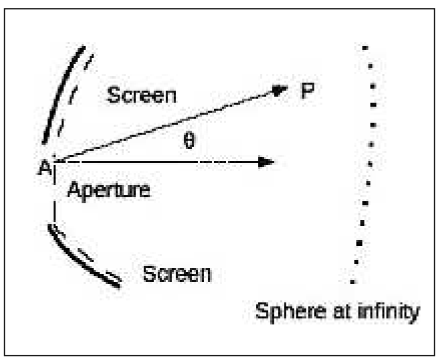

Figure 4. The Kirchoff integral. The solid line is the screen, and the dashed line a surface drawn just inside it, but covering the aperture as well. The field at the point P is only contributed by points like A on the aperture, which are assumed to be illuminated by the incident wave, The secondary wave is proportional to the incident wave, and to $\frac{\exp ikr}{r} \times (1 + \cos\theta)$, where θ is the angle between the normal to the incident wave at A, and the direction AP. The contributions from the inside of the screen and the sphere at infinity vanish.

a) On the side of the screen away from the source, the field and its first derivative (taken perpendicular to the screen) vanish. This seems reasonable because it is a shadow region

b) On the aperture, the field is equal to the field of the source, as if it was in free space. This assumption is common with the Fresnel theory

c) On a spherical surface far away from the screen, on the right-hand side of the figure one will have a spherical wave which is outgoing, since there are no sources further to the right to produce incoming waves.

The final formula given by Kirchoff looks rather similar to that of Fresnel. It represents significant progress, because the secondary wave emerges naturally from the mathematics, and does not have to be postulated. Further, the wave carries a factor $1 + \cos\theta$, where θ is the angle between the normal to the wavefront and the direction from the secondary source to the point where we are calculating the field. This means that the amplitude of the secondary wave is maximum normal to the wavefront, and is zero in the backward direction. The good news is that the theory agrees very well with experiments on diffraction.

The bad news is that the final formula contradicts assumption (a). If we evaluate the integral at a point just behind the screen, it is not zero. It is true that the field at the screen calculated this way is usually small if the aperture is large compared to the wavelength. This is because there is considerable cancellation of the contributions from different parts of the aperture, since the path varies by many wavelengths. However this is a serious problem for an aperture of the order of, or smaller than, the wavelength.

The challenge of improving the formula was taken up by Sommerfeld. As a mathematical physicist, he knew that

The final formula given by Kirchoff looks rather similar to that of Fresnel. It represents significant progress, because the secondary wave emerges naturally from the mathematics, and does not have to be postulated.

if a solution of the wave equation and its first derivative vanish on a surface, the solution has to vanish everywhere. His first attempt was therefore to modify Kirchoff's theory to remove this defect. By using the same identity in a different way, he was able to assume that the function was zero, but not the first derivative. This solution was also found by Rayleigh. Unlike Kirchoff's theory, it works only for plane screens. It does not remove the basic inconsistency pointed out earlier. So the theory is restricted to aperture sizes much greater than the wavelength. Agreement with experiment is no better than the Kirchoff solution.

There is a deeper physical reason why the theory of diffraction should be formulated in quite a different way. One can take a fundamental point of view that all metals, or even transparent media like glass, should be regarded as atoms embedded in vacuum. The field at any point, whether it be in front of a screen or behind it, is the sum of the field due to all sources, each radiating in vacuum. The primary source is the incoming wave. Each atom responds to the electric field which it sees, and becomes a source itself. This is truly a secondary source, not the fictitious one of Huygens! The strength of each secondary source depends on the field seen by each atom. This is not only the field of the incoming wave but also the field radiated by other atoms. We cannot simply postulate what the field is in any given region, without knowing what is happening everywhere.

One can take a fundamental point of view that all metals, or even transparent media like glass, should be regarded as atoms embedded in vacuum. The field at any point, whether it be in front of a screen or behind it, is the sum of the field due to all sources, each radiating in vacuum.

Figure 5 shows a metal sheet with no holes, reflecting a plane electromagnetic wave. We normally say that the field is 'unable to reach' the other side because it is blocked by the screen. But from our new viewpoint, the physical origin of the reflected wave lies in currents flowing in the screen, which act as sources. These same sources also radiate in the opposite direction, into the region behind the screen. Why do we not see any field there? *It is because this secondary field plays the role*

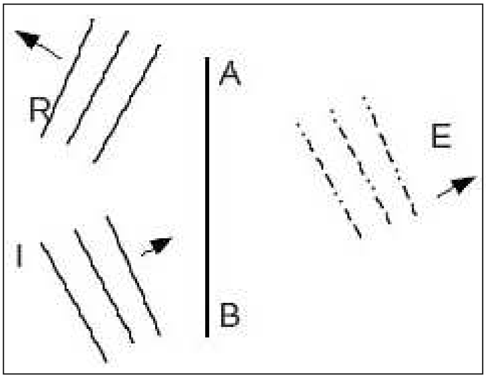

Figure 5. The Ewald–Oseen extinction theorem, illustrated by reflection from a perfectly conducting plane AB. A source far to the left and below generates the incident wave *I*. The field due to *I* sets up currents in the plane AB which radiate the reflected wave *R*. But the same currents also radiate the wave *E*, which travels in the region behind the plane, and is exactly minus the wave radiated by the source. This destructive interference is responsible for the 'obvious' fact that the wave does not penetrate the metal.

of cancelling the incident wave! Among basic physics textbooks, the *Feynman Lectures on Physics*, particularly Volume I, (33–6, the 6th Section in Chapter 33) emphasise, and use this viewpoint. Even when a light wave falls on glass, the atoms of the glass produce the reflected and refracted waves and cancel the incident wave in the glass! This deeper understanding of wave propagation in a medium was originally developed by Ewald (a student of Sommerfeld) in Germany, and independently by Oseen in Sweden, and goes by the name of the extinction theorem.

But from our new viewpoint, the physical origin of the reflected wave lies in currents flowing in the screen, which act as sources. These same sources also radiate in the opposite direction, into the region behind the screen. Why do we not see any field there? *It is because this secondary field plays the role of cancelling the incident wave*!

We therefore have to view the screen as a source of radiation, and the solution to the problem should take into account the currents flowing in the screen. It is no longer obvious that the field in the aperture will be the undisturbed field, nor that the field behind the screen is zero (it would be, if there were no aperture).

The formal way of doing this is to use boundary conditions on the screen for the wave equation, which recognise that it is a perfect conductor. A rigorous solution to the diffraction problem has to have the following features.

a) The field near the source should be an outgoing spherical electromagnetic wave. If the source is at infinity, then we can say that far away to the left of the screen, we have an incoming plane wave.

We should have only outgoing spherical waves at infinity. This is called the `Sommerfeld radiation condition'. Sommerfeld realised that without such a condition, Maxwell's equations would not have a unique solution.

b) The field should satisfy the correct boundary conditions on the screen. For a perfectly conducting metal screen, these conditions require the tangential component of the electric field, and the normal component of the magnetic field, to be zero,

c) We should have only outgoing spherical waves at infinity. This is called the 'Sommerfeld radiation condition. The term not entirely fair to Kirchoff, who used a similar condition! This condition makes physical sense since we can think of currents flowing on the metallic screen as the additional sources, excited by the original source to the left in *Figure* 4, and there is no source to the far right. Sommerfeld realised that without such a condition, Maxwell's equations would not have a unique solution.

It is one thing to write down (a), (b) and (c), quite another to satisfy them! We now go back to the simplest possible problem in diffraction, the single straight edge depicted in *Figure* 1.

We will now simply write down Sommerfeld's solution, and comment on it. Deriving it is quite another matter, and we can only sketch the line of thought in *Box* 2. Sommerfeld's textbook *Optics* gives full details for readers who are (very) mathematically inclined.

We first exhibit the Fresnel–Kirchoff solution of *Box* 1. We express the variable inside the Fresnel integral $Fr(-y/d_{\mathrm{F}})$, in terms of polar co-ordinates, with origin at the edge.

$$U_{\mathrm{FK}} = Fr(-y/d_{\mathrm{F}}) = Fr\left(\frac{-r\sin\theta}{\sqrt{2r\cos\theta/k}}\right).$$

Remarkably, the Sommerfeld solution has the same Fresnel integral occuring in two terms. Notice that the arguments are different.

Box 2. The Sommerfeld Solution

A plane wave travelling at an angle β to the x-axis takes the form $\exp(ik_x x + ik_y y)$ or $\exp(ik(r\cos\theta\cos\beta + \sin\theta\sin\beta)) = \exp(ikr\cos(\theta-\beta))$.

The general solution is a superposition of such solutions with a coefficient $A(\beta)$. Lesser mortals would have assumed that $A(\beta)$ is periodic in β with period 2π. Sommerfeld's insight was to choose an $A(\beta)$ with period 4π, i.e., he used trigonometric functions of $\beta/2$! This is why the solution containing $\theta/2$. This means that if we start on one side of the plane screen $\theta = 0$ and go around the origin (edge) to the other side, $\theta = 2\pi$, we do not get the same value. This is actually desirable, because these are the two boundaries of the problem, one faces the incident wave, and the other does not. In fact, the inspiration came from Riemanns theory of multivalued functions of a complex variable, such as $\sqrt{(z)}$ which changes sign as we go around the origin, from just above the x-axis to just below. Another circuit brings us back to the original value. One can say that Sommerfeld solved the wave equation on a 'Riemann surface' with two sheets, of which one is the physical space, The method generalises to a wedge with an angle $2\pi/n$, which requires $(n+1)$ sheets

Even advanced textbooks like *Classical Theory of Fields* by Landau and Lifshitz give the result for the wedge without a derivation – not surprising given how many pages it occupies in Sommerfeld's *Optics*. Poincaré is reputed to have described the solution as 'very ingenious' – high praise from a man interested in optics (as in everything, see *Resonance*, February 2000) and not lacking ingenuity himself!

$$U_{\rm S} = Fr\left(-2\sqrt{\frac{kr}{2}}\sin(\theta/2)\right) - Fr\left(-2\sqrt{\frac{kr}{2}}\cos(\theta/2)\right).$$

In the forward direction, small θ, the arguments of $U_{\rm FK}$ and the first term of $U_{\rm S}$ reduce to $\sqrt{\frac{kr}{2}}\theta$. The first term therefore agrees with $U_{\rm FK}$ and so all the experimental agreement with the earlier theories remains intact.

How about the second term? On the boundary of the geometric shadow, $\theta = 0$ and $\cos(\theta/2) = 1$. As we move away from the edge by more than the wavelength, the argument of Fr is large and negative, and hence this term is small. But the second term is important, because this is the solution in all of space, not just in front of the screen. It is needed to satisfy the condition that $U_{\rm S}$ vanishes at $\theta = 0$ and $\theta = 3\pi/2$, the two faces of the screen. And believe it or not, the combination of the

Following Sommerfeld's own reasoning in deriving it, now available in his textbook `Optics' is like watching a magician at work.

two terms takes care of the incoming wave at negative x, as well as that reflected by the screen!

Just verifying that the Sommerfeld solution satisfies all the conditions of the problem is a good exercise for the reader. Following Sommerfeld's own reasoning in deriving it, now available in his textbook *Optics*, is like watching a magician at work. It takes all of pages 249–265. He takes care to motivate every step. But at the final stage, even the master teacher throws up his hands. This is where he transforms his integrals over the complex plane into the more familiar Fresnel integrals. In his own words – "Unfortunately, this transformation is somewhat lengthy and of a largely formal character". We try and do a little more justice to the magic in *Box* 2.

Epilogue

Diffraction/scattering by metal obstacles has major technological applications today. The antennas used in communication and mobile phones are really conducting scatterers. The radar signal reflected by an airplane is scattering by 180 degrees, and computing it, or reducing it, is clearly of interest in military circles. The number of analytic solutions is still small, but powerful numerical methods now exist for solving this class of problems. And all of them follow the path opened up by Sommerfeld in this work and its follow up. (He gave an exact solution for a dipole above a conducting sphere modeling the earth, inspired by the introduction of radio communication!)

This solution, important as it is, gives just a small glimpse of Sommerfeld's abilities as a mathematical physicist. His contributions to the birth of quantum theory, and as a teacher, are covered elsewhere in this issue.

Address for Correspondence
Rajaram NItyananda
Azim Premji University,
PES Institute of Technology Campus
Pixel Park, B Block
Electronics City, Hosur Road
(Beside NICE Road)
Bengaluru 560 100
Email: rajaram.nityananda@gmail.com

Suggested Reading

[1] **Arnold Sommerfeld, *Lectures on Theoretical Physics*: *Optics,* Levant Books, Kolkata, 2006.**