In addition to the oral programme, nine poster presentations were made. Dr. David Williams (MSSL) presented a study of the eruption of a kink-unstable filament in region NOAA 10696. Dr. Adam Rees (Imperial) described multi-spacecraft observations of a complex solar-wind–ICME interaction. The complexity of the ICMEs demonstrates the necessity, in order to model fully these phenomena, to understand the source regions, the launch process, and the interaction with the solar wind. Dr. Chris Goff (MSSL) described the eruption of a flux rope and the rising of a plasmoid by using multi-wavelength observations of a CME. Dr. Jackie Davies (RAL) showed multi-satellite observations of the near-Earth plasma sheet and flank magnetopause during the passage of a CME past the Earth. Prof. L. Culhane (MSSL) showed that it is possible for confined flares to produce a disruption of the streamer belt.

Dr. Robert Bentley (MSSL) described recent advances in the European grid of solar observations, while Silvia Dalla (Manchester) presented the capabilities of the *ASTROGRID* framework, the UK's contribution to a global Virtual Observatory, for data analysis for Sun–Earth connection studies. A science case studying the geo-effectiveness of CMEs, based on solar, interplanetary, and geomagnetic data was demonstrated. Dr. Richard Stamper (RAL) described preparations for the International Heliophysical Year in 2007. The overall scientific objective of the IHY is to advance our understanding of the heliophysical processes that govern the Sun, Earth, and heliosphere; the Sun-to-Earth connection is thus a central part of its remit. Finally, Dr. Mike Hapgood (RAL) presented a poster describing the monitoring of the Sun–Earth connection for research and applications. He stressed the importance of the monitoring data, and its crucial rôle in providing both the context for and the link to space-weather applications. — LOUISE HARRA & CHRIS OWEN.

UNDERSTANDING ASTRONOMICAL REFRACTION

By Andrew T. Young
Astronomy Department, San Diego State University

Introduction

For the past two centuries, monographs[1,2,3,4] and textbooks[5,6,7,8] on spherical astronomy have all presented astronomical refraction in much the same way. The differential equation for the refracted ray is developed; series expansions are introduced that allow calculation of numerical values in the part of the sky where most astronomical observations are made; and the region near the horizon is usually ignored. Because these series expansions diverge before reaching the horizon, the few authors who treat refraction near the horizon have used entirely different expansions than the ones valid near the zenith, so that no unified picture emerges. In any case, the mathematical transformations used to evaluate the integrals entirely disguise the physics.

This standard treatment, while sufficient for the calculation and use of refraction tables, completely violates the spirit of Hamming's motto[9] that "The purpose of computing is insight, not numbers". In fact, the textbook presentation of refraction not only hides the physics of refraction behind changes of independent variable, but also misleads the reader by emphasizing small and moderate zenith distances, where refraction behaves quite differently than it does near the horizon. For example, most astronomers suppose that refraction is always proportional to the refractivity of air at the observer, even at the horizon. This approximation is useful at large altitudes, but is not exact anywhere; nor is it even roughly correct near the horizon. The result is a widespread misunderstanding of astronomical refraction, exemplified by Simon Newcomb's widely quoted[10,3,11] statement[6] that "There is, perhaps, no branch of practical astronomy on which so much has been written as this and which is still in so unsatisfactory a state". The truth of Newcomb's remark is underscored by the recent discovery[12] of both conceptual and numerical errors in Newcomb's textbook. Similarly, a textbook of spherical astronomy[13] has recommended, and the US Naval Observatory adopted[14], a refraction formula that is in error by more than Cassini's homogeneous model in *every* part of the sky. These errors result from a traditional emphasis on calculating refraction in a restricted part of the sky, while excluding the apparently uninteresting (but conceptually essential) region near the horizon. To overcome these mistakes, it is necessary to consider refraction more generally, paying particular attention to low, and even negative, altitudes; for it is only in this region that the *structure* of the atmosphere influences refraction appreciably, and effects that are present (but numerically negligible) near the zenith become large and obvious.

Those whose interest is confined to numerical values will find Fletcher's superb review[15] and Bennett's numerical comparisons[16], possibly supplemented by some more recent discussions[17,18,19] of approximate formulae, to be sufficient. Although calculating refraction is no longer "the foundation of astronomy", as Isaac Newton[20] called it, it remains essential for telescope pointing and control systems. Furthermore, refraction is needed to determine airmasses in correcting photometric observations for extinction, because the argument of the airmass function is refracted rather than geometric zenith distance[18]. Finally, refraction errors usually set the limit of accuracy in satellite geodesy, and in the use of the *Global Positioning System*. So a good understanding of refraction is required in observational astronomy, astrophysics, geophysics, and geodesy. Traditional textbooks do not provide this understanding, so a clearer treatment of refraction is needed. This article is meant to fill that need. As overemphasis on numerical calculations has obscured the optics of refraction, it is helpful to begin with some basic principles of symmetry, geometry, and physics.

Symmetry principles

Reversibility: The most basic symmetry property is the reversibility of light rays: light follows the same path between two points, regardless of the direction of propagation. This allows us to trace rays from the observer backward to their source — an extremely useful technique in what follows.

Path symmetry: If the atmosphere is horizontally stratified, so that the surfaces of constant density are concentric spheres — a good approximation — the path of a ray of light is symmetrical about its lowest point, where it is nearest the

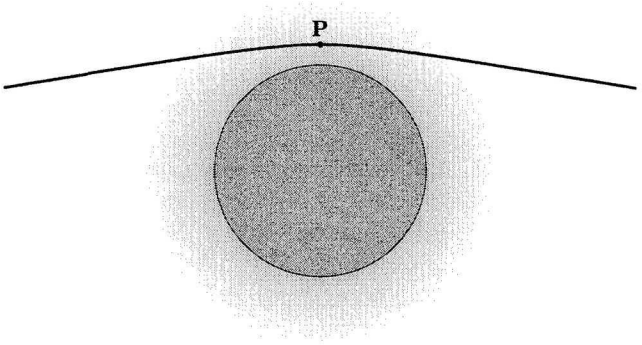

FIG. 1

The path of a ray (heavy line) is symmetrical about its perigee point, **P**. The shading represents the increasing density of the atmosphere toward the Earth's surface. The entire diagram is symmetrical about a vertical line through **P**, where the ray is horizontal. The height of the atmosphere and the curvature of the ray are greatly exaggerated.

Earth's centre (see Fig. 1). This point, **P**, is sometimes called the *apex* of the ray's trajectory; but as it is the lowest rather than the highest point, and need not even be the point of maximum curvature, the term *perigee* seems more appropriate.

Because of symmetry, the ray makes equal angles a at the points $\mathbf{O}_1$ and $\mathbf{O}_2$ where it crosses a level surface at a height h above the Earth's surface (see Fig. 2). If the Earth's radius is R_E, this surface has radius $R = R_E + h$. An observer at either of these crossings sees an object on the ray at an altitude a above the astronomical horizon when looking away from **P**, or at a depression of a when looking toward **P**.

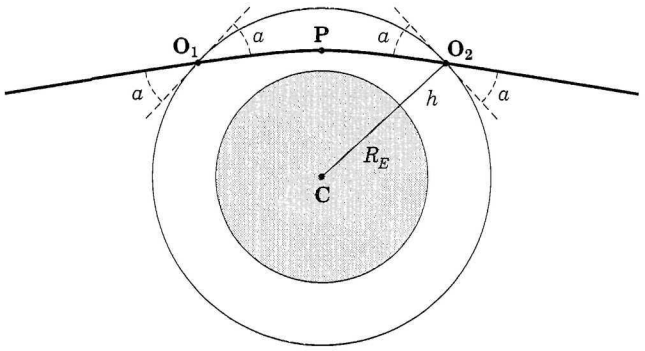

FIG. 2

The intersections $\mathbf{O}_1$ and $\mathbf{O}_2$ of the ray and a level surface of radius R at height h above the surface of the Earth, with radius R_E and centre at **C**. Dashed lines represent the planes of the astronomical horizons at $\mathbf{O}_1$ and $\mathbf{O}_2$.

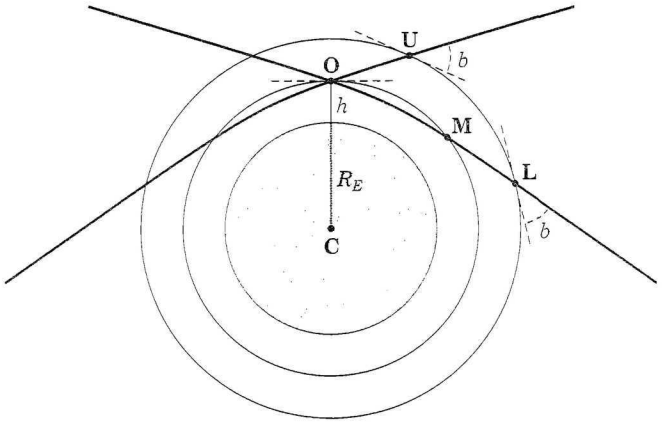

FIG. 3

Two rays (heavy curves) with the same perigee distance, seen by an observer at **O**. The entire figure is again symmetrical about the vertical line **OC**, but only the right half of the figure is labelled to avoid clutter. The observer's distance from the centre of the Earth is $R_o = R_E + h$, the radius of the circle through **O**. The outer circle is a level surface above the observer, at some larger radius $R_2 > R_o$. The plane of the observer's horizon, and those of the local horizons at **U** and **L** where the upper and lower rays cross the upper surface, are shown by dashed lines. The height h is smaller than in the previous figures, to keep the angles reasonably small.

This symmetry about **P**, where the ray is horizontal, means that no special method is needed to compute refraction below the astronomical horizon. If we can calculate the horizontal refraction from **P** up to $\mathbf{O}_2$, the refraction from $\mathbf{O}_1$ to $\mathbf{O}_2$ is just twice that. Then the total astronomical refraction at a depression of a is this double amount, plus the astronomical refraction at an altitude of a (*i.e.*, the ray-bending to the right of $\mathbf{O}_2$) — a fact first emphasized by Bouguer[21]. That fact was used extensively in the fourth paper[12] in this series.

Rotational symmetry: If the surfaces of constant density are concentric spheres, the shape of a refracted ray is independent of the position where it crosses one of these surfaces; the rays can be rotated rigidly about the centre of the Earth. So suppose we rotate two copies of Fig. 2 about **C**, so that the points $\mathbf{O}_1$ and $\mathbf{O}_2$ coalesce, as at **O** in Fig. 3. An observer at **O** sees two rays with the same shape and the same perigee height, one (**OU**) above the local horizon and the other (**OML**) an equal angle below it. The angular altitudes of the rays at **O** are the angles a in Fig. 2 (left unmarked in Fig. 3 to avoid clutter). These two rays cross a level surface **UL** above the observer at equal angles b. So rays at equal distances above and below the astronomical horizon meet any level surface above the observer at equal angles.

The equality of the angles b on the upper and lower rays is obvious in Fig. 3, where the symmetry of the rays is evident. In particular, one sees that the segment **OU** of the upper ray is identical to the segment **ML** of the lower ray. But if the parts of the rays to the left of **O** were omitted, the symmetry would be concealed, and it might seem surprising that the angles b are identical, because the ray segments **OU** and **OL** are so different in size and appearance.

Wegener's Principle: The fact that rays with apparent altitudes of $+a$ and $-a$ at the observer both meet any higher layer at the same angle b has important consequences. The equality of the b values for rays symmetrically placed above

and below the observer's astronomical horizon means that a plot of b as a function of a is symmetrical about $a = 0$, the observer's astronomical horizon. So the horizon ray itself meets any level surface above the observer at a local minimum of b (*i.e.*, a locally maximal angle of incidence, measured from the local normal).

This fact was emphasized by Alfred Wegener[22]. Because of this extremum in incidence angle, the refraction contribution from the whole atmosphere above the upper layer is nearly constant within a small zone of sky centred on the observer's astronomical horizon. We may call this *Wegener's Principle*.

Thus, the upper atmosphere plays essentially no rôle in the rapid variations of refraction with angular altitude near the horizon that distort the setting Sun. Therefore Wegener[22] explained distorted sunsets by treating in detail only the refraction produced near eye level, and used standard tables for the refraction produced by the bulk of the atmosphere. Even though most of the atmosphere ordinarily produces most of the astronomical refraction, sunset distortions are almost entirely due to atmospheric structure near and below eye level. In general, Wegener's Principle lets us separate the atmosphere into a large, boring upper part, and a small lower region that produces interesting effects near the horizon.

Geometry

Flattening: One of these interesting effects at the horizon is the flattening of the setting Sun, which Kepler[23] first treated quantitatively using Tycho's empirical refraction table. The flattening is due to a *gradient* in refraction at the horizon that raises the lower limb more than the upper limb. This refraction gradient is due entirely to the density gradient of the air just below the observer. To see why, consider first Fig. 4a, in which the lower air has constant density, so that the rays are straight. Just as in Fig. 3, the ray above the horizon at **O** corresponds to the ray extending to the right at **M**, so we can regard them as the *same* ray, rotated about **C** by the angle between the two rays — *i.e.*, an angle of $2a$. This angle is equal to the central angle **OCM**, because the ray **OM** is straight.

Equivalently, if we draw the perpendicular from **C** to the perigee point **P**, the angle **OCP** is equal to a, because both are complements of angle **COP**. Again, the central angle **OCM** equals the apparent angular separation $2a$ of the two rays at the observer. Because the refraction of the two rays is equal, and occurs entirely above the observer, their angular separation outside the atmosphere is also $2a$. Therefore the apparent separation of the rays seen by the observer at **O** is exactly the same as their separation above the atmosphere, and celestial objects are not flattened at the horizon. The angular magnification of objects at the astronomical horizon is unity.

Curvature and magnification: Now consider a ray passing through the same points, **O** and **M**, but bent by refraction (Fig. 4b). The curvature of the ray decreases the angles it makes with the local horizons at **O** and **M**. Once again, the refraction above the observer's level is exactly the same for the two rays at **O**. Now, however, the refraction of the lower ray at **O** is larger, by just the ray-bending θ between **O** and **M**. Thus the separation of the rays at the observer is smaller than their separation outside the atmosphere by this angle, θ.

Of course, the angular separation of the rays *outside* the atmosphere is still the central angle **OCM**, as can be seen from the similarity of the upper ray at **O**

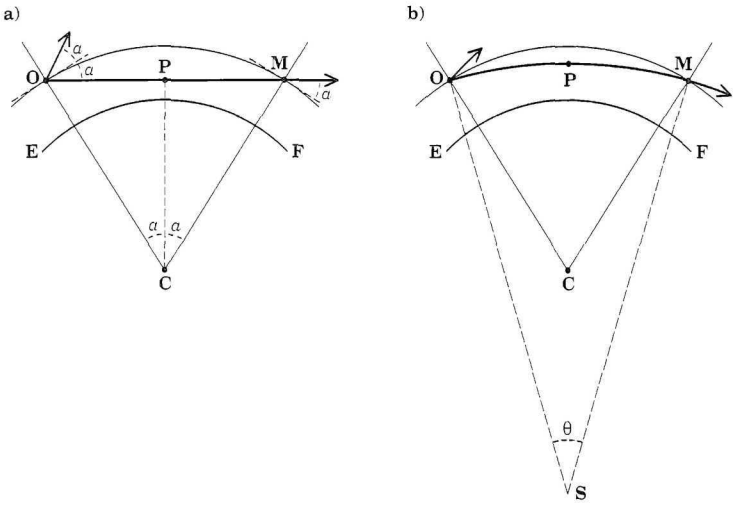

FIG. 4

a) A straight ray **OPM** in a constant-density atmosphere above the surface **EF** of the Earth, with centre at **C**. The upper ray at **O** corresponds to the ray **OU** in Fig. 3; the lower ray **OM** corresponds to **OML** in Fig. 3. Dashed lines again indicate the planes of the local horizons at **O** and **M**, and the angles a are indicated as in Fig. 2. b) A curved ray in an atmosphere with a density gradient below the observer. **S** is the centre of curvature of the refracted ray **OPM**, drawn here as the arc of a circle: note that **OS** is perpendicular to the ray **OPM** at **O**, and **SM** is perpendicular to this ray at **M**.

and the ray to the right of **M**, as before. So the apparent angular separation of the two rays at **O** has been diminished by exactly θ. As the true angular separation of the rays at infinity is **OCM**, and the apparent angular separation of the rays at **O** is (**OCM** – θ), the angular magnification m at the horizon is (**OCM** – θ)/**OCM**. But this ratio depends only on the ratio of the curvature of the ray to that of the Earth. The curvature of the ray is 1/**OS**, and that of the Earth is 1/**OC**; so the ratio of the curvatures is **OC**/**OS**. As the angles are small — a fraction of a degree — the length **OM** is practically the same, whether it is measured along the level surface or the ray. So, as the arc length **OM** is just the angle times the radius, the ratio θ/**OCM** = **OC**/**OS**. If this ratio is k, then m = (1 – k). That is, if the ray curvature is half that of the Earth, k = ½ and m = ½. If the ray curvature were ⅓ that of the Earth, k = ⅓ and m = ⅔. In the real atmosphere, k is about ⅙; so the Sun at the horizon should be flattened by about ⅙ of its horizontal diameter. (Kepler actually got ⅙ by interpolating in Tycho's table.) The ratio k (or its reciprocal, K = 1/k) is used in surveying to correct observed terrestrial altitudes for refraction.

If the ray curvature equals the Earth's, k = 1 and the magnification is zero: the Sun is flattened into an infinitely thin line at the horizon. This actually occurs at the upper edge of a duct, where rays are trapped below a layer that produces strong bending. (Ducts will be treated in more detail later.) Such phenomena were noticed by Chappell[24], whose photographs of sunsets at Lick Observatory often showed a "final singular long line, which oddly enough is substituted for the small tip of light that could reasonably be expected as the final glimpse of a bright descending sphere". To relate magnification at the

horizon to local atmospheric structure, we need the value of k, or the actual value of the ray's radius of curvature. This curvature depends on the vertical gradient of refractivity, and hence on the gradient of density, at the observer.

Ray bending

Density gradients: Consider an atmosphere at rest, in hydrostatic equilibrium. If the atmosphere were isothermal, the density would simply be proportional to the pressure at every level. Because the isothermal scale height is $H = kT/mg$, where m is the mass of an average molecule and g is the acceleration of gravity, H is very nearly 8 km. Then the density would diminish by nearly 1 part in 8000 for every metre of height. The refractivity $(n - 1)$ is very nearly proportional to the density of air, so the refractivity would also decrease by 1 part in 8000 per metre.

Before computing the ray bending due to this density gradient, let us consider the temperature gradient required to keep the density constant. If the surface temperature is 300 K, we would need to decrease the temperature by 1 part in 8000 per metre, giving a lapse rate of 300 K/8000 m = 0·0375 K/m, or 37·5°C/km, to maintain a uniform-density atmosphere in hydrostatic equilibrium. (This would be unstable against convection, but that is of no concern here.) Rays in this constant-density atmosphere would be straight, as in Fig. 4a. The value usually given[2,22] as the lapse rate corresponding to constant density is 34·2 K/km, corresponding to STP. As the value is proportional to the assumed surface temperature, which is usually above freezing, we can adopt 35 K/km as a more typical value. (Newcomb's value[6] of 1° in 34 metres at 10°C is too small; perhaps he meant 34°/km.)

Circular rays: Next, consider a ducted ray, which exactly follows the curve of the Earth. This case is easier to understand if one remembers that 'rays' of light are a non-physical abstraction; the closest thing to a 'ray' in reality is the central line of a beam of light. (This avoids the confusion that some people[25,26] have had in trying to apply the sine-law of refraction to horizontal 'rays' — although such problems had been correctly treated a century earlier[27,28].) Fig. 5 shows a horizontal beam of light, whose wavefronts are everywhere vertical. On the left,

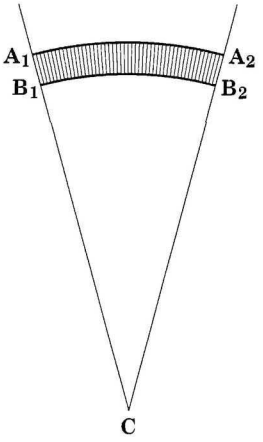

FIG. 5

A collimated horizontal beam of light, following the curve of the Earth, with centre at **C**. A few vertical wavefronts are shown; the surface of the Earth is omitted.

the vertical wavefront is $\mathbf{A}_1\mathbf{B}_1$; on the right, it has moved to $\mathbf{A}_2\mathbf{B}_2$. Obviously, the upper side of the beam moves farther (from $\mathbf{A}_1$ to $\mathbf{A}_2$) than the lower, which goes only from $\mathbf{B}_1$ to $\mathbf{B}_2$. The upper and lower edges of the beam traverse their distances in the same time, and contain the same number of wavelengths. So the speed at each edge must be proportional to its distance R from the centre of the Earth, **C**. But the speed of light in some medium, such as air, is inversely proportional to n, the refractive index of the medium. As the speed is proportional to R, the refractive index at each radial distance from **C** must be inversely proportional to R; so the product nR must be constant. This is the condition for a beam (or ray) to remain horizontal as it bends around the Earth.

Now consider the temperature gradient required to produce this condition. As R is about 6400 km, nR will remain constant if n decreases by 1 part in $6{\cdot}4 \times 10^6$ for each metre increase in height. But the refractivity $(n - 1)$, which is proportional to the air density, is only about 1/3200 of n; so the density must decrease by $3200/6{\cdot}4 \times 10^6$ m, or 1 part in 2000 per metre of height. The decrease in density due to the pressure gradient alone is 1 part in 8000 per metre, or ¼ of the required amount. So the temperature gradient must supply the remaining 3 parts in 8000. At 300 K, the temperature must increase upward by 3/8000 of the 300 K, or 900/8000 = 0·1125 degree per metre. As this gradient has the opposite sign from the usual lapse rate, it is a temperature inversion; the lapse rate is negative.

The argument presented here is crude, but close to the truth. For comparison with the rough value of −0·1125 K/m just derived, Lehn[29] gives −0·1127 K/m as the critical lapse rate, Wegener[22] gives −0·114 K/m, Newcomb[6] gives −117° per km and de Graaff Hunter[30] gives −0·066°F/foot, which corresponds to −0·116 K/m. Let us adopt −115 K/km in what follows. Notice that this critical temperature gradient is over 17 times larger in magnitude than the 6·5 K/km lapse rate of the Standard Atmosphere[31]. That means that the standard lapse rate has hardly any effect on ray curvature, which is due almost entirely to the pressure gradient under normal circumstances. (This fact justifies the use of an isothermal model at the start of this section.)

Bending and lapse rate: Now that we know the lapse rates required to produce straight rays and rays that circle the Earth indefinitely, we can calculate the bending that corresponds to any given lapse rate. The curvature of a horizontal ray is proportional to the vertical density gradient of the atmosphere. The density is inversely proportional to the temperature, so the density gradient is the negative of the temperature gradient or lapse rate, offset by the contribution of the pressure gradient. So, if a lapse rate of 35 K/km would produce straight rays ($k = 0$), and −115 K/km would produce circular rays ($k = 1$), a lapse rate of γ will produce a relative curvature of

$$k = \frac{\gamma - 35}{-115 - 35} = (35 - \gamma)/150.$$

Consequently, the standard lapse rate of 6·5 K/km corresponds to a ray curvature $K = 1/k$ about 5·3 times less than the Earth's curvature, while an isothermal atmosphere would produce a ratio of only 4·3. On the other hand, a convective atmosphere, with a lapse rate near 10 K/km, would give a curvature ratio of 6. As the atmosphere is near convective equilibrium during the day, this explains the typical flattening of the setting Sun. Surveyors and geodesists usually assume[32] a still larger value, $K = 7$, because their observations are usually made on warm afternoons at moderate elevations above sea level, and the higher temperature and lower pressure than assumed above both decrease the

curvature of the refracted ray. Lapse rates in the free atmosphere are limited by convection to the adiabatic lapse rate, though it can be exceeded within a few metres of a warm surface, which inhibits convection. But there is no limit in thermal inversions, where lapse rates exceeding a degree per metre are common. An inversion gradient of 20 K/m has been measured directly[33], corresponding to $K = 133$ and a radius of curvature of 48 km for a horizontal ray.

Historical remarks: This method of calculating the relation between lapse rate and radius of curvature of a ray is similar in spirit to that offered by Thomas Young[34], though he used the 'projectile hypothesis' for the propagation of light (*i.e.*, the emission model of Descartes and Newton).

The relation between the density gradient at the observer and the gradient of refraction at the astronomical horizon was first proved by J. B. Biot[35,36]. After mentioning the theorem (first proved by Oriani[37,38], and discussed in detail below) that the refraction out to zenith distances of 74° is approximately independent of the structure of the atmosphere, Biot[35] said, "But what has not been noticed is that there exists, ... the singularity of always being realised, in all possible constitutions of atmospheres, not just approximately, like that which we have just mentioned, but in an absolute and rigorous manner. ... Besides the unexpected singularity of finding an element of the horizontal refraction, independent of the state of distant layers, and of obtaining it, in all possible cases, without integration; besides the connection which results between the increase of refraction near the horizon and the equally observable variations of the refractive power starting from the bottom layer, the theorem which I am announcing has still other useful applications." As the magnification at the horizon is so obviously related to the refraction gradient, we may fairly attribute the magnification theorem to Biot, though he did not mention this particular "useful application" of his discovery.

The relation between ray pairs that are symmetrical about the astronomical horizon applies to all finite differences, as well as to the infinitesimal ones required to demonstrate Biot's magnification theorem. Because the atmosphere above the observer contributes equally to the refraction of the two rays, the difference in refraction at these two altitudes depends only on atmospheric structure between eye level and the height where the lower ray is horizontal. This explains why the inverse problem (of determining the temperature profile from the refraction profile) is well-posed below the astronomical horizon, even though it is ill-posed above it. The history of this problem, and the methods of solving it explicitly, have been discussed by Bruton and Kattawar[39,40].

Refraction

Approximations: The practical calculation of refraction always involves approximations. What level of accuracy is useful? Positional observations are rarely more precise than 0·1 arc second near the zenith, and the errors grow with about the square root of the airmass. So there is no practical use for calculations much better than a second of arc or so at the horizon, where the refraction is about 2000 arcsec; or a part in a thousand at moderate zenith distances, where the refraction is on the order of 100 arcsec. Generally, a part in a few thousand is the useful limit of accuracy for astronomical refraction calculations.

The sine law: We have been able to find the flattening of the setting Sun without calculus, and without actually calculating the refraction. It is even possible to calculate refraction for a simple atmospheric model without calculus,

as Cassini[41] did, before calculus had been invented (see below for details). Calculating refraction requires the sine law discovered empirically about 1600 by Harriot[42,43], rediscovered by Snel some 20 years later, and finally published in 1637 by Descartes[44], who had read Snel's unpublished manuscript. Although Descartes, Newton, Laplace, and many others pretend to 'derive' the law of refraction from theoretical considerations, it is really an experimental fact that all theories of light must accommodate.

The law of refraction is simply that

$$\frac{\sin\theta_1}{\sin\theta_2} = n,$$

where the angles of incidence (θ_1) and refraction (θ_2) are measured from the normal to the surface separating any two media. The constant ratio n is the *relative* index of refraction of the media. It is conventional to call the index of any material relative to a vacuum the *absolute* index of the material. This makes the refractive index of a vacuum exactly unity.

The plane-parallel model: Newton[45] showed that the sine law may be extended to a series of plane-parallel layers, so that the product $n \sin z$ (where the local zenith distance z is the angle from the normal at any interface) is conserved throughout the stack of layers. As he put it, "the Sum of all the Refractions will be equal to the single Refraction which it would have suffer'd in passing immediately out of the first Medium into the last." That is, refraction in a plane-parallel atmosphere is the same as in a single homogeneous layer having the refractive index at the observer.

In this single-slab model (Fig. 6), the angle of refraction inside the atmosphere is identical to the observed zenith distance z_o, and the law of refraction is just

$$n \sin z_o = \sin z_t,$$

where z_t is the object's true (unrefracted) zenith distance; so

$$z_t = \arcsin(n \sin z_o).$$

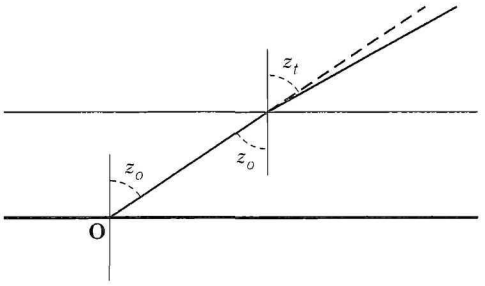

FIG. 6

Refraction in a plane-parallel slab. The observer is at **O**, on the surface of the (flat) Earth. Refraction occurs at the top of the slab of atmosphere. The dashed line is the extension of the observer's line of sight; the angle between it and the ray (solid) above the atmosphere is the astronomical refraction.

As the refraction $r = (z_t - z_o)$, it is exactly

$$r = \arcsin(n \sin z_o) - z_o, \tag{1}$$

which is clearly a nonlinear function of both n and z_o — particularly near the horizon, where the angles are near 90°, and the sine and arcsine functions have their greatest curvature.

On the other hand, near the zenith, the angles are small, and the small-angle approximation $\sin x \approx x$ is useful. This linearization gives

$$z_t \approx n z_o,$$

so that

$$r = z_t - z_o \approx n z_o - z_o = (n - 1) z_o.$$

Some textbooks[6,8] carry the small-angle approximation one order further, expanding $\sin z_t = \sin(z_o + r) = \sin z_o \cos r + \cos z_o \sin r$. Then, since r is always small (so that its cosine can be set to unity), one obtains $r \approx \sin r \approx (n - 1) \tan z_o$. This still hides the actual nonlinear dependence on refractivity.

However, Delambre[5] shows that $\tan(r/2)$ can be developed in a power series in $\tan z_o$. In this series, the coefficients of the terms involve successive powers of $(n^2 - 1)$, which he uses for the refractivity instead of $(n - 1)$ — a minor modification. The pairing of higher powers of the refractivity with higher powers of $\tan z_o$ shows how the rule that the refraction is approximately proportional to the refractivity breaks down near the horizon. Furthermore, the Earth's curvature makes the coefficient of the asymptotic formula for refraction near the zenith slightly different from the $(n - 1)$ of this plane-parallel model, even when the tangent approximation is usable.

A peculiarity of the plane-parallel model is that rays with grazing incidence ($z_t = 90°$) at the top of the atmosphere are seen at an apparent zenith distance $z_o = \arcsin(1/n)$, which corresponds to an angular altitude of about 1° 23′. Rays

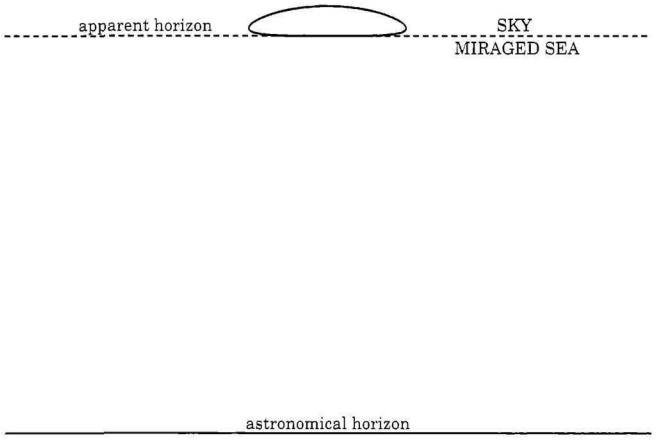

FIG. 7

The form of the setting Sun in the plane-parallel atmosphere. The 'horizon surface' (dashed) is nearly three solar diameters *above* the astronomical horizon; below it is a superior mirage of the Earth's surface.

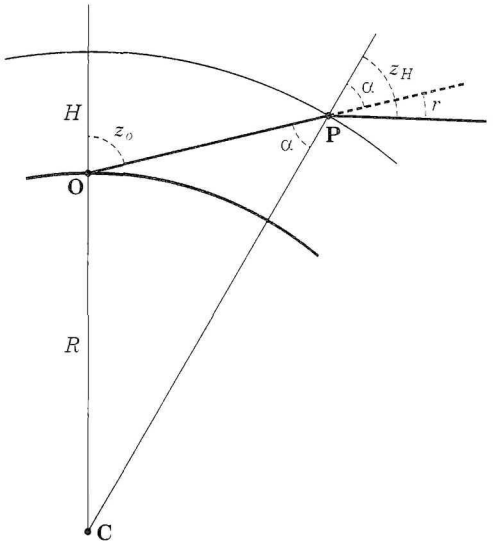

FIG. 8

Refraction at the top of a uniform spherical atmosphere. The refraction occurs at **P**, a height H above the observer at **O**. The dashed line is the extension of the refracted ray **OP**.

closer to the observer's horizon suffer total internal reflection at the upper surface of the slab atmosphere; in this zone of the sky, no astronomical objects are visible. Instead, a superior mirage of terrestrial objects would appear there, because the observer would be inside a duct. Such mirages were treated by Wegener[22] for a more realistic model, in which a smaller refractive-index jump produces the reflection, and the Earth's curvature is taken into account. Not only would the setting Sun disappear at the top of the duct, well above the astronomical horizon but, in addition, the vertical tangent of the arcsin function at an argument of unity causes infinite compression of the solar image at this elevated 'horizon surface' — see Fig. 7. (Such highly-flattened images are in fact seen just above ducts in the real atmosphere; see Fig. 8 of ref. 46 for an example.) The fact that the setting Sun is visible at the actual sea horizon, below the astronomical one, contradicts this model, and shows that the Earth's surface is convex, not flat.

Cassini's homogeneous model: It is slightly more realistic to bend the uniform slab to fit the curvature of the Earth (see Fig. 8). The ray **OP** inside the atmosphere of constant refractive index n is straight. In triangle **OPC**, the law of sines gives

$$\frac{\sin\alpha}{R} = \frac{\sin z_o}{R+H},$$

so

$$\sin\alpha = \frac{R\sin z_o}{R+H}.$$

The law of refraction, applied at **P**, is

$$\sin z_H = n \sin \alpha,$$

where z_H is the local zenith distance of the ray outside the homogeneous atmosphere. We now have expressions for $\sin \alpha$ and $\sin z_H$.

But $z_H = \alpha + r$, where r is the refraction; so

$$r = z_H - \alpha = \arcsin\left(\frac{nR \sin z_o}{R + H}\right) - \arcsin\left(\frac{R \sin z_o}{R + H}\right) \qquad (2)$$

is the refraction for this model. Again, the nonlinear dependence on n is obvious. This result is exact, and is obtained with just trigonometry.

The homogeneous model was first worked out by G. D. Cassini[41], and so is often called 'Cassini's model', though it was also assumed by Kepler[23], and can be traced back to Ptolemy[47]. Some authors[4,7] call Eqn. (2), or some approximation to it, 'Cassini's formula'; but Cassini never published a formula, just a verbal description of how the model works, with a table derived from it.

This model is surprisingly accurate out to moderate zenith distances, if we choose values of n and H that reproduce the actual conditions (refractive index, pressure, and density) at the observer. Ivory[48] first noticed that "The simple hypothesis of Cassini seems hardly to have met from astronomers with the attention it deserves; for, if we use accurate elementary quantities in the computation, it will determine the refractions to the extent of 74° from the zenith with the same degree of exactness as any of the other methods, without even excepting the formula of Laplace." Radau[2] also recognized its accuracy. Its errors (compared to the Standard Atmosphere) are only[12] 51 milli-arcsec at 74° zenith distance, 17 mas at 70°, and still smaller higher in the sky. The error remains below a second of arc out to $z_o = 81°$; but beyond that, the model quickly becomes useless, having an error of 13 arc minutes at the horizon.

The reduced height, H: As was pointed out above, the atmosphere would have constant density if the lapse rate were about 35°/km. This homogeneous atmosphere comes to an abrupt end where $T \to 0$. Rather than use this rough temperature gradient to compute the height of the atmosphere, it is more instructive to invoke hydrostatic equilibrium. The pressure at the bottom of the atmosphere is the weight per unit area of the material, *i.e.*,

$$p_0 = \rho_0 g H,$$

where ρ_0 is the density at the surface and g is the acceleration of gravity; so

$$H = p_0/\rho_0 g\ .$$

This expression for H is equivalent to the earlier version involving temperatures, if one assumes the ideal gas law. The "height of the homogeneous atmosphere" is a rather cumbersome phrase; Radau's term[2] "reduced height" is more concise. This height, which we first encountered in calculating ray curvatures, appears often in refraction theory, even in non-uniform atmospheres.

The refractive invariant

Geometric invariants: In the plane-parallel case, simple geometry relates the angle of refraction at one surface to the angle of incidence at the next interface:

these angles are equal. Within a homogeneous layer, the ray is straight, so the value of $\sin z$ remains constant; we may call $\sin z$ a *geometrical invariant* within such a layer. As the refractive index n is also constant, the product $(n \sin z)$ is conserved within each layer. But the law of refraction makes $(n \sin z)$ the same on both sides of each interface; so $(n \sin z)$ is conserved throughout the whole plane-parallel atmosphere. We may call it the *refractive invariant* for the plane-parallel model.

Refraction in a spherical atmosphere still involves a conserved quantity, but it is not $(n \sin z)$, because the curvature of the atmosphere makes z change along a straight ray within a layer of constant n. The angles at successive interfaces differ, even though the intervening layer is homogeneous. We need a geometric relation between these angles that takes account of the layer's curvature. So, what is the geometrical invariant for a straight ray in a curved layer? In Fig. 9, the length p of the perpendicular from the centre of the Earth **C** to the ray is $(R \sin z)$ at any point **P** on the ray, if z is the local zenith distance of the ray at **P**. Physicists like to call p the *impact parameter*; it is our required geometric invariant.

Refractive invariant: Now consider a single refraction at the top of a uniform layer with refractive index n. In Fig. 10, the ray is refracted at $\mathbf{P}_1$, where the local zenith distance on the vacuum side is z_1. The law of refraction tells us that z_2, the angle of refraction inside the atmosphere, is given by

$$n \sin z_2 = \sin z_1.$$

If we multiply this equation by R_1, the radius of the refracting surface, we have

$$nR_1 \sin z_2 = R_1 \sin z_1;$$

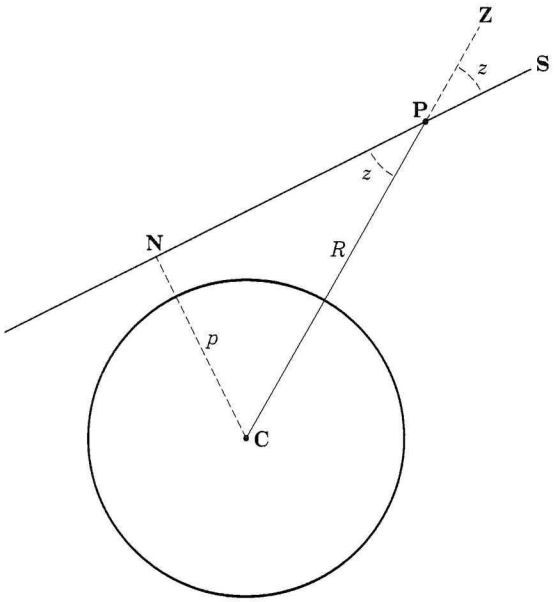

FIG. 9

The geometrical invariant is $p = R \sin z$ along an unrefracted ray, **SPN**. The local zenith distance is z at any point **P**, a distance R from the centre of the Earth, **C**.

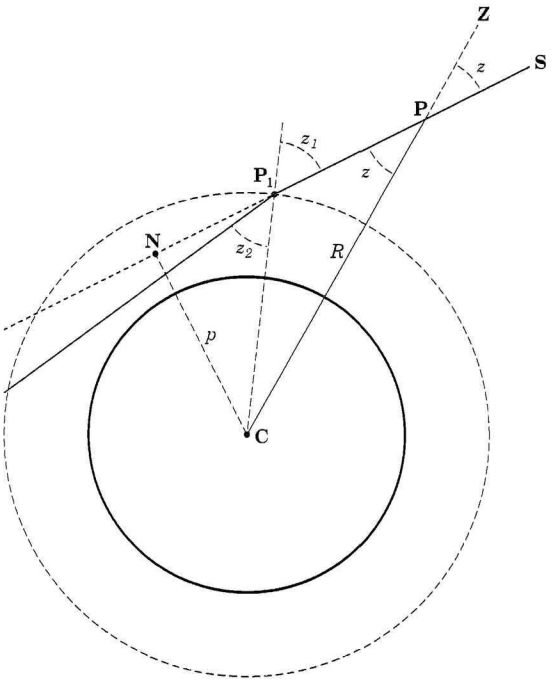

FIG. 10

The refractive invariant is $nR \sin z$ along a refracted ray, $\mathbf{SPP_1N}$, in a uniform atmosphere. The local zenith distance at point **P** is z. Outside the dashed surface, the refractive index is unity; inside, it is n. The ray is refracted at $\mathbf{P}_1$, where the angle of incidence is z_1, and the angle of refraction is z_2.

but $(R_1 \sin z_1)$ is just p, the geometric invariant for the incident ray. So we have

$$nR_1 \sin z_2 = p.$$

And of course a geometric invariant applies to the refracted part of the ray, as well: the product $(R \sin z)$ remains constant along the refracted ray inside the atmosphere. But then so does $(nR \sin z)$, if n is constant; and if $(nR \sin z) = p$ at $\mathbf{P}_1$, it must remain equal to p inside the refracting atmosphere. Thus $(nR \sin z) = p$ all along the ray, both inside and outside the atmosphere. This is the refractive invariant for the spherical model.

If we add a second refracting surface inside the first, the same argument can be repeated, just as it was for the plane-parallel atmosphere by Newton (see Smart[8] for the details). The geometric invariant $(R \sin z)$ in each spherical layer takes the place of the geometric invariant $(\sin z)$ for each plane-parallel layer, but the rest of the argument remains the same as before: the law of refraction allows us to propagate the refractive invariant from layer to layer, and to show that it remains conserved throughout the whole atmosphere, no matter how many layers there are. So the product $(nR \sin z)$ remains constant all along the ray, even in the limit of an infinite number of refractions; $(nR \sin z) = p$ is the refractive invariant. It involves only geometry and the law of refraction.

Applications: If we know the refractive invariant for a ray, we can calculate its local zenith distance z at any distance R from the centre of the Earth, provided

we have an atmospheric model that supplies the refractive index n as a function of R. For, if $nR \sin z = p$, we can solve for $\sin z = p/(nR)$, if p, n, and R are all known. A particularly convenient use of the refractive invariant is to find the radius R_{hor} where a ray is horizontal. At that point, $z = 90°$, $\sin z = 1$, so we have $nR_{hor} = p$. This implicit equation for R_{hor} is easily solved numerically by assuming that the product nR is a locally linear function of R.

Conversely, if we know where a ray is horizontal, we immediately have its refractive invariant, and can calculate its slope at every point in the atmosphere. For example, the horizon ray must be horizontal where it touches a level surface, such as the sea; this allows calculation of the (refracted) dip of the sea horizon seen from any higher elevation. If the values of n and R at the observer are n_o and R_o, and the refractive invariant for the ray is the nR product at the sea surface, we must have $n_o R_o \sin z_o = (nR)_{surface}$, from which $\sin z_o$ and hence z_o, the zenith distance of the horizon at the observer, can be found. Of course, z_o is just 90° plus the dip of the horizon; so $\sin z_o = \cos d$, where d is the dip. The refractive invariant has the important consequence that images of astronomical objects are necessarily single and erect above the astronomical horizon. Conversely, the inverted and multiple images of mirages are possible only below the horizon. This theorem was first found by Biot[49], though it has been repeatedly forgotten and rediscovered; Meyer[50] gives a simple proof by contradiction.

Suppose two different rays from an object point **P** could arrive at the observer's eye at **O** from above the horizon. As the observer would see the same object in two different directions, the rays have different zenith distances at **O**; so they have different refractive invariants (as nR is the same for both rays at the eye). However, if both rays connect **O** and **P**, the one with bigger slope at the eye must have a smaller slope somewhere else, as the mean slopes of the two rays must be equal. But nR remains equal for both rays at *every* level in the atmosphere; so the ray with the greater slope at the eye has the greater slope everywhere. As it can never have the smaller slope required to reach **P**, there cannot be two rays above the horizon.

We can avoid the contradiction if one of the rays passes through a range of heights that the other does not — either above the height of **P**, or below the eye at **O**. In either case, the ray must become horizontal at some limiting height, where nR equals the ray's refractive invariant. This can only happen for astronomical objects if that range of heights is below the observer; then the ray must be horizontal somewhere below eye level. The symmetry of rays about their perigee points means that this ray arrives at the observer from below the astronomical horizon: it belongs to the inferior mirage. (Terrestrial objects can have the second ray horizontal above the observer; this ray is ducted, and belongs to a superior mirage.) A similar argument shows that the single image above the astronomical horizon is erect. For, if the image of an object is to appear inverted, rays from the top and bottom of the object must cross, somewhere between object and observer, to reach the eye in the inverted order. Then the crossing point takes the place of **P** in the above argument; as such a point cannot exist, the rays cannot cross, and the image must be erect, if it is above the horizon. But of course the intersection can occur if one of the rays arrives from below the astronomical horizon; and in fact mirages of astronomical objects do occur there. Evidently the refractive invariant gives a special significance to the product nR. If we plot nR as a function of height for any atmospheric model, we can determine the local zenith distance of any ray

whose refractive invariant is known. Clearly, a ray cannot penetrate into regions where nR exceeds its refractive invariant, for that would require $\sin z > 1$.

Dip diagram: Plotting nR as a function of R or height in the atmosphere produces a useful diagram, whose properties were discussed in an earlier paper[51]. It allows a simple graphical determination of z along a ray. Because $(nR)_{horizon}/(nR)_{observer} = \sin z_{horizon}$ is the cosine of the dip, this is sometimes called a *dip diagram*. Any ray can be represented as a horizontal line at $nR = p$ in the dip diagram; it intersects the sloping curve that represents the atmospheric model at the point where the ray is horizontal, so that $\sin z = 1$. The ray is confined to heights where $nR > p$, so that $\sin z < 1$; that is, the ray cannot cross the model curve. Ordinarily, n decreases so slowly with increasing R that the product nR increases monotonically. However, as was mentioned above, the condition for a ray that follows the Earth's surface is $(nR) = const.$, which means the curve representing the atmospheric model is locally horizontal in the dip diagram. This occurs if the dip diagram has a local maximum or minimum.

Ducting: Because a ray must have $p < nR$, a local minimum in the dip diagram creates a region at smaller heights where rays with $p > (nR)_{min}$ can be trapped (see Fig. 11). This is a duct. For observers within the duct, celestial objects are blocked by a zone of sky centred on the astronomical horizon. The angular half-width Δz of the forbidden zone of sky is given by $\cos(\Delta z) = (nR)_{min}/(nR)_{observer}$. The symmetry of this 'blank strip'[22] about the astronomical horizon is due to the equality of the angles b in Fig. 3; see Fig. 8 of ref. 46 for photographs of an example. This symmetry can also be regarded as a consequence of the symmetry of the sine function about 90°: equal angles above and below the astronomical horizon have the same value of $\sin z$.

In the schematic dip diagram of Fig. 11, the heavy curve **ABCDE** shows the run of nR as a function of height in the atmosphere. The duct extends from the point **A** at height h_1 to **E** at h_2, where the horizontal line marked p_{min} is tangent to the local minimum in nR. All horizontal rays between these two heights are trapped in the duct, because $p = nR \sin z$ for a ray must be less than nR in the

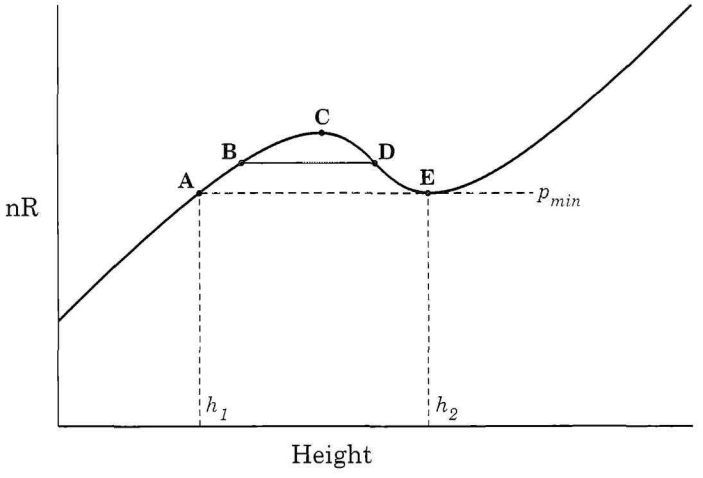

FIG. 11

A schematic dip diagram for a duct. The model atmosphere is represented by the curved line **ABCDE**; the duct extends from h_1 to h_2 in height.

atmosphere at every height. The dashed line **AE** has the minimum refractive invariant, p_{min}, that can be ducted. A ray such as **BD** within the duct oscillates between the heights of **B** and **D**; at those heights, the ray is horizontal, because ($p_{ray} = p_{model}$) implies $\sin z = 1$.

The ray **C** is horizontal at the height where the duct has maximum angular width, *i.e.*, at the height h_{max} where nR has a local maximum. If the value of nR at **C** is p_{max}, the angular halfwidth of the blank strip as seen by an observer at h_{max} is $\arccos(p_{min}/p_{max})$; this corresponds to the ray **AE**. When the ray **BD** passes through this height, its angular slope is $\arccos(p_{BD}/p_{max})$. For more detailed discussion of dip diagrams, see ref. 51.

The refraction integral

The differential of refraction: Obviously, the refractive invariant contains enough information to calculate the slope of a ray at every height in an atmosphere, if the run of nR with height is known: the refraction is just the total change in slope of the ray. Knowing the slope at every point, we can write down the differential equation for the refraction. The only problem is that the dip diagram (or its equivalent, the model atmosphere) gives the slope of the ray with respect to the *local* zenith, whose direction changes along the ray; see Fig. 12, which shows the differential triangle at a distance R from the centre of the Earth. The zenith distance of the ray at R is z, and at $R + dR$, it is $z + dz$. The differential of refraction is

$$dr = dz + d\theta,$$

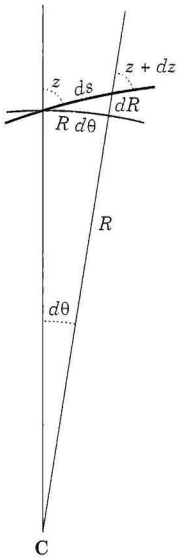

FIG. 12

The differential triangle for refraction. The heavy curved line represents the refracted ray, with zenith distance z at a distance R from the Earth's centre **C**, and zenith distance $z + dz$ at $R + dR$. The element of path length ds subtends an angle $d\theta$ as seen from **C**; $d\theta$ is the change in direction of the local zenith in the interval ds.

where $d\theta$ is the differential change in the direction of the local zeniths, *i.e.*, the angle at the centre of the Earth subtended by the element of path length ds. (If the ray follows the curve of the Earth, so that z remains constant, dr is just $d\theta$.)

The tangent of the ray's altitude is the cotangent of its zenith distance, so the little differential triangle gives

$$\frac{dR}{R\,d\theta} = \cot z.$$

We will shortly need dR/R, so we rearrange this equation:

$$\frac{dR}{R} = \frac{d\theta}{\tan z}.$$

Now use the fact that the refractive invariant ($nR \sin z$) is constant, so that its logarithmic derivative is zero:

$$\frac{dn}{n} + \frac{dR}{R} + \frac{d(\sin z)}{\sin z} = 0,$$

or

$$\frac{dn}{n} + \frac{dR}{R} + \frac{\cos z}{\sin z}\,dz = \frac{dn}{n} + \frac{dR}{R} + \frac{dz}{\tan z} = 0.$$

Then combine the value of $dR/R = d\theta/\tan z$ with the fact that $d\theta = dr - dz$, so that $dR/R = (dr - dz)/\tan z$, and substitute this into the last equation:

$$\frac{dn}{n} + \frac{dr - dz}{\tan z} + \frac{dz}{\tan z} = 0.$$

Finally, cancel the two dz terms, and solve for dr:

$$dr = -\tan z\,\frac{dn}{n}.$$

Physically, the minus sign occurs because n decreases as R increases. The factor $\tan z$ shows how sensitive refraction is to the *local* zenith distance along the ray: where the ray is horizontal, $\tan z$ becomes infinite. Although this infinity requires transforming dr to handle horizontal rays, the present form with $\tan z$ is the most informative expression for the refraction differential.

Integrating the refraction: If we leave the differential of refraction in the form just derived, the whole refraction is just

$$r = \int_{n_o}^{1} dr = -\int_{n_o}^{1} \tan z\,\frac{dn}{n} = \int_{1}^{n_o} \tan z\,\frac{dn}{n},$$

where n_o is the value of the refractive index at the observer, and 1 is its value above the atmosphere.

The independent variable here is n, the refractive index. Note that n varies only from 1·0000 in space to not quite 1·0003 at sea level. Because the

refractivity $(n - 1)$ is very nearly proportional to the density ρ of the air, $dn \propto d\rho$. So n is a linear function of the density. It is instructive to plot the refraction integrand as a function of n, bearing in mind the linear relation between n and density. When we do so, we find that the refraction integrand (and hence the refraction itself) behaves very differently in different parts of the sky.

Small zenith distances: Fig. 13 shows the refraction integrand for the Standard Atmosphere[31], for a few zenith distances up to 60 degrees. In this range, the refraction is less than 2 minutes of arc, so the ray is nearly straight. As 2 minutes is 1/1800 of 60°, a straight-line approximation to the ray should nearly meet the required degree of accuracy out to this zenith distance. If we neglect ray curvature entirely, the local zenith distance of the ray at any level is independent of the atmospheric structure. In this approximation, $\tan z$ can be computed from the geometric invariant $p_g = R \sin z$. Then at height h, $(R_E + h) \sin z_h = R_E \sin z_o$ (assuming the observer at the Earth's surface); so

$$\sin z_h = \frac{R_E}{R_E + h} \sin z_o.$$

But for small z, $\sin z \approx \tan z \approx z$; so we can also write

$$\tan z_h \approx \frac{R_E}{R_E + h} \tan z_o.$$

Furthermore, the atmosphere is so shallow that $h \ll R_E$ everywhere: there is practically no refraction above 100 km height, so $h/R_E \leq 1/64$. This means that $\tan z$, which provides most the variation of the refraction integrand, hardly varies by 1% across Fig. 13. Because the integrand is so nearly constant, we can replace both n and $\tan z$ with their average values, leaving only dn inside

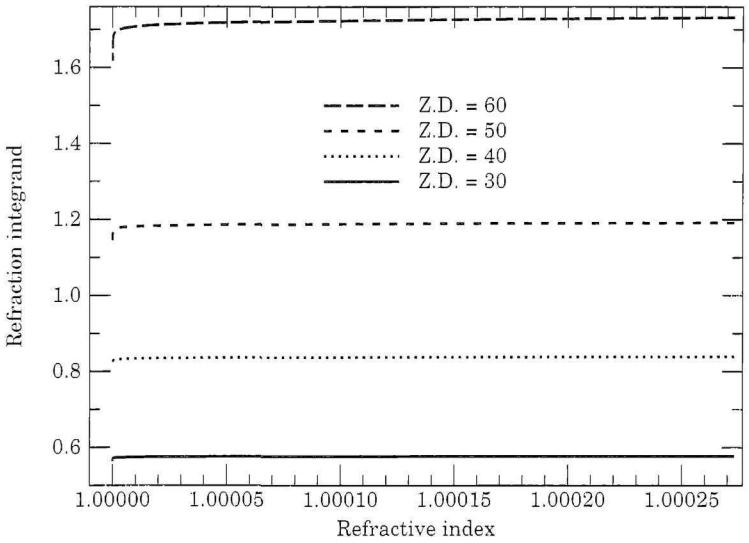

FIG. 13

The refraction integrand for the Standard Atmosphere, for zenith distances at the observer from 30 to 60 degrees. The Earth's surface (at the right edge) corresponds to $n = 1.00027$, and the top of the atmosphere (left edge) to $n = 1.00000$.

the integral. The refraction becomes simply

$$r = \frac{\langle \tan z \rangle}{\langle n \rangle} \int_1^{n_o} dn = (n_o - 1) \frac{\langle \tan z \rangle}{\langle n \rangle}.$$

The mean $\langle n \rangle$, which is $(n_o + 1)/2 \approx 1.00014$, obviously corresponds to the atmospheric level where the density is half the surface density, about 6.7 km above sea level. If the integrand were a straight line, the mean $\langle \tan z \rangle$ would correspond to this same level. But Fig. 13 shows that the integrand is slightly concave downward, because of the nonlinear dependence of h, and hence z, on density; so the mean $\langle \tan z \rangle$ is smaller, and corresponds to a higher level. This effective height turns out to be the reduced height $H \approx 8$ km. At that height, the mean value $\langle \tan z \rangle = \tan z_H$ is

$$\tan z_H = \frac{R_E}{R_E + H} \tan z_o.$$

So the refraction at small zenith distance is well represented by

$$r = \left(\frac{R_E}{R_E + H} \right) \left(\frac{2}{n_o + 1} \right) (n_o - 1) \tan z_o,$$

or just

$$r = \left(\frac{R_E}{R_E + H} \right) (n_o - 1) \tan z_o,$$

if we neglect an error of 1 part in 7000 and set $\langle n \rangle = (n_o + 1)/2$ to unity. This approximation represents the refraction of the Standard Atmosphere within 0.1 arcsec to 48.8° zenith distance, and within 1 second to nearly 68°. The coefficient $(n_o - 1)$ $[R_E/(R_E + H)]$ is often called the *refraction coefficient* or the *refraction constant*. Notice that it differs by the factor $R_E/(R_E + H)$ from the coefficient $(n_o - 1)$ in the tangent approximation for the plane-parallel atmosphere. This curvature correction factor is less than unity by about $H/R_E \approx 1/800 = 0.00125$, which is larger than the acceptable relative error; so it must be taken into account.

Physically, the curvature of the Earth decreases refraction (compared to the flat case) because the change in direction of the local vertical along the ray reduces angles of incidence, and hence local values of $\tan z$, in the upper atmosphere. The curvature correction factor in $\langle \tan z \rangle$ represents the average effect of the tilted verticals along the ray relative to the observer's zenith. The greater the reduced height, the smaller is the refraction for a fixed n_o. For example, if we raise the temperature of the atmosphere, we must increase the surface pressure to keep ρ_o and hence n_o fixed. This requires a larger mass of gas above the observer. At every $R > R_E$, there is now a greater density than before, so n is (slightly) higher. This makes $\sin z$ and hence $\tan z$ smaller; so the left side of the integrand (corresponding to the upper atmosphere) moves downward in the plot, decreasing the area under the curve (*i.e.*, the refraction). The warmer atmosphere has a bigger reduced height, H, and its curvature correction factor $R_E/(R_E + H)$ is correspondingly less. On the other hand, if we keep both the temperature and pressure at the observer fixed, so that the refractivity at the observer stays fixed, changes above the observer are constrained by hydrostatic equilibrium. We can move gas up and down by

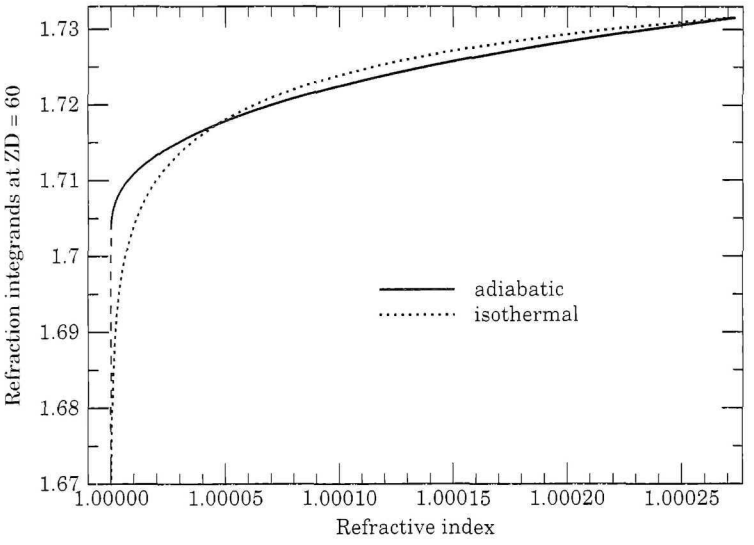

FIG. 14

The refraction integrand for adiabatic and isothermal atmospheres, at 60° zenith distance. The top curve in Fig. 13 would fall between these; note the greatly expanded vertical scale here. The dashed vertical line at $n = 1{\cdot}0000$ marks the lower limit of the refraction integration.

altering the temperature profile, but the total mass of gas in the column, and the reduced height, H, remain the same. Then changes in the integrand in one region are nearly balanced by opposite changes in another region, and the refraction is almost unaffected.

Fig. 14 compares the integrands for isothermal and adiabatic atmospheres with the same conditions at the observer. Both atmospheres have the same temperature and pressure at the bottom; the adiabatic model has a lapse rate of 10 K/km. The bottom part of the adiabatic model, at the right and centre of the figure, is cooler and denser than the isothermal model at the same height; so nR is larger, and consequently $\sin z$ and $\tan z$ are smaller in this region for the adiabatic model. As $\tan z$ is the dominant variable in the integrand, the adiabatic integrand lies *below* the isothermal one in this part of the diagram. But, because the upper parts of the adiabatic model are much colder than the isothermal one, the layers of lowest density are much closer to the surface of the Earth: the adiabatic model terminates below 30 km. In these low-density layers, at the left side of Fig. 14, n is very nearly unity, and R dominates the nR product. So in this region, $\sin z$ and $\tan z$ are larger for the adiabatic model than for the isothermal one. The two curves cross where the density is about $\frac{1}{6}$ of the surface density ($n \approx 1{\cdot}00004$). The areas between the two curves are almost exactly equal on either side of this crossing. That is, the areas under the two curves — *i.e.*, the refractions in the two models — are almost exactly equal. If the average ordinates of the two parts were the same, the cancellation produced by the crossover would be exact. Near the zenith, the integrands are in fact nearly flat, and this cancellation is nearly perfect: all models give almost exactly the same refraction, regardless of the temperature profile. Even at 60° zenith distance, the average ordinates on the two sides of the crossing differ by only about 1% (note the expanded scale of Fig. 14). So the imbalance is very slight here: the

refraction is about 98 seconds, but the two models differ by less than 0·002 arcsec — about 2 parts in 10^5.

Moderate zenith distances: However, at larger zenith distances, the disparity between the upper and lower parts of the atmosphere rapidly increases, because of the growing change of $\tan z$ along the ray within the atmosphere. At $z_o = 85°$, the curves for the adiabatic and isothermal models cross at $\tan z \approx 9$, and the average ordinates for the left and right portions are about 8 and 10 — a difference about 20 times larger than at 60°. The mean refraction is 9·6 arc minutes; the difference between the models has risen to nearly 4 arcsec, about 7 parts in 10^3, which is quite significant. Fig. 13 shows that at large zenith distances, the decrease of $\tan z$ in the upper atmosphere gives appreciable slope to the integrand, especially at the upper left corner of the curve, which is concave downward. The unequal weighting of upper and lower regions when this slope is appreciable makes the refraction sensitive to atmospheric structure: a density gradient in the lower atmosphere, where $\tan z$ is larger, contributes more to the refraction than the same gradient higher in the atmosphere. However, this sensitivity is constrained by the refractive invariant. A change in atmospheric density at a given R makes a proportional change in the refractivity $(n - 1)$; but that is, at most, only a part in 3000 of n itself, which appears in the invariant $nR \sin z$. Thus, changes in refractivity at a given R produce changes in $\sin z$ that are thousands of times smaller. As $\tan z$ is less than 10 times larger than $\sin z$ at 84° zenith distance, the changes in $\tan z$ also remain small until $\tan z / \sin z = \sec z$ appreciably exceeds 10.

Series expansions: In this region, where the sensitivity to atmospheric structure is small, it is tempting to extend the $\tan z$ approximation, which works well near the zenith, to a power series in $\tan z$. Delambre[5] gives numerous examples of such developments, based on trigonometric identities. The more usual approach[1,2,6,7,8] is that introduced by Lambert[52]: first, replace the $\tan z$ in the integrand by $\sin z / \sqrt{1 - \sin^2 z}$. Then, replace $\sin z$ by its equivalent p/nR by using the refractive invariant $p = (n_o R_o \sin z_o)$, so that the refraction integral $r = \int_1^{n_o} \tan z \, \frac{dn}{n}$ becomes

$$r = \int_1^{n_o} \frac{p}{\sqrt{(nR)^2 - p^2}} \, \frac{dn}{n} ,$$

which of course looks much more intimidating if we write out p in terms of n_o, R_o, and $\sin z_o$, as is customary:

$$r = \int_1^{n_o} \frac{n_o R_o \sin z_o \, dn}{n \sqrt{(nR)^2 - (n_o R_o \sin z_o)^2}} .$$

The final step is to expand the square root in the denominator by using the binomial theorem, and introduce some closed-form relation between n and R to express the series terms as functions of a single variable. The expansion produces very complex expressions even for simple atmospheric models. The series, involving powers of a 'small quantity' such as $[(H/R) \tan^2 z]$, can be integrated termwise, but even then the individual terms involve integrals that must themselves often be approximated by series expansions. This process allows the numerical calculation of refraction tables out to zenith distances of about 82°, but at the expense of mathematical abstractions that obliterate all traces of the physics. Some people have carried it to ridiculous extremes:

Bauernfeind[53] extended the series to the 28th power of sec z. Worse yet, the series expansions are only semi-convergent, as was first pointed out by Ivory[54]. Unfortunately, the concept of semi-convergence had only been introduced a dozen years earlier by Legendre[55], and Ivory's warning escaped the notice of most astronomers.

Strictly speaking, these Lambert series diverge at *every* zenith distance — not just at the horizon, where $\tan z$ blows up. The real difficulty lies in the coefficients of the terms, which increase like factorials with the order of the term. So, no matter how small $\tan z$ may be, the higher terms eventually increase without limit: the series diverges. Because only odd powers of the small argument appear in the series, the unwary often suppose that convergence is rapid. This odd-power property is sufficiently un-obvious that Bradley laboriously established empirically[56] that the coefficient of $\tan^2 z$ is negligible. But it is simply a result of symmetry: if the refraction is regarded as a continuous function through the zenith, with zenith distances counted positive on one side and negative on the other, it is obvious that the refraction must also change sign on passing through the zenith, where it is zero. Therefore the refraction is an odd function, and can involve only odd powers of $\tan z$, which itself is odd. What makes the series useful numerically is its alternating signs (due to the binomial expansion of a negative power — the square root is in the denominator), which allow truncation after a few terms, so long as $\tan z$ is not too large. Then the partial sums are accurate enough for practical work, which only requires three or four significant figures. But, as successive terms decrease more and more slowly, many are required for high accuracy, even at moderate zenith distances. (For example, two terms approximate the refraction of a realistic atmosphere considerably less accurately[12] than does the Cassini model, at *all* zenith distances.) And expansions fail entirely around a zenith distance of 82°, where the smallest term in the series becomes unacceptably large.

Oriani's theorem

Evidently, this is not a very instructive approach to the problem. However, it does produce one remarkable and well-known result, which can be neatly demonstrated[7] by setting $R/R_o = 1 + s$, so that s is just height measured in Earth radii. Note that s is always small: it is 0·01 at 64 km height, and we can neglect the refraction above $s = 0{\cdot}02$. So, set $R = (1 + s)\, R_o$ in the refraction integral, and keep only the first-order terms in s, so that $(nR)^2 \approx (1 + 2s)\,(nR_o)^2$. Then, cancelling R_o factors in numerator and denominator, we have

$$r = \int_1^{n_o} \frac{n_o \sin z_o \, dn}{n \sqrt{n^2 + 2n^2 s - (n_o \sin z_o)^2}}.$$

The argument of the square root can be rewritten as

$$(n^2 - n_o^2 \sin^2 z_o) + 2n^2 s = (n^2 - n_o^2 \sin^2 z_o) \left(1 + \frac{2n^2 s}{n^2 - n_o^2 \sin^2 z_o} \right);$$

then the refraction integral becomes

$$r = \int_1^{n_o} \frac{n_o \sin z_o \, dn}{n \sqrt{n^2 - n_o^2 \sin^2 z_o}} \left(1 + \frac{2n^2 s}{n^2 - n_o^2 \sin^2 z_o} \right)^{-1/2}.$$

Now, expand the expression in large parentheses on the right, using the binomial theorem; keep only the first-order term in s, and integrate the resulting terms separately:

$$r = \int_1^{n_o} \frac{n_o \sin z_o \, dn}{n\sqrt{n^2 - n_o^2 \sin^2 z_o}} - \int_1^{n_o} \frac{s n n_o \sin z_o \, dn}{(n^2 - n_o^2 \sin^2 z_o)^{3/2}} .$$

The first term is an elementary integral, whose value,

$$\arcsin(n_o \sin z_o) - z_o ,$$

is exactly the refraction (Eqn. 1) for the plane-parallel atmosphere. Thus the second integral can be regarded as the first-order correction for atmospheric curvature. This correction term can be evaluated by setting both n and n_o to unity in its integrand. (The error made is of higher order, as this term is already of order s.) Then the correction term becomes

$$- \frac{\sin z_o}{\cos^3 z_o} \int_1^{n_o} s \, dn .$$

Next, we use the Gladstone-Dale rule that the refractivity $(n - 1)$ is proportional to the density, ρ. But if $(n - 1) = c\rho$, then $dn = c \, d\rho$. This converts the correction term to

$$- c \frac{\sin z_o}{\cos^3 z_o} \int_o^{\rho_o} s \, d\rho .$$

Finally, integration by parts gives

$$- c \frac{\sin z_o}{\cos^3 z_o} \int_o^{s_{max}} \rho \, ds ,$$

where s_{max} is the largest normalized height that contributes appreciably to the refraction — essentially, the top of the atmosphere. But this integral of density through the whole atmosphere is just the mass of a unit column; so this last integral is proportional to the surface pressure at the observer, or to the ratio H/R_o. Furthermore, the factor $\sin z_o / \cos^3 z_o = \tan z_o \sec^2 z_o$; and if we replace $\sec^2$ with $(1 + \tan^2)$, this factor is just $(\tan z_o + \tan^3 z_o)$. The plane-parallel term can also be expressed as a sum of tangent and tangent-cubed terms, if we expand its arcsine in a Taylor series and neglect higher powers of the refractivity. So the sum of the two terms is of the familiar form

$$r = A \tan z_o - B \tan^3 z_o ,$$

where the coefficients A and B involve only conditions at the observer and are independent of the density distribution. This result was first proved by Oriani[37,38], who stated that "This expression depends on no hypothesis about either the law of heat in the atmosphere or about the density of the air at various distances from the surface of the Earth". Laplace[1] provided a more rigorous and complete proof of Oriani's theorem.

One important practical consequence of Oriani's theorem is that all possible atmospheres in hydrostatic equilibrium produce the same astronomical refraction, up to the zenith distance where the $\tan^5 z_o$ term becomes appreciable. (For example, it explains the near-perfect compensation shown in Fig. 14.) Therefore it is convenient to derive the exact expressions for the coefficients A and B from (for example) Cassini's model — as Ball[7] does. This limiting zenith distance is typically in the range 70° to 74°, depending on the size of the $\tan^5 z_o$ term and the required accuracy. Oriani's theorem explained why all previous refraction calculations had given very similar results out to about 74° — a fact that had puzzled many earlier workers.

Of course, the $\tan^5 z_o$ term does depend on atmospheric structure, so different models diverge rapidly beyond this limit. Because $\tan z$ is inversely proportional to altitude a at large zenith distances, the initial differences are approximately inversely proportional to a^5. So the divergence increases by a factor of 3 between 70° and 74°, and is still faster beyond that. In fact, one sees from the derivation of the series expansion that successive terms involve integrals of consecutive powers of s with respect to n; but because dn is proportional to $d\rho$, these integrals are all of the form

$$\int s^j \, d\rho, \quad j = 1,2,3,\ldots,$$

i.e., they are successive height-moments of the density distribution.

In an exponential atmosphere, $d\rho \propto \exp(-s)\, ds$, and these moments become essentially

$$\int_0^\infty x^j e^{-x} \, dx = j!$$

(Indeed, it was while studying the refraction integral for an exponential atmosphere that Kramp[57] developed the theory of factorials and introduced the function we know today as the Γ function.) Hydrostatic equilibrium forces the real atmosphere to be nearly exponential, so the coefficients of Lambert's series-expansion terms increase nearly factorially. This is why the series diverges for all zenith distances. As the $\tan^5 z_o$ term depends on the second moment of the density distribution, it is similar for all realistic model atmospheres. This term depends mainly on the average lapse rate in the troposphere, so the differences in refraction between 70° and 80° for different models depend almost entirely on this mean lapse rate. The similarity of the second moments of all realistic models means that their refractions differ only a little out to 80° zenith distance, somewhat beyond the range where Oriani's theorem guarantees complete independence from atmospheric structure.

Refraction near the horizon

Beyond Oriani: The independence of refraction from atmospheric structure, in the region where Oriani's theorem applies, is due to the negligible variation in $\tan z$ along the ray. At moderate altitudes, $\tan z$ is nearly constant (*cf.* Fig. 13), so the refraction depends on a nearly equally-weighted average density or temperature gradient through the whole atmosphere. But near the horizon, $\tan z$ is large and nearly equal to $\sec z = 1/\cos z = 1/\sin a \approx 1/a$. So the $\tan z$ weighting along the ray changes significantly if the local altitude a changes along the ray.

The transition from small to large variation of $\tan z$ along the ray can be found from the refractive invariant.

How near the horizon is 'near'? The refractive invariant is $(nR) \sin z$, so fractional changes in $\sin z$ correspond to similar fractional changes in (nR). Now, $\sin z = \cos a \approx 1 - a^2/2$, for small a; so the upper and lower parts of the atmosphere are equally weighted if their fractional difference in (nR) is comparable to $a^2/2$. Because the refractivity is less than 3×10^{-4}, changes in (nR) are mostly due to changes in R. Half the mass of the atmosphere is above the level where the pressure is half the surface pressure, near 6·7 km height. This corresponds to a fractional change in R of about 1 part in 10^3; but the fractional change in n there is only $1{\cdot}5 \times 10^{-4}$, nearly an order of magnitude smaller. So the lower atmosphere becomes disproportionately important when $a^2/2 \approx 10^{-3}$, corresponding to $a \approx 0{\cdot}05$ radians or 2·6°. This agrees with the altitude where nocturnal inversions are found to become important[12]: numerical integrations show large differences in refraction among different models only below 2 or 3 degrees altitude.

Fig. 15 shows how $\tan z$ blows up in the lowest layers near the horizon. The top two curves, for altitudes of 1° and 2°, have altitudes below the critical value just calculated; and it is just these that show a nonlinear increase in $\tan z$ in the lowest layers (right-hand side). In this zone of sky, $\tan z$ is much larger near the observer than in the upper half of the atmosphere. In fact, the refraction integrands all have nearly the same values at the left side of the figure; they differ by only a factor of 2 at the tropopause. The upper atmosphere contributes a nearly fixed amount to the refraction at all altitudes near the horizon. This behaviour follows from the refractive invariant: $\sin z > 0{\cdot}995$ at the observer for all these curves, so the local zenith distance depends more on R than on z_o above the height where $R_o/R > 0{\cdot}995$ (about 32 km). This is another example of Wegener's Principle: the horizon ray meets the tropopause (or any surface

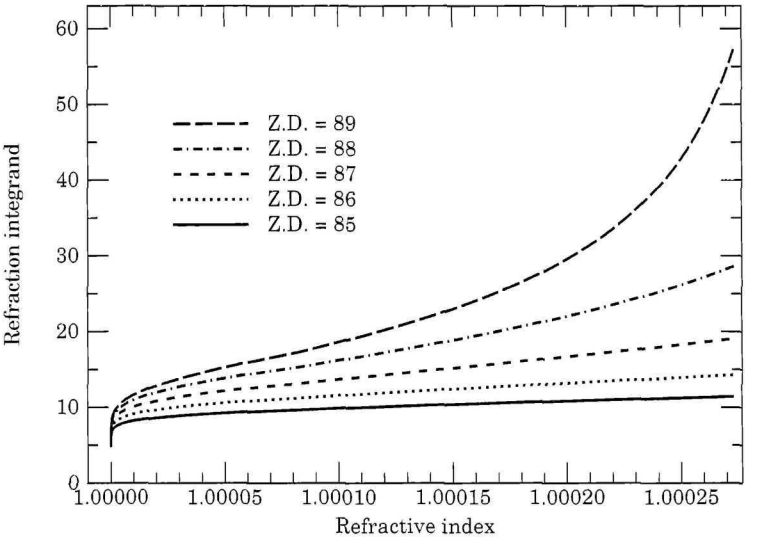

FIG. 15

The refraction integrand for the Standard Atmosphere, for zenith distances at the observer from 85 to 89 degrees. (*cf.* Fig. 13.)

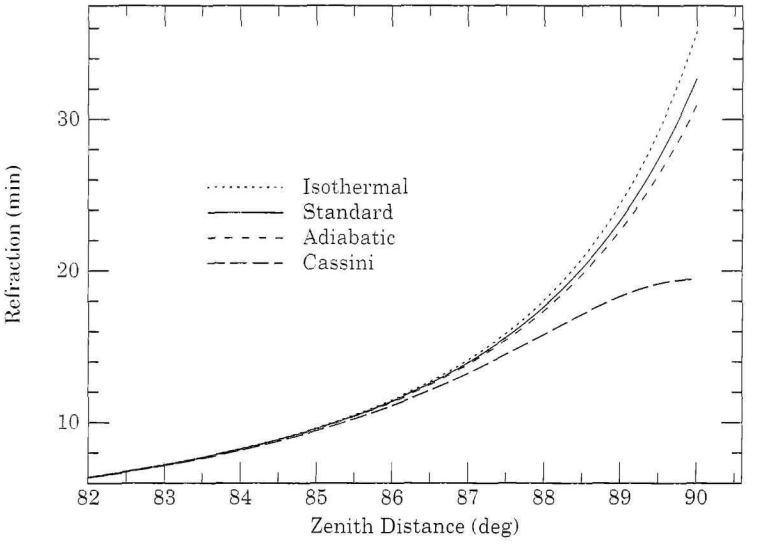

Fig. 16

The refraction for the Standard Atmosphere and three other polytropic models compared. All models have the same pressure and temperature at the observer; only their lapse rates differ.

overhead) at a local maximum value, so nearby rays also meet that surface at almost the same angle, and have similar refractions at greater heights. So the heavily-weighted lowest layers dominate the refraction at low altitudes. And the lapse rate in the boundary layer becomes more and more important in the total refraction (not just its derivative, as Biot's theorem tells us) as we approach the horizon[12]. Fig. 16 compares low-altitude refractions for the Standard Atmosphere and a few other canonical models with constant lapse rates. Because horizontal-ray curvature depends on the lapse rate, Biot's theorem gives the slope of each curve at the horizon; the zero slope of Cassini's model there can also be regarded as an example of Wegener's Principle. Oriani's theorem forces the curves to converge at the left side of Fig. 16. These curves, which meet only at the zenith, form a one-parameter family that illustrates many of the principles discussed above.

Refraction below the horizon: Between the astronomical horizon and the sea horizon is a zone in which rays are horizontal somewhere below eye level. This zone includes the whole disc of the setting Sun for observers above 220 m in the Standard Atmosphere, and even for lower observers with thermal inversions. Below the astronomical horizon, the refraction is dominated by the perigee layer, where the ray is horizontal. In this region, we invoke Wegener's Principle to separate the refraction contributions of the layers above and below eye level. To avoid problems with the infinite value of $\tan z$ at the perigee point, it is easier to work with ray bending in this lower region than with the refraction integral in the whole atmosphere; Wegener's principle shows that the upper part contributes only a small variation with altitude. If the ray were straight (*cf.* Fig. 4a), the thickness t of the layer between the heights of the observer and the perigee would be just **OC**$-$**PC** or $R\,(1-\cos a) \approx Ra^2/2$. If the lapse rate in this layer is constant, the ray is nearly a circular arc; then it can be shown[22, 51] that

the same result applies, if we use an effective curvature $1/R_{eff}$ corresponding to the *difference* in curvatures of the ray and the Earth:

$$\frac{1}{R_{eff}} = \frac{1}{R} - \frac{1}{\mathbf{OS}}$$

(see Fig. 4b). In terms of the ratio of curvatures $k = \mathbf{OC}/\mathbf{OS}$ used in the discussion of magnification, the effective radius of curvature is

$$R_{eff} = R/(1 - k).$$

Then

$$t = R_{eff}\,(1 - \cos a) \approx \frac{Ra^2}{2(1 - k)}\ .$$

The ray-bending θ produced by the layer is just k times the central angle **OCM**, which can be found from the distance from **O** to the perigee point **P**:

$$\mathbf{OP} = \mathbf{OC}\,\sin(\mathbf{OCM}/2) \approx \mathbf{OC}\cdot(\mathbf{OCM}/2).$$

But **OP** can also be found from the formula for the distance to the horizon[51]

$$\mathbf{OP} = \sqrt{2tR/(1 - k)}\ ;$$

so, equating the two expressions for **OP**, solving for the central angle **OCM**, and multiplying by k to get θ, we have the refraction contributed by the layer of thickness t:

$$\theta = 2k\sqrt{\frac{2t}{(1 - k)R}}\ .$$

This is very large when k is close to unity. For example, a layer with a lapse rate of $-0{\cdot}1$ K/m has $k = 0{\cdot}9$; a 1-m thickness of this layer contributes 10 minutes of arc to the refraction of a ray tangent to its lower surface. And because of the square root, even one centimetre of this layer contributes a minute of arc of refraction. As heat conduction forces the lapse rate to be continuous, there is always a place at the top of a duct where $k \to 1$ and the refraction becomes infinite. This is actually seen by an observer above the duct: the Sun flattens out into a line above the duct, where it gradually fades from view as the extinction also grows without limit. Evidently, the atmospheric structure must be resolved to better than a millimetre if precise results are to be calculated near the top of a duct below eye level.

Mirages and ducting: A constant lapse rate produces an erect but somewhat compressed image of the sky at the astronomical horizon. The air near the observer acts much like a prism, deviating the image of the setting Sun (for example) in proportion to the ray curvature. However, when the lapse rate changes with height — that is, when the temperature profile is curved — the refraction changes rapidly with apparent altitude. Above the astronomical horizon, where $\tan z$ varies relatively slowly along the ray, the refraction integral averages out the contributions from different layers so well that the refraction cannot change faster than the zenith distance. But below the horizon, where

rays can be horizontal, the perigee layer dominates the refraction, which can change very rapidly with altitude. In particular, if the lapse rate below eye level decreases rapidly with height, as at the base of a low-lying thermal inversion, or immediately above a warm surface, the refraction decreases with increasing zenith distance. When refraction decreases faster than zenith distance increases, we see objects that are actually higher in the sky as we look lower: images become inverted, and we have a mirage. Such mirages[46] do not require ducting; they merely require rapid changes in lapse rate with height, so that the atmosphere acts like a positive, cylindrical lens, producing a real, inverted image of a zone of sky. The zigzag limb of the low Sun often displays a stack of thermal inversions below eye level; O'Connell's book on the green flash[58] shows many fine examples. If there is a duct below eye level, the image of the sky becomes discontinuous where the line of sight is tangent to the top of the duct. The refraction increases without limit just above this boundary, but is finite just below it. The resulting image discontinuities and inversions[22] produce the 'Chinese-lantern' effect on the setting Sun. O'Connell shows a few examples of such discontinuities, which he calls "surfaces of separation".

Calculating refraction near the horizon: Because Lambert's series expansion diverges more rapidly with increasing zenith distance, and becomes useless numerically well above the horizon, a different approach is required at low altitudes. The most useful one is that invented by Biot[36], and independently rediscovered by Auer & Standish[59]. It avoids the divergence of the integrand at the horizon (because of the $\tan z$ factor) by changing the variable of integration from n to the local zenith distance z. This method works well so long as the ray curvature differs appreciably from the Earth's. However, if the ray curvature approaches the Earth's, z remains nearly constant along the ray; the interval of z corresponding to some interval of height becomes vanishingly small, and the integrand must become enormous to keep the area under the curve finite. Mathematically, this is because the denominator of the transformed integrand is proportional to the difference of the curvatures. Computationally, the problem is numerical instability: the denominator is the small difference of two nearly-equal quantities, so the calculation is swamped by round-off error. It therefore becomes unusable when ducting occurs — as was already foreseen by Biot.

Infinite refraction: This 'corner case' (of horizontal rays with curvatures equal to the Earth's) is obviously intractable: it corresponds to rays that circle the Earth endlessly, giving infinite refraction. This situation was beautifully analyzed by Kummer[60], who discovered that an infinite number of infinitely thin images of the whole sky are produced — if we neglect extinction.

Refraction and extinction

Laplace's extinction theorem: The relation between extinction and refraction was established by Laplace[1] two centuries ago. His result is equation [8599] in Nathaniel Bowditch's admirable translation[61] of the *Mécanique Céleste*; in Bowditch's words, it is just: "... the logarithm of the intensity of light, of any heavenly body, is proportional to its refraction, divided by the cosine of its apparent altitude." Or, remembering that $\cos a = \sin z$, we can say that the refraction r is proportional to the product $M \sin z$, where M is the airmass.

To see why this is so, consider that the differential of refraction is the element of path length ds times the component of the refractivity gradient normal to the ray. As the gradient is vertical, the projection factor is just $\sin z$; and the

refractivity gradient is proportional to the density gradient, $d\rho/dh$. So the refraction is

$$r \propto \int \left(\frac{d\rho}{dh} \right) \sin z \, ds .$$

But the airmass is just proportional to the integral of the density itself:

$$M \propto \int \rho \, ds .$$

We would have everything needed to obtain Laplace's theorem if $d\rho/dh$ in the refraction integral were just proportional to ρ in the airmass integral. In general, they are not proportional. But Laplace imposes a fairly mild condition: he assumes the atmosphere is isothermal, so the density decreases exponentially with height. And of course the derivative of an exponential is proportional to the exponential itself; so we have what we need. There is also the problem that $\sin z$ is not constant along the ray. However, it is nearly constant along the ray if the zenith distance is not too large, because z is nearly constant. And, near the horizon, $\sin z$ is nearly unity, even though z varies by a few degrees as the ray traverses the atmosphere. As a result, the variation of $\sin z$ is not a serious problem; as Laplace demonstrates, the theorem is a moderately good approximation.

Although the real atmosphere is not isothermal, most of the refraction (and airmass) is contributed by the bottom few kilometres[12], in which the temperature varies by only a few per cent. So Laplace's extinction theorem is really fairly accurate. As a trivial example, consider the simple approximations for the plane-parallel atmosphere: $r \propto \tan z$ and $M \approx \sec z$. Their ratio is just $\sin z$, as expected. Laplace's result has also been verified for more realistic model atmospheres[62]. A handy corollary of Laplace's theorem is that near the horizon, where $\sin z \approx 1$, the refraction is very nearly proportional to the airmass.

Series expansions: Lambert's series-expansion method can also be applied to airmass calculations[63]. Apart from the $\sin z$ factor, the terms are similar; however, because airmass has the density where refraction has the density *gradient*, the moments that appear in the coefficients of the terms are all one order higher in the airmass series. In particular, the coefficients begin with the first moment, not the zeroth moment, so there is no term in the airmass series that is independent of atmospheric structure, and we have nothing comparable to Oriani's theorem.

Who cares about extinction? Although refraction and extinction are intimately related, they have traditionally belonged to different fields. Refraction has been the concern of astrometrists; extinction, that of photometrists. However, refraction has to be taken into account in ground-based photometry — not only for telescope pointing, but because the blue image of a star lies above the red one, so that aperture errors and centring may differ, depending on the passband being measured. This is particularly a problem for the large-airmass observations needed to determine the extinction. Furthermore, it is the refracted and not the true zenith distance that is the independent variable in the airmass tables, so the wrong airmasses will be used if zenith distances calculated from times and positions are not corrected for refraction[18]. On the other hand, extinction is a problem for astrometry, because the refraction is wavelength-

dependent; and the effective wavelength depends on atmospheric reddening, varying with zenith distance. Finally, the tropospheric corrections required in *GPS* calculations are more closely allied to airmass than to refraction. Consequently, all observers should understand both airmass and refraction.

Discussion

The physics of atmospheric refraction is simple. But it is obscured by the lengthy semi-convergent series expansions that were introduced by Lambert[52]. Because astronomers have concentrated on ridding observational data of the nuisance of atmospheric refraction, rather than regarding it as a subject to be understood in its own right, much effort has been mis-directed. For example, many people, including Simon Newcomb[6], have worried that the poorly-understood structure of the upper atmosphere was a major source of uncertainty in refraction tables; in fact[12] it is quite unimportant. The 19th-Century emphasis on analytical rather than numerical methods led to an emphasis on atmospheric models that made the series expansions tractable — even after Biot[36] had shown that accurate numerical integrations require only modest computational effort. Most of the competing models were polytropes[64,3] of various degrees. These all have a constant lapse rate, so they terminate with an absolute temperature of 0 K at a finite height — a feature that worried many workers.

Cassini's homogeneous model can be regarded as a polytrope of index zero; the Simpson[65]–Bradley[66]–Meyer[67] model has polytropic index one; and the isothermal model has an infinite polytropic index. The efforts of Laplace[1] and Ivory[48,54] went into constructing a sort of interpolation or hybrid formula between these extremes. Because of the poorly-determined position of absolute zero on existing temperature scales, Ivory decided the best polytropic index (to put his result in modern terms) was 3 or 4. Later, Bauernfeind[53] worked out the polytrope of index 5; and Radau[2] gives results for 4, 5, and 6. These workers all struggled with the conflict between the small height of the polytropic model, which ends at $(m + 1)\,H$ if m is the polytropic index, and the much greater height of the atmosphere inferred from meteors, aurorae, and twilight phenomena. Today, we understand that this is due to the isothermal stratosphere, which greatly extends the height of the upper atmosphere without[12] affecting its refraction. Indeed, Ivory's model gradually tails off into a nearly-isothermal upper extension, which he recognized was poorly constrained by refraction data. Another problem these workers had was that the mean tropospheric lapse rate produces less refraction within 2° of the horizon than is observed. This, we now understand[12], is due to the nocturnal thermal inversion, whose importance was first urged by Oppolzer[68], but without success.

If refraction corrections are needed for precise astrometric observations, Cassini's model is more than good enough[12] — as, indeed, is suggested by Oriani's theorem[37,38]. If the boundary layer were featureless, Biot's magnification theorem[35,36] would suffice to relate the local lapse rate to the Sun's flattening at the horizon. If the details of refraction near the horizon are required, as in explaining sunset mirages[22,24,46] and their associated green flashes[69,58,70,71], one must take account of the complex thermal structure in the boundary layer. And because of the dispersion that causes green flashes, the atmospheric reddening, and hence the linkage between refraction, airmass, and effective wavelength of observation, must always be kept in mind.

References

(1) P. S. Laplace, *Traité de Mécanique Céleste, tome 4, liv. X, Ch. III* (J. B. M. Duprat, Paris), 1805.
(2) R. Radau, *Annales de l'Observatoire de Paris*, **16**, B.1, 1882.
(3) I. G. Kolchinskii, *Refraktsiya Sveta v Zemnoi Atmosfere* (Naukova Dumka, Kiev), 1967.
(4) A. V. Alexeev, M. V. Kabanov, I. F. Kushtin & N. F. Nelyubin, *Opticheskaya refraktsiya v zemnoi atmosferye* (Nauka, Novosibirsk), 1983, p. 58.
(5) J. B. J. Delambre, *Astronomie Théorique et Pratique, Tome Premier* (Courcier, Paris), 1814.
(6) S. Newcomb, *A Compendium of Spherical Astronomy* (Macmillan, New York), 1906, Ch. 8.
(7) R. Ball, *A Treatise on Spherical Astronomy* (Cambridge University Press), 1915.
(8) W. M. Smart, *Text-Book on Spherical Astronomy* (Cambridge University Press), 1962, Ch. 3.
(9) R. W. Hamming, *Numerical Methods for Scientists and Engineers (2nd edition)* (McGraw-Hill, New York), 1973.
(10) W. F. Meggers & C. G. Peters, *ApJ*, **50**, 56, 1919.
(11) C. A. Murray, *Vectorial Astronomy* (Adam Hilger, Bristol), 1983.
(12) A. T. Young, *AJ*, **127**, 3622, 2004.
(13) R. M. Green, *Spherical Astronomy* (Cambridge University Press), 1985.
(14) R. C. Stone, *PASP*, **108**, 1051, 1996.
(15) A. Fletcher, *J. Inst. Navigation (London)*, **5**, 307, 1952.
(16) G. G. Bennett, *J. Inst. Nav.*, **35**, 255, 1982.
(17) Th. Saemundsson, *S&T*, **72**, 70, 1986.
(18) A. T. Young, *Appl. Opt.*, **33**, 1108, 1994.
(19) A. D. Wittmann, *AN*, **318**, 305, 1997.
(20) F. Baily, *An Account of the Revd. John Flamsteed* (Lords Commissioners of the Admiralty, London), 1835, p. 149.
(21) P. Bouguer, *Mém. Acad. Roy. Sci., (for the year 1749)*, 75, 1753.
(22) A. Wegener, *Annalen der Physik*, **57**, 203, 1918.
(23) J. Kepler, *Optics: Paralipomena to Witelo & Optical Part of Astronomy* (Green Lion Press, Santa Fe, NM), 2000, p. 144.
(24) J. F. Chappell, *PASP*, **45**, 281, 1933.
(25) C. V. Raman & S. Pancharatnam, *Proc. Indian Acad. Sci. A*, **49**, 251, 1959.
(26) J. F. Davis & T. B. Greenslade, *Physics Teacher*, **29**, 47, 1991.
(27) A. Bravais, *Ann. Chim. Phys.*, **46**, 492, 1856.
(28) J. Thompson, *Brit. Assoc. Adv. Sci. Report*, **42**, 41, 1872.
(29) W. H. Lehn, *Amer. J. Phys.*, **69**, 598, 2001.
(30) J. de Graaff Hunter, *Professional paper — No. 14: Formulæ for atmospheric refraction and their application to terrestrial refraction and geodesy* (Survey of India, Dehra Dun), 1913.
(31) Committee on Extensions to the Standard Atmosphere, *US Standard Atmosphere, 1976* (US Government Printing Office, Washington, D.C.), 1976.
(32) G. Bomford, *Geodesy* (Clarendon Press, Oxford), 1971.
(33) B. B. Balsley *et al.*, *J. Atmos. Sci.*, **60**, 2496, 2003.
(34) Emeritus [Thomas Young], *Quart. J. Sci. Lit. & Arts*, **11**, 174, 1821.
(35) J. B. Biot, *C. R. Acad. Sci.*, **3**, 237, 1836.
(36) J. B. Biot, *Additions a la Connaissance des Tems, (for the year 1839)*, 3, 1836.
(37) B. Oriani, *Ephemerides astronomicae anni 1788: Appendix ad ephemerides Anni 1788* (Appresso Giuseppe Galeazzi, Milano), 1787, pp. 164–277.
(38) B. Oriani, *Opuscula Astronomica ex Ephemeridibus Mediolanensibus ad annos 1788 & 1789 excerpta* (Joseph Galeatium, Mediolani [Milan]), 1787, pp. 44–107.
(39) W. D. Bruton & G. W. Kattawar, *Appl. Opt.*, **36**, 6957, 1997.
(40) W. D. Bruton & G. W. Kattawar, *Appl. Opt.*, **37**, 2271, 1998.
(41) G. D. Cassini, *[correspondence on refraction]*, in *Ephemerides Novissimæ Motuum Coelestium Marchionis Cornelii Malvasiæ* (ex typographia Andreæ Cassiani, Mvtinæ impensis avthoris), 1662.
(42) J. W. Shirley, *Amer. J. Phys.*, **19**, 507, 1951.
(43) J. Lohne, *Centaurus*, **6**, 113, 1959.
(44) R. Descartes, *Discourse on Method, Optics, Geometry, and Meteorology; Translated, with an Introduction, by Paul J. Olscamp (Revised Edition)* (Hackett Publishing, Indianapolis), 2001.
(45) I. Newton, *Opticks* (Dover, New York), 1952, pp. 271–273.
(46) A. T. Young, G. W. Kattawar, & P. Parviainen, *Appl. Opt.*, **36**, 2689, 1997.
(47) A. M. Smith, *Trans. Amer. Philos. Soc.*, **86**, no. 2, 1996.
(48) J. Ivory, *Phil. Mag.*, **57**, 321, 1821.
(49) J. B. Biot, *Recherches sur les réfractions extraordinaires qui ont lieu près de l'horizon* (Garnery, Paris), 1810.

(50) R. Meyer, in F. Linke & F. Möller (eds.), *Handbuch der Geophysik* (Gebr. Borntraeger, Berlin), 1942–1961, Kap. 13, pp. 769–821.
(51) A. T. Young & G. W. Kattawar, *Appl. Opt.*, **37**, 3785, 1998.
(52) J. H. Lambert, *Les proprietés remarquables de la route de la lumière* (N. van Daalen, La Haye), 1759.
(53) C. M. Bauernfeind, *AN*, **62**, 209, 1864.
(54) J. Ivory, *Phil. Trans. Roy. Soc.*, (**113**), 409, 1823.
(55) A. M. Legendre, *Exercices de Calcul Intégral, sur divers ordres de Transcendantes et sur les Quadratures* (Courcier, Paris), 1811, p. 294.
(56) S. P. Rigaud, *Supplement to Dr. Bradley's Miscellaneous Works: with an account of Harriot's astronomical papers* (Oxford University Press), 1833.
(57) Chr. Kramp, *Analyse des Réfractions Astronomiques et Terrestres* (E. B. Schwikkert, Leipsic), 1799.
(58) D. J. K. O'Connell, *The Green Flash and Other Low Sun Phenomena* (North Holland, Amsterdam), 1958.
(59) L. Auer & E. M. Standish, *AJ*, **119**, 2472, 2000.
(60) E. E. Kummer, *Monatsber. Kgl. Preuss. Akad. Wiss. Berlin*, **5**, 405, 1860.
(61) P. S. Laplace, *Celestial Mechanics. Translated, with a commentary, by Nathaniel Bowditch* (Chelsea Pub., Bronx, New York), 1966.
(62) L. K. Kristensen, *AN*, **319**, 193, 1998.
(63) A. Bemporad, *Mitt. Grossherzogl. Sternwarte Heidelberg*, No. 4, 1, 1904.
(64) R. Emden, *Met. Zs.*, **40**, 173, 1923.
(65) T. Simpson, *Mathematical Dissertations on a Variety of Physical and Analytical Subjects* (T. Woodward, London), 1743, pp. 46–61.
(66) N. Maskelyne, *Phil. Trans. Roy. Soc. (Lond.)*, **54**, 263, 1764.
(67) T. Mayer, *Tabulæ motuum Solis et Lunæ novæ et correctæ* (Typis Gulielmi et Johannis Richardson, Londini), 1770.
(68) E. von Oppolzer, in W. Valentiner (ed.), *Handwörterbuch der Astronomie* (Verlag von Eduard Trewendt, Breslau), 1901, Vol. IIIb, pp. 548–601.
(69) G. Dietze, *Zeitschr. f. Meteorologie*, **9**, 169, 1955.
(70) A. T. Young, *Optics and Photonics News*, **10**, 31, 1999.
(71) A. T. Young, *JOSA A*, **17**, 2129, 2000.

'BEST TIME' FOR THE FIRST VISIBILITY OF THE LUNAR CRESCENT

By A. H. Sultan
Physics Department, Sana'a University, Yemen

The concept of 'best time' for the first visibility of the thin crescent moon developed by Bruin, Schaefer, and Yallop did not consider the elevation of the site of observation. Our first estimation — after analyzing some documented observations — is that the 'best time' is directly proportional to site elevation and inversely proportional to the Moon's altitude. For moderate-elevation sites (less than 1000 m) the crescent could first be seen shortly after sunset. However, for higher elevations (around 2000 m) the crescent could first be seen shortly before moonset.

By using our first-visibility photometric model, the extensive data of Blackwell's 1946 experiment, and the measured twilight-sky brightness of our site (1990 m), we find that the optimum lunar altitude for first visibility is about 2 degrees, no matter what the lunar elongation is.