Light and information

D. Gabor – Imperial college, London

1972

This article is the substance of a Ritchie lecture, delivered by the author on March 2, 1951 at the University of Edinburgh. The contents of the lecture became known to a wider wider audience through the distribution of a limited number of mimeogaphed notes, which have since become widely quoted in the literature. This wish has been often expressed that a permanent record of the lecture should be made generally available. We are glad to meet this wish.

1 Introduction

Light is our most powerful source of information on the physical world. Anthropologists have often emphasized that the privileged position of Man is due as much to his exceptionally perfect eye, as to his large brain. I was much impressed by a remark of Aldous Huxley, that we owe our civilization largely to the fact that vision is an objective sense. An animal with an olfactory sense or with hearing, however well developed, could never have created science. A smell is either good or bad, and even hearing is never entirely neutral; music can convey emotions with an immediateness of which the sober visual arts are incapable. No wonder that the very word "objective" has been appropriated by optics. But on the other hand it is probably the peculiar character of vision which is chiefly responsible for one of the most deep-rooted of scientific prejudices; that the world can be divided into an outer world and into an "objective" observer, who observes"what there is", without influencing the phenomena in the slightest.

In this lecture an attempt will be made to discuss optics from the point of view of information theory. But before doing this, I must start with a disclaimer. I do not want to give the impression that we have now a valuable new epistemological principle, which we want to hand over to the physicist. Nothing irritates the physicist more than when the philosopher tries to look over his shoulder and to give him advice, and this is hardly surprising in view of the past record of philosophers, from Aristotle to Hegel, to mention only those who are safely dead. Information theory does not originate from philosophers, but also from a group of outsiders; from mathematically interested electrical engineers, and mathematicians interested in communications. They may not be quite as suspect of conceit as philosophers, but it will be as well to point out from the start that the point of view of information theory was never quite absent from physics, and has been growing stronger and stronger in modern physics long before information theory became fashionable. Again and again in the course of this lecture I shall be able to point out the work of physicists in this direction. But having said this, I may be allowed also to say that the points of view of information theory, consequently applied, may yet prove of appreciable heuristic value to physics.

What then are the points of view of information theory? I want to say that I am stating my own views, not necessarily shared by others who are working in this field. There are two steps in the approach. In the first step we specify the *degrees of freedom* of the phenomenon, in such a way that we operate always with discrete degrees, and in all practical cases with a finite number of them. This, in MacKay's useful terminology specifies the *structural* aspects of information.

Once we have found the right coordinates, the second step is to specify the phenomenon by attaching a measure to each coordinate. But it is essential that we must never expect an *exact* measure. we must take account of the fact that in every physical measurement there is an unavoidable amount of uncertainty, fluctuation or "noise", so that the best we can do is to specify the measure between certain limits, with a certain probability. A convenient way of doing this is to lay down a certain"scale of distinguishable steps", also called a "proper scale". This means that we proceed along the scale in steps roughly equal to the uncertainty. Of course some sort of convention must be made regarding what one considers as distinguishable, e.g. by agreeing that if one says that the value is in a certain interval, this means that on repetition of the experiment one would find this interval say in 50% of the cases. Once such a convention is made – and practically it is easy to fix one in most cases – the measurement is expressed by an *integer*, by the order of the interval, counted from the lowest step.

Thus in information theory every phenomenon is described by *a finite number of integers*. There is no continuity, except in the probabilities. There is no need to emphasize how close this view comes to the method of quantum physics, and the authors of information theory do not wish to plead ignorance of this fact. On the contrary this was always emphasized, especially in a paper by MacKay, and those of the author. The "structural" information, i.e. the free coordinates of the phenomenon to be studied, has been also called the "*a priori*" part of it. What is meant by this can perhaps be illustrated by Eddington's famous "parable of the fishing net" (Eddington [1939], pp. 16, 62). – If an ichthyologist casts a net with meshes two inches wide for exploring the life on the ocean, he must not be surprised if he finds that "no sea-creature is less than two inches long". Similarly, if one tries to explore atmospherics by means of a radio set with a bandwidth of a thousand cycles, there is no need to look out for surges with a "base" of less than a millisecond. But one must be very careful with the word "*a priori*". We do not always know our instrument as well as the ichthyologist ought to know his net, and the specification of the free coordinates of the instrument requires physical knowledge, and not only the knowledge of formal logic, as may be suggested by the word "*a priori*". Later in this lecture there will be opportunity for showing that an important part of our knowledge of light is in fact embodied in the system of "free coordinates", suitable for its description.

The handling of the metrical information (sometimes called *a posteriori*), in information theory has a distinctive feature which may be briefly mentioned. The number which appears as the result of the measurement is often considered as a selection from a number of possible values. Historically this may be attributed to the fact that the first authors in the field which became later known as "communication theory", Nyquist, Kupfmuller, Hartley, were interested in telegraphy, where the signals are in fact selections from a certain discrete set. This view may appear a little strange to the physicist, but he may remember that once he has set his galvanometer, every possible reading is a selection from the distinguishable marks on his scale. At any rate if we include the reading "off the scale". Nor is this concept such a stranger to physics as it might appear at first sight, as we shall see later when we come to the discussion of light, information and thermodynamics.

2 Geometrical Optics

After a few, rather unsuccessful attempts of the ancients, the laws of light were first formulated round the turn of the 16th century in the form of geometrical optics. This is built on the concept of a "ray of light" which for a long time was naively identified with a geometrical line (sometimes a curve.)

From the point of view of information theory this is a completely unsatisfactory departure. Every point of an object plane sends out a double infinity of rays, and if we had a perfect lens, which is no impossibility in geometrical optics, we can unite this whole pencil of rays in one point of an image plane, and study the object plane point-for-point. But there is no need for a perfect lens. Let us take instead a *camera obscura*, with a "point-hole", and we have automatically perfect representation. The number of "free coordinates" is infinite in this naive views; we have not only an infinity of points or rays, but it is *transfinite* infinity.

It is evident how strongly these naive views, and the crude experiments on which they are based are responsible for our belief in a continuous geometry. It was, of course, a very sound instinct which led Snell, Descartes and others to base the infant theory of light on what appeared to them the safe foundation of euclidean geometry. To this day we cannot do without the concept of a continuous space, though it is no longer euclidean. Attempts to eliminate it appeared to Einstein as promising as "breathing in a vacuum". Some day it may be possible to discard it, but the time has not come yet, and we shall have to use continuous space as a background, though it will soon become evident that what we can physically distinguish in it are not points, but at the best small, diffuse patches. Yet, geometrical optics gives at least a hint which way to look for a basis in applying information theory to light. Information is something which is propagated from the object to the image without destruction, if the imaging system is a perfect one; thus we must look for the invariants of the imaging process. Moreover we must look for a geometrical invariant for the structural specifications; one which exists as soon as we set up the image plane, the object plane and the lens system, irrespective of what object we put in the plane, and how we illuminate it.

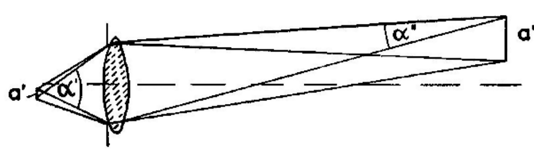

Figure 1: The Smith-Lagrange invariant in geometrical optics

But there exists only one of this type; the Smith-Lagrange invariant (Fig. 1). This is the product of any small line element at right angles to the optic axis with the angular divergence of the rays which issue from any one of its points and pass through the lens aperture. This is the same for the object as for its image,

$$a^{'}\alpha^{'} = a^{''}\alpha^{''}$$

This holds exactly true only for perfect imaging, but – excluding certain types of lens errors – it will be true also for less perfect ones if we restrict both the elements, and the divergences to very small values. We can now write conveniently

$$dSd\Omega = \text{inv.} \tag{1}$$

where $d\Omega$ is the solid angle of a very narrow cone of rays, and dS is the projection of the area of a very small element, viewed from the direction of the cone. It will be seen later that this is, in fact, an important cue.

3 Classical Wave Optics

After Snell, Descartes and Fermat the next great progress came with Christiaan Huygens, who formulated what we would call nowadays the scalar wave theory of light. It is known that this had to be replaced later by the "vector theory" of Young and Fresnel, and that their mechanical vectors had to be reinterpreted by Maxwell as electromagnetic ones, but these steps, important as they were, are not as fundamental from the point of view of information theory as Huygens' step from rays to waves. But we must not forget that Newton, though he opposed the wave theory, supplied what is perhaps the most important element in it, by his celebrated experiments in which he decomposed light into spectral colours, and showed that these could not be further decomposed. This made it possible later, in the hands of Young and of Fresnel, to associate a characteristic length, the wavelength, with every spectral colour. It is this characteristic length which changes the picture completely from the point of view of information theory.

In wave optics the concept of a "ray" is not at all elementary. Its place is taken by the simplest solution of the wave equation; the plane, monochromatic wave. Unfortunately, like most of the simple concepts with which we do our thinking, this turns out to be a very remote abstraction from reality because it must be infinite in extension. But we must retain it because of its mathematical simplicity, with much the same reservation which we have made about geometry. Let us therefore consider for a start what appears the simplest case; a plane, monochromatic wave with wavelength λ impinging on a plane object. But in order to represent such a wave mathematically we must make another questionable assumption. Consider for simplicity "scalar light" with amplitude u. (In the vector theory we can instead consider any cartesian component of the vectors involved.) This must satisfy the wave equation

$$\Box\, u \equiv \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2}{\partial t^2} \right) u = 0$$

which expresses the fact that light propagates in all directions with the velocity c. But if we want to satisfy this equation, we must assume that the wave which is periodic in space with period λ is periodic in time with a frequency $\nu = c/\lambda$. This introduces an air of unreality into classical wave optics, because the frequency of light has never been measured in any optical experiment, nor the phase of this hypothetical vibration. What we measure are always wavelengths and *relative* phases, which are entirely determined by geometry. The behaviour of light *in time* appears in wave optics as an *ad hoc* construction, so contrived as to account for the velocity of propagation. But we will accept it for the present, because for long wavelengths, for radio waves, frequency and phase become really measurable quantities, and the vibrations can be followed in time by means of oscillographs. Why frequency should be measurable for long waves but not for short ones is a question to which classical wave theory has evidently no answer, and which we must leave for later.

Consider now that such a wave, whose mathematical expression is

$$u_0 = e^{2\pi i (z/\lambda - \nu t)}$$

falls in the z-direction on a plane object in the plane $z = 0$ (Fig. 2). Immediately behind the object the amplitude will be given by some expression of the form

$$u(x, y, +0, t) = t(x, y) e^{2\pi i \nu t} \tag{2}$$

$t(x, y)$ is the complex "amplitude transmission" of the object. There is no need here to discuss its meaning, and how it is related to physical properties of the object, because in this experiment the function $t(x, y)$ *is* the object. That is to say it contains everything that we can expect to find out about the object; in fact, as we shall see in a moment, it contains much more, and only a small part of it is actually observable.

The amplitude being given by eq. (2) immediately behind the object, the problem is to calculate it for any z. One could solve it by using the method of Huygens and Fresnel, by superimposing the elementary spherical wavelets issuing from all surface elements of the object. But another method, connected with the name of Fourier, but which, I believe, was first introduced into optics by Rayleigh is far more appropriate.

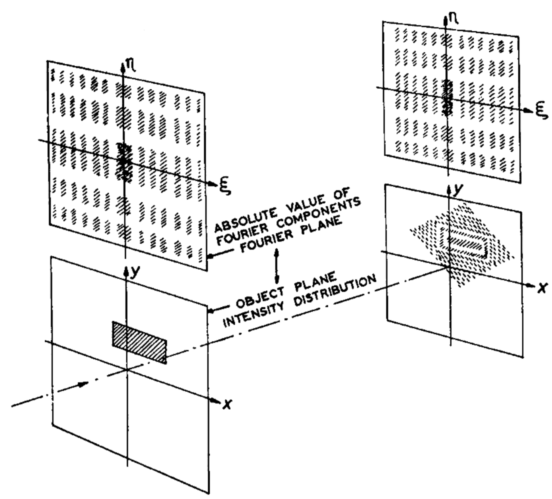

Figure 2: Propagation of light waves

One starts by decomposing the transmission function $t(x, y)$ into its Fourier components, by the formula

$$t(x, y) = \int\int_{-\infty}^{+\infty} T(\xi, \eta) e^{2\pi i (x\xi + y\eta)} d\xi d\eta. \tag{3}$$

Each Fourier component represents a simple periodic infinite "standing wave" of transmission, with the periods $1/\xi$ and $1/\eta$ in the x and y direction.

Thus the "Fourier variables" ξ and η can be interpreted as *wave numbers* in the plane $z = 0$. The (complex) amplitude $T(\xi, \eta)$ of these components is called the *Fourier transform* of $t(x, y)$.

By the principle of superposition (first noticed for water waves by Leonardo da Vinci), the amplitude at any point x, y, z can be calculated by determining separately the wave issuing from every one of the Fourier components, and summing them. The calculation – carried out in Appendix I – gives a very simple and significant result : those Fourier components whose period in the object plane is longer than a wavelength will be propagated as plane waves, while those with a shorter period will be continued as exponentially damped "*evanescent* waves", which means that they will be practically damped out in a matter of few wavelength at most.

It is intuitively clear that if the Fourier components below a wavelength are cut out, all details of the object (that is to say of $t(x, y)$) which are finer than about half a wavelength will be cut out with them. Thus we arrive at the first significant result of wave theory, that light with a wavelength λ will under no circumstances carry with it information on detail below $\frac{1}{2}\lambda$

We obtain a very clear idea of the propagation of the remaining information if we follow the transformation of the amplitude u with increasing distance z from the object plane in space, and simultaneously in "Fourier space". This is illustrated in Fig. 2, but for simplicity only the intensity is shown, i.e. the squared absolute value of the amplitude u, and the modulus of its Fourier transform. The striking feature is that while the intensity pattern changes rapidly, so that the object soon becomes unrecognizable, the modulus of the Fourier transform does not change at all. This can be easily understood if it is remembered that each point ξ,η the Fourier pattern corresponds to a certain direction in space, in which the corresponding plane wave is propagated, and this does not of course change in free space. The phase (argument), of the Fourier component changes with z, but here again we have a law which is very much simpler than the one for the phase change of u: – The phase factor of T depends only on z, ξ and η, i.e. it is independent of all other points in the Fourier diagram, and it can be easily calculated, as shown in Appendix I. This is the advantage of the method of Fourier transforms, which does not apply to the simple case only which we have here considered, and which is finding increasing applications in instrumental optics, after having been for many years one of the chief mathematical tools of communication engineering.

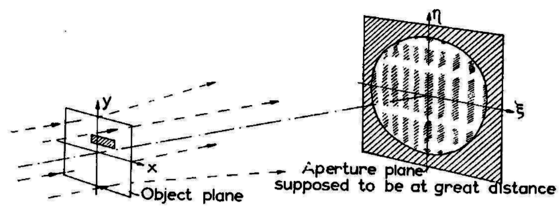

Figure 3: Connection between Fourier variables and angular variables

If the distance z of the screen on which we observe the intensity is further increased, all resemblance to the object is gradually lost, and finally the intensity pattern becomes identical with the Fourier modulus diagram. This is illustrated in Fig. 3. This happens at a distance so large that every plane wavelet issuing from the object can be considered as a "ray". If we use a lens-between the object and the screen, there is no need to go to infinity, we find the same conditions very nearly realized in the rear focal plane of the lens. This is the plane in which to place a ray-limiting aperture, if one wants the same angular limitation for every point of the object. In this simple case it is quite evident that we lose some further information, because we have cut out all Fourier components outside a certain area.

We are now in a position to answer the first question of information theory, the question of the degree of freedom, or of "free coordinates"' We can reformulate this question in the form: –"How many in- dependent variables are necessary to express as much of the function $t(x, y)$ as we can learn from an optical image, under certain conditions of ray limitation?" Consider first, for simplicity, the last example, in which the Fourier variables were all limited to the same region (by an aperture at a large distance), independently of the space coordinate x, y. We now build up the complicated beam which issues from the object out of elementary beams, every one of which has a non-zero Fourier transform only inside the allowed region, and we try to expand $t(x, y)$ in a series of these. We find that we get into difficulties, because if the Fourier spectrum is sharply cut off, as assumed, these beams will spread out at the base, i.e. in the object plane, to infinity, hence we cannot have, as we wished, a sharply limited object. without going into technicalities which have been dealt with elsewhere (Gabor [1946]) we will only mention that there exists a correlation of the form

$$\frac{\text{smallest eff. beam area x solid ang. of divergence}}{\text{square of wavelength}} \geq 1 \quad (4)$$

and that the smallest possible value of this ratio is achieved for the rotationally symmetrical "gaussian elementary bean" illustrated in Fig. 4. This smallest value is of the order unity, with any reasonable definition of the quantities which figure in the numerator in eq. (4).

Thus we see that so long as the product of object area and Fourier area is of the order unity or smaller, we cannot even start to answer the question regarding the degrees of freedom, because we cannot construct even one elementary beam to satisfy the cut-off conditions.

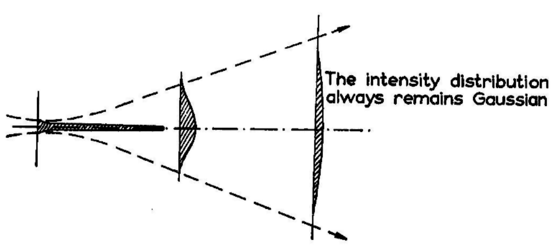

Figure 4: Gaussian elementary beam

The question is evidently of a statistical nature, and can be answered with an accuracy of order $1/M$ if the product of object area and Fourier area is of the order M; a large number. But with this qualification we can give an answer to the question : — A monochromatic beam of light has F degrees of freedom, where

$$F = 2 \times 2 \times \text{object area} \times \text{accessible Fourier area} \quad (5)$$

because it takes this number of independent terms to build up what remains of $t(x, y)$ inside the object area, after cutting out the Fourier components outside a certain area[1]. The first factor 2 is due to the fact that each term has an arbitrary complex coefficient, equivalent to two real data, the other is due to the vector nature of light. In principle light can transmit two independent images, polarized at right angles to one another. This result is essentially contained in an important paper by Max von Laue [1914], though not in connection with the transmission of information by light. It may be mentioned that the theorem has not yet been proved with a rigour which would satisfy mathematicians, but physicists have their own standards in these matters.

The result is illustrated in Fig. 5. The information space has really four dimensions x, y, ξ and η but in the simple case where the solid angle Ω is independent of x, y three dimensions suffice. The theorem can be evidently generalized: The degree of freedom is 2 (or 4) times the volume of the information space available.

So far we have talked of the stationary case only, i.e. of a steady, unchanging image. What happens if the object is moving or changing?

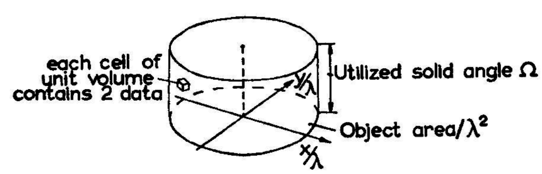

Figure 5: Information space

We can give an answer immediately, by availing ourselves of the now fairly generally known results of communication theory. Every degree of freedom can

[1] Appendix II contains two examples of such series expansions for the "non-redundant" representation of what is left of $t(x, y)$ after cutting off Fourier components. Physicists will need no reminding of how similar that is to the procedure in quantum mechanics, especially in Dirac's formulation.

be conceived as a separate and independent communication line, which has

$$(2)\Delta\nu\Delta t$$

degrees of freedom in a frequency interval $\Delta\nu$, and the time interval, (observation time) Δt. The factor (2) in brackets is to be used if the "temporal phase" is measurable. The author has shown in a recent paper (Gabor [l950]) that in the case of light this is possible only with quite extraordinary intensities, combined with high spectral purity, which it may never be possible to realize with existing light sources. But in the region of radio waves phase is easily measurable, and the factor 2 is justified. — We can now write down our result for the degree of freedom of *any* beam of light (which need no longer be monochromatic or coherent[2]) in the general form

$$2 \times 2 \times (2) \int\int\int\int\int\int dx dy d\xi d\eta d\nu dt, \qquad (6)$$

or, in terms of the cross-section dS and the solid angle $d\Omega$

$$2 \times 2 \times (2) \int\int \frac{dS}{\lambda^2} d\Omega d\nu dt \qquad (7)$$

This is evidently a significant quantity, because $\frac{dS}{\lambda^2} d\Omega$ and $d\nu dt$ are both relativistic invariants. But the result is hardly written down before doubts arise whether it can really stand on its own legs. We have already seen that the bracketed factor 2 becomes physically real only at very high intensities. But another question, even more elementary is suggested by eqs. (6) and (7) : What happens if we do not cut off the area or the angular variables sharply, as we have assumed up to now, but e.g. just almost cut out a part of the waves, by an *almost* black filter? Are we still allowed to measure the information space just as if it were fully accessible? This is a familiar dilemma in problems of a statistical nature, to which classical theory has no answer. The weighting factor which is evidently necessary will have to come from another side. But before approaching this question, we will sharpen the dilemma, by an example which throws into relief the logical insufficiency of the classical scheme.

[2] *cf.* Appendix III

4 The Paradox of "Observation without Illumination"

The classical theory of light claims validity at all levels of intensity, however small. This appears a harmless assumption. Combined with the elementary experience that in fact every observation requires a certain minimum, finite light sum, one would at first sight conclude only that one has to wait a correspondingly long time for an observation.

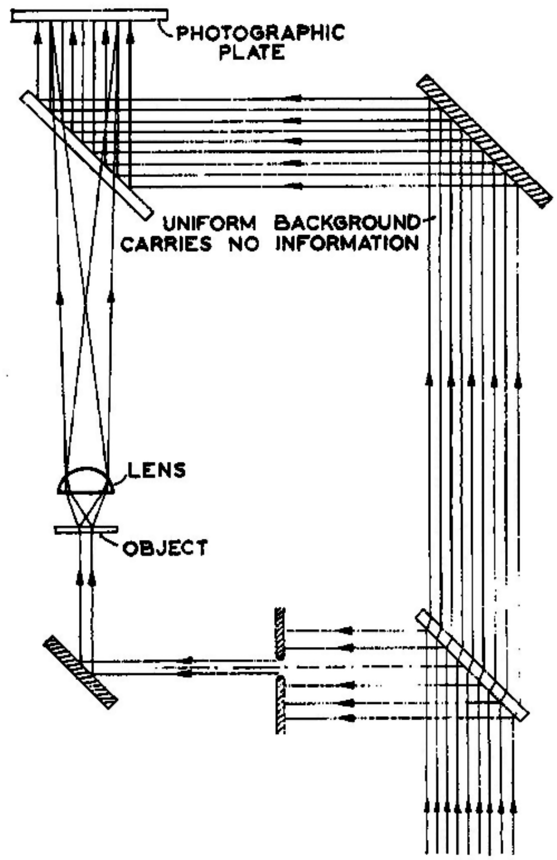

Figure 6: Observation "without illumination"

But it will now now be shown that if the classical theory were true, however large the minimum energy, we could make an observation with alight sum passing through the object which could be made as small as we like.

Let us take a Zehnder-Mach interferometer, as shown in Fig. 6, in which coherent light is divided into two very unequal parts. Only a very small fraction is directed through the branch which contains the object; the rest is branched through the other term, and united with the weak beam only at the receptor, which may be e.g. a photographic plate. Thus we have divided the tight into two parts; a weak one which carries all the information, and a strong one which carries almost all the energy.

For simplicity let us fix our attention on one resolvable element of the object, say a square whose edges are equal to the resolution limit, so that the result of the observation is expressed by a single number; the light sum through the image of the element during the observation time. Let us call A_0 the amplitude in the strong, uniform background, a the amplitude which the image-carrying beam would produce by itself. As the two are coherent, the resulting intensity is

$$I = A_0^2 + a^2 + 2A_0 a \cos\phi \tag{8}$$

if ϕ is the phase angle between the two which depends on the optical paths and also on the phase delay in the object. Similarly we have the relation between the resulting and the partial light sums

$$S = S_0 + s + 2(S_0 s)^{\frac{1}{2}} \cos\phi \tag{9}$$

This is the sum of the large uniform background term S_0, known beforehand, the light sum s which has penetrated through the object, and which can be assumed as very small, and an interference term. This can happen to be zero if the two amplitudes are in quadrature, but if necessary we can repeat the experiment with quarter-wave plate introduced into one branch or the other. The absolute expectation value of this term is

$$\frac{4}{\pi}(S_0 s)^{\frac{1}{2}} \tag{10}$$

which means that we can *amplify* the effect of the weak image carrying beam roughly in the ratio $(S_0/s)^{1/2}$. It is true that the contrast is still small, of the order $(s/S_0)^{1/2}$, but as the background is known and uniform, we can subtract it. Subtraction is particularly easy if electrical methods are used; one takes the image on a television screen and suppresses the d.c. component in the transmission. But it can be done also with photographic plates if the grain is fine enough to be negligible, e.g. by using the Foucault-Toepler "schlieren" method. We are allowed to neglect the grain, because however large the area and the corresponding minimum light sum, and however small s we can make so large enough for the product (10) to become observable. Thus, in the limit, we could make an observation with as small a total illumination of the object as we like.

Instinct of course tells us that this cannot be true. The weak point in the argument is evidently the subtraction of the strong but uniform background. The argument would break down if, in increasing the intensity in the background, we would, at the same time, increase its *uncontrollable fluctuations* to such an extent that in the end the interference term (10), which indicates the object, could not be told against the background of "noise". But we could eliminate the imperfections of the apparatus unless these fluctuations arose *from the nature of light itself.*

Let us now make the reasonable assumption, that the experiment is bound to fail if the light sum s which has gone through the object element is smaller than a certain minimum energy ε_0, because for

$$s < \varepsilon_0$$

the interference term (10), becomes smaller than the root of the mean fluctuation square of the background, i.e.

$$S_o s < (\delta \bar{S_0})^2.$$

Assume that equality, i.e. possible observation is just achieved for $s = \varepsilon_0$, this can be written

$$\delta(\bar{S_0/}\varepsilon_0)^2 = S_0/\varepsilon_0. \tag{11}$$

S_0/ε_0 is a pure number, the light sum in the background in units of the minimum energy which makes an elementary observation possible. But eq. (11) is Poisson's "law of rare events". It could be exactly accounted for by the hypothesis that monochromatic light arrives in quanta of some size ε_0 which arrive at random, subject only to the condition that an average of S_0/ε_0 arrives during the observation time. *No observation can be made with less than one quantum passing through the observed object.*

5 A Further Paradox : "A Perpetuum Mobile of the Second Kind"

This chapter is skipped, for it is very intricate and not so much interesting for the discussion.
It will be translated in a further version of this document (the interested reader can find the original version of the present article here `http://antoine.wojdyla.fr/assets/archive/gabor1951-original.pdf`)

6 The Metrical Information in Light Beams

We can now return to the problem of the information content of light, which we had to leave in a rather unsatisfactory state. Classical theory enabled us to count the degrees of freedom, but it did not provide a metric. In quantum theory we can *count* light energy: in terms of photons, which provides a natural measure.

In classical theory there is no upper limit to field intensities, and quantum theory, at least for the present, retains this feature by allowing *any* number of photons in one cell, i.e. in any one degree of freedom. It is interesting to consider for a moment to what an extent we can avail ourselves in practice of this generous theoretical permission. We consider three powerful sources in different parts of the electromagnetic spectrum

A *power generating station*, 10 000 kW, 50 ± 0.01 cycles/sec puts about

$$10^{41} \text{ photons}$$

into a single cell[3].

A *large magnetron*, in pulsed operation on 10 cm wavelength, $3 \times 10^9 \pm 0.5 \times 10^6$ cycles, though with an instantaneous power only ten times smaller than the generating station, produces

$$10^{24} \text{ photons per cell.}$$

This is 10^{15} times smaller than in the first case, but still a large enough number to make electrical engineers indifferent to quantum theory. But a powerful high pressure mercury lamp, emitting 1 watt per cm^3 arc area in the form of the green line $\lambda = 5461 \pm 10$ Angstroms achieves less than

$$10^{-3} \text{ photons per cell,}$$

i.e. the best it can do is about one photon for a thousand cells! Hence light optics, when it comes to metrical problems is entirely outside the classical region. The classical theory has given us the formula

$$2 \times 2 \times (2) \int\int \frac{dS}{\lambda^2} d\Omega d\nu dt \qquad (12)$$

for the degrees of freedom in an arbitrary light beam, called also the "number of logons". We can now consider every logon separately from the point of view of *information capacity*. It is convenient to define this as the logarithm of the number of distinguishable steps $s(n)$ if 1, 2, ... photons are packed into it, up to a level n. This problem was the subject of a recent investigation by the author [1950] where it was found that the number of distinguishable steps is, approximately,

$$s(n) = 2\left(\frac{n}{1 + 2n_T}\right)^2$$

and that the factor (2) in eq. (7) must be suppressed[4]. n_T is here the number of thermal photons

[3] see appendix V.

[4] It can be directly verified that this is the number of distinguishable energy levels, by using an extension of Einstein's law for the energy fluctuations in a Hohlraum. One might ask whether, at high quantum levels, this is the same as the number of distinguishable *states*, because classical theory associates two quantities with every level: an amplitude and a phase. These are also considered as observable in the quantum theory of radiation, but only to a certain accuracy, determined by the uncertainty relation. Born and Rosenfeld [1933, 1950] have proved that these measurements can be indeed carried out, if no restriction is imposed on the particles used in the imaginary experiments, i.e. if one admits test bodies composed of nuclear matter, or even denser. On the other hand, I have found *l.c.* that if one uses electrons, and the type of electronic amplifier which appears the most promising for this purpose, the total information contained in the best possible measurements of

at the temperature T at which the observations are made, which is, by Planck's law

$$n_T = 1/(e^{h\nu/kT} - 1) \tag{13}$$

Now consider the case of a beam with many degrees of freedom, as given by (7), the beam which is issuing from an object under a microscope. We define the information capacity again as the *logarithm of the number of distinguishable states.* In order to calculate this by combining the number $s_i(n_i)$ for the different degrees of freedom i we must have some condition for the n_i the photons which may appear in i. We obtain such a condition in the simplest and most natural form if we separate the "time cell", $\Delta\nu\Delta t$ from the integrand, and put it equal to unity. In this case all elementary beams contained in eq. (7) are necessarily *coherent.* We now imagine that the object has been illuminated with N photons, in the same unit time cell, which means of course "coherent illumination". Thus a total maximum of N photons can appear in the beam issuing from the object, this will be the case if the object is not absorbing but has only "phase contrast". The problem is not clearly given; in how many ways can we distribute N photons over F degrees of freedom, and what is the total number of distinguishable patterns, formed by combinations of distinguishable steps? The quantity defined in this way is very close to, though not quite identical with Max Planck's [1924] definition of the *entropy* of a quantized system as k times the logarithm of the "probability" P, which is defined as the number of ways in which a given energy can be distributed over the states. The difference is only that we have replaced "states" by "distinguishable states".

The calculation, carried out in Appendix VI, gives the asymptotic formula, valid for large N and F

$$\log P = \tfrac{1}{2} F \log \frac{2\pi e N}{F(1+2n_T)} \tag{14}$$

for the maximum information capacity in a beam with F degrees of freedom and containing N photons.

amplitude and phase will be, at most, equal to total number of distinguishable energy levels, i.e. to $s(n)$. The reason for this remarkable divergence from the "ideal" experiments is the shot effect in electron beams. If one assumed that, at least for very high n, eq. (12) has to be replaced by a linear law, this would merely restore the factor (2) which we have suppressed.

This formula still has the weakness that it gives equal weight to all degrees of freedom. What happens if we do not cut out some of the degrees but weaken them by absorbing screens? These screens are a part of our experimental set-up, they are part of our a priori information. We can answer the question immediately by associating a transmission coefficient τ_i with the i-th degree of freedom, which is a real, positive number, smaller than or at most equal to unity. Eq. (14) now changes into

$$\log P = \tfrac{1}{2} \sum_{i=1}^{F} \log \frac{2\pi e N \tau_i}{F(1+2n_T)} \tag{15}$$

This formula at last answers the objections to the classical theory; the degrees of freedom are properly weighted. It may be noted that the formula is an asymptotic one, it must not be extended to τ_i so small that an added degree of freedom might appear to make a negative contribution, which happens if the argument of the logarithm falls below unity, i.e. we must cut off at

$$\tau_i \frac{N}{F} = \bar{n_i}\tau_i >= \frac{1}{2\pi e}(1+2n_T) \tag{16}$$

For zero thermal noise, $n_T = 0$ this limit is about 17 times smaller than that given by MacKay's intuitive rule: "Adding a degree of freedom is useless if it will contain in the mean less than about 'one metron per logon' ". The reason for this appreciable discrepancy is that if the logon is one of many, and a large energy is distributed over them, it can still make a useful contribution in the cases where it receives an energy above the average, in other words by making use of the fluctuations.

It may be pointed out that the entropies (14) and (15) which have a very close relation with what goes under this name in statistical mechanics, are not to be identified with Shannon's "entropy, or measure of information in communication theory". The relations between them are discussed in Appendix VI.

7 Conclusion

This, I believe, does not by any means exhaust what information theory can give to the physics of light.

I have mentioned the unavoidable increase of disorder which every observation must create, but I could not go into the question of the unavoidable disorder which an observation creates in the object itself. This question was first raised by Bohr and by Heisenberg, and most important further developments are due to L. de Broglie [1947]. It is a problem of the greatest interest to those who, like the author, are engaged to extend the limits of microscopic vision. I hope to have shown that information theory is of some heuristic use in physics, by asking the right sort of questions. But even if this were questioned, another advantage is, I believe, evident beyond doubt. This is that it prepares the mind for quantum theory, whose strange methods are so difficult to assimilate for those who have been too long engaged in classical physics. As we must now give up all hope of ever understanding the physical world on classical lines, it is gratifying that in information theory we appear to have the right tool for introducing the quantum point of view into classical physics.