ELSEVIER

Physics Reports 356 (2002) 229–365

PHYSICS REPORTS

www.elsevier.com/locate/physrep

The mathematical physics of rainbows and glories

John A. Adam

Department of Mathematics and Statistics, Old Dominion University, Norfolk, VA 23529, USA

Received May 2001; editors: J. Eichler, T.F. Gallagher

Contents

E-mail address: jadam@odu.edu (J.A. Adam).

0370-1573/02/$ - see front matter
PII: S0370-1573(01)00076-X

Abstract

A detailed qualitative summary of the optical rainbow is provided at several complementary levels of description, including geometrical optics (ray theory), the Airy approximation, Mie scattering theory, the complex angular momentum (CAM) method, and catastrophe theory. The phenomenon known commonly as the glory is also discussed from both physical and mathematical points of view: backward glories, rainbow-glories and forward glories. While both rainbows and glories result from scattering of the incident radiation, the primary rainbow arises from scattering at about 138° from the forward direction, whereas the (backward) glory is associated with scattering very close to the backward direction. In fact, it is a more complex phenomenon physically than the rainbow, involving a variety of different effects (including surface waves) associated with the scattering droplet. Both sets of optical phenomena—rainbows and glories—have their counterparts in atomic, molecular and nuclear scattering, and these are addressed also. The conceptual foundations for understanding rainbows, glories and their associated features range from classical geometrical optics, through quantum mechanics (in particular scattering from a square well potential; the associated Regge poles and scattering amplitude functions) to diffraction catastrophes. Both the scalar and the electromagnetic scattering problems are reviewed, the latter providing details about the polarization of the rainbow that the scalar problem cannot address. The basis for the complex angular momentum (CAM) theory (used in both types of scattering problem) is a modification of the Watson transform, developed by Watson in the early part of this century in the study of radio wave diffraction around the earth. This modified Watson transform enables a valuable and accurate approximation to be made to the Mie partial-wave series, which while exact, converges very slowly at high frequencies. The theory and many applications of the CAM method were developed in a fundamental series of papers by Nussenzveig and co-workers (including an important interpretation based on the concept of *tunneling*), but many other contributions have been made to the understanding of these beautiful phenomena, including descriptions in terms of so-called diffraction catastrophes. The rainbow is a fine example of an observable event which may be described at many levels of mathematical sophistication using distinct mathematical approaches, and in so doing the connections between several seemingly unrelated areas within physics become evident.

PACS: 42.15.–i; 42.25.–p; 42.25.FX; 42.68.M$_j$

Keywords: Rainbow; Glory; Mie theory; Scattering; Complex angular momentum; Diffraction catastrophe

Like the appearance of a rainbow in the clouds on a rainy day, so was the radiance around him. This was the appearance of the likeness of the glory of the Lord...
The Book of Ezekiel, chapter 1, verse 28 (New International Version of the Bible).

1. Introduction

1.1. Structure and philosophy of the review

The rainbow is at one and the same time one of the most beautiful visual displays in nature and, in a sense, an intangible phenomenon. It is illusory in that it is not of course a solid arch, but like mirages, it is nonetheless real. It can be seen and photographed, and described as a phenomenon of mathematical physics, but it cannot be located at a specific place, only in a particular direction. What then *is* a rainbow? Since many levels of description will be encountered along the way, the answers to this question will take us on a rather long but fascinating journey in the footsteps of those who have made significant contributions to the subject of "light scattering by small particles". Let us first 'listen' to what others have written about rainbows and the mathematical tools with which to understand them.

"Rainbows have long been a source of inspiration both for those who would prefer to treat them impressionistically or mathematically. The attraction to this phenomenon of Déscartes, Newton, and Young, among others, has resulted in the formulation and testing of some of the most fundamental principles of mathematical physics." K. Sassen [1].

"The rainbow is a bridge between two cultures: poets and scientists alike have long been challenged to describe it... Some of the most powerful tools of mathematical physics were devised explicitly to deal with the problem of the rainbow and with closely related problems. Indeed, the rainbow has served as a touchstone for testing theories of optics. With the more successful of those theories it is now possible to describe the rainbow mathematically, that is, to predict the distribution of light in the sky. The same methods can also be applied to related phenomena, such as the bright ring of color called the glory, and even to other kinds of rainbows, such as atomic and nuclear ones." H.M. Nussenzveig [2].

"Probably no mathematical structure is richer, in terms of the variety of physical situations to which it can be applied, than the equations and techniques that constitute wave theory. Eigenvalues and eigenfunctions, Hilbert spaces and abstract quantum mechanics, numerical Fourier analysis, the wave equations of Helmholtz (optics, sound, radio), Schrödinger (electrons in matter), Dirac (fast electrons) and Klein–Gordon (mesons), variational methods, scattering theory, asymptotic evaluation of integrals (ship waves, tidal waves, radio waves around the earth, diffraction of light)—examples such as these jostle together to prove the proposition." M.V. Berry [3].

The three quotations above provide a succinct yet comprehensive survey of the topic addressed in this review: the mathematical physics of rainbows and glories. An attempt has been made to provide complementary levels of description of the rainbow and related phenomena; this mirrors

to some extent the historical development of the subject, but at a deeper level it addresses the fact that, in order to understand a given phenomenon as fully as possible, it is necessary to study it at as many complementary levels of description as possible. In the present context, this means both descriptive and mathematical accounts of the rainbow and related phenomena, the latter account forming the basis of the paper: it is subdivided into the various approaches and levels of mathematical sophistication that have characterized the subject from the investigations of Déscartes down to the present era.

There are several classic books and important papers that have been drawn on frequently throughout this paper: the book by Greenler on rainbows, halos and glories [4] has proved invaluable for the descriptive physics in this introduction; the article by Nussenzveig [2] from which the second quotation is taken is an excellent introduction to both the physics and the qualitative description of the various mathematical theories that exist for the rainbow. The two papers [5,6] by the same author constitute a major thread running throughout this article, but particularly so in Section 5 (complex angular momentum theory). Van de Hulst's book on light scattering by small particles [7] is a classic in the field, and for that reason is often cited, both in this article and in many of the references. Indeed, to quote from Section 13.2 in that book

> "The rainbow is one of the most beautiful phenomena in nature. It has inspired art and mythology in all people, and it has been a pleasure and challenge to the mathematical physicists of four centuries. A person browsing through the old literature receives the impression that a certain affection for this problem pervades even the driest computations."

The very next sentence in the above citation is most à propos, since it also accurately reflects *this* author's hopes for the *present* article

> "The writer hopes that the following report, though it has to be concise and must leave out most of the history, will to some extent demonstrate this mathematical beauty."

Another important book, less comprehensive and mathematically complete than [7], but nonetheless extremely valuable from both physical and mathematical viewpoints, is that by Tricker on meteorological optics [8]. The book by Bohren and Huffman [9] provides a great deal of information on the absorption and scattering of light by small particles and diverse applications, with many references. At the time of writing, the most comprehensive and up-to-date book on diffraction effects in semi-classical scattering theory and its applications is that by Nussenzveig [10]. It is a seminal work, and incorporates many of the topics addressed here (amongst others), and should be consulted by anyone wishing to study the subject from an expert in the field.

No attempt is made here to discuss the historical development of the theory of the rainbow except insofar as it is germane to the context; that topic is superbly treated in the book by Boyer [11]. A noteworthy and somewhat technical account of more recent historical developments can be found in the review article by Logan [12]. It is a survey of early studies in the scattering of plane waves by a sphere. This is extremely important and interesting from a historical perspective: Logan provides 103 references, and in so doing notes that the literature appears to be characterized by writers who, it seems, failed to recognize the significance of the contributions made by their predecessors and contemporaries. There are several instances of the "rediscovery of the wheel" (so to speak), and the author has performed a valuable task in identifying the "lost" contributions to the subject.

In the rest of this introduction, elementary physical and mathematical descriptions (e.g. geometrical optics) of the rainbow are provided. Section 2 addresses the capacity of Airy's theory to account for, among other things, the primary bow and its associated supernumeraries. In Section 3 glories (backward, forward and rainbow-modified) are described; Section 4 addresses the semiclassical description of atomic, molecular and nuclear rainbows and glories, as well as diffractive/tunneling effects in these areas and in particle physics. Section 5 is devoted to a summary of the complex angular momentum (CAM) theory for the scalar scattering problem; a fascinating and very powerful tool for understanding many of the subtleties of wave scattering by both impenetrable and transparent (or even absorbing) spheres. It is appropriate to include the topic of the glory again in this section, because much of the CAM theory developed by Nussenzveig addressed both rainbows and glories. In Section 6 further developments are discussed, including the full electromagnetic problem (Mie scattering theory), polarization, alternative models, the effects of non-spherical droplets, and a comparison of theories. A brief account of rainbows and glories in atomic, molecular and nuclear scattering is also provided. Section 7 contains an account of the relevant aspects of 'diffraction catastrophe' scattering to the problem at hand. The summary in Section 8 precedes four appendices: Appendix A summarizes aspects of classical scattering relevant to the present article, and previews rainbow and glory scattering, forward peaking and orbiting, all of which reappear in a semiclassical context. Appendix B provides brief details of the Airy and Fock functions referred to in the article. Appendix C contains a brief account of the Watson transform and its modification to become the basis of the CAM theory discussed earlier, while Appendix D is a brief summary of the Chester–Friedmann–Ursell (CFU) method.

1.2. The rainbow: elementary physical features

Let us return to the question asked above: what then *is* a rainbow? It is sunlight, displaced by reflection and dispersed by refraction in raindrops. It is seen by an observer with his or her back to the sun (under appropriate circumstances). As shown in Fig. 1(a), the primary rainbow, which is the lowest and brightest of two that may be seen, is formed from two refractions and one reflection in myriads of raindrops (the path for the *secondary* rainbow is shown in Fig. 1(b); Fig. 1(c) illustrates notation referred to in the text). For our purposes we may consider the path of a ray of light through a single drop of rain, for the geometry is the same for all such drops and a given observer. Furthermore, we can appreciate most of the common features of the rainbow by using the ray theory of light; the wave theory of light is needed to discuss the finer features such as the supernumerary bows (discussed below).

The first satisfactory explanation for the existence and shape of the rainbow was given by René Déscartes in 1637 (he was unable to account for the colors however; it was not until thirty years later that Newton remedied this situation). Déscartes used a combination of experiment and theory to deduce that both the primary and secondary bow (larger in angular diameter and fainter than the primary) are caused by refraction and reflection in spherical raindrops. He correctly surmised that he could reproduce these features by passing light through a large water-filled flask—a really big "raindrop". Since the laws of refraction and reflection had been formulated some 16 years before the publication of Déscartes' treatise by the Dutch scientist Snell, Déscartes could calculate and trace the fate of parallel rays from the sun impinging on

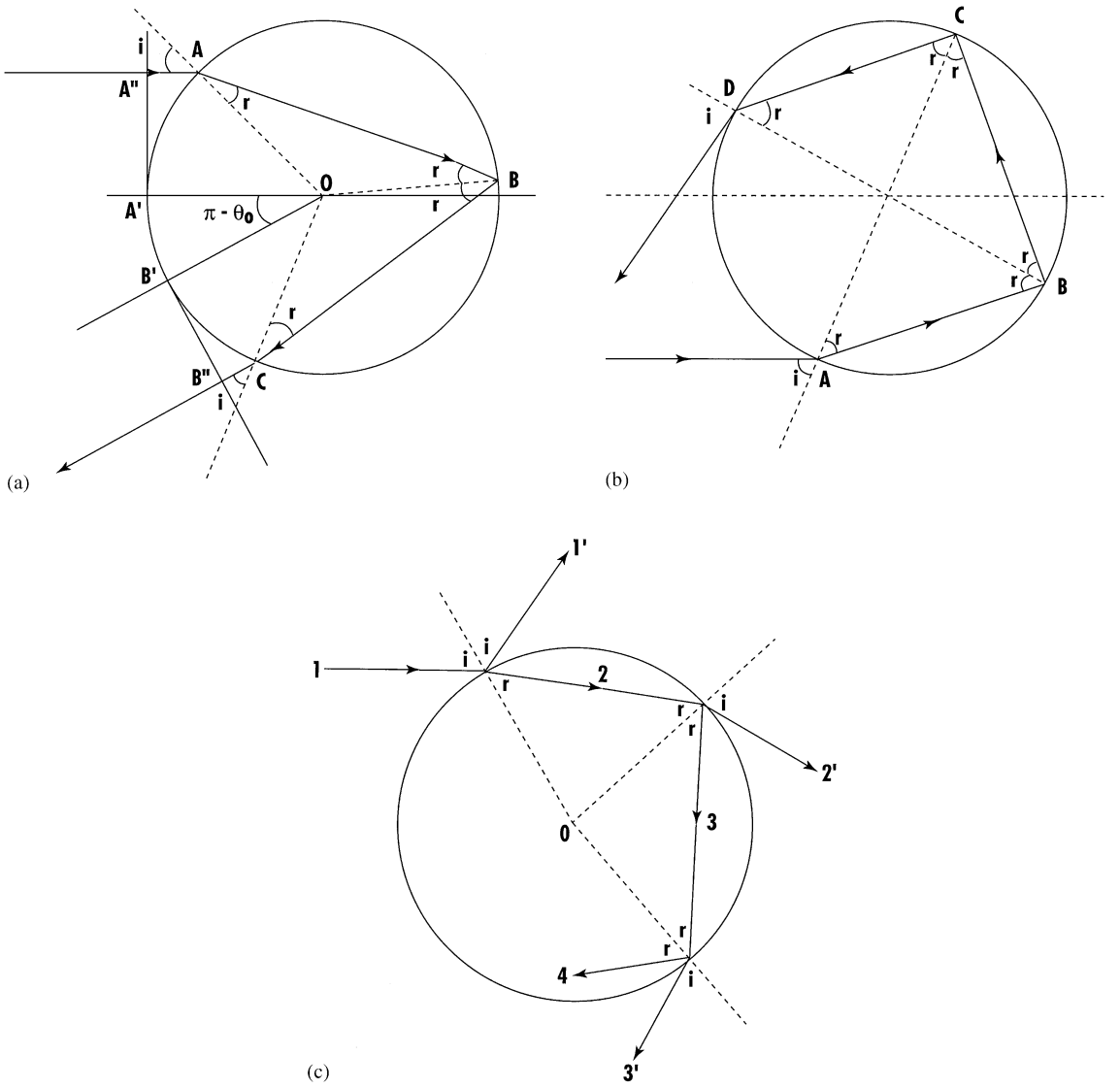

Fig. 1. (a) the basic geometry of ray paths in a spherical raindrop for the primary rainbow; O is the center of the drop, and i, r, respectively, are the angles of incidence and refraction. The incident and exiting 'wavefronts' $A'A''$ and $B'B''$ are also shown (refer to Section 2.1), (b) the corresponding geometry for the formation of the secondary rainbow, (c) the geometry of an incident ray (1) showing the externally reflected ray ($1'$), the transmitted ray (2) and externally refracted ray ($2'$), etc. All externally transmitted rays are denoted by primed numbers.

a spherical raindrop. As can be seen from Fig. 2, such rays exit the drop having been deviated from their original direction by varying but large amounts. The ray along the central axis (#1) will be deviated by exactly 180°, whereas above this point of entry the angle of deviation decreases until a minimum value of about 138° occurs (for yellow light; other colors have slightly different minimum deviation angles). For rays impinging still higher above the axis the

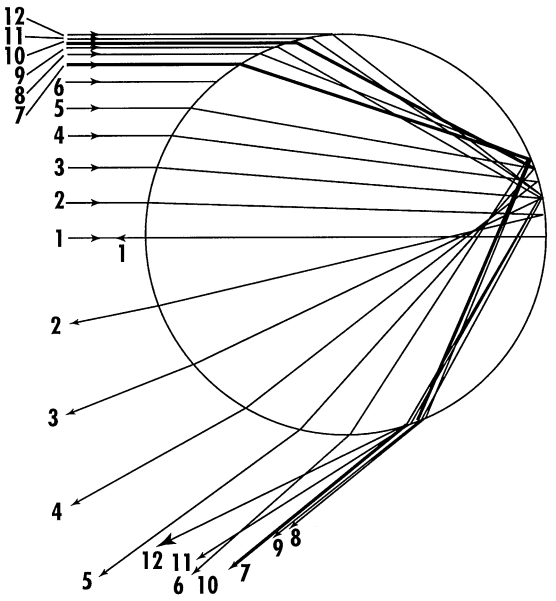

Fig. 2. The paths of several rays through a spherical raindrop illustrating the differences in total deflection angle for different angles of incidence (or equivalently, different impact parameters). The ray #7 is called the rainbow ray and this ray defines the minimum angle of deflection in the primary rainbow. (Redrawn from [4].)

deviation angle increases again. The ray of minimum deviation (#7 in Fig. 2) is called the *rainbow ray*. The significant feature of this geometrical system is that the rays leaving the drop are not uniformly spaced: those "near" the minimum deviation angle are concentrated around it, whereas those deviated by larger angles are spaced more widely (see also Fig. 3).

Put differently, in a small (say half a degree) angle on either side of the rainbow angle ($\simeq 138^\circ$) there are more rays emerging than in any other one degree interval. It is this concentration of rays that gives rise to the (primary) rainbow, at least as far as its light intensity is concerned. In this sense it is similar to a caustic formed on the surface of the tea in a cup when appropriately illuminated. The rainbow seen by any given observer consists of those deviated rays that of course enter his eye. These are those that are deviated by about 138° from their original direction (for the primary rainbow). Thus the rainbow can be seen by looking in any direction that is about 42° away from the line joining one's eye to the shadow of one's head (the antisolar point); the 42° angle is supplementary to the rainbow angle. This criterion defines a circular arc (or a full circle if the observer is above the raincloud) around the antisolar point and hence all raindrops at that angle will contribute to one's primary rainbow. Of course, on level ground, at most a semi-circular arc will be seen (i.e. if the sun is close to setting or has just risen), and usually it will be less than that: full circular rainbows can be seen from time to time at high altitudes on land or from aircraft. In summary, the primary rainbow is formed by the deflected rays from all the raindrops that lie on the surface of a cone with vertex (or apex) at the eye, axis along the antisolar direction and semi-vertex angle of 42°. The same statement holds for the secondary rainbow if the semi-vertex angle is about 51° (the supplement of a 129° deviation).

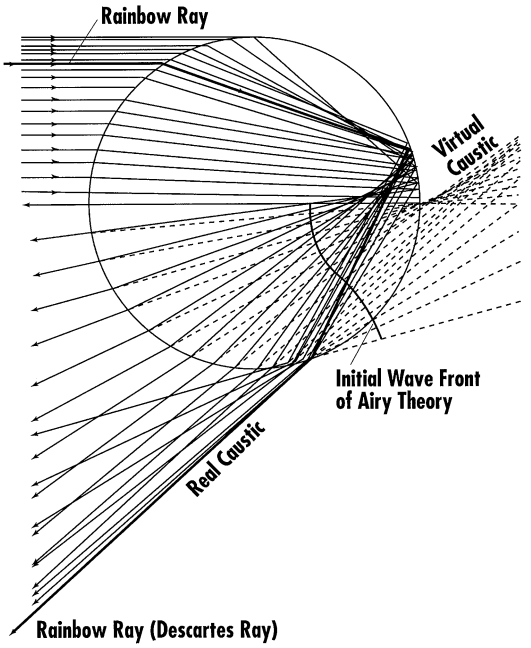

Fig. 3. A more detailed version of Fig. 2 showing the "Airy wavefront" (which is perpendicular to all rays of class 3' in Fig. 1(c)). Also shown, following Nussenzveig [2] are the caustics (one real, one imaginary) of these rays; these are envelopes of the ray system. (Redrawn from [2].)

These cones will be different for each observer, so each person has his or her own personal rainbow.

Up to this point, we have been describing a generic, colorless type of rainbow. Blue and violet light get refracted more than red light: the actual amount depends on the index of refraction of the raindrop, and the calculations thereof vary slightly in the literature because the wavelengths chosen for "red" and "violet" may differ slightly. Thus, for a wavelength of 6563 Å (Angstrom units; $1\ \text{Å} = 10^{-10}$ m) the cone semi-angle is about 42.3°, whereas for violet light of 4047 Å wavelength, the cone semi-angle is about 40.6°, about a 1.7° angular spread for the primary bow. A similar spread (dispersion) occurs for the secondary bow, but the additional reflection reverses the sequence of colors, so the red in this bow is on the *inside* of the arc.

In principle more than two internal reflections may take place inside each raindrop, so higher-order rainbows (tertiary, quaternary, etc.) are possible. It is possible to derive the angular size of such a rainbow after any given number of reflections (Newton was the first to do this). Newton's contemporary, Edmund Halley found that the third rainbow arc should appear as a circle of angular radius about 40° around the sun itself. The fact that the sky background is so bright in this vicinity, coupled with the intrinsic faintness of the bow itself would make such a bow almost impossible, if not impossible to see (but see [13]). Jearl Walker has used a laser beam to illuminate a single drop of water and traced rainbows up to the 13th order, their positions agreeing closely with predictions. Others have traced 19 rainbows under similar laboratory conditions [14].

The sky below the primary rainbow is often noticeably brighter than the sky outside it; indeed the region between the primary and secondary bow is called Alexander's dark band (after Alexander of Aphrodisias who studied it in connection with Aristotle's (incorrect) theory of the rainbow). Raindrops scatter incident sunlight in essentially all directions, but as we have seen, the rainbow is a consequence of a "caustic" or concentration of such scattered light in a particular region of the sky. The reason the inside of the primary bow (i.e. inside the cone) is bright is that all the raindrops in the interior of the cone reflect light to the eye also, (some occurring from direct reflections at their surfaces) but it is not as intense as the rainbow light, and it is composed of many colors intermixed. Similarly, *outside* the secondary bow a similar (but less obvious) effect occurs (see Figs. 4.3(b) and 4.7 in [15]). Much of the scattered light then, comes from raindrops through which sunlight is refracted and reflected: these rays do not emerge between the 42° and 51° angle. This dark angular band is not completely dark, of course, because the surfaces of raindrops reflect light into it; the reduction of intensity, however, is certainly noticeable.

Another commonly observed feature of the rainbow is that when the sun is near the horizon, the nearly vertical arcs of the rainbow near the ground are often brighter than the upper part of the arc. The reason for this appears to be the presence of drops with varying sizes. Drops smaller than, say 0.2–0.3 mm (about $1/100$ in) are spherical: surface tension is quite sufficient to keep the distorting effects of aerodynamical forces at bay. Larger drops become more oblate in shape, maintaining a circular cross section horizontally, but not vertically. They can contribute significantly to the intensity of the rainbow because of their size ($\approx 1/25$ in, i.e. ≈ 1 mm or larger) but can only do so when they are "low" on the cone, for the light is scattered in a horizontal plane in the exact way it should to produce a rainbow. These drops do not contribute significantly near the top of the arc because of their non-circular cross section for scattering. This will be discussed in more detail in Section 6.4. Small drops, on the other hand contribute to all portions of the rainbow.

Drop size, as implied above, can make a considerable difference to the intensity and color of the rainbow. The "best" bows are formed when the drop diameter is $\gtrsim 1$ mm; as the size decreases the coloration and general definition of the rainbow becomes poorer. Ultimately, when the drops are about 0.05 mm or smaller in diameter, a broad, faint, white arc called a fogbow occurs. When sunlight passes through these very tiny droplets the phenomenon of diffraction becomes important. Essentially, due to the wave nature of light, interactions of light with objects comparable to, (or not too much larger than) a typical wavelength, a light beam will spread out. Thus the rainbow colors are broadened and overlap, giving rise in extreme cases to a broad white fogbow (or cloudbow, since droplets in those typically produce such bows). As Greenler points out, these white rainbows may sometimes be noted while flying above a smooth featureless cloud bank. The rainbow cone intersects the horizontal cloud layer in a hyperbola if the sun's elevation is less than 42° (or an ellipse if it is greater than 42°) as familiarity with the conic sections assures us. This phenomenon can also be seen as a "dewbow" on a lawn when the sun is low in the eastern sky. Other related phenomena of interest, such as reflected-light rainbows (produced by reflection from a surface of water behind the observer) and reflected rainbows (produced by reflection from a surface of water in front of the observer) can be found in the book by Greenler [4].

Another feature of rainbows produced by smaller drops is also related to the wave nature of light. This time the phenomenon is interference and it produces supernumerary bows. These are a

series of faint pink or green arcs (2–4 perhaps) just *beneath* the top of the primary bow, or much more rarely, just above the top of the secondary bow. They rarely extend around the fully visible arc, for reasons that are again related to drop size. Two rays that enter the drop on either side of the rainbow ray (the ray of minimum deviation) may exit the drop in parallel paths; this certainly *will* happen for appropriately incident rays. By considering the wavefronts (perpendicular to the rays), it is clear that if the incident waves are in phase (i.e. crests and troughs aligned with crests and troughs) the emerging rays will *not* be in phase, in general. Inside the drop they travel along paths of different length. Depending on whether this path difference is an integral number of wavelengths or an odd integral number of half-wavelengths, these waves will reinforce each other (constructive interference) or annihilate each other (destructive interference). Obviously partially constructive/destructive interference can occur if the path difference does not meet the above criteria exactly. Where waves reinforce one another, the intensity of light will be enhanced; conversely, where they annihilate one another the intensity will be reduced. Since these beams of light will exit the raindrop at a smaller angle to the axis than the Déscartes ray, the net effect for an observer looking in this general direction will be a series of light and dark bands just inside the primary bow.

The angular spacing of these bands depends on the size of the droplets producing them (see Fig. 4). The width of individual bands *and* the spacing between them decreases as the drops get larger. If drops of many different sizes are present, these supernumerary arcs tend to overlap somewhat and smear out what would have been obvious interference bands for droplets of uniform size. This is why these pale blue or pink or green bands are then most noticeable near the top of the rainbow: it is the smaller drops that contribute to this part of the bow, and these may represent a rather narrow range of sizes. Nearer the horizon a wide range of drop size contributes to the bow, but as we have seen, at the same time it tends to blur the interference bands.

There are many complementary levels of mathematical techniques with which one can describe the formation and structure of the rainbow. In this section we examine the broad features using only elementary calculus; one account of the basic mathematics is described in [16] (and a useful graphical account is provided in [17]); a thorough treatment of the related physics of multiple rainbows may be found in the article by Walker [14]. More mathematical details are found in two classical works, one by Humphreys [18] and the other by Tricker [8]. In [16] the description of the location and color of the rainbow is approached as an exercise in mathematical modeling: the laws of reflection and refraction are stated along with the underlying assumptions about the deviation of sunlight by a raindrop (e.g. the sphericity of the drop; parallel light "rays" from the sun, neither of which are strictly true, of course; the raindrop being fixed during the scattering process, amongst other 'axioms'). At this elementary level of description involving geometrical optics, dispersion, geometry, trigonometry and calculus of a single variable, a reasonably satisfactory "explanation" of the main features of the rainbow is possible. Incorporation of wave interference and diffraction, amongst other things, is necessary to take the description further, and as we shall see, this can lead to very sophisticated models indeed.

A most fascinating aspect of the topic at hand is the fact that rainbows and glories can also be produced in atomic, molecular and nuclear scattering experiments, thus illustrating at a profound level the wave/particle duality of matter and radiation. Rays in geometrical optics become

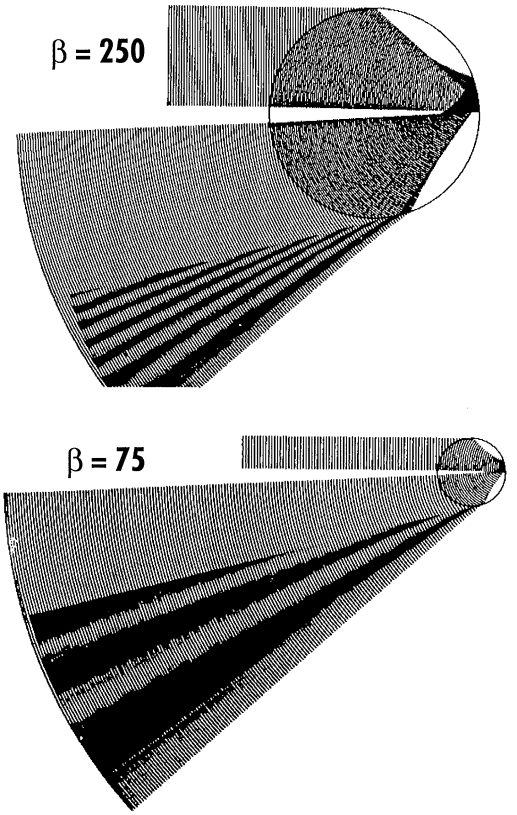

Fig. 4. A wavefront version of the ray geometry shown in Figs. 2 and 3; based on work by Fraser [141] this shows the folded wavefronts emanating from a spherical raindrop of radius a for different size parameters $\beta = 2\pi a/\lambda = ka$. Here k is the wavenumber of the light; note that the fringes correspond to the supernumerary bows, and that larger drops (here $\beta = 250$) produce narrow, closely spaced supernumeraries whereas smaller drops ($\beta = 75$) produce broader, more widely spaced ones.

particle trajectories, and the refraction of such rays corresponds to particle deflection under the influence of atomic or nuclear forces. This puts a new perspective on Hamilton's recognition of the analogy between geometrical optics and classical particle mechanics, an analogy that has had a powerful influence on the development of mathematical physics since the mid-19th century.

There are obvious differences between "trajectories" in geometrical optics, which are line segments with discontinuities in direction, and those in particle scattering systems, which are curved and smoothly varying, in fact, differentiable with respect to arc length or impact parameter. This latter parameter is the distance of an incident ray or particle trajectory from the central axis of the droplet or scattering object, and in the case of the optical rainbow it lies between zero and the droplet radius. In scattering problems it can extend in principle to infinity. There is a one-to-one correspondence between impact parameter and deflection angle (see Appendix A), and there is a unique trajectory for which the (local) angular deflection is a maximum.

By analogy this is the "rainbow angle" for the interaction because a concentration of scattered particles arises near this angle. However, the analogy extends much further than this, for as will be noted in Section 4, Ford and Wheeler [19] carried out a wave–mechanical analysis of atomic and nuclear rainbows, showing that interference between trajectories emerging in the same direction gives rise to *supernumerary bows* (peaks in intensity). Furthermore, it has been possible to formulate an analogue of Airy's theory (Section 2) for particle scattering. In 1964, Hundhausen and Pauly [20] made the first observations of an atomic rainbow with an experiment in which sodium atoms were scattered by mercury atoms. The "primary" bow (there is no secondary) and two supernumeraries were detected; subsequent experiments have revealed structure at yet finer scales. Just as careful observations of optical rainbows yield information on the scatterering centers—the raindrops, the refractive index, etc., so too do experiments such as these. The atomic (and for that matter, molecular and nuclear) rainbow angle depends on the *strength* of the interaction, specifically the attractive part, and the *range* of the interaction in turn determines the positions of the supernumerary bows. This topic will be discussed at greater length in Section 6.5.

Returning to the optical rainbow, it is important to note that the theories of Déscartes, Newton and even Young's interference theory all predicted an abrupt transition between regions of illumination and shadow (as at the edges of Alexander's dark band when rays only giving rise to the primary and secondary bows are considered). In the wave theory of light such sharp boundaries are softened by diffraction—and this should have been Young's conclusion, incidentally, for his was a wave theory [2]. In 1835, Potter showed that the rainbow ray can be interpreted as a *caustic*, i.e. the envelope of the system of rays comprising the rainbow. The word caustic means "burning curve", and caustics are associated with regions of high intensity illumination (as we shall below, geometrical optics predicts an infinite intensity there). When the emerging rays comprising the rainbow are extended backward through the drop, another caustic (a virtual one) is formed, associated with the real caustic on the illuminated side of the drop. Typically, the number of rays differs by two on each side of a caustic at any given point, so the rainbow problem is essentially that of determining the intensity of (scattered) light in the neighborhood of a caustic (see Fig. 3, and also Fig. 5). This was exactly what Airy attempted to do several years later in 1838 [21]. The principle behind Airy's approach was established by Huygens in the 17th century: Huygens' principle regards every point of a wavefront as a secondary source of waves, which in turn defines a new wavefront and hence determines the subsequent propagation of the wave. Airy reasoned that if one knew the amplitude distribution of the waves along any complete wavefront in a raindrop, the distribution at any other point could be determined by Huygens' principle. However, the problem is to find the initial amplitude distribution. Airy chose as his starting point a wavefront surface inside the raindrop, the surface being orthogonal to all the rays which comprise the primary bow; this surface has a point of inflection wherever it intersects the ray of minimum deviation—the rainbow ray. Using the standard assumptions of diffraction theory, he formulated the local intensity of scattered light in terms of a "rainbow integral", subsequently renamed the *Airy integral* in his honor (it is related to the now familiar Airy function; see Appendix B. Note that there is another so-called Airy function in optics which has nothing to do with this phenomenon [22]). Qualitatively, the intensity distribution so produced is similar to that associated with the shadow of a straight edge, particularly when external reflection is included (see [23, Chapter 8]).

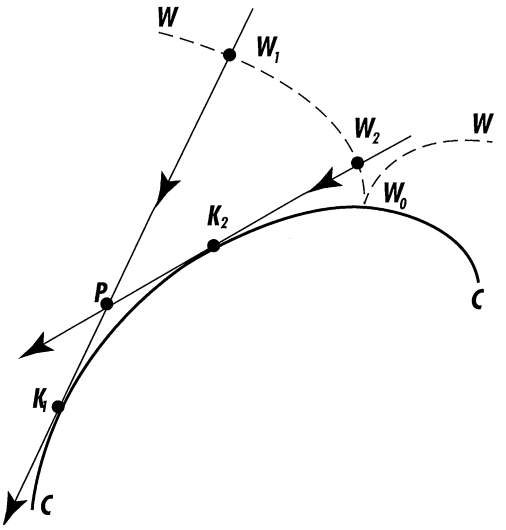

Fig. 5. A caustic curve (bold curve CC) showing rays and corresponding wavefronts on the "lit" side; in simplistic terms the caustic is a shadow boundary, but in reality there are transition regions that smoothly blend regions of darkness and light (see Section 5).

One severe limitation of the Airy theory for the optical rainbow is that the amplitude distribution along the initial wavefront is unknown: based on certain assumptions it has to be guessed. There is a natural and fundamental parameter, the *size parameter*, β, which is useful in determining the domain of validity of the Airy approximation; it is defined as the ratio of the droplet circumference to the wavelength of light (λ). In terms of the *wavenumber* k this is

$$\beta = \frac{2\pi a}{\lambda} = ka\ ,$$

a being the droplet radius. Typically, for sizes ranging from fog droplets to large raindrops, β ranges from about 100 to several thousand. Airy's approximation is a good one only for $\beta \gtrsim 5000$. In the light of these remarks it is perhaps surprising that an exact solution *does* exist for the rainbow problem, as it is commonly called. The more precisely expressed version is the 'scattering of plane electromagnetic waves by a homogeneous transparent sphere', which is in turn related to such scattering by an impenetrable sphere. This latter problem has historical significance in connection with the diffraction of radio waves around the earth, amongst other applications. The same type of formulation arises in the scattering of sound waves by a sphere, studied by Lord Rayleigh and others in the 19th century. The solution to this problem is expressible in terms of an infinite set of partial waves called, not surprisingly, a partial wave expansion. The corresponding result for electromagnetic wave scattering was developed in 1908 by Mie [24]. It is obviously of interest to determine under what conditions such an infinite set of terms (each of which is a somewhat complicated expression) can be truncated, and what the resulting error may be by so doing. However, in confirmation of the conjecture that there is no such thing as a free lunch, it transpires that the number of terms that must be retained is of the same order of magnitude as the size parameter, i.e. up to several thousand

for the rainbow problem! The "why is the sky blue?" scattering problem, on the other hand—Rayleigh scattering—requires only *one* term because the scatterers are molecules which are much smaller than a wavelength, so the simplest truncation—retaining only the first term—is perfectly adequate. Although in principle the rainbow problem can be "solved" with enough computer time and resources, numerical solutions by themselves (as Nussenzveig points out) offer little or no insight into the physics of the phenomenon.

Fortunately help was at hand. The now-famous Watson transformation, developed by Watson [25] is a method for transforming the slowly converging partial-wave series into a rapidly convergent expression involving an integral in the complex angular-momentum plane (see Appendix C). But why *angular momentum*? Although they possess zero rest mass, photons have energy $E = hc/\lambda$ and momentum $E/c = h/\lambda$ where h is Planck's constant and c is the speed of light in vacuo. Thus, for a non-zero impact parameter b_i, a photon will carry an angular momentum $b_i h/\lambda$, (which can in fact assume only certain discrete values). Each of these discrete values can be identified with a term in the partial wave series expansion. Furthermore, as the photon undergoes repeated internal reflections, it can be thought of as orbiting the center of the raindrop. Why *complex* angular momentum? This allows some powerful mathematical techniques to "redistribute" the contributions to the partial wave series into a few points in the complex plane—specifically poles (called Regge poles in elementary particle physics) and saddle points. Such a decomposition means that instead of identifying angular momentum with certain discrete real numbers, it is now permitted to move continuously through complex values. However, despite this modification, the poles and saddle points have profound physical interpretations in the rainbow problem; (this is known in other contexts: in many wave propagation problems poles and branch points of Green's functions in the complex plane can be identified with trapped and freely propagating waves, respectively, or in terms of discrete and continuous spectra of certain Sturm–Liouville operators (see e.g. [26–28] and references therein). A thoughtful and somewhat philosophical response to the question '*why complex angular momentum*?' appears in a quotation at the very end of this review (Section 8).

Mathematical details will be provided later in Section 5, but for now it will suffice to point out that since *real* saddle points correspond to ordinary rays of light, we might expect complex saddle points to correspond in some sense to *complex rays* (whatever that might mean). Again, in problems of wave propagation, imaginary components can often be identified with the damping of some quantity—usually the amplitude—in space or time. An elementary but perfect example of this can be found in the phenomenon of (critical) total internal reflection of light at a glass/air or water/air boundary: the "refracted" ray is in effect a surface wave, propagating along the interface with an amplitude which diminishes exponentially away from that interface. Such waves are sometimes identified as "evanescent", and are common in many other non-quantum contexts [26]. Typically, the intensity is negligible on the order of a wavelength away from the interface; this defines a type of "skin depth". Quantum-mechanical tunneling has a similar mathematical description, and, directly related to the rainbow problem, complex rays can appear on the shadow side of a caustic, where they represent the damped amplitude of diffracted light.

In the light of this (no pun intended) it is interesting to note that there is another type of surface wave: this one is associated with the *Regge pole* contributions to the partial-wave expansion. Such waves are initiated when the impact parameter is the drop diameter, i.e.

the incident ray is tangential to the sphere. These waves are damped in the direction of increasing arc-length along the surface, i.e. tangentially, (as opposed to radially, in the case of an evanescent wave) because at *each* point along its circumferential path the wave penetrates the spherical drop at the critical angle for total internal reflection, in a reversal of the normal event. Subsequently it will re-emerge as a surface wave after taking one or more of these shortcuts. In [2] Nussenzveig points out that this "pinwheel" path for light rays was considered by Kepler in 1584 (see [11,29]) as a possible explanation for the rainbow, but he did not pursue it because he could not reproduce the correct rainbow angle.

It is important to note that the Watson transformation was originally introduced in connection with the diffraction of radio waves around the earth [25], alluded to above. In 1937, Van der Pol and Bremmer [30,31] applied it to the rainbow problem, but, notwithstanding a claim by Sommerfeld that the problem had now been brought to a beautiful conclusion [32], they were able only to establish that the Airy approximation could be recovered as a limiting case. (Their papers contain some very beautiful and detailed mathematics, however, and are significant contributions nonetheless.) In some profound and highly technical papers Nussenzveig subsequently developed a modification of the transformation [33,5,6] (a valuable summary of his work up to that point may be found in [34]; references to more recent work will be made throughout the review) and was able to bring the problem to the desired conclusion using some extremely sophisticated mathematical techniques without losing sight of the physical implications.

In the simplest Cartesian terms, on the illuminated side of the rainbow (in a limiting sense) there are two rays of light emerging in parallel directions: at the rainbow angle they coalesce into the ray of minimum deflection, and on the shadow side, according to geometrical optics, they vanish (this is actually a good definition of a caustic curve or surface; refer again to Fig. 5). From the above general discussion of real and complex rays it should not be surprising that, in the complex angular momentum plane, a rainbow in mathematical terms is *the collision of two real saddle points*. But this is not all: this collision does not result in the mutual annihilation of these saddle points; instead a single complex saddle point is born, corresponding to a complex ray on the shadow side of the caustic curve. This is directly associated with the diffracted light in Alexander's dark band.

1.3. The rainbow: elementary mathematical considerations

As already noted, the primary rainbow is seen in the direction corresponding to the most dense clustering of rays leaving the drop after a single internal reflection. Correspondingly higher-order rainbows exist (in principle) at similar "ray clusters" for k internal reflections, $k = 2, 3, 4, \ldots$. Although only the primary and secondary rainbow are seen in nature, Walker [14] has determined the scattering angles, widths, color sequences and scattered light intensities of the first 20 rainbows using geometric optics (he also examined the associated interference patterns, discussed in Sections 2 and 6 below). Referring to Figs. 1(a) and (b), consider the more general angular deviation D_k between the incident and emergent rays after k internal reflections. Since each such reflection causes a deviation of $\pi - 2r$, ($r = r(i)$ being the angle of refraction, itself a function of i, the angle of incidence) and two refractions always occur, the

total deviation is

$$D_k(i) = k(\pi - 2r) + 2(i - r)\ . \tag{1.1}$$

Thus, for the primary rainbow ($k = 1$), $D_1(i) = \pi - 4r + 2i$, and for the secondary rainbow ($k = 2$), $D_2 = 2i - 6r$ (modulo 2π). Seeking an extremum of $D_k(i)$, and using Snell's law of refraction twice (once in differentiated form) it is readily shown that for all k, $D_k(i)$ is an extremum when

$$\frac{\cos i}{\cos r} = \frac{N}{k+1} \tag{1.2}$$

or after some rearrangement,

$$\cos i = \left\{\frac{N^2 - 1}{k(k+2)}\right\}^{1/2}\ , \tag{1.3}$$

where N is the index of refraction for the incident ray (assumed monochromatic for the present). For this value of i, i_m say, $D_k = D_k^m$, so that for $k = 1$ in particular

$$D_1^m = \pi + 2i_m - 4\arcsin\left(\frac{\sin i_m}{N}\right)\ . \tag{1.4}$$

That this extremum is indeed a minimum for $k = 1$ follows from the fact that

$$D_k''(i) = \frac{2(k+1)(N^2 - 1)\tan r}{N^3 \cos^2 r} \tag{1.5}$$

(see [18]) which is positive since $N > 1$ and $0 < r < \pi/2$, except for the special case of normal incidence, so the clustering of deviated rays corresponds to a *minimum* deflection. Note also that the generalization of Eq. (1.4) above to k internal reflections is

$$D_k^m(i_m) = k\pi + 2i_m - 2(k+1)\arcsin\left(\frac{\sin i_m}{N}\right)\ . \tag{1.6}$$

Returning to the primary rainbow for which $k = 1$, since the minimum angle of deviation depends on N, we now incorporate dispersion into the model. For red light of wavelength 6563 Å in water [8], $N \approx 1.3318$, and for violet light of wavelength 4047 Å, $N \approx 1.3435$. These correspond to $D_1^m \approx 137.7^\circ$ (i.e. a rainbow angle—the supplement of D_1^m, of $\approx 42.3^\circ$—see below) and $D_1^m \approx 139.4^\circ$ (a rainbow angle of $\approx 40.6^\circ$), respectively. This gives a theoretical angular width of $\approx 1.7^\circ$ for the primary rainbow; in reality it is about 0.5° wider than this [2] because this is the angular diameter of the sun for an earth-based observer, and hence the incident rays can deviate from parallellism by this amount (and more to the point, so can the emergent rays as a simple differential argument shows). The above-mentioned *rainbow angle* for a given D_k^m (or $D_{\min}$) is the supplementary angle $180^\circ - D_k^m$; this is the angle of elevation relative the sun-observer line and for the primary rainbow is about 42°. Light deviated at angles larger than this minimum will illuminate the sky inside the rainbow more intensely than outside it, for this very reason. For this reason, the outside "edge" of the primary bow will generally be more sharply defined than the inner edge. For the secondary rainbow ($k = 2$) $D_2^m = D_{\min} \approx 231^\circ$

and the rainbow angle in this case is $D_2^m - 180^\circ$, or approximately 51° for light of an orange color for which we choose $n = \frac{4}{3}$ (sometimes, as in [8] the angle of minimum deviation is defined as the complement of $D_{\min}$, which for $k = 2$ is $\approx 129^\circ$). Each additional reflection of course is accompanied by a loss of light intensity because of transmission out of the drop at that point, so on these grounds alone it would be expected that the tertiary rainbow ($k = 3$) would be difficult to observe. In this case $D_{\min} \approx 319^\circ$, and the light concentration therefore is at about an angle of 41° from the incident light direction. This means that it appears *behind* the observer as a ring around the sun. Due to (i) the increased intensity of sunlight in this region, (ii) the fact that the angles of incidence i_m increase with k (see (1.3)) and result in a reduction of incident intensity per unit area of the surface, and hence also a reduction for the emergent beam, (iii) higher-order rainbows are wider than orders one and two, (iv) the presence of light reflected from the outer surface of the raindrops (direct glare), (v) light emerging with no internal reflections (transmitted glare), and (vi) the reduced intensity from three reflections, it is not surprising that such rainbows (i.e. $k \geqslant 3$) have not been reliably reported in the literature (but see [13]).

At this point it is useful to note the alternative valuable and very succinct treatment of the geometrical theory of the rainbow provided by Jackson [35]. Instead of using the angles of incidence i and refraction r, the author works in terms of the impact parameter b (normalized by the drop radius a so the fundamental variable is $x = \sin i = b/a$; see also the appendix in [16]). For the primary rainbow in particular ($k = 1$) Eq. (1.1) becomes

$$\theta = \pi + 2\arcsin x - 4\arcsin(x/N)\,, \tag{1.1a}$$

where the deviation angle D_1 has been replaced by θ, which is the standard variable used in the study of scattering cross sections (see Appendix A and Section 1.5 below; it is useful to interpret the rainbow problem in terms of scattering theory even at this basic level of description). Noting that $\mathrm{d}\theta/\mathrm{d}i = \cos i\,\mathrm{d}\theta/\mathrm{d}x$ for given N, it follows that extrema of $\theta(i)$ can be expressed in terms of extrema of $\theta(x)$; hence

$$\theta'(x) = \frac{2}{\sqrt{1-x^2}} - \frac{4}{\sqrt{N^2-x^2}} \tag{1.1b}$$

and

$$\theta''(x) = \frac{2x}{(1-x^2)^{3/2}} - \frac{4x}{(N^2-x^2)^{3/2}}\,. \tag{1.1c}$$

By requiring $\theta'(x_0) = 0$ it follows that $x_0 = \sqrt{(4-N^2)/3}$ from which result (1.3) is recovered for $k = 1$. Obviously, this result can be generalized to other positive integer values of k. Note also that

$$\theta''(x_0) = \frac{9(4-N^2)^{1/2}}{2(N^2-1)^{3/2}} > 0 \quad \text{for } 1 < N < 2\,,$$

so the angle of deviation is indeed a minimum as expected. We note two other aspects of the problem discussed in [35] (this article will be mentioned again in connection with Airy's theory of the rainbow in Section 2.1). First, it is clear that for $x \approx x_0$ the familiar quadratic "fold" for

$\theta(x)$ is obtained:

$$\theta \approx \theta_0 + \theta''(x_0)(x - x_0)^2/2 \,, \tag{1.1d}$$

where $\theta_0 = \theta(x_0)$. This result will be required in Section 2.1. The second comment concerns dispersion: since $\theta = \theta(x, N)$ it follows from (1.1a) that

$$\frac{\partial\theta}{\partial N} = -4\,\frac{\partial}{\partial N}\,[\arcsin(x/N)] = \frac{4x}{N\sqrt{N^2 - x^2}}$$

and so at the rainbow angle ($\theta_0 \approx 138^\circ$ corresponding to $x_0 = \sqrt{(4 - N^2)/3} \approx 0.86$ for $N = \frac{4}{3}$)

$$\frac{\partial\theta}{\partial N} = \frac{2}{N}\sqrt{\frac{4 - N^2}{N^2 - 1}} \,. \tag{1.1e}$$

This can be used to estimate the angular spread of the rainbow ($\Delta\theta$) given the variation in N over the visible part of the spectrum (ΔN); as noted above $\Delta\theta$ is slightly less than 2° for the primary bow. Note also that since

$$\frac{\mathrm{d}\theta}{\mathrm{d}\lambda} \approx \frac{\mathrm{d}\theta}{\mathrm{d}N}\,\frac{\mathrm{d}N}{\mathrm{d}\lambda}$$

for given x_0 (i.e. neglecting the small variation of x_0 with wavelength λ) and $\mathrm{d}N/\mathrm{d}\lambda < 0$, it follows that $\mathrm{d}\theta/\mathrm{d}\lambda < 0$, and so the red part of the (primary) rainbow emerges at a smaller angle than the violet part, so the latter appears on the underside of the arc with the red outermost.

Thus far we have examined only the variation in the deviation of the incident ray as a function of the angle of incidence (or as a function of the normalized impact parameter). Even within the limitations of geometrical optics, there are two other factors to consider. The first of these is the behavior of the coefficients of reflection and refraction as a function of angle of incidence, and the second is to determine how much light energy, as a function of angle of incidence, is deviated into a given solid angle after interacting with the raindrop. This might be thought of as a classical "scattering cross section" type of problem. The first of these factors involves the Fresnel equations, and though we shall not derive them here an excellent account of this can be found in the book by Born and Wolf [23] (see also [22,36]: another classic text).

1.4. Polarization of the rainbow

Electromagnetic radiation—specifically light—is propagated as a transverse wave, and the orientation of this oscillation (or "ray") can be expressed as a linear combination of two "basis vectors" for the "space", namely, mutually perpendicular components of two independent linear polarization states [2]. Sunlight is unpolarized (or randomly polarized), being an incoherent mixture of both states, but reflection can and does alter the state of polarization of an incident ray of light. For convenience we can consider the two polarization states of the light incident upon the back surface of the drop (i.e. from within) as being, respectively, parallel to and perpendicular to the plane containing the incident and reflected rays. Above a critical angle of incidence (determined by the refractive index) both components are totally reflected, although some of the light does travel around the surface as an "evanescent" wave; this shall be discussed

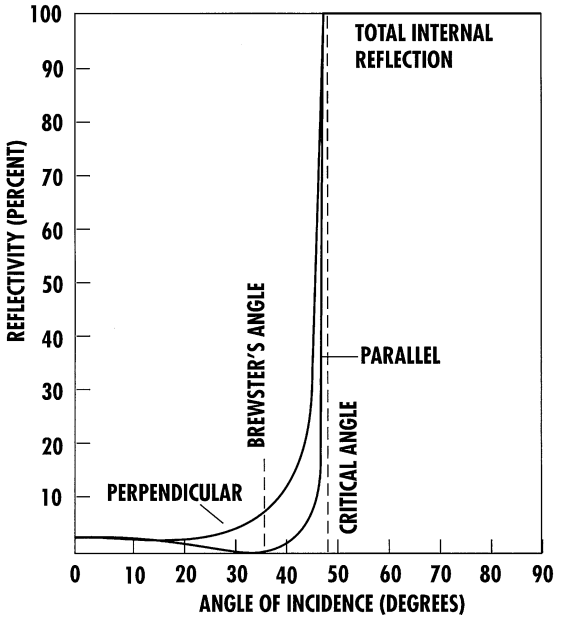

Fig. 6. The reflectivity of light impinging on a water/air interface from within the denser medium (i.e. from within the raindrop) as a function of the angle of incidence, resulting in the polarization of the rainbow. There is always some reflection of the component perpendicular to the plane of reflection (the plane containing the incident and reflected rays); the parallel component however, is completely transmitted at the Brewster angle. At the critical angle for total internal reflection, surface waves can be created, travelling along the surface of the drop for several degrees of arc (see Sections 3 and 5). (Redrawn from [2].)

in more detail in a later section. This critical angle of incidence, i_c, from *within* the drop is defined by the expression

$$i_c = \arcsin(N^{-1}) \,, \tag{1.7}$$

where N is the refractive index of the water droplet (more accurately, N is the *ratio* of the refractive indices of water and air, both relative to a vacuum). Taking a generic value for N of $\frac{4}{3}$ (orange light) we find $i_c \approx 48.6^\circ$. At angles i less than i_c the parallel component of polarization is reflected less efficiently than its perpendicular counterpart, and at the *Brewster angle* it is entirely transmitted, leaving only perpendicularly polarized light to be reflected (partially) from the inside surface of the drop. Since the Brewster angle, as shown below, is close to the angle of total internal reflection i_c, the light which goes on to produce the rainbow is strongly perpendicularly polarized. At the Brewster angle the reflected and refracted rays are orthogonal (and hence complementary), and therefore it follows from Snell's law that

$$i_B = \arctan(N^{-1}) \tag{1.8}$$

(remember this is from *within* the drop), from which $i_B \approx 36.9^\circ$, differing from i_c by about 12° (see Fig. 6 and also [2]). It is perhaps worth pointing out here that the discussion by Tricker [8] is for an *air-to-water* interface, so that N^{-1} in Eq. (1.8) above should be replaced by N; the corresponding value i_B above is the supplement of the above angle, i.e. $\approx 53.1^\circ$.

The coefficient of reflection of light depends on its degree of polarization: Consider first the case of light polarized perpendicular to the plane of incidence, with the amplitude of the incident light taken as unity. It follows from the Fresnel equations that the fraction of the incident intensity reflected is [23]

$$\frac{\sin^2(r-i)}{\sin^2(r+i)} \tag{1.9}$$

and

$$\frac{\tan^2(r-i)}{\tan^2(r+i)}\,, \tag{1.10}$$

if the light is polarized parallel to the plane of incidence (there is an apparent discrepancy between the terminology of Tricker [8] and Walker [14] at this point: in particular the diagrams in [8] for the Fresnel coefficients as a function of i for an air-to-water interface have the opposite sense of polarization to that in [22,23] and other texts). It follows that for a ray entering a drop and undergoing k internal reflections (plus 2 refractions), the fraction of the original intensity remaining in the emergent ray is, for perpendicular polarization,

$$I_{k1} = \left[1 - \left(\frac{\sin^2(r-i)}{\sin^2(r+i)}\right)\right]^2 \left(\frac{\sin^2(r-i)}{\sin^2(r+i)}\right)^{2k} \tag{1.11}$$

and for parallel polarization

$$I_{k2} = \left[1 - \left(\frac{\tan^2(r-i)}{\tan^2(r+i)}\right)\right]^2 \left(\frac{\tan^2(r-i)}{\tan^2(r+i)}\right)^{2k}\,. \tag{1.12}$$

The total fraction is the sum of these two intensities. For angles of incidence close to normal, for which $i \approx Nr$, it follows that the single-reflection intensity coefficients for reflection and refraction become, from Eq. (1.9)

$$\left(\frac{N-1}{N+1}\right)^2 \tag{1.13}$$

and

$$\frac{4N}{(N+1)^2}\,, \tag{1.14}$$

respectively. For the choice of $N = \frac{4}{3}$ under these circumstances it follows that the reflection and refraction coefficients are approximately 0.02 and 0.98, respectively.

An alternative but entirely equivalent formulation of the Fresnel equations is applied to the polarization of the (primary) rainbow in [35], using the relative amplitude equations for polarization both perpendicular to ($\mathbf{E}_\perp$) and parallel to the plane of incidence ($\mathbf{E}_{||}$) [36]. For each polarization, the emerging relative amplitude (i.e. scattered : incident) is determined by the product of the relative amplitudes corresponding to the three interfaces, namely transmission at an air/water interface, reflection at a water/air interface and transmission at a water/air interface (for rays 1, 2, 3 and 3′, respectively, in Fig. 1(c)). Care must be taken to identify the correct refractive indices and angles of incidence, both of which depend on which medium the ray is

about to enter. The calculations (omitted in [35]) are straightforward but nonetheless worthy of note, and are summarized below. There are three interfaces to consider for the primary rainbow analysis: *air/water transmission* (with transmission coefficients $t_s^{(1)}$ and $t_p^{(1)}$ for the vector **E** perpendicular and parallel to the plane of incidence, respectively); *water/water reflection* (with corresponding reflection coefficients $r_s^{(2)}, r_p^{(2)}$), and *water/air transmission* (with transmission coefficients $t_s^{(3)}$, $t_p^{(3)}$). Each will be considered in turn; in what follows n denotes the refractive index of the 'incidence' medium, n' that of the 'transmission' medium, and as usual, i and r represent the angles of incidence and refraction. N is the refractive index of water; that of air will be taken as unity. The formulae for the various coefficients are taken from [36].

1.4.1. Air/water transmission

$n=1$, $n'=N=\frac{4}{3}$; at the (primary) rainbow angle (according to geometrical optics)

$$\cos i = \sqrt{\frac{N^2-1}{3}}$$

from which the remaining quantities follow, so that

$$t_s^{(1)} = \frac{2\cos i}{\cos i + \sqrt{N^2 - \sin^2 i}} = \frac{2}{3}$$

and

$$t_p^{(1)} = \frac{2N\cos i}{N^2\cos i + \sqrt{N^2 - \sin^2 i}} = \frac{2N}{2+N^2}\ .$$

1.4.2. Water/water reflection

$n=N=\frac{4}{3}$, $n'=1$; in this case

$$\sin r = \frac{1}{N}\sqrt{\frac{4-N^2}{3}}\ ,$$

so that

$$r_s^{(2)} = \frac{N\cos r - \sqrt{1-N^2\sin^2 r}}{N\cos r + \sqrt{1-N^2\sin^2 r}} = \frac{1}{3}$$

and

$$r_p^{(2)} = \frac{\cos r - N\sqrt{1-N^2\sin^2 r}}{\cos r + N\sqrt{1-N^2\sin^2 r}} = \frac{2-N^2}{2+N^2}\ .$$

1.4.3. Water/air transmission

Again, $n=N=\frac{4}{3}$, $n'=1$; and in this final situation

$$t_s^{(3)} = \frac{2N\cos r}{N\cos i + \sqrt{1-N^2\sin^2 r}} = \frac{4}{3}$$

and

$$t_p^{(3)} = \frac{2N\cos r}{\cos i + N\sqrt{1-N^2\sin^2 r}} = \frac{4N}{2+N^2}\ .$$

These coefficients are combined according to respective products

$$\text{(i)}\ t_s^{(1)} r_s^{(2)} t_s^{(3)} = \tfrac{2}{3}\,\tfrac{1}{3}\,\tfrac{4}{3} = \tfrac{8}{27}$$

and

$$t_p^{(1)} r_p^{(2)} t_p^{(3)} = \frac{2N}{2+N^2}\,\frac{2-N^2}{2+N^2}\,\frac{4N}{2+N^2} = 2\left(\frac{2-N^2}{2+N^2}\right)\left(\frac{2N}{2+N^2}\right)^2$$

so that the required relative amplitudes are

$$(1)\quad \frac{E_{\text{scatt}}}{E_{\text{inc}}} = \tfrac{8}{27}\ (\text{for } \mathbf{E}_\perp)\quad \text{and (1.15)}$$

$$\frac{E_{\text{scatt}}}{E_{\text{inc}}} = 2\left(\frac{2N}{N^2+2}\right)^2\left(\frac{2-N^2}{2+N^2}\right)\quad (\text{for } \mathbf{E}_{||})\ .$$

It follows that the *relative intensity* of the perpendicular polarization (for $N = \frac{4}{3}$) is 8.78% (i.e. $I_\perp \approx 0.0878 I_0$) and for the parallel polarization it is 0.34% (i.e. $I_{||} \approx 0.0034 I_0$; note therefore, that $I_{||} \approx 0.039 I_\perp$, so the primary bow is about 96% perpendicularly polarized).

Notice from Eq. (1.10) that the fraction of the intensity of reflected light is *zero* (for parallel polarization) when the denominator in the coefficient vanishes, i.e. when $i+r = \pi/2$; at this point the energy is entirely transmitted, so the internally reflected ray is completely perpendicularly polarized. From this it follows that

$$\sin i = N \sin r = N \cos i\ ,$$

whence

$$\tan i = \tan i_B = N$$

for *external* reflection, and Eq. (1.8) for *internal* reflection. In the latter case, for $N = \frac{4}{3}$, this yields $i_B \approx 36.9^\circ$, the Brewster polarizing angle, as pointed out above. Note that as $i+r$ "passes through" $\pi/2$, the tangent changes sign (obviously being undefined at $\pi/2$); this corresponds to a phase change of 180° which will be noted when the so-called "glory" is discussed later. Notwithstanding the differences in terminology, it is also shown in [8], using Eqs. (1.11) and (1.12) that the primary rainbow is about 96% polarized (as shown above using Eqs. (1.15)), and the secondary bow somewhat less so: approximately 90%.

1.5. The physical basis for the divergence problem

In this subsection the physics of the *differential scattering cross section* is discussed (see the definition of this in Appendix A). Referring to Fig. 7, note that for light incident upon the raindrop in the incidence interval $(i, i+\delta i)$ the area "seen" by the incident rays is

$$(2\pi a \sin i)(a \cos i)\delta i$$

or

$$\pi a^2 \delta i \sin 2i\ .$$

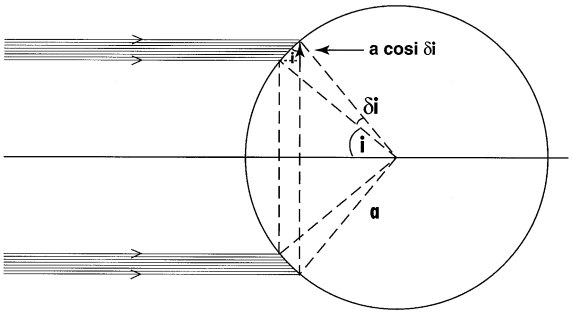

Fig. 7. The geometry leading to the derivation of the geometrical optics intensity function (1.15), which predicts infinite intensity at the rainbow angle. (Redrawn from [8].)

These rays are then scattered into an interval $(\theta, \theta + \delta\theta)$ say, measured from the direction of incidence, and this results in solid angle of

$$(2\pi \sin\theta)\delta\theta$$

being occupied by these rays. Thus the energy (relative to incoming energy) entering unit solid angle per unit time after internal reflection (and neglecting losses from refraction and reflection) is

$$\frac{a^2 \sin 2i}{2 \sin\theta} \left(\frac{\delta i}{\delta\theta}\right) . \tag{1.15}$$

Passing to the limit, the Cartesian condition for the occurrence of the rainbow is of course that $\mathrm{d}\theta/\mathrm{d}i = 0$, but now we are in possession of a geometric factor which modulates non-zero values of this derivative. This factor is undefined when $\theta = 0$ or π (these cases occurring when $i = 0$), but for the case of a single internal reflection with both i and θ "small", using $\theta = \pi - D = 4r - 2i$, we have

$$\lim_{i \to 0^+} \frac{\sin 2i}{\sin\theta} = \frac{2i}{\theta} = \frac{2i}{4r - 2i} = \frac{N}{2 - N} . \tag{1.16}$$

Since this ratio is just 2 when $N = \frac{4}{3}$ it is clear that the intensity of the "backscattered" light is not particularly large when $\theta = 0$. This excludes, of course, any diffraction effects. Note that for large angles of incidence (i.e. $i \approx 90^\circ$), the angle θ (as defined in Fig. 7) never returns to zero, being given instead by the expression

$$\theta = 4 \arcsin(N^{-1}) - 180^\circ , \tag{1.17}$$

which for $N = \frac{4}{3}$ is approximately 14.4° according to geometrical optics (more will be said about such grazing incidence rays when the glory is discussed in Section 3). Of course, it has already been noted that at the rainbow angle (corresponding to $\theta \approx 42^\circ$) the intensity is predicted to become infinite according to the same theory!

2. Theoretical foundations

2.1. The supernumerary rainbows; a heuristic account of Airy theory

These bows are essentially the result of interference between rays emerging from the raindrop close to the rainbow angle (i.e. angle of minimum deviation). In general they will have entered at different angles of incidence and traversed different paths in the denser medium; there is of course a reduction in wavelength inside the drop, but the overall effect of different path lengths is the usual "diffraction" pattern arising as a result of the destructive/constructive interference between the rays. The spacing between the maxima (or minima) depends on the wavelength of the light and the diameter of the drop—the smaller the drop the greater the spacing (see Fig. 4). Indeed, if the drops are less than about 0.2 mm in diameter, the first maximum will be distinctly below the primary bow, and several other such maxima may be distinguished if conditions are conducive.

Although this phenomenon is decidedly a "wavelike" one, we can gain some heuristic insight into this mechanism by considering, following Tricker [8], a geometrical optics approach to the relevant rays and their associated *wavefronts*. This is done by considering, not the angle of minimum deviation, but the point of *emergence* of a ray from the raindrop as the angle of incidence is increased. A careful examination of Fig. 1(a) reveals that as the angle of incidence is increased away from zero, while the point of entry moves clockwise around the drop, the point of emergence moves first in a counterclockwise direction, reaches an extreme position, and then moves back in a clockwise direction. This extreme point has significant implications for the shape of the wavefront as rays exit the drop in the neighborhood of this point. Referring to this figure, we wish to find when angle $BOA = 4r - i$ is a maximum. This occurs when

$$\frac{\mathrm{d}r}{\mathrm{d}i} = \frac{1}{4}$$

and

$$\frac{\mathrm{d}^2 r}{\mathrm{d}i^2} < 0 \,.$$

The first derivative condition leads to

$$\frac{\cos^2 i}{N^2 - \sin^2 i} = 16 \tag{2.1}$$

or

$$i = \arccos\left(\frac{N^2 - 1}{15}\right)^{1/2} , \tag{2.2}$$

whence for $N = \frac{4}{3}$, $i \approx 76.8^\circ$, well past the angle of incidence corresponding to the rainbow angle, which is $i \approx 59.4^\circ$ for the same value of N. As can be seen from Fig. 8(a), if YEY' represents the ray which emerges at minimum deviation, rays to either side of this are deviated through larger angles as shown. By considering the corresponding wavefront WEW' (distorted from the wavefront XOX' (see Fig. 8(b)) corresponding to a parallel beam of rays) it is clear that we are dealing with, locally at least, a cubic approximation to the wavefront in the neighborhood

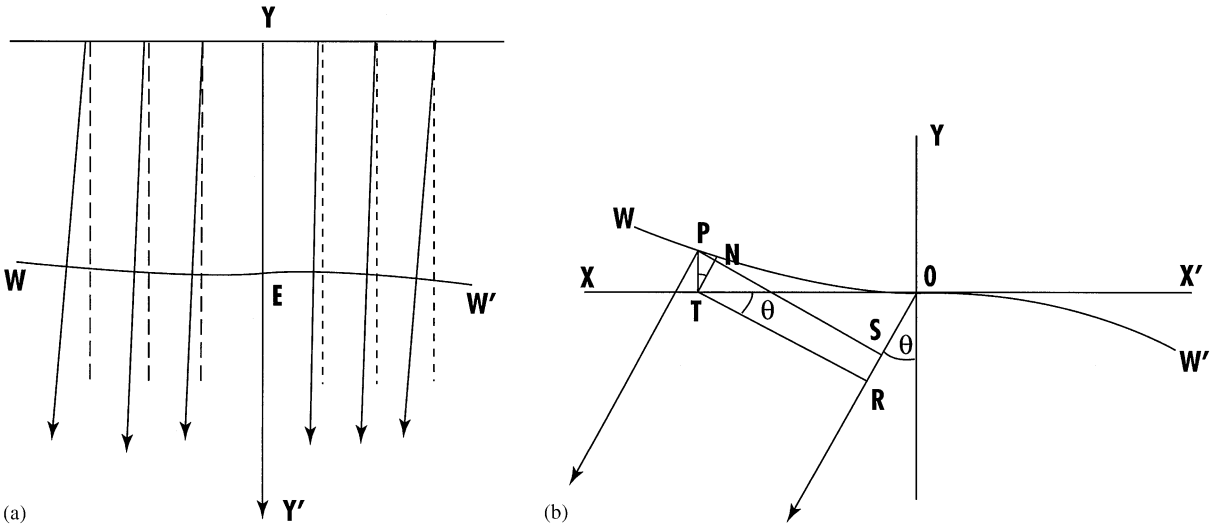

Fig. 8. (a) The ray of minimum deviation YY' and nearby rays emerging from the (locally) cubic wavefront WW', (b) the geometry of the wavefront WW' leading to the Airy integral form for the amplitude (2.8). (Redrawn from [8].)

of the point E, which is itself a point of inflection. With point E as the origin of coordinates and YY', XX' as the coordinate axes, it is clear that, as drawn, the equation of the wavefront WEW' can be written as

$$y = c'x^3 ,$$

since $y'(0) = 0$, where c' is a constant with dimensions of $(\text{length})^{-2}$. It is reasonable to expect that the linear dimensions of the wavefront will be related to the size of the raindrop from whence it comes, so in terms of the drop radius a we write

$$y = \frac{c}{a^2}x^3 , \tag{2.3}$$

c now being dimensionless. At this point, we are in a position to deduce the form of the famous "rainbow integral" introduced by Sir George Airy in 1838 [21]. Using Fig. 8(b) we can derive an expression for the amplitude of the wave in a direction making an angle θ with that of minimum deviation. By considering the path difference η between the points $P(x, y)$ and O we have

$$\eta = OS = OR - RS = OR - TN = x\sin\theta - y\cos\theta$$

or

$$\eta = x\sin\theta - \frac{cx^3}{a^2}\cos\theta . \tag{2.4}$$

In relative terms, if the amplitude of a small element δx of the wavefront at O is represented by $\sin\varpi t$, then that from a similar element at P is $\sin(\varpi t + \delta)$, where

$$\delta = \frac{2\pi\eta}{\lambda} = k\eta ,$$

λ being the wavelength, and k being the wavenumber of the disturbance. For the whole wavefront the cumulative disturbance amplitude is therefore the integral

$$\mathscr{A} = \int_{-\infty}^{\infty} \sin(\varpi t + k\eta)\,\mathrm{d}x$$

or, since $\sin k\eta$ is an odd function,

$$\mathscr{A} = \sin \varpi t \int_{-\infty}^{\infty} \cos k\eta\,\mathrm{d}x\ . \tag{2.5}$$

In terms of the following changes of variable:

$$\xi^3 = \frac{2kcx^3\cos\theta}{\pi a^2} \tag{2.6}$$

and

$$m\xi = \frac{2kx\sin\theta}{\pi}\ , \tag{2.7}$$

the above integral may be written in the canonical form

$$\left(\frac{\pi a^2}{2kc\cos\theta}\right)^{1/3} \int_{-\infty}^{\infty} \cos\frac{\pi}{2}(m\xi - \xi^3)\,\mathrm{d}\xi\ . \tag{2.8}$$

This is Airy's rainbow integral, first published in his paper entitled "On the Intensity of Light in the neighbourhood of a Caustic" [21]. The significance of the parameter m can be noted by eliminating ξ from Eqs. (2.6) and (2.7) to obtain

$$m = \left(\frac{2ka}{\pi}\right)^{2/3} \left(\frac{\sin^3\theta}{\cos\theta}\right)^{1/3}\ , \tag{2.9}$$

which for sufficiently small values of θ (the angle of deviation from the rainbow ray) is proportional to θ^3.

A graph of the rainbow integral qualitatively resembles the diffraction pattern near the edge of the shadow of a straight edge, which has the following features: (i) low-intensity, rapidly decreasing illumination in regions that geometrical optics predicts should be totally in shadow, and (ii) in the illuminated region (as with diffraction), a series of fringes, which correspond to the supernumerary bows below the primary rainbow. The first maximum is the largest in amplitude, and corresponds to the primary bow; the remaining maxima decrease rather slowly in amplitude, the period of oscillation decreasing also. The underlying assumption in this approach is that diffraction arises from points on the wavefront in the neighborhood of the Déscartes ray (of minimum deviation); provided that the drop size is large compared to the wavelength of light this is reasonable, and is in fact valid for most rainbows. It would not be as useful an assumption for cloud or fog droplets which are considerably smaller than raindrops, but even then the drop diameter may be five or ten times the wavelength, so the Airy theory is still useful. Clearly, however, it has a limited domain of validity.

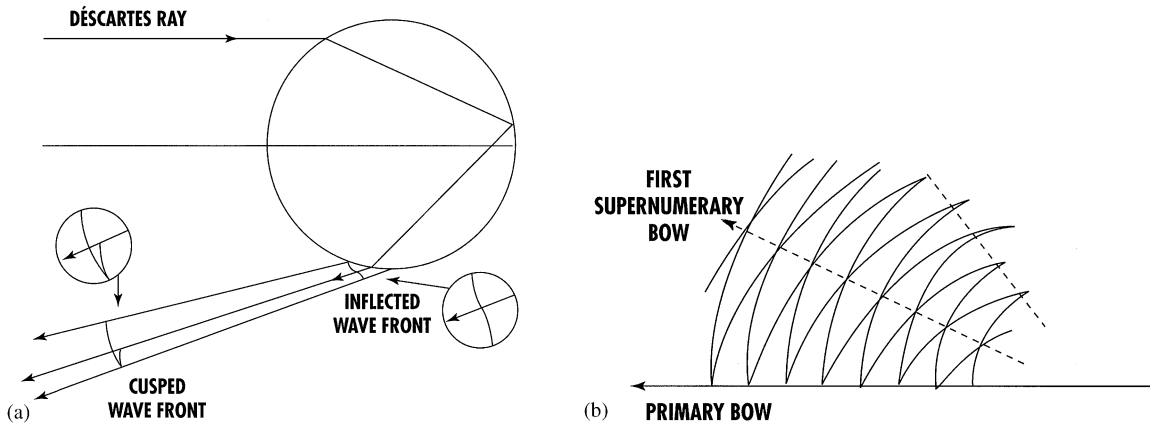

Fig. 9. (a) Details of the cusped wavefront upon its emergence from the raindrop. (b) Interference between the two arms of the cusped wavefronts (greatly exaggerated) can be expected to produce the supernumerary bows. (see Fig. 4 and [8], from which these figures are redrawn.)

Another feature of the rainbow problem that has been neglected thus far is the three-dimensional nature of the diffraction: the wavefront is a *surface* in three dimensions, not merely a curve in two. The factor in Eq. (2.8) above, namely

$$\left(\frac{\pi a^2}{2kc\cos\theta}\right)^{1/3} = \left(\frac{\lambda a^2}{4c\cos\theta}\right)^{1/3}, \tag{2.10}$$

requires modification. In fact, it is necessary to multiply this factor by $(a/\lambda)^{1/2}$ (see Appendix *I* in [8] for a discussion of this factor based on Fresnel zones). Since the angle θ will be small, this results in an amplitude proportional to $(a^7/\lambda)^{1/6}$ or equivalently, an intensity proportional to $(a^7/\lambda)^{1/3}$. Thus, the relatively strong dependence of intensity upon drop size is established within the Airy regime: other things being equal, larger drops give rise to more intense rainbows. Note also that the λ-dependence is important to determine the intensity distribution with wavelength.

The diffraction pattern may be thought of as arising from interference of two arms of a cusped wavefront [8] (see Figs. 9(a) and (b); but note also the discussion in [37]). A set of maxima and minima occur, lying between the direction of the incident light and the Déscartes ray. A change of phase occurs when light passes a focus (see Fig. 9(a)), so it is to be expected that the Déscartes ray would not correspond exactly to the direction of maximum intensity (which is displaced inwards), and this is indeed the case, as more complete theory shows. Furthermore, it is not the case that the intensities should be the same along both arms of the cusped wavefront, which implies that the minima will not in general be of zero intensity; in Airy's theory there is assumed to be no variation in intensity along the wave, and as a result the intensity minima are zero.

A more mathematical description has been provided in [35] which is both brief and valuable. Referring to Fig. 1(a), we write the phase ϕ accumulated along the critical ray between the surfaces $A'A''$ and $B'B''$ is, in terms of the relative impact parameter ($x = \sin i = b/a$) introduced in

Section 1.3,

$$\phi(x) = ka\{2(1-\cos i) + 4N\cos r\}$$
$$= 2ka\{1 - \sqrt{1-x^2} + 2\sqrt{N^2-x^2}\}\ .$$

The first term inside the first set of brackets is the sum of the distances from both A'' and B'' to the surface of the drop; the second term is N times the path length interior to the drop. Noting that

$$\phi' x(x) = 2ka\left\{\frac{x}{\sqrt{1-x^2}} - \frac{2x}{\sqrt{N^2-x^2}}\right\}$$

and comparing this with the result derived in Section 1 for the derivative with respect to x of the deflection angle $\theta(x)$ it is seen that

$$\phi'(x) = kax\theta'(x)\ .$$

Near the critical angle $\theta = \theta_0$ (and hence near $x = x_0$) this result may be written in terms of the variable $\xi = x - x_0$, i.e.

$$\phi'(\xi) = ka[x_0\theta'(\xi) + \xi\theta'(\xi)]\ ,$$

whence

$$\phi(\xi) = \phi_0 + ka\left[x_0(\theta-\theta_0) + \xi\theta(\xi) - \int_0^{\xi'}\theta(\xi')\,\mathrm{d}\xi'\right]\ .$$

Using the result, previously derived, that near $\theta = \theta_0$,

$$\theta(\xi) \approx \theta_0 + \theta''(\theta_0)\xi^2/2 + O(\xi^3)\ ,$$

it is readily shown that

$$\phi(\xi) \approx \phi_0 + ka[x_0(\theta-\theta_0) + \theta''(\theta_0)\xi^3/3 + O(\xi^4)\ ,$$

so that for two rays, each one close to, but on opposite sides of the critical ray, with equal and opposite ξ-values, it follows that their phase difference δ is

$$\delta = \phi(\xi) - \phi(-\xi) \approx 2ka\theta''(\theta_0)\xi^3/3\ .$$

When a phase difference is equal to an integer multiple of 2π then the rays interfere constructively in general; however, in this instance (and as noted above) an additional $\pi/2$ must be added because a focal line is passed in the process (see [7, chapters 12 and 13]). Hence for constructive interference

$$\theta_K - \theta_0 \approx \left(\frac{3\pi(K+\frac{1}{4})}{ka}\right)^{2/3}\left[\frac{\theta''(\theta_0)}{8}\right]^{1/3}, \quad K = 1,2,\ldots\ .$$

Clearly, within this approximation, $\theta_K - \theta_0 \propto (ka)^{-2/3}$ which is quite sensitive to droplet size. If the droplets are large enough, the supernumerary bows lie inside the primary bow and thus

are not visible; Jackson demonstrates that the maximum uniform droplet size rendering them visible is $a \simeq 0.28$ mm [35]. If the droplets are not uniform in size, the maxima may be washed out by virtue of the spread in sizes; if the droplets are all very small ($a < 50\ \mu$m) the various colors are dispersed little and "whitebows" or "cloudbows" result.

To capitalize on the expressions for the phase function $\phi(x)$ and $\theta - \theta_0$, refer now in figure (*) to the wavefront along BB'; following [35], it can be shown using the Kirchoff integral for diffraction that the amplitude of the scattered wave near $\theta = \theta_0$ is

$$\psi_{\text{scatt}} \propto \int_{-\infty}^{\infty} e^{ika[(\theta-\theta_0)\xi - \theta''\xi^3/6]}\, d\xi\ .$$

This can be expressed in the form of an Airy integral $\text{Ai}(-\eta)$ (essentially equivalent to Eq. (2.8), which is the form given in [7]), where

$$\text{Ai}(-\eta) = \frac{1}{\pi}\int_0^{\infty} \cos\left(\frac{1}{3}\xi^3 - \eta\xi\right)\, d\xi\ ,$$

where

$$\eta = \left(\frac{2k^2a^2}{\theta''(\theta_0)}\right)^{1/3}(\theta - \theta_0)\ .$$

For more details of the Airy function and the relationship between the two forms, see Appendix B. For large values of $\eta > 0$ the dominant term in the asymptotic expansion for $\text{Ai}(-\eta)$ is

$$\text{Ai}(-\eta) \simeq (\pi^2\eta)^{-1/4}\sin\left(\frac{2}{3}\eta^{3/2} + \frac{\pi}{4}\right)$$

corresponding to slowly decaying oscillations on the "bright" side of the primary bow (for $\eta < 0$, $\text{Ai}(-\eta)$ decreases to zero faster than exponentially; this is the "shadow" side of the primary bow). Noting that [35]

$$\langle|\text{Ai}(-\eta)|^2\rangle = (2\pi\sqrt{\eta})^{-1}\ ,$$

where $\langle . \rangle$ means the average value of its argument, it may be verified that

$$\langle|\text{Ai}(-\eta)|^2\rangle = \frac{1}{2\pi}\left(\frac{\theta''^2(\theta_0)}{4ka}\right)^{1/3}\sqrt{\frac{2}{\theta''(\theta_0)(\theta-\theta_0)}}\ .$$

Near $\theta = \theta_0$, the classical differential scattering cross section has been found to be

$$\frac{d\sigma}{d\Omega}_{\text{class}} \approx \left(\frac{a^2x_0}{\sin\theta_0}\right)\sqrt{\frac{2}{|\theta''(\theta_0)(\theta-\theta_0)|}}$$

(the factor a^2 appearing because x is dimensionless). By comparing this directly with the mean square Airy intensity, an approximate expression for the "Airy differential cross section" can be inferred, namely

$$\frac{d\sigma}{d\Omega}_{\text{Airy}} \approx \left(\frac{2\pi a^2x_0}{\sin\theta_0}\right)\left(\frac{4ka}{\theta''^2(\theta_0)}\right)^{1/3}|\text{Ai}(-\eta)|^2\ .$$

If the intensity loss from two refractions and one reflection are ignored (they cannot be) this cross section is about 30 times as great as that for an isotropic differential cross section ($d\sigma/d\Omega = a^2/4$) at the peak of the primary bow when $ka = 10^3$. This must be multiplied by the scattered/incident ratios calculated in Section 1.4 when these intensity losses are included. This renders the two cross sections comparable in magnitude [35].

The extension of the Airy theory to incorporate the superposition of light of all colors was first carried out by Pernter and Exner [38] in 1910, and later by Buchwald [39], who corrected some errors in the former work. The methods and results for the Airy–Pernter theory are well summarized by Tricker, and only a few specific results will be noted here. From the basic diffraction theory discussed so far, it is reasonable to expect supernumerary bows to exist *outside* the secondary bow (in the region of low-intensity illumination). Because of the width of the secondary—more than twice that of the primary—the displacements of the individual color interference patterns will be larger, and whiter, weaker supernumerary bows are to be expected, which, coupled with the lower intensity of the secondary bow means that such supernumeraries are rarely observed. Another point to be mentioned here involves polarization: changes in polarization can be neglected only if the plane of incidence of sunlight on the drop varies by a small angle, which in turn means that integral (2.8) should extend (in practice) only over those portions of the wavefront in the neighborhood of the Déscartes ray. This is clearly another limitation of the theory as originally formulated. Malkus et al. attempted to improve upon the Airy–Pernter theory by developing a better approximation to the shape of the wavefront (amongst other things), subsequently applying their theory to cloud (as opposed to rain) droplets, with mixed success [40].

2.2. Mie scattering theory

Mie theory is based on the solution of Maxwell's equations of electromagnetic theory for a monochromatic plane wave from infinity incident upon a homogeneous isotropic sphere of radius a. The surrounding medium is transparent (as the sphere may be), homogeneous and isotropic. The incident wave induces forced oscillations of both free and bound charges in synchrony with the applied field, and this induces a secondary electric and magnetic field, each of which has a components inside and outside the sphere [41] (see also [9,10]).

In what follows reference will be made to the intensity functions i_1, i_2, the Mie coefficients a_n, b_n and the angular functions π_n, τ_n. The former pair are proportional to the (magnitude)^2 of two incoherent, plane-polarized components scattered by a single particle; they are related to the scattering amplitudes S_1 and S_2 in the notation of Nussenzveig [10]. The function $i_1(\beta, N, \theta)$ is associated with the electric oscillations *perpendicular* to the plane of scattering (sometimes called horizontally polarized) and $i_2(\beta, N, \theta)$ is associated with the electric oscillations *parallel* to the plane of scattering (or vertically polarized). The scattered wave is composed of many partial waves, the amplitudes of which depend on $a_n(\beta, N)$ and $b_n(\beta, N)$. Physically, these may be interpreted as the nth electrical and magnetic partial waves, respectively. The intensity functions i_1, i_2 are represented in Mie theory as a spherical wave composed of two sets of partial waves: electric (a_n) and magnetic (b_n). The first set is that part of the solution for which the radial component of the magnetic vector in the incident wave is zero; in the second

set the corresponding radial component of the electric vector is zero. A given partial wave can be thought of as coming from an electric or a magnetic multipole field, the first wave coming from a dipole field, the second from a quadrupole, and so on [41]. The angular functions $\pi_n(\theta)$ and $\tau_n(\theta)$ are, as their name implies, independent of size (β) and refractive index (N). These are investigated in some detail by Penndorf [41] in connection with scattering in the forward, backward and side directions for $N = 1.33$.

For a point $\mathscr{P}$ located a distance r from the origin of coordinates, at polar angle θ and azimuthal angle ϕ the scattered intensities I_θ and I_ϕ are, respectively,

$$I_\theta = \left(\frac{i_2}{kr}\right)^2 \cos^2\phi \tag{2.11}$$

and

$$I_\phi = \left(\frac{i_1}{kr}\right)^2 \sin^2\phi\ , \tag{2.12}$$

where $i_j = |S_j|^2$, $j = 1, 2$ and the amplitude functions S_j are given by

$$S_1 = \sum_{n=1}^{\infty} \frac{2n+1}{n(n+1)}[a_n\pi_n(\cos\theta) + b_n\tau_n(\cos\theta)] \tag{2.13}$$

and

$$S_2 = \sum_{n=1}^{\infty} \frac{2n+1}{n(n+1)}[a_n\tau_n(\cos\theta) + b_n\pi_n(\cos\theta)]\ , \tag{2.14}$$

where n is the order of the induced electric or magnetic multipole. The Legendre functions $\pi_n(\cos\theta)$ and $\tau_n(\cos\theta)$ are defined in terms of the associated Legendre functions of the first kind, $P_n^1(\cos\theta)$ as

$$\pi_n(\cos\theta) = \frac{P_n^1(\cos\theta)}{\sin\theta} \tag{2.15}$$

and

$$\tau_n(\cos\theta) = \frac{\mathrm{d}}{\mathrm{d}\theta}P_n^1(\cos\theta)\ . \tag{2.16}$$

The scattering coefficients a_n and b_n are defined in terms of the so-called Ricatti–Bessel functions of the first and second kinds, respectively, these being

$$\psi_n(z) = \sqrt{\frac{\pi z}{2}}J_{n+1/2}(z) \quad \text{and} \quad \chi_n(z) = (-1)^n\sqrt{\frac{\pi z}{2}}J_{-(n+1/2)}(z)\ .$$

Also note that a Riccati–Hankel function is readily defined by

$$\xi_n(z) = \psi_n(z) + \mathrm{i}\chi_n(z)\ .$$

Then it follows that a_n and b_n can be written

$$a_n = \frac{\psi_n(\beta)\psi_n'(\alpha) - N\psi_n(\alpha)\psi_n'(\beta)}{\xi_n(\beta)\psi_n'(\alpha) - N\psi_n(\alpha)\xi_n'(\beta)} \tag{2.17}$$

and

$$b_n = \frac{\psi_n(\alpha)\psi_n'(\beta) - N\psi_n(\beta)\psi_n'(\alpha)}{\xi_n'(\beta)\psi_n(\alpha) - N\psi_n'(\alpha)\xi_n(\beta)}\,, \tag{2.18}$$

where the usual dimensionless size parameters are $\beta = ka$ and $\alpha = N\beta$. This form can be simplified by the introduction of phase angles; and this results in considerable simplification if the refractive index is real. Further details of the analysis may be found in the book by van de Hulst [7]; for an alternative but entirely equivalent formulation see [10]. Van de Hulst demonstrated that the Mie formulae lead, for large values of β, to a principle for localizing rays and separating diffracted, refracted and reflected light (in the sense of geometrical optics). This is called the *localization principle* and will be utilized in connection with the theories of the glory in Section 3 and the CAM approach to rainbows and glories in Section 5. The principle asserts that the term of order n in the partial wave expansion corresponds approximately to a ray of distance $(n + \frac{1}{2})/k$ from the center of the particle (this is just the impact parameter). When $\beta \gg 1$, the expansions for the S_j may be truncated at $n + \frac{1}{2} \simeq \beta$ (in practice, $n_{\max} \sim \beta + 4\beta^{1/3} + 2$; see [5,10,42]) and the remaining sum is separated into two parts: a diffracted light field component independent of the nature of the particle, and reflected and refracted rays dependent on the particle (see also [43–46]). In the former case both amplitude functions are equal to the quantity

$$\beta^2 \frac{J_1(\beta \sin\theta)}{(\beta \sin\theta)}\,, \tag{2.19}$$

a result we shall have occasion to refer to later.

3. Glories

3.1. The backward glory

Mountaineers and hill-climbers have noticed on occasion that when they stand with their backs to the low-lying sun and look into a thick mist below them, they may sometimes see a set of colored circular rings (or arcs thereof) surrounding the shadow of their heads. Although an individual may see the shadow of a companion, the observer will see the rings only around his or her head. This is the phenomenon of the glory, sometimes known at the anticorona, the brocken bow, or even the specter of the Brocken (it being frequently observed on this high peak in the Harz mountains of central Germany). A brief but very useful historical account (based on [38]), along with a summary of the theories, can be found in [47]. Early one morning in 1735, a small group of people were gathered on top of a mountain in the Peruvian Andes, members of a French scientific expedition, sent out to measure a degree of longitude, led by Bouguer

and La Condamine; a Spanish captain named Antonio de Ulloa also accompanied them. They saw an amazing sight that morning. According to Bouguer [38] this was

> "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... The closeness of the shadow allowed all its parts to be distinguished: arms, legs, the head. 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 gave a similar description and also drew a picture. In his account he said

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

During the 19th century, many such observations of the glory were made from the top of the Brocken mountain in central Germany, and it became known as the "Specter of the Brocken". It also became a favorite image among the Romantic writers; it was celebrated by Coleridge in his poem "Constancy to an Ideal Object". Other sightings were made from balloons, the glory appearing around the balloon's shadow on the clouds. Nowadays, while not noted as frequently as the rainbow, it may be seen most commonly from the air, with the glory surrounding the shadow of the airplane. Once an observer has seen the glory, if looked for, it is readily found on many subsequent flights (provided one is on the shadow side of the aircraft!). Some beautiful color photographs have appeared in the scientific literature [4,48,49].

The phenomenon in simplest terms is essentially the result of light backscattered by cloud droplets, the light undergoing some unusual transformations en route to the observer, transformations not predictable by standard geometrical optics (see the summary of contributing factors listed below). In a classic paper ([50]; later incorporated into his book [7]) van de Hulst put forward his theory of the glory, though at that time the term *anti-corona* was commonly used for this phenomenon. Noting that glories typically occur for cloud drop diameters of the order of 25 μm ($\beta \approx 150$), he used the asymptotic forms for the Legendre functions (spherical harmonics) π_n and τ_n above, namely

$$\pi_n(\cos\theta) = (-1)^{n-1}\tfrac{1}{2}n(n+1)\{J_0(z) + J_2(z)\} \tag{3.1}$$

and

$$\tau_n(\cos\theta) = (-1)^{n}\tfrac{1}{2}n(n+1)\{J_0(z) - J_2(z)\}\ , \tag{3.2}$$

where as before, n is the order of the induced electric or magnetic multipole, $z = (n+\frac{1}{2})\gamma$ and $\gamma = \pi - \theta$ is small (note that in [42] z is defined as $n\gamma$). Under these circumstances the amplitude functions $S_1(\theta)$ and $S_2(\theta)$ in the backward direction have the form

$$S_1(180^\circ) = -S_2(180^\circ) = \sum_{n=1}^{\infty}\frac{1}{2}(2n+1)(-1)^n(b_n - a_n)\ . \tag{3.3}$$

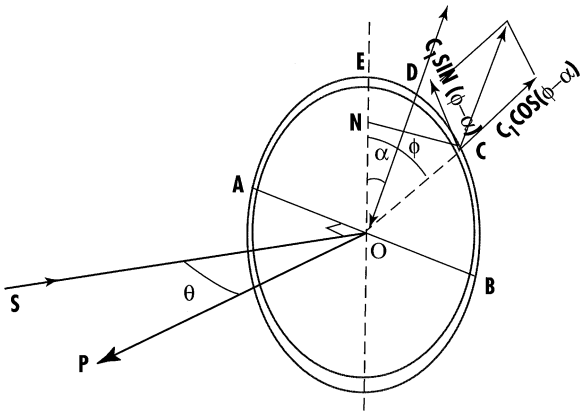

Fig. 10. An annular wavefront *ABCDE* emerging from a water droplet, giving rise to diffracted light at *P* (in particular). *SO* is the direction of the incident light, orthogonal to *AOB*. The geometry of this figure is used to determine the intensity of the glory in a given direction (see Eq. (3.8)). (Redrawn from [8].)

Debye (see details in [51]) noted that terms in series (2.13) and (2.14) are comparable up to $n=\beta$ and decrease rapidly thereafter, and for the moderately large values of β of interest here, most terms in the above sums will nearly cancel out each other; however, this will **not** be the case when near some value $\tilde{N}$ say, terms of the order

$$\ldots\tilde{N}-3,\ \tilde{N}-2,\ \tilde{N}-1,\ \tilde{N},\ \tilde{N}+1,\ \tilde{N}+2,\ \tilde{N}+3,\ldots$$

have nearly equal phases, resulting in a significant total amplitude. This is essentially a principle of "stationary phase" for sums, as opposed to integrals. Van de Hulst defines

$$c_1=\sum_n(2n+1)(-1)^n b_n \quad \text{and} \quad c_2=\sum_n(2n+1)(-1)^{n-1}a_n\ , \tag{3.4}$$

where the sums are to be taken only over those terms in the immediate "neighborhood" of $n=\tilde{N}$, further terms being neglected. For small γ values the total amplitude functions can be written subsequently as

$$S_1(180^\circ-\gamma)=\tfrac{1}{2}c_2\{J_0(u)+J_2(u)\}+\tfrac{1}{2}c_1\{J_0(u)-J_2(u)\} \tag{3.5}$$

with the c_1 and c_2 being interchanged for the amplitude $S_2(180^\circ-\gamma)$, where $u=N\gamma$.

The basic calculations for this phenomenon are based on the intensity of diffracted light arising from an annular (in fact toroidal) wavefront, the radius of which is of the same order of magnitude as the size of the drop [8]. By considering incident light as depicted in Fig. 10, and the disturbance arising from an element of the wavefront at point C in the figure, it can be shown that [8] the amplitude of the component polarized parallel to OE from this element is

$$a_1=c_1\sin(\phi-\alpha)\sin\phi+c_2\cos(\phi-\alpha)\cos\phi\ ,$$

where c_1 and c_2 are the amplitudes of the emergent polarizations of the parallel (to plane of incidence) and transverse electric fields, respectively. The path difference between points C and

E in the direction OP is $\delta = r\sin\phi\sin\theta$, and the phase difference is $2\pi\delta/\lambda \equiv x\sin\phi$. The total amplitude polarized parallel to OE from the whole wavefront is, after some reduction (and using antisymmetry of the cosine between the first and second quadrants)

$$A_1 = c_1\cos\alpha\int_0^{2\pi}\frac{1-\cos 2\phi}{2}\cos(x\sin\phi)\,\mathrm{d}\phi + c_2\cos\alpha\int_0^{2\pi}\frac{1+\cos 2\phi}{2}\cos(x\sin\phi)\,\mathrm{d}\phi\ ,$$

which to within a multiplicative factor is

$$A_1 = \cos\alpha(c_1\{J_0(x) - J_2(x)\} + c_2\{J_0(x) + J_2(x)\})\ . \tag{3.6}$$

Similarly, for the perpendicular polarization

$$A_2 = \sin\alpha(c_1\{J_0(x) + J_2(x)\} + c_2\{J_0(x) - J_2(x)\})\ . \tag{3.7}$$

For unpolarized light the direction of polarization changes rapidly and randomly, so little is lost by replacing the terms $\cos^2\alpha$ and $\sin^2\alpha$ in the intensity by their averages ($\frac{1}{2}$) to obtain

$$I = A_1^2 + A_2^2 = (c_1 + c_2)^2 J_0^2(x) + (c_1 - c_2)^2 J_2^2(x)\ . \tag{3.8}$$

The observational consequences of various choices of c_1 and c_2 (e.g. $c_1 = \pm c_2; c_1 = 0$) can be found in Section 13.33 of [7]. Tricker [8] makes the choice $c_2 = -0.204c_1$ based on a rainbow with 96% polarization in the plane of incidence ($c_2^2/c_1^2 = 4/96$), and incorporating the change in phase as the Brewster angle is exceeded (see also [50]). Fig. 10 is based on his calculations. Further descriptive (and also historical) details may be found in [52].

More can be said about the glory from a purely geometrical optics viewpoint. Recall that, for the heuristic description of the rainbow (Section 1.5) an important geometrical factor in the amplitude was shown to be

$$\left|\frac{\sin 2i}{\sin\theta}\frac{\mathrm{d}i}{\mathrm{d}\theta}\right|^{1/2}\ .$$

As noted already, rainbows correspond (at this level of description) to the condition $\mathrm{d}\theta/\mathrm{d}i = 0$, i.e. infinite intensity is predicted at the rainbow angle when interference is neglected. Such a divergence *also* occurs when $\sin\theta = 0$ while $\sin 2i \neq 0$, which occurs for $\theta = 0$ and π. The features associated with the forward direction are blended with the stronger coronae caused by the diffraction of light around the spherical drop [7], but the backward direction corresponds to the glory. Notice from Fig. 11 that the plane wavefront of the incident wave is emitted as a circular front, which has a *virtual focus* at F; when the figure is rotated about the axis of symmetry, the emitted wavefront is seen to be *toroidal*. Furthermore, this wavefront apparently emerges from the focal circle defined by Huygens' principle can be used to describe the interference pattern corresponding to this wavefront. Note that for a refractive index $N < \sqrt{2}$ there can be no such off-axis rays with less than four internal reflections contributing to the glory [7], but the attenuation for these rays is such that their contribution is negligible [53]. Thus according to ray theory, only the axial rays can contribute to this phenomenon.

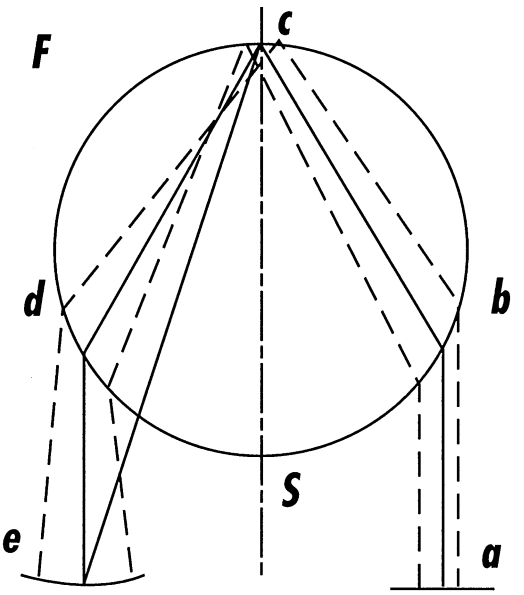

Fig. 11. Geometry of rays giving rise to the glory (neglecting surface-wave contribution). A plane wavefront is incident upon a spherical drop: several ray paths are drawn, one of which is the path *abcde*. It can be seen that rays adjacent to the central one correspond to a curved emerging wavefront, with a virtual focus at F. Because of cylindrical symmetry, the figure must be rotated about the axis cs; the outgoing rays (which mutually interfere) define a toroidal wavefront from the virtual circular source described by the point F. (Redrawn from [7].)

Bryant and co-workers have carried out experimental and numerical investigations of the glory, and have developed some interesting phenomenological models. In [54] the scattered intensity of the axial rays is given as

$$I_a = \frac{R\beta^2}{2(2-N^2)}[(N^2-2N+2)+N(2-N)\cos 4N\beta]\,, \tag{3.9}$$

where R is the reflection coefficient at normal incidence. This axial contribution is not sufficient to explain the glory phenomenon [7] so Bryant and Cox [54] use the exact Mie theory to elucidate information about surface wave contributions to the glory. (It should be pointed out that in addition to surface waves there is also the possibility of such waves taking "short cuts" through the sphere; this is addressed below.) In particular they calculated scattered intensities at 90° and 180° of incident light polarized both normal to and parallel to the plane of scattering for a size parameter $\beta \simeq 200$. The total cross section contains large spikes at the 180° intensity; this feature, much reduced in magnitude, also remains at 90°, with spikes occurring alternately in one polarization but not the other (see Fig. 2 in [54]). The periodicity of these spikes in β is twice that of the total cross section and the 180° intensity. In the latter there is a sinusoidally varying component of period near 0.8. This "*ripple*" in the extinction curve (i.e. the ratio of total cross section to geometric cross section vs β) thus becomes enormously amplified in the differential cross section at 180°. It is well known that the cross sections are extremely sensitive to small changes in the input parameters [10].

Mention has been made already of the *localization principle*: terms of order n in the Mie expansion correspond to a ray passing the origin at a distance $(n+\frac{1}{2})/k$ because as is known in connection with partial-wave analysis, the nth partial wave is associated with angular momentum

$$M = \sqrt{n(n+1)}\hbar \sim b_n = (n+\tfrac{1}{2})\hbar$$

(the last approximation being valid for large values of n). The appropriate impact parameters are concentrated within the annular region $[n\lambda/2\pi, (n+\frac{1}{2})\lambda/2\pi]$, and for the mid-range impact parameter b_n, $M = pb_n$, where p is the incident linear momentum, which in the semiclassical approximation is associated classically with a particle of momentum h/λ. Hence

$$b_n \approx \frac{(n+\frac{1}{2})}{k} .$$

The convergence of the Mie expansion for $n \gtrsim \beta = ka$ is (physically) a consequence of the fact that the rays no longer interact with the scattering sphere (again, though, this is not the complete picture; as has been noted in the introduction, tunneling to the surface can occur for impact parameters greater than a, giving rise to the sharp resonances found in the ripple: this will be further elaborated in Section 5). The results for the scattering at 180° indicate that they are determined almost entirely by rays nearly tangential to the surface together with a smaller contribution from the axial rays.

Based on these results, an interesting phenomenological model for the surface waves was developed in [54]. The idea is that a tangentially incident bundle of rays is trapped on the spherical surface and propagates around it with constant attenuation due to reradiation. In terms of unit amplitude at its point of tangential entry at 0°, at 90° it is A, A^2 at 180°, A^3 at 270°, and so on, where $|A| < 1$. The scattering amplitude at 0° can be represented as

$$S(0^\circ) = \alpha(1 + A^4 + A^8 + A^{12} + \cdots) + B , \tag{3.10}$$

where α is a real constant, and the term B represents the contribution to zero-degree scattering from other mechanisms. Clearly

$$S(0^\circ) = \alpha(1 - A^4)^{-1} + B , \tag{3.11}$$

so writing $A^4 = be^{i\phi}$ and using the optical theorem [7] in the form

$$Q = 4\,\mathrm{Re}[S(0^\circ)]\beta^{-2} , \tag{3.12}$$

Q being the extinction coefficient, it is found that

$$Q = Q_0 + \frac{4\alpha}{\beta^2}\left(\frac{1 - b\cos\phi}{1 + b^2 - 2b\cos\phi}\right) , \tag{3.13}$$

where Q_0 is determined by the term B. The authors then consider a region of resonance near $\phi = 2m\pi$, m being an integer and ϕ being replaced by $\phi \bmod 2\pi$. Thus it follows that if $b = e^{-g}$ and ϕ and g are both small

$$Q(\phi) \approx \frac{D_1}{1 + (\phi/g)^2} + D_2 , \tag{3.14}$$

where D_1 and D_2 are insensitive to small changes in ϕ. The period of this peak is 0.809, so $\phi = 2\pi\beta/0.809$, and the authors fit the equation for $Q(\phi)$ to their numerical results for Q, yielding a value for $g = 0.008$. This means that $b \approx 0.99$, and this represents the attenuation around one circumference of the drop. Being so small, it implies that the peak is an interference maximum involving tens, perhaps even hundreds of circuits by the surface wave. A similar argument for $\theta = 180^\circ$ yields

$$S(180^\circ) = \alpha A^2(1 - A^4)^{-1} + B' \tag{3.15}$$

with a corresponding expression for the intensity $|S(180^\circ)|^2$; if the 90° and 270° scattering amplitudes are combined then

$$S(90^\circ/270^\circ) = \alpha A(1 - A^2)^{-1} + B'' \tag{3.16}$$

and from the definition of ϕ it is clear that the resonances at 90° have twice the period that is exhibited by $S(180^\circ)$ and Q.

In a subsequent comparison of experimental results with Mie theory, Fahlen and Bryant [55] developed a two-parameter scalar model in an attempt to gain insight into the scattering mechanisms involved in the scattering data they acquired. Their model again involved surface waves; such waves transport energy tangential to the interface between two media, and arises when that energy strikes such an interface (between two dielectric media) from the more dense side at or exceeding the critical angle for total reflection (allowing for a limiting process for the initial interaction from the less dense side). It is known from several experiments [56,57] that as a surface wave travels along the interface, it continuously reradiates energy back into the denser medium at the angle of critical refraction. It is also attenuated by "spraying" off energy tangentially as it propagates around the drop surface. Part of the energy refracted back into the drop (the short cut component) is reflected internally and the remaining part interferes with the surface wave; this occurs over the course of many revolutions around the drop. For a drop of radius a it is shown that the total optical path length for a ray performing M revolutions around the drop and n jumps (short cuts) across it is

$$l(n,M) = 2\pi Ma + \theta a + 2na(\tan\alpha - \alpha)\,, \tag{3.17}$$

where θ is the scattering angle subtended by both the chord and the arc AB (see Fig. 12). Denoting the amplitude of such a ray by $A(n,M)$, the probability (statistical weight in [55]) for n jumps by $P(n,M)$ and the surface wave attenuation by $\exp(-\gamma z)$, z being the surface distance defined by

$$z = 2\pi Ma + \theta a - 2n\alpha a\,,$$

then

$$A(n,M) = P(n,M)\exp[-\gamma z + 2\pi i l(n,M)/\lambda]\,. \tag{3.18}$$

The (complex) total amplitude A_t is the sum of all possible paths, i.e.

$$A_t = \sum_{M=0}^{\infty}\sum_{n=0}^{n^*} A(n,M)\,, \tag{3.19}$$

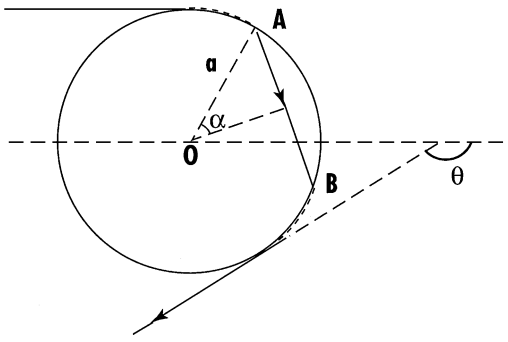

Fig. 12. A general representation of both surface-wave (to A and from B) and "shortcut" components contributing to the glory formed from water droplets. (Redrawn from [55].)

where $n^* \leqslant (2\pi M + \theta)/2\alpha$ must be an integer. If $P(n,M) = \exp(-pn)$ and $g = \gamma\lambda/2\pi$ then it may be shown that

$$\mathrm{Re}\, A_t = \sum_{M=0}^{\infty} \sum_{n=0}^{n^*} \exp[-pn - g\beta(\theta + 2\pi M - 2n\alpha)] \cos[\beta(\theta + 2\pi M + 2n(\tan\alpha - \alpha)] \tag{3.19a}$$

with a similar expression (involving a sine term) for $\mathrm{Im}\, A_t$.

The two parameters in this mathematical model are p (related to the jump probability) and g (related to the attenuation of surface wave energy). By varying them to obtain the best agreement between the experiment and the Mie theory curve for intensity vs size parameter β, Fahlen and Bryant find values $p = 0.01$ and $g = 0.004$. For these values the agreement is good: the correct periodicity is present and also the sharp intensity resonances, even though no allowance has been made for polarization. The combination of surface wave propagation and short cuts across the drop are consistent with the more technical theory of Nussenzveig [6] and contain many (but not all) of the essential components for a satisfactory explanation of the glory. As mentioned above in connection with tunneling to the surface, a complete account of the glory has now been afforded by more recent work of Nussenzveig and coworkers; this will be reviewed in Sections 5.6 and 5.8.

3.2. Rainbow glories

The topic of rainbow-enhanced forward and backward glory scattering (rainbow glories) was studied by Langley and Morrell using both the geometrical optics approximation and Mie theory [58] (there are many further useful references in this paper). For certain values of the refractive index N in spherical scattering objects, a higher-order rainbow can coincide with a forward or backward glory to produce exceptionally strong scattering in those directions. The relevant values of N can be found from geometrical optics, but as already seen, in order to characterize the glory adequately it is necessary to consider the effects of surface waves, tunneling, internal resonances and high-order rainbows [53,59,60] along with complex angular momentum techniques [61,10] as noted in the previous subsection.

The total deviation angle D undergone by a ray is, in the present notation,

$$D = 2\mathrm{i} - 2rP + (P-1)\pi \ . \tag{3.20}$$

Forward and backward scattering correspond to $D = L\pi$, so from this relation and Snell's law the glory condition is found to be

$$N = (\csc r)\cos\{rP - \pi(P-L)/2\} \ . \tag{3.21}$$

The ray paths may be conveniently designated by the ordered pair (P,L) where P is the number of chords the ray makes inside the sphere and L is the number of times the ray crosses the optical axis. The authors demonstrate that for a given N there may be $0, 1$ or 2 glory rays of type (P,L) with the following ranges:

3.2.1. Ray

$$\sec(L\pi/2P) > N > 0, \quad P-1 > L \ ,$$

$$\csc(\pi/2P) \leqslant N \leqslant P, \quad P-1 = L \ .$$

3.2.2. Rays

$$\sec(L\pi/2P) \leqslant N \leqslant N^*, \quad P-1 > L > 0 \ .$$

N^* is the refractive index for which a rainbow coincides with a glory; using the (necessary) rainbow condition $\mathrm{d}D/\mathrm{d}i = 0$ is, on using Snell's law and Eq. (3.20),

$$\tan i = P\tan r \ . \tag{3.22}$$

The requirement that both the rainbow and glory conditions be satisfied is therefore

$$P\tan r = B(\tan rP)^B \tag{3.23}$$

where $B = (-1)^{P-L-1}$. For a given (P,L) the solution of this equation for r when substituted in (3.21) gives the value of N^*. A table of eight such values can be found in [58] along with detailed figures and calculations for the physical optics approximation and the Mie theory; we note several features here. By comparing incremental areas of the wavefront entering and exiting the sphere, the parameter $A = 2(P\tan r - \tan i)$ occurs, where A is zero at the rainbow angle. Using this parameter, the radius of curvature of the wavefront at the exit plane is defined to be

$$Q(N) = \frac{A}{A + \sin i} \ .$$

$Q = 0$ corresponds to a cubic wavefront, but the glory wavefront may be diverging ($Q > 0$) or converging ($Q < 0$) at the exit plane, depending on the value of N. Many of the mathematical details follow from the earlier work [62,63]. Normal glory scattering leads to an irradiance with size parameter dependence $\mathcal{O}(\beta)$; for a rainbow-enhanced glory this becomes $\mathcal{O}(\beta^{4/3})$, based on the physical optics approximation, and the Mie theory calculations confirm this at the predicted values of N. The polarization properties of rainbow glories are found to be similar to those for normal glory scattering.

3.3. The forward glory

The optical glory is a strong enhancement in near-backward scattering of light from water droplets with size parameter $\beta \gtrsim 100$, arising from a combination of almost grazing incident rays, diffraction of and interference between surface waves, ray shortcuts across the drop interior, and complex rays in the shadow of high-order rainbows. The enhancement is associated with an axial focusing effect of order $\sqrt{\lambda_G}$, where $\lambda_G = kb$ is the glory angular momentum [7], but for water droplets the major contribution to the phenomenon is from complex trajectories. Nussenzveig and Wiscombe [64] studied the *forward* (optical) glory, and noted the following necessary conditions for such a glory to be observed in Mie scattering: (i) the contributing ray paths should involve the smallest possible number of internal reflections to minimize attenuation by reflection and absorption along the optical path (the latter occurring if $\mathrm{Im}\, N > 0$); (ii) the incident rays should be nearly tangential to minimize energy losses from internal reflections (so the internal reflectivity should be as large as possible); and (iii) the effect must be large enough not to be swamped by the intense forward diffraction peak (which is $O(\beta)$).

The smallest number of internal reflections that can lead to a forward glory is two; this requires $\mathrm{Re}\, N \geqslant 2$. For $\mathrm{Re}\, N = 2$, tangential incidence gives rise to an internal geometrical resonance path [6] in the shape of an equilateral triangle. For $\mathrm{Re}\, N < 2$, of interest in this report, the triangle does not close, resulting in an angular gap ζ to be bridged by surface rays, (where for simplicity of notation we write N for $\mathrm{Re}\, N$) [64]

$$\zeta = 6 \arcsin(1/N) - \pi \,. \tag{3.24}$$

For $N = 1.33$, $\zeta \simeq 112.5^\circ$, but for $N = 1.85$, $\zeta \simeq 16^\circ$. This is indicative of the sensitivity of $\zeta(N)$, and as pointed out by Khare [42], for large values of p ($p-1$ being the number of internal reflections) this becomes extreme, and care must be taken in employing asymptotic expressions for the Debye amplitudes (note that if $\zeta < 0$, the contribution is from rainbow terms, not surface waves [53]).

In their calculations comparing the exact Mie theory with the complex angular momentum approximation, Nussenzveig and Wiscombe note that the difference between the extinction efficiencies does not show the irregular ripple fluctuations seen in the backward glory, but instead a regular, nearly sinusoidal oscillation arising from interference between the forward diffraction peak and the forward glory contributions. Comparison of the predicted oscillation periods δ from both methods shows excellent agreement; the complex angular momentum theory predicting that [61]

$$\delta = 2\pi/(6\sqrt{N^2 - 1} + \zeta) \tag{3.25}$$

Note that for the next geometrical resonance (an inscribed square), occurring for $N = \sqrt{2}$, the corresponding δ has a factor of 8 instead of 6 in (3.25), and the corresponding ζ is given by

$$\zeta = 8 \arcsin(1/N) - 2\pi \,. \tag{3.26}$$

Concerning polarization, the authors state that for scattering angles $\theta \ll \beta^{-1/3}$ and $\beta\theta$ not $\gg 1$, both scattering amplitudes tend to be dominated by the Airy pattern of Fraunhofer diffraction by a circular disk, namely $J_1(\beta\theta)/\beta\theta$, which gives rise to the forward diffraction peak. The amplitudes are polarized in a similar manner to those in the backward glory: one component

exhibiting the Airy pattern behavior and the other with a $J_1'(\beta\theta)$ angular dependence. In short, the forward glory is strongly polarized, with the electric (parallel) component dominating, characteristic of surface waves [61].

4. Semi-classical and uniform approximation descriptions of scattering

In a primitive sense, the semi-classical approach is the "geometric mean" between classical and quantum mechanical descriptions of phenomena; while one wishes to retain the concept of particle trajectories and their individual contributions, there is nevertheless an associated de Broglie wavelength for each particle, so that interference and diffraction effects enter the picture. The latter do so via the transition from geometrical optics to wave optics. The differential scattering cross section is related to the quantum scattering amplitude $f(k,\theta)$ and this in turn is expressible as the familiar partial wave expansion [65]. The formal relationship between this and the classical differential cross section is established using the WKB approximation, and the principle of stationary phase is used to evaluate asymptotically a certain phase integral (see [10], Section (1.2) and below for further details). A point of stationary phase can be identified with a classical trajectory, but if more than one such point is present (provided they are well separated and of the first order) the corresponding expression for $|f(k,\theta)|^2$ will contain interference terms. This is a distinguishing feature of the 'primitive' semiclassical formulation, and has significant implications for the four effects (rainbow scattering, glory scattering, forward peaking and orbiting) noted in Appendix A. The infinite intensities (incorrectly!) predicted by geometrical optics at focal points, lines and caustics in general are "breeding grounds" for diffraction effects, as are light/shadow boundaries for which geometrical optics predicts finite discontinuities in intensity. Such effects are most significant when the wavelength is comparable with (or larger than) the typical length scale for variation of the physical property of interest (e.g. size of the scattering object). Thus a scattering object with a "sharp" boundary (relative to one wavelength) can give rise to *diffractive scattering* phenomena.

Under circumstances appropriate to the four critical effects noted above, the primitive semiclassical approximation breaks down, and diffraction effects cannot be ignored. Although the angular ranges in which such critical effects become significant get narrower as the wavelength decreases, the differential cross section can oscillate very rapidly and become very large within them. As such the latter are associated with very prominent features and in principle represent important probes of the potential, especially at small distances. The important paper by Ford and Wheeler [66] contained *transitional asymptotic approximations* to the scattering amplitude in these 'critical' angular domains, but they have very narrow domains of validity, and do not match smoothly with neighboring 'non-critical' angular domains. It is therefore of considerable importance to seek *uniform asymptotic approximations* that by definition do not suffer from these failings [68]. Fortunately, the problem of plane wave scattering by a homogeneous sphere exhibits all of the critical scattering effects (and it can be solved exactly, in principle), and is therefore an ideal laboratory in which to test both the efficacy and accuracy of the various approximations. Furthermore, it has relevance to both quantum mechanics (as a square well or barrier problem) and optics (Mie scattering); indeed, it also serves as a model for the scattering

of acoustic and elastic waves, and was studied in the early 20th century as a model for the diffraction of radio waves around the surface of the earth [67].

The existence of a rainbow angle, θ_R (or θ_0 as has also been used above) is not just relevant to the meteorological phenomenon; along with the glory it has counterparts in atomic, molecular, heavy-ion and nuclear scattering (though the glory is usually a *forward* glory in these situations). These fundamentally important aspects of 'scattering problems' will be summarized in Section 7, but a few comments are in order at this juncture. In general, whenever a plane wave is scattered by a spherically symmetric potential there is an associated scattering angle θ_R (measured from the forward direction) for which the differential cross section $\sigma(\theta)$ is large [68]. Just as with the optical rainbow, on the "shadow" side of θ_R the scattering drops rapidly to zero; on the "lit" side $\sigma(\theta)$ oscillates. Ford and Wheeler [66] referred to this as *rainbow scattering* whenever it occurs in other scattering situations, where provided $\theta_R \neq 0, \pi$ it is called the rainbow angle. In particular rainbow scattering occurs in the intermolecular collisions arising during molecular beam experiments, where "molecular rainbows" can provide information about the intermolecular forces ([20,69]; an early non-mathematical account can be found in [70]); this is often referred to as an *inverse problem* in the mathematics literature. The inverse problem in scattering theory refers to the construction of the potential directly from the measured data, e.g. in molecular scattering, a set of phase shifts may be obtained from the observed angular and energy dependence of the cross section, and the potential is deduced from these phase shifts. There is an enormous amount of literature on the subject of inverse problems in general, ranging from the mathematical theory (e.g. questions of existence and uniqueness of the potential) to the practical aspects of implementing the theory. Since uniqeness is not usually guaranteed in problems of this type, there has been interest in identifying all equivalent potentials for a given set of phase shifts. For a summary of both aspects of the problem prior to 1974 (but with emphasis on practical inversion procedures) see the excellent review by Buck [71] where many further references can be found. A discussion of more recent developments in the field is found in Section 6.5.

The types of phenomena identified thus far typically arise when the scattering "center" is much larger than the wavelength of the incident beam (as in the meteorological rainbow); under these circumstances a large number of partial waves contribute to the scattering, giving rise to significant variations in the differential cross section. Under these conditions, as pointed out by Berry [68], the system is close to both *the geometrical optics limit of electromagnetism and the classical limit of quantum mechanics*. However, the calculations are not the same for optical and atomic/molecular/nuclear rainbows because in the former case the refractive index is discontinuous (or at least changes significantly over distances much smaller than a wavelength, i.e. at the drop surface) and in the latter case the potential is continuous (this latter point has been made already in connection with truncated potentials). There are several approximations that have been used to describe rainbow scattering (to be understood in the general sense in this section). In [68] these are classified as (i) the classical approximation, (ii) the "crude semi-classical" (or WKB) approximation, (iii) the Airy approximation, and (iv) the uniform approximation. Mention has been already made of these, but now more details will be included.

The classical approximation can be obtained by smoothing out the oscillations in the asymptotic form of the exact cross section for angles far from θ_R and analytically continuing the result to angles near θ_R; this is basically equivalent to Déscartes' theory in the case of the

optical rainbow, and as in that case, it diverges at θ_R. The 'crude semi-classical' approach is the unsmoothed asymptotic form of $\sigma(\theta)$ far from θ_R and it also diverges if continued analytically to θ_R. The Airy approximation here means the application to continuous potentials of the previously discussed account of the light intensity in the neighborhood of a caustic and developed by Airy for the optical rainbow problem. This gives the form of $\sigma(\theta)$ close to θ_R but rapidly becomes inaccurate as θ deviates from θ_R by more than a degree or two. Since it is such an important paper, we will defer discussion of the Ford/Wheeler contribution [66] until later, and content ourselves with summarizing here the seminal contribution of Berry [68] to the fourth classification—the uniform approximation—which in some respects is very similar to (and pre-dates) the application of the Chester et al. [72] saddle-point method to the rainbow problem.

In his valuable paper Berry discusses the reason for the existence of all these different approximations. Each is valid in a restricted angular domain, and the asymptotic expansion of the scattering amplitude $f(\theta)$ with respect to the (very) small parameter $\hbar$ (when $|\theta-\theta_R|$ is large) changes its form from a series in powers of $\hbar^{1/2}$ to a series in powers of $\hbar^{1/3}$ (when $|\theta-\theta_R| \ll 1$). This phenomenon is familiar for functions defined by ordinary differential equations, which is why the term 'crude semi-classical' is associated with the WKB approximation. By contrast, the Airy approximation is called a *transitional* approximation: it 'heals the wound' between one region and another in a smooth, well-defined manner [73]. The uniform approximation, as its name implies, is *valid for the whole variable domain* (see [74,75]). Berry transforms the eigenfunction expansion of $f(\theta)$ by the *Poisson sum formula* (see Appendix C) to give a series of integrals (as does Rubinow in an important but infrequently noted paper [76], and Nussenzveig [33]; indeed, in 1950 Pekeris [77] suggested that the Poisson formula is the natural mathematical tool to use when considering an eigenfunction expansion in a region where ray/classical concepts are appropriate). As noted earlier in connection with the Mie series, such a transformation is exceedingly helpful in practical terms because the standard Faxen–Holtsmark formula for $f(\theta)$, while exact, converges extremely slowly for θ in the neighborhood of θ_R; often many thousand partial waves are required in this region.

The application of the Poisson sum formula allows $f(\theta)$ to be written in terms of integrals which, in the notation of [68], may be referred to as $I_m^{\pm}$ (though we do not write them down here; see Eqs. (4)–(9) in [68]). The integrands of $I_m^{\pm}$ are rapidly oscillating functions, and it is known that the main contributions to the integrals come from the neighborhoods of stationary points on the positive real axis (when they exist) [78]. We will describe the basic features of this analysis below, but it suffices here to note that the points of stationary phase are defined in terms of a quantity $\Theta(l)$ (this is in fact the classical deflection function—see below and Appendix A); rainbow scattering arises whenever $\Theta(l)$ has an extremum. As we will see in Section 5, there are in fact two real stationary (saddle) points if $\theta < \theta_R$; if $\theta > \theta_R$ they are complex conjugates, having moved away from the real axis, and only one of these contributes significantly to the amplitude. At $\theta = \theta_R$ they coalesce into a third-order saddle point, for which the semiclassical approach is not valid (recall that the direction θ_R defines a caustic direction). The uniform approximation, based on the work of Chester et al. [72] comes, so to speak, to the rescue. Instead of treating the saddle points separately, as in semi-classical methods, or as essentially coincident, as in the Airy approximation, this method maps their *exact* behavior onto that of the stationary points of the integrand in the Airy function. The resulting equations

for the rainbow cross section σ_R are rather complicated, but Berry shows that on both the lit and shadow sides of θ_R they reduce in the appropriate limits to that deduced from the Airy approximation.

Ford and Wheeler [66] identify four mathematical approximations that can be considered to define the term "semi-classical" approximation. The differential scattering cross section into unit solid angle at θ is

$$\sigma(\theta) = |f(\theta)|^2 \,, \tag{4.1}$$

where $f(\theta)$ is expressed in the form

$$f(\theta) = \frac{1}{2ik} \sum_{l=0}^{\infty} (2l+1)(e^{2i\eta_l} - 1) P_l(\cos\theta) \,, \tag{4.2}$$

η_l being the phase shift for the lth partial wave. The first approximation is:

(i) η_l is replaced by its WKB approximate value, namely

$$\eta_l = \frac{\pi}{4} + \frac{\pi l}{2} - r_0 + \int_{r_0}^{\infty} (\kappa(r) - k)\, dr \,, \tag{4.3}$$

where

$$\kappa(r) = \left[\frac{2m(E-V)}{\hbar^2} - \frac{(l+\frac{1}{2})^2}{r^2} \right]^{1/2} \tag{4.4}$$

and r_0 is the turning point of the (classical) motion, defined by $\kappa(r) = 0$ (see Appendix A). Physically, this approximation is equivalent to the requirement that the potential V be "slowly varying", i.e.

$$\left| \frac{1}{kV} \frac{dV}{dr} \right| \ll 1 \,. \tag{4.5}$$

As Ford and Wheeler point out, the most important property of the WKB phase shift is its simple relation to the classical deflection function $\Theta(l)$:

$$\Theta(l) = 2\frac{d\eta_l}{dl} \,. \tag{4.6}$$

This allows a correspondence of sorts to be made between the quantum and classical results.

(ii) The second approximation concerns the replacement of the Legendre polynomials by their aymptotic forms for large values of l. Thus (a)

$$P_l(\cos\theta) \cong [\tfrac{1}{2}(l+\tfrac{1}{2})\pi \sin\theta]^{-1/2} \sin\left[(l+\tfrac{1}{2})\theta + \frac{\pi}{4}\right], \quad \sin\theta \gtrsim l^{-1} \tag{4.7}$$

and (b)

$$P_l(\cos\theta) \cong (\cos\theta)^l J_0[(l+\tfrac{1}{2})\theta], \quad \sin\theta \lesssim l^{-1} \,. \tag{4.8}$$

For this approximation to be valid, many l-values must contribute to the scattering at a given angle (and the major contribution to the scattering amplitude comes from values of $l \gg 1$). The third approximation is

(iii) Replacement of the sum over l by an integral with respect to the same variable. Again, this means that many partial waves should contribute to the scattering and also that $\eta_l(l)$ should vary slowly and smoothly. Approximations (i) and (ii) render both η_l and P_l continuously differentiable functions of their arguments (in general), so that (iii) is appropriate.

Approximations (i), (ii)(a) and (iii) (for angles θ not close to 0 or π) yield the semi-classical form of the scattering amplitude

$$f_{\rm sc} = -k^{-1}(2\pi\sin\theta)^{1/2}\int_0^\infty (l+\tfrac{1}{2})^{1/2}[e^{i\phi_+} - e^{i\phi_-}]\,dl\,, \tag{4.9}$$

where the phase functions $\phi_\pm(l,\theta)$ are defined by

$$\phi_\pm(l,\theta) = 2\eta_l \pm (l+\tfrac{1}{2})\theta \pm \frac{\pi}{4} \tag{4.10}$$

and the result $\sum_l (2l+1)P_l = 0$ has been used [79]. A fourth approximation, not always necessary, is that

(iv) It may be necessary for the integral in (iii) above to be evaluated by the method of stationary phase, or the method of steepest descent, as we have seen already. Mott and Massey [79] examined the integral under circumstances such that there is only *one* point of stationary phase (corresponding to $\Theta(l)$ varying monotonically between 0 and $\pm\pi$). This is not the case in rainbow or glory scattering, where $\Theta(l)$ is not monotone everywhere in its domain. Ford and Wheeler study five special features of scattering in the semiclassical approximation; we mention only the interference/rainbow/glory features in keeping with the emphasis of this review. The mathematical details can be found in [79].

Because (classically) several incident angular momenta may correspond to the same scattering angle at a given energy, the total cross section will in general be the sum of different contributions from the different branches; and when these are well separated, each will give an independent contribution to the scattering amplitude which may be evaluated by the method of stationary phase. Then, formally, the semi-classical amplitude becomes

$$f_{\rm sc} = \sum_j (\sigma_{\rm cl})_j^{1/2} e^{i\beta_j}\,, \tag{4.11}$$

where β_j is a phase angle for the jth contributing branch. If $j = 1,2$ only, then

$$\sigma_{\rm sc} = |(\sigma_{\rm cl})_1^{1/2} + (\sigma_{\rm cl})_2^{1/2} e^{i(\beta_2-\beta_1)}|^2 \tag{4.12}$$

so that $\sigma_{\rm sc}$ will oscillate between minimum and maximum values defined by the quantities $[(\sigma_{\rm cl})_1^{1/2} \pm (\sigma_{\rm cl})_2^{1/2}]^2$, with "wavelength" $\Delta\theta = 2\pi/|l_2 - l_1|$, the $l_{1,2}$ being the points of stationary phase.

The rainbow angle corresponds to the singularity arising in the classical cross section when $d\Theta(l)/dl$ vanishes (the cross section contains a factor $(d\Theta(l)/dl)^{-1}$). Near a rainbow angle, $\Theta(l)$ may be approximated by the quadratic function

$$\Theta(l) = \theta_R + q(l - l_R)^2\,, \tag{4.13}$$

where the terms have obvious meanings. On the "lit" or bright side of the rainbow angle the classical intensity is

$$\sigma_{\rm cl} = \left(\frac{l_R + \frac{1}{2}}{k^2 \sin\theta_R} \right) |q(\theta - \theta_R)|^{-1/2} \tag{4.14}$$

and on the "dark" or shadow side, the classical intensity is zero (if there are no additional contributing branches of the deflection function $\Theta(l)$). This of course is identical with the predictions of geometrical optics for the optical rainbow. In terms of the phase shift η_l however,

$$\eta_l = \eta_R \pm (\theta_R/2)(l - l_R) + (q/6)(l - l_R)^3 \,, \tag{4.15}$$

where the $\pm$ is necessary for the antisymmetry (see Eq. (4.10) for $\phi_\pm$ above). For $\theta_R > 0$ the dominant contribution to integral (4.9) will come from the term containing the factor $\exp(\mathrm{i}\phi_-)$; this yields essentially the same functional form for the scattering differential cross section as does the original Airy theory for the intensity of the optical rainbow. Thus

$$f_{\rm sc} = [2\pi(l_R + 1/2)/k^2 \sin\theta]^{1/2} q^{-1/3} \mathrm{e}^{\mathrm{i}\delta} \mathrm{Ai}(x) \tag{4.16}$$

where

$$x = q^{-1/3}(\theta_R - \theta) \,,$$

is a measure of the deviation from the rainbow angle, and

$$\delta = 2\alpha_R - \pi/4 + (l_R + 1/2)(\theta_R - \theta) \tag{4.17}$$

for $\Theta(l_R) > 0$. If $\Theta(l_R) < 0$, then $\theta_R - \theta$ must be replaced by its negative in the expressions for x and δ. The quantity α_R is the intercept of the tangent line to the $(\eta_R, l_R + 1/2)$ curve with the vertical axis, i.e.

$$\alpha_R = [\eta_l - (l + 1/2)(\mathrm{d}\eta_l/\mathrm{d}l)]_{l=l_R} \tag{4.18}$$

(see Fig. 1 in [66]). The form of the Airy integral used there is

$$\mathrm{Ai}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\left[\mathrm{i}xu + \frac{1}{3}\mathrm{i}u^3 \right] \mathrm{d}u \,. \tag{4.19}$$

The differential cross section near the rainbow angle then has the form

$$\sigma_{\rm sc} = [2\pi(l_R + 1/2)/k^2 \sin\theta] |q^{-2/3}| \mathrm{Ai}^2(x) \,. \tag{4.20}$$

If other branches of $\Theta(l)$ do contribute at θ_R, then $f_{\rm sc}$ above must be combined with the other contributing amplitudes before the absolute square is taken (this in fact arises (theoretically) for magnetic pole scattering, and also for ion-atom scattering (see a detailed discussion in [66])).

Provided $0 < \Theta(l) < \pi$, or $-\pi < \Theta(l) < 0$, $\sigma_{\rm sc}$ can be described entirely in terms of the classical cross section, together with interference effects (not discussed here) and rainbow scattering. If however $\Theta(l)$ passes smoothly through 0, $\pm\pi$, etc. then the vanishing of $\sin\Theta(l)$ for $l, |\mathrm{d}\Theta/\mathrm{d}l| < \infty$ leads to a singularity in the cross section for both forward and backward

scattering. The optical/meteorological terminology is again borrowed for this situation: it is referred to as a *glory*. In the case e.g. of a backward glory, near $\Theta(l)=\pi$, we may write the approximate form

$$\Theta(l)=\pi+a(l-l_g) \tag{4.21}$$

and then $\sigma_{\rm cl}$ in the backward direction is just the sum of two equal contributions from $\Theta<\pi$ and $\Theta>\pi$, i.e.

$$\sigma_{\rm cl}=\frac{2l_g}{k^2|a|(\pi-\theta)}\ . \tag{4.22}$$

From (4.21) the phase shift, for those values of l that contribute most to the glory, may be written as

$$\eta_l=\eta_g+\frac{\pi}{2}(l-l_g)+\frac{a}{4}(l-l_g)^2 \tag{4.23}$$

and after some rearrangements of the integral in the expression for $f_{\rm sc}$, it can be shown that the glory cross section is

$$\sigma_{\rm sc}=(l_g+1/2)^2(2\pi/k^2|a|)J_0^2(l_g\sin\theta)\ , \tag{4.24}$$

when there are no interference effects. For a forward glory the result is similar; there is a different phase term but this makes no contribution of course to $\sigma_{\rm sc}$. Thus the singularity in $\sigma_{\rm cl}$ is replaced by a finite peak in both forward and backward directions of

$$\sigma_{\rm sc}^{\rm max}=(l_g+1/2)^2(2\pi/k^2|a|)\ . \tag{4.25}$$

The Bessel function oscillations may be interpreted as resulting from interference between contributions from the two branches of $\Theta(l)$ near a glory, i.e. $\Theta>\pi$ (or 0) and $\Theta<\pi$ (or 0). Ford and Wheeler point out that when the intensity is averaged over several such oscillations (using the result $\langle J_0^2(x)\rangle=(1/\pi x)$) then $\sigma_{\rm sc}$ reduces to the classical expression $\sigma_{\rm cl}$. For further early references to atomic and molecular scattering, see [80–87]; more recent work in these contexts and in nuclear scattering is reviewed in Section 6.5.

5. The complex angular momentum theory: scalar problem

5.1. The quantum mechanical connection

An important summary of the main results for high frequency scattering by a transparent sphere has been provided by Nussenzweig [34]. A rather comprehensive set of references to this and related work up to 1991 can be found in his book [10]; there are several more recent papers, along with those of co-workers (e.g. [88,89]; see also an early paper on square well and barrier potentials [90]).

As has been pointed out already, in terms of the size parameter $\beta = ka$, the wavenumber k, the real refractive index N and the impact parameter for the lth partial wave

$$b_l = \frac{(l+1/2)}{k}\,, \tag{5.1}$$

there are certain assumptions necessary for reasonable approximations to be made to the scattering problem. The presence of absorption can be incorporated by allowing N to be complex-valued, but that will not be necessary here; although according to [5] it should not be unduly difficult to carry out.

The essential mathematical problem for scalar waves can be thought of either in classical terms, e.g. the scattering of sound waves, or in wave-mechanical terms, e.g. the non-relativistic scattering of particles with momentum $p = \hbar k$ by a square potential well (or barrier) of radius a and depth (or height) V_0, i.e.

$$\begin{aligned} &(2) \qquad V(r) = -V_0 \quad (0 \leqslant r \leqslant a)\,, \\ &\qquad\quad V(r) = 0 \quad (r > a)\,. \end{aligned} \tag{5.2}$$

The governing time-independent radial Schrödinger equation for this potential is

$$\left[\frac{\mathrm{d}^2}{\mathrm{d}r^2} - \frac{l(l+1)}{r^2} + \frac{2m}{\hbar^2}(E - V(r))\right]\psi(r) = 0\,, \tag{5.3}$$

where $\psi(r)$ is the scalar radial wavefunction and the third term represents the centrifugal barrier [91] and $E = p^2/2m$ is the eigenvalue parameter (energy: note that expressed in units for which $\hbar = 2m = 1$, E is merely equal to k^2). In the $l = 0$ case in particular, this equation has a useful "classical" form if the wavenumber k is used, i.e.

$$\left[\frac{\mathrm{d}^2}{\mathrm{d}r^2} + k^2 N^2(r)\right]\psi(r) = 0\,, \tag{5.4}$$

where the refractive index is, for the potential well [5,76,92]

$$N = \left(1 + \frac{2mV_0}{\hbar^2 k^2}\right)^{1/2}\,, \tag{5.5}$$

m being the mass of the "particle". It follows that $N > 1$ corresponds, for fixed k (or energy), to a potential well (crudely, the raindrop can "trap" the 'probability waves' to some extent) while $N < 1$ corresponds to a barrier (inhibiting the waves to a lesser degree, as with an air bubble in water). It is the former case that will concern us here. Other related interpretations are possible: for a fixed V_0, N is dispersive, being frequency-dependent, and if N is fixed then the potential is proportional to the energy, as noted earlier.

Partial waves for which $b_l \lesssim a$ (i.e. $l + \frac{1}{2} \lesssim \beta = ka$) are substantially distorted by the spherical scatterer, so that at least β terms are required for reasonable accuracy (in fact it transpires that a better measure of the number of terms is given by $l_+ \sim \beta + c\beta^{1/3}$, where $c \gtrsim 3$, [93]). This can be interpreted in terms of the centrifugal barrier penetration: from the radial equation (5.3)

the effective potential is (see Appendix A)

$$U(r) = V(r) + \frac{\hbar^2 l(l+1)}{2mr^2} \ . \tag{5.6}$$

In fact, if the WKB approximation is to be applied, a correction is in order, as pointed out by Berry [94]; the term $l(l+1)$ should be replaced by $(l+\frac{1}{2})^2$. This is because the ordinary WKB solutions are not valid near the origin (because $U(r)$ varies rapidly there) and so cannot satisfy the boundary condition that the reduced radial wavefunction is zero at the origin. However, as shown by Langer [74] and Bertocchi et al. [95], the WKB approximation *is* valid near the origin provided the above change is made in the effective potential $U(r)$. Under these circumstances the modified WKB wavefunction is

$$\psi(r) \propto \frac{\exp\{\pm \mathrm{i} \int^r \kappa(r)\, \mathrm{d}r\}}{\{\kappa(r)\}^{1/2}} \ , \tag{5.7}$$

where from (4.4)

$$\kappa^2(r) = \frac{2m}{\hbar^2}\left\{E - V(r) - \frac{\hbar^2(l+\frac{1}{2})^2}{2mr^2}\right\} \ . \tag{5.8}$$

Note also that $U(r)$ is discontinuous at $r=a$, introducing a barrier, so $b_l > a$ corresponds to an energy level below the top of the centrifugal barrier. Berry states that the transmissivity of the barrier up to $r=a$ (from below) is

$$T_l = \frac{4N}{N^2+1}\exp(2\psi_l) \ . \tag{5.9}$$

From where does this particular multiplicative factor come? The transmissivity T for a step-barrier of heights V_1 and $-V_0$ with barrier thickness $\bar{a}$ (see Fig. 13) is obtained in the usual manner by matching the wavefunction and its first derivative, and is found to be [65] (recall that, as defined above, $V_0 < 0$ for a barrier):

$$T = 4\frac{K}{k}\left[\left(1+\frac{K}{k}\right)^2 \cosh^2\kappa\bar{a} + \left(\frac{\kappa}{k} - \frac{K}{\kappa}\right)^2 \sinh^2\kappa\bar{a}\right]^{-1} \ , \tag{5.10}$$

where

$$\frac{\hbar^2k^2}{2m} = E + V_0, \quad \frac{\hbar^2K^2}{2m} = E \quad \text{and} \quad \frac{\hbar^2\kappa^2}{2m} = V_1 - E \ .$$

The centrifugal barrier is, to a first approximation, a potential of this type; specifically it is extremely narrow because the high l-value imposes a steep decline beyond $r=a$. This of course means that $\bar{a} \ll 1$ and hence $\sinh^2\kappa\bar{a} \ll 1$, $\cosh^2\kappa\bar{a} \approx 1$ and under these circumstances

$$T = \frac{4N}{(1+N)^2} \ , \tag{5.11}$$

where

$$N = \frac{k}{K} = \left(1+\frac{V_0}{E}\right)^{1/2} \tag{5.12}$$

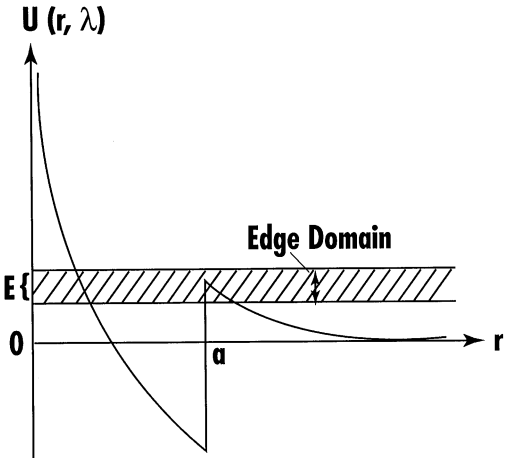

Fig. 13. The location of the "edge domain" (rays passing close to the "edge" of the sphere), defined by $\beta - c\beta^{1/3} = l_- \lesssim l \lesssim l_+ = \beta + c\beta^{1/3}$ for the lth partial wave. The constant c is of order one. The potential $U(r,\lambda)$ is the sum of the centrifugal barrier term and the square-well potential (see Eqs. (5.6), (A.5) and [47]). The angular momentum $\lambda = l + \frac{1}{2}$. In the edge domain the energy associated with these λ is, as shown, within a small neighborhood of the top of the barrier.

(There appears to be a small typographical error in [94], perpetuated in [6], assuming this potential is the basis for the statement in the former paper.) In practice, of course, the centrifugal "spike" barrier varies with energy, so this expression for T has to be modified by the "tunneling factor", defined below, which takes this into account. At the other extreme, for a thick barrier, we find that the transmissivity is dominated by the exponential term

$$\exp(-2\kappa\bar{a}) ,$$

which rapidly approaches zero as $\bar{a}$ increases.

The expression for T_l exploits the optical/quantum analogy: the refractive index N being the ratio k/K as noted above. The phase integral ψ_l is defined in terms of the turning points at $r = a$ and $r = r_1$, i.e. for which $\kappa(a) = 0 = \kappa(r_1)$. Thus

$$\psi_l = -\int_a^{r_1} |\kappa(r)|\, \mathrm{d}r \tag{5.13}$$

from which it follows that, in the interval (a, r_1),

$$\kappa^2 = k^2 - \frac{(l+\frac{1}{2})^2}{r^2} < 0 . \tag{5.14}$$

In terms of the variable $x = kr$ the phase integral can be written as

$$\psi_l = -\int_\beta^{l+1/2} \left\{ \left(l+\frac{1}{2}\right)^2 - x^2 \right\}^{1/2} \frac{\mathrm{d}x}{x} . \tag{5.15}$$

In particular, near the top of the barrier this integral may be approximated straightforwardly using, for example, the Trapezoidal Rule. For just two trapezoids

$$\int_a^b f(x)\,\mathrm{d}x \approx \frac{b-a}{4}\left[f(\beta)+2f\left(\beta+l+\frac{1}{2}\right)+f\left(l+\frac{1}{2}\right)\right]\,. \tag{5.16}$$

After some algebra this yields the expression

$$\psi_l \approx -\frac{3\sqrt{2}}{4}\left[\left(l+\frac{1}{2}-\beta\right)\Big/\beta^{1/3}\right]^{3/2}\,. \tag{5.17}$$

(In [6] the numerical factor is given as $-2\sqrt{2}/3$, the reciprocal of the value above: it is not clear if this is merely a typographical error or due to a different method of approximating the integral; in either case it is close to unity and does not alter the conclusions that follow.) From Eq. (5.9) it follows that the transmissivity for impact parameters b_l exceeding a is significant only when $\psi_l \approx O(1)$, i.e. effectively in the range $\beta < l < l_+$ (see Fig. 13).

5.2. The poles of the scattering matrix

The dimensionless scattering amplitude $f(k,\theta)$ is expressed as a partial-wave expansion

$$f(k,\theta) = (\mathrm{i}ka)^{-1}\sum_{l=0}^{\infty}\left(l+\frac{1}{2}\right)[S_l(k)-1]P_l(\cos\theta)\,, \tag{5.18}$$

S_l being the scattering matrix element for a given l and P_l being a Legendre polynomial of degree l. In [90] the expression for S_l has been derived for a (spherical) square well/barrier: for the former this is

$$S_l = -\frac{h_l^{(2)}(\beta)}{h_l^{(1)}(\beta)}\left\{\frac{\ln' h_l^{(2)}(\beta) - N\ln' j_l(\alpha)}{\ln' h_l^{(1)}(\beta) - N\ln' j_l(\alpha)}\right\}\,, \tag{5.19}$$

where using Nussenzveig's notation, $\ln'$ represents the logarithmic derivative operator, j_l and h_l are spherical Bessel and Hankel functions respectively; $\beta = ka$, as noted earlier, is the dimensionless external wavenumber, and $\alpha = N\beta$ is the corresponding *internal* wavenumber [5].

Poisson's sum formula in the form (see Appendix C)

$$\sum_{l=0}^{\infty} f\left(l+\frac{1}{2}\right) = \sum_{m=-\infty}^{\infty}(-1)^m\int_0^{\infty} f(\lambda)\mathrm{e}^{2\mathrm{i}m\pi\lambda}\,\mathrm{d}\lambda \tag{5.20}$$

is next used to rewrite (5.18) as

$$f(\beta,\theta) = \frac{\mathrm{i}}{\beta}\sum_{m=-\infty}^{\infty}(-1)^m\int_0^{\infty}[1-S(\lambda,\beta)]P_{\lambda-1/2}(\cos\theta)\mathrm{e}^{2\mathrm{i}m\pi\lambda}\lambda\,\mathrm{d}\lambda\,, \tag{5.21}$$

where, in terms now of *cylindrical* Bessel and Hankel functions [10],

$$S(\lambda,\beta) = -\frac{H_\lambda^{(2)}(\beta)}{H_\lambda^{(2)}(\beta)}\left\{\frac{\ln' H_\lambda^{(2)}(\beta) - N\ln' J_\lambda(\alpha)}{\ln' H_\lambda^{(1)}(\beta) - N\ln' J_\lambda(\alpha)}\right\}\,. \tag{5.22}$$

For fixed β, $S(\lambda,\beta)$ is a meromorphic function of the complex variable λ, and in particular in what follows, it is the *poles* of this function that are of interest. They are defined by the condition

$$\ln' H_\lambda^{(1)}(\beta) = N \ln' J_\lambda(\alpha)\ . \tag{5.23}$$

In view of relationship (5.5) between the refractive index N and the square well depth $-V_0$, these zeros may be identified with the Regge poles for a square potential well ($N > 1$) or barrier ($N < 1$). Several studies have been made of such poles for these potentials [96–100], both for $N > 1$ and $N < 1$. Nussenzveig [90] carried out a detailed investigation of the distribution of such poles, particularly for the case of $\beta \gg 1$; some of this material is relevant to the rainbow scattering problem so a brief summary of the pertinent details will be provided here for $N > 1$, but as is noted in [5], the physical interpretation is especially simple for $N \gg 1$. This implies that $\alpha \gg \beta \gg 1$, and renders it appropriate to use the asymptotic expansions for the Bessel and Hankel functions. Physically, this corresponds to an optically dense material or an extremely deep potential well (with, therefore, many energy levels). Mathematical details are provided in [98,99].

Under these approximations the complex λ-plane is subdivided into seven regions, though the Regge poles themselves can be associated with (i) broad resonances, (ii) narrow resonances and (iii) surface waves. Firstly, there exists a series of poles very close to the real axis; optically they correspond to so-called free modes of vibration of a dielectric sphere. The associated high internal reflectivity (since N is large) and high centrifugal barrier (since l is large) guarantee long lifetimes: the resonance appears when the pole is in close proximity to a physical value of λ, i.e. $\lambda = l + \frac{1}{2}$. Quantum mechanically, a deep well surrounded by a high barrier gives rise to sharp resonances. All these poles have their counterparts in the β- (or k-) plane of course, from which the penetration factors $\exp(2\psi_l)$ and resonance widths Γ_n can be established.

Another class of poles exists, this time not so close to the real λ-axis, but with almost constant imaginary part. These poles are associated with broad resonances above the top of the centrifugal barrier; again in the β-plane the location of the corresponding poles provides information on the penetration factors and resonance widths. A third set of poles, this time near $\lambda = \beta$, have an extremely important physical interpretation, and to clarify this, we write down the approximation for these, namely

$$\lambda_n \approx \beta + \frac{x_n e^{i\pi/3}}{(2/\beta)^{1/3}} + \frac{i}{(N^2-1)^{1/2}}\ , \tag{5.24}$$

where $-x_n$ is the nth zero of the Airy function $\mathrm{Ai}(z)$, i.e. $\mathrm{Ai}(-x_n) = 0$. The first and second terms occur in the expression for *Regge poles for an impenetrable sphere* [10,30], so the poles are related to surface waves. The second term contains a damping contribution which depends on the radius of curvature, and represents damping due to propagation along a curved surface. The final term represents damping due to refraction of the surface waves into the sphere, and by virtue of the temporary assumption that $N \gg 1$, this term is small, so the surface wave damping is determined primarily by geometry rather than refraction. For even larger values of $|\lambda|$ there are poles which continue to be associated with surface waves by virtue of their similarity to those

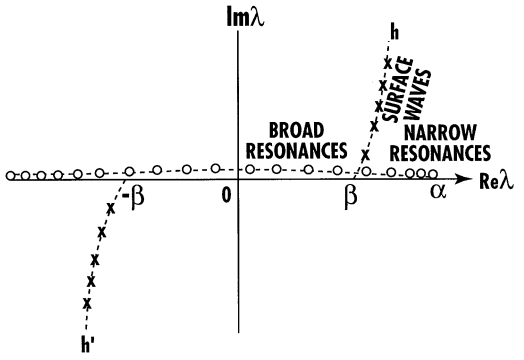

Fig. 14. The Regge poles of the scattering function $S(\lambda,\beta)$ for $\alpha = N\beta \gg 1$ and $\beta \gg 1$ illustrated schematically. The broad and narrow resonances are referred to as Class I poles in [5], and the surface waves are Class II poles. (Redrawn from [5].)

arising in the impenetrable sphere problem. Schematically these are shown in Fig. 14; Nussenzveig points out that they fall into two well-differentiated classes: those located near the real λ-axis (broad and narrow resonances), and those which lie on curves almost orthogonal to the first class in quadrants I and III (surface waves). These poles are not surprisingly referred to as *Class I* and *Class II poles*. In [99] Class I poles are called "physical" because of their correspondence, for certain values of energy, with the stationary states of a potential well. They behave somewhat similarly to Regge trajectories for Yukawa-type potentials [100]. For very deep wells those poles with large positive real parts move along the real axis at negative energies (giving rise to bound states) and move off the axis into the first quadrant, at positive energies (giving rise to resonances). At finite energy, the number of Class I poles in the right half-plane is bounded above. Physically, they are associated with the potential interior, i.e. for $r < a$, which is why they resemble Yukawa-type Regge trajectories.

Class II poles are, by contrast, referred to as "unphysical" in [99]. Even at finite energies the infinite number of them in the first quadrant have unbounded real parts; as $\beta \to 0$ they all move towards the origin. The trajectories of both classes of poles are briefly described in [99]. Physically, as we have seen, they are associated with surface waves, and are therefore almost completely determined by the geometrical shape of the scatterer. These contrasting pole behaviors provide a clue as to why the scattering amplitudes for finite-ranged potentials are so very different in analytic behavior from those potentials with tails extending to infinity (e.g. Yukawa-type potentials, or integral superpositions thereof, of the form $r^{-1}\exp(-\alpha r)$). Physically, it seems reasonable to suppose that by "cutting off" such a potential sufficiently far from the origin, the properties so obtained will differ insignificantly (both mathematically and physically) from those without compact support (i.e. infinite range potentials). However, the above discussion indicates that such a truncation induces the presence of surface waves: the presence now of an *infinite* set of Class II poles in addition to the finite set of Class I poles gives rise to an essential singularity (at finite energy) at infinity in the momentum transfer plane, so that the usual (Mandelstam) representation [92] is no longer valid. Even a sufficiently rapid exponential decrease in the potential is unable to mimic the cut-off potential (i.e. one

with compact support); the phrase 'sufficiently rapid' implies a length scale much shorter than the wavelength under consideration, and this is not true at 'sufficiently high' energies, whereas cut-off potentials can support surface waves at arbitrarily high energy.

5.3. The Debye expansion

The Watson transformation is applicable to the problem of scattering by an impenetrable sphere; this is discussed in detail elsewhere [30]; see also [10] and Appendix C. However, in the case of a transparent sphere, a large number of Regge poles lie in the neighborhood of the real axis, so if expression (5.21) for $f(\beta,\theta)$ were reduced to rapidly convergent contour integrals plus a series of residues from the Regge poles, such a series would of necessity converge very slowly (physically, the difference between the two cases arises because wave penetration occurs in the transparent sphere). The number of such poles is of the order of $(N-1)\beta$, as may be seen from Eq. (2.37) in [6], which means that the minimum number of residue terms required is at least of the same order as the number of terms needed in the original partial-wave series (a large number of such waves can be near resonance at high frequencies). In [33], the contour integrals were evaluated by the saddle point method, the contributions from which correspond to those arising from geometrical optics, at least to a first approximation. This is perfectly acceptable for the impenetrable sphere, for which only direct reflection occurs, but in the present case an incident ray is partially reflected and partially transmitted, the latter in turn being partially (internally) reflected and partially transmitted, and so on. Thus an infinite series of multiple internal reflections (or interactions with the surface) corresponds to an infinite number of saddle points, which must be appropriately superimposed to enable the external solution to be constructed. The Debye series is just this: *a representation of the scattering problem in terms of surface interactions*; Debye adopted this method for a circular cylinder [101], and later van der Pol and Bremmer applied it to the sphere [30,31]. This is done by means of a related but fictitious problem; i.e. by regarding the surface $r=a$ as an interface between two unbounded homogeneous media and matching the solutions to the (radial) equation to determine the internal and external spherical reflection and transmission coefficients. The interaction of an incoming spherical wave with the sphere is then described in terms of a sequence of surface interactions (partial reflections and transmissions) with interior propagation occurring in an alternate fashion. In particular, the external and internal spherical reflection coefficients R_{22} and R_{11} are given by Eqs. (5.25) and (5.26) for the scalar scattering problem (for the full electromagnetic problem there is a corresponding coefficient for each polarization; see [10,102] for details):

$$R_{22}(\lambda,\beta) = -\left(\frac{\ln' H_\lambda^{(2)}(\beta) - N \ln' H_\lambda^{(2)}(\alpha)}{\ln' H_\lambda^{(1)}(\beta) - N \ln' H_\lambda^{(2)}(\alpha)}\right) \tag{5.25}$$

and

$$R_{11}(\lambda,\beta) = -\left(\frac{\ln' H_\lambda^{(1)}(\beta) - N \ln' H_\lambda^{(1)}(\alpha)}{\ln' H_\lambda^{(1)}(\beta) - N \ln' H_\lambda^{(2)}(\alpha)}\right) . \tag{5.26}$$

From conservation of energy (or probability in the quantum–mechanical context) the following relationships hold:

$$|R_{22}(\lambda,\beta)|^2+\left|\frac{H_\lambda^{(2)}(\beta)}{H_\lambda^{(2)}(\alpha)}T_{21}(\lambda,\beta)\right|^2=1 \tag{5.27}$$

and

$$|R_{11}(\lambda,\beta)|^2+\left|\frac{H_\lambda^{(1)}(\alpha)}{H_\lambda^{(1)}(\beta)}T_{12}(\lambda,\beta)\right|^2=1\ , \tag{5.28}$$

where T_{ij} is the spherical transmission coefficient from medium i to medium j ($i,j=1,2$; the interior and exterior of the sphere are denoted by indices 1 and 2, respectively). In the limit as the radius a of the sphere tends to infinity, the above four coefficients reduce to the standard Fresnel coefficients for reflection and transmission for a plane interface at normal incidence, i.e.

$$R_{11}\to\frac{N-1}{N+1},\qquad T_{12}\to\frac{2N}{N+1} \tag{5.29}$$

and $R_{22}=-R_{11}$, $T_{21}=T_{12}$.

The scattering function $S(\lambda,\beta)$ (Eq. (5.22)) is to be expressed in terms of surface interactions; in order to accomplish this the external reflection coefficient R_{22} is subtracted from S; after some rearrangement this can be written in terms of the quantity.

$$\rho(\lambda,\beta)=\frac{H_\lambda^{(1)}(\alpha)}{H_\lambda^{(2)}(\alpha)}R_{11}(\lambda,\beta) \tag{5.30}$$

as

$$S(\lambda,\beta)=\frac{H_\lambda^{(2)}(\beta)}{H_\lambda^{(1)}(\beta)}\left(R_{22}(\lambda,\beta)+T_{21}(\lambda,\beta)T_{12}(\lambda,\beta)\frac{H_\lambda^{(1)}(\alpha)}{H_\lambda^{(2)}(\alpha)}\sum_{p=1}^{\infty}[\rho(\lambda,\beta)]^{p-1}\right)\ . \tag{5.31}$$

This is the celebrated *Debye expansion*, arrived at by expanding the expression $[1-\rho(\lambda,\beta)]^{-1}$ as an infinite geometric series. The first term inside the parentheses represents direct reflection from the surface; the pth term represents transmission into the sphere (from the term T_{21}) subsequently "bouncing" back and forth between $r=a$ and 0 a total of p times (from the ratio of Hankel functions preceding the summation), with $p-1$ internal reflections at the surface (via the R_{11} term in ρ; the case $p=1$ corresponds to direct transmission). The final factor in the second term, T_{12}, corresponds to transmission to the outside medium. The multiplicative factor outside the parentheses is phase term due to the surface interaction taking place at $r=a$ rather than $r=0$. The origin behaves as a perfectly reflecting "boundary" because of the regularity of the wavefunction there. In summary, therefore, *the pth term of the Debye expansion represents the effect of* $p+1$ *surface interactions.*

Regarding convergence of the expansion, it follows from Eqs. (5.28) and (5.30) for real λ that

$$|\rho(\lambda,\beta)|=|R_{11}(\lambda,\beta)|<1\ .$$

Nussenzveig [6] also shows that

$$\lim_{\lambda \to \pm\infty} |\rho(\lambda, \beta)| = 1 \ , \tag{5.32}$$

so that any improper integrals for $\lambda \in (0, \infty)$ must be interpreted in the limiting sense of $\lambda \in (0, \varpi), \varpi \to \infty$. In this spirit we may write

$$f(\beta, \theta) = f_0(\beta, \theta) + \sum_{p=1}^{\infty} f_p(\beta, \theta) \ , \tag{5.33}$$

where

$$f_0(\beta, \theta) = \frac{\mathrm{i}}{\beta} \sum_{m=-\infty}^{\infty} (-1)^m \int_0^{\infty} \left[1 - \frac{H_\lambda^{(2)}(\beta)}{H_\lambda^{(1)}(\beta)} R_{22} \right] P_{\lambda - 1/2}(\cos\theta) \exp(2\mathrm{i}m\pi\lambda) \lambda \, \mathrm{d}\lambda \tag{5.34}$$

and

$$f_p(\beta, \theta) = -\frac{\mathrm{i}}{\beta} \sum_{m=-\infty}^{\infty} (-1)^m \int_0^{\infty} U(\lambda, \beta) [\rho(\lambda, \beta)]^{p-1} P_{\lambda - 1/2}(\cos\theta) \exp(2\mathrm{i}m\pi\lambda) \lambda \, \mathrm{d}\lambda \tag{5.35}$$

for $p \geqslant 1$. In this last expression

$$U(\lambda, \beta) = T_{21}(\lambda, \beta) \frac{H_\lambda^{(1)}(\alpha)}{H_\lambda^{(2)}(\alpha)} \frac{H_\lambda^{(2)}(\beta)}{H_\lambda^{(1)}(\beta)} T_{12}(\lambda, \beta) = U(-\lambda, \beta) \ . \tag{5.36}$$

In practice, contributions to the integrals in (5.33) for large values of λ become vanishingly small, so the effect of condition (5.32) is negligible. Alternative but equivalent representations to $f(\beta, \theta)$ are also possible (see [5] for details).

While convergence of the Debye expansion is a fundamental mathematical property, once established it is necessary to ask the practical question: how *rapid* is the convergence? The earlier expression (5.21) for $f(\beta, \theta)$ involves a slowly converging residue series because of the existence of many Regge poles close to the real axis. What happens when the modified Watson transformation is applied now, i.e. how are the poles for each term distributed in the complex-λ plane? A distinct but related question concerns how fast the Debye expansion itself converges.

The poles for the Debye expansion differ from the Regge poles associated with expression (5.23); they are defined by

$$\ln' H_\lambda^{(1)}(\beta) = N \ln' H_\lambda^{(2)}(\alpha) \tag{5.37}$$

corresponding to the replacement of standing waves (Regge poles) by travelling waves (Debye interactions). Interestingly, the locations of the poles are the same for each term, but their *order* varies: they are of order $p + 1$ for the pth term ($p = 0, 1, 2, \ldots$). There are now two classes of poles to consider, each being symmetrically located with respect to the origin, denoted in [5] by λ_n and λ_n' respectively. In the right half-plane, for $N|\beta - 1| \gg \beta^{1/3}$, using earlier notation

$$\lambda_n \approx \beta + \mathrm{e}^{\mathrm{i}\pi/3}(x_n/\gamma) - \mathrm{i}/M \tag{5.38}$$

for $N > 1$, (the only situation which concerns us in this paper). Similarly,

$$\lambda'_n \approx \alpha + e^{-i\pi/3} N^{1/3}(x_n/\gamma) + N/M \tag{5.39}$$

for $N > 1$.

The set $\{\lambda_n\}$ differs little from that corresponding to the impenetrable sphere problem, so it is closely associated with surface waves. This is actually the case for the set $\{\lambda'_n\}$ as well: the λ'_n lie in the fourth quadrant, and arise from the Debye expansion; however, for $N > 1$ their contributions are negligible (which is not the case for $N < 1$) and so they will not be discussed further here.

In general, when the modified Watson transformation is applied to each term in the Debye expansion, the dominant contributions to the asymptotic behavior of each term will be from saddle points and residue series, the former corresponding to rays in geometrical optics. Thus for each term there is a finite number of saddle points, though this number does increase for each successive term. The residue-series terms converge rapidly (the imaginary part of λ_n increases rapidly with n), and so overall a rapidly convergent asymptotic expansion for each term of the Debye expansion can be obtained.

Returning to the question of convergence of the Debye series itself, the saddle-point contributions converge at a rate determined by the damping produced at each internal reflection. This in turn is determined by the Fresnel reflection coefficient at the interface, and that depends on the impact parameter and the value of N. For most directions (excluding the cases $N \ll 1$ and $N \gg 1$) this coefficient is small and as a consequence the Debye series converges quite rapidly. The convergence rate will be enhanced if the scattering sphere is not transparent, i.e. if absorption occurs in the interior; mathematically this can be accommodated by allowing N to be a complex number.

Van de Hulst [7] has estimated that for water droplets ($N \approx 1.33$ for orange light) more than 98.5% of the total intensity of incident radiation is accounted for by the rays $1', 2'$ and $3'$ in Fig. 1(c); these correspond to the first three terms in the Debye expansion. Thus the remainder, less than 1.5%, must be distributed among the higher-order terms and the contributions from the residue series. This is not to say, however, that they do not contribute significantly; they can be concentrated within narrow angular domains about certain directions (see Section 5.4 below). Indeed, the phenomenon of the "glory" (see Section 3 and [47]) is a consequence of these residue-series contributions dominating the geometrical ray contributions in a particular angular region. These contributions in general converge considerably more slowly than their saddle-point counterparts. This can be appreciated on physical grounds by considering what happens to the incident ray as the impact parameter increases. Under these circumstances the reflection coefficient increases towards unity (total reflection) as glancing incidence (for $N > 1$) is approached. At glancing incidence these rays are of course totally reflected within the approximation of geometrical optics, but these are precisely the limiting rays responsible for the *excitation of surface waves.* The consequence of high reflectivity inevitably implies slow convergence of these surface-wave contributions.

In [5], Nussenzveig examined in detail the first *two* terms of the Debye expansion (and in [6] the third term and effects associated with higher-order terms were discussed in connection with both the optical rainbow and glory). The first two terms are associated respectively with direct

reflection from the surface of the scatterer and direct transmission (with no internal reflection), consistent with the physical interpretation of the Debye series. The asymptotic expansions were carried out up to (but not including) correction terms of order $(ka)^{-2}$. The behavior of the first term is similar to that found for the impenetrable sphere problem, exhibiting (for $N > 1$) a forward diffraction peak, a geometrical reflection (or lit) region, and a transition region in which the scattering amplitude is expressible in terms of generalizations of *Fock functions* (see Eq. (4.67) in [5] and Appendix 2), the latter being defined in [33] as

$$\mathscr{F}(\tau) = \int_0^\infty \frac{e^{i\tau x}}{\mathrm{Ai}(xe^{i\pi/3})}\,dx + \int_0^\infty \frac{e^{-i\tau x}}{\mathrm{Ai}(xe^{-i\pi/3})}\,dx\ . \tag{5.40}$$

The second term gives rise to a lit region, a shadow region and again, a Fock-type transition region. By a process termed *critical refraction*, surface waves are able to take "shortcuts" across the sphere. The application of the modified Watson transform to the third term in the Debye expansion of the scattering amplitude is examined by Nussenzveig in [6], along with the effect of higher-order terms. *It is this term which is associated with the phenomena of the rainbow and the glory.*

The theory for the first two terms is uniformly valid for all $N > 1$; this is not the case for the third term however. It transpires that there are five different ranges of the refractive index to be considered: each of them requires a separate treatment [6,34]. These subdivisions can be found directly from geometrical optics considerations; corresponding to different values of N there are different angular regions characterized by the number of rays—0, 1, 2 or 3—in a particular direction within that region. The only range of significance here is $1 < N < \sqrt{2}$; the refractive index of water for the optical spectrum falls within this range, and from now on all statements made pertain to it. From a consideration of geometrical optics, there are three different angular regions to examine: (i) a 0-ray (shadow) region located near the forward region, $0 \leqslant \theta < \theta_R$; (ii) a 1-ray region near the backward direction, $\theta_L < \theta \leqslant \pi$, where as shown below, $\theta_L = 4\arccos(1/N)$ and (iii) a 2-ray region between them, $\theta_R < \theta < \theta_L$. However, the presence of (non-geometrical optics) transition regions leads to a total of *six* regions to examine for $\theta \in (0, \pi)$; the three additional regions arise because of diffraction effects at the boundaries between the above three regions. The transition region near $\theta = \pi$ corresponds to the "*glory*" region, and the 2-ray/0-ray (light/shadow or caustic) transition region corresponds to the "*rainbow*" region. There is a penumbra region around the 1-ray/2-ray boundary.

The so-called "normal" regions can be described as follows. The 1-ray region contains the single geometrical optics contribution *plus* surface waves excited at the 2-ray/1-ray shadow boundary; these are associated with diffracted rays that take two "shortcuts" across the scattering sphere. This 2-ray/1-ray transition region corresponds to one of the normal "Fock-type" regions. The 2-ray region contains two real saddle points (one for each ray direction in this region) which become complex in the 0-ray (shadow) region; here the scattering amplitude cannot be reduced to a pure residue series.

The rainbow light/shadow transition region is associated physically with the confluence of a pair of geometrical rays and their transformation into "complex rays"; mathematically this corresponds to a pair of real saddle points merging into a complex saddle point. Then the problem is to find the asymptotic expansion of an integral having two saddle points that move towards or away from each other. The generalization of the standard saddle-point technique to

include such problems was made by Chester et al. [72] and using their method, Nussenzveig was able to find a uniform asymptotic expansion of the scattering amplitude which was valid throughout the rainbow region, and which matched smoothly onto results for neighboring regions (see Appendix D). The lowest order approximation in this expansion turns out to be the celebrated *Airy approximation* which, despite several attempts to improve upon it, was the best approximate treatment prior to the analyses of Nussenzveig and coworkers. However, Airy's theory had a limited range of applicability as a result of its underlying assumptions; by contrast the uniform expansion is valid over a much larger range. This new approach also enables a quantitative theory of the glory—a strong enhancement of scattering in the near-backward direction, as noted in Section 3—to be developed. The intensity predicted by geometrical optics is not nearly sufficient to account for this effect. In 1947, Van de Hulst conjectured that the glory is due to surface waves that make two shortcuts across the sphere ([50], see also [7]), but no quantitative analysis had been provided. Nussenzveig used the modified Watson transform to advantage in the backward scattering direction, and evaluated the residue-series contributions there, showing that they are indeed capable of accounting for the enhancement in the backward intensity, thus confirming Van de Hulst's conjecture (as far as the third term of the Debye series is concerned; but many other terms contribute to the glory, as summarized in Section 5.7 below). The enhancement arises from the focusing of diffracted rays on the axis which compensates for the exponential damping over the surface of the sphere, and illustrates the phenomenon of "Regge-pole dominance" of the scattering amplitude.

Nevertheless, the "third term" theory, highly successful as it is, is not able to satisfactorily account for the backward-scattered intensity *as a function of* β; it transpires that contributions from higher-order terms in the Debye expansion are necessary to explain the results obtained by numerical summation of the partial-wave series (and note that the *fourth* term in the Debye expansion corresponds to the *secondary* rainbow). As noted in Section 3, Bryant and Cox [54] found a very complicated fine structure consisting of a quasiperiodic pattern with prominent but irregular peaks, and it is this type of behavior that requires higher-order terms to be invoked. Despite some mathematically technical difficulties, it was shown that the quasiperiodic fine structure could be accounted for in this way, and also resonance effects predicted. Furthermore, the same effects can be shown to be responsible for the "ripple" in the total cross section for $N > 1$ (see [10,88,89] and Section 5.6 for more details).

Explicitly, the third term in the Debye expansion is, from (5.35),

$$f_2(\beta,\theta) = -\frac{\mathrm{i}}{\beta}\sum_{m=-\infty}^{\infty}(-1)^m\int_0^{\infty}U(\lambda,\beta)\rho(\lambda,\beta)P_{\lambda-1/2}(\cos\theta)\exp(2\mathrm{i}m\pi\lambda)\lambda\,\mathrm{d}\lambda\,. \tag{5.41}$$

After some rearrangement, and a shift in the path of integration to the above real axis, this can be written as

$$f_2(\beta,\theta) = -\frac{\mathrm{i}}{2\beta}\int_{-\infty-\mathrm{i}\epsilon}^{\infty+\mathrm{i}\epsilon}U(\lambda,\beta)\rho(\lambda,\beta)P_{\lambda-1/2}(\cos\theta)\exp(\mathrm{i}\pi\lambda)\frac{\lambda}{\cos\pi\lambda}\,\mathrm{d}\lambda\,, \tag{5.42}$$

where $\varepsilon > 0$. Furthermore, this integral can be decomposed into

$$f_2(\beta,\theta) = f_{2,0}^{+} + f_{2,r} = f_{2,0}^{-} - f_{2,r}\,, \tag{5.43}$$

where

$$f_{2,0}^{\pm}(\beta,\theta) = \pm\frac{\mathrm{i}}{\beta}\int_{-\infty\pm\mathrm{i}\epsilon}^{\infty\mp\mathrm{i}\epsilon}\rho U Q_{\lambda-1/2}^{(2)}(\cos\theta)\lambda\,\mathrm{d}\lambda \tag{5.44}$$

and

$$f_{2,r}(\beta,\theta) = \frac{1}{2\beta}\int_{-\infty-\mathrm{i}\epsilon}^{\infty+\mathrm{i}\epsilon}\rho U\exp(2\mathrm{i}\pi\lambda)\frac{\lambda\,\mathrm{d}\lambda}{\cos\pi\lambda}\,. \tag{5.45}$$

In expression (5.44), $Q_{\lambda-1/2}^{(2)}$ is a Legendre function of the second kind. Details of these manipulations can be found in [33,6]. It can also be shown from the nature of the integrand in (5.44) that there is always some neighborhood of the imaginary axis where that integrand diverges to infinity. Since this is true for any value of θ, it follows that there is no domain of θ-values in which $f_{2,0}^{\pm}$ (and hence $f_2(\beta,\theta)$) can be reduced to a pure residue series.

5.4. Geometrical optics régimes

As previously noted, the third term of the Debye expansion corresponds to rays that are transmitted out of the transparent sphere after one internal reflection. There are in fact three possible types of ray trajectories depending on the value of the refractive index N. The one of concern here is for N in the domain $(1,2)$; in fact it is the subdomain $(1,\sqrt{2})$ that will be examined below. The other domains are of course $(0,1)$ and $(2,\infty)$; again, further subdivisions occur, so that there are in total *five* subdomains:

$$(0,1),\quad (1,\sqrt{2}),\quad (\sqrt{2},[6\sqrt{3}-8]^{1/2}),\quad ([6\sqrt{3}-8]^{1/2},2)\quad\text{and}\quad (2,\infty)\,.$$

The mathematics based on ray theory has been discussed earlier so we merely summarize it here. In terms of the angles of incidence, refraction and deflection (or scattering), θ_i, θ_r and θ $(=i,r,$ and $D_k,$ respectively, in Section 1.3)

$$\sin\theta_i = N\sin\theta_r,\quad 0\leqslant\theta_i,\ \ \theta_r\leqslant\pi/2,\ \ 0\leqslant\theta\leqslant\pi$$

and

$$\theta = 2(\theta_i - 2\theta_r) + \pi\,. \tag{5.46}$$

For the deflection angle to be no greater than π radians it is necessary that $\phi = \theta_i - 2\theta_r \leqslant 0$; the limiting angle of incidence θ_i^a such that $\phi = 0$ satisfies the implicit condition

$$\theta_i^a = 2\arcsin\left(\frac{\sin\theta_i^a}{N}\right) \tag{5.47}$$

or explicitly

$$\cos\theta_i^a = \tfrac{1}{2}(N^2-2) \tag{5.48}$$

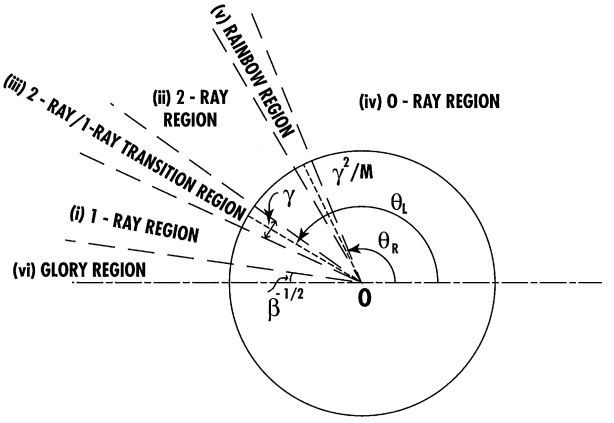

Fig. 15. A schematic representation of the six angular regions associated with both rainbow and glory scattering from a spherical water droplet. The glory, rainbow and 2-ray/1-ray transition region are greatly exaggerated for clarity. The angular size of these regions are noted (see text for details). This is valid for $1 < N < \sqrt{2}$. The rainbow angle is θ_R and θ_L is the 2-ray/1-ray boundary angle. (Redrawn from [6].)

(excluding the zero impact parameter case $\cos\theta_i^a = 1$). Since $\theta_i \in [0, \pi/2]$ it follows that such an angle θ_i^a only exists for $N \in [\sqrt{2}, 2]$. In terms of ϕ, it can be seen that

$$\frac{\mathrm{d}\phi}{\mathrm{d}N} = \frac{2\sin\theta_i/N^2}{\sqrt{1-(\sin\theta_i/N)^2}} > 0\ ,$$

i.e. ϕ is an increasing function of N. Since $\phi = 0$ for $N = \sqrt{2}$, it follows that the range of interest here corresponds to $\phi < 0$, i.e. $N \in (1, \sqrt{2})$. Note also that the limiting incident ray is

$$\theta_i = \frac{\pi}{2}, \quad \theta_r = \arcsin\frac{1}{N} = \theta_l\ , \tag{5.49}$$

the corresponding scattering angle is

$$\theta_L = 2\pi - 4\arcsin\frac{1}{N} = 4\arccos\frac{1}{N}\ . \tag{5.50}$$

Finally, as noted in Section 1.3, the *rainbow* angle (the minimum value of the scattering angle) is, in the present notation,

$$\theta_R = \pi + 2\theta_{i,R} - 4\arcsin\left(\frac{\sin\theta_{i,R}}{N}\right)$$

where

$$\theta_{i,R} = \arccos\left(\frac{N^2-1}{3}\right)^{1/2}\ .$$

The angle θ_L is a 1-ray/2-ray shadow boundary; the rainbow appears around the angle θ_R, which is a 2-ray/0-ray shadow boundary (see Fig. 15).

5.5. Saddle points

The location of the saddle points is a continuing theme throughout the papers [5,6,33] and subsequent work from the author (summarized in [10]); there are many details and subtleties that cannot be entered into in a review of this nature. Nevertheless, the basic features of the process can be described, because there are similarities between the methods employed for the impenetrable sphere problem and for each of the first three terms in the Debye expansion (and in principle for the higher-order terms). The method of steepest descent (or saddle-point method) is described in many texts on asymptotic methods (e.g. [103,104]; a brief but excellent summary may be found in [23]) so it will suffice here to note that the integrals of interest are all reducible to the following generic form (neglecting multiplicative factors)

$$\int \mathrm{C}(w_1, w_2, \beta, \theta) \exp[\mathrm{i}\beta\omega(w_1, w_2, \theta)]\, \mathrm{d}w_1 \ , \tag{5.51}$$

where $\lambda = \beta \sin w_1$ and $\lambda = \alpha \sin w_2$.

The derivation of the specific forms for $\mathrm{C}(w_1, w_2, \beta, \theta)$ and $\omega(w_1, w_2, \theta)$ involve the Debye asymptotic expansions for $H_\lambda^{(1,2)}(\beta)$ [33]; these expansions are valid in an oblong-shaped region in the λ-plane (see [105] for an explanation); this means that the method applied to integrals of type (5.51) above can locate both real and complex saddle points (if any exist) in this region. It transpires that

$$\frac{\partial\omega}{\partial w_1} = 2\cos w_1 \left[2w_2 - w_1 - \frac{1}{2}(\pi - \theta)\right] \ , \tag{5.52}$$

which yields the saddle points $w_1' = \theta_i$, $w_2' = \theta_r$ (on using the deflection angle formula (5.46)), and hence

$$\lambda' = \beta \sin\theta_i = \alpha \sin\theta_r \ . \tag{5.53}$$

In an appendix to [33], Nussenzveig presents the details calculating the real and complex roots of (5.52) and (5.53); this involves some very complicated algebraic manipulation so a summary will suffice here.

For $\theta \in (\theta_L, \pi)$, the 1-ray region in Fig. 15, there is only saddle point, and this is real. As θ decreases from π in this interval the saddle point moves from the origin along the positive real axis to the point z_L. At $\theta = \theta_L$ another saddle point appears at $\lambda'/\beta = 1$, and as θ continues to decrease from θ_L to θ_R, the two saddle points move toward each other (this is something of an oversimplification; refer to [6], Section 3 for details). They merge at the rainbow angle θ_R, so for θ in the interval (θ_R, θ_L), there are two real saddle points, corresponding to the 2-ray region in Fig. 15. The remaining interval is $\theta \in (0, \theta_R)$. Here the two saddle points become complex, leaving the real axis orthogonally, and tracing complex-conjugate trajectories (see Fig. 16). This is associated with the 0-ray region in Fig. 15. Note that these results are valid only for the domain illustrated in Fig. 15; the exterior of this domain corresponds to the geometrical shadow region; a different representation of integral (5.51) would be necessary to trace the behavior of the saddle points in this region. In fact, in this region the contribution to the scattering amplitude from the complex saddle points no longer dominates the other contributions, so such analysis is not necessary.

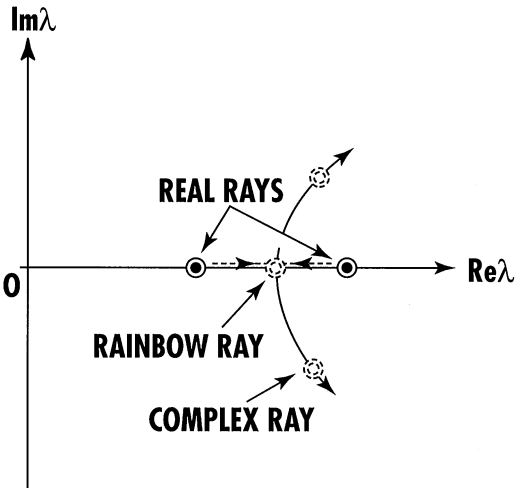

Fig. 16. The coalescence of two real saddle points in the complex λ-plane as the rainbow angle θ_R is approached from below (i.e. from the lit side). At θ_R the points collide and subsequently move away from each other along complex-conjugate directions as θ increases away from θ_R into the shadow region. It is the lower complex saddle point that contributes to the wave field in this region.

It has been noted already that there is a total of six angular regions in $(0,\pi)$ of interest in the full scattering problem for $1 < N < \sqrt{2}$; the regions $[0,\theta_R),(\theta_R,\theta_L)$ and $(\theta_L,\pi]$ discussed above, together with their transition regions and a transition region near the backward direction (the glory region). We defer discussion of this and the rainbow region and briefly discuss the nature of the remaining four so-called "normal" regions.

5.5.1. The 1-ray region

This is the region

$$\theta - \theta_L \gg \gamma = (2/\beta)^{1/3}, \qquad \pi - \theta \gg \beta^{-1/2} ,$$

where the latter restriction is to enable the asymptotic expansion for $Q^{(2)}_{\lambda-1/2}(\cos\theta)$ to be utilized in (5.44), leading to the integral representation (5.51). The path of steepest descent crosses the real axis at the (unique) saddle point at an angle of $\pi/4$, as it passes from the third quadrant through the fourth to the first quadrant. The third term in the Debye expansion (referred to here as f_2 for brevity) is dominated by the WKB approximation, representing the contribution from one geometrical ray (transmitted after one internal reflection; see Fig. 1(a)).

5.5.2. The 2-ray region

This is the region

$$\theta_L - \theta \gg \gamma, \quad \theta - \theta_R \gg \gamma^2/M ,$$

where the second restriction is related to the condition for validity of the Debye asymptotic expansion. Essentially, it requires that the "range" of the (now two) saddle points [103] be much smaller than their separation distance (see Appendix D). In this region the path of steepest

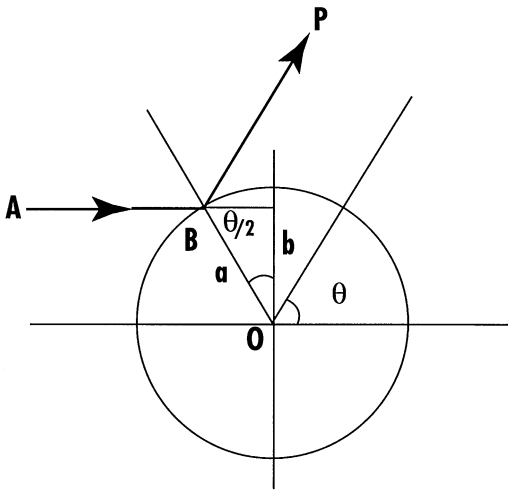

Fig. 17. An incident ray AB is reflected in the direction θ to P; this provides a physical interpretation of the left saddle point associated with equation $\tilde{\lambda} = \beta\cos(\theta/2)$; b is the impact parameter of the ray AB reflected to P. (See [33,34].)

descent must cross the new saddle point at an angle of $-\pi/4$, so a new path of integration is necessary. Again, f_2 is dominated by the WKB approximation, which now represents contributions from the two saddle points; physically, this corresponds to the existence of two different impact parameters giving two parallel geometric rays ([34]; see also Fig. 17).

5.5.3. The 0-ray region

This is the region

$$\theta_R - \theta \gg \gamma^2/M$$

and here the saddle points become complex. In view of this, the path of integration must now be taken over the one which induces an exponential decreasing contribution in the shadow region; this turns out to be the lower saddle point. This may be interpreted as a "complex ray" [106]. In this shadow region, f_2 cannot be reduced to a pure residue series because of the complex saddle points. Towards the lower end of the above domain, the amplitude is dominated by the background integral, which in turn is dominated by the lower saddle point (see Fig. 16). Thus "complex rays" are born, corresponding to super-exponential damping in the shadow (but with a different factor from those associated with the poles λ_n) [39]. The damping factor is proportional to

$$\exp[-c(|\theta - \theta_R|/\gamma^2)^{3/2}]\,,$$

where γ is a positive constant. As one moves more deeply into the shadow from the penumbra region the residue contribution, being more weakly damped, eventually dominates the complex-ray contribution.

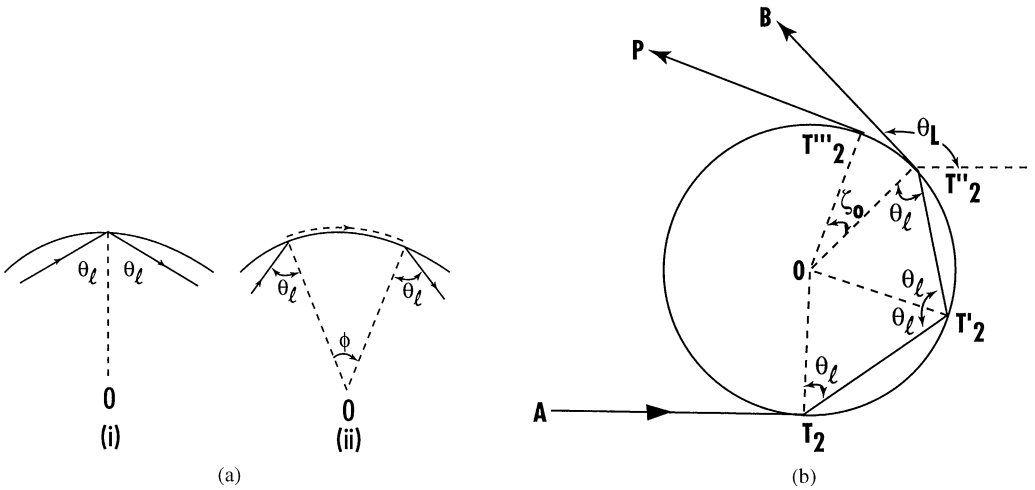

Fig. 18. (a) The two types of vertices associated with an internally incident ray at the critical angle $\theta_l = \arcsin(1/N)$. In (i) the surface interaction is just total internal reflection; in (ii) it is critical refraction to the outside, followed by a surface wave arc of angle ϕ (which may take any value) and then critical refraction back into the sphere. Combinations of these two types of vertices provide several classes of "diagrams" with $p-1$ vertices that contribute to a Debye term of order p. (Redrawn from [47].) (b) Diffracted rays $T_2''T_2'''P$ excited at the shadow boundary $T_2''B$. (Redrawn from [34].)

5.5.4. The 1-ray/2-ray transition region

This is the region

$$|\theta - \theta_L| \lesssim \gamma\,,$$

which is a normal (Fock-type) transition between the 1-ray and 2-ray regions. At such a shadow boundary, surface waves are excited (see Fig. 18). A tangentially incident ray undergoes critical refraction and one subsequent internal reflection before emerging tangentially to define the shadow boundary. The surface waves excited at this point of emergence correspond to residues at the poles λ_n and propagate into the shadow (1-ray) region. Nussenzveig points out that there are two types of such diffracted rays, corresponding to one or two surface-wave arcs (see Fig. 18) before reemergence as a surface wave. Both types of wave must of course be accounted for in any contribution summations [34].

Now we are in a position to discuss the two remaining and extremely physically significant regions:

5.5.5. The rainbow region (the 2-ray/0-ray transition region)

This is the region

$$|\theta - \theta_R| \lesssim \gamma^2/M\,.$$

The mathematical problem here involves the asymptotic evaluation of an integral arising in the 2-ray region, but in a domain where its two saddle points are very close to each other. The

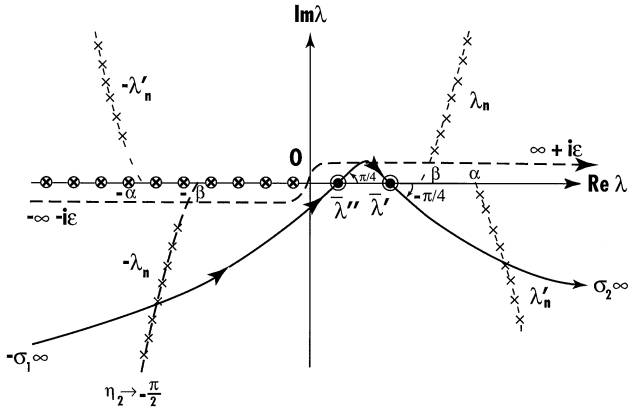

Fig. 19. The deformed contour of integration (solid line) in Eq. (5.44) for the 2-ray region, i.e. $\theta_R < \theta < \theta_L$. Plain crosses correspond to the poles of ρU, crossed circles to poles of $Q^{(2)}_{\lambda-1/2}$ and dots in circles to saddle points. The original contour is shown as a broken line; it is deformed into the steepest-descent path, giving rise to residues at the poles $-\lambda_n$ and λ'_n. See [6], Section 4 for details of the corresponding analysis for the rainbow region.

integral is the "geometrical optics" term

$$f_{2,g}(\beta,\theta) = -\frac{\mathrm{i}}{\beta}\int_{-\sigma_1\infty}^{\sigma_2\infty} \rho U Q^{(2)}_{\lambda-1/2}(\cos\theta)\lambda\,\mathrm{d}\lambda\ , \tag{5.54}$$

where the path of integration in (5.44) from $(-\infty - \mathrm{i}\varepsilon, \infty + \mathrm{i}\varepsilon)$ is deformed into the path of steepest descent (Fig. 19) denoted by $(-\sigma_1\infty, \sigma_2\infty)$. (This entails moving across the poles $-\lambda_n$ and λ'_n so the residue contributions also differ from those in the 1-ray region.)

The relevant theory for this asymptotic evaluation (briefly reviewed in Appendix D) has been developed by Chester et al. [72]. The above integral may be rewritten as

$$f_{2,g}(\beta,\theta) = 2\mathrm{e}^{-\mathrm{i}\pi/4} N \left(\frac{2\beta}{\pi\sin\theta}\right)^{1/2} F(\beta,\theta)\ , \tag{5.55}$$

where

$$F(\beta,\theta) = \int_{-\sigma_1'\infty}^{\sigma_2'\infty} g(w_1)\exp[2\beta f(w_1,\theta)]\,\mathrm{d}w_1 \tag{5.56}$$

and

$$f(w_1,\theta) = \mathrm{i}\left[2N\cos w_2 - \cos w_1 + \left(2w_2 - w_1 - \frac{\pi-\theta}{2}\right)\sin w_1\right] \tag{5.57}$$

and

$$g(w_1) = (\sin w_1)^{1/2}\cos^2 w_1 \cos w_2 \left[\frac{\cos w_1 - N\cos w_2}{(\cos w_1 + N\cos w_2)^3}\right][1 + \mathcal{O}(\beta^{-1})]\ . \tag{5.58}$$

The limits $-\sigma_1'\infty$ and $\sigma_2'\infty$ are the images of $-\sigma_1\infty$ and $\sigma_2\infty$ in the w_1-plane, respectively. The two saddle points are given by $w_1' = \theta_i'$ and θ_i''. In the present domain,

$$\theta = \theta_R + \varepsilon, \quad |\varepsilon| \ll 1\ .$$

If we take $\varepsilon > 0$ initially, then the saddle points are real. To illustrate the complexity of the calculations, note that after a great deal of analysis it can be shown that [6]

$$f_{2,g}(\beta,\theta) = 4e^{i\pi/4}N\left(\frac{\pi}{\sin\theta}\right)^{1/2}(2\beta)^{1/6}\exp[2\beta A(\epsilon)]\{p_0(\epsilon)\mathrm{Ai}[(2\beta)^{2/3}\zeta(\epsilon)] ,$$
$$-\frac{q_0(\epsilon)}{(2\beta)^{1/3}}\mathrm{Ai}'[(2\beta)^{2/3}\zeta(\epsilon)]\}[1+\mathcal{O}(\beta^{-1})] , \tag{5.59}$$

where

$$A(\varepsilon) = i[N(\cos\theta'_r + \cos\theta''_r) - \tfrac{1}{2}(\cos\theta'_i + \cos\theta''_i)] , \tag{5.60}$$

$$\tfrac{2}{3}[\zeta(\varepsilon)]^{3/2} = i[N(\cos\theta'_r - \cos\theta''_r) - \tfrac{1}{2}(\cos\theta'_i - \cos\theta''_i)] , \tag{5.61}$$

$$p_0, q_0 = \frac{e^{-i\pi/4}}{4N}\zeta^{\pm 1/4}\left[\left(\frac{\sin\theta''_i}{2\cos\theta''_i - N\cos\theta''_r}\right)^{1/2}(2N\cos\theta''_i\cos\theta''_r)^{3/2}\frac{(\cos\theta''_i - N\cos\theta''_r)}{(\cos\theta''_i + N\cos\theta''_r)^3}\right.$$
$$\left. \pm \left(\frac{\sin\theta'_i}{N\cos\theta'_r - 2\cos\theta'_i}\right)^{1/2}(2N\cos\theta'_i\cos\theta'_r)^{3/2}\frac{(\cos\theta'_i - N\cos\theta'_r)}{(\cos\theta'_i + N\cos\theta'_r)^3}\right][1+\mathcal{O}(\beta^{-1})] \tag{5.62}$$

and the $\pm$ signs correspond to p_0 and q_0, respectively. Also recall that θ'_1 and θ''_1 are the saddle points. The expressions for $A(\epsilon)$ and $\zeta(\epsilon)$ correspond to half the sum and half the difference of the optical paths through the sphere, respectively. Very close to the rainbow angle, i.e. for $|\epsilon| \ll 1$, the approximations for the above quantities may be substituted into the expression for $f_{2,g}(\beta,\theta)$ to yield a still very complicated expression (see [6] for details); the result is a good approximation for $|\epsilon| \ll \gamma$, containing as it does, the rainbow region. The dominant contribution to the third Debye term for the scattering amplitude $S_{j2}(\beta,\theta)$ turns out to be proportional to $\beta^{7/6}$. The above approximation can still be employed over part of the 2-ray region, where the Airy functions may be replaced by their asymptotic expansions for negative arguments, which we will not state here. The resulting expression is oscillatory (containing a sinusoidal term) and is associated with the interference between the two geometrical ray contributions, giving rise to the *supernumerary bows* referred to in the introduction. Other expressions, valid for still larger scattering angles in the 2-ray region, may be derived which coincide exactly with the results obtained by the saddle-point method, so the Chester–Friedman–Ursell method leads to a uniform asymptotic expansion, matching smoothly the result obtained by the saddle-point method in its domain of validity.

In the 0-ray region, within the domain $\gamma^2/M \ll |\epsilon| \ll \beta^{-1/2}$, $\epsilon < 0$ so the asymptotic expansions for Airy functions with positive arguments may now be used, resulting in an expression for the complex-ray contribution (described above) and giving rise to Alexander's dark band between the primary and secondary bows. Further away from the rainbow angle, the residue-series contributions dominate the amplitude in the deep shadow region.

Airy's approximation may be recovered from the $|\epsilon| \ll 1$ version of Eq. (5.59) for $f_{2,g}(\beta,\theta)$, as expected; this is

$$f_{2,g}^{(\mathrm{Airy})}(\beta,\theta) = -\frac{16}{27}\frac{(3)^{7/12}}{(2)^{1/3}}\mathrm{e}^{-\mathrm{i}\pi/4}\frac{N^2(N^2-1)^{1/4}}{(8+N^2)^{1/2}(4-N^2)^{2/3}}\beta^{1/6}$$
$$\times\exp\left\{\frac{\mathrm{i}\beta}{\sqrt{3}}[6(N^2-1)^{1/2}+(4-N^2)^{1/2}(\theta-\theta_R)]\right\}$$
$$\times\mathrm{Ai}\left[-\frac{(N^2-1)^{1/2}}{(4-N^2)^{1/6}}\left(\frac{2\beta}{3}\right)^{2/3}(\theta-\theta_R)\right]\ . \tag{5.63}$$

Van de Hulst [7] claims that this approximation is a useful quantitative theory only for $\beta > 5000$ (or drops of radius greater than about 0.5 mm) and $|\epsilon| < 0.5^\circ$; although Huygen's principle may still be applied for $\beta > 2000$, "a quantitative theory of the rainbow for the entire gap $30 < \beta < 2000$ is still lacking". The above contributions from Nussenzveig and coworkers bridge this gap *and more*; indeed, to quote from [10, Section 8.6],

> the CAM theory bridges the gap between short-wavelength and long-wavelength scattering, remaining applicable all the way down to $\beta \sim 1\ldots$ For $\beta > 100$, the accuracy of the CAM approximation is better than 1 ppm, i.e. it becomes, for all practical purposes, more accurate than the 'exact' partial-wave expansion!

(For further comments on accuracy see Section 5.7 below). For β of the order of a few hundred, within the domain $|\epsilon| \ll \gamma$, the corrections to Airy's theory can attain several percent; this increases with deviation from the rainbow angle.

5.6. The glory

The post-Airy study of both the rainbow and the glory may be divided into two phases, expressed succinctly as "$\lesssim$ CAM" and "$\gtrsim$ CAM". The dividing line for the purposes of this review occurs around the late 1970s/early 1980s, and although what follows below pertains more to the first phase, it contains within it much that is relevant to the second phase, since during the latter many of the implications of the CAM method for rainbows and glories were "fine-tuned" to a high degree. The details that follow are based predominantly on the analysis in [6], which provides a convenient overview of the results in the first phase of study. The 'second phase' is summarized in Section 5.7 below.

The angular region where the glory is observed is $\theta = \pi - \epsilon$, $0 \leqslant \epsilon \lesssim \beta^{-1/2}$, i.e. in the neighborhood of the backward region. It is a strong enhancement in this direction due to very small water droplets, with β value up to a few hundred ($\beta \sim 160$ corresponds to water droplets with average diameter ≈ 0.028 mm). The effect is concentrated within a very small solid angle around the 180° direction, corresponding to a narrow peak in the back-scattered intensity. This explains in part the unusual meteorological manifestations that have been observed (see the introduction for more details). It is *not* observed for larger water droplets ($\beta \sim 1000$). Other important features of the phenomenon are as follows:

If ε_i denotes the angular radius of the ith dark ring (as observed in yellow light) then the approximate bounds arise [6,7]:

$$0.35 \lesssim \epsilon_1/\epsilon_2 \lesssim 0.45, \quad 1.6 \lesssim \epsilon_3/\epsilon_2 \lesssim 1.7$$

and $\epsilon_1 = \mathcal{O}(\beta^{-1})$, which is a measure of the narrowness of the backward peak. The above ratio bounds are different from those associated with another diffraction effect: diffraction coronae, which correspond to the *forward* diffraction peak. The intensity of the rings in the glory decreases more slowly away from the center than it does in diffraction coronae. The radii and brightness of the rings may change with time, even during the course of a single observation. There are indications that the glory, like the rainbow, is strongly polarized.

To recapitulate some points made earlier in Section 3, Bryant and Cox [54] made some numerical calculations of the intensity at or near 180° for the *electromagnetic problem*; the intensity was computed as a function of β, at intervals of 0.005, near $\beta = 200$ and 500. The results showed a great deal of fine structure; in particular they found a rapidly varying quasiperiodic pattern superimposed on a more slowly varying background for the back-scattered intensity. The period $\delta\beta$ for these fluctuations was found to be in the approximate range $0.81 \lesssim \delta\beta \lesssim 0.82$. Even within a single period, they found irregular peak widths ranging from $\sim$0.01 to $\sim$0.1, with intensities changing by a factor of $\sim$100, resulting in large spikes superimposed on the background. They also found that the *total cross section* showed fluctuations similar to those of the background intensity, but greatly reduced in magnitude; at 90° intermediate-sized fluctuations were noticed but with periods twice those given above for $\delta\beta$.

The scalar theory developed by Nussenzveig [5,6] was in principle capable of explaining the above features, with the exception of polarization effects; however, the details of the intensity variations require that higher-order terms in the Debye expansion be taken into consideration. Van de Hulst [50] conjectured that surface waves are responsible for the glory: specifically diffracted rays excited at the 2-ray/1-ray shadow boundary which travel along a surface arc of about 14° (for water) before emerging in the backward direction, having taken two shortcuts across the sphere (see Fig. 18(c)). Using the formalism of the modified Watson transform $f_{2,g}(\beta,\theta)$ may be written as [6]

$$f_{2,g}(\beta,\theta) = \frac{\mathrm{i}}{\beta}\int_0^{\sigma_1\infty} \rho U P_{\lambda-1/2}(-\cos\theta)\mathrm{e}^{\mathrm{i}\pi\lambda}\tan(\pi\lambda)\lambda\,\mathrm{d}\lambda\,, \tag{5.64}$$

upon employing symmetry arguments. This background integral represents near-central rays which are reflected back after one internal transmission; the surface waves are "generated" by the residue-series contributions $f_{2,\mathrm{res}}$ at the poles λ_n. In [34] it is pointed out that the evaluation of $f_{2,\mathrm{res}}$ in the β-domain under consideration is difficult because it involves computing a residue at a triple pole of Hankel functions with complex index in a region where several terms in their asymptotic expansion may have to be retained. Nevertheless, for $\beta = 130$ and $\theta = 180^\circ$,

$$f_{2,\mathrm{res}}(130,\pi) \approx -0.165 + 0.483\mathrm{i}\,.$$

The geometric ray contribution to this order of approximation consists of two terms: the contribution from the directly reflected ray ($f_{0,g}$) and that from the central ray which undergoes

one internal reflection ($f_{2,g}$). Thus (for $N = 1.33$)

$$f_g = f_{0,g}(130, \pi) + f_{2,g}(130, \pi) \approx 0.101 + 0.176\mathrm{i} \ .$$

The "exact" scattering amplitude can be computed from its partial-wave expansion with the result that

$$f(130, \pi) \approx -0.011 + 0.798\mathrm{i} \ .$$

Noting that $|f_g|^2 \approx 0.04$ while $|f|^2 \approx 0.64$, it is clear that the geometrical optics or lowest WKB approximation accounts for only about 7% of the total intensity *and* has the wrong phase. The remainder must be accounted for by the residue-series contributions; this is what has been referred to above as strong "Regge-pole dominance" of the scattering amplitude (but in contrast to earlier circumstances, this is a lit region as opposed to a shadow region where the amplitude would vanish in the WKB approximation). Indeed, the above $f_{2,\mathrm{res}}$ term alone corrects the phase in the right direction, and

$$|f_g + f_{2,\mathrm{res}}|^2 = |-0.064 + 0.659\mathrm{i}|^2 \approx 0.44 \ ,$$

which is about 70% of the correct value. This implies that the Van de Hulst conjecture is substantially correct. The remaining difference must come from residue-series contributions of higher-order terms in the Debye expansion. These are not damped as rapidly as the geometrical optics terms, due to the high internal reflection coefficient. The angular distribution for these contributions near the backward direction follows from the approximation [107]

$$P_{\lambda_n - 1/2}(-\cos\theta) \approx J_0(\beta[\pi - \theta]) = J_0(\beta\epsilon) \ , \tag{5.65}$$

where, as noted earlier,

$$\lambda_n \approx \beta + \mathrm{e}^{\pi\mathrm{i}/3}\gamma^{-1}x_n$$

and J_0 is a Bessel function of the first kind of order zero. This of course corresponds to an intensity distribution near the backward direction proportional to $J_0^2(\beta\epsilon)$, which is not unreasonable when compared with the observations [7]. In fact, this will explain the slow intensity decrease at large angles, in contrast with diffraction coronae, for which the intensity is proportional to $J_1^2(\beta\theta)/(\beta\theta)^2$. In the former case, the intensity decreases like $(\beta\epsilon)^{-1}$, whereas in the latter case it decreases like $(\beta\theta)^{-3}$. Furthermore, the ratios of the dark-ring radii in the glory are given by the ratios of the corresponding zeros of $J_0(x)$, namely

$$\epsilon_1/\epsilon_2 \approx 0.44, \quad \epsilon_3/\epsilon_2 \approx 1.6 \ ,$$

which compares quite favorably with the observational results quoted earlier from [7].

There are two other pieces of evidence which tend to support the general correctness of the Van de Hulst surface-wave conjecture, the numerical evidence being that when Bryant and Cox [54] examined the partial sums of the Mie series as a function of the number of partial waves retained, they found in the near-backward direction that the most important contribution by far comes from the edge domain, $l_- \lesssim l \lesssim l_+$, which corresponds to nearly grazing incident rays. The other piece of evidence is experimental; Fahlen and Bryant [55] observed a luminosity of the circumference of a water droplet as seen from the backward direction, analogous to the

luminosity of a diffracting edge as seen from the shadow. This provides direct evidence for the existence of intense surface-wave contributions along the 180° direction.

With the surface-wave explanation in principle fairly well established, we can now ask why the exponential damping of these waves along a 14° arc would not prevent them from making such a significant contribution. It turns out that, while the specific surface waves referred to by Van de Hulst are nonetheless significant, such contributions from higher-order terms in the Debye expansion (by now not surprisingly) also play an important role in this phenomenon. We will not delve further into the mathematical details here, merely sufficing ourselves with some qualitative comments. It can be shown that the total residue-series contribution has the same order of magnitude as the second term in the Debye expansion [6]. Certain "resonance" phenomena may occur among diffracted rays, wherein each diffracted ray returns to its starting point after an integer number of shortcuts taken through the sphere. Thus for $N=\sqrt{2}$, (outside the scope of interest in this article) a resonance occurs after four shortcuts, i.e. the internal piecewise path is a square. For $N=[\cos(\pi/5)]^{-1}\approx 1.236$, the corresponding resonance path is a regular pentagon. The character of oscillations in the scattering amplitude can be explained theoretically, in part at least, by the deviation (or otherwise) of the refractive index from one or other of these resonance values. In fact the intensity is very sensitive to the values of β and N, which may also explain the variability of the glory referred to earlier (see further comments on this below). The magnitude of the residue-series-dominated resonance peaks predicted by the theory is bounded by

$$|f_{\rm res}(\beta,\pi)| \lesssim \beta^{1/3}\ ,$$

which can be appreciated by considering the approximation to the partial-wave series

$$f_{\rm res}(\beta,\pi)\approx \frac{1}{{\rm i}\beta}\sum_{l=l_-}^{l=l_+}(-1)^l\left(l+\frac{1}{2}\right)[S_l(k)-1]\ . \tag{5.66}$$

It follows from the unitarity condition for the S-matrix that

$$|S_l(k)-1|\leqslant 2\ ,$$

where the upper bound is attained when the lth partial wave is resonant. Using this and the expression for the *edge domain* (or edge strip)

$$l_-\sim\beta-c\beta^{1/3}\lesssim l\lesssim l_+\sim\beta+c\beta^{1/3},\quad c=\mathcal{O}(1)\ ,$$

leads to the above result. Physically, it arises because most of the partial waves within the edge domain are close to resonance (see Fig. 13). We have seen that the diffracted-ray resonances are associated with poles of the Debye expansion; resonances in particular partial waves correspond to Regge poles close to the real axis. Thus it appears that there is an interesting two-way relationship between the Debye and the partial-wave expansions: surface-wave resonances may be interpreted as the cumulative effect of many partial waves close to resonance, and vice versa.

For the *forward* scattering amplitude, an irregular fluctuation known as the "ripple" is known to occur, which Van de Hulst suggested may be a type of "forward" glory [7]. This is reflected in the total cross-section, via the optical theorem. By summing the contributions for $\beta=130$

from the forward diffraction peak (f_d), the Fock correction terms (f_F) and the geometrical-optic component (f_g, neglecting non-central rays), it is found that

$$f_d(130,0) + f_F(130,0) + f_g(130,0) \approx -2.142 + 69.093\mathrm{i}$$

for $N = 1.33$. The corresponding "exact" result from the partial-wave series is

$$f(130,0) = -2.529 + 68.988\mathrm{i}\ ,$$

so that the residue-series contribution must of necessity be

$$f_{\mathrm{res}}(130,0) \approx -0.387 - 0.105\mathrm{i}\ ,$$

which is in fact smaller than any of the three terms in the approximate sum above.

The total cross section σ_{tot} is related to $f(\beta,0)$ via the optical theorem, i.e.

$$\sigma_{\mathrm{tot}} = \frac{4\pi a^2}{\beta}\mathrm{Im}\, f(\beta,0)\ . \tag{5.67}$$

In terms of the above notation we may write, in an obvious manner

$$\sigma_{\mathrm{tot}} = \sigma_d + \sigma_F + \sigma_g + \sigma_{\mathrm{res}}\ .$$

The diffraction, Fock and geometric terms give rise to a slowly varying background (which in the absence of the geometric contribution is monotonically decreasing) which approaches σ_d asymptotically as $\beta \to \infty$. Since $f_d = \mathrm{i}\beta/2$ and hence $\sigma_d = 2\pi a^2$, this is equivalent to the well-known result that the asymptotic cross-section is twice the geometrical cross section. The geometrical-optic contribution gives a set of relatively slow oscillations with period

$$\delta_g\beta = \frac{\pi}{N-1}$$

and amplitude decreasing like β^{-1}, superimposed on the background. Physically, these oscillations arise from interference between waves diffracted around the sphere and those transmitted through it (geometrically).

No ripple is to be expected for $N < 1$ (and none is observed); diffracted rays cannot take any shortcuts through the sphere in this case. However, damping of the ripple is to be expected for an absorbing sphere (complex N, $\mathrm{Re}\,N > 1$), and such attenuation has indeed been observed in numerical computations [54]. For small values of β (here meaning $\beta \lesssim 4$) the ripple peaks may be explained as resonances in successively higher partial waves. As β increases however, more than a single partial wave may be near resonance, and the ripple may be thought of more reasonably as a surface-wave phenomenon. In a similar manner to the discussion of the glory, these two pictures are actually symbiotic in that each effect in one description corresponds to a collective effect produced by several terms in the other one. There is no reason why ripple effects should be confined to just the backward and forward directions, of course. As pointed out earlier, they have been observed in numerical calculations [54] for $\theta = 90^\circ$ with corresponding period *twice* that found for $\theta = 0^\circ$ and 180°. The amplitude of the ripple contribution has been estimated [6]; for low values of β and $N = 1.33$ the dominant contribution arises from the

diffracted rays taking four shortcuts (the "square" resonance mentioned earlier). Under these circumstances the contribution to the scattering cross section from the pole λ_1 is

$$\sigma_{\text{res}} \sim A(\beta)\beta^{-2/3} \exp(-\zeta_4 \operatorname{Im} \lambda_1) \sin(4 \operatorname{Re} \delta + \chi) , \tag{5.68}$$

where the various terms are as follows: A is a bounded amplitude factor, ζ_p is the minimum angle described by surface waves after p shortcuts through the sphere before emerging in the direction θ, $\delta = \delta(N, \beta, \lambda_1, \zeta_2)$ is a complex number describing the period of the ripple oscillation, and χ is a constant phase term. From this expression it follows that the relative amplitude of the ripple component is at most of order $\beta^{-2/3}$ at $\theta = 0^\circ$, and at most of order $\beta^{1/3}$ at $\theta = 180^\circ$ (see [6]) where the ripple is dominant. In directions far from these two angles Nussenzveig concludes that the relative amplitude of the ripple is of order $\beta^{-1/6}$, i.e. the geometric mean of the values in the backward and forward directions. This is in good agreement with numerical estimates made by Penndorf [41], who found that the ripple is present in all directions, with increasing relative amplitude as θ increases from 0° to 180°. He estimated the average relative amplitude as 0.1, 5 and 500, respectively, at $\theta = 0^\circ, 90^\circ$ and 180°.

5.7. Summary of the CAM theory for rainbows and glories

The modified Watson transformation is a very effective tool in the high-frequency domain ($\beta^{1/3} \gg 1$, $\beta^{1/3}|N-1|^{1/2} \gg 1$) for extracting the complete asymptotic behavior of the scattering amplitude in any direction from the partial-wave expansion. In general, the total amplitude in any direction is found by adding the contributions from the first three terms in the Debye expansion and incorporating, where appropriate, the higher-order correction terms. There are six subdivisions of the interval $(0, \pi)$ that arise in the rainbow/glory problem (i.e. corresponding to the third term in the Debye expansion) for $N > 1$. Essentially the regions consist of those predicted by geometrical optics and transition regions; the dominant term in the lit regions is *usually* that provided by the geometrical-optic contribution, which is the first-order WKB approximation. (The first correction term is the second-order WKB approximation.) In the shadow regions the surface-wave contributions usually dominate, and this is usually pictured in terms of geometrical diffraction. Surface waves are excited by tangentially incident rays (for $N > 1$) for both the impenetrable and transparent sphere problems, though in the latter case, as noted earlier, there are two types of surface interactions from within the sphere wherein shortcuts may occur across the sphere.

There are three types of transition problem of interest here: (i) normal transitions (of the generalized Fock type); (ii) the rainbow; (iii) the glory. Type (i) transitions also arise for the impenetrable sphere, and in both cases their angular width is of order γ. They include the region around the forward diffraction peak. The rainbow region (ii) is associated with the transformation of a pair of real rays into complex rays, and a uniform asymptotic expansion can be developed for the amplitude which contains the Airy theory as a special case, but which is valid for a considerable domain beyond that approximation. Finally, the glory region (iii) is an example of "Regge-pole dominance" of the scattering amplitude in near-backward directions. The glory is due to surface waves taking shortcuts through the sphere, but with some corrections (from higher-order terms) to the original idea proposed by Van de Hulst. With the exception of polarization, the CAM version of the theory enables an almost complete description

of the glory to be made. The features so explained arise from competition between four different phenomena: (a) exponential damping of surface waves as they travel along the spherical surface; (b) focusing of diffracted rays along the axis, which enhances the back-scattered contribution; (c) high internal reflectivity of diffracted rays at the surface, which means that many internal reflections must be taken into consideration in general; (iv) resonance effects associated with nearly closed circuits after four successive short-cuts. Estimates of higher-order residue-series contributions can also be made. The quasiperiodic intensity fluctuations associated with these resonant diffracted rays are present in all directions, but dominate near the backward direction in terms of large intensity variations that characterize the glory region. The contribution from surface-wave terms decreases with increasing β, eventually (for large enough β) permitting the geometrical-optic terms to become dominant again. The electromagnetic problem is discussed in some detail by Khare and Nussenzveig in [47,108] (see also [53]).

Nussenzveig critiqued his ('phase one') papers from mathematical, numerical and physical points of view [6] in the sense of suggesting further improvements, extensions and applications, which were later addressed in [10] (and references therein). Nevertheless, several of his earlier comments in this regard bear repeating here: in the neighborhood of transition points between different ranges of N, a confluence of saddle points may occur near $\theta = \pi$, leading to a mixture of rainbow and glory effects; this has been discussed more fully in [47]. Furthermore only the scattering amplitude has been studied—the behavior of the wavefunction itself has not been discussed (as it *has* been for the impenetrable sphere problem [33]). The extension of the theory to complex N to represent an absorptive sphere is of interest, both from a mathematical point of view (the convergence of the Debye series would be improved owing to increased damping), and because of potential applications to both nuclear physics (in the optical model) and atomic physics (see Section 6.5). Indeed, there was also early evidence for nuclear glory scattering [109]. Extension of the theory to both different-shaped scattering objects and *inhomogeneous* bodies (see for example [110]) is of considerable interest (both of which correspond to more general potentials than the square well/barrier in the quantum–mechanical context).

The *second phase* of the analysis of the glory (from about 1980 onward) filled in the gaps that were present in the early studies. The later work showed that higher-order terms in the Debye expansion, both from surface waves and the shadow of higher-order rainbows, cannot be neglected [10]. Noting that the glory is one of the most complicated of all scattering phenomena [10], Nussenzveig lists (along with their explanation) a total of 15 observational and numerical features associated with optical glories (some of which are conspicuously different from *coronae* (which arise from a forward diffraction peak); indeed, as already noted, van de Hulst named them *anti-coronae* [50]). Briefly summarized, they are:

(i) variability of the glory rings; (ii) their angular distribution; (iii) polarization (parallel polarization appears to be dominant in the outer rings); (iv) the angular width of the glory region is $O(\beta^{-1})$; (v) features around $\beta = 10^2$; (vi) features around $\beta = 10^3$; (vii) intensity enhancement (at least an order of magnitude greater than that predicted by geometrical optics; this is due to dominant diffraction effects); (viii) ripple fluctuations; (ix) quasiperiodicity; (x) background intensity; (xi) effects of absorption or size averaging; (xii) other periodicities; (xiii) quasichaotic features; (xiv) average gain factor (the ratios of polarized intensities to their limiting value as predicted by geometrical optics for a totally reflecting sphere); and (xv) edge origin (dominant contributions to the backscattered intensity arise from the edge strip).

Within the glory region, cross-polarization effects also contribute to the intensity enhancement. Concerning surface waves, van de Hulst's $p=2$ term is an important contributor to the optical glory, but which higher-order Debye terms might contribute significantly to backscattering? It is known that in general such terms are strongly damped by multiple internal reflection except within the edge strip(s)

$$\beta \leqslant \lambda \leqslant \beta(1+c_+\gamma^2)$$

and

$$\beta(1-c_-\gamma^2) \leqslant \lambda \leqslant \beta\ ,$$

i.e. near the top of the centrifugal barrier (as noted in Section 5.1 the constants $c_\pm$ are of order unity; note also that 'physical' values of λ correspond to $\lambda = l+\frac{1}{2}$; in Sections 4 and 5 and [5,6], the edge strip is stated in terms of l; in [10] the above form is used). It transpires that Debye terms of orders up to a few times $\beta^{1/3}$ (undergoing a large number of internal reflections in a manner analogous to orbiting) can yield appreciable contributions from these regions. In addition there are enhancements due to axial focusing (of order $\beta^{1/2}$ arising from constructive interference along the axis from toroidal wavefronts where the edge strip acts as a virtual ring source) and proximity to an edge backward glory ray or a near-backward higher-order rainbow (of order $\beta^{1/6}$). The latter arise from near-edge incidence; the *10th-order* rainbow is predominant, being formed very close to $\theta=\pi$, and for $\beta \gtrsim 200$ this is the leading term. However, the width of the main rainbow peak decreases as β increases, and complex-ray damping in the rainbow shadow reduces the backward gain factor [10]. The quasiperiodic features referred to above are consequences of interference oscillations among Debye components (and corresponding shortcuts through the droplets); the quasichaotic features are identified with the sensitive dependence of lower-order geometric resonances (for closed or nearly closed orbits) on the droplet size and physical parameters, as are the resonances in the total amplitude (the ripple fluctuations). The slowest ripple fluctuations are also major contributors to the glory; they correspond to resonances in the total scattering amplitude (s). Because an infinite number of branches of the deflection function contibute to this component of the glory (in keeping with the concept of orbiting), they cannot be represented as a finite sum of terms in the Debye expansion. The resonances arise from complex Regge poles near $\lambda=\beta$ (see Fig. 19) as a result of tunneling around the edge strip. Thus (and leading naturally to the next subsection) it has been stated that [10]:

> ...all leading contributions to the glory arise from complex critical points (poles or saddle points), so that this beautiful and impressive meteorological effect is produced almost entirely by *light tunneling* on a macroscopic scale.

5.8. A synopsis: diffractive scattering, tunneling effects, shape resonances and Regge trajectories [89]

For a sphere of radius a and dimensionless size parameter $\beta=ka\gg 1$ (k being the incident wavenumber) the diffractive scattering amplitude in direction θ is [65]

$$f_d(k,\theta)=\frac{\mathrm{i}a}{\theta}J_1(\beta\theta)\ , \tag{5.69}$$

where J_1 is the Bessel function of the first kind of order one. Nussenzveig [10] notes that the angular width of the forward diffraction peak (and the spacing of secondary peaks) is very narrow, being $O(\beta^{-1})$. The total (elastic) diffractive cross section [65] is

$$\sigma_d \approx 2\pi \int_0^\infty |f_d(k,\theta)|^2 \theta \, \mathrm{d}\theta = 2\pi a^2 \int_0^\infty \frac{J_1^2(z)}{z} \, \mathrm{d}z = \pi a^2 \,, \tag{5.70}$$

which is the cross sectional area of the sphere. The forward peak arises from axial focusing: secondary waves with different azimuths interfere constructively along the axis. The domain of validity of the above expression is $0 \leqslant \theta \leqslant \beta^{-1}$, and the question therefore arises: how can the domain be extended to larger angles of diffraction? One approach was developed by Fock [112], based on the idea of "transverse diffusion" along a wavefront [113,114]; an alternative approximation is Keller's *geometric theory of diffraction* [106,115]. Diffracted rays are introduced as extremals of the optical path, and follow the laws of geometrical optics. A smooth scattering object (e.g. a sphere) can support such rays along segments of the boundary; tangentially incident rays excite such diffracted rays. Considered as surface waves, their amplitudes are damped exponentially with angle of travel (θ) around the surface because of tangential shedding of radiation (in acoustics these waves are sometimes referred to as *creeping modes*). As noted above, competing approximations can be conveniently tested using the model of scattering (both by direct reflection and diffraction, in this instance) by an impenetrable sphere of radius a (the transparent sphere is not necessary since it is only the "blocking effect" that is being described here; obviously the transparent sphere model is crucial for understanding the complete scattering problem).

The domain of validity of the semiclassical approximation to the reflection amplitude is determined by the condition that the first Fresnel zone for the reflected ray should not intersect the discontinuity at the light/shadow boundary on the surface of the sphere [114]. This imposes the condition

$$\theta \gtrsim 2\left(\frac{2}{\beta}\right)^{1/3} = 2\gamma \,,$$

where γ is a parameter measuring the angular width of the penumbral region. This is much broader than the classical diffraction domain $\theta \lesssim \beta^{-1}$ (since $\beta \gg 1$), but both the convergence of the WKB expansion (used in the semi-classical approximation) and the surface wave contributions are rapid *only for* $\theta \gg \gamma$. It is not surprising therefore that the greatest analytical and numerical challenges are associated with the domain between these two angular regions. Fock's theory yields only a transitional asymptotic approximation here, and is based on an inadequate physical picture (because the diffusion coefficient is *pure imaginary*, the diffusion equation should be replaced by the Schrödinger equation, with corresponding changes in interpretation—see below). A uniform semiclassical approximation for potentials with long-range tails was provided by Berry [116] (though recall that the forward diffraction peak in this case is not caused by edge diffraction), and for finite-range potentials of the type discussed here, Nussenzveig and co-workers applied complex angular momentum (CAM) theory to obtain the requisite uniform approximation [117–119]. The essential (and novel) feature of his approach is to regard diffraction *as a tunneling process.*

The basic procedure used is to transform the partial-wave expansion for the scattering amplitude by means of a modified version of the Poincaré–Watson transform (via the Poisson summation formula [33]). In this process, the scattering function $S_l(k)$ is extended to continuous values of angular momentum λ in the complex λ-plane (at this juncture for the impenetrable sphere only; the extension to the transparent sphere was developed later [5,6]). The resulting expression for the scattering amplitude $f(k,\theta)$ can be considered to be a superposition of "pseudoclassical paths" (generalizations of diffracted rays) [120] associated with all values of λ (away from the forward and backward scattering directions). The qualification is to distinguish such paths from the corresponding classical paths, the envelope of which is the sphere $r = r_0(\lambda)$, r_0 being the outermost radial turning point (diffracted rays are pseudoclassical paths with $r_0(\lambda) = a$; they are composed of incoming and outgoing classical paths). The resulting integral superposition over λ arises from the classical "action" along such paths (the action being the analog of the optical path).

In appropriate units (i.e. $\hbar = 2m = 1$), the impact parameter associated with angular momentum λ is $b = \lambda/k$ (this follows from the so-called *localization principle* [7,10]) tangential rays (grazing or edge rays) correspond to the threshold value $\lambda = \beta$ with energy k^2 located at the top of the potential barrier in the effective potential (see Figs. 29(a), (c) and also Fig. 13). So-called below-edge rays ($\lambda < \beta = ka$) 'hit' the sphere, and above-edge rays ($\lambda > \beta$) pass outside with a radial turning point given by $r_0(\lambda) = b(\lambda)$; it is the contributions from these regions close the edge that give rise to significant modifications of classical diffraction. When $\lambda > \beta$ tunneling through the 'triangular' barrier can occur between $r = r_0(\lambda)$ and a (see below for further discussion of this); as might be expected, if the potential is approximated by a linear potential, the solutions of the Schrödinger equation are expressible in terms of Airy functions (more precisely, as *generalized Fock functions*; see Appendix B). By deforming the path of integration for the scattering amplitude into the complex plane, the integral can be evaluated approximately from a small number of dominant contributions. The *outer approximation* is valid for $\theta \gg \gamma$, being the sum of the WKB expansion (for reflection) and small surface wave corrections. For values of θ below this range (including zero) the *uniform approximation* is composed of three dominant terms, namely uniform versions of the classical diffraction amplitude, above-edge amplitude (incorporating the effects of tunneling through the centrifugal barrier) and below-edge amplitude (see [121]). The CAM approximation is obtained from a smooth match of the outer and uniform approximations, and is valid for $0 \leqslant \theta \leqslant \pi$. It is also extremely accurate [10,121] and bridges the gap between short and long wavelength scattering (remaining accurate down to $\beta \cong 1$).

There is an interesting consequence from the effects of above-edge tunneling: each contribution is weighted by the uniform tunneling penetration amplitude through the centrifugal barrier to the surface, and in contrast to classical diffraction, represents a *nonlocal* interaction with range $b - a \approx (\lambda_0^2 a)^{1/3}$, i.e. a weighted geometric mean between the wavelength λ_0 and the radius a. Furthermore, it should come as no surprise by now that near-edge tunneling is also very significant in rainbow and glory scattering and also orbiting. We are reminded in [122] that descriptions of tunneling always involve analytic extensions to complex values of some parameter, and CAM is a natural vehicle for this. In rainbow scattering, for example, penetration into the shadow zone (tunneling) occurs through a complex ray; in the optical glory surface waves (launched by tunneling) and complex rainbow shadow rays make dominant contributions

to this phenomenon; orbiting also contributes to the glory, and to resonance effects, which are well-illustrated by the problem of scattering by a transparent sphere (with real refractive index $N > 1$). Now the effective potential has a discontinuity at $r = a$, essentially because of the 'addition' of a potential well to the centrifugal potential. Thus the 'spike' corresponds to a barrier surrounding a well, and suggests the existence of resonances, particularly between the top of the former and bottom of the latter, where there are three turning points. Such resonances are called "shape resonances" (and sometimes these are called "morphology-dependent resonances" [123]); they are quasibound states in the potential well that escape by tunneling through the centrifugal barrier. The widths of these resonances depends on where they are located; the smaller the node-number of the radial wavefunction within the well, the deeper that state lies in the well. This in turn determines the width (and lifetime) of the state, because the tunneling amplitude is "exponentially sensitive" to the barrier height and width [122]. Since the latter decreases rapidly with the depth of the well, the smaller is the barrier transmissivity and the lowest-node resonances become very narrow for large values of β. The lifetime of the resonance (determined by the rate of tunneling through the barrier) is inversely proportional to the width of the resonance, so these deep states have the longest lifetimes [123,88] (with theoretical details in [89]).

Some calculations are in order at this point. Powerful analogies abound in physics; of particular interest here are the well-known analogies of Hamiltonian mechanics with geometrical optics and wave mechanics with wave optics [10]. In the former case, a non-relativistic particle of energy $E\ (=k^2)$ in a central potential $V(r)$ has associated with it in the optical realm an index of refraction N given by

$$N = \sqrt{1 - \frac{V(r)}{E}}\ . \tag{5.71}$$

The analog of the optical path being the *action*, Fermat's principle corresponds to the principle of least action. In particular, if the potential is a square well for which $V(r) = -V_0$, $V_0 > 0$, for $0 < r < a$, then the depth of the well is $V_0 = (N^2 - 1)k^2$; this is therefore a potential which is *directly proportional to the energy* in contrast to the standard quantum mechanical problem in which $V(r)$ is a fixed function, independent of the energy. Thus it follows that if the energy k^2 is varied, so does the depth of the potential well (or barrier, if $V_0 < 0$). The effective potential corresponding to the square well is (in units with $\hbar = 2m = 1$)

$$\begin{aligned} (3)\quad U(r) &= \frac{\lambda^2}{r^2} - (N^2 - 1)k^2 \quad \text{for } 0 < r < a \\ &= \frac{\lambda^2}{r^2} \quad \text{for } a \leqslant r < \infty \end{aligned} \tag{5.72}$$

(see Fig. 29). Note that as k^2 is reduced, the bottom of the potential rises (and for some value of k the energy will coincide with the bottom of the well [123]); however, at the top of the well, $U(a) = \lambda^2/a^2$ is independent of k^2, but if k^2 is increased, it will eventually coincide with the top of the well. Consider a value of k^2 between the top and the bottom of the well: within this range there will be three radial turning points, the middle one obviously occurring at $r = a$, and

the largest at $r=b$ for which $U(b)=\lambda^2/b^2$. The smallest of the three ($r_{\min}$) is found by solving the equation

$$k^2=\frac{\lambda^2}{r_{\min}^2}-(N^2-1)k^2\ ,$$

to obtain

$$r_{\min}=\frac{\lambda}{Nk}=\frac{b}{N} \tag{5.73}$$

in terms of the impact parameter $b(\lambda)=\lambda/k$. By applying Snell's law for given b, it is readily shown that the distance of nearest approach to the center of the sphere is just $r_{\min}$; indeed, there are in general many nearly total internal reflections (because of internal incidence beyond critical) within the sphere between $r=b/N$ and a. This is analogous to orbiting in a ray picture; the very low leakage of these states allows the resonance amplitude and energy to build up significantly during a large resonance lifetime which in turn can lead to non-linear optical effects [10]. In acoustics these are called "whispering gallery modes" [124,125].

The energy at the bottom of the well (i.e. $\lim_{r\to a^-}U(r)$) corresponding to the turning point at $r=a$ is determined by the impact parameter inequalities $a<b<Na$, or in terms of $\lambda=kb$,

$$U(a^-)=\left(\frac{\lambda}{Na}\right)^2<k^2<\left(\frac{\lambda}{a}\right)^2=U(a^+)\ , \tag{5.74}$$

which is the energy range between the top and bottom of the well (and in which the resonances occur). To cross the "forbidden region" $a<r<b$ requires tunneling through the centrifugal barrier and near the resonance energies the usual oscillatory/exponential matching procedures lead to very large ratios of internal to external amplitudes (see Fig. 29(c)); these resonances correspond to "quasibound" states of light (that would be bound in the limit of zero leakage) [88].

Mathematically, the resonances are complex eigenfrequencies associated with the poles λ_n of the scattering function $S(\lambda,k)$ in the first quadrant of the complex λ-plane; these are known as Regge poles (for real β). Corresponding to the energy interval $[U(a^-),U(a^+)]$ the real parts of these poles lie in the interval $\beta=ka<\mathrm{Re}\,\lambda_n<N\beta=Nka$ (i.e. impact parameters $\in[a,Na]$ which is of course the tunneling region). The imaginary parts of the poles are directly related to resonance widths (and therefore lifetimes); in fact Nussenzveig has coined the term "life angle" instead of lifetime, since the leakage gives rise to angular damping [92]. As n decreases, $\mathrm{Re}\,\lambda_n$ increases and $\mathrm{Im}\,\lambda_n$ decreases very rapidly (reflecting the exponential behavior of the barrier transmissivity). As β increases, the poles λ_n trace out Regge trajectories, and $\mathrm{Im}\,\lambda_n$ tends exponentially to zero. When $\mathrm{Re}\,\lambda_n$ passes close to a "physical" value, $\lambda=l+1/2$, it is associated with a resonance in the lth partial wave; the larger the value of β, the sharper the resonance becomes for a given node number n. More specific mathematical details are provided below.

It has been noted several times that Mie scattering cross sections have a very complicated structure as functions of β; this "ripple" is extremely sensitive to parameter changes in the cross sections (quasichaotic) at all angles, being greatly enhanced in backscattering; but it does contain some quasiperiodic features. It is very much related to the sensitivity of the resonances to variations in barrier shape noted above. A detailed CAM analysis [88,89] identifies the ripple

with the residue contribution from *Regge trajectories*, and it is to these we now briefly turn. The poles of the $S^j(\lambda,\beta)$ function (the scalar version of which is given by Eq. (5.22)) are the roots of the following complex transcendental equation [10]

$$\ln' H_\lambda^{(1)}(\beta) + \frac{1}{2\beta} = Ne_j \left\{ \ln' J_\lambda^{(1)}(\alpha) + \frac{1}{2\alpha} \right\} \tag{5.75}$$

in the notation of Section 5, where '$\ln'$' refers to the logarithmic derivative, and for the magnetic (electric) polarization, e_1 (e_2) $= 1$ (N^{-2}). For real β, the Regge poles are the zeros of this expression in the complex λ-plane; for real $\lambda = l + \frac{1}{2}$, the poles may also be found in the complex β-plane, although this is not particularly relevant here. It is the Regge poles close to the real λ-axis, with $\beta < \mathrm{Re}\,\lambda < \alpha$, that are of interest (see Fig. 14). The imaginary part of these poles is exponentially small. Their approximate location was obtained in [5,88], and if N is real, their real parts (in the lowest order of approximation) are determined by the implicit relation

$$2\int_{\lambda/Nk}^{a} \left(N^2k^2 - \frac{\lambda^2}{r^2} \right)^{1/2} \mathrm{d}r = (2n+1)\pi \quad (n = 0, 1, 2, \ldots)\,, \tag{5.76}$$

where the integrand plays the role of an effective radial wavenumber within the well. The integral is a radial phase integral between the inner turning point and the surface; it is noted in [10] that the above condition is like a Bohr–Sommerfeld quantization condition [20] for 'bound' states of electromagnetic energy. The integer n plays the role of a quantum (or 'family') number. To the same degree of approximation, the imaginary part of the Regge poles is proportional to the centrifugal barrier penetration factor

$$\exp[-2\Psi(a,\lambda/k)] \equiv \exp\left[-2\int_a^{\lambda/k} \left(\frac{\lambda^2}{r^2} - k^2 \right)^{1/2} \mathrm{d}r \right]\,, \tag{5.77}$$

where Ψ is the radial phase integral between the surface and the outer turning point. Higher-order approximations can be obtained [88]. As β varies, each complex Regge pole $\lambda_{nj}(\beta)$ describes a Regge trajectory (a path in each of the $\mathrm{Re}\,\lambda/\beta$ and $\mathrm{Im}\,\lambda/\beta$ planes). When $\mathrm{Re}\,\lambda_{nj}(\beta) = l + 1/2$ a resonance occurs for polarization j in the lth partial wave. The lifetime of the resonant state is the inverse of its width and proportional to the inverse of the transmissivity of the barrier. As such, the dominant influence on resonant widths is that of tunneling through the centrifugal barrier.

We revisit non-relativistic quantum scattering of an impenetrable sphere by considering the generalization of the above expression (in connection with the above-edge amplitude). In the notation of Nussenzveig and Wiscombe [118], the uniform approximation to the dimensionless scattering amplitude $F(\beta,\theta) \equiv f(k,\theta)/a$ is [117]

$$F = F_d + \tilde{F}_s + F_{e^+} + F_{e^-}\,, \tag{5.78}$$

where the first term is the uniform version of the classical (blocking) diffraction amplitude. The second term $\tilde{F}_s$ is the surface-wave contribution from waves that have traversed at least half the circumference, and is usually small in comparison with the last two terms which cumulatively

are referred to as the *edge amplitude* F_e, arising from incident rays with impact parameters in the edge domain. The '$\pm$' identifies the above/below edge amplitudes respectively. Of primary interest here is the former amplitude, F_{e^+}. By employing uniform asymptotic expansions for both the Legendre function $P_{\lambda-1/2}(\cos\theta)$ and cylindrical functions (see [117] for more details) it is found that

$$F_{e^+}(\beta,\theta) = -\int_0^\infty \mathscr{H}(x,\phi)\mathscr{P}(\theta,\beta\cosh\phi)(\phi^{-1}\cosh\phi)x^{1/2}\,\mathrm{d}x\ , \tag{5.79}$$

where

$$\mathscr{P}(\theta,\lambda) \equiv \left(\frac{\theta}{\sin\theta}\right)^{1/2}\left\{[1+O(\lambda^{-2})]J_0(\lambda\theta) - \frac{1}{8}(\theta^{-1}-\cot\theta)\frac{J_1(\lambda\theta)}{\lambda} + O(\lambda^{-3})\right\}\ ,$$

$$\mathscr{H}(x,\phi) \equiv \frac{\mathrm{e}^{-\mathrm{i}\pi/6}\,\mathrm{Ai}(x) - \sigma(x,\phi)\mathrm{Ai}'(x)}{\mathrm{Ai}(\mathrm{e}^{2\mathrm{i}\pi/3}x) + \mathrm{e}^{-\mathrm{i}\pi/6}\sigma(x,\phi)\,\mathrm{Ai}'(\mathrm{e}^{2\mathrm{i}\pi/3}x)}$$

with

$$\sigma(x,\phi) \equiv \frac{5\mathrm{e}^{-\mathrm{i}\pi/6}}{24\beta x^{1/2}\sinh\phi}\left[\frac{1}{3(\phi\coth\phi-1)} - \coth^2\phi + \frac{3}{5}\right]$$

and

$$x \equiv \left[\frac{3\beta}{2}(\phi\cosh\phi - \sinh\phi)\right]^{2/3}\ .$$

This independent variable is equivalent to $\gamma(\lambda-\beta) \equiv \gamma k(b-a)$, where b is the impact parameter associated with angular momentum λ; note that $|x| \lesssim 1$ defines the edge strip [10]. For $x \gg 1$, $\mathscr{H}(x,\phi)$ is approximated by

$$\mathrm{e}^{\mathrm{i}\pi/3}\mathscr{H}(x,\phi) \approx \exp\left(-\frac{4}{3}x^{3/2}\right) = \exp\left[-2\int_a^{\lambda/k}\left(\frac{\lambda^2}{r^2}-k^2\right)^{1/2}\mathrm{d}r\right]\ . \tag{5.80}$$

(There is a minor misprint in Eq. (7) of [118].) Hence F_{e^+} is a pure tunneling amplitude: it describes uniform tunneling of the outside rays through the centrifugal barrier to the surface (with which they interact) and back.

The below-edge amplitude F_{e^-} is given by an integral similar to (5.79), and the integrand again has behavior dominated by a nonlinear exponent. For real paths, it corresponds to strong reflection above but close to the top of the centrifugal barrier (containing the effects of surface curvature on reflection); but an alternative representation in terms of complex angles [117] demonstrates that this is equivalent to tunneling with the super-exponential decay factor (5.80) (for a brief summary of the contributions from both paths of steepest descent and stationary phase describing the 'birth' of a reflected wave in the WKB region, see [10]). Thus, the full edge amplitude F_e can be interpreted as a tunneling amplitude. For $\theta \lesssim \gamma$, it can be expanded in powers of θ/γ, leading to generalized Fock functions, and the recovery of the Fock approximation. This breaks down, not surprisingly, for $\theta \gg \gamma$, and as is characteristic of transitional approximations, cannot be matched with large angle results. However, it *can* be

smoothly matched with the *outer approximation* $F = F_r + F_s$, where F_r is the reflection amplitude given by the WKB approximation (with corrections up to the second order) and F_s is the surface wave amplitude containing *all* surface-wave contributions.

6. The electromagnetic problem

6.1. Polarization

The scalar theory obviously has nothing to say with regard to polarization. We have noted, however, that the major techniques introduced in the Debye/Nussenzveig scalar approach are applicable also to the electromagnetic problem [59,10]. Much, indeed most of the analysis for the scalar problem carries over to this more general one. The main difference, as noted in Section 1.4, is that we now must deal with the *Mie solution* [24] and the fact that there are now *two* scattering amplitudes to consider: one for each of the magnetic and electric polarizations. In a change from the f notation for the scattering amplitude, Khare and Nussenzveig [47] use $S_1(\beta,\theta)$ and $S_2(\beta,\theta)$ to denote these, respectively (these forms were used in Section 3.1 also), and we will respect that notation in this section as well. As pointed out in [47,61], the intensities $|S_j(\beta,\theta)|^2$ $(j=1,2)$ and the phase difference $\delta = \arg S_1 - \arg S_2$ completely characterize the scattering. The modified Watson transform is applied to each term in the Debye (multiple internal reflection) expansion of the Mie series, i.e. for

$$S_j = S_j^{(0)} + S_j^{(1)} + S_j^{(2)} + \cdots = \sum_{p=0}^{\infty} S_{j,p}(\beta,\theta), \quad j=1,2\ , \tag{6.1}$$

where $S_{j,p}$ is associated with waves that undergo $p-1$ internal reflections. As in the scalar case, the primary *rainbow* is associated with the third Debye term $S_j^{(2)}$, which is associated with rays undergoing a single internal reflection in the spherical drop. Again, the rainbow region is a 2-ray/0-ray transition region, represented in the complex angular momentum plane by the confluence of two real saddle points then becoming complex; the method of Chester et al. (CFU method: see [72] and Appendix D) once more provides a uniform asymptotic expansion in this situation.

The dominant contribution to $S_j^{(2)}$ is given by the same type of expression(s) as stated in Eqs. (5.55)–(5.58) above. The saddle points in the 2-ray region, θ_i' and θ_i'', correspond to the two angles of incidence associated with geometrical rays emerging from the drop in the direction θ. The path of integration is the same as in the scalar case. For $|(2\beta)^{2/3}\zeta| \gg 1$ (see Eq. (5.59)), the result matches smoothly with those in the neighboring angular regions, in contrast with that from Airy's theory. Some of the CFU coefficients are considerably smaller for polarization "2" (parallel; electric) because the angle of incidence for rainbow rays is close to Brewster's angle, and this suppresses much of the contribution from that polarization (indeed, for $N=\sqrt{2}$, the rainbow angle *is* Brewster's angle (see [111]: in this paper, which is in German, a variety of rainbow features is discussed, including the refractive indices for which rainbows of all orders are completely polarized, together with brief comments on the 'appearance' of rainbows in literature, art, folklore and religion).

There is of course a contribution to the rainbow region from the ray directly reflected at the drop surface, represented by the first Debye term $S_j^{(0)}(\beta,\theta)$, and this was included in the numerical studies undertaken by Khare and Nussenzveig [59]. They compared the exact Mie results (obtained, recall by summing over a number $\gtrsim\beta$ of partial waves) for $N=1.33$, $50\leqslant\beta\leqslant1500$, and $136^\circ\leqslant\theta\leqslant142^\circ$ with both the Airy approximation and the results from system (5.55)–(5.58). The main correction to the Airy theory is provided by the $\mathrm{Ai}'(-x)$ term (where $x=-(2\beta)^{2/3}\zeta$) in Eq. (5.59). For polarization 1, within the main rainbow peak $|x|\lesssim1$, the corrections to the Airy theory are small, but for the supernumerary arcs ($x\gg1$; secondary peaks) they become significant. For polarization 2 there are large corrections to the Airy theory, because the $\mathrm{Ai}'(-x)$ term is dominant throughout the complete range of interest. Because there is a change in sign of the amplitude reflection coefficient at Brewster's angle, one such correction is that the secondary-peak maxima and minima should be interchanged for the two polarizations (see also [7]). The results are compared in several figures in [59]; in particular for $\beta=500$ oscillations with period $\approx(300/\beta)^\circ$ are superimposed on the main peak and part of the secondary (see Fig. 2 in [59]); this results from interference with the direct reflection term. Such interference remains appreciable up to $\beta=1500$ (the upper value investigated; this is true even close to the rainbow angle θ_R), so the direct reflection term has been subtracted out in their Fig. 3, which is reproduced here as Fig. 20. For the same value of β the phase difference δ is also plotted: the Airy theory fails here, even close to θ_R, whereas the exact results and asymptotic theory agree well throughout. Again, rapid oscillations arise from interference with direct reflection. The authors point out that the oscillatory nature of the deviations between the asymptotic theory and the exact solution is consistent with interpreting them as "ripple"; this, as we have noted above, is present in all directions (but dominant in the glory) is due to grazing incident rays which are in turn associated with Regge-pole-type contributions (surface waves) and higher-order Debye terms.

Concerning the *glory*, the same techniques apply, but for $\theta=\pi$ the scattering amplitudes are related by

$$S_1(\beta,\pi)=S^M(\beta)+S^E(\beta)=-S_2(\beta,\pi)\,, \tag{6.2}$$

where the contributions from the magnetic multipoles (M) and the electric multipoles (E) are of the same order (this is called the cross-polarization effect [7,53,108]). Again, the object is to find which p values in the Debye expansion for each polarization yield significant contributions at a given value of θ, and thence to represent these by suitable asymptotic expansions via the modified Watson transform. There are in general four sets of contributions from (i) geometrical-optic rays (real isolated saddle-points in the λ-plane); (ii) surface waves (complex Regge poles); (iii) rainbow terms (confluent saddle points) and (iv) Fock-type transition region contributions (interpolating smoothly between (i) and (ii)). As we have seen in the scalar case, the contributions from geometrical optics are too insignificant to account for the glory; and while van de Hulst conjectured that $p=2$ surface waves are the dominant feature in describing the phenomenon, as we have already seen, higher-order Debye terms need to be incorporated for a complete description.

For $p\gg1$, the important contributions come from the edge domain $|\lambda-\beta|\lesssim\beta^{1/3}$, where the internal reflectivity is close to one. However, there are also contributions from *higher-order*

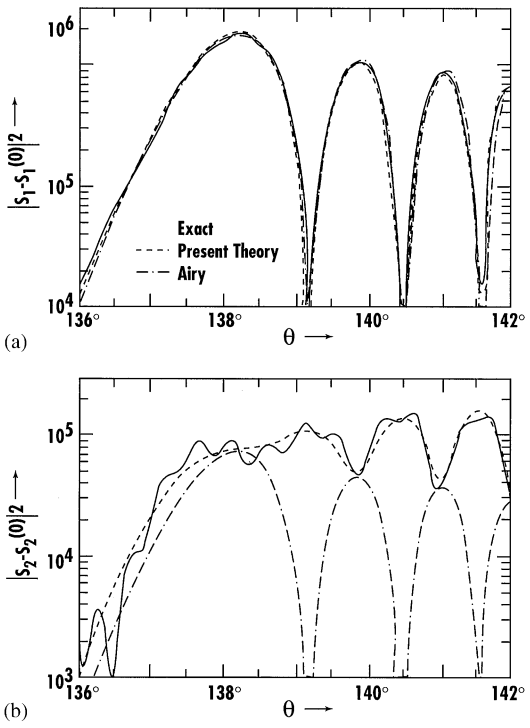

Fig. 20. A comparison of the intensities predicted by the exact Mie theory, the complex angular momentum (CAM) theory and the Airy theory for $\beta = 1500$, redrawn from [59]. The direct reflection term has been subtracted out. (a) refers to magnetic polarization intensity and (b) to the electric polarization intensity.

rainbows formed near $\theta = \pi$, due primarily to the $\beta^{1/6}$ rainbow enhancement factor [61] (but see comments in [42]) which persists at considerable angular distances from the particular rainbow angle θ_R^p (because the width of the rainbow region increases with p). The backward direction falls within the shadow side of the relevant rainbows; here the amplitude is exponentially damped with exponent proportional to $\beta(|\zeta_p|\,p^{-1})^{3/2}$ for large values of p. The corresponding exponent for the surface-wave terms is $\beta^{1/3}\zeta_p$; note that

$$\zeta_p \equiv \pi - p\theta_t \,(\mathrm{mod}\, 2\pi)\,, \tag{6.3}$$

where $\theta_t = 2\arccos(1/N)$, $-\pi \leqslant \zeta_p \leqslant \pi$. If $\zeta_p > 0$ this represents the angle described by a surface wave before emerging in the backward direction; if $\zeta_p < 0$ this corresponds to the approximate deviation $\varepsilon_R^p = \pi - \theta_R^p$ from the rainbow angle. Khare and Nussenzveig [108] conjectured, in view of the damping exponents, that the dominant Debye contributions arise from those values of p for which $-\theta_t \leqslant \zeta_p < \theta_t$, and that for surface waves the lowest ζ_p dominate, whereas rainbow terms are dominated by the lowest values of $|\zeta_p|/p$. They carried out numerical comparisons (neglecting the direct-reflection term $p = 0$) with the exact results for

$$N = [\cos(11\pi/48)]^{-1} \approx 1.33007 \tag{6.4}$$

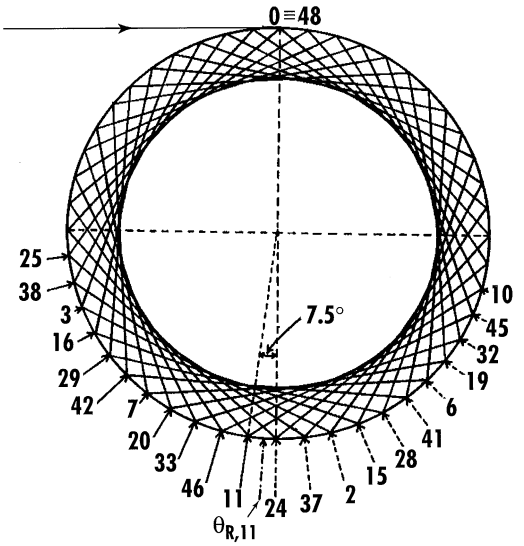

Fig. 21. The path of a tangentially incident ray for a refractive index given by $N = \sec(11\pi/48) \simeq 1.33007$ (see Eq. (6.4)). The values of p (corresponding to the pth Debye term) at each vertex are indicated, and the arrows point in the directions corresponding to the surface angles (ζ_p) traversed. Surface-wave terms are indicated by dotted arrows, rainbow terms by solid arrows. The length of the arrows corresponds qualitatively to the ordering of terms (by increasing ζ_p for surface waves and $|\zeta_p|/p$ for rainbow terms).

for which $\zeta_p = 0$, i.e. a tangentially-incident ray becomes resonant by forming a closed 48-sided star-shaped regular polygonal path inside the spherical droplet (see Fig. 21).

6.2. Further developments on polarization: Airy theory revisited

In this subsection we examine an important and mathematically simpler approach to this problem formulated by Können and de Boer [126]. It is based on the classical Airy theory but with an important difference: *a position-dependent amplitude term is introduced for the polarized rainbow* (as opposed to the common rainbow—terms to be precisely defined below). As Airy showed, the wavefront emerging from a raindrop is well approximated by a cubic function near the geometric ray of minimum deviation (the Déscartes ray), and the interference pattern of light from such a wavefront is described by the Airy function Ai(z), where $z \propto \theta_R - \theta$, i.e. the deviation of the ray from the rainbow angle (Section 2). The Airy approximation is valid provided the deviation angle is only a degree or two, and the raindrop diameter is $\gtrsim 0.3$ mm for visible light.

As already noted, light from an optical rainbow is strongly polarized; the direction of the **E** vector is tangential to the rainbow. In the primary rainbow, light is internally reflected once, and for the Déscartes ray this reflection occurs close to the Brewster angle; at this angle light polarized parallel to the plane of incidence is completely transmitted (none reflected) leaving only light polarized perpendicular to this plane in the reflected contribution. This is the tangential component referred to in [126]; the parallel polarization is referred to as radial. Since

the reflection angle for the Déscartes ray is several degrees away from the Brewster angle, some of the radial component *is* reflected to produce a correspondingly much weaker rainbow; it is this that we shall refer to as the *polarized rainbow*, the more intense common rainbow being tangentially polarized. It is interesting to note that the supernumerary bows (intensity maxima) in the polarized rainbow are shifted to the locations of the *minima* in the common rainbow; this was first observed in fog bows by Bricard [127]. From the geometrical optics viewpoint, supernumerary bows are caused by interference between two rays with different impact parameters (and hence path lengths inside the drop) and the same scattering (deviation) angle. If the respective angles of incidence for these two rays, i_1, i_2 say, are such that $i_1 < I_B < i_2$, where I_B is the Brewster angle, then these two rays have an additional phase difference of 180° (see the very useful descriptions in Können's book [128]). This does not occur for the corresponding rays in the common rainbow, so it is clear that constructive interference in the latter is associated with destructive interference in the polarized rainbow, and this is consistent with the numerical calculations of Khare and Nussenzveig [59]. Note also that this contrast will *not* be present in the secondary bow because of the two internal reflections in the drop.

Near the Brewster angle the Fresnel coefficient of reflection for parallel-polarized light varies substantially (in relative terms), so that the wave amplitude cannot be considered constant along the wavefront, even near the rainbow angle. This contrasts with the case for perpendicular polarization, for which the Fresnel coefficient while varying, is never zero and so (in relative terms again) the wave amplitude may be treated as constant to a first approximation. This is the basis of the Airy approximation and while not completely satisfactory, provides a reasonable representation of the common rainbow as we have seen.

The expression for the cubic wavefront has been noted earlier. A fuller account of its derivation may be found in the book by Humphreys [18] to which we now refer. By considering small deviations from the angles of incidence and refraction corresponding to the Déscartes ray, and with judicious use of Maclaurin's theorem, Humphreys deduces that in a coordinate system centered on the point of inflection (see Eq. (2.3) above, where $h = 3c$)

$$y = \frac{h}{3a^2}x^3 \,, \tag{6.5}$$

where in terms of the *order* n of the rainbow and the refractive index N

$$h = \frac{(n^2+2n)^2}{(n+1)^2(N^2-1)}\left\{\frac{(n+1)^2-N^2}{N^2-1}\right\}^{1/2} \,, \tag{6.6}$$

a being as before the radius of the drop. For the primary rainbow ($n=1$) and $N=1.33$, $h \approx 4.9$; for the secondary rainbow ($n=2$), $h \approx 28$. Obviously h is not independent of N. Let I_B, R_B and I_R be the Brewster angles of incidence, reflection and the angle of incidence for minimum deviation respectively, and furthermore let $\alpha = i - I_R$ and $\theta_d = \theta - \theta_R$, where i and θ refer to general angles of incidence and scattering, respectively. Then after some tedious but elementary manipulations [18] we may write

$$x = a\alpha \cos I_R$$

and the integral superposition of the wave amplitude in the direction θ_d as

$$\mathscr{A} = \int_{-\infty}^{\infty} k(x)\sin(\omega t - \delta)\,\mathrm{d}x = A\sin\omega t - B\cos\omega t\ , \tag{6.7}$$

where $k(x)$ is the wave amplitude per unit length along the wavefront, ω is the angular frequency of the (monochromatic) light and δ is the phase shift for a wave in direction θ_d from position x with respect to a wave emanating from $x=0$ in the same direction. The expression for δ is ([18], see also Section 2)

$$\delta = \frac{2\pi}{\lambda}\left(\frac{hx^3}{3a^2}\cos\theta_d - x\sin\theta_d\right)\ , \tag{6.8}$$

where λ is the wavelength of the light. We define the intensity $\mathscr{I} = A^2 + B^2$ since (6.7) can be written in the form $\mathscr{A} = (A^2+B^2)^{1/2}\sin(\omega t - \phi)$.

For the common rainbow, the Airy assumption is that $k(x)$ is uniform throughout (and chosen here as unity without loss of generality), and $B=0$ since δ is an odd function of x, so for small values of θ_d the standard Airy intensity is obtained as A^2 where

$$A = \int_{-\infty}^{\infty} \cos\frac{2\pi}{\lambda}\left(\frac{hx^3}{3a^2} - x\theta_d\right)\mathrm{d}x\ . \tag{6.9}$$

The situation is more complicated for the polarized rainbow. For the primary rainbow $A_\perp$ and $A_{||}$ refer to the amplitudes of light with the **E** vector perpendicular and parallel to the plane of incidence, respectively. Then by Fresnel's equations (for unit amplitude incident light)

$$A_\perp = \left[1 - \frac{\sin^2(r-i)}{\sin^2(r+i)}\right]\left[\frac{\sin(r-i)}{\sin(r+i)}\right] \tag{6.10}$$

and

$$A_{||} = -\left[1 - \frac{\tan^2(r-i)}{\tan^2(r+i)}\right]^2\left[\frac{\tan(r-i)}{\tan(r+i)}\right]\ . \tag{6.11}$$

Note again that $A_{||}$ changes sign if i passes through I_B. If $\gamma = i - I_B$, $\varepsilon = r - R_B$ then using Snell's law it follows for small γ and ε that

$$\varepsilon = \frac{\cos I_B}{N\cos R_B}\gamma = \frac{\gamma}{N^2}$$

from which

$$k(x) = \frac{A_{||}}{A_\perp} \simeq \frac{\gamma+\varepsilon}{\cos^3(I_B - R_B)}$$

or

$$k(x) \simeq \frac{(1+N^{-2})\gamma}{\cos^3(I_B - R_B)} \approx 1.77\gamma = \frac{1.77}{a\cos I_R}(x + x_0)\ , \tag{6.12}$$

where $x_0 = a(I_R - I_B)\cos I_R$. Upon substituting all these into the expressions for A and B we find the expected results for the common rainbow, i.e. $B = 0$ and

$$A = 2\pi \left(\frac{a^2\lambda}{2\pi h}\right)^{1/3} \mathrm{Ai}(z)\,, \tag{6.13}$$

where

$$uz = -\frac{2\pi x\theta_d}{\lambda}, \quad u^3 = \frac{2\pi h x^3}{\lambda a^2}\,,$$

while for the polarized rainbow

$$A = 2\pi(1.77)(I_R - I_B)\left(\frac{a^2\lambda}{2\pi h}\right)^{1/3} \mathrm{Ai}(z) \tag{6.14}$$

and

$$B = -2\pi\frac{1.77}{a\cos I_R}\left(\frac{a^2\lambda}{2\pi h}\right)^{2/3} \mathrm{Ai}'(z) \tag{6.15}$$

(the prime referring to derivative with respect to z). The variable z may also be written as

$$z = -\left(\frac{4\pi^2 a^2}{h\lambda^2}\right)^{1/3} \theta_d. \tag{6.16}$$

If the intensity distribution for the common rainbow is expressed in arbitrary units as $\mathscr{I}_1 = [\mathrm{Ai}(z)]^2$, then for the values for N ($\frac{4}{3}$) and λ (0.6 μm) chosen in [126], the intensity of the polarized rainbow is

$$\mathscr{I}_2 = [1.77(I_R - I_B)\,\mathrm{Ai}(z)]^2 + \left(\frac{a^2\lambda}{2\pi h}\right)^{2/3}\left[\frac{1.77}{a\cos I_R}\,\mathrm{Ai}'(z)\right]^2$$

or

$$\mathscr{I}_2 = 0.0376\mathscr{I}_1 + 0.232a^{-2/3}[\mathrm{Ai}'(z)]^2\,, \tag{6.17}$$

where $z = -4.92a^{2/3}\theta_d$ if θ_d is expressed in degrees. The intensities $\mathscr{I}_1$ and $\mathscr{I}_2$ are shown in Fig. 22 for $\beta = 2\pi a/\lambda = 1500$ (corresponding to a drop size of ∼0.14 mm for the chosen value of λ). The results agree well with the numerical calculations of Khare and Nussenzveig [59]. The shift between locations of maxima for the two polarizations remains present for much smaller drop sizes also [126], although the Airy theory can give qualitative information only in this regime.

Können and de Boer also use the asymptotic forms for the Airy function and its derivative to determine an upper bound for the droplet size where the maxima and minima are interchanged for the two polarization directions. For the parameters they choose, the two intensity distributions are in phase if $a > 0.6$ mm and out of phase if $a < 0.6$ mm. It is interesting to note from Eq. (6.14) that when $I_R = I_B$, i.e. the Déscartes ray is reflected at the Brewster angle, then $\mathscr{I}_2$ is proportional to $[\mathrm{Ai}'(z)]^2$, and is out of phase with the common rainbow at all droplet sizes. This can only occur for $N = \sqrt{2}$ as noted earlier, but this effect is still noticeable for sufficiently small drops (i.e. $\lesssim 0.6$ mm).

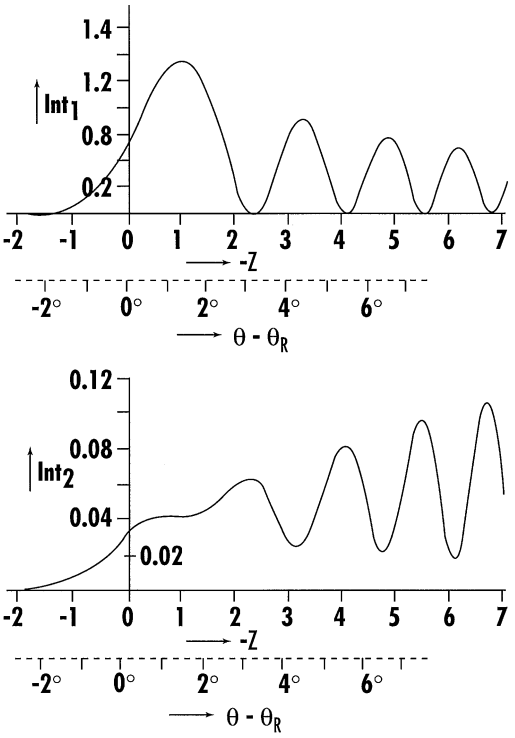

Fig. 22. The light intensity distribution as a function of $\theta - \theta_R$ for both rainbow polarizations: (a) the "common" rainbow and (b) the polarized rainbow. This corresponds, respectively, to the **E** vector (a) tangential and (b) radial with respect to the rainbow arc. The scattering angle is θ and the rainbow angle is θ_R. The calculations are for size parameter $\beta = 1500$ which corresponds to a droplet radius of approximately 0.14 mm in visible light. Note the differences in the positions of the supernumerary bows in the two figures. (Redrawn from [126].)

Returning to the choice $k(x) = 1$, made earlier for the classical Airy approximation to the common rainbow, it is of course possible to extend that theory by including the variation in the reflection coefficient for $A_\perp$ also. This in fact contributes only a small correction to the Airy theory: the term in $\mathrm{Ai}'(z)$ is never dominant. However, the minima in this case are never zero, which is also in agreement with the results reported in [59]. Similar calculations were performed for the secondary rainbow; deviations from the Airy theory are less pronounced, even for the polarized rainbow, where the expression for $\mathscr{I}_2$ contains terms in $\mathrm{Ai}^2(z)$ and $\mathrm{Ai}''(z)$ only; since these terms are proportional (see e.g. [129]) the intensity oscillations remain in phase with those of the common secondary rainbow.

Finally, the degree of polarization can be easily determined using this formalism. At the rainbow angle $\theta_d = 0$, so from Eq. (6.17) it follows that

$$\frac{\mathscr{I}_2}{\mathscr{I}_1} = R(a) = 0.036 + 0.00465a^{-2/3}\,, \tag{6.18}$$

which for large a compares well with the geometrical optics result [111] of $R = 0.039$, especially as $a \to \infty$. However, since R as written here increases monotonically as $a \to 0$, for small values of

a it is more appropriate to consider R as the ratio of the maxima of the intensity distributions for $z \neq 0$. By taking values of z equal to -3.25 (common rainbow) and -4.08 (polarized rainbow) as representative maxima (the first supernumerary bows) Können and de Boer obtain

$$R(a) = 0.0087a^{-2/3}\left[\frac{\mathrm{Ai}'(-4.08)}{\mathrm{Ai}(-3.25)}\right]^2 = 0.032a^{-2/3}\,, \tag{6.19}$$

$R(0.14) = 0.12$ (a is in mm) which is to be compared with the prediction of 0.054 from the above asymptotic-type geometrical optics result. This implies that the polarized rainbow is more readily observable for small droplets (e.g. as in fogbows) than is predicted by geometrical optics.

6.3. Comparison of theories

There have been several comparisons between some or all of the following: geometrical optics, Mie theory and the Airy approximation. Ungut et al. [43] studied the scattered light from transparent spherical particles with a range of diameters from 1 to 100 μm, and in the forward scattering angle range 0°–20°, and compared the results of Mie theory and geometrical optics. For the parameter ranges considered they concluded that the latter is a reasonable approximation to the former, especially if the particles are assumed to deviate from perfect sphericity in any cross-sectional area by a change in the diameter $\delta d \simeq \lambda$. More recently, Wang and van de Hulst [42] have given a detailed computational account of an efficient and accurate method of evaluating the Mie coefficients for $\beta \lesssim 50\,000$, corresponding to drop diameters up to $\sim$6 mm for visible light. They also compare their Mie theory calculations with the Airy approximation (generalized to an arbitrary number of internal reflections; this number is denoted by $p - 1$, $p \geqslant 1$, consistent with the notation in [6] and Section 5 above). Specifically they find that for drop diameters ~ 1 mm there is agreement in all details, including the polarization and the position and intensity of the supernumerary maxima; this agreement is still fairly good down to a diameter of ~ 0.1 mm. It transpires that the higher the value of the refractive index, the closer is the agreement between the two theories for the primary rainbow ($p = 2$) for drop diameters as small as 0.02 mm. They also investigate Alexander's dark band and conclude that in addition to the presence of externally reflected light ($p = 1$) there is a significant contribution from the $p = 6$ and 7 rainbows. The positions of these are very sensitive to the wavelength of the light.

We note some other points from this paper, the first being of historical interest. The authors point out that since 1957 (when Van de Hulst's book [7] first appeared) the large gap that existed between small β-values, for which Mie theory was computationally feasible, and large β-values, for which the Airy theory is valid, is much closer to being closed today. This gap was $30 \lesssim \beta \lesssim 2000$, and since that time (as detailed earlier) the complex angular momentum theory ([10] and many references therein) and the development of highly efficient computational schemes have helped to close the gap (from the upper and lower ends, respectively). Now, reliable Mie computations are feasible for β up to *at least* 50 000. Details of these can be found in [42].

We have alluded earlier to the curvature term $h/3a^2$ for the cubic wavefront approximation of Airy (see Eqs. (6.5) and (6.6)), where h is expressed in terms of the order of the rainbow and the refractive index. Note that the curvature of the wavefront is greater for higher values of p (and hence n) and smaller values of a (or, for light of a given wavelength, for smaller values

of β), so that the angular spacing between two successive supernumerary peaks increases with increasing p (or n) and/or with decreasing a. In both [7] and its relatively recent extension [42] an expression is given for the rainbow peak intensity: in particular there is a size dependence of $\sim x^{7/3}$ which is consistent with an amplitude dependence of $\sim x^{7/6}$ in Eq. (2.4) of [61], so there is no discrepancy in this regard between the two approaches, i.e. between the Airy theory and the complex angular momentum theory. For a drop with diameter 3.2 mm, the authors also found excellent agreement between the Mie and the Airy theories (such a large drop would not, of course retain its spherical shape in the atmosphere). The primary rainbow was found to be ≈ 7.7 (red light) to ≈ 8.4 (violet light) times more intense than the secondary rainbow, and all the Airy maxima up to about the tenth supernumerary bow were at the predicted angles and had the correct intensities. Also the angular separation of the rainbows was found to decrease as the light shifted from violet to red: the shift in the secondary ($p=3$) being twice that in the primary ($p=2$) rainbow.

For a drop diameter of 0.1 mm there is still (surprisingly, as the authors point out) good agreement between the Mie and Airy intensities for the first maxima in cases $p=2$ and 3. Amongst other features, the dark band is still seen in the violet but diminishes in the red, and the Airy theory predicts somewhat larger shifts in the higher supernumerary bow positions for $p=2$ and 3. Tests for even smaller sizes (and various refractive indices) indicate that the Airy theory predicts well the features of the first peak of the primary rainbow. In the dark band, where the intensity may drop to below 0.5% of that in the primary peak, the intensity is well explained as contributions from the $p=0$ (external reflection) and the $p=6$ and 7 rainbows. The overall conclusion drawn by Wang and van de Hulst is that the Airy theory provides simple and reliable information over a wide range of sizes, and furthermore this may be useful in determining properties of particles in laboratory, meteorological or astronomical experiments.

Related to these results is a set of observations made earlier by Sassen [1] in connection with angular scattering and rainbow formation in *pendant drops*. These are laboratory-produced near-spherical drops with certain similarities to the shape of distorted raindrops. The latter become increasingly oblate and unstable with increasing size, and tend to orient themselves uniformly in space as they fall. Measurable raindrop distortions have been noted for drops as small as 0.3 mm in diameter [130] and are significant for diameters $\gtrsim 1$ mm. As might therefore be expected, rainbow features often vary with the position of the generator of the rainbow "cone" as the observer views different parts of the rainbow. This is because of the change in geometric cross-section of the drop as the angle of observation changes (for more on this see Section 6.4). Such variations have been observed [131]. Obviously in any given situation it is expected that there will be a distribution in raindrop size. It is frequently the case that rainfall from convective showers (conditions under which rainbows are frequently observed) contain "supermillimeter" drops, so information on near-spherical drops may well help to explain occasional anomalous features (i.e. unlikely to occur for purely spherical drops) such as reports of tertiary rainbows [14,132–134] at 42° (while standard theory predicts their existence, they are considered too faint and too close to the sun's position to be observable, although these reports might refer to reflected or reflected-light rainbows [4]).

The pendant drops discussed in [1] are more akin in shape to the familiar "teardrop" shape so beloved of cartoonists. Nevertheless, they combine spherical and non-spherical scattering mechanisms that are considered relevant to those for aspherical raindrops for the following

two reasons: (i) they retain a circular cross section in a horizontal plane and (ii) display a reduction in surface curvature (compared to a spherical drop) along the vertical sides of the drop. These are exactly the same features displayed by the "hamburger-bun" shapes of larger raindrops (see further references in [1] for details of numerical and theoretical simulations of realistic raindrop shapes). Angular scattering experiments revealed that such near-spherical drops generate exceptionally well-defined primary, secondary and associated supernumerary bows, but unexpectedly strong higher-order bows as well, perhaps because of the reduced curvature in the central elongated regions, thus providing relatively more surface area within the drop available for generating rainbow rays. As with the later work of Wang and van de Hulst [42], it was found that the precise angular positions of at least the first 10 intensity maxima for the primary and secondary rainbows were predicted by the Airy theory. Interestingly, it is claimed that the theory fails to account for the position of the sixth-order rainbow ($p=7$), but in [42] it was found that this rainbow, along with that for $p=6$ is found just inside Alexander's dark band.

Dave [135] presented results of Mie computations for large non-absorbing spheres and compared them with those corresponding to geometrical and physical optics (i.e. including diffraction effects). This was done for monochromatic radiation of wavelength $\lambda=0.4$ μm and water spheres of radii $a=6.25, 12.5,$ 25 and 50 μm, corresponding to size parameters, respectively, of $\beta=31.25\pi, 62.5\pi, 125\pi$ and 250π, well below the region of applicability (and prior to the emergence of the complex angular momentum theory) of the classical ray and Airy theories [7]. Dave considered three regimes using geometrical optics: forward scattering, the rainbow region and the glory region. An approximate value for the scattered intensity in the first regime can be found by replacing the sphere by a circular disk of radius a [7]. This unpolarized diffracted radiation has intensity

$$I_f=\beta^4\left[\frac{J_1(\beta\sin\theta)}{\beta\sin\theta}\right]^2,$$

relative to the incident radiation intensity. The maximum intensity of $\beta^4/4$ occurs of course at $\theta=0^\circ$; successive minima and maxima occur as θ increases away from zero. If the amplitudes of the reflected and refracted are comparable with that of this diffracted component, there can be a significant increase in intensity and interference between these components. (This is sometimes referred to as "anomalous diffraction" mentioned earlier; further details may be found in [7].) Similar interference phenomena may of course occur in the rainbow region on either side of the Déscartes ray of minimum deviation, if two such rays emerge in parallel directions—leading at this level of description to an explanation of the supernumerary bows. Apart from van de Hulst's suggestion of surface wave involvement as responsible (in part) for the glory, there was at that time no satisfactory detailed account of this phenomenon (the work by Nussenzveig was published shortly after [135] appeared). This paper [135] provides many figures and related results; it was concluded on the basis of the four size parameters studied that ray optics and Mie theory agree well only if $\beta\sim 800$ (and presumably $\beta\gtrsim 800$, though Dave's largest value was $\beta=785.4$). He concluded that a satisfactory explanation of the glory should await a more substantial understanding of surface waves and the interaction between radiation fields (true), and noted also the lack of a simple explanation for (i) anomalous diffraction in the forward direction (for $\theta>2^\circ$) and (ii) the appearance of several distinct maxima outside the regions of primary and secondary rainbows.

The four approaches to the theory of the rainbow discussed so far are, in historical order: geometrical optics, Airy's approximation, Mie theory and complex angular momentum theory. Crudely, the first of these is independent of size of the drop, but only valid for large size parameters, e.g. $\beta \gtrsim 5000$. Airy's approximation is valid for $\beta \gtrsim 5000$ and for $|\theta - \theta_R| \lesssim 0.5^\circ$; the Mie solution is exact but requires the summation of a number of partial waves greater than β (this is no longer the problem it used to be for $\beta \gtrsim 100$ as noted above in connection with [42]). The complex angular momentum theory provides a very good substitute for Mie theory if $\beta \gtrsim 50$ (in fact it is still very accurate for $\beta \gtrsim 1$ [10]) but is mathematically very technical. It was with these restrictions in mind that Mobbs [136] developed another theory, based on Huygens' principle (as is Airy's theory) in which he states that the physics is straightforward and only simple computation is required for a substantial range of drop sizes. A virtual wavefront leaving the drop after one internal reflection is examined, as in Airy's theory, and a set of equations is developed which completely defines this wavefront. Mobbs also defines, in a similar manner, a virtual wavefront for the externally reflected ray. The distributions of amplitude (for both polarizations) depend on the drop geometry, the wavefront and the Fresnel amplitude coefficients. Sixteen Fresnel zones were constructed across the wavefront, each of which has an amplitude (uniform across the zone) characteristic of the total light flux passing through that zone. Thus the wavefronts consisted of 16 annular rings around the (obvious) central axis (z) of the drop; the boundaries corresponding approximately to equally spaced values of the angle of incidence. The basic mathematical features of this approached will be briefly summarized here (and in the opinion of this author, with regard to Mobbs' earlier comment about the CAM theory, his calculations are not entirely without mathematical complexity either!).

Suppose that the nth zone, $1 \leqslant n \leqslant 16$, lies in the angle of incidence interval $[i_n, i_{n+1}]$. The area orthogonally presented to the incident light between these angles is readily shown to be

$$\mathsf{a}_n = \frac{\pi a^2}{2}(\cos 2i_n - \cos 2i_{n+1}) \tag{6.20}$$

and the area of the zone is, in a cylindrical polar coordinate system,

$$A_n = 2\pi \int_{i_n}^{i_{n+1}} R \left[1 + \left(\frac{\mathrm{d}z}{\mathrm{d}R}\right)^2\right]^{1/2} \frac{\mathrm{d}R}{\mathrm{d}i}\,\mathrm{d}i\ . \tag{6.21}$$

Define

$$G_{p,n} = \left(\frac{\mathsf{a}_n}{A_{p,n}}\right)^{1/2}\ ,$$

where the subscript p refers to externally reflected light ($p = 0$) and internally reflected light for the primary rainbow ($p = 2$). The rainbow rays undergo two refractions (transmissions) and a single reflection. So using the Fresnel amplitude coefficients the amplitude corresponding to the nth zone is given by

$$\mathscr{A}_{2,n}^{(1)} = G_{2,n}\frac{\sin(2i)\sin(2r)\sin(i-r)}{\sin^3(r+i)} \tag{6.22}$$

for polarization 1 (magnetic) and

$$\mathscr{A}_{2,n}^{(2)} = G_{2,n}\frac{4\sin(2i)\sin(2r)\tan(i-r)}{[\sin(2i)+\sin(2r)]^2\tan(r+i)} \tag{6.23}$$

for polarization 2 (electric). The corresponding amplitudes for externally reflected light are

$$\mathscr{A}_{0,n}^{(1)} = -\,G_{0,n}\frac{\sin(i-r)}{\sin(i+r)} \tag{6.24}$$

and

$$\mathscr{A}_{0,n}^{(2)} = G_{0,n}\frac{\tan(i-r)}{\tan(i+r)}\,. \tag{6.25}$$

Mobbs then applies Fresnel's formulation of Huygen's principle by examining the path difference between rays that meet the wavefronts such that $\tau = \pi - \theta_R$. The disturbance from an element of area $\mathrm{d}S$ of the wavefront near a point P, a large distance s from the drop can then be expressed in terms of the $\mathscr{A}$ coefficients above, some complex exponential factors in time and space, and an obliquity factor, details of which may be found in [136]. By suitable formal integrations, using results of the type [105]

$$\int_0^{\pi}\cos(n\theta - x\sin\theta)\,\mathrm{d}\theta = \pi J_n(x), \quad n = 0,1,2,\ldots \tag{6.26}$$

and after considerable reduction, the amplitude coefficients $S_p^{(j)}(\theta,\beta)$ can be written in terms of the rather formidable expression

$$\begin{aligned} S_p^{(j)}(\theta,\beta) = {} & \sum_{n=1}^{16}\pi\mathscr{A}_{p,n}^{(j)}(\lambda^{-1}\tan\tau)\mathrm{e}^{-\mathrm{i}kl_p\sin\tau}[R_pJ_1(kR_p\sin\tau)\mathrm{e}^{\mathrm{i}kz_p\cos\tau}]_{R_{p,i_n}}^{R_{p,i_{n+1}}} \\ & -\sum_{n=1}^{16}\pi\mathscr{A}_{p,n}^{(j)}(k\lambda^{-1})\mathrm{e}^{-kl_p\sin\tau}\left(\int_{i_n}^{i_{n+1}}R_p\left\{\left[1+\left(\frac{\mathrm{d}z_p}{\mathrm{d}R_p}\right)^2\right]^{1/2}+\sec\theta\right\}\right. \\ & \left.\times\mathrm{e}^{\mathrm{i}kz_p\cos\tau}J_0(kR_p\sin\tau)\,\frac{\mathrm{d}R_p}{\mathrm{d}i}\,\mathrm{d}i\right)\,, \end{aligned} \tag{6.27}$$

where the path difference calculated is between light from the point $(l_p,0,0)$ and the point $P(R_p,\phi,z_p)$. The addition of the contributions from the internally and externally reflected light is carefully considered in terms of the phase difference ψ between the two; then the total amplitude coefficients $S^{(j)}(\theta,\beta)$ are

$$\begin{aligned} S^{(j)}(\theta,\beta) = {} & \mathrm{Re}(S_2^{(j)}) + \mathrm{Re}(S_0^{(j)})\cos\psi - \mathrm{Im}(S_0^{(j)})\sin\psi \\ & +\mathrm{i}[\mathrm{Im}(S_2^{(j)}) + \mathrm{Re}(S_0^{(j)})\sin\psi + \mathrm{Im}(S_0^{(j)})\cos\psi]\,. \end{aligned} \tag{6.28}$$

Concerning the validity of this approach, Mobbs makes an interesting point: its validity is based on the assumption that the wavefronts can be determined by constructing rays (orthogonal

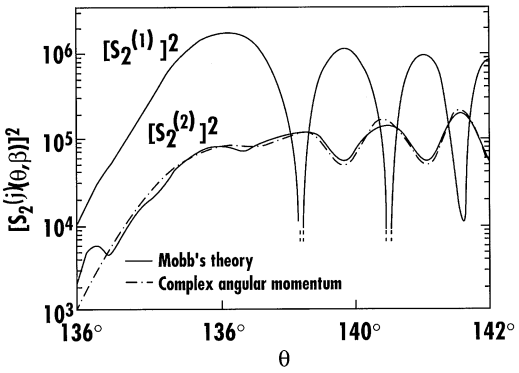

Fig. 23. The light intensity distribution as a function of θ for polarizations 1 and 2 (excluding externally reflected light) for $\beta = 1500$, in which the results of the complex angular momentum theory and Mobbs' theory are compared. Agreement for polarization 1 is excellent and the differences for polarization 2 are small. (Redrawn from [136].)

to the wavefronts), and *on this basis* Tricker [8] has attempted to determine the validity of the Airy approximation. In fact his argument uses the Fresnel zone concept also, but Mobbs points out that the fundamental assumption in the Airy theory is (instead) that the amplitude is uniform across the wavefront. However, his result *is* applicable here: the radius of the drop should be large compared with 4λ, or $\beta \gg 8\pi$. In practice, Mobbs also found an upper limit to β determined by the difficulty of evaluating the sixteen integrals for large values of β. For the range of scattering angles examined, namely $136^\circ \leqslant \theta \leqslant 142^\circ$, the agreement between this theory and the complex angular momentum approach is very good for $\beta = 500$ and 1500 (the latter excluding externally reflecting light, which superimposed small rapid oscillations on the main features of the internally reflected light; see Fig. 23). Even for $\beta = 50$ the agreement is fairly good for both polarizations. This approach represents an interesting alternative to the complex angular momentum theory, and a feasible one when surface waves and higher-order reflection terms are not important; however the power, generality and beauty of the latter theory (not to mentions its broad range of accuracy) cannot be refuted, even if it is considered, in some applications, to be the proverbial "sledge-hammer cracking a nut".

Lynch and Schwartz [137] (see also [138]) used Mie theory to study the optics of rainbows and fogbows (i.e. rainbows formed in clouds and fogs, as opposed to rain showers) for monodisperse drop radii in the range 3–300 μm (for light of wavelength $\lambda \sim 0.5$ μm this corresponds to $38 < \beta < 3800$). This upper limit was chosen because of the tendency of larger drops to oscillate or become aerodynamically distorted. Furthermore, drop sizes below 3 μm exhibit no discernible rainbow. The range of scattering angles in the calculations was $110^\circ < \theta < 160^\circ$. In their calculations they included a realistic solar illumination spectrum (as modified by the earth's atmosphere) and the finite angular size of the sun. Fogbows tend to be paler than rainbows (being formed from smaller drops, $a \sim 50$ μm), and are also wider; the maximum intensity occurs a few degrees higher than the 138° scattering angle for the primary rainbow. Amongst other goals, the authors were interested in obtaining a quantitative understanding of the transition between rainbows and fogbows. They calculated the scattering intensity function $I(\theta)$ from an appropriate integral involving the ratio of the scattered-to-incident radiation for unpolarized

light; this latter quantity is the scattering matrix element $S_{11}(a,\lambda,\theta)$ in their notation. (To use their results to model a rainbow formed in a rain shower would require $I(\theta)$ to be weighted with the drop distribution $n(a)$ and an additional cross-section factor.) In particular, for the ranges considered the authors found that:

(i) the rainbow attains a minimum angular width for $a = 100\ \mu$m. Indeed, as a decreases the widths of both primary and secondary bows increase, their contrasts (as defined in [137]) decrease and their scattering angles of maximum brightness move away from Alexander's dark band.

(ii) the supernumerary bows for the primary rainbow attain maximum visibility for $a = 50\ \mu$m; they grow wider and further apart as a decreases. For values of a above about 56 μm their separation is less than the sun's angular diameter of 0.5° and their amplitudes are greatly reduced as a result of the concomitant smoothing effect; below about 36 μm the wider bows lose contrast. There are other factors influencing this of course, such as drop size, shape and distribution [139,140].

(iii) Alexander's dark band maintains a constant angular width of about 9.5°, apparently because the primary and secondary bows, while widening grow away from the dark band with decreasing drop size; its angular width, however, is a function of drop size;

(iv) the maximum linear polarization of both types of bow is also a function of drop size, reaching a maximum of 90% for the primary bow and 50% for the secondary bow when $20\ \mu\text{m} < a < 100\ \mu$m, and

(v) the secondary fogbow is observable only when $a > 10\ \mu$m. The authors further suggest that a fogbow be defined as any rainbow formed in drops for which $a < 35\ \mu$m.

In 1983, Fraser published a paper with the intriguing title "Why can the supernumerary bows be seen in a rainshower?" [141]. In order to see why there should be such a question in the first place a little background will be useful. The supernumerary bows are essentially an interference phenomenon, as Young pointed out in the Bakerian Lecture of 1802 [142]. The Airy approximation predicts, as we have seen, that the spacing between adjacent bows, and the width of the primary bow should increase with decreasing drop radius. Since in reality however, normal rainshowers have a spectrum of drop radii (spanning more than an order of magnitude [143]) it would seem that no consistent pattern of supernumeraries should be seen, contrary to what is frequently observed. These bows form near the top of the bow, and so any explanation of their existence should incorporate the fact that the light in this portion passes through the vertical cross section of often-distorted drops.

While the standard ray approach (supplemented by the wave concept of interference) is a perfectly reasonable way to explain the existence of supernumerary bows for two sets of rays emerging in the same direction from monodisperse drops, Fraser points out that "a more compelling visualization is obtained if the rays are replaced by waves", and this is certainly the case (see Fig. 4). The ray picture cannot of course account for different drop sizes, whereas the wave picture does, and indicates that large drops yield tightly spaced supernumeraries with narrow maxima, while small drops yield broader maxima with wider spacing. Fraser uses the Airy approximation in his calculations, pointing out that it does adequately identify the positions of the maxima for drop radii in excess of 0.05 mm (50 μm), which is certainly outside the fogbow range according to [137]. The angle of minimum deviation is affected by drop distortion: Möbius showed that the deviation angle of the Déscartes ray would in fact be increased [144].

Fraser used ellipsoidal drops in his calculations, noting that for the smaller drop sizes of interest to him, this is an excellent approximation. It appears that the minimum angle of deviation is also weakly dependent on solar elevation (presumably because of the different-sized drops that contribute to the light distribution at the top of the rainbow), but only one such elevation (0°) was used since this effect is negligible here. Since smaller drops yield larger angles for the Déscartes ray, it is apparent that there is an intermediate drop size that will yield a *minimum deviation* angle $\mathscr{D}$, i.e. in formal terms

$$\mathscr{D}(a) = \min_a \{\min_i D(i)\}$$

and it these drops which cause the supernumeraries that are observed. This drop size turns out to be in the range 0.1–0.2 mm for the primary bow and in the range 0.2–0.3 mm for the first two supernumeraries. No other drops make a significant contribution. On the basis of his results Fraser predicts that if supernumerary bows are seen in a rain shower, their separation will be $\sim 0.75^\circ$, independent of the particular size distribution of the raindrops, and found from photographs typical separations of $\sim 0.70^\circ$.

In addition, Fraser [145] has pointed out that the explanation by Voltz [150] for the absence of supernumerary bows near the base of the primary rainbow (when the sun is low in the sky) also applies to the intensification of the red light contribution in the same location. (He also states that the rainbow is a fine example of a phenomenon in which theory tends to color observations.) Near the top of a rainbow the red color is much less intense that towards the base, i.e. the more nearly vertical portions of the bow. It is only in these regions that the incident light, which is deviated in a more nearly horizontal plane, encounters a circular cross section in a flattened drop. Large drops will still contribute to the rainbow, whereas such drops in the vertical plane displace the rainbow angle inward toward the antisolar point because of their distorted cross section in that plane (this was in fact first demonstrated by Möbius [144]: elliptical drops lower the top of the bow. Voltz estimates that drops smaller in radius than 0.25–0.5 mm are sufficiently spherical to contribute to all portions of the rainbow; larger than this they will contribute only to the vertical portions, which in turn explains why the rainbow is brightest there—the larger drops substantially increase the intensity—and why the supernumeraries are absent—the intensity maxima overlap. The complete spectrum of "colors of the rainbow" may only occur near the foot of the bow (along with the pot of gold, etc.).

6.4. Non-spherical (non-pendant) drops

Subsequently, and perhaps inspired by the work of Fraser, Können discussed the same type of phenomena in more mathematical detail, but specifically for the *secondary rainbow* [147]. They (i.e. supernumeraries for the secondary bow) are seen only on extremely rare occasions [148]. Können points out that the Déscartes ray for the secondary rainbow shows very little dependence on the drop oblateness (for visible light) due to the fact that oblateness-induced changes in the two refractions and two interior reflections essentially cancel out, and so in view of Fraser's explanation for supernumeraries associated with the primary bow, they are unlikely to be seen for the secondary bow (for most solar elevations there is now no stationary point for $\mathscr{D}(a)$). However, for solar elevations exceeding about 35° the flattening of the drops can result in a

minimum for $\mathscr{D}(a)$, and this may be sufficient to cause the first supernumerary—particularly for red light—to become visible (see [147] for a detailed explanation). Können approximates a falling drop by an oblate spheroid, with its two semi-major axes of length a horizontal and its semi-minor axis of length b vertical. The equivalent (i.e. same volume) spherical drop radius is $r=(a^2b)^{1/3}$ and for $r<1.2$ mm this is a reasonable approximation (and an excellent one if $r<0.5$ mm). Following Möbius [144] the drop ellipticity is defined as

$$\rho=\frac{a-b}{a+b}\,. \tag{6.29}$$

For drops falling at terminal velocity (assumed here) Green [149] derived an approximate relationship between the ratio b/a and r. It is given by

$$ba^{-1}=[(4/17)(17B/4+1)^{1/2}+13/17]^{-3/2}\,, \tag{6.30}$$

where $B=\rho' r^2 g'/\sigma$ is called the Bond number; ρ', g' and σ are respectively the drop density, effective gravity ($g'=(\rho-\rho_{\mathrm{air}})\rho^{-1}g$) and surface tension coefficient. By substituting this into the expression above for ρ and expanding in a Maclaurin series about $r^2=0$, Können obtains the further approximation $\rho\simeq 0.05r^2$ if r is expressed in millimeters. He states that this simple result is useful for $r<1.5$ mm. According to [144] the deviation between the rainbow angles for spheroidal and spherical drops is

$$\Delta\theta_R\equiv\theta_R(\text{spheroid})-\theta_R(\text{sphere})\,,$$

i.e.

$$\Delta\theta_R=\frac{180^\circ}{\pi}16\rho\sin\beta\cos^3\beta\cos(2h-42^\circ)\,, \tag{6.31}$$

where here β is the angle of refraction of the Déscartes ray for spheres, h is the (angular) solar elevation and 42° ($=180^\circ-\theta_R(\text{sphere})$) is the angular distance from the bow to the antisolar point. (Können also notes a sign error, and significant consequences therefrom, present in [152]). For angles of observation other than for the top of the bow, (i.e. a non-vertical scattering plane) the corrections to ρ and h are also provided in [147]. In the visible range the angle β ($\simeq 40^\circ$) varies by only about 1°, so equation (6.31) is almost independent of wavelength. Furthermore, it is consistent with Fraser's conclusions that $\Delta\theta_R$ is a strong function of drop radius r but only a weak function of h ($h<42^\circ$). Thus to a good approximation,

$$\Delta\theta_R=13^\circ r^2\cos(2h-42^\circ)\equiv Cr^2, \tag{6.32}$$

where $10^\circ\leqslant C\leqslant 13^\circ$. The result corresponding to (6.31) for the secondary rainbow is

$$\Delta\theta_R=-\frac{180^\circ}{\pi}64\rho\sin\beta\cos^3\beta\cos 2\beta\cos(2h-51^\circ)\,, \tag{6.33}$$

where the negative sign results from changing the scattering angle interval to $[180^\circ,360^\circ]$, which enables the results for the primary and secondary bows to be presented in a more convenient and consistent manner. Note that in this case $\Delta\theta_R$ is zero when $\beta=45^\circ$; it is a straightforward

matter to show from geometrical optics that

$$\cos 2\beta = \frac{5N^2 - 9}{4N^2}$$

and so this occurs when $N = \sqrt{9/5} \simeq 1.342$, which for water corresponds to violet light. This of course means that now there *is* a strong dependence of $\Delta\theta_R$ on wavelength for light this wavelength, which is readily formulated using differentials. It turns out to be necessary to include terms up to ρ^2 for the secondary rainbow, the result corresponding to (6.32) above being

$$\Delta\theta_R = C_1 r^2 + C_2 r^4, \tag{6.34}$$

where the coefficients C_1 and C_2 are defined in [147]. The author further demonstrates that the stationary behavior for the secondary rainbow near its first supernumerary occurs if $r \simeq 0.7$ mm; this in turn leads to a yet simpler form for (6.34), namely $\Delta\theta_R = C' r^2$, where $C' = C_1 + 0.5C_2$. C' is a strong function of wavelength (via C_1) and solar elevation (via C_2).

For falling drops not exceeding about 1.8 mm in radius ($\rho \approx 0.1$), the Airy theory can still be applied if the rainbow angle for spheres is replaced by that for spheroids using the expression for $\Delta\theta_R$. For larger eccentricities this is not the case, because the Airy intensity distribution (which may be thought of as corresponding to a *fold diffraction catastrophe*) may be transformed into a *hyperbolic umbilic diffraction catastrophe* in any given drop of large enough ρ. This is because two additional rays with skew paths through the drop can also contribute to the rainbow interference pattern [151]. (Diffraction catastrophes will be discussed in more detail in Section 7). This has been observed to occur for the primary rainbow when the scattering plane is approximately horizontal [147]. For reasons discussed in that paper, it is conjectured that no such transformation occurs for the secondary rainbow (though it may occur for other refractive indices). Furthermore this effect can be neglected if there is a broad distribution of drop sizes, because for the Marshall-Palmer distribution [152] (for example) less than 1% of the drops exceed 1.5 mm in radius, so the effect is swamped by that from the smaller drops.

Returning to the regime for which the analytic form from the Airy theory is still appropriate, the following expression arises for the intensity distribution near the secondary rainbow angle θ_R ($\simeq 231^\circ$)

$$I(r,\theta) \propto r^{7/3}\,\mathrm{Ai}^2[f(r,\theta;\lambda)]\,, \tag{6.35}$$

where $f(r,\theta;\lambda)$ (not to be confused with the scattering amplitude) is defined as:

$$f(r,\theta;\lambda) = -3.1\left(\frac{500r}{\lambda}\right)^{2/3}[\theta - \theta_R(\text{sphere}) - Cr^2]\,. \tag{6.36}$$

The same functional form applies to the primary rainbow also with some differences in the argument of the Airy function. The above argument is stationary at $r = r_s$ where

$$4Cr_s^2 = \theta - \theta_R(\text{sphere})\,.$$

If $C > 0$, the stationary behavior of f is in the oscillatory region of the Airy function, which means that the integral of Eq. (6.35) over a broad drop-size distribution also has an oscillatory

component, and the spacing between consecutive maxima in the intensity distribution can be inferred from this. The angular spacing $\theta_{1,2}(\lambda,h)$ between the first and second supernumerary bows is given in [147] by the expression

$$\theta_{1,2}(\lambda,h)=0.71^{\circ}C^{1/4}(\lambda,h)\left(\frac{\lambda}{640}\right)^{1/2}. \tag{6.37}$$

For the primary bow with $C=13^{\circ}$, $\theta_{1,2}=0.88^{\circ}$ at $\lambda=640$ nm, consistent with the results of Fraser [141]; for the secondary bow with $C=1^{\circ}$ at the same wavelength, $\theta_{1,2}=0.71^{\circ}$.

In order to simulate the intensity distribution of a typical secondary bow produced by a rain cloud, Können assumed the drop-size distribution to be of the Marshall-Palmer type [152]

$$\frac{\mathrm{d}n}{\mathrm{d}r}\propto\exp(-6r)\,,$$

where n is the number of particles of radius r (in mm). This relation was then multiplied by (6.35) and integrated over both r and the solar disk. Amongst Können's conclusions were the following: (i) the intensity of the main peak of the secondary bow is a factor of 4 greater at the top than at the base; (ii) at low solar elevations, no supernumeraries (for the secondary bow) should ever be observed; and (iii) at higher solar elevations the first supernumerary should be observable, especially at the red end of the visible spectrum if a filter is used. The primary reason for the difficulty of observing the supernumerary bows is that the secondary rainbow angle is much less sensitive to the flattening of rain drops than is the corresponding angle for the primary bow. At low solar elevations, an increase in drop oblateness shifts the secondary rainbow pattern toward smaller deviation angles, which is in the wrong direction for producing supernumeraries. At higher solar elevations this is not necessarily the case, but even then the background light may swamp an otherwise visible supernumerary bow.

Lock [153] discusses the observability of supernumerary bows and atmospheric glories, pointing out that one of the chief factors influencing this is the spatial coherence width of sunlight as received at the earth. Sunlight in fact possesses little spatial (or temporal) coherence, and this places limits on both the number of supernumerary bows that can be produced adjacent to the primary (for a given size of drop) and on the size of drops that can produce a glory. It is shown by Wood [154] that the diameter of the area over which the phase is constant is $d\approx R\lambda/2r$ for an object of apparent diameter $2r/R$ (r being the actual radius if circular) observed at a distance R. For sunlight with $\lambda\approx 0.5$ μm, $d\approx 50$ μm, so if the distance between the incident light rays exceeds d then the lack of coherence results in no interference phenomena, i.e. supernumeraries or glories. The idea behind Lock's investigation is to test the consistency of this approach with the observations by calculating the optical path difference between rays exiting the drop parallel to each other. According to these calculations, only the first two or three supernumerary bows have incident rays that are separated by less than the value of d given above (in fact Lock uses $d\approx 40$ μm, so only the first supernumerary bow meets his criterion). Interestingly, the intensity distribution calculated for scattering by a sphere of radius 250 μm with partial coherence qualitatively resembles the pattern for diffraction by an edge, as opposed to the standard Airy pattern for coherent radiation (cf. Figs. 1 and 2 in [153]).

As far as the glory is concerned, Lock demonstrates that the peak observability of the glory occurs for droplets with radii in the range 10–20 μm. Some pertinent comments are made

about the nature of the surface wave effect in this phenomenon that are usefully stated here. It is known that the wave-interference factor for glory backscattering (exiting the drop as a toroidal wave front) is [7,50]

$$F = (c_1 + c_2)^2 J_0^2\left(\frac{\beta r}{z}\right) + (c_1 - c_2)^2 J_2^2\left(\frac{\beta r}{z}\right) , \tag{6.38}$$

at a distance z behind the scatterer, r being the radial coordinate. The quantities c_1 and c_2 are the products of the transmission and internal reflection coefficients for the incident transverse magnetic and electric fields, respectively. (There is also a geometric scaling factor $2\pi a\,\Delta r \propto a^{4/3}$ associated with the edge region Δr of the drop.) As noted earlier, for water drops there must be a surface wave traversing a 14° arc to account for the glory, so the coefficients c_1 and c_2 must also include terms to account for the creation and propagation of surface waves, and so they will not be constants in fact, for they will depend on the droplet size. It is shown in [47] that damping of surface waves increases with drop size, with

$$c_1 \approx c_2 \propto \exp(-0.4\beta^{1/3}) .$$

For a perfectly coherent light source the contribution to the intensity of the glory from the surface wave is proportional to [47] $\beta^{8/3}\exp(-0.8\beta^{1/3})$ which has a relative maximum at $\beta = 1000$ or $a \approx 88$ μm. This may be understood as follows: If a increases (below this maximum value), the geometric scaling factor likewise increase and this allows more rays to contribute to the glory scattering. Above this maximum value of a the damping of the surface waves reduces the contribution [153]. Based on estimates of Van de Hulst [7], Lock uses in his calculations the forms $c_1 = -0.2\exp(-0.4\beta^{1/3})$ and $c_2 = 1.0\exp(-0.4\beta^{1/3})$. Again, significant quantitative differences are found for the case of partial coherence versus coherence in the spatial domain, and it appears that in the presence of a wide distribution of drop sizes the glory is primarily produced by a small range (10–20 μm).

Since many rainbows are formed in association with thunderstorms, it is natural to enquire as to the effects on them of strong (vertical) electric fields. A study by Gedzelman [155] does just that, noting that any force that can alter the shape of a raindrop will also affect the rainbow. Indeed, rainbows have been observed to quiver during thunderstorms [148]. A vertical electric field is known to polarize the drops and stretch them vertically [156,157] which of course is the opposite effect to that of aerodynamic forces on falling drops. The extremely large vertical fields in the vicinity of thunderstorms ($E \gtrsim 10^5\ \mathrm{V\,m^{-1}}$) may produce significant and hence measurable changes on drop shape. Any net charge present on the drops also affects their shape by reducing the effective surface tension, thus enhancing any propensity to depart from sphericity, though this is realistic only for the largest drops, and since these do not contribute to the supernumeraries at the top of the bow, electrical charge should not affect the visible appearance of the rainbow significantly.

If electric charge is neglected, the formula used in [155] for the eccentricity of a raindrop in the presence of a vertical electric field of intensity E is (on correcting a minor typographical error)

$$e = \frac{a-b}{a+b} = \frac{3r(2\rho g r - 3\epsilon E^2)}{16\sigma} , \tag{6.39}$$

where r is the drop radius, ρ is the effective density for water in air, ε is the permittivity and σ is the surface tension of water under these conditions. The quantities a and b are, respectively, horizontal and vertical semi-axes. This formula is valid only for small eccentricities and hence small drops (see [156] for a more general formula for e) because it neglects the progressive flattening of the base for larger drops; nevertheless it is a reasonable approximation for drops of radius $\lesssim 1$ mm. Note that drops with radii smaller than $3\varepsilon E^2/2\rho g$ will be prolate spheroids ($e < 0$); for $E = 4 \times 10^5$ V m^{-1} this implies $r = 0.2$ mm. Fields of this magnitude are only to be found in thunderstorms in general; the typical fair-weather value is about three orders of magnitude smaller and will be unlikely to alter the visible rainbow.

Gedzelman performed a similar set of calculations to Können [147] for non-spherical drops, including solar elevation and drop size distribution, and integrating over the angular width of the solar disk. He concluded that the electric field elevates the top of the bow by decreasing the scattering angles of the first three Airy maxima (because of reduced drop oblateness); it also flattens the minima in these scattering angles, thus producing a brighter bow because it increases the effective range of drop sizes contributing to the bow. However, it also decreases the intensity contrasts of and spacing between the supernumeraries near the top of the rainbow.

Shipley and Weinman [158] carried out a detailed numerical study of scattering by large dielectric spheres, noting that their average light scattering properties can be adequately described by diffraction theory in the forward direction and geometrical optics at most other angles. They compared their results for $N = 1.333$ and size parameters $200 \leqslant \beta \leqslant 4520$ using Mie theory with the extinction efficiency approximation of Van de Hulst [7], the standard forward diffraction formula and with geometrical optics. Excellent agreement was found for the first two comparisons, and as expected for the case of geometrical optics, agreement was found with the exact averaged Mie calculations (for $\beta \sim 4520$), but neither the forward diffraction peak nor the backward glory can be predicted (they did not discuss the rainbow problem in their paper). They investigated also the magnitude and angular structure of the glory, verifying both the results of Bryant and Cox [54] and the predictions of Nussenzveig concerning the backscatter oscillations and phase function oscillations (three predicted oscillations were verified, and a fourth was discovered for $200 \leqslant \beta \leqslant 2000$).

Liou and Hansen [159] carried out a similar set of comparison calculations, but this time for both non-absorbing and absorbing spheres for several refractive indices in the range $1.1 \leqslant \mathrm{Re}\, N \leqslant 2.0$. They were particularly interested in the intensity and polarization for single (as opposed to multiple) scattering by polydisperse spheres for use in cloud microstructure studies. For Re $N = 1.33$ and 1.50 they determined that the ray theory and Mie theory are in close agreement if the size parameter $\beta \gtrsim 400$.

6.5. Rainbows and glories in atomic, nuclear and particle physics

> "Do we have enough imagination to see in the spectral curves the same beauty we see when we look directly at the rainbow?" [207]

While the majority of this review is devoted to the mathematical physics of the *optical* rainbow and glory (albeit with significant quantum mechanical connections, such as tunneling),

a synopsis of rainbows and glories 'writ small' is entirely appropriate. A useful summary can be found in Chapter 16 of [10], where the following three arenas are noted:

(i) In *atomic and molecular scattering*, observed effects include diffractive and rainbow scattering, glories, orbiting and shape resonances;

(ii) In *nuclear heavy-ion scattering*, semiclassical effects such as nuclear rainbows, backward and forward nuclear glories, and surface waves in diffractive scattering.

(iii) In *particle physics*, one type of phenomenological model based on the tunneling of surface waves is very consistent with observations of high-energy proton–proton scattering.

A summary of the literature in these three contexts up to about 1991 can be found in [10]; more recent "key" references (where many further relevant citations are to be found) will be identified below. A common potential used to describe the interaction between two neutral atoms is the *Lennard–Jones* $(12,6)$ potential

$$V(r) = V_0 \left[\left(\frac{r_{\min}}{r} \right)^{12} - 2 \left(\frac{r_{\min}}{r} \right)^{6} \right] ,$$

where V_0 is the depth of the well and $r_{\min}$ is the position of the potential minimum; this describes a repulsive central core within a long-range attractive potential. In addition to derivations of CAM representations of the scattering amplitude [160–163], Regge poles and trajectories have been computed for several interatomic potentials [164,165] (see also related earlier work, previously discussed [19,68,166–168]). The first atomic rainbow was observed in 1962 by Beck [169]. In [170] differential scattering cross sections were studied; the main peaks correspond to the (primary) rainbow and several supernumeraries. By contrast to the optical rainbow, where the (geometrical-optic) rainbow angle depends only on the refractive index N, there is significant 'energy dispersion' in the atomic case because of the energy (k^2) dependence of

$$N_{\text{atomic}} = \sqrt{1 - \frac{V(r)}{k^2}} .$$

The results are like spectral curves for different colors; hence the reason for the comment at the beginning of this subsection. Note also that the size parameters are also functions of energy. Thus scattering can be used as a probe of the depth V_0 of the attractive well in the interatomic potential. Considered as diffraction features, the supernumeraries (which depend on the size parameter) provide information on the *range* of the potential. But things get even better: there are also 'rapid quantum oscillations' arising from interference between attractive and repulsive trajectories. These probe different parts of the potential so that, cumulatively, they can provide information on the *shape* of the potential [171–173]. Furthermore, rainbow effects have been detected in rotational and vibrational excitations in various contexts [174,175]. Forward glory paths and interference oscillations also exist, as do glory undulations (observed in the total cross section for several noble gas systems [176]). Their spacing is a measure of the strength-range product $V_0 r_{\min}$ [177]. There is also a physical interpretation for the total number of glory oscillations [178,179]: since the phase difference between successive maxima is π, and this approaches the s-wave phase shift as $E \to 0$, the total number of oscillations is the same (semiclassically) as the number of multiples of π contained in the zero-energy phase shift. It follows from *Levinson's theorem* [180,92], that this is equal to the *number of bound states* of the potential (see [181] for improved estimates of glory undulations based on a uniform

approximation). The addition of a centrifugal barrier can induce shape resonances, as noted for the optical rainbow [182]; orbiting resonances also play a role in atomic scattering [183–185]. A major aspect of atomic and molecular scattering experiments is of course to use the observational data and relevant mathematical techniques to solve the 'inverse problem', i.e. to reconstruct the original potential [71,172,173].

Similar types of remarks apply to nuclear heavy-ion collisions; semi-classical techniques are developed in this context by Brink [186]. However, in contrast to atomic collisions, the Coulomb interaction must be incorporated into the theory (and the background integral is no longer small [187]); allowance must also be made for strong absorption for distances less than a critical radius R_s (called the strong absorption radius) Nuclear interactions are commonly described via the complex Woods–Saxon optical potential

$$V(r) = \frac{V_0}{1 + \exp[(r - R_r)/a_r} - \frac{\mathrm{i}W_0}{1 + \exp[(r - R_i)/a_i} \ .$$

The presence of the imaginary part describes absorption from the elastic channel; $\mathrm{Re}\, V(r)$ describes a smooth potential well of depth $O(V_0)$ and depth $O(R_r)$; a_r is a parameter describing 'surface diffuseness'. Similar interpretations apply to $\mathrm{Im}\, V(r)$. When $R_i < R_r$ the potential is said to be *surface transparent*. The effective potential for a given angular momentum is composed of the long-range repulsive Coulomb potential (dominant at large distances), the centrifugal barrier and the complex nuclear optical potential $V(r)$. (An optical model potential describes the effective interaction between two particles whose centers of mass are a distance r apart, and for which the solutions of the one-body Schrödinger equation can be used in an asymptotic sense.) Another significant parameter in nuclear scattering is the Sommerfeld parameter

$$n = \frac{kZ_1Z_2e^2}{2E}$$

for a collision with center-of-mass energy E between two nuclei with charge numbers Z_1 and Z_2. When n is small enough, the Coulomb interaction is not dominant, and there are similarities to scattering from an impenetrable sphere (and its concomitant diffractive features [186,187]). For larger values of n (as the impact parameter decreases, and neglecting absorption), the changing relationship between the Coulomb repulsion and the nuclear attraction leads to a *maximum* in the deflection function: a *Coulomb rainbow*. As the impact parameter further decreases, the 'battle' is now between the nuclear attraction and the centrifugal barrier; this time the *minimum* of the deflection function corresponds to a *nuclear rainbow* [10]. Not surprisingly, the effect of absorption is to damp out contributions from paths positioned below R_s; uniform rainbow approximations incorporating absorption have been developed [188,189]. A so-called 'nearside–farside' decomposition was introduced by Fuller [189] and extensively reviewed more recently by Hussein and McVoy [190]; the Coulomb rainbow is a nearside feature because such trajectories are repulsive, and nuclear rainbows are farside features because the latter paths are attractive. In particular, a nuclear rainbow occurs in the scattering of α-particles by ^{90}Zr [190,191].

Fuller and Moffa [192] extended CAM theory to include the Coulomb interaction. For a surface-transparent potential with weak absorption, the scattering amplitude may be decomposed

into contributions from *barrier waves* (a direct reflection term corresponding to reflection at the Coulomb barrier) and *internal waves* (corresponding to the effects of the nuclear potential); see [193,194]. In 1982, Hussein et al. [195] suggested that forward glories (observed of course in optics and atomic scattering) might also occur in heavy-ion nuclear scattering. In optical scattering, there is interference with forward diffraction, but in the nuclear case interference with Coulomb scattering can also arise; this complicates the situation considerably because the latter has a singular amplitude [196] as the forward direction is approached. Under these circumstances, a modification of the total cross section was proposed: it is the *sum-of-differences* cross section

$$\sigma_{\mathrm{sod}}(\theta_0) \equiv 2\pi \int_{\theta_0}^{\pi} \left(\frac{\mathrm{d}\sigma_{\mathrm{C}}}{\mathrm{d}\Omega} - \frac{\mathrm{d}\sigma_{\mathrm{el}}}{\mathrm{d}\Omega} \right) \sin\theta \, \mathrm{d}\theta \, ,$$

representing the cumulative effects of the difference between the Coulomb and elastic differential cross sections, extending down to a small angle θ_0. If this angle is small enough, it has been shown that [197,198]

$$\sigma_{\mathrm{sod}}(\theta_0) \approx \sigma_R - \frac{4\pi}{k} |f_N(0)| \sin \left[\arg f_N(0) - 2\sigma_0 + 2n \ln \sin \frac{\theta_0}{2} \right] \, ,$$

σ_R being the total reaction cross section, $f_N(0)$ the forward nuclear scattering amplitude, σ_0 the s-wave Coulomb phase shift and n the Sommerfeld parameter. There are correction terms to this formula which can be ignored if θ_0 is very small [199–201]. The above result is the charged-particle analog of the optical theorem

$$\sigma = \frac{4\pi}{k} \operatorname{Im} f(k, 0)$$

where $\sigma_{\mathrm{sod}}(\theta_0)$ expresses the removal of the flux from Coulomb paths by absorption and elastic scattering within a narrow forward cone [196]. Note the singular behavior of $\sigma_{\mathrm{sod}}(\theta_0)$ as $\theta_0 \to 0$ as evidenced by the $\ln \sin \theta_0/2$ term: the oscillations are of constant amplitude but increasingly fast frequency. If observed, these could be a characteristic signature of a forward nuclear glory, but in an interesting series of papers [199–201] (particularly the last paper, from a mathematical standpoint) Ueda et al. argue that forward nuclear glory scattering is not taking place in almost any heavy-ion collisions at low energies, despite the above 'signature'. They proposed an alternative mechanism, summarized in the almost-poetic title *Glory* "*in the shadow of the rainbow*" [201]. This refers to the shadow effect of a nuclear rainbow, which leads to a very similar angular distribution to that of the forward glory (being described by the zeroth-order Bessel function, the (fixed) frequency of which is given by the nuclear rainbow angular momentum *instead* of the glory angular momentum).

As far as particle physics is concerned (i.e. high-energy nucleon–nucleon scattering), the concept of tunneling associated with surface waves has been exploited to describe *hadronic diffraction* [202,203]. In elastic collisions, hadrons (elementary particles that interact via the strong force) can be treated as extended objects in that there is a domain of interaction with a characteristic shape and size (in the center-of-mass system). Typically this is a prolate spheroid, with a constant transverse radius R_T (of order 1 fm) and a longitudinal radius R_L that is directly

proportional to the momentum k, i.e. $R_L \propto kR_T^2$. Under appropriate circumstances the interaction can then be treated as the scattering of a plane wave by a k-dependent impenetrable prolate spheroid, with the scattering amplitude expressible in terms of a (by now, familiar) background integral plus residue series from Regge-like poles [204,203] (representing surface waves over the spheroid with associated tunneling features). In [203] this model is extended to include an extra parameter to represent blurring of the sharp boundary of the interaction region.

We conclude this subsection by noting two papers providing (amongst other features) very useful background information for heavy-ion scattering in particular. Brandan et al. [205] provide a valuable description of semi-classical analysis in the context of optical potential models, and more recently Brandan and Satchler provided a clear review of the physics of 'light heavy-ion' scattering [206].

7. The rainbow as a diffraction catastrophe

An alternative way of describing the rainbow phenomenon is by way of catastrophe theory, the rainbow being one of the simplest in catastrophe optics. A review of this subject has been made by Berry and Upstill [208] wherein may be found an introduction to the formalism and methods of catastrophe theory as developed particularly by Thom [209], but also by Arnold [210]. The books by Gilmore [211] (Chapter 13 of which concerns caustics and diffraction catastrophes) and Poston and Stewart [212] are noteworthy in that they also provide many applications. The terminology "diffraction catastrophe" was introduced by Trinkaus and Drepper [213] who were particularly interested in two-dimensional diffraction and the corresponding *inverse problem*: what information can be obtained from the observed diffraction catastrophe pattern? Connor [214] has utilized the ideas of catastrophe theory to semiclassical collision theory (atomic and molecular). Specifically, has applied catastrophe theory to the study of molecular collisions; in particular, the cusp catastrophe is applied to the theory of cubic or cusped rainbows [215–217], of interest in atom–molecule collisions [215,216] and the scattering of atoms from surfaces [217–219].

Optics is concerned, to a great degree with families of rays filling regions of space; the *singularities* of such ray families are *caustics* (see Figs. 3, 5 and 24(a),(c)). For optical purposes this level of description is important for classifying caustics using the concept of *structural stability*: this enables one to classify those caustics whose topology survives perturbation. Structural stability means that if a singularity S_1 is produced by a generating function ϕ_1 (see below for an explanation of these terms), and ϕ_1 is perturbed to ϕ_2, the correspondingly changed S_2 is related to S_1 by a *diffeomorphism* of the control set C (that is by a smooth reversible set of control parameters; a smooth deformation). In the present context this means, in physical terms, that distortions of the raindrop shape of incoming wavefronts from their 'ideal' spherical or planar froms does not prevent the formation of rainbows [10], though there may be some changes in the features. Another way of expressing this concept is to describe the 'system' as well-posed in the limited sense that small changes in the "input" generate correspondingly small changes in the "output". For the elementary catastrophes [209], structural stability is a generic (or typical) property of caustics. Each structurally stable caustic has a characteristic diffraction pattern, the wave function of which has an integral representation in terms of the standard

polynomial describing the corresponding catastrophe. From a mathematical point of view these diffraction catastrophes are especially interesting because they constitute a new hierarchy of functions, distinct from the special functions of analysis [208].

Before summarizing the mathematical details necessary for the purposes of this review, it is appropriate to discuss in qualitative terms the basic rainbow diffraction catastrophe. As known from geometrical optics, the scattering deviation angle D has a minimum corresponding to the rainbow angle (or Déscartes ray) when considered as a function of the angle of incidence i. Clearly the point $(i, D(i))$ corresponding to this minimum is a singular point (approximately $(59^\circ, 138^\circ)$) insofar as it separates a two-ray region ($D > D_{\min}$) from a zero-ray region ($D < D_{\min}$) at this geometrical-optics level of description. This is a singularity or *caustic point.* The rays form a directional caustic at this point, and this is a *fold catastrophe* (symbol: A_2), the simplest example of a catastrophe (see Fig. 24(a); however, it has been noted by McDonald [37] that the caustic is more complicated *inside* the drop than had previously been realized: the wavefront locus exhibits an odd cusp on a segment of the virtual caustic lying wholly within the raindrop). It is the only stable singularity with *codimension* one (the dimensionality of the control space (one) minus the dimensionality of the singularity itself, which is zero). In space the caustic surface is asymptotic to a cone with semiangle 42° (see Fig. 24(c)). Alternatively, as in [208], one can regard i as a state (or behavior) variable and D as a control variable; $i(D)$ is of course multivalued (or not a function, depending on one's preference); and it seems more causally appropriate to consider D as the state variable.

Diffraction is discussed in [208] in terms of the scalar Helmholtz equation

$$\nabla^2\psi(\mathscr{R}) + \kappa^2 N^2(\mathscr{R})\psi(\mathscr{R}) = 0 \tag{7.1}$$

for the complex scalar wavefunction $\psi(\mathscr{R})$. The concern in catastrophe optics is to study the asymptotic behavior of wave fields near caustics in the short-wave limit $\kappa \to \infty$ (semi-classical theory). In a standard manner, $\psi(\mathscr{R})$ is expressed as

$$\psi(\mathscr{R}) = a(\mathscr{R})\mathrm{e}^{\mathrm{i}\kappa\chi(\mathscr{R})}\,, \tag{7.2}$$

where the modulus a and the phase $\kappa\chi$ are both real quantities. To the lowest order of approximation χ satisfies the Hamilton–Jacobi equation and ψ can be determined asymptotically in terms of a phase-action exponent (surfaces of constant action are the wavefronts of geometrical optics). The integral representation for ψ is

$$\psi(\mathscr{R}) = \mathrm{e}^{-\mathrm{i}n\pi/4}\left(\frac{\kappa}{2\pi}\right)^{n/2}\int\ldots\int b(s;\mathscr{R})\exp[\mathrm{i}\kappa\phi(s;\mathscr{R})]\,\mathrm{d}^n s\,, \tag{7.3}$$

where n is the number of state (or behavior) variables s and b is a weight function. In general there is a relationship between this representation and the simple ray approximation (see [208] for further details). According to the principle of stationary phase, the main contributions to the above integral for given $\mathscr{R}$ come from the stationary points, i.e. those points s_i for which the gradient map $\partial\phi/\partial s_i$ vanishes; caustics are *singularities* of this map, where two or more stationary points coalesce. Because $\kappa \to \infty$, the integrand is a rapidly oscillating function of s so other than near the points s_i, destructive interference occurs and the corresponding contributions are negligible. The stationary points are well separated provided $\mathscr{R}$ is not near a caustic;

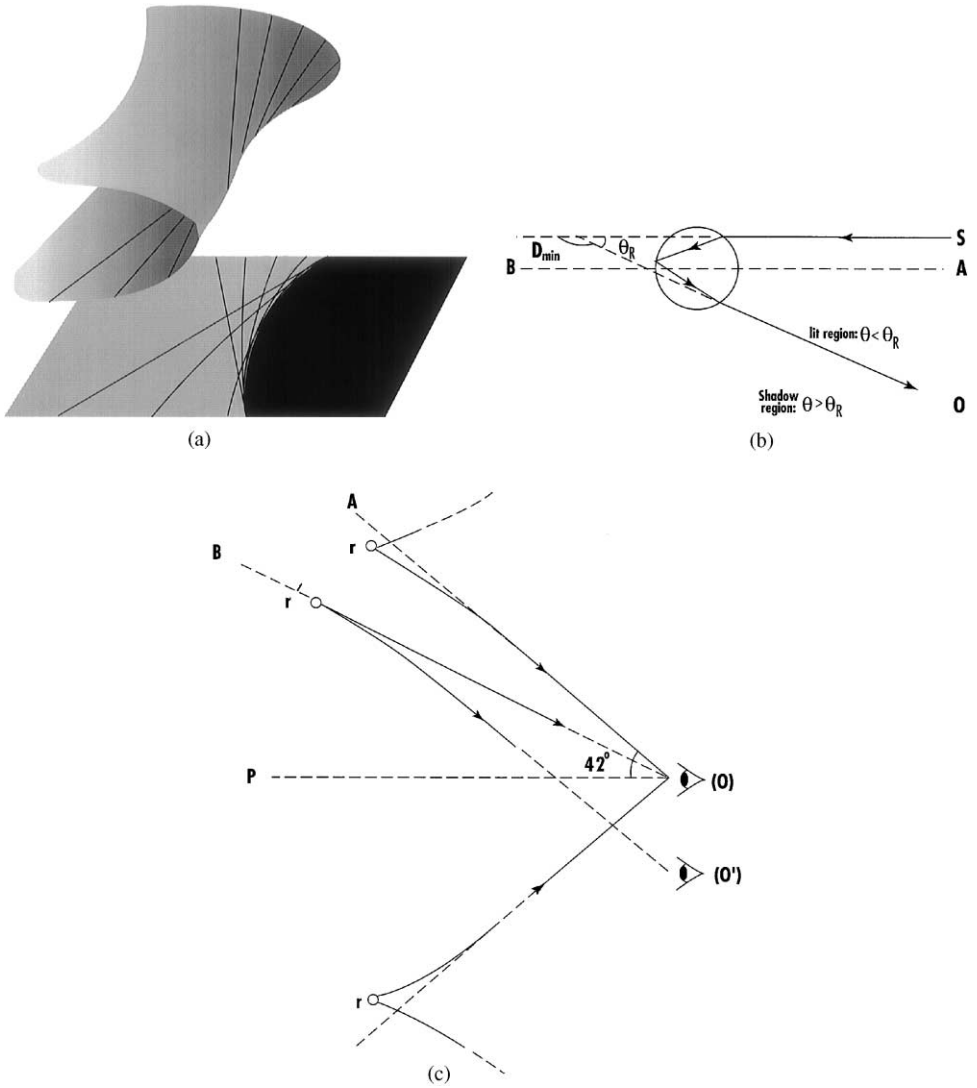

Fig. 24. (a) A generic representation of the fold catastrophe showing the caustic separating the shadow region from the lit or illuminated region. A detailed analysis (see [6]) reveals the underlying nature of the transition zone between the two regions. (Modified from [212].) (b) The shadow and lit regions from a spherical raindrop according to geometrical optics. The observer is at O; AB is the axis of symmetry for rays from the sun (S). (c) A schematic representation of the caustic associated with rainbow scattering as observed at O. The caustic (greatly exaggerated near the raindrops r and r') is asymptotic to a cone of semi-vertex angle $\approx 42^\circ$; in practice the cone extends to within a few diameters of the drop. Each "color" corresponds to a different (but concentric) cone from each drop, so the observer sees every "color" from a different drop (see Fig. 12.13 in [212]). Rays from near the drop at r' in direction BO contribute to the increased illumination below the rainbow seen by the observer at O. The same drop contributes to the rainbow seen by the observer at O$'$. The figure has axial symmetry about the line PO. It is the existence of the fold caustics that enhance the intensity of scattered light at the rainbow angles (one for each color, though of course there is a continuum rather than a discrete set of angles; in this sense the rainbow has an infinite spectrum of "colors"). If such caustics were not present, there would be no rainbow, merely an averaging out to white light again.

the simplest form of stationary phase can then be applied, and yields a series of terms of the form

$$\psi(\mathscr{R}) \approx \sum_{\mu} a_{\mu} \exp[\mathrm{i} g_{\mu}(\kappa, \mathscr{R})] \,, \tag{7.4}$$

where the details of the g_{μ} need not concern us here. Near a caustic, however, two or more of the stationary points are close (in some appropriate sense) and their contributions cannot be separated without a reformulation of the stationary phase principle to accommodate this (i.e. via the CFU method [72]), or by using diffraction catastrophes. The problem is that the "ray" contributions can no longer be considered separately; when the stationary points approach closer than a distance $\mathcal{O}(\kappa^{-1/2})$ the contributions are not separated by a region in which destructive interference occurs. When such points coalesce, $\phi(s;\mathscr{R})$ is stationary to higher than first order, and quadratic terms as well as linear terms in $s - s_{\mu}$ vanish. This implies the existence of a set of displacements ds_i, away from the extrema s_{μ}, for which the gradient map $\partial\phi/\partial s_i$ still vanishes, i.e. for which

$$\sum_{i} \frac{\partial^2 \phi}{\partial s_i \partial s_j} ds_i = 0 \,. \tag{7.5}$$

The condition for this homogeneous system of equations to have a solution (i.e. for the set of control parameters C to lie on a caustic) is that the Hessian

$$H(\phi) \equiv \det\left(\frac{\partial^2 \phi}{\partial s_i \partial s_j} \right) = 0 \,, \tag{7.6}$$

at points $s_{\mu}(C)$ where $\partial\phi/\partial s_i = 0$ (again, details can be found in [208]). The caustic defined by $H = 0$ determines the bifurcation set for which at least two stationary points coalesce (in the present circumstance this is just the rainbow angle). In view of this discussion there are two other ways of expressing this: (i) rays coalesce on caustics, and (ii) caustics correspond to singularities of gradient maps.

To remedy this problem the function ϕ is replaced by a simpler "normal form" Φ with the same stationary-point structure; the resulting diffraction integral is evaluated exactly. This is where the property of structural stability is so important, because if the caustic is structurally stable it must be equivalent to one of the catastrophes (in the diffeomorphic sense described above). The result is a generic diffraction integral which will occur in many different contexts. The basic diffraction catastrophe integrals (one for each catastrophe) may be reduced to the form

$$\Psi(C) = \frac{1}{(2\pi)^{n/2}} \int \cdots \int \exp[\mathrm{i}\Phi(s;C)] \, \mathrm{d}^n s \,, \tag{7.7}$$

where s represents the state variables and C the control parameters (for the case of the rainbow there is only one of each, so $n = 1$). These integrals stably represent the wave patterns near caustics. The *corank* of the catastrophe is equal to n: it is the minimum number of state variables necessary for Φ to reproduce the stationary-point structure of ϕ; the *codimension* is the dimensionality of the control space minus the dimensionality of the singularity itself. It is

interesting to note that in "ray" catastrophe optics the state variables s are removed by differentiation (the vanishing of the gradient map); in "wave" catastrophe optics they are removed by integration (via the diffraction functions). For future reference we state the functions $\Phi(s;C)$ for both the fold (A_2) and the cusp (A_3) catastrophe; the list for the remaining five elementary catastrophes can be found in Table 1 of [208]. For the fold

$$\Phi(s;C)=\tfrac{1}{3}s^3+Cs \tag{7.8}$$

and for the cusp

$$\Phi(s;C)=\tfrac{1}{4}s^4+\tfrac{1}{2}C_2s^2+C_1s\ . \tag{7.9}$$

As emphasized in [208], the diffraction catastrophes $\Psi(C)$ provide transitional approximations, valid close to the caustic and for short waves. They are increasingly inaccurate far from the caustic, but the theory of uniform approximation (mentioned throughout this review but initially in Section 4 regarding the work of Berry on semi-classical scattering [68]) can be obtained by deforming the standard integrals (see references in [208] for further details). The technique is more general than catastrophe optics, since it can be applied to structurally *unstable* caustics, the glory being an example, the codimension of which is infinite [208]. This last statement needs some clarification. Because (i) the wavefront associated with the glory possesses circular symmetry, and (ii) there is an infinite number of topologically different ways in which the (continuous group) symmetry of a circle may be broken, it follows that in order to find all the structurally stable caustics "near" the glory requires infinitely many parameters [220]; indeed, the singularity index ν (see below) for the glory is $\frac{1}{2}$ which means that it is a stronger singularity than the elementary catastrophes with codimensions 1–3 (see the table in Section 3 of [220]). The forward diffraction peak from localized scattering objects is an even stronger structurally unstable caustic than the glory; for further details on short-wave fields and Thom's theorem, again refer to [220].

By substituting the cubic term (7.8) into (7.7) we see that

$$\Psi(C)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp[\mathrm{i}(s^3/3+Cs)]\,\mathrm{d}s=\sqrt{2\pi}\,\mathrm{Ai}(C)\ , \tag{7.10}$$

which has been encountered several times before! For $C<0$ there are two rays (stationary points of the integrand) whose interference causes oscillations in $\Psi(C)$; for $C>0$ there is one (complex) ray that monotonically and exponentially decays to zero. This describes diffraction near a fold caustic. Some 50 years after Airy introduced this function [21] to study diffraction along the asymptote of a caustic (although he did not express it in these terms), and provided a fundamental description of the supernumerary bows, Larmor [221] obtained power laws for the variation of fringe spacing with wavelength of light and the curvature of the caustic. Note that the corresponding integral for the cusp catastrophe is frequently referred to as the *Pearcey integral* (based on [222]). The scaling properties found by Airy for the intensity near the caustic and by Larmor for the fringe spacing can be deduced by considering, from (7.3) the "physical"

fold diffraction catastrophe

$$\Psi(C') = \left(\frac{\kappa}{2\pi}\right)^{1/2} \int_{-\infty}^{\infty} \exp[\mathrm{i}\kappa(s'^3/3 + C's')]\,\mathrm{d}s' \,, \tag{7.11}$$

where, following [208], s' is proportional to distance along the initial wavefront and C' is proportional to distance from the fold caustic. Using the scaling $\kappa s'^3 = s^3$ implies that

$$\Psi(C') = \frac{\kappa^{1/6}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp[\mathrm{i}(s^3/3 + \kappa^{2/3} C's)]\,\mathrm{d}s \,. \tag{7.12}$$

When this is compared with a suggested scaling law between the physical wave function $\psi(C')$ (dependent on physical controls C', independent of κ, e.g. position or time) and the diffraction catastrophe $\Psi(C)$, where C represents the standard control parameters, i.e. in general,

$$\psi(C') = \kappa^{v} \Psi(\kappa^{-\sigma_i}, C'_j) \,, \tag{7.13}$$

then it follows that $v = \frac{1}{6}$ and $\sigma_1 = -\frac{2}{3}$. The exponent v is a measure of the divergence of $|\psi|$ as $\kappa \to \infty$ at the most singular control point $C = 0$ (it is therefore called the *singularity index* [210]). The σ_i (there being only one for the rainbow problem) measure the fringe spacings in the different control directions C_j as $\kappa \to \infty$ [223]. Thus near a fold catastrophe the intensity $|\psi|^2$ is $\mathcal{O}(\kappa^{1/3})$ [21] and the fringe spacing is $\mathcal{O}(\kappa^{-2/3})$ [221]. Note that in contrast to [208] the exponent of κ in the argument of Ψ here is negative; it is written in this way without loss of generality.

It is noted in [208] that wavelength scaling laws outside catastrophe optics do not necessarily follow the functional form (7.13), an example being the glory (Section 3). As noted in that section, the glory, crudely put, is a result of the intense backscattering of light from spherical water droplets. For refractive indices $\sqrt{2} \leqslant N \leqslant 2$ (obviously excluding water) rays emerging in the backward direction after one or more internal reflections would form a (structurally *unstable*) focal line extending to infinity with intensity $|\psi|^2 \propto \kappa$. For water a tangentially-incident ray emerges after one internal reflection at an angle of 14° to the backward direction with amplitude $\mathcal{O}(\kappa^{-1/6})$. These surface waves creep around the surface of the drop by diffraction, suffering wavelength-dependent attenuation as they do so. These rays define a focal line, giving rise to a backscattered wave amplitude [6,208]

$$|\psi| \propto \kappa^{1/3} \exp[-h(\kappa a)^{1/3}] \,, \tag{7.14}$$

where a is the drop radius and h is a constant. Note that the glory disappears in the geometrical-optics limit.

A very interesting set of papers published by Marston and coworkers [224–228] take these ideas to the next level: generalized rainbows arising via the cusp diffraction catastrophe for scattering from spheroidal drops (see also [229] for a more classical approach to rainbow scattering). Using acoustically levitated drops whose shapes were closely approximated by oblate spheroids with the short (symmetry) axis vertical. These drops were observed to scatter in the horizontal rainbow region with patterns corresponding to the *hyperbolic–umbilic diffraction catastrophe* (in Arnold's classification, D_4^+). An important parameter is the axis ratio $q = D/H$ where D is the diameter in the horizontal plane, and H is in the vertical plane. For q sufficiently close to unity, the fold diffraction catastrophe (A_2) is observed, as expected. As q increases, a

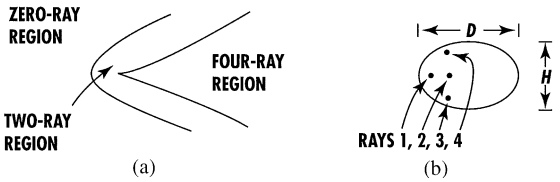

Fig. 25. (a) The partitioning of angular regions according to the hyperbolic–umbilic diffraction catastrophe, arising in connection with rainbow scattering from spheroidal drops (see [224,230] for details). (b) The rays in the horizontal plane (1,2) and the skew rays (3,4) contribute to the generalization of the rainbow for scattering from spheroidal drops. (Redrawn from [224].)

cusp diffraction catastrophe (A_3) enters the field for large θ scattering, and the arcs of the A_2 pattern become noticeably bent (see the photographs in [224,225]). For q at or very close to a critical value, found experimentally to be 1.305 ± 0.016, the fold and cusp patterns merge to form a focal section of the D_4^+ pattern, namely the hyperbolic–umbilic catastrophe. For q beyond this value, the pattern returns to separate A_2 and A_3 patterns (e.g. for $q \simeq 1.37$ and $D = 1.40$ mm in [225]). The A_3 pattern is described by the Pearcey diffraction integral mentioned above.

Why does the D_4^+ pattern occur? It transpires that for the spheroid there are two once-reflected and twice-refracted rays that are *not confined* to the horizontal plane P, but which merge with the Déscartes ray when

$$q = q_c = \left[\frac{3N^2}{4(N^2-1)}\right]^{1/2} \simeq 1.311 \quad \text{for } N = 1.332 \tag{7.15}$$

(see below for discussion of this result), clearly consistent with the above experimental determination. These skew rays are consistent with there being at most four stationary points of the phase function within the D_4^+ diffraction integral. When $q \neq q_c$ the scattering pattern is divided into three regions corresponding to 0-, 2- and 4-ray regions (see Fig. 25); these reflect the number of non-degenerate points of stationary phase of the integral. Three rays merge for $q = q_c$ (the two skew rays and one equatorial ray) in the direction of the cusp point in the focal direction $(\theta, \zeta) = (\theta_c, 0)$, where the first and second coordinates refer to the horizontal and vertical scattering angles, respectively, and θ_c is the cusp location at merging. Marston derives a parametric representation of $q(\theta_c)$ which is in good agreement with the experimental data:

$$q = N[2(N^2 - \sin^2 i)^{1/2}((N^2 - \sin^2 i)^{1/2} - \cos i)]^{-1/2} \tag{7.16}$$

and

$$\theta_3 = 180^\circ + 2\mathrm{i} - 4\arcsin(\sin i/N)\,. \tag{7.17}$$

His theory predicts that as grazing incidence is approached ($i \to 90^\circ$)

$$q \to q_T = N(2N^2 - 2)^{-1/2} \simeq 1.070 \quad \text{and} \quad \theta \to \theta_T \simeq 166^\circ\,.$$

This is a Fock-type transition region [5] for the equatorial ray 1 (see Fig. 25); in a spherical drop the 2-ray region is confined to the angular interval $\theta_R \leqslant \theta \leqslant 166^\circ$. For $q_T < q < q_c$, θ_c

corresponds to the merging of rays 1, 3 and 4. As $i \to 0^\circ$,

$$q \to q_L = \left[\frac{N}{2N-2}\right]^{1/2}, \tag{7.18}$$

where q_L corresponds to a "lips" event [208,230]. The interval $q_c < q < q_L$ corresponds to the merging of rays 2–4. Both types of merging have been photographically confirmed.

The scaling laws for the intensity (more accurately, irradiance [22]) I at the focal angles of the A_2, A_3 and D_4^+ catastrophes are

$$I \propto \left(\frac{D}{\lambda}\right)^p \left(\frac{D}{R}\right)^2 \tag{7.19}$$

for fixed q, where $p = \frac{1}{3}, \frac{1}{2}$ and $\frac{2}{3}$, respectively, and R is the far field distance. Marston suggests that measurements of $\theta_c - \theta_R$ may be used for the inverse problem of determining q, even if D is not well known, since θ_c is independent of $D \gg \lambda$.

A fascinating prediction is made on the basis of the results in these papers: for a low-altitude sun, oblate-shaped drops might exhibit a "cusped rainbow" (A_3 focus) in addition to the usual A_2 rainbow. It is suggested that this may occur for $D \gtrsim 1.8$ mm. Further details concerning coloration and technical details concerning, for example, the opening rate of the cusp caustic may be found in [226–228], and references therein.

Immediately following (i.e. adjacent to) the paper by Marston and Trinh [224] was one by Nye [230] who provided further details of the catastrophe-theoretic interpretation of their experimental and analytical results. By examining the point of entry of the incident ray a vertical distance ϵ above the horizontal plane, he shows that to first order in ϵ the direction cosines of the incident ray, normal to the surface and refracted ray are the following rows of a determinant:

$$\begin{vmatrix} \sin i & 0 & \cos i \\ 0 & \varepsilon/\rho & 1 \\ \sin r & \varepsilon/s & \cos r \end{vmatrix},$$

where $s = D\cos r$ and ρ is the radius of curvature of the surface in the vertical plane, $H^2/2D$. For these three directions to be coplanar the determinant must be zero; this condition yields the result

$$q_c = \frac{D}{H} = \sec r = \left[\frac{3N^2}{4(N^2-1)}\right]^{1/2}, \tag{7.20}$$

which reconfirms the analysis of [224]. Nye further predicts the occurrence of a new phenomenon (a lips event, in the terminology of [132]) as $N \to 2$. For further details consult the paper by Nye [230].

When rainbows are viewed in the laboratory at large distances (compared with the size of the drop) from single water droplets they are frequently referred to as colored *glarespots* [231–234]. Lock studied their formation using Mie theory and the results from the complex angular momentum theory of Nussenzveig [231]. These caustic surfaces are formed by the light rays

emerging from the droplet and the focusing properties of the eye. Some of the emerging rays in the lit region form a real caustic beginning near the surface of the drop and ending on the Déscartes ray; the remainder, traced backwards, form a virtual caustic on the rear side of the drop and extending backward with the Déscartes ray as asymptote. These caustics then act as real or virtual sources of the glare spots. In his paper Lock remarks that when an observer is far from a droplet, he sees the Mie scattered intensity, whereas close to the droplet he sees the square of the Fourier transform of the Mie scattered electric field. He suggests that this Fourier-transforming property of the eye may account for the occasional reports of the tertiary rainbow seen in the atmosphere. Noting that the Mie infinite series of partial waves is an exact solution for the scattering of plane electromagnetic waves from a sphere, Lock points out that all scattering effects that can possibly occur are "hidden somewhere or another" in the Mie amplitude. The extraction of this information has been at times problematical, but by Fourier transforming the Mie fields, many signals (with their own periodicity) superposed in the spatial domain, become separated in the Fourier domain. This helps to determine the properties of glare spots that are determined by geometrical considerations and those which may not be; in particular a glare spot produced by rays at grazing incidence appears in the Fourier transformed Mie fields was noted, the physical mechanism of which was unclear.

8. Summary

In summarizing the different aspects of the rainbow discussed in this review, it seems appropriate to identify some categories that, although a little vague and not mutually exclusive, may serve to describe that phenomenon in a variety of different contexts. These are somewhat heuristic in character, but nonetheless may serve as a reminder of the many complementary levels of description that can be appreciated from the viewpoint of mathematical physics.
A rainbow is:

(1) A concentration of light rays corresponding to a minimum (for the primary bow) of the deviation or scattering angle $D(i)$ as a function of the angle of incidence i; this minimum is identified as the Déscartes or rainbow ray.

(2) A caustic, separating a 2-ray region from a 0-ray (or shadow) region.

(3) An integral superposition of waves over a (locally) cubic wavefront (the Airy approximation).

(4) In part, an interference problem (the origin of the supernumerary bows).

(5) A coalescence of two real saddle points.

(6) (i) A result of scattering by (or interactions with) an effective potential comprising a square well and a centrifugal barrier.

(6) (ii) An example of Regge-pole dominance.

(7) Associated with tunneling in the edge domain (see Section 5.6 and Fig. 13).

(8) Defined, to a considerable extent by the behavior of the scattering amplitude for the third term in the Debye series expansion.

(9) A tangentially polarized circular arc.

(10) A fold diffraction catastrophe.

Since much of this review is devoted to an exposition of the complex angular momentum theory of the rainbow (and to a lesser extent the glory), it is of interest also to summarize the differences as well as similarities between the predictions of the Airy theory (of particular historical importance) and CAM theory:

8.1. The rainbow according to CAM theory

While confirming some aspects of the Airy theory, the method of complex angular momentum predicts differences also: six major features of the CAM theory are summarized in [10]:

(i) *Rainbow enhancement*: The scattering amplitudes predicted by geometrical optics (the zeroth-order WKB term) are $O(\beta)$; CAM theory predicts $O(\beta^{7/6})$, or a maximum rainbow intensity enhancement of $O(\beta^{1/3})$, in agreement with the Airy theory.

(ii) *Rainbow width*: The angular width is $O(\gamma^2)$, much narrower than the penumbral angular width of $O(\gamma)$; hence only a fraction of the total scattered intensity "inhabits" a rainbow.

(iii) *Rainbow polarization*: CAM theory predicts that the parallel polarization intensity is about 4% of the corresponding 'perpendicular' one; thus there is almost complete dominance of the latter (for which the electric field vector is tangential to the rainbow arc) in agreement with observations [146,147]. While this is basically consistent with the Airy theory, there are deviations from the latter as functions of the angle of observation and the size parameter.

(iv) *Angular distribution and supernumerary peaks*: The Airy theory is valid only for large size parameters and small deviations from θ_R, within the rainbow peak. Corrections to the theory are nevertheless small here, however, for the perpendicular polarization, only becoming significant for the supernumeraries. Large deviations occur for the parallel polarization because of the dominance of a term involving the derivative of the Airy function Ai. Indeed, CAM theory predicts (correctly, see [127]) that the peaks for the parallel polarization are located near the minima for the perpendicular component, and vice-versa (because constructive interference for the latter become destructive for the former, the reflection amplitudes for the parallel polarization change sign after going through zero at the Brewster angle).

(v) *Uniform approximation*: The dominant contribution to the third Debye term in the rainbow region is a *uniform asymptotic expansion*. On the bright side of the rainbow, this result matches smoothly with the WKB approximation in the 2-ray region. On the dark side, it does so again, this time with the damped complex saddle-point contribution. Put differently, diffraction into the shadow side of a rainbow occurs by *tunneling*.

(vi) *Higher-order rainbows*: The uniform CAM theory can be extended naturally to higher-order bows. For a Debye term of order p, the critical angle of incidence for a rainbow is, (from Eq. (1.3), where $p = k + 1$)

$$i_p = \arccos\left(\frac{N^2 - 1}{p^2 - 1}\right)^{1/2} .$$

This angle increases monotonically with p, so that higher-order rainbows arise from incidence *within the edge strip*. Note also that, since the angular width of the rainbow is $O(p\beta^{2/3})$, i.e. growing linearly with p for large values of p, the rainbow peak flattens with increasing p

values. Interestingly, while the intensity of higher-order rainbows decreases with p by virtue of losses by internal reflection, this becomes less pronounced as p increases because the reflectivity increases as the edge is approached. The deviations from the Airy theory increase with p for both polarizations.

In many respects, the rainbow is, in physical terms at least, a simpler phenomenon than the glory. Khare and Nussenzveig [47] list eight different physical effects which together or in part can contribute to this second beautiful occurrence. These are:

(i) *The edge effect*: This is the effect of incident rays in the domain (again, see Section 5.5 and Fig. 13). These rays are both above and below the barrier of the effective potential in this domain, and correspond, respectively, to rays that are incident on the drop from within the Fock penumbral region, and rays that reach the surface by tunneling (barrier penetration).

(ii) "*Orbiting*": This applies to rays that penetrate the drop at angles close to critical; they undergo nearly total internal reflection and consequently are able to make many circumlocutions within the drop with appreciable damping. Mathematically, many higher-order Debye terms contribute to this effect.

(iii) *Axial focusing*: The axial symmetry of the problem and the peripheral nature of the leading contributions to the glory give rise to an amplitude enhancement $\sim \beta^{1/2}$.

(iv) *Cross-polarization*: This occurs because of interference between contributions of comparable magnitude from the electric and magnetic polarization components, leading to additional structure in the glory pattern.

(v) *Surface waves*: These represent one of the two categories of leading contributions to the glory (the other being complex rainbow rays: see (vi) below). These arise from Regge–Debye poles. In addition to strictly surface waves, waves that subsequently take (or have previously taken) additional shortcuts through the drop also must be accounted for. A limiting case is a "glory ray" contribution, an example of which is the $p=24$ case for N given by Eq. (6.4), which yields a Fock-type term (see Fig. 21).

(vi) *Complex rainbow rays*: These arise on the shadow side of higher-order rainbows formed near the backward direction (for a description of these higher-order rainbows and their location relative to the observer see [14,132,133]). Mathematically, they are complex saddle points and physically they occur because of the $\beta^{1/6}$ rainbow enhancement factor [61] and the flattening of the rainbow peak for higher-order rainbows.

(vii) *Geometrical resonances*: These are closed (or nearly so) quasi-periodic orbits associated with both the quasi-periodicity of the glory pattern and the narrow resonances or "spikes". Such orbits play an important role in the semi-classical description of quantization, and the study of resonances and bound states [235].

(viii) *Competitive damping*: In general, the glory pattern and its polarization are functions of both β and θ, reflecting the interplay between several competing damping effects. The outcome of such competition determines whether or not the leading term is electric or magnetic in character, surface-wave or rainbow-type, and the ordering of Debye contributions. The different types of damping have been noted earlier: surface waves are radiation-damped in tangential directions; rainbow terms are damped as complex rays in the shadow of a caustic; and all terms are subject to damping as a result of multiple internal reflections. For $\beta \sim 10^2$ surface-wave effects are generally dominant; rainbow effects are dominant for $\beta \sim 10^3$.

In [47] the authors make an important point in connection with these many different types of contribution (see also [235–237]):

> "The glory shows for the first time in optics that diffraction effects due to complex orbits (surface waves, complex rainbow rays) may be strongly dominant over geometrical optic ray contributions (real orbits). The moral this suggests is that complex extremals of Feynman path integrals may have to be taken into account in more general situations."

Continuing this theme, at the end of his book [10] Nussenzveig asks the question (raised at the beginning of this article): *why complex angular momentum?* He draws the conclusion that, in part at least, it is because of the intimate relationship that exists between tunneling and analytic continuation, noting that all descriptions of the former make use of the latter (in one form or another). It is perhaps surprising, initially, that such 'physical' and 'mathematical' topics should be so closely related, but a moment's reflection on both the historical and current relationship between the two subjects indicates that this is fascinating but very natural. Indeed, to quote from [10]:

> "Ultimately, this [relationship] goes back to Euler's great discovery of the connection between oscillations and the exponential function, exemplified by what has been called one of the most beautiful formulas of mathematics, his synthesis of analysis, algebra, geometry and arithmetic: $e^{i\pi} = -1$."

It is fitting that "one of the most beautiful formulas of mathematics" should provide the basis for a richer understanding of two of the most beautiful meteorological phenomena known. Nussenzveig explains further why CAM theory is especially well suited to the treatment of semiclassical problems:

> "angular momentum is conjugate to the scattering angle, and the localization principle provides a natural physical interpretation of the Poisson representation, linking Huygens' principle with pseudoclassical paths. This allows one to combine physical insights derived from classical wave theory and from quantum potential scattering in order to deal with previously intractable dynamical features of diffraction.
>
> Thus, the beautiful mathematical theory of analytical continuation provides the key to a deeper understanding of some of the most beautiful phenomena displayed in the sky, and also manifested in so many other ways—through all scales of size—revealing the underlying unity of nature."

It is quite remarkable that, despite the tremendous development of mathematical physics since the time of Newton, he was able to express similar thoughts almost 300 years earlier when he stated in 'Question 31' of "Opticks" that

> "Nature is very consonant and conformable to herself." [241]

Acknowledgements

It is a pleasure to acknowledge the assistance of several individuals during the preparation of this manuscript. In particular, I am grateful to Professor H.M. Nussenzveig for permission

to use some figures from some of his publications; I have found the book [10] to be a valuable resource, especially with regard to what one might describe as 'macroscopic tunneling', but also as a source of references to rainbows and glories in atomic, molecular and nuclear scattering. I am also grateful to Professor M.S. Hussein who provided me with references to some of his most recent publications. Both the anonymous reviewer and the editor, Professor J. Eichler provided valuable feedback and advice which helped to improve the manuscript considerably. In addition, Professor T.F. Gallagher offered advice in the early stages of preparation of this article. My colleague Professor Mark Lesley drew my attention to the statement by Newton (quoted above); I am delighted that his enthusiasm for this project has been almost as great as mine. Finally I would like to thank all the staff at the editorial office in Amsterdam, particularly Yvonne van Lieshout who patiently encouraged me through the entire editorial process.

Appendix A. Classical scattering; the scattering cross section

Consider the problem of a particle of mass m moving under the influence of a central field; the path of the particle lies in one plane and its angular momentum M is conserved [238,239], i.e. in standard notation,

$$M = mr^2\dot{\theta} = \text{constant} . \tag{A.1}$$

It is well-known that this implies Kepler's second law, namely that the radius vector of the particle sweeps out equal areas in equal times. If $V(r)$ is the potential field in which the particle moves, then the total energy E of the particle may be written as

$$E = \tfrac{1}{2}m(\dot{r}^2 + r^2\dot{\theta}^2) + V(r) = \tfrac{1}{2}m\dot{r}^2 + \frac{M^2}{2mr^2} + V(r) , \tag{A.2}$$

from which

$$\dot{r} = \sqrt{\left\{\frac{2}{m}[E - V(r)] - \left(\frac{M}{mr}\right)^2\right.} . \tag{A.3}$$

This expression can be integrated to yield the time t explicitly, or from (A.1)

$$\theta = \int \frac{M}{r^2\sqrt{\{2m[E - V(r)] - (M/r)^2}}\,\mathrm{d}r + \text{constant} , \tag{A.4}$$

which is the general solution of the equation of the path $\theta(r)$. It also follows from (A.1) that θ varies monotonically with time since $\dot{\theta}$ never changes sign. The quantity

$$\frac{M^2}{2mr^2} + V(r) = U(r) \tag{A.5}$$

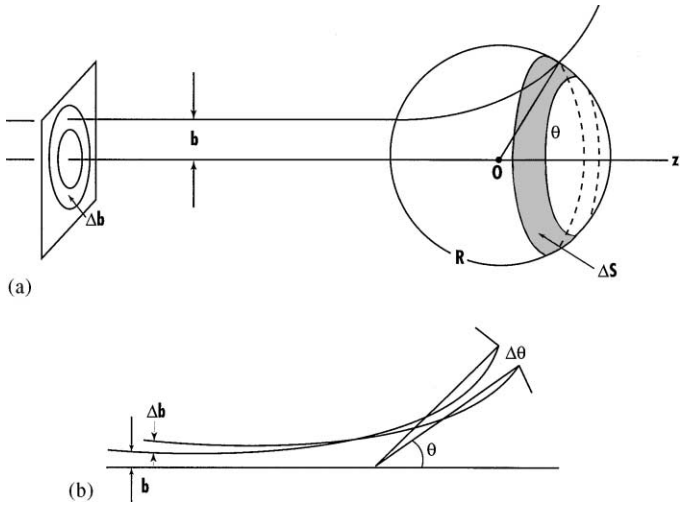

Fig. 26. (a) The distribution of scattered particles incident from a narrow annulus of width Δb, b being the impact parameter, located on the surface of a large sphere with center at the scattering center. (b) Two orbits for a repulsive potential with slightly different values of the impact parameter b.

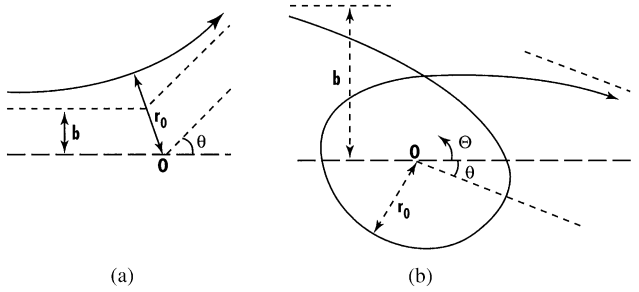

Fig. 27. Relationships between the impact parameter b, the classical distance of closest approach (or outermost radial turning point) r_0, the deflection angle Θ and the scattering angle θ. In (a) the interaction is attractive, and $\Theta = \theta$; in (b) it is repulsive, and $\Theta = -2\pi - \theta$. (Redrawn from [10].)

is called the "effective potential", being the sum of the central potential $V(r)$ and the "centrifugal" potential. Values of energy for which $E = U(r)$ are intimately related to the limits of the motion since at these values $\dot{r} = 0$, thus defining the *turning points* of the path. The radial motion of the particle ranges from $r = \infty$ to the largest root r_0 of $\dot{r} = 0$ (thus defining the *classical distance of closest approach* (or outermost radial turning point), and then from r_0 to $r = \infty$. The total (classical) deflection angle (see Figs. 26 and 27) is therefore

$$\Theta(M) = \pi - 2M \int_{r_0}^{\infty} \frac{\mathrm{d}r}{r^2 \sqrt{\{2m[E - V(r)] - (M/r)^2}} \,. \tag{A.6}$$

The *impact parameter* b is defined as $b = M/mv_0$, where v_0 is the initial speed of the particle, so $\Theta = \Theta(b)$ also. For an entirely repulsive potential, $\Theta \in [0, \pi]$, but as pointed out in [10], it can take arbitrarily large negative values for an attractive one, since the particle may "circle" around the scattering center many times before finally emerging. The relationship between the deflection angle Θ and the *scattering angle* θ is

$$\Theta = \pm\,\theta - 2n\pi, \quad n = 0, 1, 2, \ldots$$

(see Fig. 27) where is chosen such that $\theta \in [0, \pi]$, its physical range of variation.

For particles with impact parameters in the range $[b, b + \Delta b]$ the area of the associated annuls is $\Delta\sigma = 2\pi b|\Delta b|$; if these particles are scattered within an "onion ring" of area $\Delta S = 2\pi R^2 \sin\theta|\Delta\theta|$ (see Fig. 26), or equivalently within a solid angle $\Delta\Omega = 2\pi \sin\theta|\Delta\theta|$, then the *differential scattering cross section* is defined to be (in the appropriate limiting sense)

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = \frac{b \csc\theta}{|\mathrm{d}\theta/\mathrm{d}b|} = \frac{b \csc\theta}{|\mathrm{d}\Theta/\mathrm{d}b|}\ . \tag{A.7}$$

The distribution is independent of the azimuthal angle ϕ for a central force field. Note that for a repulsive potential inparticular, $\mathrm{d}\theta/\mathrm{d}b < 0$ because the larger the impact parameter the smaller will be the angle through which the particles are scattered in general. Note that the total scattering cross section is obtained by integrating (A.7) over all solid angles, i.e.

$$\sigma = \int \frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\mathrm{d}\Omega = \int_0^{b_{\max}} 2\pi b \left|\frac{\mathrm{d}b}{\mathrm{d}\theta}\right| \mathrm{d}\theta = \pi b_{\max}^2\ , \tag{A.8}$$

where $b_{\max}$ is the maximal impact parameter, i.e. if for $r > b_{\max}$ the potential vanishes (as it does for square well or barrier potentials) then particles with an impact parameter greater than $b_{\max}$ are not deflected at all (at least, classically; tunneling does occur as pointed out in the body of the article). The scattering cross section is then just the geometrical cross section of the region; if the scattering potential has no such finite cut-off then the classical scattering cross section is infinite. This is related to a sharp forward peaking of the differential cross section [239, Chapter 5], and is one of the classical singularities in $\mathrm{d}\sigma/\mathrm{d}\Omega$ discussed below.

It is pointed out in [10] that there may exist several trajectories which give rise to the same scattering angle, so (A.7) must be generalized to sum over all the impact parameters that lead to the same angle:

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}(\theta) = \sum_j \frac{b_j(\theta)\csc\theta}{|\mathrm{d}\theta/\mathrm{d}b_j|} \tag{A.7a}$$

where the derivative on the right-hand side is equivalent to the Jacobian of the transformation relating θ and b. Some of these trajectories are illustrated in Fig. 28 for a typical potential type found in atomic and nuclear scattering problems, namely an attractive outer region with a repulsive central core (also shown schematically in the figure, along with a typical deflection function $\Theta(b)$). It is also noted in [10] (see also [200]) that there are three different types of singularities (or caustics—a caustic is an envelope of a family of rays) associated with directions in which the differential cross section diverges. These are as follows:

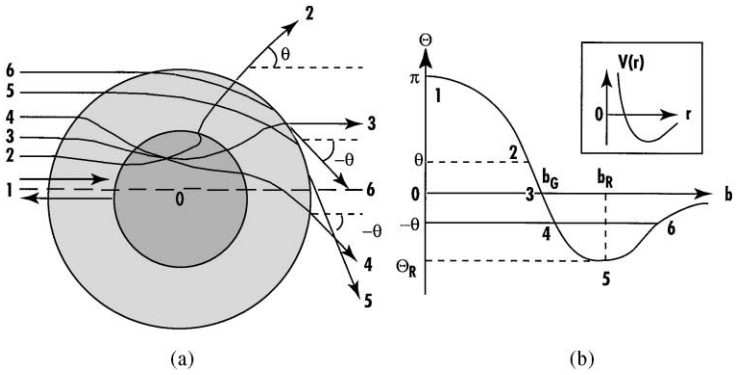

Fig. 28. (a) Several different paths for particle interaction with a potential composed of a repulsive inner core (hatched region) with an attractive outer shell. (b) The corresponding deflection angle as a function of the impact parameter b, $\Theta(b)$. The inset shows the typical potential shape $V(r)$. Note that the repulsive path 2 and the attractive paths 4 and 6 result in the same magnitude of the scattering angle $|\theta|$. Path 3 (when $b = b_G$) is a 'glory path' and path 5 (when $b = b_R$) is a 'rainbow path'. (Redrawn from [10].)

(I) *Rainbow scattering*: This is a caustic arising when $\theta = \theta_R$, the rainbow angle; it occurs when

$$\left(\frac{\mathrm{d}\theta}{\mathrm{d}b}\right)_{\theta_R} = 0\,, \tag{A.9}$$

so that the deflection function passes through an extremum. Its name arises from the analogy with the optical rainbow. In the "heavy-ion" scattering this caustic is referred to as Coulomb or nuclear rainbow scattering if the extremum is a maximum or minimum, respectively [200], and is manifested in terms of an angular distribution given by the Airy function (see Section 2.1). Clearly in view of (A.9) for θ sufficiently close to θ_R we may write

$$\theta(b) \approx \theta_R + \tfrac{1}{2}\theta''(b_R)(b - b_R)^2\,,$$

where $b = b_R$ when $\theta = \theta_R$. Under these circumstances, the classical differential scattering cross section is

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} \approx \left(\frac{b}{\sin\theta}\right)_R \sqrt{\frac{2}{|\theta''(b_R)(\theta - \theta_R)|}}$$

which diverges like $|\theta - \theta_R|^{-1/2}$.

(II) *Glory scattering*: This is a caustic caused by the vanishing of $\sin\theta$ in Eq. (A.7). In terms of the deflection angle Θ for non-zero impact parameter b_{G}

$$\Theta(b_{\mathrm{G}}) = n\pi \quad (n = 0, -1, -2, \ldots)$$

so that the differential cross section diverges like $(\sin\theta)^{-1}$ in the precisely forward direction (n even: forward glory scattering) or backward direction (n odd: backward glory scattering).

Eq. (A.6) implies that $\Theta = \pi$ can only occur when $b = 0$; this case does not constitute glory scattering. The characteristic angular distribution behaves like a zeroth-order Bessel function [200], where it is pointed out that nuclear rainbow scattering and backward glory scattering are expected to be good probes of nuclear interaction at relatively short distances, as is forward glory scattering at relatively large distances.

(III) *Forward peaking*: As alluded to above, scattering potentials with tails extending asymptotically to zero at infinity will lead in general to a divergent forward differential cross section. This is because of the accumulation of small deflections for particles with arbitrarily large impact parameters, i.e. from (A.7a) in which some $b_j \to \infty$, $\theta \to 0$ while $|\mathrm{d}\theta/\mathrm{d}b_j|$ remains bounded. Even for cut-off potentials, some forward anomaly is still to be expected. There is a fourth situation that may occur when the effective potential $U(r) = U_M(r)$ for a given value of M has a relative maximum at $r = r_0$:

(IV) *Orbiting*: In this case

$$U_M(r) = E; \quad (\mathrm{d}U_M/\mathrm{d}r)_{r_0} = 0 \ .$$

This is interpreted as the existence of an unstable circular orbit with radius r_0 [10]. Under these circumstances the integral (A.6) diverges logarithmically at its lower limit if $U''_M(r_0) \neq 0$. A simple calculation yields the result that $\Theta(M) \to -\infty$ as $r \to r_0$; this means that a particle with this energy and angular momentum will spiral indefinitely around the scattering center, and taking an infinitely long time to reach the top of the "barrier". In addition, the inverse function $b(\theta)$ becomes infinitely many-valued, so an infinite number of branches contribute to the differential cross section (A.7a). If $U_M(r)$ has a local minimum for $r < r_0$ then particles with energy $E < U_M(r_0)$ can sustain oscillatory orbits within the potential "well", and the unstable circular orbit functions as a separatrix between bounded and unbounded trajectories [10] (see also [240]). Fig. 29 illustrates connections between the effective potential, tunneling and the ray picture in rainbow and glory formation.

A.1. Semi-classical considerations: a précis

In a sense, the semi-classical approach is the "geometric mean" between classical and quantum mechanical descriptions of phenomena; while one wishes to retain the concept of particle trajectories and their individual contributions, there is nevertheless an associated de Broglie wavelength for each particle, so that interference and diffraction effects enter the picture. The latter do so via the transition from geometrical optics to wave optics. The differential scattering cross section is related to the quantum scattering amplitude $f(k, \theta)$ and this in turn is expressible as the familiar partial wave expansion [65]. The formal relationship between this and the classical differential cross section is established using the WKB approximation and the principle of stationary phase is used to evaluate asymptotically a certain phase integral (see [10, Section 1.2] for further details). A point of stationary phase can be identified with a classical trajectory, but if more than one such point is present (provided they are well separated and of the first order) the corresponding expression for $|f(k, \theta)|^2$ will contain interference terms. This is a distinguishing feature of the 'primitive' semi-classical formulation, and has significant implications for effects (I)–(IV) noted above. The infinite intensities (incorrectly) predicted

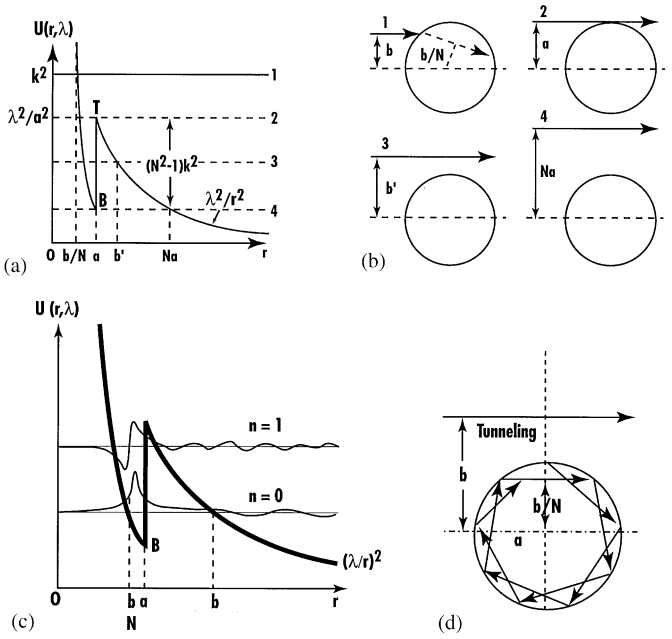

Fig. 29. (a) The effective potential $U(r,\lambda)$ for a transparent sphere of radius a (similar to Fig. 13 but with $U(r,\lambda) > 0$ for all r and redrawn to show four 'energy levels', respectively, above the top of the potential well, at the top, in the middle and at the bottom of the well. (b) The corresponding incident rays and impact parameters. Case 2 shows a tangentially incident ray; note that in case 1 the refracted ray is shown. It passes the center at a distance of $l = b/N$; that this is the case is readily shown from simple geometry: from Snell's law of refraction, $\sin i = N \sin r = b/a$ and since $l = a \sin r$, the result follows. (c) Similar to (a), but with resonant wave functions shown, corresponding to 'family numbers' $n = 0$ and 1 (the latter possessing a single node). (d) The 'tunneling' phenomenon illustrated for an impact parameter $b > a$, being multiply reflected after tunneling, between the surface $r = a$ and the caustic surface $r = b/N$ (the inner turning point). ((a) and (b) redrawn from [10]; (c) and (d) from [88].)

by geometrical optics at focal points, lines and caustics in general are "breeding grounds" for diffraction effects, as are light/shadow boundaries for which geometrical optics predicts finite discontinuities in intensity. Such effects are most significant when the wavelength is comparable with (or larger than) the typical length scale for variation of the physical property of interest (e.g. size of the scattering object). Thus a scattering object with a "sharp" boundary (relative to one wavelength) can give rise to *diffractive scattering* phenomena.

Under circumstances like those in I–IV above, the primitive semiclassical approximation breaks down, and diffraction effects cannot be ignored. Although the angular ranges in which such critical effects become significant get narrower as the wavelength decreases, the differential cross section can oscillate very rapidly and become very large within these ranges. As such the latter are associated with very prominent features and in principle represent important probes of the potential, especially at small distances. The important paper by Ford and Wheeler [19] contained *transitional asymptotic approximations* to the scattering amplitude in these 'critical' angular domains, but they have very narrow domains of validity, and do not match smoothly

with neighboring 'non-critical' angular domains. It is therefore of considerable importance to seek *uniform asymptotic approximations* that by definition do not suffer from these failings. Fortunately, the problem of plane wave scattering by a homogeneous sphere exhibits all of the critical scattering effects (and it can be solved exactly, in principle), and is therefore an ideal laboratory in which to test the accuracy and efficacy of the various approximations. Furthermore, it has relevance to both quantum mechanics (as a square well or barrier problem) and optics (Mie scattering); indeed it also serves as a model for the scattering of acoustic and elastic waves, and was studied in the early twentieth century as a model for the diffraction of radio waves around the surface of the earth [25] (and see Appendix C).

Appendix B. Airy functions and Fock functions

There are several different notational forms for *Airy functions* in the literature. In the form introduced by Airy [21], the 'rainbow integral' was defined as [10,242]

$$\mathrm{Ai}(z)=\frac{3^{1/3}}{\pi}\int_0^\infty \cos(\xi^3+3^{1/3}z\xi)\,\mathrm{d}\xi\ , \tag{B.1}$$

where the above form $y=\mathrm{Ai}(z)$ is a particular solution (for real z) satisfying the "wound healing" differential equation

$$\frac{\mathrm{d}^2y}{\mathrm{d}z^2}+zy=0 \tag{B.2}$$

(so-called because it provides a transition between discontinuous potentials or other spatial features). A second linearly independent solution is denoted by $\mathrm{Bi}(z)$. The Airy functions $\mathrm{Ai}(z)$ and $\mathrm{Bi}(z)$ are entire transcendental functions of z and are real when z is real. They may be expressed as linear combinations of Bessel functions of order $\pm\frac{1}{3}$; either in terms of modified Bessel functions (for argument z) or ordinary Bessel functions (for argument $-z$). Their derivatives are correspondingly expressible in terms of Bessel functions of order $\pm\frac{2}{3}$ (Bowman et al.). The more general form of (B.1) for complex values of z is (see also (B.6) below), together with a rather more complicated expression for $\mathrm{Bi}(z)$.

$$\mathrm{Ai}(z)=\frac{1}{2\pi\mathrm{i}}\int_{\infty\mathrm{e}^{-\mathrm{i}\pi/3}}^{\infty\mathrm{e}^{\mathrm{i}\pi/3}}\exp\left(\frac{1}{3}\xi^3-z\xi\right)\,\mathrm{d}\xi\ .$$

The *general* solution of (B.2) may be written as

$$y(z)=z^{1/2}[c_1J_{1/3}(\zeta)+c_2Y_{1/3}(\zeta)]\ , \tag{B.3}$$

where

$$\zeta=\tfrac{2}{3}z^{3/2}\ .$$

The asymptotic expansion of $\mathrm{Ai}(z)$ for large $|z|$ is [243]

$$\mathrm{Ai}(z)\approx\tfrac{1}{2}\pi^{-1/2}z^{-1/4}\mathrm{e}^{-\zeta}[1+O(\zeta^{-1})] \tag{B.4}$$

for $|\arg z| < \pi$, and

$$(4)\quad \mathrm{Ai}(-z) \approx \pi^{-1/2} z^{-1/4} \left\{ \sin\left(\zeta + \frac{\pi}{4}\right) [1 + O(\zeta^{-1})] - \cos\left(\zeta + \frac{\pi}{4}\right) O(\zeta^{-1}) \right\} . \tag{B.5}$$

An alternative form for the Airy integral, used by van de Hulst [7], is

$$f(\tilde{z}) = \int_0^\infty \cos\left[\frac{\pi}{2}(\tilde{z}\xi - \xi^3)\right] \mathrm{d}\xi .$$

Jackson [35] uses the form [243]

$$\mathrm{Ai}(-z) = \frac{1}{\pi} \int_0^\infty \cos\left(\frac{\xi^3}{3} - z\xi\right) \mathrm{d}\xi \tag{B.6}$$

where

$$z = (2k^2 a^2/\theta''(\theta_0))^{1/3} (\theta - \theta_0)$$

(where $\tilde{z} = (12/\pi^2)^{1/3} z$, and $f(\tilde{z}) = (2\pi^2/3)^{1/3} \mathrm{Ai}(-z)$).

The *Fock functions* are defined as [244]

$$(5)\quad f(z) = \frac{\mathrm{e}^{-\mathrm{i}\pi/6}}{2\pi} \int_\Gamma \frac{\mathrm{e}^{\mathrm{i}z\xi}}{\mathrm{Ai}(\xi \mathrm{e}^{2\mathrm{i}\pi/3})} \mathrm{d}\xi$$

$$g(z) = -\frac{\mathrm{e}^{\mathrm{i}\pi/6}}{2\pi} \int_\Gamma \frac{\mathrm{e}^{\mathrm{i}z\xi}}{\mathrm{Ai}'(\xi \mathrm{e}^{2\mathrm{i}\pi/3})} \mathrm{d}\xi , \tag{B.7}$$

where Γ is a contour starting at infinity in the angular sector $\pi/3 < \arg \xi < \pi$, passing between the origin and the pole of the integrand nearest the origin, ending at infinity in the angular sector $-\pi/3 < \arg \xi < \pi/3$. They can be generalized to include integer powers of ξ in the integrand. To leading order, their asymptotic behavior is [244]

$$f(z) \sim 2\mathrm{i}z \exp(-\mathrm{i}z^3/3), \quad z \to -\infty$$

$$\sim 0, \quad z \to \infty$$

and

$$g(z) \sim 2 \exp(-\mathrm{i}z^3/3), \quad z \to -\infty$$

$$\sim 0, \quad z \to \infty .$$

Appendix C. The Watson transform and its modification for the CAM method

The Watson transformation was originally motivated by the desire to understand diffraction of radio waves around the Earth (prior to the discovery of the ionosphere). The topic of primary interest was therefore the neighborhood of the Earth's surface in the shadow zone of the transmitter. Once the transformation to the complex l-plane had been made, with the

corresponding freedom of path deformation, rapidly convergent asymptotic estimates could be made by paths suitably chosen to emphasize a few dominant pole contributions as opposed to dealing with those from many partial waves. In the shadow region the dominant (complex) poles λ_n are now called Regge poles, and their imaginary part (associated with angular damping [47]) grows rapidly with n, so as we have seen, only those few poles closest to the real axis need to be considered. This leads to a rapidly convergent residue series, each term of which corresponds to a "creeping wave" generated by tangentially incident rays which are radiation damped as they travel around the surface of the sphere.

In the lit region Watson's method fails: in addition to Regge poles the now-dominant "background integral" in the l-plane (or equivalently the λ-plane, where $\lambda = l + \frac{1}{2}$) must be considered. The significant contributions to this integral come from saddle points which are, in general, complex. When these are real, however, they correspond to the rays of geometrical optics and provide the associated contributions; by taking account of higher-order terms in the saddle-point method, the WKB series can be recovered to any desired accuracy. There is a penumbral region between the lit and shadow regions: this was investigated by Fock [245], who was able to describe the behavior of the wave field in this region in terms of a new function (subsequently named the *Fock function*; see Appendix B). It is in this region, near the edge, that the creeping waves are generated. The Fock function interpolates smoothly between the geometrically lit (WKB) region and the diffracted ray or creeping wave region. While subsequent attempts were made to apply the method to both impenetrable and transparent spheres ([30,31]; see [47] for further references), and significant advances were made, there was no theory going beyond the classic Airy approximation in a comprehensive manner for all scattering angles.

In a very comprehensive treatment of wave and particle scattering [239], Newton mentions many of the topics contained in this review (see for example, Chapters 3, 5, 13 and 18 therein). In particular, the latter chapter summarizes the CAM method as originally formulated by Regge, and the reader is referred there for details of the Watson transform (unmodified), questions of uniqueness, Regge poles and the Mandelstam representation. In this appendix an outline due to Frautschi [246] is followed for the Watson transform, and further commentary on the modified transform is added from [47]. There are many equivalent but seemingly different representations of the Watson–Regge transform (as it is called in [239]; see for example [247], Chapter 15 and references in [248]); a common one will be adopted here.

We consider the partial-wave expansion of the scattering amplitude in form (5.18) (omitting the drop diameter or well radius a):

$$f(k,\theta) = \frac{1}{2ik}\sum_{l=0}^{\infty}(2l+1)(S_l(k)-1)P_l(\cos\theta) \tag{C.1}$$

and consider the analytic continuation of this expression to the complex l-plane, $\mathrm{Re}\, l > -\frac{1}{2}$. This partial-wave expansion may be rewritten as an integral $\mathscr{I}(k,\theta)$ in this domain such that each term in the summation is the residue of a pole in the contour integration; thus

$$\mathscr{I}(k,\theta) = \frac{1}{4k}\int_{\mathscr{C}'}(2l+1)(S_l(k)-1)\frac{P_l(-\cos\theta)}{\sin\pi l}\,\mathrm{d}l\ , \tag{C.2}$$

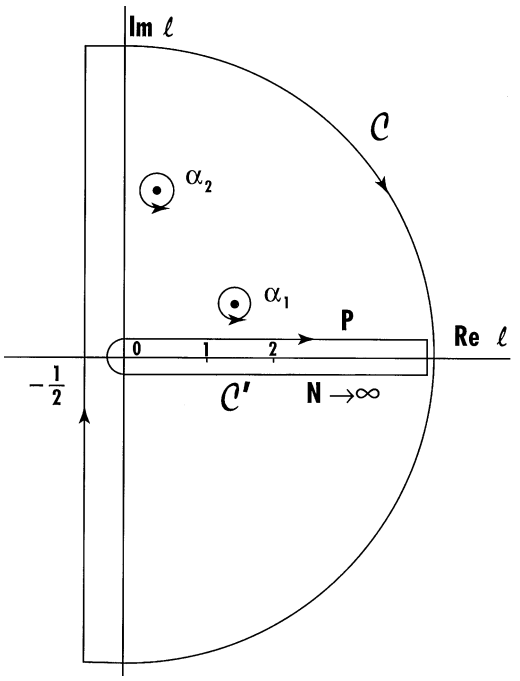

Fig. 30. The (classical) Watson transformation contour for the scattering amplitude before (P) and after deformation away from the real l-axis. A corresponding contour applies in the complex λ-plane also.

where $\mathscr{C}'$ is the contour shown in Fig. 30. This representation follows from the fact that near the pole of the integrand at $l=n\geqslant 0$,

$$\frac{P_l(-\cos\theta)}{\sin\pi l}\approx\frac{P_n(-\cos\theta)}{(-1)^n\pi(l-n)}=\frac{P_n(\cos\theta)}{\pi(l-n)}$$

and the orientation of the contour $\mathscr{C}'$, which encloses all the integers $l=0$–$\mathscr{N}$, and the limit $\mathscr{N}\to\infty$ is taken. The original expansion (C.1) is not useful for $z=\cos\infty\to\infty$ because it converges only at small z, i.e. within the Lehmann ellipse [239,246–248], the difficulty lying with the asymptotic behavior of $P_l(z)$. Regge overcame this problem [249] by shifting the contour of integration from around the real l-axis to the vertical line $l=-\frac{1}{2}+\epsilon+i\,\mathrm{Im}\,l$; this is satisfactory provided the contribution from the semicircular contour vanishes as $l\to\infty$ and any additional singularities accrued by the shift in contour are sufficiently well-behaved. Regge investigated this representation for superpositions of Yukawa potentials at large values of l; he found that the number of poles for $\mathrm{Re}\,l>\mathscr{N}$ is finite, and that the semi-circular contribution vanished at large l for small z subject to certain technical details (see also [250–252]). Then $\mathscr{I}(k,\theta)$ can be written as a sum over poles plus a "background integral" along $\mathrm{Re}\,l=-\frac{1}{2}+\epsilon$:

$$\mathscr{I}(k,\theta)=\frac{1}{4k}\int_{-i\infty-1/2+\epsilon}^{i\infty-1/2+\epsilon}(2l+1)(S_l(k)-1)\frac{P_l(-\cos\theta)}{\sin\pi l}\,dl+\sum_i\frac{\beta_i(k)}{\sin\pi\alpha_i(k)}P_{\alpha_i(k)}(-\cos\theta)\tag{C.3}$$

in terms of the pole contributions $\alpha_i(k)$ with factor β_i. This is sometimes called the Sommerfeld–Watson representation, and now the pole furthest to the right (α_1 say, with $\mathrm{Re}\, l > -\frac{1}{2}$) dominates the integral, i.e. as $z = \cos\theta \to \infty$,

$$\mathscr{I}(k,z) \sim \frac{\beta_1(k)}{\sin \pi\alpha_1(k)} z^{\alpha_1} \ . \tag{C.4}$$

In general, as noted in Section 5, for potentials which are not superpositions of Yukawas (e.g. square well or other finite-range potentials), the contribution from the semicircular contour does not vanish (but see [250]). The Yukawa potentials thus are of great importance for the Mandelstam representation [92] (which cannot exist unless $\mathscr{I}$ is bounded by a finite power of z; (C.4) is also of significance in connection with relativistic scattering [246]).

To elucidate part of the physical significance of (C.3), consider the inner product

$$\mathscr{I}_l(k) = \frac{1}{2} \int_{-1}^{1} P_l(z) \mathscr{I}(k,z)\, \mathrm{d}z \ ,$$

along with the relation

$$\frac{1}{2} \int_{-1}^{1} P_l(z) P_\alpha(-z)\, \mathrm{d}z = \frac{\sin \pi\alpha}{\pi(\alpha - l)(\alpha + l + 1)} \tag{C.5}$$

where l is an integer and α is complex [243]. (This reduces to the standard result $(-1)^l/(2l+1)$ as $\alpha \to l$). Consider next the contribution of one Regge pole at $l = \alpha$ to $\mathscr{I}_l$, namely

$$\frac{\beta(k)}{\pi(\alpha(k) - l)(\alpha + l + 1)} \ . \tag{C.6}$$

If $\alpha \approx l$, this can provide a large contribution which may change rapidly with k (and hence energy), so if there exists an energy E_l such that $l = \mathrm{Re}\, \alpha$, then expanding α in a Taylor series about E_l yields

$$\alpha \simeq l + (E - E_l) \frac{\mathrm{d}\, \mathrm{Re}\, \alpha(E_l)}{\mathrm{d}E} + \mathrm{i}\ \mathrm{Im}\, \alpha(E_l) \ . \tag{C.7}$$

Then (B.6) can be rewritten in Breit–Wigner form as

$$\frac{\beta}{\pi(\alpha + l + 1)[\mathrm{d}\, \mathrm{Re}\, \alpha(E_l)/\mathrm{d}E]\{E - E_l + \mathrm{i}\Gamma/2\}} \tag{C.8}$$

where

$$\frac{\Gamma}{2} = \frac{\mathrm{Im}\, \alpha(E_l)}{\mathrm{d}\, \mathrm{Re}\, \alpha(E_l)/\mathrm{d}E} \ ,$$

I is the resonance half-width. Interpretation of these results to trace the trajectory of the Regge pole may be found in [246,249].

As noted in Section 5, a detailed analysis of a more effective version of the Watson transformation appeared in 1965 [33], along with many subsequent applications (see Section 5 for details). This *modified Watson transformation* is based on the application of the *Poisson sum*

formula [253,254] to the scattering amplitude or the wave field [33]; one form of this formula is:

$$\sum_{l=0}^{\infty} \varpi\left(l+\frac{1}{2},\beta,\theta\right) = \sum_{m=-\infty}^{\infty} (-1)^m \int_0^{\infty} \varpi(\lambda,\beta,\theta)\exp(2im\pi\lambda)\,\mathrm{d}\lambda \tag{C.9}$$

with an appropriate choice of deformed path in the complex λ-plane. This choice, which is different for different angular regions, leads also to Regge pole contributions and saddle-point-dominated background integrals. Khare and Nussenzveig point out that a saddle-point on the real axis is also a point of stationary phase in (C.9), and this characterizes an extremal path, i.e. a ray in geometrical optics, or a classical orbit in the particle context [47]. The integer m in (C.9) has the topological significance of a *winding number*, which is associated with the number of circumlocutions performed by a path around the center of the sphere. The details and subtleties of the Regge poles for the problem of scattering by a transparent sphere, the basis of the rainbow and the glory, are discussed in detail in Section 5.2, along with the significance of the Debye expansion in regaining rapid convergence of the residue series.

A succinct derivation of the Poisson sum formula has been provided in the book by Brink [186], and in view of its importance in CAM theory in particular, the formal proof is presented here (in a slightly modified form). From the theory of (complex) Fourier series, with standard Dirichlet conditions applying to $f(x)$ [255], it follows that

$$f(x) = \sum_{m=-\infty}^{\infty} A_m \mathrm{e}^{-2m\pi\mathrm{i}x/L}\,,$$

where

$$A_m = \frac{1}{L}\int_b^{b+L} f(\lambda)\mathrm{e}^{2m\pi\mathrm{i}\lambda/L}\,\mathrm{d}\lambda\,.$$

Note that since m and if l are integers,

$$f(x+l) = \sum_{m=-\infty}^{\infty} A_m \mathrm{e}^{-2m\pi\mathrm{i}x/L}\mathrm{e}^{-2m\pi\mathrm{i}x/L}$$

then choosing $L=1$ it follows that

$$f(x+l) = \sum_{m=-\infty}^{\infty} A_m \mathrm{e}^{-2m\pi\mathrm{i}x}\,,$$

where

$$A_m = \int_l^{l+1} f(\lambda)\mathrm{e}^{2m\pi\mathrm{i}\lambda}\,\mathrm{d}\lambda\,,$$

having chosen $b=l$. Therefore in particular

$$\begin{aligned}(6)\quad f\left(l+\frac{1}{2}\right) &= \sum_{m=-\infty}^{\infty} A_m \mathrm{e}^{-m\pi\mathrm{i}} = \sum_{m=-\infty}^{\infty} (-1)^m A_m \\ &= \sum_{m=-\infty}^{\infty} (-1)^m \int_l^{l+1} f(\lambda)\mathrm{e}^{2m\pi\mathrm{i}\lambda}\,\mathrm{d}\lambda\,. \end{aligned} \tag{C.10}$$

Hence

$$\sum_{l=0}^{\infty} f\left(l+\frac{1}{2}\right) = \sum_{m=-\infty}^{\infty} (-1)^m \sum_{l=0}^{\infty} \int_l^{l+1} f(\lambda) \mathrm{e}^{2m\pi \mathrm{i}\lambda}\, \mathrm{d}\lambda = \sum_{m=-\infty}^{\infty} (-1)^m \int_0^{\infty} f(\lambda) \mathrm{e}^{2m\pi \mathrm{i}\lambda}\, \mathrm{d}\lambda \tag{C.11}$$

and all that remains is to replace $f(\lambda)$ by $\varpi(\lambda, \beta, \theta)$ to recover (C.9).

Appendix D. The Chester–Friedman–Ursell (CFU) method

This technique is used when two saddle points in the complex λ-plane approach one another as a parameter governing the confluence gradually changes. The description given in [10] is followed here, notationally adapted to the rainbow problem. The complex integral of interest is

$$F(\tilde{\beta}, \tilde{\theta}) = \int g(\lambda) \mathrm{e}^{\tilde{\beta} f(\lambda, \tilde{\theta})}\, \mathrm{d}\lambda\,, \tag{D.1}$$

where $\tilde{\beta} = 2\beta$, the asymptotic expansion parameter is large and positive and $\tilde{\theta} = \theta - \theta_R$ is an independent parameter. In the 1-ray region (in particular), the integral is dominated by a single saddle-point $\bar{\lambda} = \bar{\lambda}(\tilde{\theta})$, in the neighborhood of which f and g are regular, so

$$f(\lambda, \tilde{\theta}) \approx f(\bar{\lambda}, \tilde{\theta}) + \tfrac{1}{2} f_{\lambda\lambda}(\bar{\lambda}, \tilde{\theta})(\lambda - \bar{\lambda})^2\,.$$

When the path of steepest descent is chosen to pass through $\bar{\lambda}$, a Gaussian-type integral occurs, and this yields the dominant contribution to the integral $F(\tilde{\beta}, \tilde{\theta})$ in a range $\Delta\lambda$ about the saddle point, where [103]

$$\Delta\lambda \sim (\tilde{\beta} | f_{\lambda\lambda}(\bar{\lambda}, \tilde{\theta})|)^{-1/2}\,.$$

The asymptotic expansion of $F(\tilde{\beta}, \tilde{\theta})$ then follows by standard techniques (a power-series expansion of g about $\bar{\lambda}$ and subsequent integration term by term [104]).

By contrast, in the 2-ray region there are two saddle points ($\bar{\lambda}', \bar{\lambda}''$ say) to be considered, and provided their ranges do not overlap, their contributions are additive. When the ranges *do* overlap, the above method is inapplicable because the series expansion of g has a radius of convergence of order $|\bar{\lambda}' - \bar{\lambda}''|$. Furthermore, at confluence (when $\tilde{\theta} = 0$), the second derivative $f_{\lambda\lambda}$ is zero, so the second term in f above becomes cubic in λ. For a decreasing sequence of *fixed* $\tilde{\theta} \neq 0$, $\tilde{\beta}$ can be chosen large enough so that the ranges still do not overlap (i.e. $\Delta\lambda \to 0$ as $\tilde{\beta} \to \infty$), but this does not yield a result *uniformly valid* in $\tilde{\theta}$. Similarly, a cubic approximation for $f(\lambda, \tilde{\theta})$ about $\tilde{\theta} = 0$ provides only a transitional approximation for $F(\tilde{\beta}, \tilde{\theta})$.

The basis of the CFU method is to transform the exponent in the equation (D.1) into an exact cubic by means of the change of variables

$$f(\lambda, \tilde{\theta}) = \tfrac{1}{3} u^3 - \xi(\tilde{\theta}) u + A(\tilde{\theta})$$

with the two saddle points transformed into $\pm \xi^{1/2}(\tilde{\theta})$. The essential feature is to design a mapping which preserves the saddle point structure [235]. Under these circumstances the mapping

$\lambda \leftrightarrow u$ is analytic and one-to-one near $u = 0$. It follows eventually from (D.1) that an asymptotic expansion of the following form is valid:

$$F(\tilde{\beta},\tilde{\theta}) = \exp[\tilde{\beta}A(\tilde{\theta})]\left\{\tilde{\beta}^{-1/3}\left[\sum_{s=0}^{M-1} a_s(\tilde{\theta})\tilde{\beta}^{-s} + O(\tilde{\beta}^{-M})\right]\mathrm{Ai}[\tilde{\beta}^{2/3}\xi(\tilde{\theta})]\right.$$

$$\left. + \tilde{\beta}^{-2/3}\left[\sum_{s=0}^{M-1} b_s(\tilde{\theta})\tilde{\beta}^{-s} + O(\tilde{\beta}^{-M})\right]\mathrm{Ai}'[\tilde{\beta}^{2/3}\xi(\tilde{\theta})]\right\} .$$

The a_s and b_s are the so-called 'CFU coefficients' determined from the expansion of the integrand after the transformation. When $|z| \gg 1$ and $|\arg z| < \pi$, it is known that [243]

$$\mathrm{Ai}'(z) \approx -z^{1/2}\,\mathrm{Ai}(z)$$

so that the Ai′ corrections in the expression above can become of the same order as the Ai terms. This uniform asymptotic expansion can be matched smoothly with the results from saddle-point analysis.

References

[1] K. Sassen, J. Opt. Soc. Am. 69 (1979) 1083.
[2] H.M. Nussenzveig, Sci. Am. 236 (1977) 116.
[3] M.V. Berry, Sci. Prog. Oxf. 57 (1969) 43.
[4] R. Greenler, Rainbows, Halos and Glories, Cambridge University Press, Cambridge, 1980.
[5] H.M. Nussenzveig, J. Math. Phys. 10 (1969) 82.
[6] H.M. Nussenzveig, J. Math. Phys. 10 (1969) 125.
[7] H.C. van de Hulst, Light Scattering by Small Particles, Dover, New York, 1981.
[8] R.A.R. Tricker, Introduction to Meteorological Optics, Elsevier, New York, 1970.
[9] C.F. Bohren, D.R. Huffman, Absorption and Scattering of Light by Small Particles, Wiley, New York, 1983.
[10] H.M. Nussenzveig, Diffraction Effects in Semiclassical Scattering, Cambridge University Press, Cambridge, 1992.
[11] C.B. Boyer, The Rainbow, from Myth to Mathematics, Princeton University Press, Princeton, NJ, 1987.
[12] N.A. Logan, Proc. IEEE 53 (1965) 773.
[13] D.E. Pedgely, Weather 41 (1986) 401.
[14] J.D. Walker, Am. J. Phys. 44 (1976) 421.
[15] D.K. Lynch, W. Livingston, Color and Light in Nature, Cambridge University Press, New York, 1995.
[16] J.D. Austin, F.B. Dunning, Math. Teacher (1988) 484 (September issue); see also S. Janke, UMAP Module 724, COMAP, Inc., Lexington, MA, 1992.
[17] R.J. Whitaker, Phys. Teacher 12 (1974) 283.
[18] W.J. Humphreys, Physics of the Air, Dover, New York, 1964.
[19] K.W. Ford, J.A. Wheeler, Ann. Phys. 7 (1959) 259.
[20] E. Hundhausen, H. Pauly, Z. Physik 187 (1965) 305.
[21] G.B. Airy, Trans. Camb. Phil. Soc. 6 (1838) 379.
[22] G.R. Fowles, Introduction to Modern Optics, Dover, New York, 1975.
[23] M. Born, E. Wolf, Principles of Optics, Pergamon, Oxford, 1965.
[24] G. Mie, Ann. Physik 25 (1908) 377.
[25] G.N. Watson, Proc. R. Soc. A 95 (1918) 83.
[26] J. Adam, Phys. Rep. 142 (1986) 263.
[27] J. Adam, J. Math. Phys. 30 (1989) 744.

[28] J. Adam, Wave Motion 12 (1990) 385.
[29] C.B. Boyer, Am. J. Phys. 18 (1950) 360.
[30] B. Van der Pol, H. Bremmer, Phil. Mag. 24 (1937) 141.
[31] B. Van der Pol, H. Bremmer, Phil. Mag. 24 (1937) 825. (See also 25 (1938) 817.)
[32] A. Sommerfeld, Partial Differential Equations in Physics, Academic Press, New York, 1964 (Appendix *II*).
[33] H.M. Nussenzveig, Ann. Phys. 34 (1965) 23.
[34] H.M. Nussenzveig, in: J.E. Bowcock (Ed.), Methods and Problems of Theoretical Physics, North-Holland, Amsterdam, 1970, p. 203.
[35] J.D. Jackson, Phys. Rep. 320 (1999) 27.
[36] J.D. Jackson, Classical Electrodynamics, 3rd Edition, Wiley, New York, 1998.
[37] J.E. Macdonald, Am. J. Phys. 31 (1963) 282.
[38] J.M. Pernter, F.M. Exner, Meteorologische Optik, W. Braumüller, Vienna, 1910.
[39] E. Buchwald, Ann. Phys. 43 (1948) 488.
[40] W.V.R. Malkus, R.H. Bishop, R.O. Briggs, NACA Technical Notes 1622 (1948).
[41] R. Penndorf, J. Opt. Soc. Am. 52 (1962) 402.
[42] R.T. Wang, H.C. van de Hulst, Appl. Opt. 30 (1991) 106.
[43] A. Ungut, G. Grehan, G. Gouesbet, Appl. Opt. 20 (1981) 2911.
[44] C.W. Querfeld, J. Opt. Soc. Am. 55 (1965) 105.
[45] A.J. Patitsas, Can. J. Phys. 50 (1972) 3172.
[46] W.J. Wiscombe, Appl. Opt. 19 (1980) 1505.
[47] V. Khare, H.M. Nussenzveig, in: U. Landman (Ed.), Statistical Mechanics and Statistical Methods in Theory and Application, Plenum Press, New York, 1977.
[48] H.C. Bryant, N. Jarmie, Ann. Phys. (NY) 47 (1968) 127. See also Chapter 8 in Light from the Sky (readings from Scientific American), Freeman, San Francisco, 1980.
[49] G.P. Können, Polarized Light in Nature, Cambridge University Press, Cambridge, 1985.
[50] H.C. van de Hulst, J. Opt. Soc. Am. 37 (1947) 16.
[51] See M. Born, E. Wolf, Principles of Optics, Pergamon, Oxford, 1965 and references therein.
[52] Y.G. Naik, R.M. Joshi, J. Opt. Soc. Am. 45 (1955) 733.
[53] V. Khare, in: A.D. Boardman (Ed.), Electromagnetic Surface Modes, Wiley, New York, 1982 (Chapter 11).
[54] H.C. Bryant, A.J. Cox, J. Opt. Soc. Am. 56 (1966) 1529.
[55] T.S. Fahlen, H.C. Bryant, J. Opt. Soc. Am. 58 (1968) 304.
[56] H. Maecker, Ann. Phys. 4 (1948) 409.
[57] B.W. Woodward, H.C. Bryant, J. Opt. Soc. Am. 57 (1967) 430.
[58] D.S. Langley, M.J. Morrell, Appl. Opt. 30 (1991) 3459.
[59] V. Khare, H.M. Nussenzveig, Phys. Rev. Lett. 33 (1974) 976.
[60] J.R. Probert-Jones, J. Opt. Soc. Am. A 1 (1984) 822.
[61] H.M. Nussenzveig, J. Opt. Soc. Am. 69 (1979) 1068.
[62] D.S. Langley, P.L. Marston, Phys. Rev. Lett. 47 (1981) 913.
[63] P.L. Marston, D.S. Langley, J. Opt. Soc. Am. 72 (1982) 456.
[64] H.M. Nussenzveig, W.J. Wiscombe, Optics Lett. 5 (1980) 455.
[65] L.I. Schiff, Quantum Mechanics, 3rd Edition, McGraw-Hill, New York, 1968.
[66] K.W. Ford, J.A. Wheeler, Ann. Phys. 7 (1959) 287.
[67] H. Bremmer, Terrestrial Radio Waves, Elsevier, Amsterdam, 1949.
[68] M.V. Berry, Proc. Phys. Soc. 89 (1966) 479.
[69] E.A. Mason, L.J. Monchick, J. Chem. Phys. 41 (1964) 2221.
[70] E.A. Mason, R.J. Munn, F.J. Smith, Endeavour 30 (1971) 91.
[71] U. Buck, Rev. Mod. Phys. 46 (1974) 369.
[72] C. Chester, B. Friedman, F. Ursell, Proc. Camb. Phil. Soc. 53 (1957) 599.
[73] M.J. Lighthill, Waves in Fluids, Cambridge University Press, Cambridge, 1987.
[74] R.E. Langer, Phys. Rev. 51 (1937) 669.
[75] S. Jorna, Proc. R. Soc. A 281 (1964) 99.
[76] S.I. Rubinow, Ann. Phys. 14 (1961) 305.

[77] C.L. Pekeris, Proc. Symp. Appl. Math. 2 (1950) 71.
[78] E.T. Copson, Asymptotic Expansions, Cambridge University Press, Cambridge, 1965.
[79] N.F. Mott, H.S.W. Massey, The Theory of Atomic Collisions, 2nd Edition, Clarendon Press, Oxford, 1949.
[80] E.A. Remler, Phys. Rev. A 3 (1971) 1949.
[81] W.G. Rich, S.M. Bobbio, R.L. Champion, L.D. Doverspike, Phys. Rev. A 4 (1971) 2253.
[82] D.A. Goldberg, S.M. Smith, Phys. Rev. Lett. 33 (1974) 715.
[83] W.A. Friedman, K.W. McVoy, G.W.T. Shuy, Phys. Rev. Lett. 33 (1974) 308.
[84] W.H. Miller, J. Chem. Phys. 48 (1968) 464.
[85] W.H. Miller, J. Chem. Phys. 51 (1969) 3631.
[86] J.F. Boyle, Mol. Phys. 22 (1971) 993.
[87] U. Buck, M. Kick, H. Pauly, J. Chem. Phys. 56 (1972) 3391.
[88] L.G. Guimarães, H.M. Nussenzveig, Opt. Commun. 89 (1992) 363.
[89] L.G. Guimarães, H.M. Nussenzveig, J. Mod. Opt. 41 (1994) 625.
[90] H.M. Nussenzveig, Nucl. Phys. 11 (1959) 499.
[91] R.M. Eisberg, Fundamentals of Modern Physics, Wiley, New York, 1961.
[92] H.M. Nussenzveig, Causality and Dispersion Relations, Academic Press, New York, 1972.
[93] W.J. Wiscombe, Appl. Opt. 19 (1980) 1505.
[94] M.V. Berry, Proc. Phys. Soc. 88 (1966) 285.
[95] L. Bertocchi, S. Fubini, G. Furlan, Nuovo Cimento 35 (1965) 596.
[96] C.J. Bollini, J.J. Giambiagi, Nuovo Cimento 26 (1962) 619.
[97] C.J. Bollini, J.J. Giambiagi, Nuovo Cimento 28 (1963) 341.
[98] A.O. Barut, F. Calogero, Phys. Rev. 128 (1962) 1383.
[99] A.Z. Patashinskii, V.L. Pokrovskii, I.M. Khalatnikov, JETP 17 (1963) 1387.
[100] R.G. Newton, The Complex j-Plane, Benjamin, New York, 1964.
[101] P.J. Debye, Ann. Phys. Ser. 4 30 (1909) 57.
[102] V. Khare, Ph.D. Thesis, University of Rochester, NY, 1975.
[103] N.G. de Bruijn, Asymptotic Methods in Analysis, North-Holland, Amsterdam, 1958.
[104] F.W.J. Olver, Asymptotics and Special Functions, Academic Press, New York, 1974.
[105] G.N. Watson, A Treatise on the Theory of Bessel Functions, Cambridge University Press, Cambridge, 1966.
[106] J.B. Keller, in: L.M. Graves (Ed.), Calculus of Variations and its Applications, Proceedings of Symposia in Applied Mathematics, Vol. 8, McGraw-Hill, New York, 1958.
[107] G. Szegö, Proc. London Math. Soc. 36 (2) (1934) 427.
[108] V. Khare, H.M. Nussenzveig, Phys. Rev. Lett. 38 (1977) 1279.
[109] J.M. Peterson, Phys. Rev. 125 (1962) 955; K.W. McVoy, L. Heller, M. Bolsterli, Rev. Mod. Phys. 39 (1967) 245; K.W. McVoy, Ann. Phys. 43 (1967) 91.
[110] P.J. Wyatt, Phys. Rev. 127 (1962) 1837.
[111] S. Rösch, Appl. Opt. 7 (1968) 233.
[112] V.A. Fock, Electromagnetic Diffraction and Propagation Problems, Pergamon Press, Oxford, 1965.
[113] M.A. Leontovich, Bull. Acad. Sci. USSR 8 (1944) 16.
[114] H.M. Nussenzveig, Comm. At. Mol. Phys. 23 (1989) 175.
[115] J.B. Keller, J. Opt. Soc. Am. 52 (1962) 116.
[116] M.V. Berry, J. Phys. B 2 (1969) 381.
[117] H.M. Nussenzveig, J. Phys. A 21 (1988) 81.
[118] H.M. Nussenzveig, W.J. Wiscombe, Phys. Rev. Lett. 59 (1987) 1667.
[119] H.M. Nussenzveig, W.J. Wiscombe, Phys. Rev. A 43 (1991) 2093.
[120] M.V. Berry, K.E. Mount, Rep. Prog. Phys. 35 (1972) 315.
[121] H.M. Nussenzveig, W.J. Wiscombe, Phys. Rev. Lett. 45 (1980) 1490.
[122] H.M. Nussenzveig, in: S. Haroche, J.C. Gay, G. Grynberg (Eds.), Atomic Physics 11, World Scientific Press, Singapore, 1989, p. 421 see also [114].
[123] B.R. Johnson, J. Opt. Soc. Am. 10 (1993) 343.
[124] F.M. Labianca, J. Acoust. Soc. Am. 53 (1973) 1137.
[125] J. Adam, Phys. Rep. 142 (1986) 263 (see Appendix 5).

[126] G.P. Können, J.H. de Boer, 18 (1979) 1961.
[127] J. Bricard, Ann. Phys. 14 (1940) 148.
[128] J.R. Meyer-Arendt, Introduction to Classical and Modern Optics, Prentice-Hall, Englewood Cliffs, NJ, 1972 (Chapter 3.1).
[129] H. Jeffreys, B.S. Jeffreys, Methods of Mathematical Physics, Cambridge University Press, Cambridge, 1966.
[130] H.R. Pruppacher, K. Beard, Q.J.R. Meteorol. Soc. 96 (1970) 247.
[131] A.B. Fraser, J. Atmos. Sci. 29 (1972) 211.
[132] J.D. Walker, Sci. Am. 237 (July 1977) 138.
[133] J.D. Walker, Sci. Am. 242 (June 1980) 174.
[134] D. Walker, Weather 5 (1950) 324.
[135] J.V. Dave, Appl. Opt. 8 (1969) 155.
[136] S.D. Mobbs, J. Opt. Soc. Am. 69 (1979) 1089.
[137] D.K. Lynch, P. Schwartz, Appl. Opt. 30 (1991) 3415.
[138] W.V.R. Malkus, Weather 10 (1955) 331.
[139] F.E. Voltz, Physics of Precipitation, American Geophysical Union, Washington DC, 1960.
[140] J.A. Lock, J. Opt. Soc. Am. A 6 (1989) 1924.
[141] A.B. Fraser, J. Opt. Soc. Am. 73 (1983) 1626.
[142] T. Young, Phil. Trans. Roy. Soc. xcii (1802) 12387.
[143] A.F. Spilhaus, J. Meteorol. 5 (1948) 108.
[144] W. Möbius, Abh. Kgl. Saechs. Ges. Wiss. Math. Phys. Kl. 30 (1907) 108.
[145] A.B. Fraser, J. Atmos. Sci. 29 (1972) 211.
[146] M.G.J. Minnaert, Light and Colour in the Open Air, Dover, New York, 1954.
[147] G.P. Können, J. Opt. Soc. Am. A 4 (1987) 810.
[148] M.G.J. Minnaert, Light and Colour in the Outdoors, Springer, New York, 1993.
[149] A.W. Green, J. Appl. Meteorol. 14 (1975) 1578.
[150] F.E. Voltz, Handbuch der Geophysik, Vol. 8, Borntrager, Berlin, 1961.
[151] P.L. Marston, E.H. Trinh, Nature 312 (1984) 529.
[152] H.R. Pruppacher, J.D. Klett, Microphysics of Clouds and Precipitation, Reidel, Dordrecht, 1978.
[153] J.A. Lock, Appl. Opt. 26 (1987) 5291.
[154] R.W. Wood, Physical Optics, Macmillan, New York, 1934.
[155] S.D. Gedzelman, J. Opt. Soc. Am. A 5 (1988) 1717.
[156] D.S. Zrnić, R.J. Doviak, P.R. Mahapatra, Radio Sci. 19 (1984) 75.
[157] R. Rasmussen, C. Walcek, H.R. Pruppacher, S.K. Mitra, J. Lew, V. Levizzani, P.K. Wang, U. Barth, J. Atmos. Sci. 42 (1985) 1647.
[158] S.T. Shipley, J.A. Weinman, J. Opt. Soc. Am. 68 (1978) 130.
[159] K-N. Liou, J.E. Hansen, J. Atmos. Sci. 28 (1971) 995.
[160] S. Bosanac, Molec. Phys. 35 (1978) 1057.
[161] J.N.L. Connor, in: M.S. Child (Ed.), Semiclassical Methods in Molecular Scattering and Spectroscopy, Reidel, Dordrecht, 1980, p. 45.
[162] K.E. Thylwe, J. Phys. A 16 (1983) 1141.
[163] K.E. Thylwe, J.N.L. Connor, J. Phys. A 18 (1985) 2957.
[164] J.N.L. Connor, J.B. Delos, C.E. Carlson, Mol. Phys. 31 (1976) 1181.
[165] S. Bosanac, J. Math. Phys. 19 (1978) 789.
[166] J.N.L. Connor, R.A. Marcus, J. Chem. Phys. 55 (1971) 5636.
[167] J.M. Mullen, B.S. Thomas, J. Chem. Phys. 58 (1973) 5216.
[168] J.N.L. Connor, D. Farrelly, D.C. Mackay, J. Chem. Phys. 74 (1981) 3278.
[169] D. Beck, J. Chem. Phys. 37 (1962) 2884.
[170] U. Buck, H. Pauly, J. Chem. Phys. 54 (1971) 1929.
[171] U. Buck, Rev. Mod. Phys. 46 (1974) 369.
[172] U. Buck, Adv. Chem. Phys. 30 (1975) 313.
[173] U. Buck, in: G. Scoles (Ed.), Atomic and Molecular Beam Methods, Vol. 1, Oxford University Press, Oxford, 1988, p. 499.

[174] A.W. Kleyn, Comm. At. Mol. Phys. 19 (1987) 133.
[175] N. Neskovic (Ed.), Rainbows and Catastrophes, Boris Kidric Institute of Nuclear Science, Belgrade, 1990.
[176] J.J.H. van den Biesen, R.M. Hermans, C.J.N. van den Meijdenberg, Physica A 115 (1982) 396.
[177] R.B. Bernstein, in: J. Ross (Ed.), Vol. 10, Advances in Chemical Physics, Wiley, New York, 1966, p. 313.
[178] see also [92].
[179] P. Kong, E.A. Mason, R.J. Munn, Am. J. Phys. 38 (1970) 294.
[180] N. Levinson, Kgl. Danske Videnskab. Selskab., Mat.-fys. Medd. 25 (9) (1949).
[181] E.A. Mason, C. Nyeland, J.J.H. van den Biesen, C.J.N. van den Meijdenberg, Physica A 116 (1982) 133.
[182] A. Schutte, D. Bassi, F. Tommassini, G. Scoles, Phys. Rev. Lett. 29 (1972) 979.
[183] S. Bosanac, Mol. Phys. 36 (1978) 453.
[184] S. Bosanac, Phys. Rev. A 19 (1979) 125.
[185] H.J. Korsch, K.E. Thylwe, J. Phys. B 16 (1983) 793.
[186] D.M. Brink, Semiclassical Methods for Nucleus-Nucleus Scattering, Cambridge University Press, Cambridge, 1985.
[187] W.E. Frahn, Diffractive Processes in Nuclear Physics, Oxford University Press, Oxford, 1985.
[188] R. da Silveira, in: N. Neskovic (Ed.), Rainbows and Catastrophes, Boris Kidric Institute of Nuclear Science, Belgrade, 1990, p. 103.
[189] M.P. Pato, M.S. Hussein, Phys. Rep. 189 (1990) 127.
[190] M.S. Hussein, K.W. McVoy, in: D. Wilkinson (Ed.), Progress in Nuclear and Particle Physics, Vol. 12, Pergamon Press, Oxford, 1984, p. 103; see also R.C. Fuller, Phys. Rev. C 12 (1975) 1561.
[191] L.W. Put, A.M.J. Paans, Nucl. Phys. A 291 (1977) 93.
[192] R.C. Fuller, P.J. Moffa, Phys. Rev. C 15 (1977) 266.
[193] D.M. Brink, N. Takigawa, Nucl. Phys. A 279 (1977) 159.
[194] N. Takigawa, Y.S. Lee, Nucl. Phys. A 292 (1977) 173.
[195] M.S. Hussein, H.M. Nussenzveig, A.C.C. Villari, J.L. Cardaso, Phys. Lett. B 114 (1982) 1.
[196] R. Lipperheide, Nucl. Phys. A 469 (1987) 190.
[197] C. Marty, Z. Phys. A 309 (1983) 261.
[198] J. Barrette, N. Alamanos, Nucl. Phys. A 441 (1985) 733.
[199] M. Ueda, N. Takigawa, Nucl. Phys. A 598 (1996) 273.
[200] M. Ueda, M.P. Pato, M.S. Hussein, N. Takigawa, Phys. Rev. Lett. 81 (1998) 1809.
[201] M. Ueda, M.P. Pato, M.S. Hussein, N. Takigawa, Nucl. Phys. A 648 (1999) 229.
[202] B. Schrempp, F. Schrempp, Phys. Lett. B 70 (1977) 88; see also B. Schrempp, F. Schrempp, Nuovo Cimento Lett. 20 (1977) 95.
[203] B. Schrempp, F. Schrempp, Nucl. Phys. B 163 (1980) 397; see also P.D.B. Collins, An Introduction to Regge Theory and High Energy Physics, Cambridge University Press, Cambridge, 1977.
[204] B.R. Levy, J.B. Keller, Canad. J. Phys. 38 (1960) 128.
[205] S.H. Fricke, M.E. Brandan, K.W. McVoy, Phys. Rev. C 38 (1988) 682.
[206] M.E. Brandan, G.R. Satchler, Phys. Rep. 285 (1997) 143.
[207] R.P. Feynman, R.B. Leighton, M. Sands, The Feynman Lectures in Physics, Addison-Wesley, Reading, MA, 1964.
[208] M.V. Berry, C. Upstill, in: E. Wolf (Ed.), Progress in Optics, Vol. 18, North-Holland, Amsterdam, 1980, p. 257.
[209] R. Thom, Structural Stability and Morphogenesis, Addison-Wesley, New York, 1989.
[210] V.I. Arnold, Russian Math. Surveys 26 (1968) 1.
[211] R. Gilmore, Catastrophe Theory for Scientists and Engineers, Wiley, New York, 1981.
[212] T. Poston, I. Stewart, Catastrophe Theory and its Applications, Pitman, Boston, 1978.
[213] H. Trinkaus, F. Drepper, J. Phys. A 10 (1977) L11.
[214] J.N.L. Connor, Mol. Phys. 31 (1976) 33.
[215] J.N.L. Connor, M.S. Child, Mol. Phys. 18 (1970) 653.
[216] J.N.L. Connor, Mol. Phys. 26 (1973) 1217.
[217] M.V. Berry, J. Phys. A 8 (1975) 566.
[218] U. Garibaldi, A.C. Levi, R. Spadacini, G.E. Tommei, Surf. Sci. 48 (1975) 649.

[219] J.N.L. Connor, Mol. Phys. 27 (1974) 853.
[220] M.V. Berry, Adv. Phys. 25 (1976) 1.
[221] J. Larmor, Proc. Camb. Phil. Soc. 7 (1891) 131.
[222] T. Pearcey, Phil. Mag. 37 (1946) 311.
[223] M.V. Berry, J. Phys. A 10 (1977) 2061.
[224] P.L. Marston, E.H. Trinh, Nature 312 (1984) 529.
[225] P.L. Marston, Opt. Lett. 10 (1985) 588.
[226] P.L. Marston, J. Acoust. Soc. Am. 81 (1987) 226.
[227] C.L. Dean, P.L. Marston, Appl. Opt. 30 (1991) 3443.
[228] H.J. Simpson, P.L. Marston, Appl. Opt. 30 (1991) 3468.
[229] P.L. Marston, Appl. Opt. 19 (1980) 680.
[230] J.F. Nye, Nature 312 (1984) 531.
[231] J.A. Lock, Appl. Opt. 26 (1987) 5291.
[232] J.D. Walker, Am. J. Phys. 44 (1976) 421.
[233] J.D. Walker, Sci. Am. 237 (1977) 138.
[234] J.D. Walker, Sci. Am. 242 (1980) 174.
[235] M.V. Berry, K.E. Mount, Rep. Prog. Phys. 35 (1972) 315.
[236] W.H. Miller, Adv. Chem. Phys. 25 (1974) 63.
[237] T.F. George, Ann. Rev. Phys. Chem. 24 (1973) 263.
[238] L.D. Landau, E.M. Lifshitz, Mechanics, Pergamon Press, Oxford, 1960.
[239] R.G. Newton, Scattering Theory of Waves and Particles, Springer, New York, 1982.
[240] W.D. Myers, in: R.L. Robinson, et al., (Eds.), Proceedings of the International Conference on Reactions between Complex Nuclei, Vol. 2, North-Holland, Amsterdam, 1974, p. 1.
[241] Sir Isaac Newton, Opticks, Q.31, Book 3, Part 1, Dover, New York, 1979, p. 1704.
[242] R. Haberman, Elementary Applied Differential Equations, 3rd Edition, Prentice-Hall, Englewood Cliffs, NJ, 1987.
[243] M. Abramowitz, I.A. Stegun (Eds.), Handbook of Mathematical Functions, Dover, New York, 1972.
[244] J.J. Bowman, T.B.A. Senior, P.L.E. Uslenghi (Eds.), Electromagnetic and Acoustic Scattering by Simple Shapes, North-Holland, Amsterdam, 1969.
[245] V.A. Fock, Diffraction of Radio Waves Around the Earth's Surface, Publishers of the USSR Academy of Sciences, Moscow, 1946.
[246] S.C. Frautschi, Regge Poles and *S*-Matrix Theory, Benjamin, New York, 1963.
[247] J.R. Taylor, Scattering Theory, Wiley, New York, 1972 (Chapter 15).
[248] J.A. Adam, Astrophys. Space Sci. 220 (1994) 179.
[249] T. Regge, Nuovo Cimento 18 (1960) 947.
[250] A.O. Barut, F. Calogero, Phys. Rev. 128 (1962) 1383.
[251] F. Calogero, Nuovo Cimento 28 (1963) 701.
[252] H. Cheng, Nuovo Cimento 45 (1966) 487.
[253] E.C. Titchmarsh, Introduction to the Theory of Fourier Integrals, Oxford University Press, Oxford, 1937.
[254] P.M. Morse, H. Feshbach, Methods of Theoretical Physics, Vol. 1, McGraw-Hill, New York, 1953.
[255] M.A. Pinsky, Introduction to Partial Differential Equations with Applications, McGraw-Hill, New York, 1984.