A Series of Fortunate Events: Serendipitous Encounters with Remarkable Materials

Dr Mark P Silverman, G A Jarvis Professor of Physics
Trinity College (Hartford CT USA)
www.mpsilverman.com

Abstract

Chance involvement in bizarrely controversial issues relating to the unexpected behaviour of more or less ordinary materials led to many of the projects I have undertaken as a physicist. Some of these unusual undertakings included (a) organic dyes and the amplification of light that does *not* pass through them, (b) left-right asymmetric materials and the resolution of conflicting claims over the validity of Maxwell's equations, (c) opaquely turbid media and the surprising capacity to see through them with polarised light, (d) radioactive materials and the radical proposition that nuclear decays are correlated by an unknown universal force, (e) exploding glass and the puzzling question of how they fragment, and (f) quantum condensates and the unresolved fundamental problem of matter distribution in the universe. To the extent that time permits, I will discuss salient features of these diverse physical systems and the materials that contributed to, or helped resolve, the associated controversies.

1.INTRODUCTION

Throughout a long career as a research scientist, beginning in microbiology, evolving through organic and physical chemistry, and ultimately leading to atomic, optical, and nuclear physics, I never actually thought of myself as a materials scientist. It was therefore a surprise, albeit pleasant one, to be invited to give a plenary lecture at a meeting devoted to the science of materials, especially as I have never studied the kinds of exotic materials like graphene or metamaterials, that are the subjects of much recent investigation. Nevertheless, taking up my host's challenge—'You've done many interesting things, so I'm sure you'll think of something.'—I re-examined some past and current projects with due attention to how remarkable and unexpected can be the behaviour of the *non* exotic, relatively common materials that played key roles.

The events and associated materials that I will discuss may be classified in terms of a series of scientific controversies :

- Controversy: Light amplification by reflection
  - Material: Rhodamine dye
- Controversy: Light reflection from chiral materials
  - Material: Camphorquinone
- Controversy: Light depolarisation by scattering
  - Material: Latex spheres of micron size
- Controversy: "Cosmogenic Force" and nuclear decay
  - Material: Na-22 (a positron emitter)

- Controversy: The explosion of Rupert's Drops
  - Material: Glass (soda lime and lead crystal)
- Controversy: The 'dark matter' of the Universe
  - Material: WIMPs? WIDGETs? Something Else?

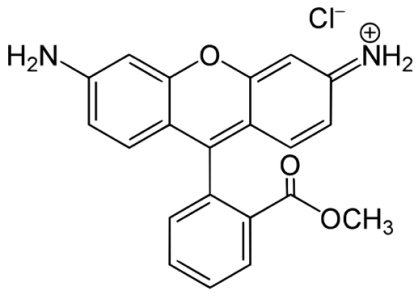
2. LIGHT AMPLIFICATION BY REFLECTION

Although Einstein proposed the existence of the process of stimulated emission in a 1917 paper, the first working laser was not actually built until 1960 (by Theodore Maiman of the Hughes Aircraft Company). The word 'laser' is an acronym for **L**ight **A**mplification by **S**timulated **E**mission of **R**adiation. In principle, all that was necessary to go from Einstein's prediction to an experimental realisation was to confine a suitable collection of excited atoms to a cavity by the addition of end mirrors, so that multiple passage of an initial quasi-monochromatic lightbeam could induce the atoms to de-excite synchronously, thereby producing coherent replicas of the initiating photons. Arthur Schawlow, who first proposed the use of mirrors in a 1958 paper with Charles Townes and is credited with the invention of the laser (which was then referred to as an 'optical **m**aser', the 'm' standing for microwave), although he did not produce a working model at the time, later described the *modus operandi* of a laser in the following colourful terms:

> *It is as if tiny mechanical men, all wound up to a certain energy and facing along the axis of the laser enclosure, were successively set in motion by other marchers and fell into step until they became an immense army marching in unison row on row (the plane wave fronts) back and forth in the enclosure. After the laser light has built up in this way it emerges through the partly reflecting mirror at one end as an intense, highly directional beam.*
>
> *—Arthur Schawlow (1968)*

The lasing material in Maiman's laser was ruby crystal, a solid. In time, lasing would be observed in a wide variety of materials in other phases as well. The first gas laser (helium-neon) was developed by Ali Javan in 1960. Liquid lasers employing organic dyes dissolved in solvent were developed independently by Peter Sorokin and F. P. Schaefer (and colleagues) in 1966. In all cases, the basic idea, captured in Schawlow's anthropomorphic picture of laser operation, was that light had to pass *through* a lasing material in order to be amplified.

In the 1970s I was using dye lasers, pumped by the UV output of a pulsed nitrogen laser, to study the properties of highly excited atomic states (Rydberg states). Dye lasers were (and are) a particularly suitable tool because they are tunable. The emission spectrum of the dye is broad, and selected wavelengths can be returned to the cavity for amplification by means of a grating. A versatile family of dyes was that of Rhodamine with a fluorescence spectrum spanning a range of about 500-800 nm. This material, in particular Rhodamine B, was to play an important role in resolution of a surprising controversy.



Rhodamine B

Suppose, in contrast to the standard description of a laser, that the excited atoms were *inside* the enclosure and the light was *outside*—and the light merely made a momentary reflective pass from the cavity surface. Could the atoms then give up their energy 'as an immense army

marching in unison'? Would the outside light be amplified as a result of stimulated emission by the 'tail' of an evanescent wave? This was the problem of amplified or enhanced reflection that came to my attention in the late 1970s. Some physicists at the time thought the answer was 'yes'; others 'no'. An experiment that seemed to show a large amplification was later attributed to entirely different phenomena.

One might wonder why the problem seemed difficult. After all, why not just substitute the refractive index of a medium with gain—which takes the form of a complex-valued expression $\tilde{n} = n(1 - i\gamma)$ with negative imaginary part—into the familiar formulas for the Fresnel coefficients[1] of a homogeneous medium and see whether some incident angles lead to numerical values greater than 1? (Note: an incident plane wave $e^{i\,2\pi(x-ct)/\lambda}$ of unit amplitude, travelling a distance Δx through a medium with the preceding complex refractive index, would emerge with a greater amplitude $e^{2\pi n\gamma\,\Delta x/\lambda}$.) This was done, but the result, however, was ambiguous.

The origin of the ambiguity, in brief, was that the Fresnel coefficients depended on a term (the normal component of the wave vector in the medium) that had to be deduced from Snel's law of refraction, which entailed a quadratic equation with two roots. The choice of one root gave rise to an amplified wave moving through the medium (i.e. *not* in the air) away from the interface, accompanied by a *non*-amplified reflected wave; the choice of the other root represented a decaying wave approaching the interface from within the medium accompanied by an amplified reflected wave. It might seem reasonable to have chosen the first solution—whereupon the reflected wave is never amplified—but this led to a difficulty for conditions of total reflection (where the light originates in the medium of greater optical density): All light ought then to be reflected for incident angles beyond critical angle, in which case there could not be an amplified wave penetrating the gain medium. On the other hand, to have chosen the second solution would have required justifying the origin of a wave that approached the interface from deep within the gain medium. Furthermore, the two solutions led to a discontinuity in reflectance precisely at critical angle.

The real origin of the problem, however, resided in the fact that an infinitely deep homogeneous medium of excited molecules (i.e. a medium with uniform gain) does not exist. Like frictionless pulleys, inextensible cords, and ideal gases, the model of a homogeneous optical medium can be useful for some purposes, but not for all. This was one of the times the model could not be used.

To resolve this conundrum, my student (Raymond Cybulski) and I took up the more realistic, but difficult, problem of a light beam reflecting from a medium in which the gain decreases exponentially with depth. Such a gain profile can be produced experimentally by illuminating the interface of the reflecting medium with another laser beam, the spatial absorption of which falls off exponentially in accordance with Beer's Law. The index of refraction of a medium with exponentially decreasing gain takes the form $\tilde{n} = n\left(1 - i\gamma e^{-z/d}\right)$ where z is the depth of penetration and d is a characteristic absorption length (or depth parameter) determined by the concentration of absorbing molecules and the absorption cross section (related to the probability of absorption).

[1] The Fresnel coefficient is a ratio of the reflected light intensity to the incident light intensity. There are two such coefficients to characterize the reflectance of light polarised either parallel to the plane of incidence (p polarization) or perpendicular to the plane of incidence (s polarisation). The plane of incidence is the plane determined by the wave vector (i.e. linear momentum) of the incident light and the outward normal to the interface of the reflecting medium.

Substitution of the exponential gain profile into Maxwell's equations (to determine the form of the light waves in the gain medium) gave rise to two rather complicated-looking differential equations, one for each state of incident light polarisation (s or p). The equation for s-polarisation was transformed to yield a Bessel function of complex-valued order as the physically acceptable solution. The equation for p-polarisation did not correspond to any recognisable form. We solved it by standard means (use of a Frobenius series), and I called the solution a 'Bussil' function by taking 3 letters each from Cy**bul**ski and **Sil**verman. Many years later, I came across an equation of similar form and found that it was a modified form of a confluent hypergeometric equation whose solutions are called Whittaker functions.

The solutions for both s and p polarisations did indeed predict amplification of reflected light, but only over a fairly restricted range of incident angles in the vicinity of critical angle, i.e. the angle θ_c at which the transmitted wave would vanish (i.e. transport no energy) were the reflecting medium to be transparent (i.e. nonabsorbing). The predicted reflectances as a function of incident angle take the following forms:

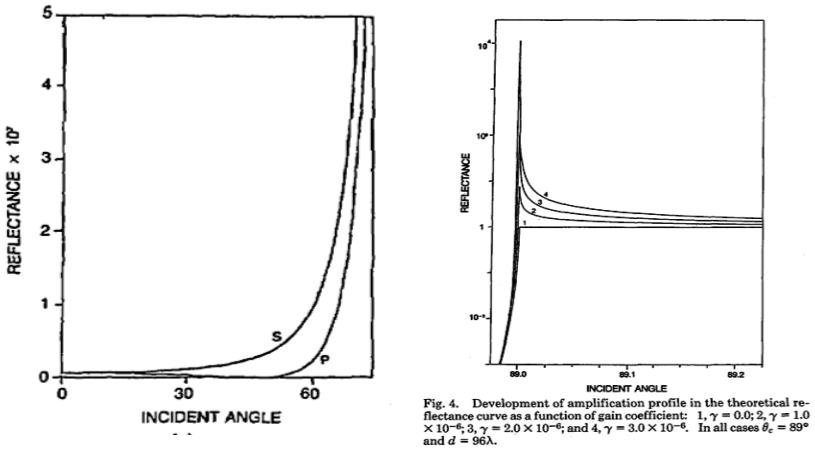


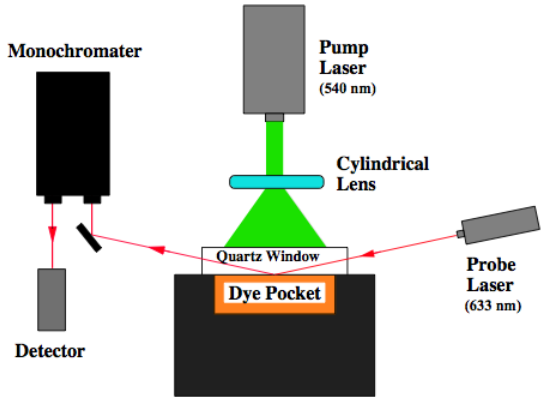
Fig. 4. Development of amplification profile in the theoretical reflectance curve as a function of gain coefficient: 1, $\gamma = 0.0$; 2, $\gamma = 1.0 \times 10^{-6}$; 3, $\gamma = 2.0 \times 10^{-6}$; and 4, $\gamma = 3.0 \times 10^{-6}$. In all cases $\theta_c = 89°$ and $d = 96\lambda$.

The left panel shows the theoretical reflectance for ordinary reflection, the p-polarised wave vanishing at the Brewster angle. The right panel shows the theoretical reflectance over the 'interesting' part of the angular range; that is, the portion extending from critical angle, which corresponded to 89°, to grazing incidence. From normal incidence to within less than a degree of critical angle, the reflectance curves are virtually the same as those calculable from the Fresnel coefficients for a homogeneous transparent medium. In the absence of gain, the reflectance would remain at 100% (a Fresnel coefficient of 1) over the angular range from critical angle to grazing incidence (see right panel). However, for a medium with gain, the reflectance overshoots unity and can become quite large depending on the gain parameter.

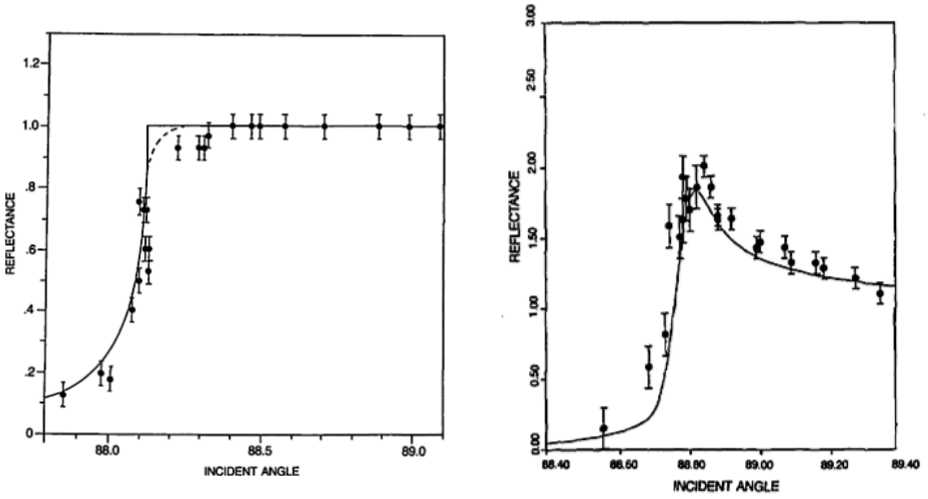
Our experiment to test this theory was carried out as follows. An amplifying medium with exponential gain was prepared from a solution of rhodamine B dye in ethyl and benzyl alcohols by excitation with a tunable pulsed dye laser (pumped by a nitrogen laser) operating at 540 nm, which is close to the center of the absorption curve of the rhodamine dye. The dye, which was contained in a pocket milled out of an aluminium block, was covered with a transparent fused-quartz window with refractive index $n_1 = 1.4570$ at the wavelength $\lambda = 633$ nm of the probe beam from a continuous helium-neon laser. Since the refractive index of ethyl alcohol is less than that of the quartz, whereas the index of benzyl alcohol is greater, we could

adjust the refractive index n_2 of the gain medium, and hence the critical angle θ_c, by adjusting the composition of the solution and then precisely controlling the temperature. The closer n_2 matched n_1, the nearer θ_c approached grazing incidence, and the greater was the predicted light amplification.

To construct a reflectance curve for a particular set of experimental conditions, we first recorded the dc signal from the He-Ne probe beam with the dye unexcited. Next, the dye was excited by the pump beam, which led to pulses of 10 ns width superposed on the dc signal. (Fluorescence from the dye was removed by passing the reflected light beam through a monochromater.) The ratio of pulse height to dc background (with fluorescence removed) yielded the reflectance coefficient *R*, which was largely independent of polarisation for large incident angles, as predicted by the theory.



The experimental results were in excellent agreement with the predictions of our theory as shown below for one set of sample measurements.



The left panel ($\theta_c = 88.12^\circ$) records the reflectance from a dye region with no pumping. The solid line corresponds to the Fresnel reflectance formulas for a transparent medium. The dashed line is the Fresnel reflectance averaged over the divergence of the probe beam, whose TEM_{00} mode had a Gaussian profile across the beam waist. The right panel records the reflectance from the pumped dye with depth parameter $d = 58\,\lambda$ and gain parameter $\gamma \sim 2.6 \times 10^{-5}$. Again, the solid curve is the theoretical reflectance averaged over the beam divergence. A maximum reflectance close to 2 was achieved.

At the time these experiments were undertaken, the existence of amplified reflection was a confused and controversial issue. Theoretical analyses disagreed with one another and, in any event, were incapable of accounting for reported observations, which were orders of magnitude

too large, non-reproducible, and attributable to other phenomena (such as photo-induced or thermally induced refractive index gradients). What our investigation showed, however, was that under well-defined conditions light amplification by reflection did indeed exist and was amenable to treatment within the framework of classical electrodynamics. Our theory gave definite, unambiguous answers to the questions of under what conditions amplification occurred and how large the effect would be. It also clarified the confusing matter of waves within the gain medium moving toward or away from the interface. This is worth examining briefly.

Since the mathematical amplitudes of the theory (Bessel and Bussil functions) are not only complex in the sense of analysis but also rather complicated, it is instructive to ask precisely why, physically, amplification occurs. As pointed out earlier, we have here a mechanism of stimulated emission by an evanescent wave—i.e. a wave that decreases exponentially into the gain medium to an extent characterised by a depth parameter *d*. In the case of total reflection from a homogeneous transparent medium, the evanescent wave transports no power—i.e. the time-averaged Poynting vector is zero. For a medium with gain, however, the reflected wave serves as the trigger to release energy stored in the pumped medium, and the transmitted wave transports energy to the surface. An examination of the series representation of the light (either polarisation) in the excited medium shows that it consists of a linear superposition of two series representing waves that travel, respectively, toward and away from the surface. There is no violation of physical law, for only an outgoing wave is to be found in the far-field region where for all practical purposes the gain has decreased to zero and the medium is transparent. Nevertheless, for amplified reflection to take place, both types of waves must be present simultaneously in the gain region, their amplitudes and phases being specified by the theory with no arbitrariness.

In the exponential gain model (i.e. where absorption of the pump beam follows Beer's law), there is an angular region, including normal incidence, for which only the transmitted wave travelling away from the interface exists. Thus, for a given set of experimental parameters, the theory predicts a lower limit to the angle of incidence at which a reflected wave can be amplified—in contrast to the simple extension of the Fresnel amplitudes to a medium with uniform gain, for which no such lower limit occurs.

When, nearly 20 years after the work was completed, I wrote of this controversy in my book *Waves and Grains*[2], I speculated that the process of enhanced reflection would have important practical applications for the telecommunications industry in its transition from a network of electrical wires to an all-optical system of light fibres. Enhanced reflection, applied at regular intervals in an optical transmission line, would make long-distance transmission practical by reducing optical losses without requiring conversion to an electronic signal for subsequent amplification and optical reconversion. This has indeed come to pass and doped fibre amplifiers employing stimulated emission is part of current technology. As I then wrote:

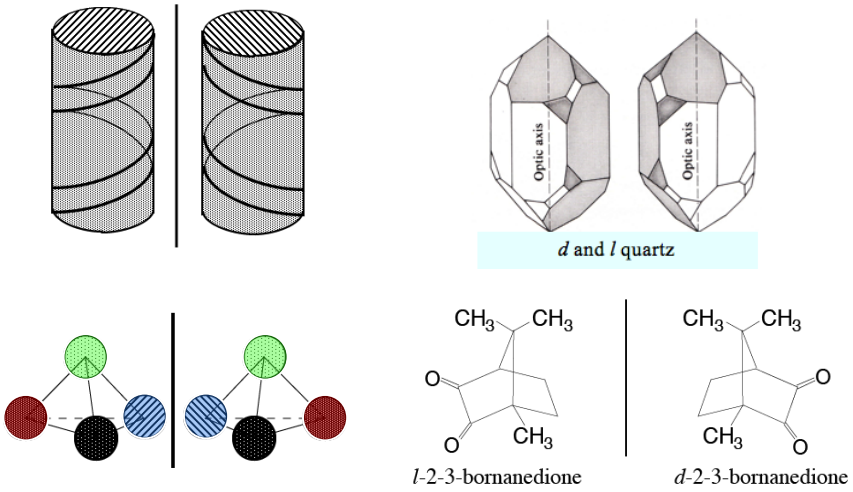
> *In Fresnel's time, the study of reflection helped elucidate the nature of light. During the twenty-first century, the same process will help carry that light around the world.*

[2] M P Silverman, *Waves and Grains: Reflections on Light and Learning* (Princeton U P, 1998).

Regrettably, the 'shelf life' of even seminal publications can be short—and few in the optical communication industry today are probably aware of our formative theory[3] and definitive experiments[4] that resolved the controversy surrounding the remarkable phenomenon of amplified reflection.

3. LIGHT REFLECTION FROM CHIRAL MATERIALS

Materials or interactions that exhibit an asymmetry with respect to handedness (left vs right) or rotational sense (clockwise vs counterclockwise) are said to be chiral from the Greek root for 'hand'. This chirality arises either because the individual molecules are chiral or because the molecular arrangement (even if consisting of achiral units) displays a chiral structure.



The upper panel shows an example of structural chirality—mirror-image helices— as exemplified by the spiral staircase arrangements of achiral silicon dioxide that make up crystalline quartz. The lower panel shows an example of molecular chirality—mirror-image tetrahedral molecules comprised of four different atoms—as would be exemplified by an 'asymmetric' carbon atom, i.e. a carbon atom linked by sp^3 chemical bonds to four different functional groups. The molecule shown at right—2-3-bornanedione also known as camphorquinone—contains such an asymmetric carbon atom (the one with the forward-pointing methyl group). This molecule played an important role in resolving the scientific controversy I will discuss shortly.

The two enantiomeric forms of a substance exhibit the same physical properties such as density, melting point, boiling point, heat capacity, etc. When interacting with non-chiral agents, they exhibit the same chemical properties such as reaction rates. However, their effects on, or reactions with, chiral agents can be markedly different. Living organisms are built from chiral molecules, such as *l*-amino acids and *d*-sugars. (*l* stands for laevo = left; *d* stands for dextro =

[3] R Cybulsky and M P Silverman, "Investigation of light amplification by enhanced internal reflection. Part I. Theoretical reflectance and transmittance of an exponentially nonuniform gain region", *J. Opt. Soc. Am.* **73** (1983) 1732-1738.

[4] M P Silverman and R Cybulsky, "Investigation of light amplification by enhanced internal reflection. Part II. Experimental determination of the single-pass reflectance of an optically pumped gain region.", *J. Opt. Soc. Am.* **73** (1983) 1739-1743.

right.) Indeed, chirality is a hallmark of the living state. Because of chiral receptors in the human olfactory system, *d*-carvone smells like spearmint, whereas *l*-carvone smells like caraway. Thalidomide is a chiral molecule. In humans, one enantiomer was a mild sedative; the other a teratogen causing serious birth defects. Pharmacology now pays careful attention to the chirality of drugs.

Light itself comes in two chiral forms referred to as left and right circular polarisations (LCP, RCP). Linear superpositions of LCP and RCP light produce orthogonal forms (e.g. vertical and horizontal) of linearly polarised light. A chiral material interacting with light exhibits properties collectively referred to as optical activity, four principal manifestations of which can be classified as follows:

- **Optical Rotation**: Rotation of the plane of incident linearly polarised light resulting from the difference in phase velocities of LCP and RCP components in the medium.
- **Circular Dichroism**: Conversion of incident linearly polarised light into elliptically polarised light due to the difference in absorption of LCP and RCP components.
- **Differential Chiral Refraction** - Refraction (bending) of the wave vectors of LCP and RCP components to different angular extents.
- **Differential Chiral Reflection** - Difference in the intensities of reflected LCP and RCP components.

All the effects are manifestations of circular birefringence: the fact that the refractive indices of LCP and RCP light are different in a chiral medium. The first two effects, which derive respectively from the real and imaginary parts of the chiral refractive indices, have long been used by chemists and physicists for quantitative identification of chiral solutes in solution, as, for example, the determination of glucose concentration. The third effect is of historical interest because it provided the means by which Augustin Fresnel first demonstrated (in the 1820's) the existence of LCP and RCP light by passing linearly polarised light through a composite prism made of alternating segments of left-handed and right-handed quartz. The fourth effect, so far as I knew when the matter of chiral optics first attracted my attention, had never been observed with a naturally chiral material.

I became interested in the subject initially because of a puzzling historical question—puzzling, perhaps, only to me for I am not aware that anyone else cared to wonder about it. I wanted to know whether Fresnel, whose name was associated with both the amplitudes for reflection *and* the discovery of circular polarisation, ever put the two themes together in his mind and sought either to calculate or measure the difference in intensity with which an optically active material would reflect incident LCP and RCP light waves. In other words, did Fresnel ever calculate the 'Fresnel coefficients' for reflection from a homogeneous, isotropic chiral material?

I had actually wondered before that about other aspects of Fresnel's remarkable contributions: for example how he was able to obtain the correct expressions for the reflection amplitudes about a half-century before Maxwell's theory of electromagnetism. To that question, I eventually found an answer (discussed in my book, *Waves and Grains*) when I held the Joliot Chair of Physics at the ESPCI (*Ecole Supérieure de Physique et de Chimie Industrielles*) in Paris and found the complete works of Fresnel—dusty and apparently untouched for decades—on the topmost shelf of an ancient wooden cabinet behind my worktable. The fact that Fresnel's collected papers comprised several thick volumes illustrated the wondrous productivity of a young man who died before the age of 40, and yet whose name was connected with an astonishing number of objects and concepts in optics: Fresnel equations, Fresnel diffraction,

Fresnel integral, Fresnel amplitudes, Fresnel coefficients, Fresnel drag, Fresnel lens, Fresnel lantern, Fresnel number, Fresnel rhomb, Fresnel zone, Fresnel zone plate, and more.

From all that I could find by or about Fresnel, I concluded that he neither calculated nor attempted to measure the Fresnel coefficients for a chiral material. So I tried the calculation myself, realising soon enough that the problem disguised a serious and subtle matter of principle that had led astray others who attempted the problem before me. The problem was that if one adopted the familiar constitutive relations—i.e. the expressions linking fields **D** and **B** to fields **E** and **H** in Maxwell's equations—there resulted Fresnel amplitudes that violated the conservation of energy. In other words, more light could reflect from the interface than was incident upon it—except that in *this* case there was no excited medium. The problem seemed insoluble and there were even speculations at the time that Maxwell's theory of electromagnetism would have to be modified to treat chiral electrodynamics satisfactorily.

The refractive indices for LCP and RCP light take the general form $\tilde{n}_{\pm} = \tilde{n}\left(1 \pm \tilde{f}\right)$, where the "+" is associated with the LCP component. The real and imaginary parts of the mean refractive index $\tilde{n} = n + i\kappa$ respectively characterise refraction and absorption of the medium with neglect of chirality, and the real and imaginary parts of the chiral parameter $\tilde{f} = f_r + if_i$ respectively characterise the phenomena of optical rotation and circular dichroism. Such a form follows from the solution of Maxwell's equations, employing either the familiar set of constitutive relations normally found in optics books

$$\left.\begin{aligned} \mathbf{D} &= \vec{\varepsilon}\mathbf{E} \\ \mathbf{B} &= \mu\mathbf{H} \end{aligned}\right\} \quad \text{Asymmetric Set,}$$

which I have termed asymmetric because the displacement field is related to the electric field by a dielectric *tensor* whereas the magnetic induction is related to the magnetic field by a *scalar* permeability, or the less familiar set of relations (obtained from a quantum mechanical calculation of the molecular moments induced by an electromagnetic field[5])

$$\left.\begin{aligned} \mathbf{D} &= \varepsilon\mathbf{E} - g\dot{\mathbf{H}} = \varepsilon\left[\mathbf{E} + if\,\hat{\mathbf{k}}\times\mathbf{E}\right] \\ \mathbf{B} &= \mu\mathbf{H} + g\dot{\mathbf{E}} = \mu\left[\mathbf{H} + if\,\hat{\mathbf{k}}\times\mathbf{H}\right] \end{aligned}\right\} \quad \text{Symmetric Set,}$$

which I have termed symmetric for reasons that should be apparent. In the symmetric set each of the fields on the left is connected to both **E** and **H** fields on the right. In the first equality, the dot over a vector signifies a time derivative; g is a more basic chiral parameter from which the parameter f in the refractive index arises. In the transformation to the second equality, obtained by use of Maxwell's equations, $\hat{\mathbf{k}} = \mathbf{k}/\left(n\omega/c\right)$ is a unit wave vector and ω is the angular frequency of the light wave. It may seem surprising and perhaps incorrect that **B** would be an *anisotropic* function of **H** even for a homogeneous, isotropic, intrinsically nonmagnetic chiral medium—but, in fact, neglect of this condition has been the fatal flaw in theoretical analyses of chiral reflection previous to mine.[6]

---

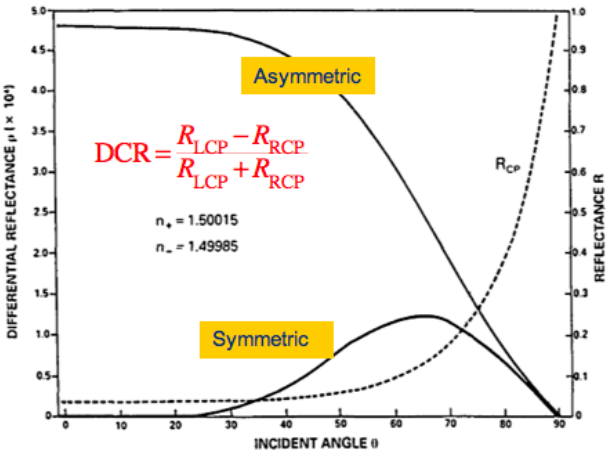
[5] H. Eyring, J. Walter, and G. E. Kimball, *Quantum Chemistry* (Wiley, New York, 1944) Chapter XVII.

[6] M P Silverman, "Specular Light Scattering from a Chiral Medium: Unambiguous Test of Gyrotropic Constitutive Relations," *Lettere al Nuovo Cimento* **43** (1985) 378-382; "Reflection and Refraction at the Surface of a Chiral Medium," *J. Opt. Soc. Am. A* **3** (1986), 830-837 and erratum *A* **4** (1987) 1145.

I have discussed in my book *Waves and Grains* the arguments given in support of the asymmetric form and why those arguments were inadequate. Suffice it to say that although both sets of constitutive relations produced the same form of the chiral refractive indices, they did not lead to equivalent expressions for the Fresnel coefficients—and therefore they were not physically equivalent. Of particular conceptual interest, was that the symmetric relations led to Fresnel coefficients that satisfied conservation of energy for all physical conditions, whereas the asymmetric relations violated energy conservation under conditions of total reflection. An example of the different outcomes of the two theoretical starting points is shown in the plot below of the differential circular reflection (also called the circular intensity difference)

$$\mathrm{DCR} = \frac{R_+ - R_-}{R_+ + R_-}$$

under conditions of ordinary reflection. The asymmetric relations lead to a maximum difference in chiral reflection at normal incidence, whereas the maximum DCR predicted by the symmetric relations occurs somewhere around Brewster's angle and, in fact, leads to *zero* difference at normal incidence.



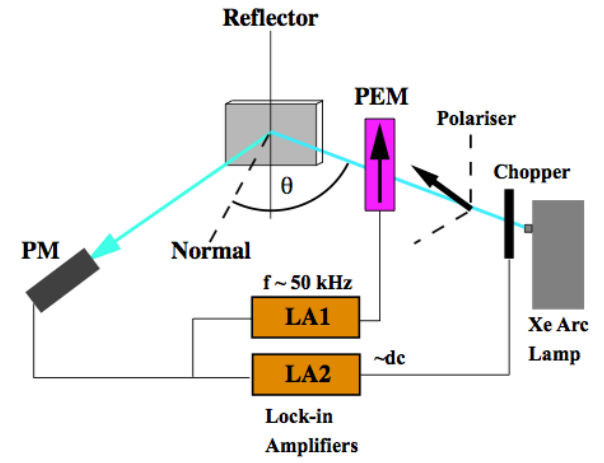
Again, one might think initially that the last prediction must surely be incorrect, for why should a chiral material *not* reflect LCP and RCP light differently at normal incidence, since the two polarisations propagate through the medium with different phase velocities? The answer is related to the particular space and time symmetries of natural (in contrast to magnetically-induced) optical activity. As a result of these symmetries, a linearly polarised light beam that passes through a transparent optically active material and is reflected back along the same path by a mirror, will emerge with *no* optical rotation of its plane of polarisation. That is, to whatever extent the plane of polarisation has been rotated on the forward trip, it will be rotated equally in the opposite sense on the reverse trip. Light reflected normally from the surface of a chiral material may be thought of as undergoing a similar reversed chiral interaction. As justified by the so-called Ewald-Oseen extinction theorem, the incident light beam does not interact with the reflecting material at the surface only, but propagates a certain depth into the material and is extinguished by the molecular dipoles, which are then induced to radiate secondary wavelets that superpose coherently to form the reflected wave. Reversal of wave helicity (i.e. the sense of rotation of the light fields relative to the wave vector) upon reflection results in an opposite chiral effect on the outgoing wave, with the result that the DCR vanishes. At larger angles of incidence the incident and reflected waves no longer overlap, and perfect cancellation of chiral asymmetry does not occur.

Having solved to my satisfaction the theoretical problem of light reflection from a chiral material, I next devised an experiment to measure the DCR from a naturally optically active material and test the theory. Because the difference in refractive indices of LCP and RCP light is so very small—the parameter f is of order 10^{-7}—and the chiral parameter enters each Fresnel coefficient (R_+, R_-) *quadratically*, it would have been impossible to detect this difference by

measuring R_+ and R_- separately and taking their difference. The sought-for signal would have been lost in the much larger accompanying noise. Perhaps Fresnel, if he thought about the problem at all, might have recognised this difficulty and was deterred from attempting such an experiment.

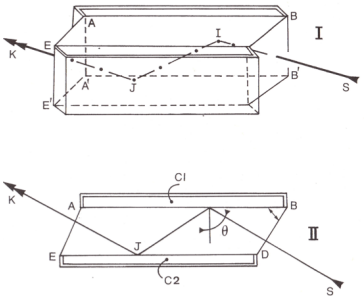
The key to a successful experiment was to measure the differential reflection simultaneously by modulating the phase of the probe beam. Theory indicated that the difference $\left(R_+ - R_-\right)$ is linearly proportional to *f*. Even so, the experimental problem was daunting. To put the level of difficulty into perspective, note that the sensitivity required to measure the DCR from a material whose molecular optical activity arose from 'ordinary' electromagnetic interactions was comparable to the sensitivity required to measure optical rotation in an atomic vapour (for example, Thallium) induced by parity non-conserving weak nuclear interactions. Nevertheless, the theory I developed together with my colleague J. Badoz suggested several ways to enhance the signal. In particular, the DCR was predicted to be significantly enhanced by multiple reflection under conditions of total reflection in the vicinity of critical angle at frequencies that fell within the absorption band of the sample.[7]

In the experimental configuration that eventually succeeded[8] light from a xenon arc lamp (at 476 nm) was mechanically chopped at a low frequency, then linearly polarised at 45° to the axis of a photoelastic modulator whose output was a light beam with polarisation oscillating between LCP and RCP states at a high rate of 50 kHz. This phase-modulated beam reflected from a chiral sample and was detected synchronously by a photomultiplier tube and two lock-in amplifiers which extracted from the photocurrent *I* signals at the modulation frequency f_m and at the mechanical chopping frequency, low enough to be considered dc or 0 Hz. In my analysis of the experiment, I had shown that the DCR was deducible experimentally from the relation



$$\mathrm{DCR} = \frac{I(f_m)/I(0)}{2J_1(\varphi_m)}$$

where the denominator is a Bessel function whose argument is the modulation amplitude, experimentally set at ~2.4 radians (for reasons discussed in *Waves and Grains*). To take advantage of enhancement by multiple reflection, a sample cell was designed in which total reflection occurred twice during passage. The chiral sample was a methanolic solution of camphorquinone with the following pertinent optical parameters at 476 nm:



[7] (a) M. P. Silverman and J. Badoz, "Large Enhancement of Chiral Asymmetry in Light Reflection near Critical Angle", *Opt. Comm.* **74** (1989) 129-133. (b) M. P. Silverman and J. Badoz, "Multiple Reflection from Isotropic Chiral Media and the Enhancement of Chiral Asymmetry", *J. Electromagnetic Waves and Applications* **6** (1992) 587-601.
[8] M. P. Silverman, J. Badoz, and B. Briat, 'Chiral Reflection from a Naturally Optically Active Medium", *Opt. Lett.* **17** (1992) 886-888.

$$n = 1.3328 \qquad f_r = -0.95 \times 10^{-7}$$
$$\kappa = 1.03 \times 10^{-4} \qquad f_i = -4.37 \times 10^{-7}$$

Measured as a function of light energy (expressed in wave numbers cm^{-1}), the experimental values of DCR were in excellent agreement with theory calculated from the chiral Fresnel coefficients derived from the symmetric constitutive relations with substitution of the optical constants deduced from measurements of absorption, rotary power, and circular dichroism as a function of wave number.

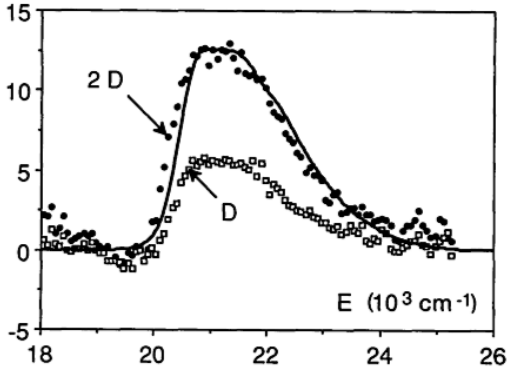
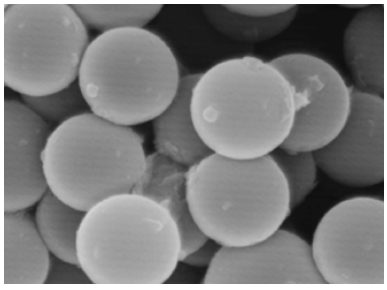


Fig. 3. Differential circular reflection DCR ($\times 10^5$) as a function of wave number E for $\theta = 67°$. D denotes DCR(1) for a single chiral reflection, and 2D denotes DCR(2) for two chiral reflections. The solid curve is the theoretical DCR(2) estimated from the optical constants of the solution.

The solid curve of the figure marks the theoretically predicted DCR for two reflections. At 21,000 cm^{-1} (476 nm), close to the maximum of the absorption band of the sample, the DCR was measured to be $D = 17 \times 10^{-5}$ with rms noise of 10^{-6} when the incident angle was set as close as experimentally possible to critical angle $\theta_c = 65.58^\circ$. In the plot shown, the incident angle was $\theta = 66.50^\circ$ and the DCR was about 6×10^{-5}. It is worth noting again that the differences in chiral refractive indices (circular birefringence) and absorption constants (circular dichroism) are numbers of the order of 10^{-7}. Under conditions of ordinary single-pass reflection this would also have been the order of magnitude of the DCR and other manifestations of chiral scattering. Thus, the observed maximum DCR of approximately 10^{-4} represented a thousandfold enhancement.

In a certain sense that was historically satisfying to me, my theoretical resolution of the problem of chiral light reflection and the ensuing first experimental observation of the difference with which a naturally optically active material reflected circularly polarised light completed the list (presented earlier) of the four basic manifestations of optical activity and provided the complement to Fresnel's experiment detecting the differential refraction of light from a chiral material. Investigations, both theoretical and experimental, of the scattering of electromagnetic waves from chiral structures have since become an intensely researched field with applications beyond the traditional ones of spectroscopy and polarimetry to communications, remote detection, radar evasion, and other interests.

4. TURBID MEDIA

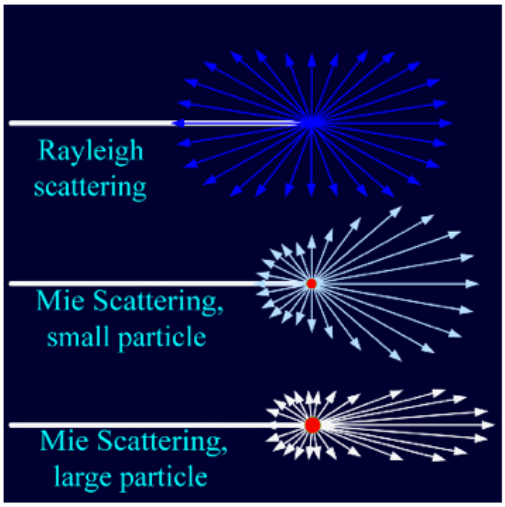


A glass of whole milk is a turbid medium. It appears opaque not because it has absorbed the light passing through, but because numerous suspended particles (globules of fat) scatter the light in all directions, thereby destroying information carried by the wave vectors concerning their initial directions of emission. Moreover, because the type of scattering (Mie) is largely independent of wavelength, all wavelengths are scattered more or less equally and the substance appears white. Turbid fluids are encountered everywhere: the atmosphere, the ocean, bodily fluids (e.g. blood), foods, industrial products, and laboratory research. In my own lab, turbid fluids with well-defined properties were made from suspensions of latex spheres of diameters on the order of

microns. Contained in a glass cuvette and illuminated with monochromatic light, the suspension appeared murky like milk.

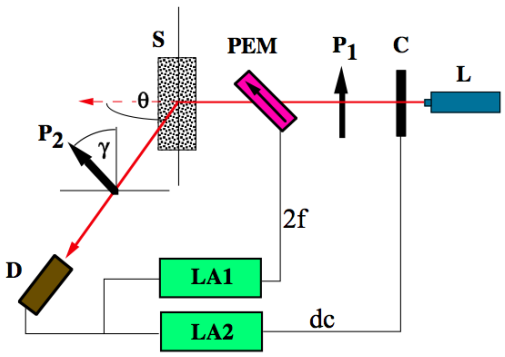
My interest in turbid fluids was motivated by doubts concerning two prevailing views at the time that light scattering (a) randomises the emerging wave vectors and destroys image information, and (b) depolarises light, thereby destroying information obtainable through optical activity. Thus, I set about to answer such questions as

- How does multiple scattering in a turbid medium affect visibility?
- Are there ways to see through a turbid medium?
- How much does scattering depolarise light?
- Can one measure optical activity in a turbid chiral medium?



In a manner of speaking that is not merely rhetorical but technically accurate, all of optics can be understood in terms of the scattering of light (a point I have emphasised in my book *Waves and Grains*). The features of the scattering, however, depend on the relative sizes of the scatterers and the wavelength. Among other things, this comparison of length scale strongly affects the angular distribution of the scattered radiation. Light scattering from particles (e.g. molecules) smaller than the wavelength constitutes Rayleigh scattering. The angular distribution of the scattered light manifests cylindrical symmetry about the light beam (incident wave vector) and mirror symmetry about a plane through the scatterer normal to the light beam (i.e. equal forward and backward scattering). Rayleigh scattering is also highly dependent on wavelength and sensitive to polarisation; the scattered intensity (or cross section) being inversely proportional to λ^4. As the particle size increases relative to the wavelength, the angular distribution tilts forward. Very large particles scatter predominantly in the forward direction. Although large-particle light scattering is often referred to as Mie scattering, the exact expressions for Mie scattering encompass all particle sizes. Large-particle scattering is largely independent of wavelength and polarisation.

Besides the attributes of the scatterer, the properties of scattered light depend on whether scattering of a photon occurs once or multiple times—i.e. whether a typical particle is illuminated only by the light of the source or also by light that was scattered by other particles. The first case, single scattering, gives rise to an exponential decrease of intensity with increasing distance from the source—that is, a Beer's law dependence. The second case, multiple scattering, leads to a much slower negative-power law decrease in intensity with distance. Optics books and introductory physics texts often attribute the blue colour of the sky to Rayleigh scattering, the idea being that blue light (475 nm) is scattered about 3.5 times more than red light (650 nm) due to the λ^4 dependence. While not exactly incorrect, the explanation is incomplete and misleading because of two important omissions. First is the occurrence of density fluctuations in the atmosphere producing fluctuations in the permittivity (dielectric constant) of the air, without which there would be no scattering. And second is the requirement of single-particle scattering. I have shown in *Waves and Grains* that Rayleigh scattering under conditions of multiple scattering can lead to rather exotic colours of the sky, depending on atmospheric pressure.



To answer the questions posed above, I undertook a series of experiments to determine the intensity, degree of polarisation, and optical rotation of light scattered from a turbid fluid, either achiral or chiral, containing achiral latex spheres as scatterers. The basic configuration involved a probe beam from a He-Ne laser that was chopped and polarised at an angle of 45° to the long axis of a photoelastic modulator (PEM), which, as described previously, impressed a sinusoidal oscillation on the waveform at a modulation frequency f_m of 50 kHz. The transmitted beam then illuminated the scattering cell containing an aqueous suspension of latex spheres of specified size and concentration. The light scattered in a particular direction (forward, lateral, backward) was, as before, detected synchronously with a photomultiplier and two lock-in amplifiers, one of which analysed the photocurrent for the dc component (due to the chopper) and the other for the first harmonic ($2f_m$) of the modulation frequency.

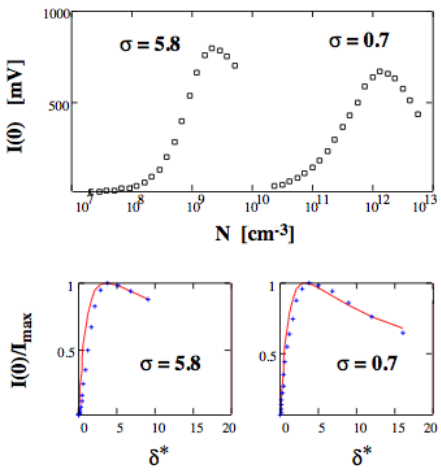
Since the heart of this experiment, like the previous one on chiral reflection, involved a PEM, I should outline briefly at this point how the device works. A piezoelectric transducer (oscillator) is bonded to a bar of fused—and therefore optically isotropic and *achiral*—quartz. Excited by the transducer, the bar vibrates longitudinally at an acoustic resonant frequency f_m (close to 50 kHz in my own PEMs) and thereby acquires a time-dependent, stress-induced linear birefringence. This dynamic birefringence results in a retardation $\varphi(t) = \varphi_m \sin(2\pi f_m t)$ between the components of an incident light field projected on the long (y) and transverse (x) axes of the quartz element. An incident wave $\mathbf{E}_i \propto \mathbf{x} + \mathbf{y}$ linearly polarised at 45° to the bar emerges with a time-varying polarisation $\mathbf{E}_t \propto \mathbf{x} + \mathbf{y}e^{i\varphi(t)}$. Depending on the modulation amplitude φ_m, the transmitted field $\mathbf{E}_t(t)$ can be made to oscillate between LCP and RCP (for $\varphi_m = \pi/2$) or between orthogonal states $(\mathbf{x} \pm \mathbf{y})$ of linear polarisation (for $\varphi_m = \pi$). Accompanied by phase-sensitive detection, the PEM enables one to measure very weak polarisations or small changes in polarisation. It has been employed in applications that span broad fields of optics: polarimetry, ellipsometry, spectroscopy, interferometry, and optical imaging.

In the experimental configuration shown, the dc signal $I(0)$ comprises two components: (a) $I^{(p)}(0)$ deriving from light that has remained polarised and (b) $I^{(u)}(0)$ deriving from light that has become unpolarised by collisions within the turbid medium. The depolarisation parameter

$$\beta \equiv \frac{I^{(p)}(0)}{I^{(u)}(0) + I^{(p)}(0)} \propto I(2f_m),$$

which quantifies the residual degree of polarisation, can be shown to be proportional to the photocurrent at the first harmonic of the modulation frequency. In the first set of experiments, my colleague Wayne Strange and I examined the intensity and depolarisation of green light ($\lambda = 544$ nm) from a He-Ne laser scattered laterally (90° to the incident beam) as a function of particle concentration N for microspheres of different diameters $2a$, ranging from 0.125 μm to about 1 μm. In terms of a dimensionless size parameter $\sigma = 2\pi a/\lambda$, the range spanned

$\sigma = 0.72$ (Rayleigh-Gans scattering[9]) to $\sigma = 5.8$ (Mie scattering). The intensity plot for particles of different size show the same general form: linear growth when the concentration is low and single scattering prevails, followed by a sharp rise to peak value with the onset of multiple scattering, and subsequent fall-off as the medium becomes increasingly opaque. The linearity is not easily discerned in the top panel for which the concentration is plotted on a logarithmic scale, but is quite obvious in the lower panels where the abscissa is the dimensionless quantity δ^*, proportional to concentration, referred to as the transport optical thickness defined by the relation



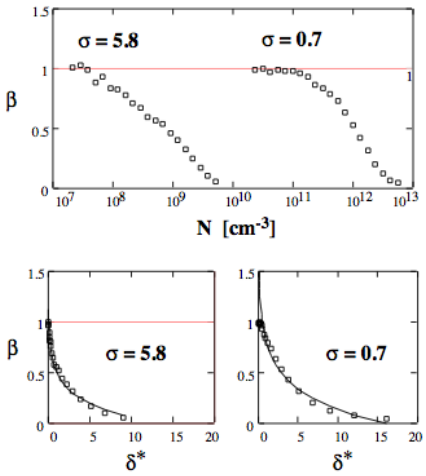
$$\delta^* = \frac{\text{sample length}}{\text{sample mean free path}} \times (1 - \text{asymmetry parameter}).$$

The sample mean free path depends inversely on the concentration, the geometric cross section of the spherical particle, and a numerical factor calculable from Mie theory. The asymmetry parameter $g \equiv \langle \cos\theta \rangle$, also determined from Mie theory, is the mean cosine of the scattering angle and ranges from 0 for Rayleigh scattering to 1 in the limit of geometric optics.

Although the range of concentration spans six powers of ten, each set of data, when normalised by the maximum intensity and plotted against δ^*, can be represented closely by a universal curve of the form

$$\frac{I(\delta^*)}{I_{\max}} = \frac{1}{1 + \bar{p}\delta^*} - e^{-\delta^*}$$

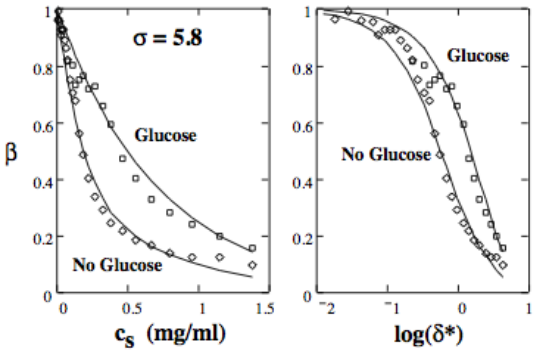
where the single parameter $\bar{p}$ depends on particle radius and can be interpreted as a measure of the probability that light incident upon a differential layer within the sample will be scattered in the opposite direction. The plots show that, irrespective of particle size, peak multiple scattering intensity occurs at $\delta^* \sim 5$.



Measurements of the depolarisation parameter β showed that (a) the scattered light remained nearly 100% polarised for low particle concentrations where single scattering dominates, (b) it took many more small particles to depolarise light to a certain extent than large particles, and (c) with the onset of multiple scattering, the decrease in β was nearly linearly proportional to the logarithm of concentration. As in the case of light intensity, plots of β conformed well to a general function of δ^*, revealing that $\delta^* \sim 10$ was an approximate upper limit at which laterally scattered light

[9] In Rayleigh scattering from a sphere of refractive index n, the particle size a is small compared to the external wavelength λ and internal wavelength λ / n, whereas for Rayleigh-Gans scattering a is small compared to $\lambda/(n-1)$, a less restrictive condition.

retained a degree of polarisation measurable with the PEM. The significance of this result is that the mean number of *depolarising* collisions was about the same for all particles irrespective of size.

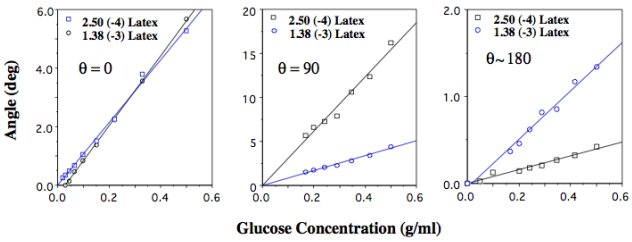


With the knowledge gained from studying light scattering in an achiral medium, I undertook to investigate the effect of light scattering in a chiral medium.[10]. The left panel shows a plot of the depolarisation parameter as a function of particle concentration for backscattering in a turbid fluid containing relatively large microspheres suspended in either pure water or a glucose solution (0.5 g/ml). Most striking is the preservation of light polarisation in the chiral fluid. The solid lines are fits to a diffusion model whereby the depolarisation parameter for linear (LP) or circular (CP) polarisation decreases exponentially with scatterer concentration or, equivalently, transport optical density—i.e. $\beta_X \propto \exp\left(-q_X \delta^*\right)$ where q_X is the depolarisation rate for X = LP or CP. Fits of the model to the data yielded q_{LP}/q_{CP} = 3.22. As shown in the right panel the model reproduced both the apparent linear variation of β over the domain of multiple scattering and the gradual approach to unity in the region of single scattering.

In retrospect, the result that linearly polarised light is depolarised after fewer scattering events than circularly polarised light should perhaps not be too surprising. All collisions that alter the azimuthal orientation of the electromagnetic field in the plane normal to the wave vector contribute to depolarising LP light, whereas only those collisions that reverse the sense of field rotation within the plane depolarise CP light. The light transmitted by the PEM in these experiments was in general elliptically polarised, i.e. contained a superposition of LP and CP states. In an optically isotropic, achiral medium like pure water, both LP and CP states propagate with the same phase velocity. However, in an isotropic chiral medium like glucose solution, the optical eigenstates must have definite helicity. Therefore LP waves decompose into independently propagating LCP and RCP waves, as first proposed by Fresnel. In this way, a chiral turbid medium acts to preserve the polarisation of light relative to an achiral turbid medium.

The effective preservation of light polarisation together with the capacity of a PEM to select those photons that remained polarised suggested to me that one should be able to observe chiral asymmetry by diffusive light scattering. This, indeed, was the case. The suite of plots below shows, respectively from left to right, the variation in optical rotation with glucose concentration for forward, lateral, and backscattered light from samples containing different concentrations of 1 μm diameter latex spheres ($\sigma = 5.8$). The optical rotation, which could be measured to a precision of 0.02° by an angular micrometer, was obtained from the first harmonic of the modulated light in the experimental configuration previously shown. (Details of the measurement procedure are given in *Waves and Grains* and the references therein to the original publications.)

[10] M. P. Silverman and W. Strange, "Light Scattering from Turbid Optically Active and Inactive Media", *Optics and Imaging in the Information Age* (Society for Imaging Science and Technology, Springfield VA, 1997) 173-180.

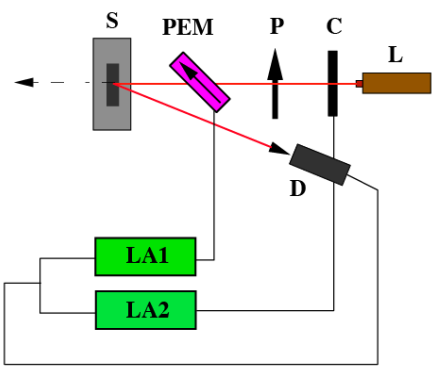


In the case of forward scattered light (detected within a 10 mrad cone of acceptance centred on the incident beam), the variation in optical rotation with glucose was weakly dependent on particle concentration and larger by ~1.4 than the corresponding rotation rate in the absence of light scattering. Such a modest increase was consistent with Mie theory, which predicted in this case of $\sigma > 1$ that the light scattering took place predominantly in the forward direction. Thus, the optical pathlengths of multiply and singly scattered light did not differ much along the incident direction.

The optical rotary power (degrees/glucose concentration) of laterally scattered light, in which no unscattered photons were received, was greater by a factor ~3.4 than the corresponding rate in *absence* of all scatterers. Thus, in marked contrast to the prevailing expectation, light scattering of the probe beam not only did not destroy the chiral signal, but appeared to enhance it. My interpretation was that multiple scattering in the glucose solution had in effect lengthened the effective optical pathlength through the chiral medium. This interpretation was supported by the visual appearance of the cuvette which was brightly illuminated throughout the less turbid sample (which had the higher optical rotation), but exhibited radiance only near the point of entry of the more turbid sample (which had the lower optical rotation).

The case of backscattering was of particular interest, since that was (and is) the geometry of choice for materials analysis and biomedical diagnostics. One might have again anticipated that no optical rotation would be observed (since light reflected at normal incidence carries no imprint of the chirality of the medium), but this is not so for Mie scattering. However, in contrast to lateral scattering, it was the suspension of greater turbidity that manifested the larger rotary power. As a consequence of predominant forward scattering, most photons escaped undetected into the cell interior. Nevertheless, because of the greater number of scattering events in the more turbid medium, those photons that did return nearly antiparallel to the incident direction incurred on average a longer optical pathlength. The larger rotary power therefore signified a greater entrapment of light.

In short, the important lesson of these chiral studies was that diffusively scattered light carried information about both the chiral structure of the ambient medium and the size and density of the scattering particles. Despite the fact that the suspensions were highly turbid, the use of the PEM to select photons whose states of polarisation were preferentially preserved, made it possible, in a manner of speaking, to 'see' inside the medium.



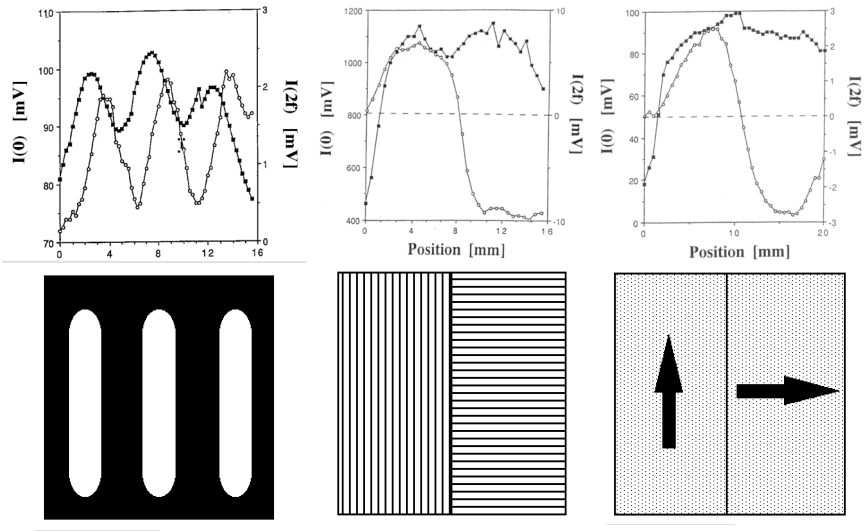
I inquired next whether it was literally possible to see inside a highly turbid medium—i.e. whether selective detection of light that has retained its phase modulation would preserve imaging information carried by its wave vector. To convert the PEM to an imaging device, polariser P2 in the previously shown configuration was removed. A small target was placed in the midplane of the scattering cell containing an aqueous suspension of latex microspheres, and the phase-modulated light of a He-Ne probe beam at 544 nm was scanned across

the flat front surface of the cell. Light reflected into the detector was then analysed by lock-in amplifiers for the photocurrent components:

- $I(0)$ — the dc signal corresponds to what a polarisation insensitive detector would see.
- $I(f_m)$ —the fundamental is proportional to the circular intensity difference $\frac{R_{LCP} - R_{RCP}}{R_{LCP} + R_{RCP}}$.
- $I(2f_m)$ —the first harmonic is proportional to the linear intensity difference $\frac{R_s - R_p}{R_s + R_p}$.

The objects hidden in the scattering cell were fabricated to test the sensitivity of the imaging procedure to different optical or topological features. Among the objects I used, were a

- Slotted Target—an absorbing slab with slotted apertures of width 2 mm and separation 3 mm on a reflective metal base;
    - Distinguishing feature: edges
- Grooved Target—a metal slab with two contiguous sections of orthogonally oriented rulings of width ≤ 1 μm and spacing 50 μm;
    - Distinguishing feature: surface texture
- Laminar Target—two linearly polarising smooth plastic strips mounted side by side with transmission axes at right angles;
    - Distinguishing feature: polarisation-selective absorption.



The outcomes, shown in the three panels for the slotted, grooved, and laminar targets from left to right respectively, were quite informative and are discussed in detail in *Waves and Grains*. Though the scattering cell appeared to the naked eye as a murky white suspension reminiscent of whole milk and revealed no evidence of the presence of the target under ambient illumination, the phase-preserving reflected light captured the essential features in each case.

For the slotted target, the peaks of the signal $I(0)$, with contrast or visibility[11] of about 14%, corresponded to the centers of the slots, where reflectance from the metal base was greatest. Interestingly, the signals $I(f_m)$ and $I(2f_m)$ attained maximum values (both positive and negative) near the steepest slopes of the $I(0)$ scan, thereby revealing a sensitivity to edges. In fact, from the signs and magnitudes of the signals, one could tell that the edges were inclined slightly to the vertical. The optical density of the medium (mean number of scattering events per photon within the medium) was about 125; the transport optical density was about 9. Yet the signal $I(2f_m)$ rendered the surface topography with a contrast of about 63%. Recall that the visibility of naked-eye viewing was 0.

The dc signal obtained from scanning the backscattered light from the grooved target showed virtually no structure apart from a slight dip at the junction of the two segments. However, the $I(2f_m)$ scan revealed a step function with positive plateau on the left (signifying a predominance of s-polarised light) and a negative plateau on the right (signifying a predominance of p-polarised light) exactly as predicted.[12] No circular intensity difference was expected, and consistent with theory no signal $I(f_m)$ was observed.

The laminar target gave results very similar to those of the grooved target. The dc signal was essentially flat across the target and revealed no information, whereas the variation in $I(2f_m)$ corresponded to preferential s-polarisation on the left and p-polarisation on the right. The interaction with the target, however, which involved selective absorption of light transmitted twice (forward and backward) through the lamina, was completely different.

Overall, these investigations of turbid media demonstrated that with phase modulation and phase-sensitive detection of light, one can recover important physical, chemical, and optical information that prevailing wisdom regarded as having been irretrievably lost by multiple scattering. With proof of principle demonstrated, there are numerous objectives for wanting to 'see' through turbid media such as underwater photography, navigation through fog, environmental monitoring in foggy or dusty atmospheres, medical imaging (e.g. of the retina through a cataract), and chemical identification and analysis (e.g. blood glucose measured via optical activity through thin skin or the eye).

5. RADIOACTIVE MATTER

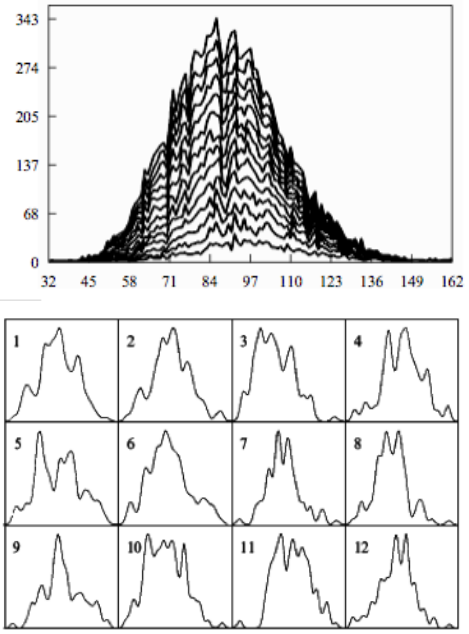
One of the most bizarre controversies that ever came to my attention involved the claims—repeated in peer-reviewed publications for well over a decade—of the existence of a 'cosmogenic force' responsible for correlating what to all appearances would be considered random fluctuations of independent stochastic processes. In the investigation of this controversy, the element sodium was to play a crucial role as both the sample of interest, in the form of the

[11] The contrast or visibility is defined as the difference between the maximum and minimum values of a signal divided by the sum.

[12] One expects the reflected light to be polarised perpendicular to the orientation of the grooves. The electric field of the light drives currents along the grooves which generate a reflected wave maximally in the plane normal to the axis of the radiating dipole. The imaging experiment with grooves comparable in size to the optical wavelength replicates at micron scales an experiment often performed in undergraduate physics labs whereby microwaves are reflected from a grating with spacing of a few centimetres.

radioactive nuclide Na-22, a beta plus emitter, and as the major component of the detector of gamma rays, which was a thallium activated sodium iodide [NaI(Tl)] scintillator.

Since the 1990's, one of my interests was to examine quantum processes for evidence of nonrandom behaviour. Although the principles of quantum mechanics were worked out in the 1920s and the quantum predictions of one kind or another had been subjected to numerous experimental tests, there were, surprisingly few direct examinations of what is arguably the most characteristic feature of the entire theory—the prediction that individual quantum events are entirely random and unpredictable. By devising different statistical tests, I examined, together with my colleague Wayne, various nuclear processes (alpha, beta, and electron-capture decay) for statistically significant indications of a deterministic cause underlying the randomness. We did not find any; the processes passed all the tests of random behaviour.



Having established a certain expertise in an area that, at least at *that* time attracted little interest or controversy, I subsequently received some unusual emails from physicists pointing out the strange claims of a group of Russian researchers[13] and requesting that I look into the matter. Eventually, I found the time to do so, and the endeavour led me and Wayne to examine the statistics of nuclear decay more deeply than we had done before. The authors' conclusions, if valid, would have necessitated a radical revision of the laws of physics as we understand them. I must leave the details of their investigations to the literature, but suffice it to say that when my attention was first drawn to their work they claimed that (a) 'macroscopic fluctuations' exhibited 'discrete states', and that (b) fluctuations in the time series of decays of radioactive nuclei were correlated and recurred with regularity. Both claims were based on the visual appearance of histograms (which, in fact, they had smoothed to give less blocky shapes). The articulations in the first plot above—which is actually a superposition of numerous histograms comprising time series that overlapped—was taken as evidence of the universal 'discrete states'. The reappearance of similar shapes, like the 'rabbit ears' in the time sequence of histograms in the second plot above, was taken as evidence of recurring, correlated fluctuations. Both plots of histograms were made from time series of measurements of the alpha emitter Pu-239.

I must note at the outset that the 'shape' of a histogram is an ill-defined concept dependent on various arbitrary choices of construction and not an invariant statistical property. It can take widely differing forms for a given set of events depending on the number and widths of the categories into which events are assigned. I discuss this further in an article[14] prompted by this controversy and a book currently in preparation[15]. Moreover, there is a branch of mathematics (Ramsey theory) that virtually guarantees that any sought-for pattern can be found in

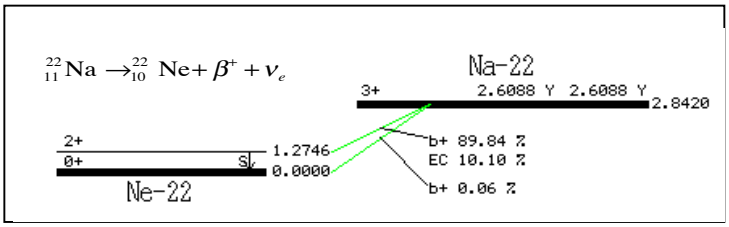
[13] S. Schnoll et al, *Phys. Usp*. **41** (1998) 1025.

[14] M. P. Silverman and W. Strange, "Search for correlated fluctuations in the β^+ decay of Na-22", *Europhysics Letters* **87** (2009) 32001 (1-6).

[15] M. P. Silverman, *A Certain Uncertainty: Nature's Random Ways* (to be published by Cambridge U.P.).

the distribution of a sufficiently large set of points. Only rigorous statistical analysis can reveal whether time series and frequency distributions actually manifest correlations.

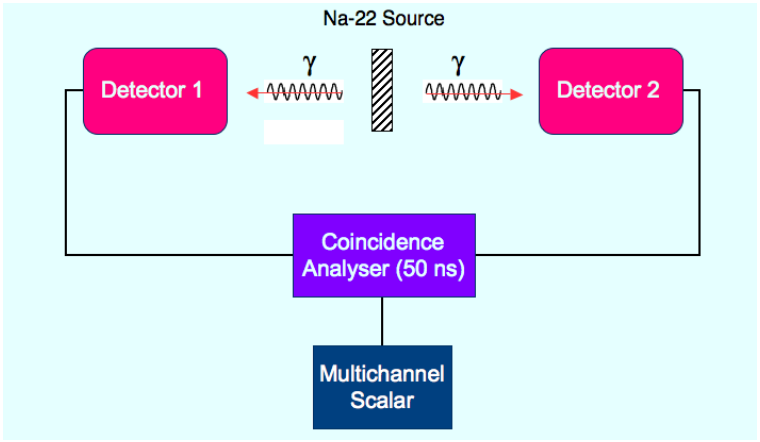
To test for randomness and correlations of nuclear decay, we performed an experiment employing the transmutation of sodium-22 (Na-22) to neon-22 (Ne-22) through positron emission:



+ $e^+ + e^- \rightarrow \gamma + \gamma$

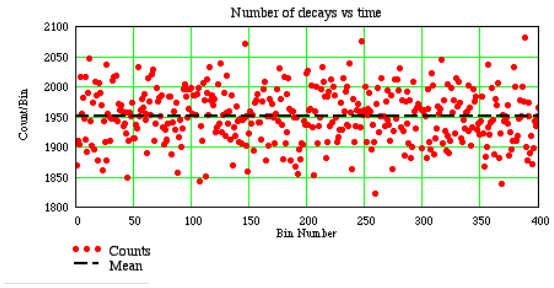
The emitted positron combined with an ambient electron to annihilate into back-to-back gamma rays which were detected in coincidence in a configuration of apparatus like that below. An electron neutrino, which was also emitted, interacts so weakly with all matter, that it escaped entirely and played no role in the experiment.

Na-22 half-life	**2.6 y**
Initial activity	**1 µci**
Gamma energy	**511 keV**
Coincidence window	**0.44 s**
Intervals	**8192 bins = 1 h**
Duration	**167 h = 7 d**
Mean count rate	**442 s^{-1}**
Background rate	**0.021 s^{-1}**

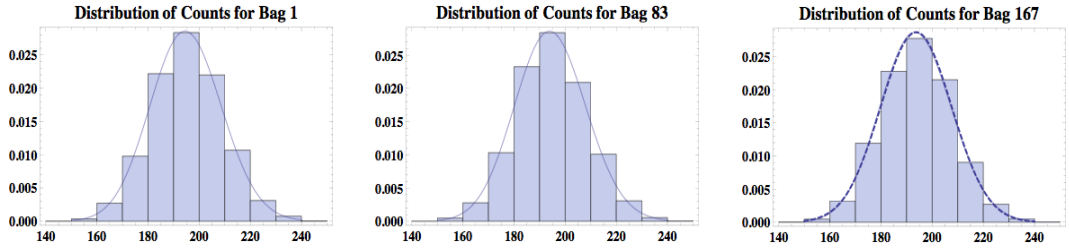


The decay of Na-22 was chosen for several reasons. First, the process should be governed by Poisson statistics; thus the parent probability function was known and all other pertinent statistical quantities could be determined analytically. Second, the transmutation was an example of a weak nuclear interaction with long half-life, so the time series of decays over the period of the experiment was very nearly stationary. Third, the decay yielded a stable nuclide of neon and a single outgoing positron which shortly afterward produced two counter-propagating 511 keV gammas. The simplicity of the final state together with spatial correlation and narrow energy uncertainty of the gammas permitted coincidence measurements with very low noise background and high signal-to-noise ratio.

The decay of Na-22 was monitored for 167 hours straight, whereby the number of decays within one coincidence window of 0.44 s constituted 1 bin of data. 8192 bins—a duration of about 1 hour—comprised 1 'bag' of data. Overall, the experiment led to more than 1 million bins of data. The scatter plot shows an example of 400 Bins of raw data (where 1 Bin = 10 bins). To confirm that the statistics remained very nearly stationary throughout



the experiment, histograms of counts were made of each bag. The 3 histograms shown represent fits of the Poisson probability function to data collected at the beginning, middle, and end of the data collection period.



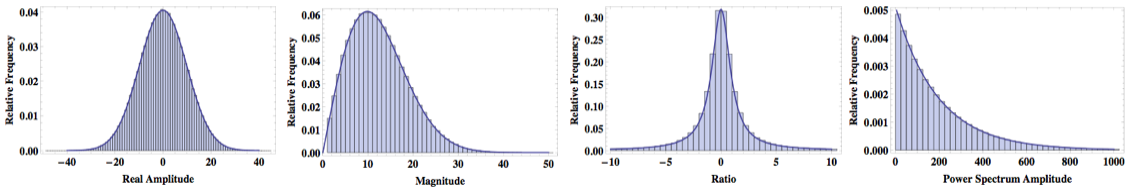
With account taken of the natural lifetime of Na-22, theoretical fits matched the histograms visually closely and passed chi-square tests with expected *P*-values.

There is more to the statistics of nuclear decay, however, than just the Poisson distribution. In a comprehensive set of tests, we examined the distributions of the Fourier amplitudes of the decay time series, the moduli and phases of these amplitudes and the power spectrum. Each derived quantity was characterised by a different theoretical distribution which the data must satisfy if the decay of Na-22 nuclei occurs randomly and independently (our 'null hypothesis'). From left to right, the suite of histograms below shows the experimentally observed distributions of

- Real part of Fourier amplitude (Gaussian distribution)
- Modulus of Fourier amplitude (Rayleigh distribution)
- Ratio of imaginary part to real part of the Fourier amplitude (Cauchy distribution)
- Power spectrum of the Fourier amplitudes (Exponential distribution)

and the match to the corresponding theoretical density function, which is the solid line forming the envelope of each histogram.

Gaussian Rayleigh Cauchy Exponential

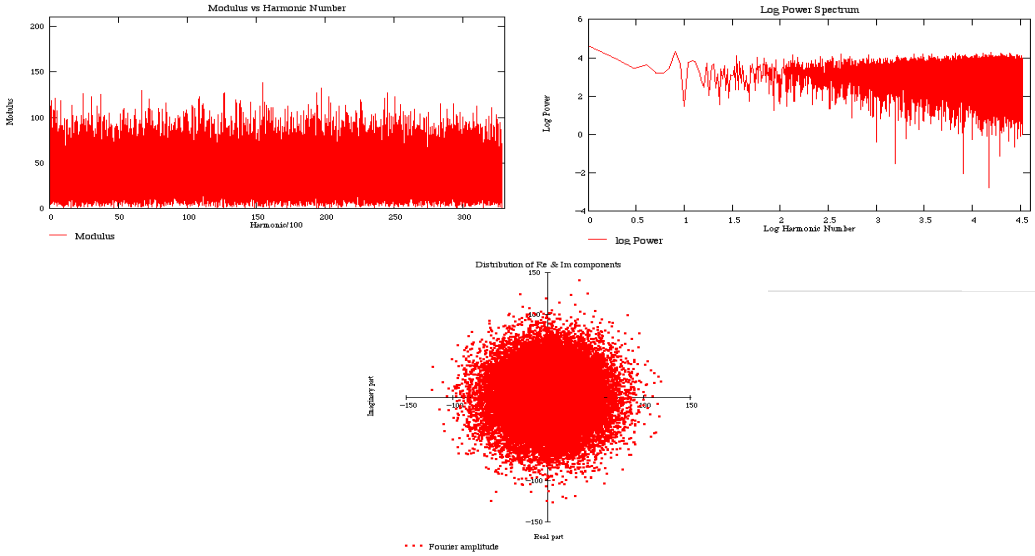


These are the actual data and not computer simulations. At the scale of the plots, virtually no fluctuations are visible, a consequence of the long length of the time series of gamma coincidences and presumably the validity of the null hypothesis. It is to be noted that all the distributions are determined by one and the same experimental quantity—mean number of counts per bin—which is fixed at the outset of the experiment. There are no other adjustable parameters.

The Fourier spectrum of the gamma coincidences can be viewed in several visually informative ways. Left to right and then down, are shown scatter plots of

- modulus vs harmonic number

- log of the power spectrum vs log of the harmonic number
- imaginary part vs real part



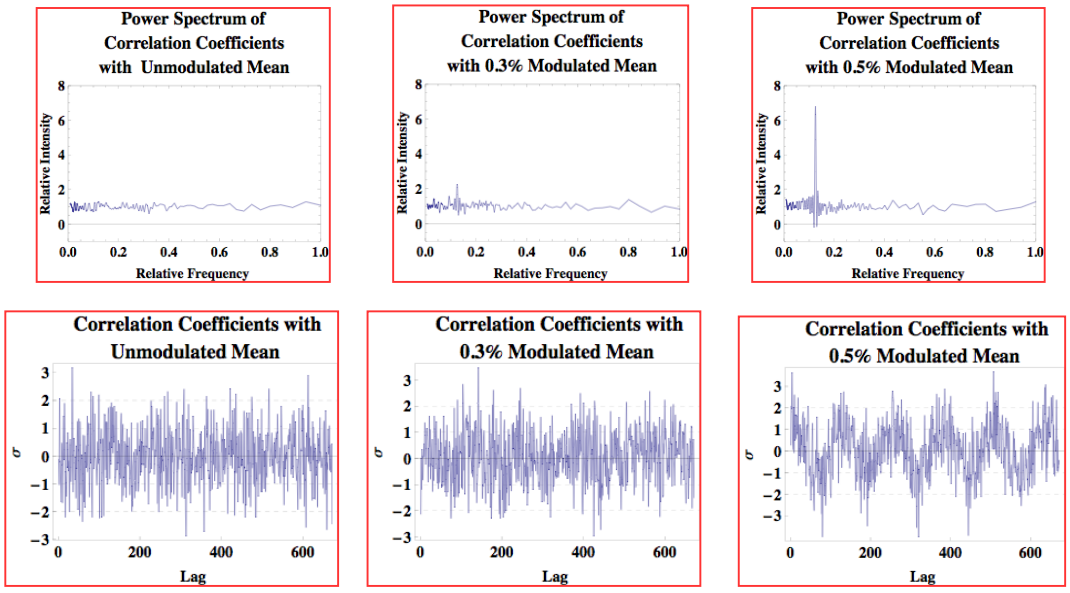
A characteristic visual feature of the first panel is the large number of high peaks scattered throughout the spectral range. However, this is to be expected if the null hypothesis is valid. For quantities subject to a Rayleigh distribution, the fluctuation (as gauged by the standard deviation) is about one-half the mean value. Statistical tests of the largest peaks show them to be of no significance. The fluctuation in peaks is even greater in a display (not shown) of the power spectrum—i.e. the squares of the moduli—because the standard deviation of an exponentially distributed random variable is equal to the mean. (Recall that for a Poisson distribution, the *variance* is equal to the mean.) For anyone used to working primarily with Gaussian distributions, the intuitive assessment of the probability of occurrence of events distributed differently may require a radical recalibration of thinking. For example, whereas the probability is about 0.3% that a normal random variable falls outside ± 3 standard deviations about the mean, the probability is 77.4% that the *largest* observed peak in a nuclear decay power spectrum of $2^{16} = 65{,}536$ peaks is at least *ten* times the mean peak.

A characteristic visual feature of the second panel is the increasing (with harmonic number) triangular wedge shape of the pattern and apparent zero slope of a line of regression through the bulk of the power spectral amplitudes. This is also to be expected if the null hypothesis is valid and the plot represents a power spectrum of white noise.

Finally, a characteristic visual feature of the third panel is the isotropic symmetry with density of points falling off radially—also to be expected if the null hypothesis is valid.

The power spectrum and correlation function of a time series of random data provide two means of discerning an underlying periodicity or regularity. Examination of the gamma coincidence data revealed no such structure with periods shorter than the duration of the experiment. The first phase of our investigation had shown that the decay of Na-22 followed very closely a Poisson distribution. Any external periodic influence on the decay, if present, must

therefore show up as a periodic variability of the mean decay count—since a Poisson distribution is uniquely determined by a single parameter, the mean. We could place an approximate limit on the sensitivity of the data to reveal a periodic component by simulating a time series with Poisson random number generator of time-varying mean: $\mu(t) = \mu_0\left(1+\beta\cos\left(2\pi t/T\right)\right)$. The figures below show the progressive change in the power spectrum (upper panels) and autocorrelation (lower panels) as the hypothetical harmonic amplitude β takes on the sequential values 0, 0.003, and 0.005 for a period T less than the duration of the experiment (i.e. time series) T_{exp}. At a threshold value $\beta \le 0.3\%$, the highest power ordinate passes statistical tests for randomness, and the harmonic variation in the autocorrelation merges with the noise. Thus, if a harmonic with larger amplitude β were present in the time series, it would have been revealed by analysis even though visual inspection of the sequence of 167 histograms (1 for each bag of data) would show no statistically significant recurrences (such as the authors of the controversial papers claimed).



A time series of duration T_{exp} does not permit one to measure a period $T > T_{\text{exp}}$. However, a partial-period component, if present, would be equivalent to a trend in the data and thereby lead to low-frequency oscillations in the power spectrum of the autocorrelation. Apart from a weak trend due to natural lifetime which we accounted for, the gamma coincidence time series revealed no other measurable trend. By resorting again to computer simulations of the autocorrelation function for partial-period components of various amplitudes, we could conclude that our experiment would have revealed a trend resulting from an external influence of period up to about $5T_{\text{exp}}$, or approximately 35 days.

Time does not permit full presentation of the statistical tests performed on the Na-22 decay data., so I will simply summarise the principal findings.

- The 'discrete states' in the histograms of Pu-239 decay were *entirely* an artifact of the mode of data presentation. The patterns could be produced by use of a Poisson random number generator, which showed beyond doubt that they arose from a mathematical

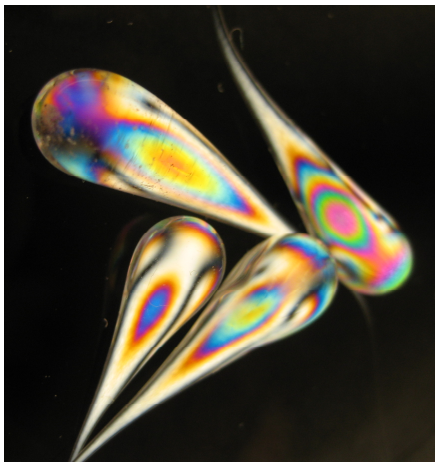
algorithm and had nothing whatever to do with correlated fluctuations of any real physical force. In fact, by creating superpositions of histograms with random number generators of different probability distributions, I could create some rather amusing artwork.

- Visual inspection of the shapes of histograms provide no reliable test of correlations of stochastic processes.
- The statistical tests we performed on the β^+ decay of Na-22 showed
    - complete consistency with white noise
    - no correlations in fluctuations in nuclear counts
    - no correlations in fluctuations of histogram frequencies
    - no periodicity in nuclear counts or count frequencies for time intervals under 167 hours, or unexplained trends with a time period under 35 days.

The authors of the controversial papers claimed to have observed visually the periodic recurrence of histogram shapes within the time periods we investigated. Had the count frequencies of our experiment been correlated, our statistical tests would have revealed this feature even under conditions where visual inspection of histograms could not.

In closing this section, I would remark that during the past few years a controversy at least as strange has arisen, whereby the intrinsic rates of certain radioactive decay processes are claimed to be correlated with solar activity. As one manifestation of this, the influenced decay rates were claimed to depend on the orbital position of the Earth about the Sun. To my knowledge no current experiment has actually produced such results. Rather, the inferences were drawn from data collected many years ago for other purposes. I know of no convincing way to explain such effects within the scope of the laws of physics as we currently understand them. Such claims can be validated only by reproducible experiments under conditions where instrumental artifacts due to the numerous oscillatory modes of the Sun are demonstrably absent or accounted for. I have no opinion on the matter at present, but am investigating the phenomenon further.

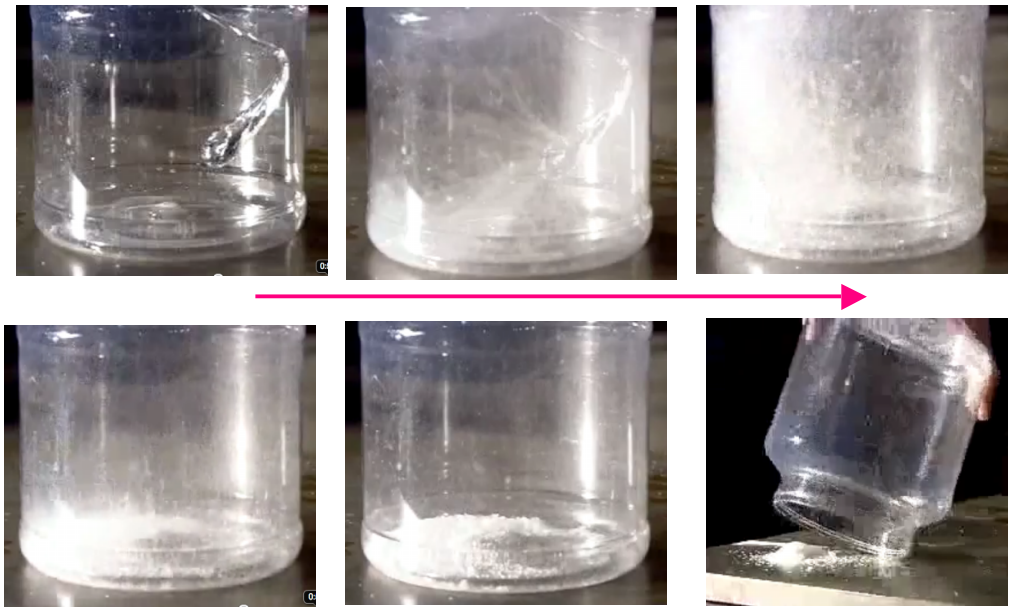
6. RUPERT'S DROPS (EXPLODING GLASS)



Over the years I have worked on a number of projects that began primarily as a source of amusement, but later turned out to be highly instructive and of more general interest to others besides myself. Some examples of these kinds of projects were (a) the 'vortex tube', which, in apparent (but not actual) violation of the 2nd Law of Thermodynamics, transformed an in-flow of room-temperature compressed air into oppositely directed out-flows of cold air and hot air by means of a mechanism *with no moving parts* (apart from the air); (b) the 'voice of the dragon', which was a corrugated tube swung in a circle over one's head to produce a wide range of musical sounds; (c) the 'world's simplest motor', which consisted only of a AA battery and a bent paper clip, as well as others. My most recent example, Rupert's glass drops, falls into this category.

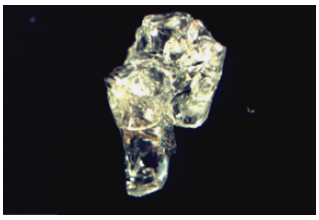
The drops are made by heating the end of a glass rod—I have used soda lime glass—clamped horizontally over a torch and letting the melted droplets fall into a bucket of cold water. As the droplet begins to melt, gravity pulls it into a teardrop shape with long tail which eventually breaks and releases the droplet. Simple as these instructions may sound, there is a finesse to the process that must be discovered or else the droplets may shatter in the water. The photograph above shows examples of soda-lime drops between crossed-polarisers. The colours signify a substantial stress-induced birefringence (a topic that I discuss at length in *Waves & Grains*) because the outside layer is under great compression and the interior of the drop under great tension by virtue of the differential rate of cooling.

The stressed drop has a remarkable property known for more than 300 years and quite possibly since the time of ancient Rome if not before then. So hard is the head of the drop that one can squeeze it tightly with a pliers or hit it with a hammer without damaging it—yet a simple snap of the thin fragile tail will cause the initially transparent glass drop to explode into a white powder of small particles. It is startling and fascinating to behold—and never fails to attract a student's attention however much his or her sense of wonder may have been dulled. Samples of the drops (and presumably the secret to making them) were said to have been brought out of Germany by Prince Rupert of the Rhine sometime during the late 17th Century and given to King Charles II of England who sent 5 of them to the Royal Society for study. My understanding from diverse readings is that Charles would offer the glass drops to members of his court in such a way that the tails were snapped and the drops exploded in their hands.



The progression of frames (screenshots from a video made by the Corning Museum of Glass) from left to right in each row shows the chronological transformation from a glass drop, whose tail is snapped, to a mound of powder.

In the ensuing centuries, many people have made and exploded these drops for amusement or even as physics demonstrations, but I am aware of only a few serious scientific investigations of their properties. My interest in the drops, once I learned of their existence, was stimulated by two questions—one immediate and obvious and the other emergent and more profound.

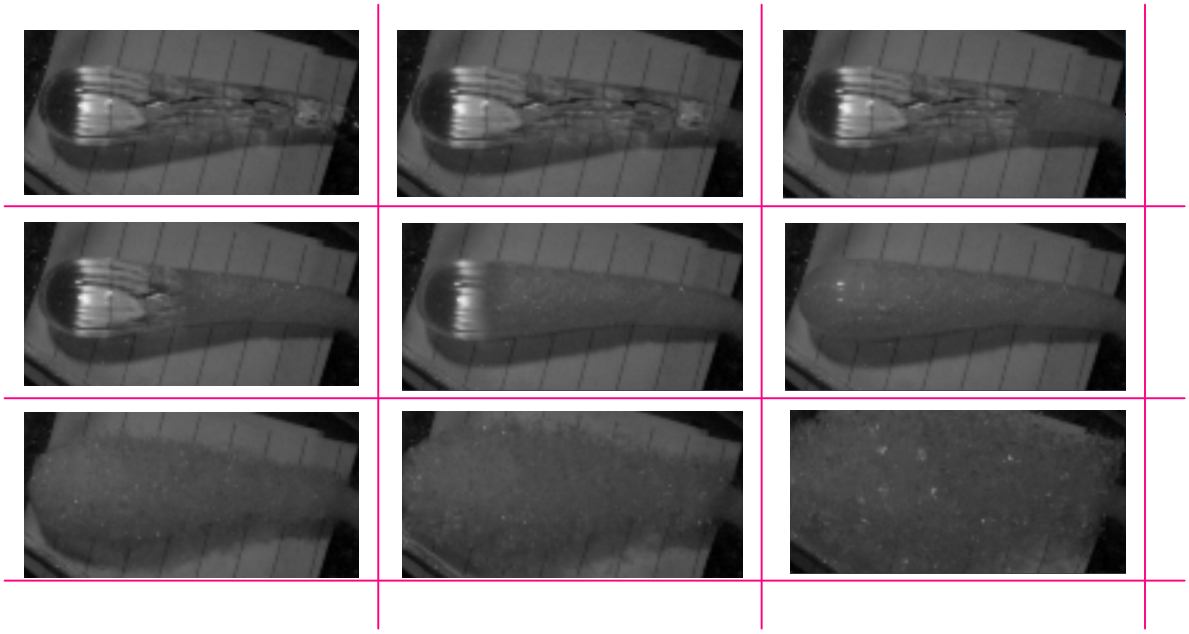


The obvious question was simply: *Why* does a Rupert's drop explode? Superficially, the answer was also simple: to relieve the stresses by going to a state of lower internal energy. Indeed, microscopic observation of the glass fragments between crossed polarisers showed the absence of stress-induced birefringence after the explosion. However, that answer is not really satisfactory. Rupert's drops are a form of tempered glass—i.e. glass, such as one finds in the side windows of motor cars, that has been toughened by thermal treatment. Tempered glass does not explode when subject to shock, but breaks into myriad small pieces such as shown in this sequence of screenshots taken from a YouTube video.

http://www.youtube.com/watch?v=eqV5W76U8Qg

The second, more profound question, is *how* does a drop explode—i.e. by what mechanism? In one of the few published reports I have seen, the explosion of a drop had been filmed with a high-speed camera. From microscopic examination of frames taken during progression of the shock wave from tail to head, the author concluded that fragmentation proceeded by crack bifurcation. I could not discern that myself from the dark and poorly resolved photographs in my copy of the paper. A similar high-speed photographic study, unpublished to my knowledge, was made by Dr Steven DeMartino at the Corning Glass Corporation, who shared his results with me. Using a camera with a frame speed reported to be 0.94 miles/second (1,513 m/s), Dr DeMartino photographed the advance of the shock front across a lead-crystal drop whose length I estimated (from a calibrated scale in accompanying photographs) to be about 1 inch (2.5 cm). The drop was superposed over a card calibrated in units of about 1/8 inch, again estimated. In the sequence of frames below, which progresses chronologically from left to right and top to bottom, one can see clearly the advance of the shock front behind which the glass is fragmented and opaque (due to multiple light scattering). From the frame rate and estimated calibration spacing, I deduced that the shock front propagated at a speed of about 1,130-1,140 m/s. For comparison, the speed of sound in lead-crystal glass falls in the range of 3,400-4,200 m/s.

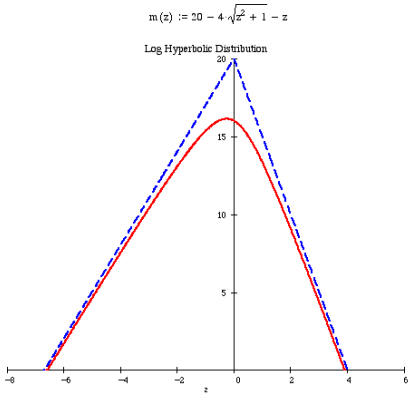
It is to be noted that only after the shock wave propagated entirely through the drop did fragments of glass begin to disperse outward from the surface, and the shape of the drop become increasingly amorphous. The Corning video continued well beyond the last frame shown (lowest right photo). By counting the number of frames that passed between the time the shock wave reached the tip of the head and the time when the surface of scattered glass fragments moved outward a distance equal to the calibration length (interval between two calibration marks), I could estimate the velocity of the dispersing fragments to be about 15-25 m/s. The delay in the onset of dispersion (i.e. the explosion) of the outer surface reflects the greater speed of propagation of the shock wave through the tensile interior compared to that through the compressed surface layer.



Despite the exceptional clarity of the Corning video, magnification of the frames did not reveal the fragmentation pattern at a microscopic level, and therefore neither confirmed nor refuted the mechanism of crack bifurcation. Shortly after I became interested in the behaviour of Rupert's drops, which was several years before contact with the Corning Glass Company, it occurred to me that a statistical study of the glass fragments could provide useful, perhaps definitive, information regarding the explosion mechanism. The research was begun with an undergraduate whose senior project was to make a sufficient number of drops for a preliminary statistical analysis. The work proceeded slowly and only a few drops were made, and the statistical results, though crude, suggested a power law distribution of particle sizes. Another generation of undergraduates contributed a few more drops. Finally, in order to expedite the work, I made a request to the Corning Museum of Glass (CMOG) if they could provide me a sample (that would ordinarily have been discarded as waste) of residual glass fragments from their public demonstrations of exploding Rupert's drops. I was informed that for safety reasons CMOG did not do these demonstrations anymore, but my request eventually brought me into correspondence with Dr DeMartino who generously sent about 500 g of fragments of lead-crystal glass from exploded Rupert's drops. (To the amusement of both of us, he had initially thought I wanted fragments from the explosion of a single 500 g drop - which, presuming such a drop could even be made, would have generated quite an explosion.)

The fragments were sieved and the mass density $m(s)$ as a function of size *s*—that is, the mass of fragments falling within the interval *ds* at *s*—was determined to follow a log hyperbolic density of the form

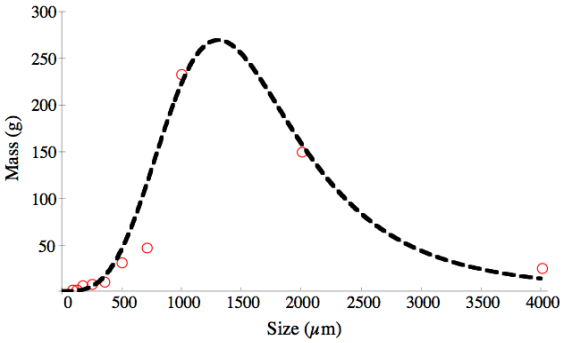
$$m\big(z(s)\big) \sim Ae^{-\alpha\delta\sqrt{z^2+1}+\beta\delta z} \quad \text{where} \quad z(s)=\frac{\ln s-\mu}{\delta}.$$

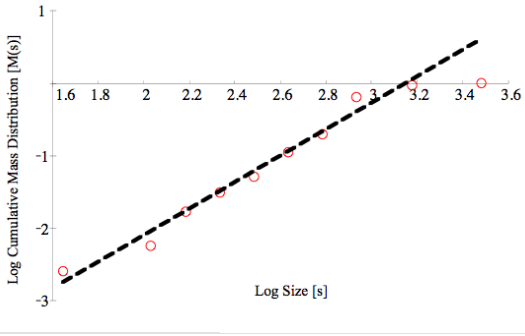


The adjustable parameters are interpretable as follows: (a) *A* is a normalisation factor, (b) μ is a location parameter, (c) δ is a scale factor, and (d) α and β together determine the left and right asymptotic slopes of the hyperbola through the relations $\varphi_{\pm}=\pm(\alpha\pm\beta)\delta$. The name of the distribution follows from

the hyperbolic curve (as a function of z) in the exponent of $m(z)$, an example of which, together with the asymptotes, is shown above for $f(z)=20-4\sqrt{z^2+1}-z$ whose numerical parameters were chosen arbitrarily. The cumulative mass distribution (also referred to simply as the mass distribution) $M(s)=\int_0^s m(s')ds'$ is the total mass of particles of size s or smaller. The plot in the left frame below shows the variation in mass fraction with size upon which is superposed a visually fit log hyperbolic density. The right frame is a log-log plot of the corresponding cumulative mass distribution with size. Of particular interest is the slope of the linear portion because this slope is related to the so-called fractal dimension D of the sample.

To say that a distribution of particles is fractal signifies that there is no characteristic length scale—i.e. the fragmentation mechanism operates in the same way at all scales. This





behaviour can only be approximate because every finite sample of real particles has a smallest size and a largest size. Nevertheless, the scale-independent character of a process may be apparent over a significant range of sizes.

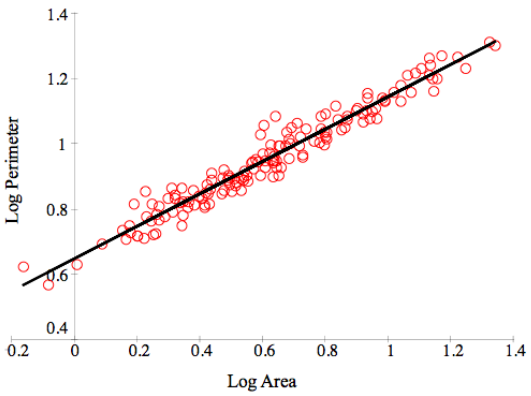
One defining characteristic of a fractal fragmentation process is that the resulting cumulative frequency distribution—the number of particles of size s or greater—follows a power law

$$N(s)=\int_s^\infty n(s')ds' \propto s^{-D}.$$

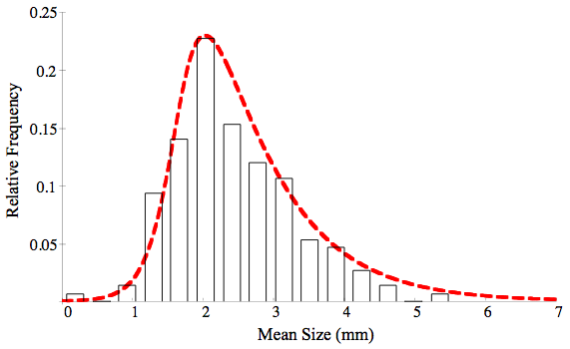
The particle density function $n(s)=-dN(s)/ds$ is the number of particles within the interval ds at s. The exponent D is called the fractal dimension, a statistical quantity that indicates the extent to which self-similar fractal objects fill a given space as their size is reduced. The mass density function $m(s)$ and particle density function $n(s)$ of a material with uniform mass per volume are related by $m(s) \propto s^d n(s)$ where d = 3 is the topological dimension of the Euclidian space in which the particles are embedded. It then follows that $n(s) \propto s^{-D-1}$ and $m(s) \propto s^{d-D-1}$, whereupon $M(s) \propto s^{d-D}$, and

$$\gamma \equiv d\log M(s)/d\log s = d-D.$$

The fractal dimension D can therefore be deduced from the slope γ of the linear portion of the plot of log cumulative mass distribution as a function of log size. Note that γ does not depend on either the units of mass and size or the base of the logarithm. My analysis of the lead-crystal Rupert's drops fragments led to a mass-based fractal dimension $D_M = 1.17 \pm 0.12$.



An alternative method, based on geometry rather than mass, of deducing the fractal dimension of a sample of fragments is to measure the perimeter P and area A of randomly chosen individual particles. The two geometrical properties are related by $P \propto A^{D/2}$, whereupon the fractal dimension is obtained from the slope $D_g = 2\,d\log P / d\log A$ of the least-squares line of regression to a scatter plot such as the one shown. The data for this plot were acquired by means of a stereomicroscope with digital camera and particle counting software that determined the length L, width W, mean size $(L+W)/2$, perimeter P, and area A of each fragment in a randomly chosen sample of 151 fragments. The fractal relation is not particularly sensitive to which linear measure of size is used. From the slope of the line of regression I obtained a fractal dimension $D_g = 0.99 \pm 0.07$. That the selection was indeed random is supported by a plot (histogram) of the relative frequency as a function of mean fragment size, the envelope of which is again reasonably well matched by a skew-symmetric log hyperbolic density function.



All told, I determined the fractal dimension of the Rupert's drops fragments three different ways (mass distribution, particle distribution, geometry), each method leading to a value statistically equivalent to unity. The unweighted mean of the three independent measurements was $D_{RD} = 1.06 \pm 0.09$.

Physicists have been studying the fragmentation of solids for many years and compiling data on their fractal dimensions. An abbreviated table (below) comprising a wide range of materials (rocks, coal, clays, etc.) and processes (abrasion, projectile impact, chemical explosions, etc.) shows a remarkable uniformity in the values of D. Although values of D in the table range from 1.89 to 3.54[16], the majority of values are in the vicinity of 2.5. (The mean of the tabulated values is 2.57.) For example, the fractal dimensions of broken coal, the chimney rubble above the PILEDRIVER nuclear explosion, and the fragments from high-velocity impact of a projectile on basalt are all near 2.5. The proximity of the fractal dimensions of widely differing materials and processes suggests a common element or universality to the various ways in which the fragments were produced. To this widely occurring value D = 2.5, must be contrasted the value $D_{RD} \sim 1.1$ of Rupert's drops. Rupert's drops differ from materials listed in the table (and many others for which the fractal dimensions have been measured) in that the vitreous material is metastable with a very high internal energy. I have calculated that the strain energy density within a Rupert's drop is about 0.24×10^6 Joule/m^3. To put this number in perspective, note that to create an equivalent strain energy density by application of an external force on initially unstressed glass would require a pressure of about 1,300 atmospheres.

[16] From the relation $M(s) \propto s^{d-D}$ one would infer that D must not exceed 3 and, indeed, reasons to doubt the accuracy of the last two values in the table have been published.

MATERIAL	FRACTAL DIMENSION D
Rupert's drops (lead-crystal glass)	$D_{Rd} = 1.06 \pm 0.09$
Crushed quartz	1.89
Disaggregated gneiss	2.13
Disaggregated granite	2.22
FLAT TOP (chemical explosion 0.2 kt)	2.42
Asteroids (theory)	2.48
PILEDRIVER (nuclear explosion 61 kt)	2.50
Broken coal	2.50
Interstellar grains	2.50
Projectile fragmentation of basalt	2.56
Sandy clays	2.61
Terrace sands and gravel	2.82
Glacial till	2.88
Stony meteorites	3.00
Ash and pumice	3.54

D L Turcotte, "Fractals in Geology and Geophysics", *Pure and Applied Geophysics* **131** (1989) 171-196

The utility of the fractal dimension, together with the observed mass and particle distributions, is that it provides an experimental measurement against which theoretical models of fragmentation can be tested. I have examined a variety of theoretical models for 'ordinary' brittle solids in equilibrium states, but none accounted for the observed distributions and fractal dimension of Rupert's drops fragments. There is something special about a stochastic process characterised by a power law $f(x) \propto x^{-D}$ with exponent $D = 1$. In the context of noise theory or signal processing, this law describes $1/f$ noise, where f is frequency. For reasons not fully understood, $1/f$ noise occurs ubiquitously in nature and elsewhere. First encountered in current fluctuations in vacuum tubes where it was called flicker noise, $1/f$ noise has since been found to describe the power spectrum of a wide range of extreme events like earthquakes, avalanches, and landslides, of biological processes involving pulsations in the heart and neuronal activity in the brain, of economic time series associated with the stock market, of intensity fluctuations in the recording of music and speech, and numerous other examples. Whether there is actually something universal about the process of an exploding Rupert's drop that links it to a particular power law remains to be found out.

In future experiments using photoelastic modulation for measuring stress-induced birefringence, I intend to study the compressive and tensile stresses within individual Rupert's drops and learn more about the mechanism of its explosion. For the present, the basic questions I posed earlier remain to be solved: Why does the drop explode and by what mechanism do the fractures propagate? Rupert's drops are amusing to work with and as perplexing in some ways today as they were in the time of Samuel Butler, who penned these lines in 1663 :

Honour is like that glassy bubble
That finds philosophers such trouble,
Whose least part crack'd, the whole does fly
And wits are crack'd, to find out why.

(The word play associating 'crack'd wits' with philosophers may still have a wide appeal today.)

7. DARK MATTER IN THE UNIVERSE

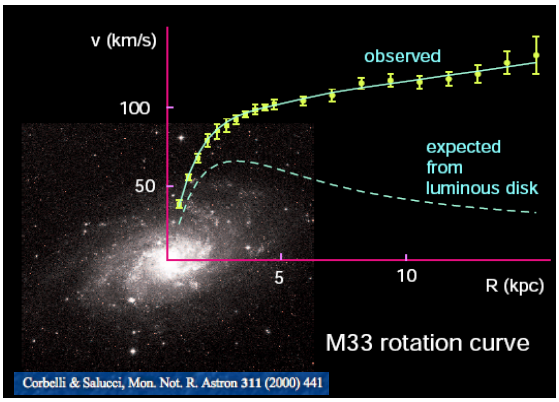
I will conclude this lecture with the strangest 'material' I ever had to think about—and which may or may not even exist. Dark matter is the hypothesised unseen substance that many physicists believe comprises about 95% of all the mass in the universe. If it does *not* exist, then there is a serious discrepancy between the amount of luminous matter detectable electromagnetically at all wavelengths and the amount of matter inferred exclusively from gravitational and cosmological observations. And if it *does* exist, then my own theoretical investigations with my colleague Ronald Mallett suggest that it must have properties radically different from those according to the prevailing theories that now guide all experimental searches for it.

The evidence in support of the existence of dark matter comes from a variety of sources among which are

- Galaxy rotation
- Gravitational lensing
- Cosmic microwave background
- Large-scale structure
- Baryon acoustic oscillations

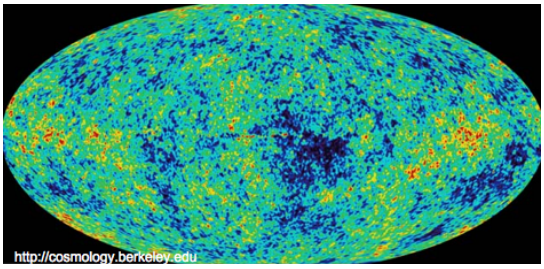
This lecture is not the place for a comprehensive discussion of the theories of dark matter and their justifications. I have written about dark matter in my book *A Universe of Atoms, An Atom in the Universe*. Let it suffice, then, if I briefly mention two of the more readily explicable arguments.

The first concerns the orbital rotation of matter such as stars and gas clouds about the centers of galaxies. As first noted by Fritz Zwicky at Caltech in the 1930s, the velocities of galaxies in the Coma cluster exceeded what could be deduced by the virial theorem from the total observable mass of the cluster. More dramatic still, were the Doppler shift measurements of Vera Rubin of outlying interstellar hydrogen clouds in the Andromeda galaxy revealing velocities again in excess of what could be deduced from the visible mass of the galaxy. According to Kepler's law



$$v(r) = \sqrt{\frac{GM}{r}}$$

the velocity of an object orbiting a mass M should decrease as the square root of the distance r from the centre of rotation. Instead, as illustrated above for the Triangulum Galaxy (M33) about 3 million light years away, the rotation curve actually continues to rise for thousands of light years outside the luminous disk. (Note: 1 kiloparsec (kp) = 3.26 light years.) The observed rotation curves for most galaxies either rise or remain more or less level with distance beyond the bulk of luminous matter, indicating the pervasive presence of unseen matter.



The second argument is cosmological. Measurements of the angular variations in temperature of the cosmic microwave background radiation (CMBR)—i.e. the relic radiation from the Big Bang diffusing through the universe once the temperature of matter dropped to a level such that protons and electrons could combine to form neutral hydrogen atoms—indicate that the mean mass density is sufficient to produce a universe of zero curvature—i.e. 'flat'. This implies the existence of a much greater mass density than what could be inferred from the presence of luminous matter. First detected by COBE (Cosmic Background Explorer) and then measured with significantly higher precision by WMAP (Wilkinson Microwave Anisotropy Probe), the CMBR data have led physicists to conclude that baryonic matter (i.e. ordinary atoms and subatomic particles) make up about 4.6% of the mass-energy of the universe, the remainder being about 23.3% dark matter (responsible for otherwise unaccountable gravitational attractions) and 72.1% dark energy (accountable for Type Ia supernova brightness variations subsequently attributed to an accelerated expansion of the universe).

In the standard model currently accepted by most theorists working on the problem, dark matter is made up of **W**eakly **I**nteracting **M**assive **P**articles or WIMPs, i.e. bosonic particles with masses estimated to be as high as 10-1000 Gev/c^2, where for reference the mass of a proton is about 1 Gev/c^2. The theory has been successful in accounting for large-scale structure, but predicted spiky galactic cores and dense substructures in galactic halos that astronomers have not observed. Moreover, despite ongoing experimental efforts to detect WIMPs directly in the laboratory, the consensus is that none seems to have been found.

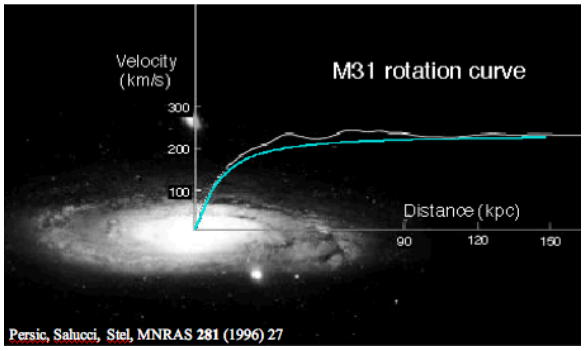
Personally, I do not believe that WIMPs exist—or, if they do, that they form the bulk of the mass of the universe. Consideration of this conundrum from a perspective analogous to that which forms the 'Standard Model of Cosmology'—i.e. symmetry breaking of a quantum field of bosons in a universe shaped by gravity as embodied in Einstein's general theory of relativity—led me to conclude that dark matter—if it *is* real stuff—would comprise a quantum condensate[17]. In keeping with the whimsicality of particle physicists (although I am not one), and in contrast to the ideas embodied in the appellation WIMPs, I have referred to such matter as WIDGETs = **W**eakly **I**nteracting **D**e**G**enerate **Et**her because of their very low predicted mass. WIDGETs have a mass, estimated by requiring the condensate coherence length to be approximately of galactic dimensions, much lower than that of electrons or even neutrinos (the existence of whose mass was inferred relatively recently from the phenomenon of neutrino oscillation). The transition from a gas of incoherent bosons to a quantum condensate occurred around the time of primordial nucleosynthesis (first few minutes after the Big Bang). Quantum mechanics prevents a condensate from collapsing to a size much below its coherence length, which thereby avoids the creation of spiky galactic centers and superfluous dense substructures.

In a Bose-Einstein condensate all the bosons have condensed into the ground state to form a macroscopic coherent quantum system. The dynamics of the system in flat spacetime are

[17] M P Silverman and R Mallett, (a) "Cosmic Degenerate Matter: A Possible Solution to the Problem of Missing Mass", *Classical and Quantum Gravity* **18** (2011) L37; (b) "Coherent Degenerate Dark Matter: A Galactic Superfluid?", *Classical & Quantum Gravity* **18** (2001) L103; (c) "Dark Matter as a Cosmic Bose-Einstein Condensate and Possible Superfluid", *General Relativity & Gravitation* **34** (2002) 633-649.

approximately described by a nonlinear Schroedinger equation referred to as the Gross-Pitaevskii equation. Applied to the condensate model of dark matter, the equation yields a solution

$$v(r) = v_\infty \sqrt{1 - \frac{\tanh(r/r_c)}{r/r_c}}$$



for the rotational speed of luminous matter orbiting the galactic centre in terms of two model-dependent parameters: v_∞ is the velocity of matter infinitely far from the centre and r_c is a characteristic length interpretable as the de Broglie wavelength of the condensate particle. The figure shows the rotation curve of the Andromeda galaxy (M31), which at 2.5 million light years away is the closest galaxy to our own (the Milky Way, also referred to simply as the Galaxy). The blue line superposed on the rotation curve is a visual fit obtained with the preceding theoretical prediction.

An interesting and unintended byproduct of the Silverman-Mallett theory was that the process of symmetry breaking, which led to the prediction of dark matter as a condensate of light bosons, also resulted in a cosmological constant in the Lagrangian. The cosmological constant was originally a term added by Einstein to his field equations of gravity in order keep the universe from collapsing under the mutual gravitational attraction of its contents. Einstein did this prior to Hubble's observations pointing to an expanding universe, and he (Einstein) subsequently withdrew the term when it became clear that the universe had a beginning and evolutionary history. The cosmological constant results in a repulsive interaction and is now (to my knowledge) the leading mechanism by which to explain the apparently unrelated hypothetical entity known as dark energy. In the Silverman-Mallett theory, the 'dark side' of nature —matter and energy—arose by a common mechanism.

The possibility of a macroscopic boson condensate was predicted by Einstein in the mid 1920s, yet so difficult had it been to achieve the requisite low temperatures that the first experimental demonstration (by Cornell and Wieman) took place in 1995, some seventy years later. At times, I have read, physicists began to doubt whether this state of matter actually existed. In his statement accompanying the press release of their experimental success, Cornell is reported to have said

- Eric Cornell (1995): *This state could never have existed naturally anywhere in the universe. So the sample in our lab is the only chunk of this stuff in the universe, unless it is in a lab in some other solar system.*

  http://jilawww.colorado.edu/www/press/bose-ein.htm

Six years later, upon completion of the Silverman-Mallett theory of dark matter, I reflected on Cornell's remark:

- Mark P Silverman (2001): *It is interesting to contemplate that, if the conclusions presented in our paper are confirmed, then the new state of matter…may well be the most abundant form of matter in the cosmos.*

  M P Silverman and R L Mallett, *Class. Quant. Grav.* (2001) **18** L1-L6

Time will tell whether Cornell or I or neither of us has speculated accurately.

In the years since working on the problem of dark matter, I have leaned increasingly toward a viewpoint that dark matter may well go the way of the electromagnetic aether. Prior to Einstein's revision of classical kinematics through his theory of special relativity, the aether was created solely to provide a medium through which electromagnetic waves could propagate. The more one investigated this medium, the stranger were the properties with which it had to be endowed in order to account for experimental observations. Ultimately, the entire structure was discarded when it was finally understood that light waves required no medium for propagation. Like the aether, dark matter was created to account for a single phenomenon: the relation between gravitational force and mass. Presuming that the quantitative description of gravity (Einstein's general relativity) was correct at large scales (thousands and millions of light years), a larger than expected gravitational attraction could only be understood in terms of the presence of an unseen amount of matter. But, again like the aether, this matter must have extraordinary properties. It cannot consist of any of the particles that make up the ordinary matter we find in our solar system and all across the universe. And it must exist in so large a proportion that it dwarfs by comparison the fraction of all other more familiar kinds of matter.

It seems to me that there must be a simpler more satisfying solution...although I don't know of one at this time—quite possibly a solution whereby the law of gravity differs from general relativity in some fundamental way that does not conflict with constraints posed by past experimental verifications. Perhaps Newton's universal constant of gravity, *G*, is not a constant, but scales in some way with energy or distance.

Physicists have often found nature to be strange—as reflected in the oft-cited comment of Haldane (who was a biologist and not a physicist) that "My own suspicion is that the Universe is not only queerer than we suppose, but queerer than we *can* suppose." I am myself inclined to agree still with the (positively phrased) first part of his suspicion [Nature is queerer than we suppose], but believe that we have long ago reached a point where human imaginations can run rampant beyond nature's more governed one.

And if some day a theory of gravity is constructed that accounts for the anomalous galactic rotation curves and all the rest, then dark matter—the alleged most extensive form of matter in the universe—will vanish in its entirety overnight. We will have lost a most unusual material, but in turn gained a world of understanding.

For more information about all the topics discussed in this lecture, see: