E. Wolf, Progress in Optics 50

Chapter 6

Light tunneling

by

H.M. Nussenzveig

Instituto de Física, Universidade Federal do Rio de Janeiro Caixa Postal 68528, Rio de Janeiro, RJ 21941-972, Brazil

ISSN: 0079-6638 DOI: 10.1016/S0079-6638(07)50006-5

Contents

§ 1. Introduction: Newton and contemporaries

The concept of tunneling was introduced in the early days of quantum mechanics (Merzbacher [2002]). It was first applied by Hund [1927], to account for hopping between wells in a double-well molecular bound state. For continuum states, the earliest application is Nordheim's [1927] treatment of thermionic emission and electron reflection by metals, modeled as electron penetration of rectangular potential barriers.

Pioneering applications to nuclear physics were independently proposed by Gamow [1928] and by Gurney and Condon [1928] to explain alpha radioactivity in terms of Coulomb barrier penetration. The huge range of observed lifetimes follows from the exponential sensitivity of the barrier penetration factor to the energy of the emitted alpha particle.

An early application to condensed matter physics was Fowler and Nordheim's [1928] theory of field emission (Eckart [1930]). For a uniform applied electric field, this amounts to penetration of a triangular potential barrier. The name "wave-mechanical tunnel effect" was first employed by Schottky [1931] in connection with photoemission from a metal–semiconductor barrier layer.

In all these instances, "tunneling" refers to the transmission through a potential barrier of a particle with insufficient energy to surmount it according to classical mechanics. It is a consequence of the wave–particle duality, one of the wave properties of matter. There is a well-known wave-mechanical analogy (Gamow and Critchfield [1949]) with the phenomenon of frustrated total reflection in optics, also treated as light tunneling (Sommerfeld [1954]). As will be seen below, light tunneling and barrier penetration are closely related.

The earliest known observation of frustrated total reflection, reported in Query 29 of his "Opticks", is due to Newton [1952], who may therefore be regarded as the discoverer of tunneling (cf. Hall [1902]). It is mentioned in his letter to Oldenburg of 9 December 1675 (Newton [1757]). He was observing the phenomenon known as Newton's rings. As described in Query 29, this was done

". . . by laying together two Prisms of Glass, or two Object-glasses of very long Telescopes, the one plane, the other a little convex, and so compressing them that they do not fully touch, nor are too far asunder."

He had observed that

“The Rays of Light in going out of Glass into a *Vacuum*, are bent towards the Glass, and if they fall too obliquely on the *Vacuum*, they are bent backwards into the Glass, and totally reflected”;
however, by compressing them as he described,

“. . . the Light which falls upon the farther Surface of the first Glass where the Interval between the Glasses is not above the ten hundred thousandth Part of an Inch, will go through that Surface, and through the Air or *Vacuum* between the Glasses, and enter into the second Glass”.

Newton was a remarkable experimenter: he made very accurate measurements of the radii of the rings, and thereby determined the corresponding thicknesses of the “Interval between the Glasses”. How did he interpret these results, given that the visual appearance of the rings immediately suggested periodicity and waves?

Newton’s views on the nature of light were as ambivalent as that nature itself. One of the most remarkable expositions of those views is his “Hypothesis explaining the Properties of Light” in the letter to Oldenburg (Newton [1757]). The dispute (Hall [1995]) following the publication of his first paper, “New Theory about Light and Colors” (Newton [1672]) rendered him extremely averse to publishing his conjectures (“Hypotheses non fingo”), so that his design in the Opticks “. . . is not to explain the Properties of Light by Hypotheses, but to propose and prove them by Reason and Experiments” (Newton [1952]).

“Were I to assume an hypothesis”, Newton writes to Oldenburg, it should be that light “is something or other capable of exciting vibrations in the aether”. These vibrations “succeed one another . . . at a less distance than the hundred thousandth part of an inch” – a correct estimate for half a wavelength of visible light.

Newton assumes that the aether density varies in different media (playing a role similar to the refractive index), with no discontinuity at an interface between two media, which he views as a continuous variation through a transition layer “of some depth”. He proceeds: “. . . refraction I conceive to proceed from the continual incurvation of the ray all the while it is passing” through this layer.

However, beyond the critical angle, the ray “must turn back and be reflected”. How does this take place? To model total reflection, in the “Principia” (Newton [1946]) he invokes the corpuscular model of light, though not committing himself to this model. In the Scholium to Proposition XCVI of Book I, he states:

“Therefore because of the analogy there is between the propagation of the rays of light and the motion of bodies, I thought it not amiss to add the following Propositions for optical uses; not at all considering the nature of the rays of light,

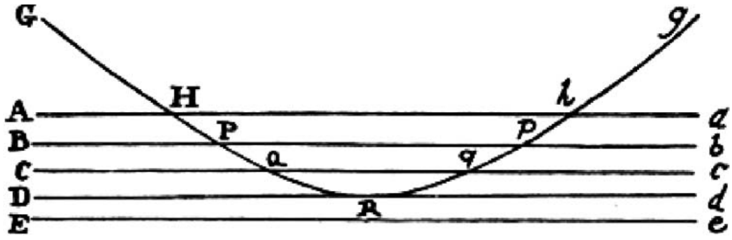


Fig. 1. Newton's model for total reflection of a light ray (Newton [1946], Book I, Proposition XCVI). The reflected ray is laterally displaced.

or inquiring whether they are bodies or not; but only determining the curves of bodies which are extremely like the curves of rays."

This statement is one of the earliest formulations of the optomechanical analogy, which plays a crucial role in the history of light tunneling. There are prior hints of this analogy in the work of the great medieval scholar known as Alhazen (Ronchi [1952]), and Descartes had derived the laws of reflection and refraction by comparison with the trajectory of a ball (Descartes [1637]): "...l'action de la lumière suit en ceci les mêmes lois que le mouvement de cette balle".

Newton's model for the transition layer is mechanically equivalent to interpolation by a linear potential field, like gravity near the earth, so that the light ray describes an arc of a parabola (fig. 1), implying, by the way, that the reflected ray also undergoes a lateral displacement – an early version of the Goos–Hänchen shift (Goos and Hänchen [1947]) which, as we will discuss later, is another manifestation of light tunneling.

A tough challenge to the corpuscular model was how to explain that light is partially reflected and partially transmitted at an interface. How does a light particle choose which path to take? To deal with this conundrum, Newton formulated the first dual model for light, endowing it with both particle and wave properties (Newton [1952], Book Two, Proposition XII):

"Every Ray of Light in its passage through any refracting Surface is put into a certain transient Constitution or State, which in the progress of the Ray returns at equal Intervals, and disposes the Ray at every return to be easily transmitted through the next refracting Surface, and between the returns to be easily reflected by it.

Definition. The return of the disposition of any Ray to be reflected I will call its Fit of Easy Reflexion, and those of its disposition to be transmitted its Fits of easy Transmission, and the space it passes between every return and the next return, the Interval of its Fits."

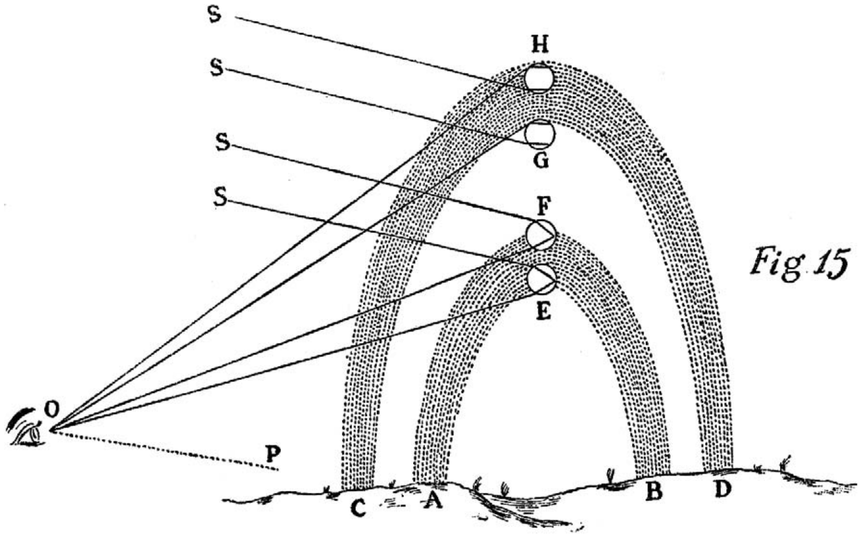
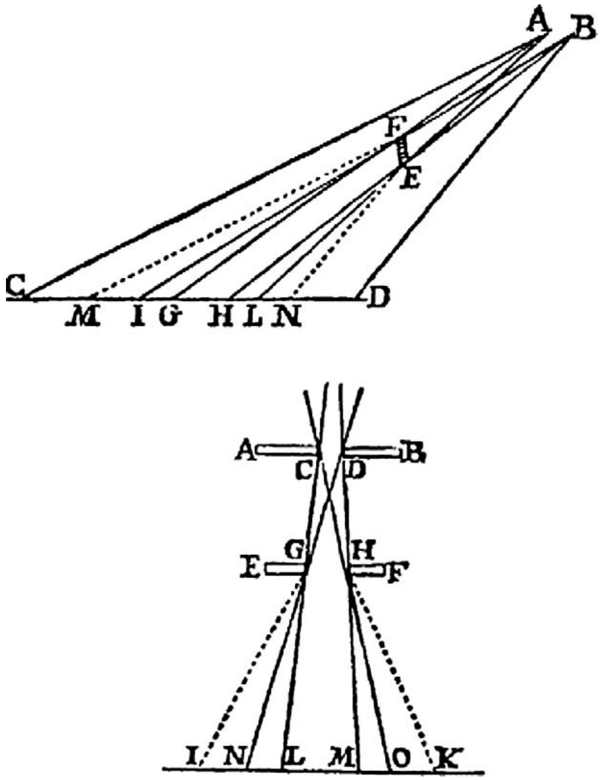


Fig. 2. Newton's explanation of the rainbow (Newton [1952], Book I, fig. 15).

In Proposition XIII, Newton states that "... light is in Fits of easy Reflexion and easy Transmission, before its Incidence on Transparent Bodies ... For these Fits are of a lasting nature" – an explicit assertion of the dual nature of light. Newton's value ("$\frac{1}{89\,000}$ of an Inch") for the "Interval of Fits" (half-wavelength) of light "in the Confine of yellow and orange", derived from his measurements of Newton's rings, is quite acceptable by present standards.

As an application of his theory of dispersion, Newton explained the colors of the rainbow (for contributions from his predecessors, see Boyer [1987]). His diagram (fig. 2) shows that the primary bow is formed by rays undergoing a single internal reflection within water droplets, and the secondary bow is formed by rays that undergo two internal reflections. The dark band between primary and secondary bows was first reported by the Greek philosopher Alexander of Aphrodisias circa 200 AD. Newton computed the angular width of the rainbow arcs, taking into account the angular diameter of the sun, and verified the results by his own measurements.

In Book III of his "Opticks", Newton discusses the diffraction of light (that he calls "Inflexion"). The discoverer of diffraction (who also gave it that name) was the Jesuit father Francesco Maria Grimaldi, who described his observations in a posthumously published book (Grimaldi [1665]). In one of them (top illustration in fig. 3(a)) he had a small opaque body lit by a pencil of sun rays and noticed the

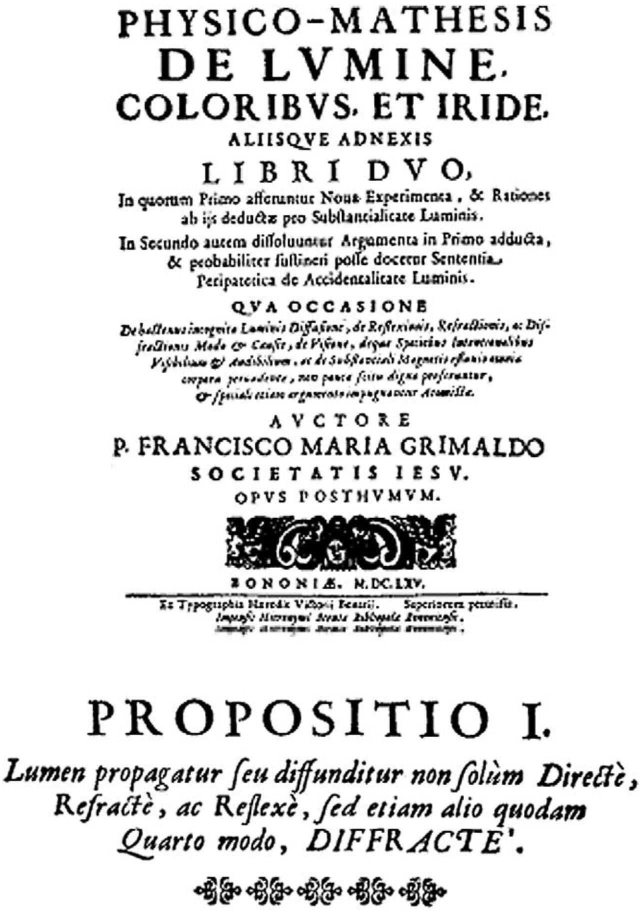


(a)

Fig. 3. (a) Grimaldi's experiments on diffraction (Grimaldi [1665]). (b) Title-page of his book and reproduction of Proposition I (*next page*).

presence of three parallel fringes beyond the penumbra (regions CM and ND). In the other one (bottom illustration in fig. 3(a)), a cone of light going through two diaphragms spreads beyond the domain ON predicted by geometrical optics. His Proposition I (fig. 3(b), bottom) states that "Light propagates and spreads not only directly, through refraction, and reflection, but also by a fourth mode, diffraction".

Newton referred to Grimaldi's experiments at the beginning of Book III and then reported his own observations of the shadows of hairs, pins, knife-edges and other very thin or sharp objects. The high accuracy of these experiments has been verified by recent tests (Nauenberg [2000]). Fresnel blamed Newton for not reporting diffraction fringes inside the shadow, but only those outside, as stated in "Opticks", Query 28:


PHYSICO-MATHESIS
DE LVMINE,
COLORIBVS, ET IRIDE,
ALIISQVE ADNEXIS
LIBRI DVO,
In quorum Primo afferuntur Noua Experimenta, & Rationes ab iis deducta pro Substantialitate Luminis.
In Secundo autem dissoluuntur Argumenta in Primo adducta, & probabiliter sustineri posse docetur Sententia Peripatetica de Accidentalitate Luminis.
QVA OCCASIONE
De hactenus incognita Luminis Diffusione, de Reflexionis, Refractionis, ac Diffractionis Modo & Causis, de Visione, deque Speciebus Intentionalibus Visibilium & Audibilium, ac de Substantiali Magnetis effluvio omnia corpora peruadente, non pauca scitu digna proferuntur, & speciali etiam argumento impugnantur Atomistae.
AVCTORE
P. FRANCISCO MARIA GRIMALDO
SOCIETATIS IESV.
OPVS POSTHVMVM.
BONONIAE. M.DC.LXV.
Ex Typographia Haeredis Victorij Benatij. Superiorum permissu.

PROPOSITIO I.
Lumen propagatur seu diffunditur non solùm Directè, Refractè, ac Reflexè, sed etiam alio quodam Quarto modo, DIFFRACTÈ.

(b)

Fig. 3. (*Continued.*)

"The Rays which pass very near to the edges of any Body, are bent a little by the action of the Body, as we shew'd above; but this bending is not towards but from the Shadow."

However, by reproducing Newton's experimental conditions, it has been verified (Stuewer [2006]) that the intensity of the diffraction lines inside the geometrical shadow was far too low for Newton to have observed them visually.

What about the explanation of diffraction? As Newton states in Book III,

"I was then interrupted, and cannot now think of taking these things into farther Consideration. And since I have not finish'd this part of my Design, I shall conclude, with proposing only some Queries in order to a farther search to be made by others:

Query 1. Do not Bodies act upon Light at a distance, and by their action bend its Rays, and is not this action (*cæteris paribus*) strongest at the least distance?

Qu. 2. Do not the Rays which differ in Refrangibility differ also in Flexibility, and are they not by their different Inflexions separated from one another, so as after separation to make the Colours in the three Fringes above described? And after what manner are they inflected to make those Fringes?

Qu. 3. Are not the Rays of Light in passing by the edges and sides of Bodies, bent several times backwards and forwards, with a motion like that of an Eel? And do not the three Fringes of colour'd Light above mention'd, arise from three such bendings?"

An illustration of Newton's proposed explanation appears in the above-mentioned Scholium in the "Principia" (fig. 4), representing diffraction by a knife-edge. According to Query 1, light rays passing outside the edge are bent into the shadow, the amount of their "Inflexion" increasing as they come closer to the edge. In Query 2, Newton proposes that "Inflexion", like refractive index, depends on colour. As will be seen below, contemporary diffraction theory provides qualitative support for all three of these queries.

Newton also conjectured that an increasing optical density of the "Aetherial Medium" might be responsible for the postulated action at a distance:

"*Qu*. 20. . . . And doth not the gradual condensation of this Medium extend to some distance from the Bodies, and thereby cause the Inflexions of the Rays

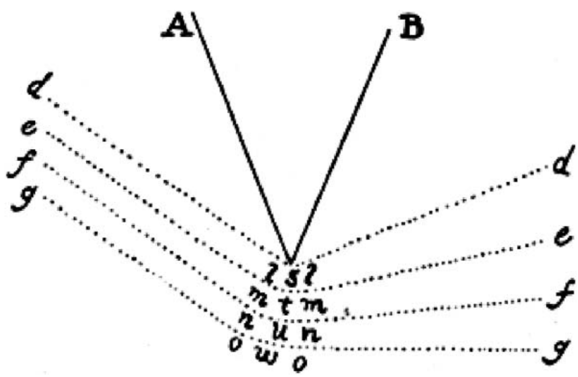


Fig. 4. Newton's interpretation of diffraction by a knife edge (Newton [1946], Book I, Proposition XCVI, Scholium). He states that ". . . the rays which fall upon the knife are first inflected in the air before they touch the knife".

of Light, which pass by the edges of dense Bodies, at some distance from the Bodies?"

In 1657, Pierre de Fermat formulated his Principle of Least Time (Fermat [1891]):

" The path really followed by light from a point A to a point B corresponds to the shortest possible time",
from which he derived the law of refraction (for reflection, this had already been done by Heron of Alexandria). Taking into account the inverse relation between velocity of light and refractive index n, and that travel time in general is stationary, not necessarily a minimum (Born and Wolf [1999]), this amounts to stationarity of the optical path between A and B,

$$\delta \int_A^B n \, \mathrm{d}s = 0, \tag{1.1}$$

where ds is the element of arc length along the path.

Newton's contemporary Christian Huygens explicitly proposed a wave theory of light (Huygens [1690]). This was actually a theory of the propagation of light pulses, excluding any idea of periodicity. Employing a billiard-ball model of the ether as scaffolding, he asserts that, in propagation, each ether particle collides with all those that surround it, so that "... around each particle there is made a wave of which that particle is the centre". This leads him to the celebrated "Huygens' Principle", illustrated in fig. 5, taken from his treatise.

Each point of a pulse front, according to Huygens, gives rise to secondary spherical pulses, the envelope of which at a given time is the propagated pulse. Secondary pulses are only effective at their point of contact with the envelope, which explains rectilinear propagation (fig. 5):

"For if, for example, there were an opening BG, limited by opaque bodies BH, GI, the wave of light which issues from the point A will always be terminated by the straight lines AC, AE, as has just been shown; the parts of the partial waves which spread outside the space ACE being too feeble to produce light there."

In contrast with Huygens' pulse theory of light, a theory of light based on oscillatory wave propagation in an elastic ether was apparently first proposed by Leonhard Euler (Euler [1746]), who emphasized the analogy with sound propagation and employed it as well to account for diffraction.

He was also responsible for a fundamental contribution to the mathematical description of wave tunneling, Euler's formula (Euler [1748]),

$$\mathrm{e}^{\mathrm{i}x} = \cos x + \mathrm{i} \sin x, \tag{1.2}$$

which connects the exponential and trigonometric functions.

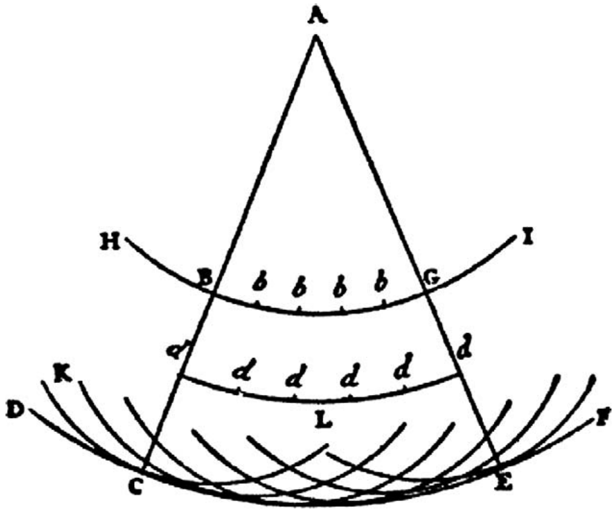


Fig. 5. Propagation of a spherical light pulse according to Huygens' Principle (Huygens [1690], fig. 6).

Its special case

$$e^{i\pi} + 1 = 0 \tag{1.3}$$

has been called "the most beautiful formula of mathematics", as it combines some of the most basic elements and operations of arithmetic, algebra, geometry and analysis.

§ 2. Classical diffraction theory

Light propagation in terms of oscillatory waves, pioneered by Euler, was taken up by Thomas Young, who formulated the associated basic principle of interference. He states it explicitly (Young [1802]):

"Radiant Light consists in Undulations of the luminiferous Ether"..."When two Undulations, from different origins, coincide either perfectly or very nearly in Direction, their joint effect is a Combination of the Motions belonging to each".

Young made several applications of his principle. Employing Newton's measurements of Newton's rings, he derived very accurate values for the wavelengths associated with different colors. He also applied it (Young [1804]) to explain the supernumerary rainbow arcs that appear just below the primary bow when relatively small and uniformly sized water droplets are present (Lee and Fraser

[2001]). However, as is discussed below, Young's interference theory of the rainbow must be amended to account for diffraction effects.

The celebrated two-slit interference experiment, illustrated with water waves, appears in fig. 6, taken from Young [1807]. An excellent modern survey of Young's contributions is given by Berry [2002].

Young's theory of diffraction differed from Newton's. He interpreted Newton's diffraction experiments with knife-edges in terms of "interference of the light reflected from the edges of the knives" (Young [1804]), so that in his view the diffracted light arises from sharp edges and object boundaries. Incident rays that just graze the edge of an aperture or obstacle would undergo "a kind of reflection". The resulting *boundary diffracted wave* would penetrate into the geometrical shadow and would interfere with incident and geometrically reflected waves in the illuminated region, accounting for the diffraction fringes observed.

The classical wave theory of light diffraction was formulated by Augustin-Jean Fresnel (Fresnel [1816]), who combined Huygens' principle with Young's principle of interference. According to Fresnel,

"... elementary waves arise at every point along the arc of the wave front passing the diffracter and mutually interfere. The problem was to determine the resultant vibration produced by all the wavelets reaching any point behind the diffracter."

Thus, what in Huygens' formulation was just a geometrical construction for wavefronts becomes a dynamical principle for wave propagation. Figure 7 illustrates the application of Fresnel's idea to diffraction of a monochromatic plane wave by an opaque half-plane. Each point of the unblocked portion of the incident plane wavefront becomes the source of a spherical secondary wave, as in Huygens' principle, giving rise to the geometrical shadow boundary (thin vertical line in fig. 7).

However, while Huygens just applied his envelope construction to get a transmitted cut-off pulse front (envelope), Fresnel lets all spherical wavelets interfere. Their resultant yields the diffracted wave, which propagates into the shadow and produces diffraction fringes around the shadow boundary. We will call Fresnel's explanation "diffraction as blocking".

Fresnel submitted his proposal (Fresnel [1816]) to the French Academy of Sciences as an entry in the competition for the Grand Prize offered for the explanation of diffraction. Most members of the award committee, chaired by Arago, favored the corpuscular theory, but the prize was awarded to Fresnel following a remarkable experimental confirmation, as reported by Arago:

"One of your commissioners, M. Poisson, had deduced from the integrals reported by [Fresnel] the singular result that the centre of the shadow of an opaque

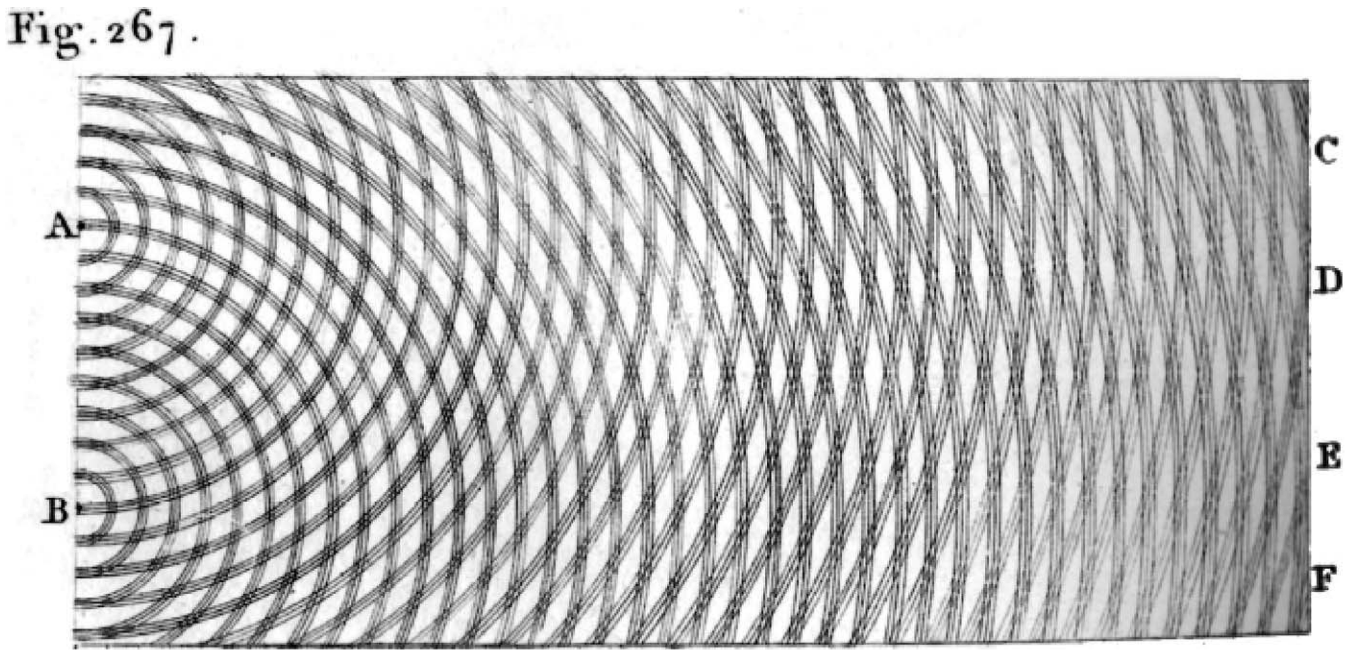


Fig. 6. Young's illustration of two-slit interference in terms of the pattern "obtained by throwing two stones of equal size into a pond at the same instant" (Young [1807]). He mentions the analogy with acoustics and optics.

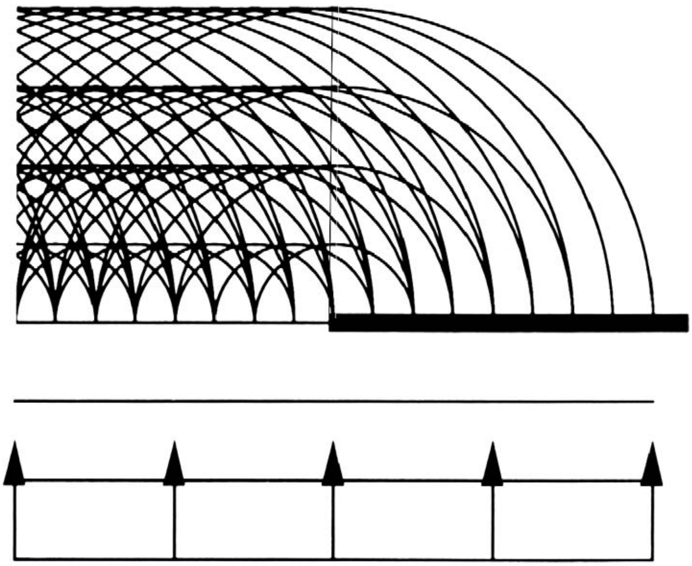


Fig. 7. According to Fresnel, diffraction by a half-plane arises from blocking of the incident plane wavefront. Secondary spherical waves propagate and interfere beyond the shadow boundary (thin vertical line).

circular screen must, when the rays penetrate there at incidences which are only a little more oblique, be just as illuminated as if the screen did not exist. The consequence has been submitted to the test of direct experiment, and observation has perfectly confirmed the calculation."

This bright spot at the center of the shadow of a circular disc became known as Poisson spot (a prior observation in 1723 by Giacomo Maraldi remained unrecognized).

A more precise mathematical formulation of the Huygens–Fresnel principle followed from the work of Gustav Kirchhoff (Kirchhoff [1882]) on scalar monochromatic wave propagation. Kirchhoff derived an exact integral representation of the wave function in terms of its boundary values over a surface (Born and Wolf [1999], Baker and Copson [1950]). For an aperture in an opaque plane screen, this surface may be taken as the plane of the screen. However, the exact boundary values are unknown.

The classical Fresnel–Kirchhoff theory of diffraction is based upon Kirchhoff's approximation, a perturbative assumption on the unknown boundary values, expected to hold when all relevant dimensions of the obstacle (aperture) as well as the wavefront are much larger than the wavelength. It is assumed that one can re-place the unknown boundary values by the incident wave over unobstructed parts of the wavefront, and that one can take them to vanish over obstructed (shadow) parts – as would be prescribed by geometrical optics.

An important application of classical diffraction theory was Airy's treatment of the rainbow problem (Airy [1838]). The primary rainbow direction is an extremal deflection angle, i.e., a caustic direction, for rays incident on a water droplet that undergo just one internal reflection (cf. Nussenzveig [1977, 1992]). Smaller scattering angles represent a shadow region for this class of rays (Alexander's dark band); at larger angles, two such rays with different impact parameters interfere, leading to Young's supernumerary arcs. The title of Airy's paper is "On the intensity of light in the neighbourhood of a caustic".

Airy applied Huygens–Fresnel theory to an S-shaped cubic wavefront within a droplet, approximating the unknown wave amplitude along it by a constant. The result was his "rainbow integral", now known as the Airy function $\mathrm{Ai}(z)$, which satisfies the differential equation

$$\mathrm{Ai}''(z) = z\mathrm{Ai}(z). \tag{2.1}$$

Figure 8 is a plot of $\mathrm{Ai}(x)$. For negative x, it has a slowly damped oscillatory behavior, with peaks related to the supernumerary arcs; for positive x, it undergoes faster than exponential damping, associated with penetration by diffraction on the shadow side of the rainbow.

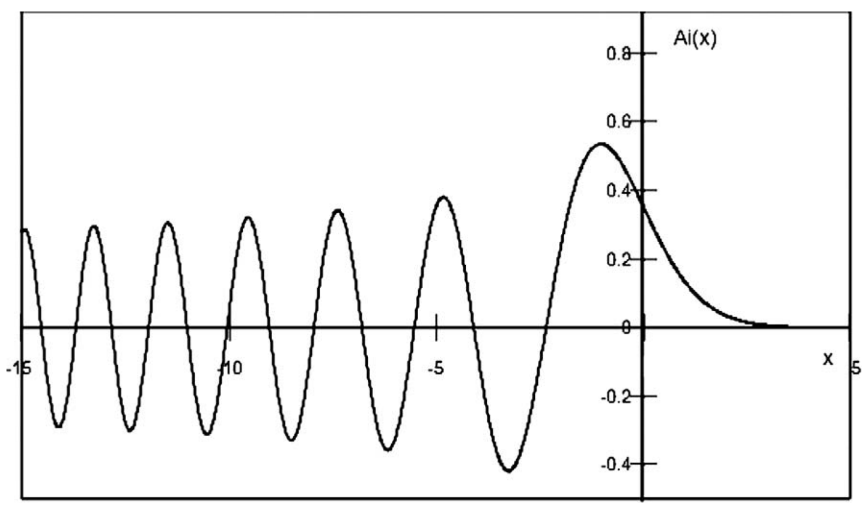


Fig. 8. Plot of $\mathrm{Ai}(x)$ for $-15 \leqslant x \leqslant 5$.

For diffraction by a (large) aperture, classical diffraction theory turns out to yield acceptable approximations. Why is that so? For scalar diffraction by an aperture in a plane screen, the diffraction amplitude (asymptotic amplitude of the outgoing spherical wave) in the direction of unit vector $\hat{\mathbf{s}}$ is given by Rayleigh's exact formula (Bouwkamp [1954])

$$f(\hat{\mathbf{s}}, \hat{\mathbf{s}}_0) = \frac{\hat{\mathbf{s}} \cdot \hat{\mathbf{s}}_0}{\mathrm{i}\lambda} \int \exp[-\mathrm{i}k(\hat{\mathbf{s}} - \hat{\mathbf{s}}_0) \cdot \mathbf{x}] u(\mathbf{x})\, \mathrm{d}^2 x, \tag{2.2}$$

where $\hat{\mathbf{s}}_0$ is the direction of the incident wave, $\lambda = 2\pi/k$ is the wavelength, the integral is extended over the plane of the screen, and $u(\mathbf{x})$ is the exact (unknown) wave function on this plane.

The short-wavelength assumption implies that dominant contributions to the amplitude arise from the near-forward domain, in which

$$\left|k(\hat{\mathbf{s}} - \hat{\mathbf{s}}_0) \cdot \mathbf{x}\right| \leqslant kd \sin\theta \sim k_\perp d \equiv 2\pi d/\lambda_\perp \ll 1, \tag{2.3}$$

where θ is the diffraction angle, d is the diameter of the aperture and $\lambda_\perp$ is the transverse wavelength.

Thus, the main diffraction pattern depends only on spatial Fourier components of $u(\mathbf{x})$ on the aperture plane with $\lambda_\perp \gg d$, which are insensitive to fine details of the aperture boundary values: they feel mainly the gross blocking effect captured by Kirchhoff's approximation (Nussenzveig [1959]). Finer details affect large diffraction angles, where the intensity is weaker.

Equivalently, one may extend the two-dimensional Fourier expansion of the exact wave function over the aperture plane (x, y) to the half-space $z > 0$ beyond it by the wave equation propagation factor $\exp(\mathrm{i}k_z z)$ $(k_x^2 + k_y^2 + k_z^2 = k^2)$, but this introduces evanescent waves, with $k_z = \mathrm{i}\sqrt{k_x^2 + k_y^2 - k^2}$, that decay exponentially with z. In this angular spectrum of plane waves (Nussenzveig [1959], Born and Wolf [1999]), finer details of the aperture distribution give rise to evanescent waves (for applications to microscopy, see Section 14.1.6). This already gives us a foretaste of light tunneling effects in diffraction.

Young's interpretation of diffraction is actually consistent with classical diffraction theory, in spite of the apparent differences. Indeed, the Fresnel–Kirchhoff representation of the diffracted wave as a surface integral over the unobstructed part of a wavefront can be converted into a line integral over the edge and interpreted in terms of Young's boundary diffraction wave (Rubinowicz [1917], Sommerfeld [1954]).

For other important contributions to the wave theory of light in the 19th century, culminating in Maxwell's electromagnetic theory, we refer to Born and Wolf [1999].

§ 3. The optomechanical analogy

Still in the first half of the 19th century, an important conceptual reformulation of optics was the development of the Hamiltonian analogy between geometrical optics and classical mechanics (Hamilton [1828]). The basis for Hamilton's reformulation is the optical path function between two points along a ray, that appears in (1.1),

$$[AB] = \int_A^B n \, \mathrm{d}s \equiv S(B) - S(A). \tag{3.1}$$

As stated by Hamilton,

"The mathematical novelty of my method consists in considering this quantity as a function of the co-ordinates of these extremities, which varies when they vary, according to a law which I have called the law of varying action; and in reducing all researches respecting optical systems of rays to the study of this single function."

The function $[AB]$ is *Hamilton's point characteristic*, and S is the *eikonal function*, which satisfies the relation

$$\nabla S = n\hat{\mathbf{u}}, \tag{3.2}$$

where $\hat{\mathbf{u}}$ is the ray direction (unit tangent vector). From (3.2) follows the *eikonal equation*

$$(\nabla S)^2 = n^2. \tag{3.3}$$

If one now considers the path of a point particle with mass m and energy E in a potential field $V(\mathbf{r})$, one finds that it can be obtained from the Hamilton–Jacobi equation (Goldstein [1957])

$$(\nabla W)^2 = 2m\big[E - V(\mathbf{r})\big], \tag{3.4}$$

provided that the initial position and the initial direction are specified, the same conditions required for (3.3).

Remembering that the refractive index is dimensionless, the analogy leads to

$$n(\mathbf{r}) = \sqrt{1 - \frac{V(\mathbf{r})}{E}}, \tag{3.5}$$

which is a real quantity in the domain accessible to the motion, $V(\mathbf{r}) < E$. Thus, the paths of the above-considered particle according to classical mechanics are identical to geometrical optic light rays in an inhomogeneous medium with refractive index (3.5).

Fermat's principle (1.1) is the analogue of Maupertuis' principle of least action. Further discussion of variational principles and of the application to electron optics can be found in Born and Wolf [1999].

The optomechanical analogy played an important role in the formulation of quantum mechanics. In one of his earliest communications on wave properties of matter, Louis de Broglie states (De Broglie [1923]): "The new dynamics of the free point particle is to the old dynamics ... as wave optics is to geometrical optics". Erwin Schrödinger made the same point in his second paper on wave mechanics (Schrödinger [1926]): "...our classical mechanics is the *complete* analogy of geometrical optics... Then it becomes a question of searching for an undulatory mechanics... working out... the Hamiltonian analogy on the lines of undulatory optics." He stated it succinctly as

"Ordinary mechanics : Wave mechanics = Geometrical optics : Undulatory optics".

The analogy can be employed in the reverse direction, by associating an effective potential field with a refractive index distribution. Light tunneling is thereby identified as the analogue of quantum tunneling through a potential barrier (Section 1).

§ 4. Modern developments in diffraction theory

4.1. The geometrical theory of diffraction

A beautiful heuristic extension of geometrical optics for treating diffraction problems, known as the geometrical theory of diffraction, was proposed by Keller [1962]. It is based upon the following main postulates:

(i) The diffracted field propagates along *diffracted rays*, that generalize the concepts of reflected and refracted rays, as well as Young's ideas on edge diffraction. They are determined by extending Fermat's Principle (1.1) of the stationary optical path, allowing the paths to include points or arcs on discontinuous boundaries, such as corners, edges or vertices.

For diffraction by a smooth obstacle, an example is illustrated in fig. 9: an incident ray PD, tangential to the surface at D, travels along a geodetic arc DT on the surface and leaves it again tangentially at T, propagating along a straight line to Q, a point in the shadow region.

(ii) The transport of amplitude and phase along a diffracted ray obeys the laws of geometrical optics.

(iii) The analogue of a reflection coefficient, relating the initial field on a diffracted ray to the incident field, is a *diffraction coefficient D*, that is supposed to

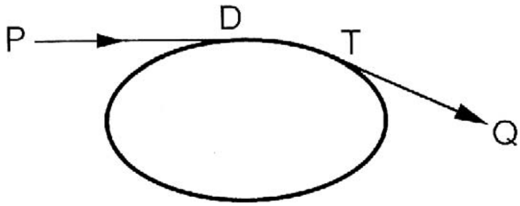


Fig. 9. The tangentially incident ray PD excites the diffracted ray DT, that leaves the surface tangentially at T to reach a point Q in the shadow.

be determined by the local geometry and nature of the boundary around the point of diffraction D.

An account of the history of the geometrical theory of diffraction is given by Keller [1979, 1985], who also discusses the shortcomings of his theory. In common with geometrical optics, it leads to singularities at focal points and caustics, including the surface of a smooth obstacle, which is a caustic of diffracted rays. As a consequence, the diffraction coefficients cannot be determined self-consistently within the theory.

To determine them, in accordance with (iii), the theory makes use of comparisons with the asymptotic short-wavelength behavior of solutions to *canonical problems* with locally similar geometry, such as diffraction by a sphere or cylinder. A survey of applications is given by Borovikov and Kinber [1994].

Transitions between domains covered by different numbers of rays are also not treated. This includes, in particular, the *penumbra region* around the light/shadow boundary.

4.2. Fock's theory of diffraction

Fock's treatment of the penumbra region, originating from his work on the propagation of radio waves around the Earth's surface (Fock [1965]) employs the conceptual picture of diffraction as *transverse diffusion*, developed by Leontovich and Fock [1946]. To explain this concept we note that, in geometrical optics, amplitude and phase are transported longitudinally along the rays, with no constraints in transverse directions (Sommerfeld [1954]), so that discontinuities at shadow boundaries are allowed. Diffraction is associated with smoothing of these discontinuities by transverse diffusion.

To see how this arises, consider monochromatic wave propagation along the z-direction and substitute the 'Ansatz'

$$u(x, y, z) = A(x, y, z) \exp(\mathrm{i}kz) \tag{4.1}$$

into the Helmholtz monochromatic wave equation. One finds

$$\frac{\partial A}{\partial z} - \frac{\mathrm{i}}{2k}\frac{\partial^2 A}{\partial z^2} = \frac{\mathrm{i}}{2k}\left(\frac{\partial^2 A}{\partial x^2} + \frac{\partial^2 A}{\partial y^2}\right), \tag{4.2}$$

where the second term on the left may be neglected for slow amplitude variation per wavelength. Traveling along with a wavefront, $z = ct$, we get

$$\frac{\partial A}{\partial t} = D\left(\frac{\partial^2 A}{\partial x^2} + \frac{\partial^2 A}{\partial y^2}\right), \qquad D \equiv \frac{\mathrm{i}c}{2k}, \tag{4.3}$$

which is Leontovich and Fock's "diffusion equation" in this simplest situation.

However, the fact that D is imaginary reveals that the proper analogy is with Schrödinger's equation, rather than the diffusion equation. Physically, what happens along the traveling wavefront is akin to the oscillatory spreading behavior of a free Schrödinger wave packet, rather than to stochastic diffusion. In view of the analogy between physical optics and wave mechanics discussed in Section 3, this is hardly surprising. As will be seen below, the Schrödinger analogy is indeed the correct one.

In Fock's theory of diffraction by a smooth body, in order to treat the penumbra, the body surface near the glancing ray is approximated by a parabolic surface, and an 'Ansatz' generalizing the above one is substituted into the Helmholtz equation in suitable coordinates, describing transverse behavior in terms of the "diffusion" analogy.

An integral representation for the solution to this local approximation leads to a new set of special functions known as *Fock functions*, related to complex Fourier transforms of the inverse of the Airy function. They are taken to represent the behavior in the penumbra region, interpolating between the illuminated and shadow domains. An exposition of Fock's theory and its applications is given by Babich and Kirpichnikova [1979].

However, as is discussed below, Fock's theory yields only a *transitional asymptotic approximation*, that cannot be extended very far on either side of the geometrical light/shadow boundary. What is needed is a *uniform asymptotic approximation*.

A very interesting independent treatment of the penumbra, equivalent to Fock's, was formulated by Pekeris [1947], in connection with microwave propagation around the Earth's horizon. He introduced an "Earth flattening transformation", representing the Earth's surface as flat, and simulating the curvature of the rays by an "atmosphere" in which the refractive index increases linearly with height above the surface. In view of (2.1), the connection with the Airy functions arises from this assumed linear dependence.

§ 5. Exactly soluble models

5.1. Exact solutions

The earliest exact solution of a diffraction problem was Sommerfeld's celebrated treatment of diffraction by a half-plane (Sommerfeld [1896, 1954]). Though physically unrealistic, because the half-plane is assumed to be infinitely thin yet opaque, it represents diffraction, for the first time, through analytic continuation. The solution is a uniform analytic function over a two-sheeted Riemann surface, with the diffracted wave penetrating into the second Riemann sheet.

The diffracted field behaves asymptotically like a cylindrical wave emanating from the edge, consistent with Young's picture of boundary diffraction. It is worth noting that Newton's Query 3 (Section 1) is also vindicated by plotting the energy flow close to the edge (Braunbek and Laukien [1952], Berry [2002]): there are indeed eel-like undulations.

Analytic continuation is also important in the solution of the integral-equation formulation of other diffraction problems by the Wiener–Hopf technique (Noble [1958]) as well as in more general scattering problems (Nussenzveig [1972], Wolf and Nieto-Vesperinas [1985]).

For a small number of separable geometries, exact solutions in the form of infinite series of eigenfunctions have been found (Bouwkamp [1954], Hönl, Maue and Westphal [1961]). However, as pointed out by Sommerfeld [1954],

"...a mathematical difficulty develops which quite generally is a drawback of this 'method of series development': for fairly large particles ($ka > 1, a =$ radius, $k = 2\pi/\lambda$) the series converge so slowly that they become practically useless".

The reason for this is discussed below.

5.2. Mie scattering

The keystone to a deeper understanding of the dynamics of diffraction, to which the rest of this work is mainly devoted, is a reformulation of a problem that has an exact solution in terms of an infinite series of eigenfunctions, allied to the unique feature of being physically realistic: the scattering of light by spherical particles.

A historical survey of the scattering of plane waves by a sphere has been prepared by Logan [1965]. An early treatment was that of Clebsch [1863], who dealt with elastic waves incident on a rigid sphere. Lord Rayleigh [1872] found the exact series solution for the scattering of sound by a sphere.

The corresponding solution for scattering of a plane electromagnetic wave by a transparent sphere was given by Lorenz [1890] (cf. Keller [2002]) and rediscovered by Mie [1908], who was concerned with explaining the colors of colloidal suspensions of metallic particles. It was also treated by Debye [1909a] in his doctoral thesis about the radiation pressure of an electromagnetic wave on a metallic sphere. Historically, referring to Lorenz–Mie scattering is better justified, but we stick with the more usual designation Mie scattering.

The Mie solution is a partial-wave series

$$S_j(\beta,\theta) = \frac{1}{2}\sum_{l=1}^{\infty}\left\{\left[1 - S_l^{(j)}(\beta)\right]t_l(\theta) + \left[1 - S_l^{(i)}(\beta)\right]p_l(\theta)\right\}$$
$$(j = 1, 2;\ i \neq j), \tag{5.1}$$

where S_j are the polarized scattering amplitudes in the direction θ for perpendicular ($j = 1$) and parallel ($j = 2$) polarizations (with respect to the scattering plane), $\beta \equiv ka$ is the size parameter, $S_l^{(i)}(\beta)$ are the S-matrix elements for magnetic ($i = 1$) and electric ($i = 2$) multipoles of order l, and t_l and p_l are angular Legendre-type functions. The S-matrix elements are rational combinations of spherical Bessel and Hankel functions and their derivatives (for their expressions and the notation, see Nussenzveig [1992]).

5.3. *The localization principle*

In quantum mechanics, the index l in the partial-wave series also labels eigenvalues of the orbital angular momentum, that are given by

$$\sqrt{l(l+1)}\hbar = \sqrt{\left(l+\frac{1}{2}\right)^2 - \frac{1}{4}}\hbar \approx \left(l+\frac{1}{2}\right)\hbar \quad \text{for } l \gg 1. \tag{5.2}$$

In semiclassical (short-wavelength) scattering (Berry and Mount [1972]), large l prevails, and the above orbital angular momentum can be associated with a corresponding impact parameter b_l by

$$b_l = \frac{(l+\frac{1}{2})\hbar}{p} = \frac{(l+\frac{1}{2})}{k}, \tag{5.3}$$

where $p = \hbar k$ is the linear momentum.

This association between a partial-wave term and the impact parameter of an incident path, already implicit in Debye's thesis, was emphasized by Van De Hulst [1957] (cf. also Roll, Kaiser, Lange and Schweiger [1998]), and it is known as the *localization principle*. A closely related wave-mechanical picture is a description of the incident beam on a transverse plane in terms of annular rings centered on the scatterer, with areas $\hbar$ consistent with the uncertainty relation, b_l being the average radius of the ring between the values l and $l+1$ (Blatt and Weisskopf [1952]).

In geometrical optics, only incident rays with impact parameters $\leqslant a$ are scattered by the sphere, so that the partial-wave series (5.1) may be cut off at a maximum value of l given by $l_{\max} \leqslant ka = \beta$. Partial waves with $l > l_{\max}$ are associated with *above-edge* incident rays. For solar radiation and cloud water and aerosol droplets, β reaches $\sim$10 000, and numerical convergence of the Mie series does not set in until values actually exceeding the above estimate for the cut off are reached, because of above-edge effects to be discussed below. This justifies Sommerfeld's comment on the convergence problem.

§ 6. Watson's transformation

An important historical incentive for trying to overcome the convergence problem of the Mie series was the treatment of radiowave propagation around the Earth, beyond the horizon. As noted by Love [1915], the value of β in this situation is of order 10^4. During the first two decades of the 20th century, rival proposals to account for the propagation were the diffraction theories and those invoking ionosphere reflection (for a detailed account, see Yeang [2003]). Diffraction theories modeled the Earth as a perfectly conducting sphere.

Henri Poincaré worked on the diffraction model from 1909 until his death in 1912, treating it in nine papers and a monograph (Poincaré [1910]). His basic idea was to convert the partial-wave series into a contour integral using the residue theorem and to obtain an asymptotic approximation of this integral.

An improved version of Poincaré's approach was developed by Watson [1918]. His transformation of the partial-wave series is equivalent to the formula

$$\sum_{l=0}^{\infty} \phi\left(l+\frac{1}{2}, x\right) = \frac{1}{2} \int_C \phi(\lambda, x) \frac{\exp(-i\pi\lambda)}{\cos(\pi\lambda)} \, d\lambda, \tag{6.1}$$

where C is a contour encircling the positive real axis and $\phi(\lambda, x)$ is a holomorphic function of λ within C that takes on the values $\phi(l+\frac{1}{2}, x)$ at the half-integers. For the Mie series, analyticity of $\phi(\lambda, x)$ follows from the properties of cylindrical and Legendre functions.

Watson deformed the path C onto a symmetric neighbourhood of the imaginary λ axis, over which the integral vanishes, because the integrand is odd. The deformation is allowed in the shadow region. In this process, the only singularities met are poles (the integrand is meromorphic), so that the transformation results in a *residue series* over the poles. The residues represent *surface waves*, traveling around the sphere and getting damped by tangential shedding of radiation, like Keller's diffracted rays in fig. 9. The angular damping is associated with the imaginary part of the poles.

Watson found that this damping was too large to account for radiowave propagation into the shadow, and in a subsequent paper (Watson [1919]) he extended his treatment to include reflection from the ionosphere, modeled as a concentric spherical surface. He showed that the results were consistent with experimental observations.

Watson was concerned with radiation from a dipole antenna on the Earth's surface. The extension of his method to the scattering of a plane electromagnetic wave by a perfectly conducting sphere was undertaken by White [1922]. Watson's original path deformation does not converge in the illuminated region. White showed that, in this region, the contour can be deformed to yield, besides the residue series, a path through a saddle point. For the asymptotic evaluation of the remaining integral, he used the method of steepest descents, that had been introduced by Debye [1909b] to deal with the analogue of the Mie series in scattering by a circular cylinder (Debye [1908]). The saddle-point contribution yields the Wentzel–Kramers–Brillouin (WKB) approximation to the reflected wave, previously obtained by Nicholson [1910].

The WKB approximation is often improperly referred to as "geometrical optics": strictly speaking, geometrical optics describes only the propagation of intensity, not phase. However, it is customary to refer to the WKB dominant term as geometrical optics (or ray optics) approximation. For the development of the WKB method, see Berry and Mount [1972] and Keller [1979]. The application to electromagnetic theory was systematized in 1944 by Luneburg (Luneburg [1964]).

The application of Watson's transformation to radio wave propagation was further elaborated by Van Der Pol and Bremmer [1937], and Bremmer [1949]. Watson's surface waves, referred to as "creeping waves", as well as White's extension to the illuminated region, were rederived by Franz and Depperman [1952], and also extended to transparent spheres (Beckmann and Franz [1957], Franz [1957]).

Applications of complex angular momentum to quantum potential scattering and high-energy physics followed from the work of Regge (1959, 1960) (see Newton [1964] and De Alfaro and Regge [1965]). Poles of the scattering amplitude in the λ plane, like those found by Watson, became known as *Regge poles* (cf. Nussenzveig [1970, 1972]).

§ 7. CAM theory of Mie scattering

7.1. The Poisson sum formula

An improved version of complex angular momentum theory, hereafter referred to as CAM, was developed by Nussenzveig [1965]. Its applications to Mie scattering, with special reference to light tunneling, will be reviewed in the remainder of this work. For more detailed expositions, see Nussenzveig [1992] (and the original references therein), Grandy [2000] and Adam [2002].

The starting point, instead of (6.1), is the Poisson sum formula (Titchmarsh [1937]), apparently first employed in this connection by Bremmer [1949],

$$\sum_{l=0}^{\infty} \phi\left(l + \frac{1}{2}, x\right) = \sum_{m=-\infty}^{\infty} \int_{0}^{\infty} \exp\left[2im\pi\left(\lambda + \frac{1}{2}\right)\right] \phi(\lambda, x)\, \mathrm{d}\lambda. \tag{7.1}$$

A beautiful physical interpretation of this Poisson representation, applied to the partial-wave expansion of the scattering amplitude in semiclassical scattering by a central potential, has been proposed by Berry and Mount [1972]. Substituting the partial-wave amplitudes by their WKB approximations, and Legendre polynomials by their asymptotic expansions (near-forward and near-backward directions excluded), and applying the stationary phase approximation to the integrals, they reinterpret (7.1) as a *sum over paths*.

Stationary phase points are roots of

$$\Theta(\bar{\lambda}) = -(2m\pi \pm \theta), \tag{7.2}$$

where Θ is the classical deflection angle associated with a classical path that takes m turns about the origin. The corresponding phase of the integrand becomes the classical action along this path.

The Poisson representation then represents a sum over a family of classical paths with continuous angular momenta. The envelope of all such paths is a *caustic sphere* with radius $r_0(\lambda)$, the classical distance of closest approach (outermost radial turning point) for angular momentum λ. The caustic is a singular solution of the equations of motion. A particle with any λ may undergo any deflection Θ by following a *pseudoclassical path*, formed by joining an incoming classical path from infinity to an outgoing one, through a piece that "coasts" along the caustic sphere and takes any number of turns about the origin.

Keller's diffracted rays represent a special case of pseudoclassical paths, in which the caustic sphere is the surface of the scatterer. The Poisson representation stands halfway between classical and wave mechanics: summing over stationary phase points yields classical path contributions, whereas summing over the topological number m leads back to "quantized" $\lambda = l + \frac{1}{2}$.

7.2. Basic tools of CAM theory

The CAM theory of Mie scattering employs the following basic tools:

(i) ***The optomechanical analogy –*** We denote by N the refractive index of the sphere. According to (3.5), the corresponding potential function is given by

$$V(r) = \begin{cases} -(N^2 - 1)k^2 & (0 \leqslant r < a), \\ 0, & (r > a) \end{cases} \tag{7.3}$$

so that it represents a rectangular well for $N > 1$ and a rectangular barrier for $N < 1$.

The partial-wave S-function for magnetic multipoles is in fact *identical* to that for quantum scattering by the potential (7.3), so that for perpendicular polarization the optical and mechanical problems are essentially the same.

(ii) ***The effective potential –*** The radial equation for the lth multipole wave yields an effective radial potential

$$U_\lambda(r) = V(r) + \frac{\lambda^2}{r^2}, \tag{7.4}$$

where $V(r)$ is given by (7.3) and the last term represents the centrifugal barrier.

(iii) ***The localization principle –*** According to (5.3), for $\beta \gg 1$, we associate angular momentum λ with incident rays having impact parameter

$$b_\lambda = \lambda / k. \tag{7.5}$$

(iv) ***Path deformations –*** Like the Poincaré–Watson method, the CAM treatment works with path deformations of the Fourier integrals of (7.1) in the λ plane. In contrast with the partial-wave series, in which a large number of terms contribute, the idea is to employ the path deformations so as to collect dominant asymptotic contributions from a small number of *critical points* in the λ plane. As happened with White's extension of the Watson transformation (Section 6), different path deformations are required in different spatial regions (e.g., shadow and illuminated regions).

(v) ***Critical points –*** The following are the main types of critical points:

(a) *Real saddle points*. These are also stationary-phase points, so that they are associated with classical paths, i.e., geometrical-optic rays. The evaluation of their asymptotic contribution by the saddle point method yields the *WKB approximation* (including higher-order correction terms).

(b) *Complex saddle points*. These are associated with *complex rays* (Chapman, Lawry, Ockendon and Tew [1997], Kravtsov, Forbes and Asatryan [1999]), that represent *dynamical tunneling effects* usually disregarded by classical diffraction theory.

(c) *Complex poles*. For the polarized amplitudes, these are Regge poles, associated with resonances. They are also related to tunneling, as will be seen.

(d) *Uniformity*. The need (iv) to employ different path deformations in different regions brings out the problem of transitions between them – where diffraction effects are found (e.g., light/shadow). To obtain smooth joining, it is essential to employ *uniform asymptotic expansions* (Berry [1969], Ludwig [1970], Olver [1974]) for the special functions that appear in the Poisson representation.

§ 8. Impenetrable sphere

8.1. Structure of the wave function

All relevant features are already apparent for a scalar field, which may represent scattering by a quantum hard sphere (Dirichlet boundary condition). It is useful to describe first the CAM results for the structure of the wave function at finite distances (Nussenzveig [1965]). The subdivision into spatial regions is represented in fig. 10.

The *deep shadow region* is the domain (shaded in fig. 10) $r/a \ll \beta^{1/3}$, $\theta_0 - \theta \gg \beta^{-1/3}$, where $\theta_0 = \sin^{-1}(a/r)$ is the shadow boundary angle. This is the only domain where the original Watson transformation is rapidly convergent, yielding exponentially damped Regge pole contributions associated with surface waves and diffracted rays. Much weaker damping is found in the shadow of a circular disc (Jones [1964]), signaling the failure of classical diffraction theory, which predicts identical behavior in both cases.

In the *deep illumination* (lit region), $\theta - \theta_0 \gg \beta^{-1/3}$, the dominant term is the WKB approximation, that includes the incident and geometrically reflected waves. Near the shadow boundary, $|\theta - \theta_0| \ll \beta^{-1/3}$, $\beta^{-1/3} \ll z/a \ll \beta^{1/3}$ (*Fresnel region* in fig. 10), we find the classical Fresnel pattern of a single straight edge, rather than that of a slit (as classical diffraction theory would predict).

In the *Fresnel–Lommel region*, $\theta \leqslant \theta_0$, $\beta^{1/3} \ll r/a \ll \beta$, classical diffraction theory does work: we get the classical Fresnel pattern of a circular disc (Lommel [1885]). In particular, along the axis, one finds the Poisson spot, reproducing the

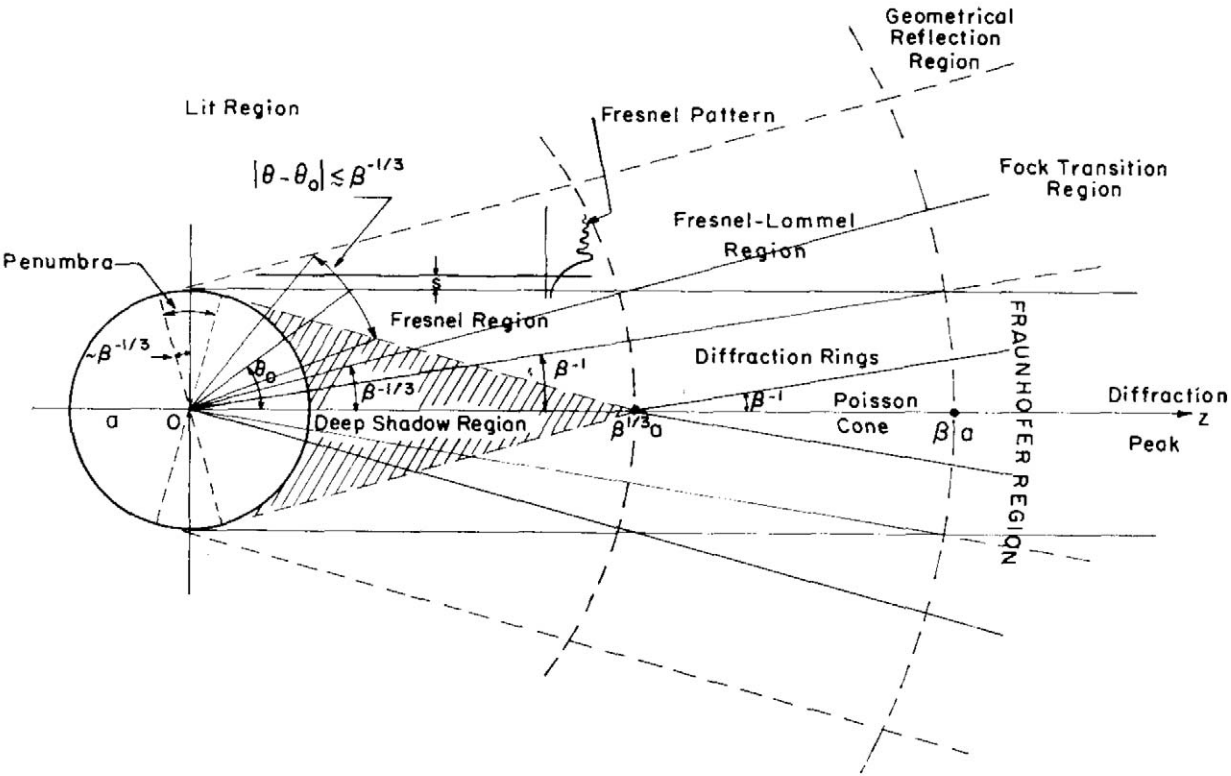


Fig. 10. Subdivision into spatial regions of the field scattered by an impenetrable sphere.

incident intensity. However, in contrast with the circular disc, it only develops at large axial distances, of order $\beta^{1/3}a$. It grows into a cone with angular opening of order β^{-1}, surrounded by diffraction rings.

In the *Fraunhofer region*, $r/a \gg \beta$, the field may be described in terms of the *total scattering amplitude* $f(k,\theta)$. Up to $\theta \sim \beta^{-1}$, the region of the *forward diffraction peak*, the amplitude is dominated by the classical Airy pattern,

$$f(k,\theta) \approx \mathrm{i}a \frac{J_1(\beta \sin\theta)}{\beta \sin\theta}. \tag{8.1}$$

For $\theta \gg \beta^{-1}$, the *geometrical reflection region* (fig. 10), the WKB approximation holds, yielding the reflected wave.

The region between the forward peak and the geometrical reflection region is the *penumbra*. A transitional asymptotic approximation in this domain yields Fock-type functions. However, it does not quite bridge the gap between the adjacent regions. For this purpose, a uniform approximation is required.

8.2. *Diffraction as tunneling*

The effective potential $U_\lambda(r)$ associated with an impenetrable sphere by (7.4) is an infinitely high wall surrounded by the centrifugal barrier. It is shown in fig. 11(i) for three values of λ, together with the associated incident rays, shown in fig. 11(ii), employing the localization principle to determine the corresponding

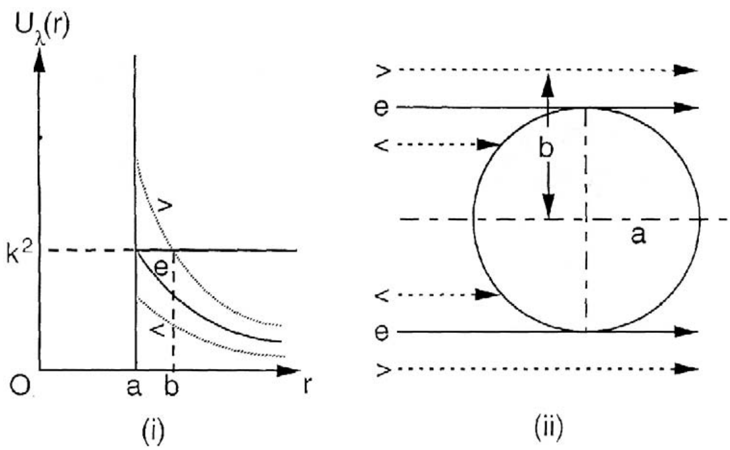


Fig. 11. (i) Effective potential for an impenetrable sphere and three different angular momenta λ: (<) below-edge; (e) edge; (>) above-edge, with turning point b. (ii) The corresponding incident rays.

impact parameter b_λ. The edge ray ($b_\lambda = a$) is associated with $\lambda = \beta$, for which the "energy" k^2 lies exactly at the barrier top. For below-edge rays ($b_\lambda < a$), k^2 is above the barrier.

For above-edge rays ($b_\lambda > a$), k^2 meets the barrier at the turning point $r = b_\lambda$. Such rays do not interact with the sphere according to geometrical optics. However, in the *above-edge domain* $\lambda - \beta = \mathrm{O}(\beta^{1/3})$, *tunneling* through the centrifugal barrier to the sphere surface leads to appreciable interaction and yields an important contribution to the scattering amplitude.

Similarly, in the *below-edge domain* $\beta - \lambda = \mathrm{O}(\beta^{1/3})$, geometrical-optic reflection from the surface is strongly distorted by interaction with the centrifugal barrier (*anomalous reflection*). The associated "peculiar interference effect" is discussed by Van De Hulst [1957], Section 17.21.

In the penumbra angular region $\theta = \mathrm{O}(\beta^{-1/3})$, which is much broader than the forward peak angular width $\theta = \mathrm{O}(\beta^{-1})$, both effects are important. Fock's theory of diffraction amounts to approximating the region near the top of the centrifugal barrier by a linear potential (equivalent to Pekeris's "Earth flattening transformation"). In view of eq. (2.1), this leads to the Airy functions that appear in Fock-type diffraction integrals. The Leontovich–Fock diffusion picture is misleading: the natural analogy is with the Schrödinger equation.

Fock's theory yields a transitional approximation because it neglects the curvature of the centrifugal barrier. CAM theory (Nussenzveig and Wiscombe (1987, 1991), Nussenzveig [1988]) leads to uniform asymptotic approximations, that match smoothly with WKB results in the geometrical reflection region.

It is possible to separate the effects of the "diffraction as blocking" picture of classical diffraction theory (Section 2) from the new contributions of tunneling and anomalous reflection by employing the localization principle to isolate above-edge and below-edge effects.

To illustrate how blocking is connected with partial-wave contributions only from rays that hit the sphere, we note that their contribution, according to Babinet's principle (Sommerfeld [1954]), is the same as that from a circular aperture, for which rays go through without interaction, yielding the forward diffraction Airy pattern

$$\sum_{l=0}^{[\beta]}\left(l+\frac{1}{2}\right)P_l(\cos\theta) = \beta\frac{J_1(\beta\theta)}{\theta} + \mathrm{O}\bigl(\beta^{-1}\bigr), \quad \beta \gg 1, \tag{8.2}$$

where $[\beta]$ denotes the largest integer contained in β and we have approximated the sum by an integral. Thus, the blocking effect arises from the cut off in the partial-wave series at $[\beta]$.

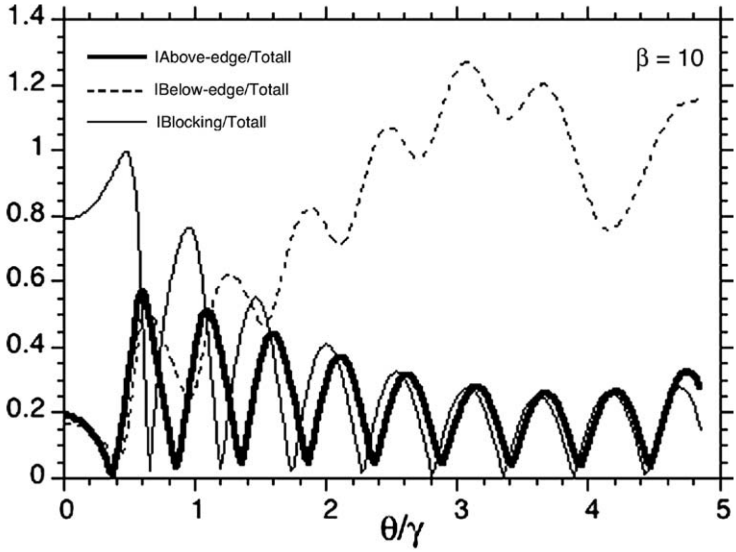


Fig. 12. Relative contributions from classical diffraction (blocking), tunneling (above-edge) and anomalous reflection (below-edge) in scattering by an impenetrable sphere; $\gamma \equiv (2/\beta)^{1/3}$. Reprinted with permission from Nussenzveig and Wiscombe [1991]. © 1991 by the American Physical Society.

The absolute values of the relative contributions from blocking, tunneling and anomalous reflection are plotted in fig. 12, where $\gamma \equiv (2/\beta)^{1/3}$, for $\beta = 10$. Classical diffraction (blocking) is dominant within the central part of the forward diffraction peak, but the three effects become comparable already at $\theta \sim \beta^{-1/3}$. At larger scattering angles, tunneling contributes as much as classical diffraction.

The tunneling range of impact parameters above the edge that contributes significantly is given by

$$b - a = \mathrm{O}\big(\lambda_0^2 a\big)^{1/3}, \tag{8.3}$$

where $\lambda_0 \equiv 2\pi/k$ is the wavelength. This weighted geometric average between wavelength and size may be thought of as the range, for an impenetrable sphere, of the effective "action at a distance" conjectured by Newton in his Query 1 (Section 1).

How much improvement is brought by the uniform CAM approximation is illustrated by the error plots in fig. 13, which represent percent errors in the extinction efficiency (ratio of the total cross-section to the geometrical cross-section) in acoustic scattering by a rigid sphere (Nussenzveig and Wiscombe [1987]), with size parameters ranging from 1 to 10. The CAM percent error is three to four

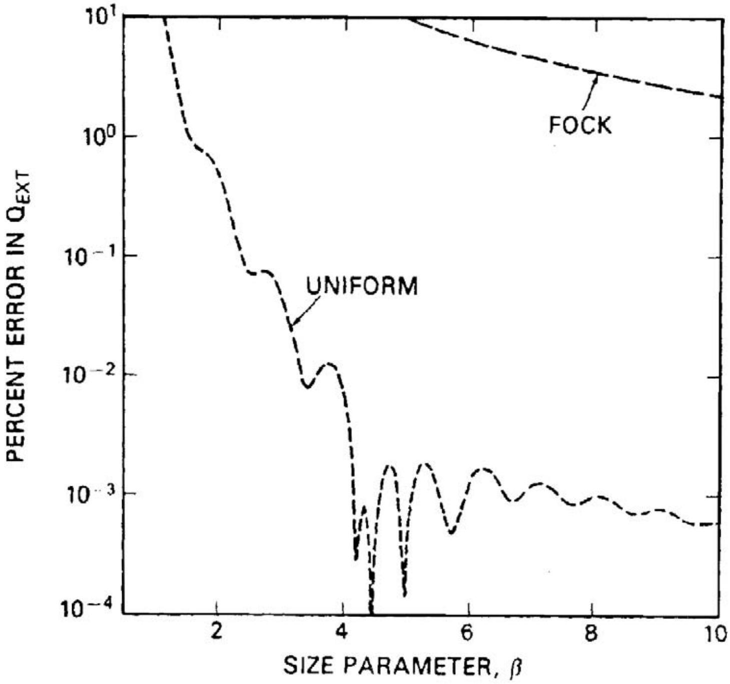


Fig. 13. Comparison between percent errors in the uniform CAM approximation and in the Fock approximation to the extinction cross-section for acoustic scattering by a rigid sphere, in the range $1 < \beta < 10$. Reprinted with permission from Nussenzveig and Wiscombe [1987]. © 1987 by the American Physical Society.

orders of magnitude smaller than that of the Fock theory, which was the best previous approximation. It is of order 10% even at $\beta = 1$ and only a few ppm at $\beta = 10$. For $\beta = 100$, it is found that the uniform CAM approximation is more accurate, for all scattering angles, than typical "exact" results from numerical partial-wave summations (1 ppm).

§ 9. Near-critical scattering

For transparent spheres with relative refractive index $N < 1$, a new diffraction effect occurs. It is manifested, for instance, in clouds of air bubbles in seawater, seen near the critical angle for total reflection: colored bands in the scattered light were observed under these circumstances by Pulfrich [1888], who compared them with a rainbow. A physical optics approximation to explain this *near-critical scat-*

tering effect, analogous to Airy's classical diffraction theory of the rainbow, was proposed by Marston [1979].

Because the Fresnel reflectivities approach total reflection with vertical slope (Born and Wolf [1999]), the scattered intensity according to geometrical optics undergoes a sharp break in slope at the critical scattering angle, a singularity which may be called a *weak caustic*.

The effective potential (7.3), (7.4) for this problem (rectangular plus centrifugal barrier for a given λ) is shown in fig. 14(a), together with four "energy levels" k^2 related with different impact parameters b and with the angle of incidence θ_1 through the localization principle $b = \lambda/k = a \sin\theta_1$. The corresponding situations are shown in fig. 14(b).

In situation 1, θ_1 is below the critical angle and the incident ray gets inside the sphere, where it undergoes multiple reflections. The radial turning point $r = b/N$ gets closer to the surface as *critical incidence* (situation 2) is approached. Situations 3 and 4 have turning points at the surface $r = a$, corresponding to total reflection; for the grazing ray 4, effects similar to those found for an impenetrable sphere take place.

The new effect occurs in situation 2. Critical incidence on a plane interface gives rise to a lateral displacement of the reflected beam, the *Goos–Hänchen shift* (Goos and Hänchen [1947], Bryngdahl [1973]). Its origin is directly connected with Newton's discovery of tunneling (Section 1): an evanescent wave penetrates into the rarer medium, running along the interface. A similar effect takes place in situation 2, but it is modified by the interface curvature (centrifugal barrier), leading to a *spherical Goos–Hänchen angular displacement* of the totally reflected beam, another light-tunneling effect.

The CAM theory of this effect (Fiedler-Ferrari, Nussenzveig and Wiscombe [1991]) yields a new type of diffraction integral, the *Pearcey–Fock integral*, that describes near-critical scattering. Results for the perpendicular polarization gain factor (ratio of polarized intensity to that for an ideal isotropic scatterer), near the critical angle for an air bubble in water with $\beta = 10^4$ are shown in fig. 15. Contributions from non-critical paths that produce rapid interference oscillations (fine structure) have been subtracted out.

The WKB curve shows the slope break at the critical angle. The physical optics approximation (POA) yields a Fresnel-like diffraction pattern, sharply disagreeing with the Mie results, while CAM produces a very good approximation in the near-critical region. The angular width of the critical region is $\delta = \mathrm{O}(\beta^{-1/2})$. The spherical Goos–Hänchen angular displacement is different for electric and magnetic polarizations.

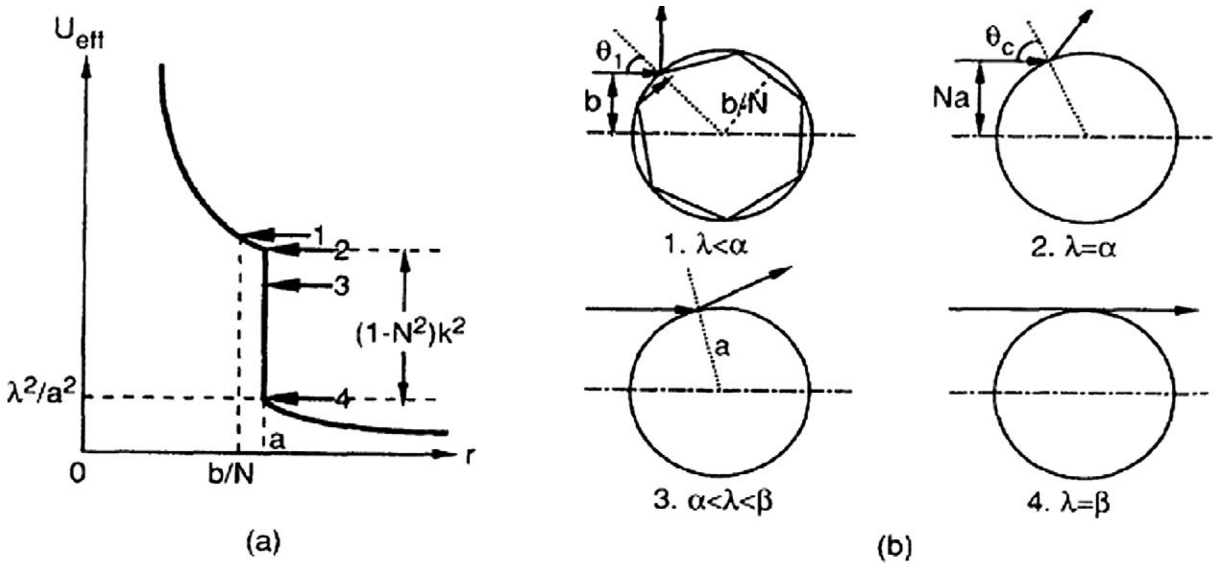


Fig. 14. (a) The effective potential for scattering by a sphere with $N < 1$, showing four "energy levels". (b) Corresponding incident rays – 1: subcritical incidence; 2: critical incidence; 3: supracritical incidence; 4: edge incidence; $\alpha \equiv N\beta$. Reprinted with permission from Fiedler-Ferrari, Nussenzveig and Wiscombe [1991]. © 1991 by the American Physical Society.

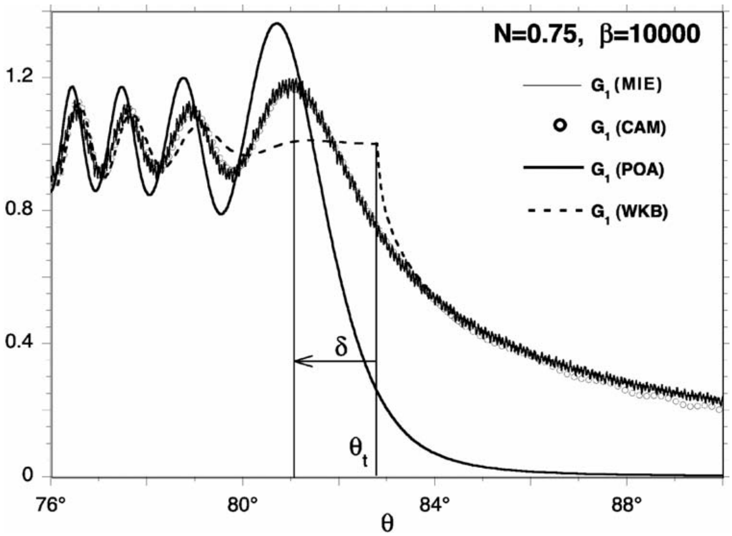


Fig. 15. Perpendicular polarization gain for $N = 0.75$, $\beta = 10\,000$, in the near-critical region: Mie result (with fine structure subtracted out), CAM (open circles), physical optics approximation POA, and WKB approximation (dashed line); θ_t = critical scattering angle; angular width of near critical region $\delta = O(\beta^{-1/2})$. Reprinted with permission from Fiedler-Ferrari, Nussenzveig and Wiscombe [1991]. © 1991 by the American Physical Society.

The Goos–Hänchen shift on a plane interface is of the order of the wavelength, so that measurements in the visible require amplification by multiple passages (Goos and Hänchen [1947]). In contrast, the spherical Goos–Hänchen angular displacement is a macroscopic effect, observable with the naked eye. Experimental observations (Tran, Dutriaux, Balcou, Le Floch and Bretenaker [1995]) for large size parameters are in very good agreement with CAM theory predictions, including the polarization dependence.

§ 10. The rainbow

10.1. The Debye expansion

Going over to $N > 1$, the chief applications are to light scattering by water droplets in the atmosphere, that gives rise to several striking meteorological optics effects (Greenler [1980], Lynch and Livingston [2001], Minnaert [1993], Tricker [1970]). The CAM treatment is based on the *Debye expansion* (Debye [1909a]),

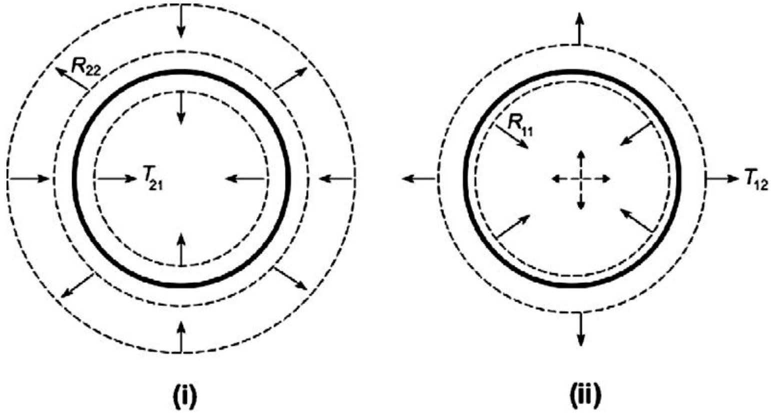


Fig. 16. Interpretation of the Debye expansion. (i) An incoming spherical multipole wave is partially reflected at the surface (reflection amplitude R_{22}) and partially transmitted (transmission amplitude T_{21}); (ii) The transmitted spherical wave goes through the center and back to the surface, where it is partially transmitted (amplitude T_{12}) and partially reflected (amplitude R_{11}).

an exact representation of the S function that parallels the ray optics description in terms of multiple internal reflections within a droplet.

We denote by 1 and 2 the interior of the sphere and the external region, respectively. The interaction of an incoming spherical multipole wave with the sphere is broken up into an infinite series of partial transmissions through the surface, going through the center, which plays the role of a perfect reflector, converting incoming into outgoing waves that are again partially reflected at the surface and partially transmitted to the outside region (fig. 16).

Each interaction with the surface gives rise to *spherical reflection and transmission coefficients* $R_{ij}^{(p)}(\lambda, \beta)$, $T_{ij}^{(p)}(\lambda, \beta)$, $(i, j) = (1, 2)$, $p = (1, 2)$, where p is the polarization index (Van Der Pol and Bremmer [1937], Nussenzveig [1969a]). They are related with the Fresnel reflection and transmission amplitudes.

Each term in the Debye expansion has a Poisson representation that is treated by the CAM methods discussed in Section 7.2: all integrands are meromorphic functions. CAM yields rapidly convergent asymptotic approximations for $\beta \gg 1$.

Background integrals are dominated by saddle-point contributions that yield the WKB approximation for ray paths with corresponding numbers of internal reflections. In different angular regions, the number of real saddle points (geometrical-optic rays) can vary, giving rise to transition regions described by penumbra Fock-type effects, as well as to rainbow-type transitions, that are discussed below.

The complex poles, named *Regge–Debye poles*, give rise to surface waves generated by glancing incident rays, as was found for an impenetrable sphere, again associated with tunneling. However, after describing any arc along the surface, they can now take "shortcuts" across the sphere (cf. fig. 25(b), below), with amplitudes determined by surface wave transmission coefficients, and the contributions from all possible such paths must be summed over.

One must also consider how fast the Debye series itself converges. Successive terms are damped roughly by the average internal reflection amplitude within the domain of angles of incidence that contributes. For below-edge rays, the internal reflection amplitude tends to be very small, implying fast convergence, except for near-glancing incidence. Thus, for water droplets, it usually suffices to consider only the first three terms of the Debye series (direct reflection, direct transmission and transmission after one internal reflection), which, in geometrical optics, contribute over 98.5% of the total intensity (Van De Hulst [1957]).

This does not prevent the intensity from becoming highly concentrated in narrow angular regions, as happens in the glory (Section 12), or in narrow size-parameter ranges (Mie resonances, Section 11). To survive internal reflection damping, such contributions must arise from near-total internal reflection, i.e., incidence near the edge. Thus, for the glory and for Mie resonances, slow convergence of the Debye expansion is expected, and one must either consider many terms or go back to the Mie series.

10.2. The primary rainbow

For rainbow history, we refer to Boyer [1987], Jackson [1999], and Lee and Fraser [2001]. We have already commented above on the contributions from Descartes, Newton, Young and Airy. The primary rainbow is contained in the third Debye term, associated with a single internal reflection. The rainbow scattering angle θ_R separates the shadow side of the primary rainbow (Alexander's dark band: no real rays with a single internal reflection) from the bright side, covered by two geometric rays with different paths (in a domain $\theta_L > \theta > \theta_R$); the interference between them gives rise to the supernumeraries.

The two real rays on the bright side are associated with two real saddle points in the λ plane, shown as black and white circles in fig. 17(a). As θ decreases from θ_L to θ_R, the two saddle points move towards each other, coalescing at the rainbow angle $\theta = \theta_R$. On the shadow side, for $\theta < \theta_R$, the two saddle points become complex, moving apart along complex conjugate paths. The contribution from the saddle point in the lower half-plane is dominant. It represents a *complex ray*, describing light tunneling into the shadow side.

Each saddle point has a *range*, its neighborhood in the λ plane that gives a significant contribution to the integral. When the two ranges overlap, as happens

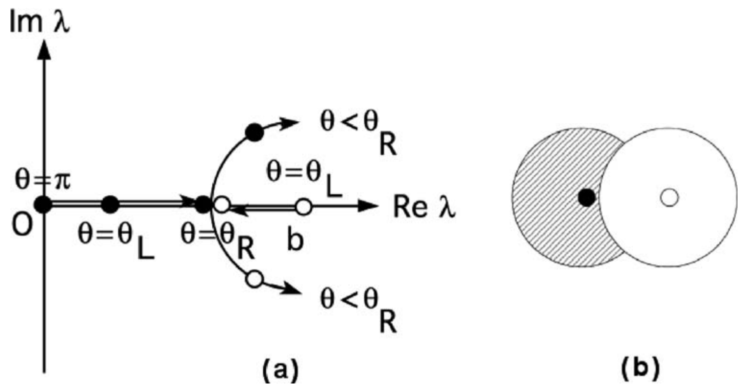


Fig. 17. (a) Saddle point trajectories in the λ plane for scattering angles θ around the rainbow scattering angle θ_R; ○ ● – saddle points. (b) Saddle points with overlapping ranges.

around the rainbow angle (fig. 17(b)), the ordinary saddle-point method must be modified, in order to yield a uniform asymptotic approximation. The uniform procedure was developed by Chester, Friedman and Ursell [1957]. It was applied by Berry [1966] to quantum potential scattering, by Nussenzveig [1969a] to scalar wave rainbow scattering, and by Khare and Nussenzveig [1974] and Nussenzveig [1979] to electromagnetic rainbow scattering. The Airy theory is a transitional lowest-order approximation (in all respects) to the uniform theory.

Interference with the Debye direct reflection contribution produces rapidly oscillatory "fine structure". Figure 18 shows a comparison, in the primary rainbow region, between parallel polarized intensities from Mie, CAM and Airy theories, after subtracting out fine structure, for water droplets with size parameter 1500 (requiring more than 1500 terms in the Mie summation). The uniform CAM approximation agrees very well with the exact theory: the small oscillatory differences arise from higher-order Debye non-rainbow contributions. However, Airy theory fails completely for this polarization, reversing the locations of supernumerary maxima and minima. For the dominant perpendicular polarization, agreement within the main rainbow peak (though not outside it) is much better (Khare and Nussenzveig [1974]).

The reason for the poor performance of Airy theory for parallel polarization is that the two interfering contributions to the supernumeraries have incidence angles on opposite sides of the Brewster angle, leading to an additional 180° phase shift not accounted for in Airy's classical diffraction theory. Experimental imaging polarimetry of the rainbow supports these results (Können and De Boer [1979], Barta, Horváth, Bernáth and Meyer-Rochow [2003]).

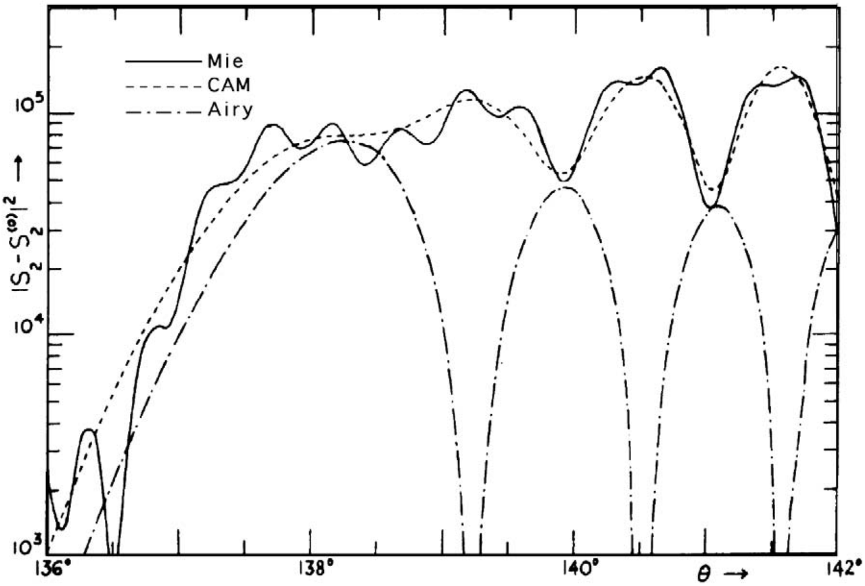


Fig. 18. Mie, CAM and Airy parallel polarized intensity near the rainbow angle for $N = 1.33$, $\beta = 1500$ (direct reflection has been subtracted out). Reprinted with permission from Khare and Nussenzveig [1974]. © 1974 by the American Physical Society.

Extended numerical comparisons between the Mie and Airy theories have been made (Wang and Van De Hulst [1991], Lee [1998]). Taking into account the size dispersion of water droplets in clouds tends to smooth out rapid interference oscillations and decrease discrepancies. The dark band, where tunneling occurs, becomes dominated by other Debye contributions, such as externally reflected light and higher-order rainbows. In terms of visual perception, rather than physical variables, differences are hard to detect.

On the other hand, applications to rainbow refractometry require more precise quantitative data. A comparative analysis of Mie, CAM and Airy theories (Saengkaew, Charinpatnikul, Vanisri, Tanthapanichakoon, Mees, Gouesbet and Grehan [2006]) shows that, for an extended angular domain which includes the secondary rainbow, and taking into account droplet size dispersion, Airy theory overestimates intensities and yields incorrect supernumerary positions at larger angles, whereas CAM agrees closely with Mie theory and requires dramatically smaller computation times.

Shape distortion of larger raindrops by air-drag forces changes rainbow features only slightly, an indication of their structural stability. Structurally stable caustics, of which the rainbow is the simplest one, are associated with "diffraction catastrophes" (Berry and Upstill [1980]).

§ 11. Mie resonances and ripple fluctuations

11.1. Mie resonances

The effective potential (7.4) for $N > 1$ is shown in fig. 19(a). It superposes a rectangular well onto the centrifugal barrier, recalling the nuclear well surrounded by the Coulomb barrier models employed in the early treatments of quantum tunneling (Section 1). This is a typical shape that gives rise to *resonances*. We consider impact parameters b in the range

$$a \leqslant b \leqslant Na \tag{11.1}$$

corresponding to incident rays that pass outside the sphere (fig. 19(b)), so that the "energy" k^2 lies between the top T of the external barrier at $(\lambda/a)^2$ and its bottom B at the well depth (7.3) below (fig. 19(a)).

Besides the outer turning point at $r = b$, there are two inner turning points, at $r = a$ and at $r = b/N$. Correspondingly (fig. 19(b)), the incident ray tunnels through the centrifugal barrier to the surface and takes a shortcut inside, so that it undergoes multiple internal reflections between the two radial turning points, the inner sphere representing a caustic for the transmitted rays.

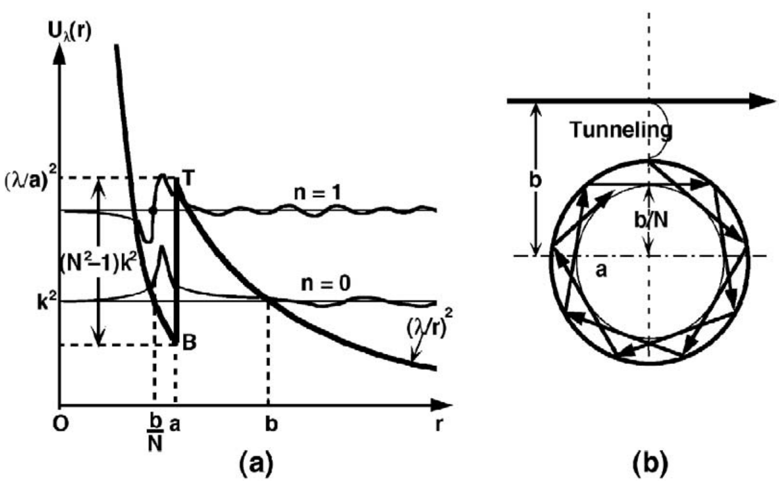


Fig. 19. (a) The effective potential for a sphere of radius a with $N > 1$, with a sketch of resonance wave functions for $n = 0$ and $n = 1$. (b) A corresponding incident ray with impact parameter $b > a$.

Since internal incidence is beyond the critical angle, reflection would be total in geometrical optics, producing a *bound state of light*. However, tunneling back to the outside through the centrifugal barrier allows radiation to escape. Within narrow neighborhoods of "resonance energies", analogous to quantum energy levels within the well, boundary condition matching leads to very large ratios of internal to external wave amplitudes, as illustrated in fig. 19(a). This provides the physical interpretation of Mie resonances (Nussenzveig [1989], Guimarães and Nussenzveig [1992, 1994], Johnson [1993]).

Mie resonances have also been referred to as "whispering gallery modes", by analogy with acoustics (Lord Rayleigh [1910]), and as "morphology-dependent resonances" (Chang and Campillo [1996]). They are also associated with the natural modes of oscillation of a dielectric sphere (Debye [1909a], Nussenzveig [1972]).

For a given angular momentum, several resonances may arise. They are characterized by a "family number" n ($= 0, 1, 2, \ldots$), the number of nodes of the radial wave function within the well in the limit of zero leakage: wave functions for $n = 0$ and $n = 1$ are sketched in fig. 19(a). The lower the value of n, the deeper the quasibound state lies within the well and the longer is the resonance lifetime, since it must tunnel across an increasing barrier width.

In CAM theory, the resonances are described by the complex Regge poles (Sections 6, 7) $\lambda_{nj}(\beta)$, where $j = 1, 2$ is the polarization index. As β varies, the poles describe *Regge trajectories*, that give unified descriptions of all resonances with the same family number n. A resonance with polarization j in partial wave l arises when $\operatorname{Re}\lambda_{nj}$ crosses the physical value $\lambda = l + \frac{1}{2}$. The resonance width follows from $\operatorname{Im}\lambda_{nj}$, which is determined by the transmissivity of the potential barrier; as above, the lowest n yield the narrowest resonances.

11.2. Ripple fluctuations

Resonance contributions to the polarized amplitudes correspond to residues at the Regge poles. The contributions from Debye terms that were discussed above play the role of background amplitudes, and their interference with resonance contributions must be taken into account in the total polarized intensities and cross-sections.

In the behavior of cross-sections as functions of size parameter, Mie resonances contribute to the "ripple", a very complex structure showing rapid quasiperiodic fluctuations and peaks with a variety of heights and widths (Shipley and Weinman [1968]), that appears at all scattering angles and is highly sensitive to refractive index changes, representing a serious nuisance in numerical computation. An exam-

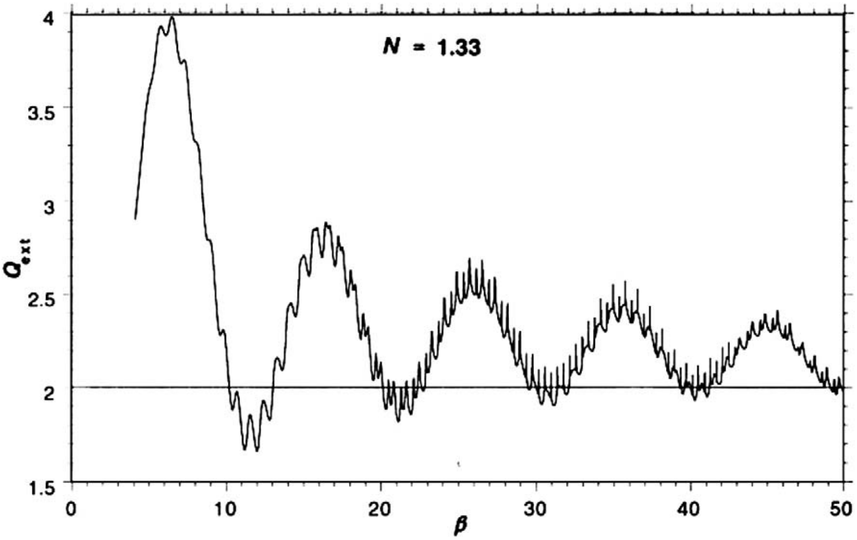


Fig. 20. Extinction efficiency for water droplets with size parameters from 4 to 50, displaying ripple fluctuations.

ple is the extinction efficiency $Q_{\text{ext}}(\beta)$ (ratio of total to geometrical cross-section) of water droplets for β between 4 and 50, shown in fig. 20.

A typical water droplet in the atmosphere, with radius of the order of 10 μm, is highly transparent in the visible and is very close to spherical shape because of surface tension, so that the Mie model is very realistic and the sharp ripple fluctuations are present. For an oil droplet of this size, they are so narrow that a fractional change in *average* radius of the order of 0.1 Å could be experimentally detected by monitoring radiation pressure (Ashkin [1980])!

In view of the extreme narrowness of the resonances for the large values of β that are required in many applications, it is important to determine resonance positions with accuracy better than the resonance width. This has been done employing uniform asymptotic approximations (Guimarães and Nussenzveig [1992]).

The quasiperiodic features of the resonances arise from the period of Regge recurrences, successive passages of a Regge trajectory near physical values of the angular momentum. The quasiperiod is approximately given by (Chylek [1990])

$$\delta\beta \sim M^{-1} \tan^{-1} M, \quad M \equiv \sqrt{N^2 - 1}, \tag{11.2}$$

which is 0.821 for $N = 1.33$.

A CAM fit to Q_{ext} within the quasiperiod $58.2 \leqslant \beta \leqslant 60.0$ (Guimarães and Nussenzveig [1992]) is shown in fig. 21. All resonances that contribute are la-

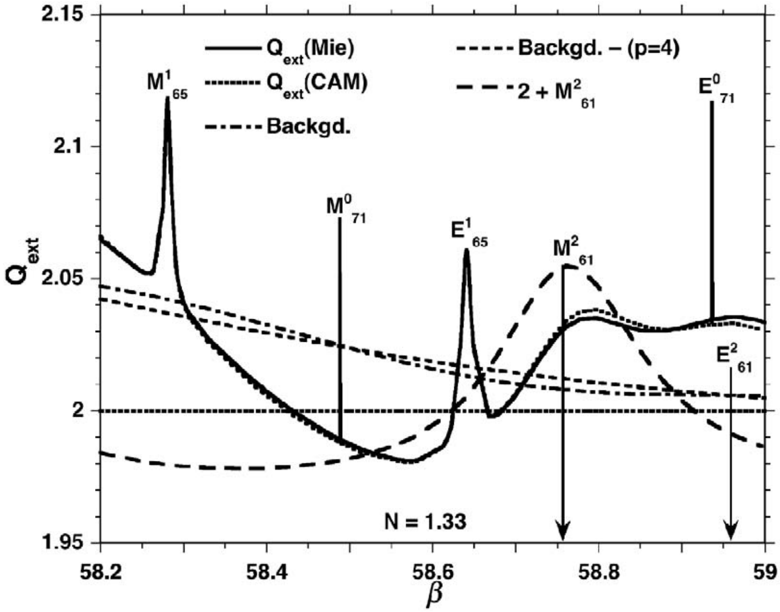


Fig. 21. CAM fit to the Mie extinction efficiency for $N = 1.33$ within a quasiperiod. The background contribution, including and excluding the $p = 4$ forward glory term, is shown. The contribution offset by 2 from the broad resonance M_{61}^2 includes negative (antiresonant) contributions.

beled by the notation $(E, M)_l^n$ (E or M polarization, lower index angular momentum, upper index family number). The CAM approximation is evaluated to order $O(\beta^{-2})$, and it fits the Mie result to this order (<0.1%), even for needle-like peaks with $n = 0$.

The high accuracy of the CAM approximation allows us to detect a *forward glory* background contribution (Nussenzveig and Wiscombe [1980a]), associated with Debye terms with 3 internal reflections. A forward glory arises from incident rays with impact parameter $b \neq 0$ that reemerge in the forward direction. It includes a real glory ray (fig. 22(a)) and surface wave contributions (fig. 22(b)). The glory enhancement is discussed in Section 13.

§ 12. Light tunneling in clouds

Among the largest uncertainties in climate modeling is the effect of clouds on solar radiation. Predictions from atmospheric models for solar absorption by water clouds are based on parametrizations (Mitchell [2000]) of Mie scattering. Since

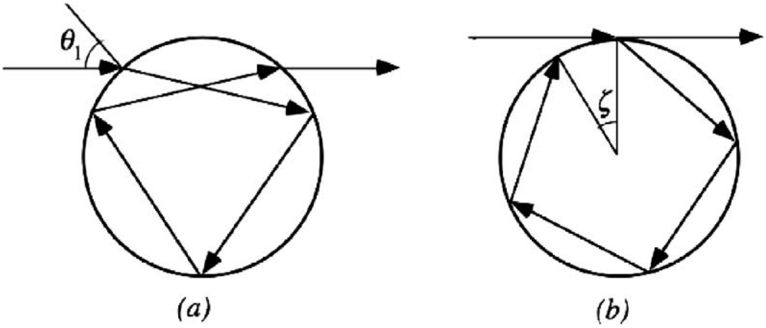


Fig. 22. Forward glory paths for the $p = 4$ Debye term, $N = 1.33$. (a) Real glory rays; (b) Surface waves with missing angle $\zeta = 30°$.

size parameters range up to several thousand, a computational step size $\Delta\beta = 0.1$ is usually adopted (Dave [1969]) to reduce the effort required.

The proliferation of resonances with widths far below 0.1 in such intervals, however, leads to aliasing errors in radiative transfer calculations. Can Mie resonances have a significant effect? Sharp resonance absorption peaks have a Lorentzian shape, and all their contributions are additive, so that their effect must be investigated.

In order to estimate the total effect of Mie resonances on absorption (Nussenzveig [2003]), one can start by looking at their contribution to the density of states. The density of states can be evaluated by quantizing the Mie modes. It is found (De Carvalho and Nussenzveig [2002]) that the ratio R of the contribution from Mie resonances to the total density of states, for $\beta \gg 1$, is asymptotically given by

$$R = (M/N)^3, \tag{12.1}$$

where M is defined by (11.2). Thus, for $N = 1.33$, approximately 29% of the total mean density of states arises from resonances.

Another asymptotic estimate is obtained from the WKB approximation, assuming $\kappa \ll 1$, where κ is the imaginary (absorptive) part of the complex refractive index. This weak absorption assumption is valid for water in the visible. The result (Nussenzveig [2003]) is

$$\frac{\langle Q_{\text{abs,res}}\rangle}{Q_{\text{abs,GO}}} = \frac{3}{4}\frac{\tan^{-1} M}{(N^3 - M^3)}\left[\left(\frac{M}{\tan^{-1} M}\right)^2 - 1\right] \tag{12.2}$$

where $\langle \cdots \rangle$ denotes an average over the quasiperiod $\delta\beta$ and the indices refer to the resonance contribution and the geometrical-optic contribution to the absorp-

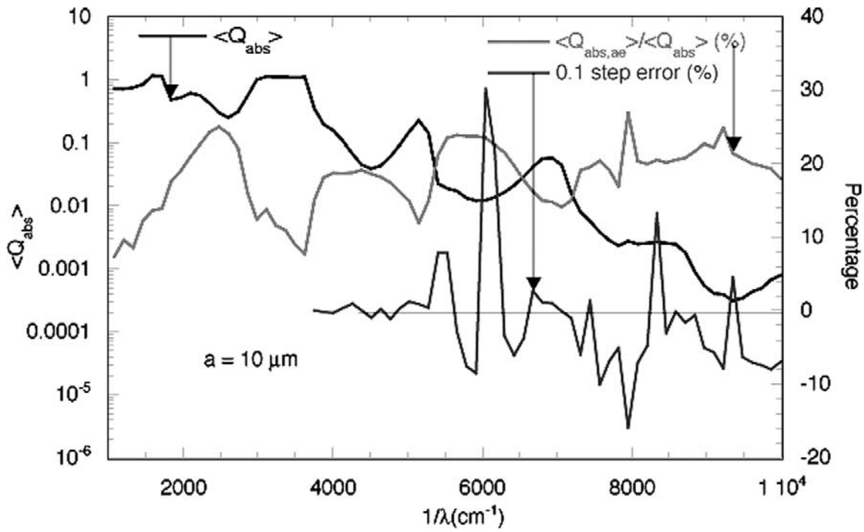


Fig. 23. Average spectral absorption efficiency of a 10-μm-radius water droplet in the near infrared (left log scale). The contribution from tunneling (% of total, right % scale) and the % error from plotting at 0.1 steps (right scale) are shown.

tion efficiency (ratio of absorption cross-section to geometrical cross-section). For $N = 1.33$, the ratio (12.2) is $\approx 15.6\%$, another indication that the contribution from resonances is significant.

A precise numerical computation of the Mie resonance contribution to solar absorption for a typical cloud water droplet with $a = 10$ μm was performed by Nussenzveig [2003]. Spectral averages $\langle Q_{\mathrm{abs}} \rangle$ over $\delta\beta$ were evaluated, as well as the tunneling (above-edge) contribution $\langle Q_{\mathrm{abs,ae}} \rangle$, that includes all resonances. For this purpose, resonances can be treated as Lorentzians, with positions and widths given by the uniform CAM approximation (Guimarães and Nussenzveig [1992]), and relatively slow-varying nonresonant terms can be plotted at 0.1 steps, greatly simplifying the numerical computation. In order to determine the aliasing percent error, the Mie evaluation was also performed at 0.1 steps for comparison.

The spectral range plotted includes most of the domain where liquid water absorption in a typical stratus cloud (Davies, Ridgway and Kim [1984]) is significant. The results are displayed in fig. 23. They show that the average tunneling contribution to absorption is of the order of 20%, consistent with the above estimates. This is a truly global, large-scale, macroscopic consequence of light tunneling: a sizable proportion of the solar darkening effect of typical liquid water clouds arises from tunneling!

How about the aliasing percent error? As seen in fig. 23 for $a = 10$ μm, besides negative errors that exceed 10% from missing contributions of sharp resonances, there are also positive errors, arising from overestimation of resonance peak areas due to the coarseness of the interval.

In order to find out possible global climate repercussions of these results, one must apply a radiative transfer model to determine how they are affected by averaging over realistic size distributions and by multiple scattering. This has been done by Zender and Talamantes [2006]. The results confirm the conclusion that the canonical computational step size $\Delta\beta = 0.1$ should not be adopted when aiming at accuracy better than 10%.

To resolve all sharp resonances requires $\Delta\beta \sim 10^{-7}$. Over 10 nm and narrower spectral bands, as employed in satellite and aircraft remote sensing, the canonical resolution may lead to absorption biases up to 70%, so that Mie resonance absorption must be included for reliable results. However, the overall increased cloud heating arising from this source is found to be less than 0.1%, a negligible effect for global climate. This results from cancellations of positive and negative errors, already seen in fig. 23 for a monodispersion, as well as from overlap between absorption by liquid droplets and by water vapor. However, the conclusion applies to pure water liquid droplets. Effects of contamination by absorbers such as soot (Chylek, Lesins, Videen, Wong, Pinnick, Ngo and Klett [1996]) remain to be considered.

§ 13. The glory

13.1. Observations and early theories

The most striking manifestation of light tunneling, and one of the most beautiful among natural phenomena, is the glory. In Philip Laven's web site www.philiplaven.com some wonderful pictures of glories and other meteorological phenomena can be found; this site is also highly recommended for its color displays of Mie and Debye fits to angular distributions and other features of these effects.

The first recorded observation of the glory was made at Mount Pambamarca in the Ecuador Highlands (formerly Peru) some time between 1737 and 1739 (Lynch and Futterman [1991]), by members of a French Academy of Sciences expedition, led by Bouguer and La Condamine, accompanied by the Spanish captain and astronomer Antonio de Ulloa[1], to make measurements concerning the Earth's shape. This is one of the expeditions about which Voltaire wrote his satirical couplet

[1] Ulloa brought back the first samples of platinum to reach Europe (Watson and Brownrigg [1749]).

"Vous avez confirmé dans ces lieux pleins d'ennui
Ce que Newton connut sans sortir de chez lui."

What they observed, as related by Bouguer, was (cf. Pernter and Exner [1910])

"...a phenomenon which must be as old as the world, but which no one seems to have observed so far... A cloud that covered us dissolved itself and let through the rays of the rising sun... Then each of us saw his shadow projected upon the cloud... What seemed most remarkable to us was the appearance of a halo or glory around the head, consisting of three or four small concentric circles, very brightly colored, each of them with the same colors as the primary rainbow, with red outermost..."

Ulloa, who drew a picture of their observation, added:

"...The most surprising thing was that, of the six or seven people who were present, each one saw the phenomenon only around the shadow of his own head, and saw nothing around other people's heads..."

A detail of a beautiful (and fanciful) engraving based on this observation, together with the frontispiece of the book where it appeared (Juan and Ulloa [1748]), is reproduced in fig. 24. One may wonder about the connection between such impressive observations and the ubiquitous representations of deities and emperors wearing halos in eastern and western iconography.

In his Nobel Lecture (Wilson [1927]), C.T.R. Wilson describes why he invented the cloud chamber:

"In September 1894 I spent a few weeks in the Observatory which then existed on the summit of Ben Nevis, the highest of the Scottish hills. The wonderful optical phenomena shown when the sun shone on the clouds surrounding the hill-top, and especially the coloured rings surrounding the sun (coronas) or surrounding the shadow cast by the hill-top or observer on mist or cloud (glories), greatly excited my interest and made me wish to imitate them in the laboratory."

Although he did not succeed in his original aim, he soon discovered that it allowed visualization of charged particle tracks!

The glory was a favorite image among romantic writers (Hayter [1973]), inspiring Coleridge's beautiful poem "Constancy to an ideal object". Besides mountain-top observations, sightings were made from balloons and nowadays glory sightings around the shadow of an airplane on the clouds (Bryant and Jarmie [1974]) have become fairly common.

Early attempts to explain the glory were based on a mistaken analogy with diffraction coronas, which arise from the forward diffraction Airy pattern (8.1). The reversal of the direction of propagation was attributed by Fraunhofer and Pernter (Pernter and Exner [1910]) to reflection from clouds, an untenable proposal.

RELACION HISTORICA
DEL VIAGE
A LA AMERICA MERIDIONAL
HECHO
DE ORDEN DE S. MAG.
PARA MEDIR ALGUNOS GRADOS DE MERIDIANO
Terrestre, y venir por ellos en conocimiento de la verdadera Figura, y Magnitud de la Tierra, con otras varias Observaciones Astronomicas, y Phisicas:

Por DON JORGE JUAN, *Comendador de Aliaga, en el Orden de San Juan, Socio correspondiente de la Real Academia de las Ciencias de Paris,* y DON ANTONIO DE ULLOA, *de la Real Sociedad de Londres: ambos Capitanes de Fragata de la Real Armada.*

PRIMERA PARTE, TOMO PRIMERO.

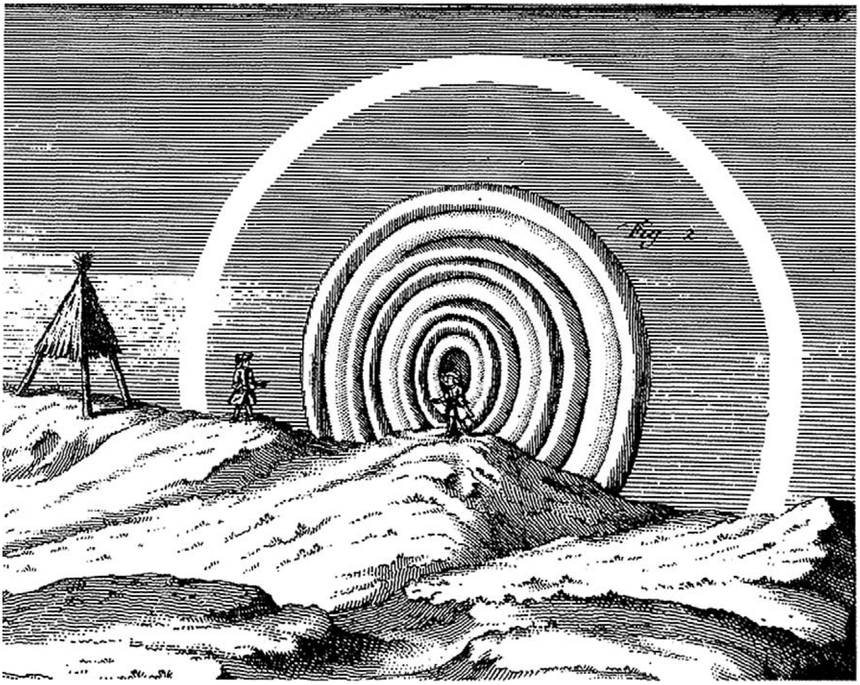
IMPRESSA DE ORDEN DEL REY NUESTRO SEÑOR
EN MADRID
Por ANTONIO MARIN, Año de M.DCC.XLVIII.

(a)

Fig. 24. (a) Title page of the book by Juan and Ulloa. (b) Detail of engraving in this book representing their observation of the glory (*next page*).

B. Ray [1923] correctly concluded from experiments with artificial clouds that the glory arises from backscattering by individual water droplets. However, he also proposed (as did Bricard [1940]) that this was produced by interference between axially reflected rays from the front and back surfaces of a droplet. As will be seen below, this contribution is negligible in the glory. A proposal by Bucerius [1946] based on what he termed "backwards diffraction" was also incompatible with observed glory features.

Thus, more than two centuries after the first reported observation, no theory existed to account for the glory, apart from Mie theory. However, the typical range of size parameters in which the glory is observed ranges from $\beta \sim 10^2$ to 10^3, so that Mie summations have to include hundreds of partial waves, and the plotting increment has to be extremely small because of the ripple (Sections 11, 12). Be-



(b)

Fig. 24. (*Continued.*)

sides, numerical results do not reveal what physical features are relevant. According to a statement attributed to Eugene Wigner, “It is very nice that the computer understands the problem, but I would like to understand it too.”

13.2. Van De Hulst’s theory

An important contribution to the theory of the glory was given by Van De Hulst [1947]. Forward glory rays were discussed in Section 11. Could the glory be produced by backward glory rays? The lowest-order such ray would require one internal reflection, as illustrated in fig. 25(a). Also shown in the figure are two neighboring rays, as well as a portion of an incident plane wavefront, converted after backscattering into a curved wavefront with a virtual focus at F.

For a generic scattering angle, this would be a virtual point source of spherical waves. However, for backscattering there is axial symmetry, so that F is part of

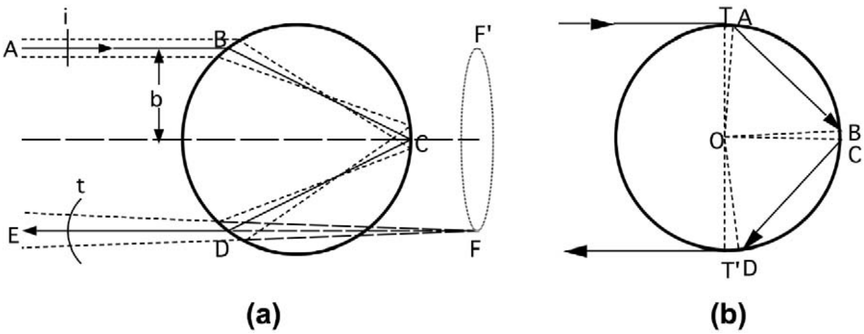


Fig. 25. (a) For a real backward glory ray ABCDE, the emerging portion of spherical wave t arising from a portion of plane incident wavefront i has a virtual focus F. The locus of virtual foci FF′ is a ring source of toroidal wavefronts, that produces the axial focusing effect. However, such a ray does not exist for $N = 1.33$. (b) The van de Hulst glory ray travels as a surface wave with shortcuts to bridge the angular gap of $\sim 15°$.

a *virtual ring source*, associated with *toroidal wavefronts*. By applying Huygens' principle to such a wavefront, van de Hulst shows that it gives rise to a backscattering amplitude enhancement by a factor of order $(kb)^{1/2}$, where b is the impact parameter of the glory ray. In terms of Young's theory of edge diffraction, this *axial focusing enhancement* is also the explanation of the Poisson spot at the center of the shadow of a circular disc.

For water droplets, however, a backward glory ray of the type shown in fig. 25(a) does not exist. The largest deviation, for tangentially incident rays, would still leave an angular gap of about 15° from the backward direction. It was suggested by van de Hulst [1957] that this gap might be bridged by a combination of surface waves and shortcuts, as illustrated in fig. 25(b). However, a quantitative procedure to derive such contributions was not available at that time.

13.3. CAM theory: background contributions

The earliest CAM treatment of the glory was in scalar scattering (Nussenzveig [1969b]). It verified that the axial-ray contribution proposed by Ray and Bricard is negligible and it provided the first evaluation of van de Hulst's surface wave term, as the residue of a Regge–Debye pole. The result for a typical value of β confirmed its relevance to the glory, but indicated the need to include higher-order Debye terms. A preliminary discussion of their contributions was given.

The extension to electromagnetic scattering (Khare [1975, 1982], Khare and Nussenzveig [1977a], Nussenzveig [1979, 1992]) verified these conclusions of

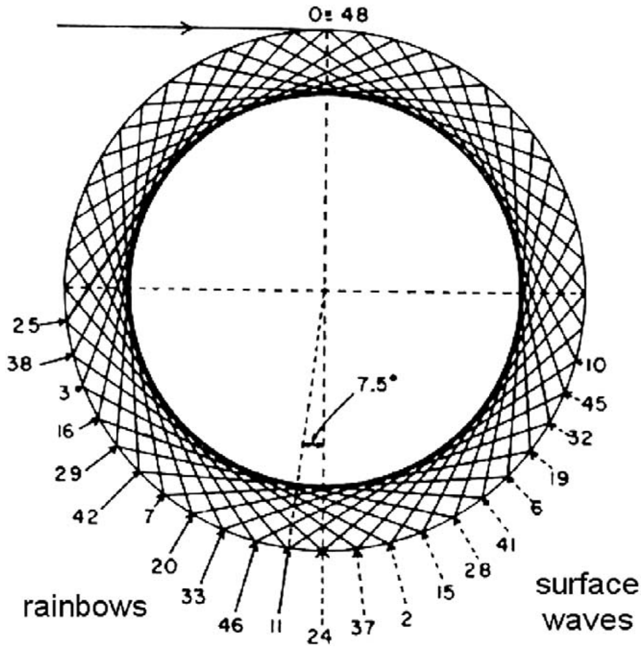


Fig. 26. For $N = 1.33007$, a tangentially incident ray yields a real backward glory ray after $p = 24$ shortcuts. The qualitative ordering of Debye background terms of order p (ignoring internal reflection damping) is represented by the length of the dashed arrows (for surface wave contributions) and full line arrows (for rainbow contributions).

the scalar treatment and determined which terms in the Debye series are most significant. It found that, besides surface-wave contributions, there are also relevant effects from higher-order rainbows formed around the backward direction. The rainbow enhancement persists at greater deviations because the width of the rainbow region increases with rainbow order.

We denote by p the order of the Debye term ($p - 1$ internal reflections). The refractive index of water is close to the value

$$N = \left[\cos(11\pi/48)\right]^{-1} \cong 1.33007, \tag{13.1}$$

for which a tangentially incident ray produces a closed orbit, a regular star-shaped 48-sided polygon (fig. 26), so that there is a real backward glory ray with $p = 24$. To the right of the vertical axis in fig. 26 lie rays that generate surface-wave contributions; to the left, those associated with rainbow contributions, where the backward direction lies on the shadow side of the rainbow.

Both types of contributions are damped by tunneling: surface-wave tunneling or rainbow shadow tunneling. The ordering of contributions taking into account this damping is indicated by the length of the arrows in fig. 26, but this does not include the additional damping by internal reflections. The different size dependence of these various damping effects leads to a change in dominance order with β. In the typical glory range, the leading surface wave term is van de Hulst's $p = 2$; the leading rainbow term is the 10th-order rainbow $p = 11$, for which the rainbow angle is 3° beyond the backward direction.

Numerical computations (Khare [1975, 1982], Khare and Nussenzveig [1977b], Nussenzveig [1992]) confirm these estimates (see also Laven [2004] and the site www.philiplaven.com). All dominant background contributions arise from tunneling in the *edge domain*

$$|\lambda - \beta| = \mathrm{O}\big(\beta^{1/3}\big), \tag{13.2}$$

where internal reflection damping is small, leading to complex interference effects among many Debye terms.

In order to assess the validity of the dominant CAM approximation to the background, that from the third Debye term, $p = 2$ (which contains the van de Hulst surface wave), we look at the *backscattering gain* $G(\beta, \pi)$, defined as the ratio of the backscattered intensity to its limiting geometrical-optic value for an isotropic scatterer (a totally reflecting sphere). In fig. 27, we compare the Mie result for

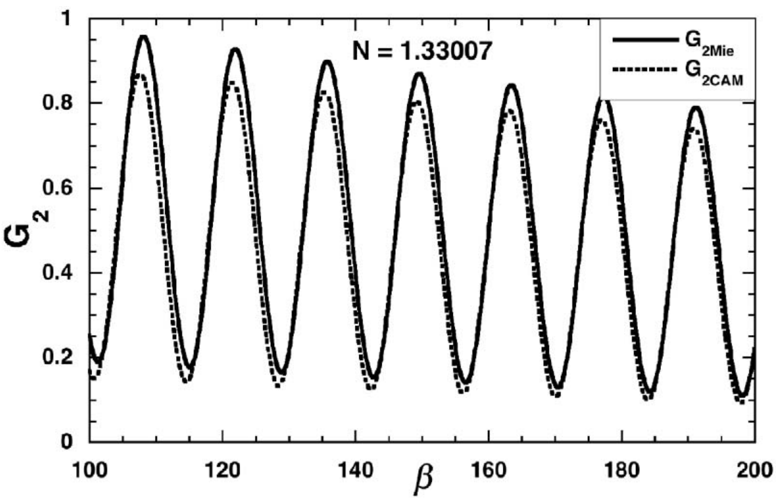


Fig. 27. Comparison between the Mie and CAM theories for the third Debye term $p = 2$. Oscillations arise from interference between the axial and van de Hulst contributions.

$G_{2,\mathrm{Mie}}$ with its lowest-order CAM approximation $G_{2,\mathrm{CAM}}$, the sum of the van de Hulst surface wave and the WKB term, for $N = 1.33007$ (cf. eq. (13.1)). The plotting range is $100 \leqslant \beta \leqslant 200$, a typical range for glory sightings (van de Hulst [1957]).

The oscillations, with period (where M is defined in eq. (11.2))

$$\Delta\beta = \frac{2\pi}{\pi + 2 - 4[N - M + \cos^{-1}(1/M)]}, \tag{13.3}$$

which for $N = 1.33$ is ≈ 14, one of the dominant periods identified in the backscattering pattern (Shipley and Weinman [1968]), arise from interference between these two contributions. We see that $G_{2,\mathrm{CAM}}$ is already a fairly good approximation to $G_{2,\mathrm{Mie}}$, even with the van de Hulst term evaluated only to lowest order, and the physical interpretation of the third Debye term is confirmed.

13.4. CAM theory: Mie resonance contributions

As was mentioned in Section 11.2, the total CAM polarized amplitudes include the background (Debye) terms discussed in Section 13.3 and the ripple contributions from Mie resonances, given by the residues at the Regge poles. We consider first the relative contributions from Mie resonances and background within a single quasiperiod.

In fig. 28 (Guimarães and Nussenzveig [1992]), the Mie result for G for $N = 1.33$, over one quasiperiod around $\beta = 58.6$, is compared with a CAM approximation that includes in the background just the van de Hulst term and the 10th-order rainbow term; direct reflection, also included, gives a much smaller contribution. We see that these terms alone increase the backscattering to values comparable with that for a total reflector.

At the sharp Mie resonances within this quasiperiod, the backscattered intensity from the transparent sphere exceeds that of a total reflector by one order of magnitude! The contributions from background and Regge terms alone are also plotted, as well as the ratio $|S_{\mathrm{B}}/S_{\mathrm{R}}|$ of the background amplitude to the Regge one, showing that they are of comparable magnitude, so that their interference must be taken into account in the backscattered intensity. The typical CAM errors for this lowest-order CAM approximation, that neglects higher-order Debye contributions to the background, is of order 20%. The fit of the backward gain, the feature that is hardest to evaluate, in view of tunneling dominance, is a crucial test of CAM theory.

The localization principle allows us to determine the relative weight of below-edge and above-edge incident ray contributions in the backscattering gain. In

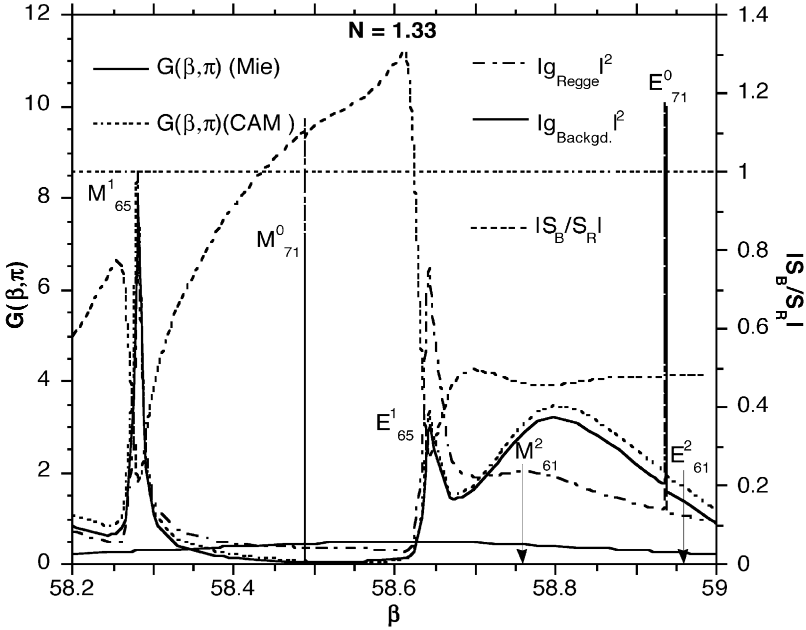


Fig. 28. Backward gain for $N = 1.33$: comparison between Mie and a CAM approximation that includes only $p = 0$, 2 and 11 background Debye terms. Contributions that would arise from background and Regge (Mie resonance) terms alone are plotted, as well as the ratio $|S_B/S_R|$ of background to Regge amplitudes.

fig. 29 (from Nussenzveig [2003]), $\langle G(\beta)\rangle$,the backscattering gain averaged over the quasiperiod (11.2) for $N = 1.33007$, is plotted in the size parameter range 5–150. We compare the Mie result $\langle G_{\text{Mie}}\rangle$ with the above-edge contribution, the below-edge contribution $\langle G_{\text{be}}\rangle$ and the geometrical-optic result $\langle G_{\text{go}}\rangle$.

In the plotted range, the average gain is of the order of 1.4, a value 40% higher than that for totally reflecting spheres and 14 times higher than the geometrical-optic result $\langle G_{\text{go}}\rangle$. Geometrical optics is totally unable to account for the glory, even at large size parameters. The complicated structure of the below-edge contribution results from interference among many higher-order Debye terms (Section 13.3). The above-edge tunneling contribution is strongly dominant, although there is appreciable interference with the below-edge contribution.

Figure 29 also shows the nonresonant CAM background approximation $\langle G_{\text{CAM,nr}}\rangle$, the average over quasiperiods of the approximation represented in fig. 28, that includes the van de Hulst term and the 10th-order rainbow term. Although $\langle G_{\text{CAM,nr}}\rangle$ contributes less than half of the total gain, it is representative

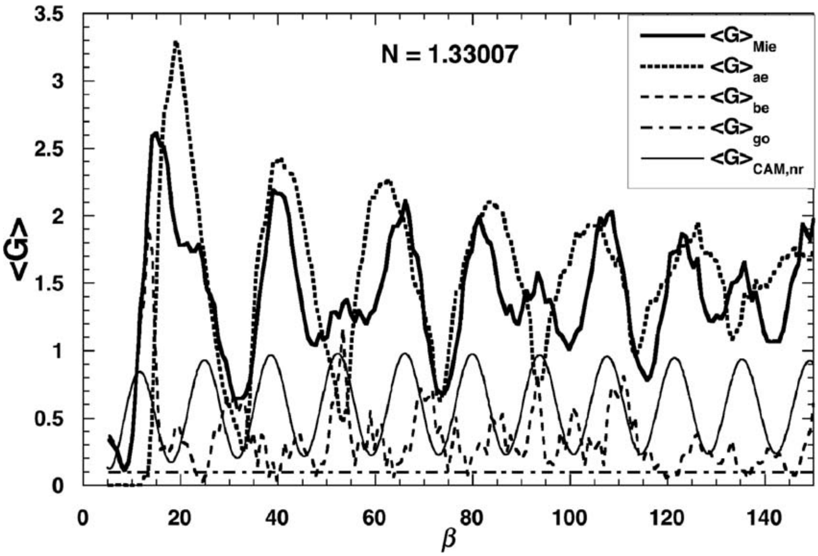


Fig. 29. Average backscattering gain factor of a water droplet with $N = 1.33007$ in the size parameter range 5–150. The Mie result is compared with the above-edge and below-edge contributions, as well as with geometrical optics. The CAM approximation to the background (nonresonant) corresponding to fig. 28 is also plotted.

of the qualitative behavior of $\langle G_{\mathrm{Mie}} \rangle$, with peaks and valleys in the two curves closely matched.

It can be verified that above-edge terms also dominate $\langle G_{\mathrm{CAM,nr}} \rangle$, so that tunneling also plays a major role in the van de Hulst surface wave. Experimental detection of the van de Hulst surface wave has been achieved in the terahertz domain (Cheville, Mcgowan and Grischkowsky [1998]).

CAM theory also yields the angular distribution and polarization of the glory. For near-backward scattering, it leads to a general approximation (Nussenzveig [1992]) of the angular dependence of the amplitudes, that follows directly from the asymptotic near-backward behavior of the angular functions in (5.1):

$$S_1(\beta, \theta) \approx 2S^M(\beta) J_1'(u) + 2S^E(\beta)\big[J_1(u)/u\big],$$
$$u \equiv \beta(\pi - \theta) \text{ not } \gg 1, \tag{13.4}$$

where $S^M(S^E)$ is the magnetic (electric) multipole contribution to the Mie amplitudes at $\theta = \pi$ and $-S_2(\beta, \pi)$ follows by interchanging $S^M \leftrightarrow S^E$; J_1 is the Bessel function.

A similar expression was proposed by Van De Hulst [1957], who tried to adjust the Bessel function coefficients by an empirical fit to observations of glory ring angular size. For natural incident light, it follows from (13.4) that the angular distribution and the polarization are given respectively by

$$\frac{i(\beta,\theta)}{i(\beta,\pi)} = J_0^2(u) + c(\beta)J_2^2(u), \qquad P(\beta,\theta) = \frac{2\sqrt{c(\beta)}J_0(u)J_2(u)}{J_0^2(u) + c(\beta)J_2^2(u)}, \tag{13.5}$$

where

$$c(\beta) \equiv \frac{S^E(\beta) - S^M(\beta)}{[S^E(\beta) + S^M(\beta)]^2}. \tag{13.6}$$

The parameter $c(\beta)$ also undergoes large ripple fluctuations (Nussenzveig [2002]). Even its size average over quasiperiods still shows substantial variation, in agreement with the fact that the angular distribution and the polarization of natural glories are highly variable. A roughly representative value is $\langle c\rangle \approx 3$, consistent with van de Hulst's empirical fits.

The corresponding angular distribution and polarization are plotted in fig. 30. They agree with the observed haziness of the first glory dark ring and with the strong radial polarization of the outer rings, opposite to that of the central ring. Polarization reversals occur in regions of near-vanishing intensity.

Tunneling contributions to Mie resonances and to the glory arise from the above-edge size parameter range

$$a < b < Na. \tag{13.7}$$

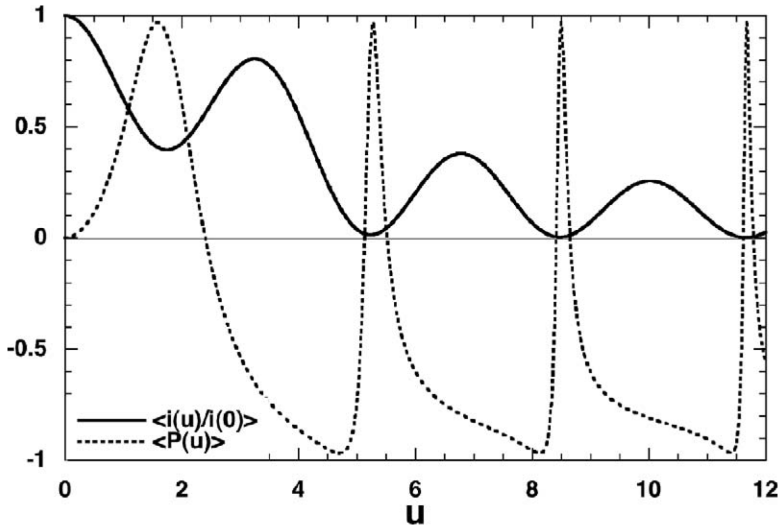


Fig. 30. Typical angular distribution and polarization in natural glories; $u \equiv \beta(\pi - \theta)$.

Comparing this with (8.3), we see that Newton's "action at a distance" effective range now extends to distances of the order of the sphere radius above its surface, independent of the wavelength. Newton's Query 20 (Section 1) might be taken to describe the effect of the centrifugal barrier on the effective potential ("refractive index"). These results provide a most striking validation of his conjectures about diffraction.

§ 14. Further applications and conclusions

14.1. Further applications

14.1.1. Average Mie efficiency factors

In applications to radiative transfer, three dimensionless parameters, defined by normalizing the various cross-sections relative to the geometrical cross-section, play an important role: the efficiency factors for extinction, absorption and radiation pressure (Van De Hulst [1957]). In Mie scattering, their behavior as functions of β is subject to ripple fluctuations, exemplified in fig. 20.

CAM approximations to size averages of these efficiencies over intervals $\Delta\beta \approx \pi$ have been evaluated (Nussenzveig and Wiscombe [1980b], Nussenzveig [1992]) and compared with corresponding Mie results. CAM contributions taken into account included the background (Debye terms), but not the ripple.

The results differ from WKB and classical diffraction theory by below-edge and above-edge corrections arising from anomalous reflection and tunneling in the edge domain (13.1). They greatly improve accuracy relative to these approximations and reduce computing time by a factor of order β with respect to Mie computations.

14.1.2. Microsphere cavities

Transparent microspheres can be regarded as optical cavities. The cavity modes are the Mie resonances; the extreme narrowness that they can attain thanks to tunneling is equivalent to very high cavity Q factors. Pioneering work to achieve this aim with silica microspheres was done by Braginsky, Gorodetsky and Ilchenko [1989], leading to the experimental realization of a Q factor of order 10^9 (Gorodetsky, Savchenkov and Ilchenko [1996]).

Applications to quantum and nonlinear optics, cavity quantum electrodynamics and chemical physics are reviewed in a book edited by Chang and Campillo [1996]; see also Fields, Popp and Chang [2000] and Von Klitzing, Long, Ilchenko, Hare and Lefèvre-Seguin [2001]. A recent report on applications to photonics and nonlinear optics (Matsko, Savchenkov, Strekalov, Ilchenko and Maleki [2005]) contains several hundred references; for a short review, see Vahala [2003].

14.1.3. Chaotic scattering

An early demonstration of the nonlinear optical effects of Mie resonances was the observation of lasing in dye-doped ethanol droplets (Tzeng, Wall, Long and Chang [1984]), arising from the large internal field between the droplet rim and the internal caustic sphere (Section 11.1). In freely falling droplets, which take on a spheroidal shape, lasing is confined to restricted angular domains around the rim (Qian, Snow, Tzeng and Chang [1986]).

The extreme variability of Mie resonance parameters to changes in size and refractive index is an indication of sensitive dependence on initial conditions, but it is not sufficient for chaos. As interpreted at the level of ray optics, however, the preferential lasing domains arising from spheroidal deformation signal the onset of chaos (Mekis, Nöckel, Chen, Stone and Chang [1995]), associated with a Kolmogorov–Arnold–Moser/Lazutkin transition (Lazutkin [1993]). In ray-optics terms, the transition to chaos occurs when the deformation allows internal incidence below the critical angle at specific regions of the boundary, leading to escape through these regions.

For other discussions of the connections between tunneling and chaos, involving also complex rays and penumbra diffraction, see Heller and Tomsovic [1993], Doron and Frischat [1995], and Primack, Schanz, Smilansky and Ussishkin [1997].

14.1.4. Photonic crystals

Mie resonances represent nearly bound states of light. Can they be rendered truly bound? This becomes possible in a crystal arrangement of microspheres. If a Mie resonance of the individual units overlaps with a collective Bragg resonance of the crystal, the light that tunnels out from a microsphere may be returned to it by the Bragg collective effect, and thus become trapped within the microsphere, corresponding to *strong localization of light* (John [1991], Chabanov and Zenack [2001]).

This intuitive picture is a highly simplified version of what takes place in photonic crystals (Joannopoulos, Meade and Winn [1995]), giving rise to *photonic band gaps* (John [1987], John and Wang [1991], Antonoyiannakis and Pendry [1997, 1999], Lidorikis, Sigalas, Economou and Soukoulis [1998], Moroz and Tip [1999]). A recent analysis is by Vandenbem and Vigneron [2005].

14.1.5. Dwell time

Complementary to the description of resonances in the frequency domain are their effects in the time domain. Since Mie resonant states decay by tunneling, this brings up the question of "tunneling time": how long does it take to tunnel through a barrier? This highly debated and controversial question, that has led to apparent paradoxes such as supposedly superluminal propagation, is an ill-posed one (Chiao and Steinberg [1997], De Carvalho and Nussenzveig [2002]).

At Mie resonances, the long whispering-gallery-like paths within the sphere greatly increase the dwell time (De Carvalho and Nussenzveig [2002]) of light in the neighbourhood of the scatterer. For a random medium with a large number of identical scatterers the velocity v_E of energy propagation is given by (Van Albada, Van Tiggelen, Lagendijk and Tip [1991], Van Tiggelen, Lagendijk, Van Albada and Tip [1992], Sheng [1990, 1995])

$$v_E \approx \frac{c}{1 + (\mathcal{T}_\mathrm{d}/\tau_\mathrm{mf})}, \tag{14.1}$$

where $\mathcal{T}_\mathrm{d}$ is the dwell time and τ_mf is the mean free time between successive scatterings. Thus, long dwell time delays can greatly reduce the velocity of energy propagation.

When such long time delays take place in an active medium, the resulting feedback enhances the gain. This can lead to mirrorless "random lasers" (Letokhov [1968], Lawandy, Balachandran, Gomes and Sauvain [1994], Burin and Cao [2004], Hu, Yamilov, Noh, Cao, Seelig and Chang [2004]).

14.1.6. Optical imaging and near-field microscopy

The evanescent waves produced by optical tunneling have very important applications in optical imaging (Bryngdahl [1973]) and near-field microscopy (Courjon and Bainier [1994], Paesler and Moyer [1996], Courjon [2003]). The basic idea is to use evanescent waves as a probe, to overcome the diffraction limit of resolution in the far field, which is of the order of the optical wavelength. For electron microscopy, this has led to the atomic force microscope and the scanning tunneling electron microscope, that have been the foundations of the nanotechnology revolution. Scanning near-field optical microscopy (Vigoureux, Girard and Courjon [1989], Richards [2003]) and photon tunneling microscopy (Guerra [1990]) are optical analogues of these techniques.

14.2. Conclusions

The history of light tunneling is closely connected with the history of diffraction theory, as might have been expected, since both deal with the penetration of light into "forbidden regions". The classical diffraction theories of Young and Fresnel, based on Huygens' Principle and on the blocking effect, account for some of the most conspicuous features of diffraction, but they omit tunneling effects. Some features of tunneling are approximately represented in Keller's and Fock's theories. Newton's insights about diffraction, expounded in his remarkable Queries, are consistent with a tunneling interpretation and are vindicated by the analysis of scattering by spheres.

The Mie problem occupies a unique position in the study of scattering and diffraction, comparable to that of the Ising model in statistical physics. It is an exactly soluble and highly nontrivial model that is also realistic. Some of its features (in particular, those connected with Mie resonances) have been evaluated and measured with a precision approaching that of quantum electrodynamics. However, as was observed by Fock [1948], there is a long way from the rigorous theoretical solution of the Mie problem to an approximate one that can reproduce and explain all the qualitative and quantitative features of the phenomena it describes.

CAM theory has led to highly accurate approximations of the variety of diffraction effects contained in the Mie solution and to new physical pictures of these effects. The main missing ingredients were connected with light tunneling. While tunneling is often regarded as a strictly quantum-mechanical effect, detectable only in nuclear and particle physics, we have seen that its macroscopic consequences can actually be detected with the naked eye, through the darkening of clouds and in the bright colored rings of the glory, an almost pure tunneling effect.

The peculiar features of tunneling have led to several apparent paradoxes. If reflection is total, how can light penetrate into the rarer medium? For pictures of how this happens, see Lotsch [1968] and Zhang and Lee [2006]. Can tunneling lead to superluminal propagation? For details of the negative answer to this question, see Chiao and Steinberg [1997] and De Carvalho and Nussenzveig [2002].

Tunneling is highly non-intuitive. Does not the glory have a simpler explanation? For a discussion of this often-posed question, see Nussenzveig [2002]. CAM theory and tunneling account for all known features of the glory (Section 13), but it remains one of the most intricate scattering phenomena.

All quantitative treatments of tunneling employ some form of analytic continuation. That an oscillatory wave, without losses, can be converted into an exponential shape and then can recover the memory of its oscillatory character ultimately goes back to Euler's beautiful formula (1.2). As Galileo famously observed, the book of nature is written in the language of mathematics.

Acknowledgements

This work was supported by the Brazilian agencies Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Instituto do Milênio de Nanociências, Instituto do Milênio de Avanço Global e Integrado da Matemática Brasileira, Fundação de Amparo à Pesquisa do Rio de Janeiro (FAPERJ) and Fundação Universitária José Bonifácio (FUJB).

References

Adam, J.A., 2002, Phys. Rep. **356**, 229–365.
Airy, G.B., 1838, Trans. Camb. Phil. Soc. **6**, 379–402.
Antonoyiannakis, M.I., Pendry, J.B., 1997, Europhys. Lett. **40**, 613–618.
Antonoyiannakis, M.I., Pendry, J.B., 1999, Phys. Rev. B **60**, 2363–2374.
Ashkin, A., 1980, Science **210**, 1081–1088.
Babich, V.M., Kirpichnikova, N.Y., 1979, The Boundary-Layer Method in Diffraction Problems, Springer, Berlin, Heidelberg.
Baker, B., Copson, E.T., 1950, The Mathematical Theory of Huygens' Principle, Clarendon Press, Oxford.
Barta, A., Horváth, G., Bernáth, B., Meyer-Rochow, V.B., 2003, Appl. Opt. **42**, 399–405.
Beckmann, P., Franz, W., 1957, Z. Naturforschg. A **12**, 533–537.
Berry, M.V., 1966, Proc. Phys. Soc. **89**, 479–490.
Berry, M.V., 1969, Sci. Prog., Oxf. **57**, 43–64.
Berry, M.V., 2002, Phil. Trans. R. Soc. Lond. A **360**, 1023–1037.
Berry, M.V., Mount, K.E., 1972, Rep. Prog. Phys. **35**, 315–397.
Berry, M.V., Upstill, C., 1980, Prog. Opt. **18**, 257–346.
Blatt, J.M., Weisskopf, V., 1952, Theoretical Nuclear Physics, John Wiley, New York.
Born, M., Wolf, E., 1999, Principles of Optics, Cambridge University Press, Cambridge.
Borovikov, V.A., Kinber, V.Y., 1994, Geometrical Theory of Diffraction, Institution of Electrical Engineers, London.
Bouwkamp, C.J., 1954, Rep. Prog. Phys. **17**, 35–100.
Boyer, C.B., 1987, The Rainbow: From Myth to Mathematics, Princeton University Press, Princeton.
Braginsky, V.B., Gorodetsky, M.L., Ilchenko, V.S., 1989, Phys. Lett. A **137**, 393–397.
Braunbek, W., Laukien, G., 1952, Optik **9**, 174–179.
Bremmer, H., 1949, Terrestrial Radio Waves, Elsevier, New York.
Bricard, J., 1940, Ann. Phys. (Paris) **14**, 148.
Bryant, H.C., Jarmie, N., 1974, Scient. Amer. **231**, 60–67.
Bryngdahl, O., 1973, Prog. Opt. **11**, 167–221.
Bucerius, H., 1946, Optik **1**, 148.

Burin, A.L., Cao, H., 2004, J. Opt. Soc. Am. B **21**, 121–131.
Chabanov, A.A., Zenack, A.Z., 2001, Phys. Rev. Lett. **87**, 153901–153904.
Chang, R.K., Campillo, A.J., 1996, Optical Processes in Microcavities, World Scientific, Singapore.
Chapman, S.J., Lawry, J.M.H., Ockendon, J.R., Tew, R.H., 1997, SIAM Rev. **41**, 417–509.
Chester, C., Friedman, B., Ursell, F., 1957, Proc. Camb. Phil. Soc. **53**, 599–611.
Cheville, R.A., Mcgowan, R.W., Grischkowsky, D., 1998, Phys. Rev. Lett. **75**, 269–272.
Chiao, R.Y., Steinberg, A.M., 1997, Prog. Opt. **37**, 345–405.
Chylek, P., 1990, J. Opt. Soc. Am. A **7**, 1609–1613.
Chylek, P., Lesins, G.B., Videen, G., Wong, J.G.D., Pinnick, R.G., Ngo, D., Klett, J.D., 1996, J. Geophys. Res. **101**, 23365–23371.
Clebsch, A., 1863, J. für Math. **6**, 195–262.
Courjon, D., 2003, Near Field Microscopy and Near Field Optics, Imperial College Press, London.
Courjon, D., Bainier, C., 1994, Rep. Prog. Phys. **57**, 989–1028.
Dave, J.V., 1969, Appl. Opt. **8**, 1161–1167.
Davies, R., Ridgway, W.L., Kim, K.E., 1984, J. Atmos. Sci. **41**, 2126–2137.
De Alfaro, V., Regge, T., 1965, Potential Scattering, North-Holland, Amsterdam.
De Broglie, L., 1923, C. R. Acad. Sci. Paris **548**, 507.
De Carvalho, C.A., Nussenzveig, H.M., 2002, Phys. Rep. **364**, 83–174.
Debye, P., 1908, Physik. Z. **9**, 775–778.
Debye, P., 1909a, Ann. Phys. (Leipzig) **30**, 57–136.
Debye, P., 1909b, Math. Ann. **67**, 535–558.
Descartes, R., 1637, Discours de la Méthode, plus la Dioptrique, les Météores et la Géométrie, Jean Maire, Leiden.
Doron, E., Frischat, S.D., 1995, Phys. Rev. Lett. **75**, 3661–3664.
Eckart, C., 1930, Phys. Rev. **35**, 1303–1309.
Euler, L., 1746, Nova Theoria Lucis et Colorum, Berlin.
Euler, L., 1748, Introductio in Analysin Infinitorum, Lausanne.
Fermat, P., 1891, Oeuvres de Fermat, t. 2, Tannery, P., Henry, C. (Eds.), Paris.
Fiedler-Ferrari, N., Nussenzveig, H.M., Wiscombe, W.J., 1991, Phys. Rev. A **43**, 1005–1038.
Fields, M.H., Popp, J., Chang, R.K., 2000, Prog. Opt. **41**, 1–95.
Fock, V.A., 1948, Phil. Mag. **39**, 149–155.
Fock, V.A., 1965, Electromagnetic Diffraction and Propagation Problems, Pergamon, Oxford.
Fowler, R.H., Nordheim, L., 1928, Proc. Roy. Soc. (London) A **119**, 173–181.
Franz, W., 1957, Theorie der Beugung Elektromagnetischer Wellen, Springer, Berlin.
Franz, W., Depperman, K., 1952, Ann. Physik **10**, 361–373.
Fresnel, A., 1816, Ann. Chim. et Phys., 239.
Gamow, G., 1928, Z. Phys. **51**, 204–212.
Gamow, G., Critchfield, C.L., 1949, Theory of Atomic Nucleus and Nuclear Energy Sources, Clarendon Press, Oxford.
Goldstein, H., 1957, Classical Mechanics, Addison-Wesley, Reading, MA.
Goos, F., Hänchen, H., 1947, Ann. Phys. Lpz. (6) **1**, 333–346.
Gorodetsky, M.L., Savchenkov, A.A., Ilchenko, V.S., 1996, Opt. Lett. **21**, 453–455.
Grandy, W.T., 2000, Scattering of Waves from Large Spheres, Cambridge University Press, Cambridge.
Greenler, R., 1980, Rainbows, Halos and Glories, Cambridge University Press, Cambridge.
Grimaldi, F.M., 1665, Physico-mathesis de Lumine, Coloribus, et Iride, Boloniae.
Guerra, J.M., 1990, Appl. Opt. **29**, 3741–3752.
Guimarães, L., Nussenzveig, H.M., 1992, Opt. Commun. **89**, 363–369.
Guimarães, L.G., Nussenzveig, H.M., 1994, J. Mod. Opt. **41**, 625–647.
Gurney, R.W., Condon, E.U., 1928, Nature **122**, 439.

Hall, A.R., 1995, All Was Light – an Introduction to Newton's *Opticks*, Clarendon Press, Oxford.
Hall, E.E., 1902, Phys. Rev., 73–106.
Hamilton, W.R., 1828, Trans. Roy. Irish Acad. **15**, 69–174.
Hayter, A., 1973, A Voyage in Vain: Coleridge's Journey to Malta in 1804, Faber and Faber, London.
Heller, E.J., Tomsovic, S., 1993, Phys. Today **46** (7), 38–46.
Hönl, H., Maue, A.W., Westphal, K., 1961, Theorie der Beugung, in: Handb. d. Physik, vol. **25/1**, Springer, Berlin.
Hu, X.H., Yamilov, A., Noh, H., Cao, H., Seelig, E.W., Chang, R.P.H., 2004, J. Opt. Soc. Am. B **21**, 159–167.
Hund, F., 1927, Z. Phys. **43**, 805–826.
Huygens, C., 1690. Traité de la Lumière, Leyden.
Jackson, J.D., 1999, Phys. Rep. **320**, 27–36.
Joannopoulos, J.D., Meade, R.D., Winn, J.N., 1995, Photonic Crystals, Molding the Flow of Light, Princeton University Press, Princeton, NJ.
John, S., Wang, J., 1991, Phys. Rev. B **43**, 12772–12789.
John, S., 1987, Phys. Rev. Lett. **58**, 2486–2489.
John, S., 1991, Phys. Today **44** (5), 32–40.
Johnson, B.R., 1993, J. Opt. Soc. Am. A **10**, 343–352.
Jones, D.S., 1964, The Theory of Electromagnetism, Pergamon Press, Oxford.
Juan, J., Ulloa, A., 1748, Relación Histórica del Viage a la América Meridional, Antonio Marin, Madrid.
Keller, J.B., 1962, J. Opt. Soc. Am. **52**, 116–130.
Keller, J.B., 1979, SIAM Rev. **21**, 229–245.
Keller, J.B., 1985, IEEE Trans. Antennas Propag. **Ap-33**, 123–126.
Keller, O., 2002, Prog. Opt. **43**, 195–294.
Khare, V., 1975, Ph. D. thesis, University of Rochester.
Khare, V., 1982, Surface waves and rainbow effects in the optical glory, in: Boardman, A.D. (Ed.), Electromagnetic Surface Modes, Wiley, New York, pp. 417–464.
Khare, V., Nussenzveig, H.M., 1974, Phys. Rev. Lett. **33**, 976–980.
Khare, V., Nussenzveig, H.M., 1977a, Phys. Rev. Lett. **38**, 1279–1282.
Khare, V., Nussenzveig, H.M., 1977b, The theory of the glory, in: Landman, U. (Ed.), Statistical Mechanics and Statistical Methods in Theory and Application, Plenum Press, New York, pp. 723–764.
Kirchhoff, G., 1882, Berl. Ber., 641.
Können, G.P., De Boer, J.H., 1979, Appl. Opt. **18**, 1961–1965.
Kravtsov, Y.A., Forbes, G.W., Asatryan, A.A., 1999, Prog. Opt. **39**, 1–62.
Laven, P., 2004, J. Quant. Spectrosc. Radiat. Transf. **89**, 257–269.
Lawandy, N.M., Balachandran, R.M., Gomes, A.S.L., Sauvain, E., 1994, Nature **368**, 436–438.
Lazutkin, V.F., 1993, KAM Theory and Semiclassical Approximations to Eigenfunctions, Springer-Verlag, Berlin.
Lee Jr., R.L., 1998, Appl. Opt. **37**, 1506–1519.
Lee Jr., R.L., Fraser, A.B., 2001, The Rainbow Bridge – Rainbows in Art, Myth and Science, Pennsylvania State University Press, SPIE Press.
Leontovich, M.A., Fock, V.A., 1946, J. Phys. USSR **10**, 13–24.
Letokhov, V., 1968, Sov. Phys. JETP **26**, 835–840.
Lidorikis, E., Sigalas, M.M., Economou, E.N., Soukoulis, C.M., 1998, Phys. Rev. Lett. **81**, 1405–1408.
Logan, N.A., 1965, Proc. IEEE **53**, 773–785.
Lommel, E., 1885, Abhand. Bayer. Akad. **15**, 233–328.
Lord Rayleigh, 1872, Proc. London Math. Soc. **4**, 253.
Lord Rayleigh, 1910, Phil. Mag. **20**, 1001–1004.

Lorenz, L., 1890, Kgl. Danske Vidensk Selsk. Skrifter **6**, 1–62.
Lotsch, H.V.K., 1968, J. Opt. Soc. Am. **58**, 551–561.
Love, A.E.H., 1915, Phil. Trans. Roy. Soc. London A **215**, 105–131.
Ludwig, D., 1970, SIAM Rev. **12**, 325–331.
Luneburg, R.K., 1964, Mathematical Theory of Optics, University of California Press, Berkeley.
Lynch, D.K., Futterman, S.N., 1991, Appl. Opt. **30**, 3538–3541.
Lynch, D.K., Livingston, W., 2001, Color and Light in Nature, Cambridge University Press, Cambridge.
Marston, P.L., 1979, J. Opt. Soc. Am. **69**, 1205–1211.
Matsko, A.B., Savchenkov, A.A., Strekalov, D., Ilchenko, V.S., Maleki, L., 2005, IPN Progress Report 42-162, 1–51. http://ipnpr.jpl.nasa.gov.
Mekis, A., Nöckel, J.U., Chen, G., Stone, A.D., Chang, R.K., 1995, Phys. Rev. Lett. **75**, 2682–2685.
Merzbacher, E., 2002, Phys. Today **8**, 44–49.
Mie, G., 1908, Ann. Phys. (Leipzig) **25**, 377–442.
Minnaert, M.G.J., 1993, Light and Colours in the Outdoors, Springer, Berlin.
Mitchell, D.L., 2000, J. Atmos. Sci. **57**, 1311–1326.
Moroz, A., Tip, A., 1999, J. Phys. Condens. Matter **11**, 2503–2512.
Nauenberg, M., 2000, Comparison of Newton's diffraction measurements with the theory of Fresnel, in: Dalitz, R.H., Nauenberg, M. (Eds.), The Foundations of Newtonian Scholarship, World Scientific, Singapore.
Newton, I., 1672, Phil. Trans. Roy. Soc. **80**, 3075–3087.
Newton, I., 1757, An hypothesis explaining the properties of light, in: Birch, Thomas (Ed.), The History of the Royal Society, pp. 247–305. http://www.newtonproject.ic.ac.uk/texts/viewtext.php?id=Natp00002&mode=normalized.
Newton, I., 1946, Principia Mathematica Philosophiae Naturalis, University of California Press, Berkeley, CA.
Newton, I., 1952, Opticks, Dover Publications, New York (reprinted from the 1730 edition).
Newton, R.G., 1964, The Complex-j Plane, Benjamin, New York.
Nicholson, J.W., 1910, Proc. London Math. Soc. **9**, 67–80.
Noble, B., 1958, Methods Based on the Wiener–Hopf Technique for the Solution of Partial Differential Equations, Pergamon, London.
Nordheim, L., 1927, Z. Phys. **46**, 833–855.
Nussenzveig, H.M., 1959, An. Acad. Bras. Cienc. **31**, 515–521.
Nussenzveig, H.M., 1965, Ann. Phys. (NY) **34**, 23–95.
Nussenzveig, H.M., 1969a, J. Math. Phys. **10**, 82–124.
Nussenzveig, H.M., 1969b, J. Math. Phys. **10**, 125–176.
Nussenzveig, H.M., 1970, Applications of Regge poles to short-wavelength scattering, in: Bowcock, J.E. (Ed.), Methods and Problems of Theoretical Physics, North-Holland, Amsterdam, pp. 203–232.
Nussenzveig, H.M., 1972, Causality and Dispersion Relations, Academic Press, New York.
Nussenzveig, H.M., 1977, Scient. Amer. **236** (4), 116–127.
Nussenzveig, H.M., 1979, J. Opt. Soc. Am. **69**, 1068–1079.
Nussenzveig, H.M., 1988, J. Phys. A **21**, 81–109.
Nussenzveig, H.M., 1989, Comments At. Mol. Phys. **23**, 175–187.
Nussenzveig, H.M., 1992, Diffraction Effects in Semiclassical Scattering, Cambridge University Press, Cambridge.
Nussenzveig, H.M., 2002, Opt. Lett. **27**, 1379–1381.
Nussenzveig, H.M., 2003, Appl. Opt. **42**, 1588–1593.
Nussenzveig, H.M., Wiscombe, W.J., 1980a, Opt. Lett. **5**, 455–457.
Nussenzveig, H.M., Wiscombe, W.J., 1980b, Phys. Rev. Lett. **45**, 1490–1494.

Nussenzveig, H.M., Wiscombe, W.J., 1987, Phys. Rev. Lett. **59**, 1667–1670.
Nussenzveig, H.M., Wiscombe, W.J., 1991, Phys. Rev. A **43**, 2093–2112.
Olver, F.W.J., 1974, Asymptotics and Special Functions, Academic Press, New York.
Paesler, M.A., Moyer, P.J., 1996, Near-Field Optics: Theory, Instrumentation and Applications, Wiley, New York.
Pekeris, C.L., 1947, J. Appl. Phys. **18**, 667–680.
Pernter, J.M., Exner, F.M., 1910, Meteorologische Optik, W. Braumüller, Vienna.
Poincaré, H., 1910, Rendiconti Circ. Mat. Palermo **29**, 169–259.
Primack, H., Schanz, H., Smilansky, U., Ussishkin, I., 1997, J. Phys. A **30**, 6693–6723.
Pulfrich, C., 1888, Ann. Phys. Chem. (Leipzig) **33**, 209–212.
Qian, S.-X., Snow, J.B., Tzeng, H.-M., Chang, R.K., 1986, Science **231**, 486–488.
Ray, B., 1923, Nature **111**, 183.
Regge, T., 1959, Nuovo Cimento **14**, 951–976.
Regge, T., 1960, Nuovo Cimento **18**, 947–956.
Richards, D., 2003, Phil. Trans. R. Soc. Lond. A **361**, 2843–2857.
Roll, G., Kaiser, T., Lange, S., Schweiger, G., 1998, J. Opt. Soc. Am. A **15**, 2879–2891.
Ronchi, V., 1952, Storia della Luce, Nicola Zanichelli, Bologna.
Rubinowicz, A., 1917, Ann. d. Phys. Lpz. **53**, 257.
Saengkaew, S., Charinpatnikul, T., Vanisri, H., Tanthapanichakoon, W., Mees, L., Gouesbet, G., Grehan, G., 2006, Opt. Comm. **259**, 7–13.
Schottky, W., 1931, Phys. Z. **32**, 833.
Schrödinger, W., 1926, Ann. Phys. (Leipzig) (4) **79**, 489–527.
Sheng, P., 1990, Scattering and Localization of Classical Waves in Random Media, World Scientific, Singapore.
Sheng, P., 1995, Introduction to Wave Scattering, Localization and Mesoscopic Phenomena, Academic Press, New York.
Shipley, S.T., Weinman, J.A., 1968, J. Opt. Soc. Am. **68**, 130–134.
Sommerfeld, A., 1896, Mathem. Ann. **47**, 317.
Sommerfeld, A., 1954, Optics, Academic Press, New York.
Stuewer, R.H., 2006, Science and Education **15**, 521–530.
Titchmarsh, E.C., 1937, Introduction to the Theory of Fourier Integrals, Oxford University Press, Oxford.
Tran, N.H., Dutriaux, L., Balcou, P., Le Floch, A., Bretenaker, F., 1995, Opt. Lett. **20**, 1233–1235.
Tricker, R.A.R., 1970, Introduction to Meteorological Optics, American Elsevier, New York.
Tzeng, H.M., Wall, K.F., Long, M.B., Chang, R.K., 1984, Opt. Lett. **9**, 499–501.
Vahala, K.J., 2003, Nature **424**, 840–846.
Van Albada, M.P., Van Tiggelen, B.A., Lagendijk, A., Tip, A., 1991, Phys. Rev. Lett. **66**, 3132–3135.
Van De Hulst, H.C., 1947, J. Opt. Soc. Am. **37**, 16–22.
Van De Hulst, H.C., 1957, Light Scattering by Small Particles, Wiley, New York.
Van Der Pol, B., Bremmer, H., 1937, Phil. Mag. **24**, 141–176, 825–864.
Van Tiggelen, B.A., Lagendijk, A., Van Albada, M.P., Tip, A., 1992, Phys. Rev. B **45**, 12233–12243.
Vandenbem, C., Vigneron, J.P., 2005, J. Opt. Soc. Am. A **22**, 1042–1047.
Vigoureux, J.M., Girard, C., Courjon, D., 1989, Opt. Lett. **14**, 1039–1041.
Von Klitzing, W., Long, R., Ilchenko, V.S., Hare, J., Lefèvre-Seguin, V., 2001, New J. Phys. **3**, 14.1–14.14.
Wang, R.T., Van De Hulst, H.C., 1991, Appl. Opt. **30**, 106–117.
Watson, G.N., 1918, Proc. Roy. Soc. (London) A **95**, 83–99.
Watson, G.N., 1919, Proc. Roy Soc. (London) A **95**, 546–563.
Watson, W., Brownrigg, W., 1749, Phil. Trans. **46**, 584–596.
White, F.P., 1922, Proc. Roy. Soc. (London) A **100**, 505–525.

Wilson, C.T.R., 1927, Nobel lecture, in: Santesson, C.G. (Ed.), Les Prix Nobel en 1927, Nobel Foundation, Stockholm.

Wolf, E., Nieto-Vesperinas, M., 1985, J. Opt. Soc. Am. A **2**, 886–890.

Yeang, C.P., 2003, Hist. Stud. Phys. Biol. Sci. **33**, 369–403.

Young, T., 1802, Phil. Trans. R. Soc. Lond. **92**, 12–48.

Young, T., 1804, Phil. Trans. R. Soc. Lond. **94**, 1–16.

Young, T., 1807, A Course of Lectures on Natural Philosophy and the Mechanical Arts, J. Johnson, London.

Zender, C.S., Talamantes, J., 2006, J. Quant. Spectrosc. Radiat. Transf. **98**, 122–129.

Zhang, Z.M., Lee, B.J., 2006, Opt. Express **21**, 9963–9970.