

1

Introduction

And the Lord saw that the light was good.

Genesis 1:4

Most probably all people, even though they belong to different cultures, would agree on the extraordinary role that light – the gift of the Sun-god – plays in nature and in their own existence. Optical impressions mediated by light enable us to form our views of the surrounding world and to adapt to it. The warming power of the sun’s rays is a phenomenon experienced in ancient times and still appreciated today. We now know that the sun’s radiation is the energy source for the life cycles on Earth. Indeed, it is photosynthesis in plants, a complicated chemical reaction mediated by chlorophyll, that forms the basis for organic life. In photosynthesis carbon dioxide and water are transformed into carbohydrates and oxygen with the help of light. Our main energy resources, coal, oil and gas, are basically nothing other than stored solar energy.

Finally, we should not forget how strongly seeing things influences our concepts of and the ways in which we pursue science. We can only speculate whether the current state of science could have been achieved without sight, without our ability to comprehend complicated equations, or to recognize structures at one glance and illustrate them graphically, and record them in written form.

The most amazing properties, some of which are completely alien to our common experiences with solid bodies, can be ascribed to light: it is weightless; it is able to traverse enormous distances of space with incredible speed (Descartes thought that light spreads out instantaneously); without being visible itself, it creates, in our minds, via our eyes, a world of colors and forms, thus “reflecting” the outside world. Due to these facts it comes as no surprise that optical effects confronted our knowledge-seeking mind with more difficult problems than those of moving material objects. Over several hundred years a bitter war was fought between two parties. One group, relying on Newton’s authority, postulated

the existence of elementary constituents of light. The other, inspired by the ideas of Huygens, fought for light as a wave phenomenon. It seemed that the question was ultimately settled in favor of the wave alternative by Maxwell's theory, which conceived light as a special form of the electromagnetic phenomena. All optical phenomena could be related without great difficulty and to a high degree of accuracy to special solutions of the basic equations of classical electrodynamics, the Maxwell equations.

However, not more than 40 years passed and light phenomena revealed another surprise. The first originated in studies of black-body radiation (radiation emitted from a cavity with walls held at a constant temperature). The measured spectral properties of this radiation could not be theoretically understood. The discrepancy led Max Planck to a theory which brought about a painful break with classical physics. Planck solved the problem by introducing as an *ad hoc* hypothesis the quantization of energy of oscillators interacting with the radiation field.

On the other hand, special features of the photoelectric effect (or photoeffect) led Einstein to the insight that they are most easily explained by the "light quantum hypothesis". Based on an ingenious thermodynamic argument Einstein created the concept of a light field formed from energy quanta $h\nu$ localized in space (h is Planck's constant and ν is the light frequency).

The newly created model was fully confirmed in all its quantitative predictions by studies of the photoeffect that followed, but there was also no doubt that many optical phenomena like interference and diffraction can be explained only as wave phenomena. The old question, Is light formed from particles or waves?, was revived on a new, higher level. Even though painful for many physicists, the question could not be resolved one way or the other. Scientists had to accept the idea that light quanta, or photons as they were later called, are objects more complicated than a particle or a wave. The photon resembles a Janus head: depending on the experimental conditions it behaves either like a particle or as a wave. We will face this particle-wave dualism several times in the following chapters when we analyze different experiments in our quest to elucidate the essence of the photon. Before this, let us take a short stroll through the history of optics.

2

Historical milestones

2.1 Light waves à la Huygens

While the geometers derive their theorems from secure and unchallengeable principles, here the principles prove true through the deductions one draws from them.

Christian Huygens
(*Traité de la Lumière*)

Christian Huygens (1629–1695) is rightfully considered to be the founder of the wave theory of light. The fundamental principle enabling us to understand the propagation of light bears his name. It has found its way into textbooks together with the descriptions of reflection and refraction which are based on it.

However, when we make the effort and read Huygens' *Treatise of Light* (Huygens, 1690) we find to our surprise that his wave concept differs considerably from ours. When we speak of a wave we mean a motion periodic in space and time: at each position the displacement (think about a water wave, for instance) realizes a harmonic oscillation with a certain frequency ν , and an instantaneous picture of the whole wave shows a continuous sequence of hills and valleys. However, this periodicity property which seems to us to be a characteristic of a wave is completely absent in Huygens' wave concept. His waves do not have either a frequency or a wavelength! Huygens' concept of wave generation is that of a (point-like) source which is, at the same time, the wave center inducing, through "collisions" that "do not succeed one another at regular intervals," a "tremor" of the ether particles. The given reason for wave propagation is that ether particles thus excited "cannot but transfer this tremor to the particles in their surrounding" (Roditschew and Frankfurt, 1977, p. 31). Therefore, when Huygens speaks of a wave, he means an excitation of the ether caused by a *single* perturbation in the wave centrum, i.e. a single wavefront spreading with the velocity of light. The plots drawn by Huygens showing wavefronts in an equidistant sequence have to be

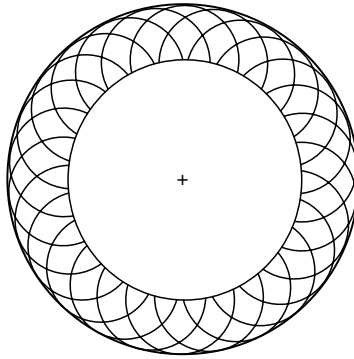


Fig. 2.1. Propagation of a spherical wave according to the Huygens principle.

understood such that it is the *same* wavefront at different times, and the regularity of the plot is caused exclusively by having chosen identical time differences.

In fact, what was correctly described in this way is white light – sunlight for instance. The time dependence of the excitation – or more precisely speaking the electric field strength component with respect to an arbitrarily chosen direction – is not predictable but completely random (stochastic).

On the other hand, it is also clear that such a theory is not able to explain typical wave phenomena such as interference or diffraction where the wavelength plays an important role. It required Newton and his ingenious insight that natural light is composed of light with different colors to come nearer to an understanding of these effects.

This should not hinder us, however, from honoring Huygens' great "model idea" known as the Huygens principle according to which each point either in the ether or in a transparent medium reached by a wave, more precisely a wavefront, becomes itself the origin of a new elementary wave, as illustrated in Fig. 2.1 for the example of a spherical wave.

The wavefront at a later time is obtained as the envelope of all elementary waves emitted at the same, earlier, moment of time. However, Huygens could not answer the question of why a *backwards running* wave is generated only at the boundary of two different media and not also in a homogeneous medium (including the ether). In fact, a satisfactory answer could not be given until Augustin Fresnel complemented the principle with the principle of interference – we use today the term Huygens–Fresnel principle – the strength of which is demonstrated when we treat theoretically the problems of diffraction. By the way, the answer to the above question is simple: the backwards running waves "interfere away."

But let us return to Huygens! Using the assumption that light propagates in two different media with different velocities, we can easily explain reflection and

2.2 Newton's light particles

5

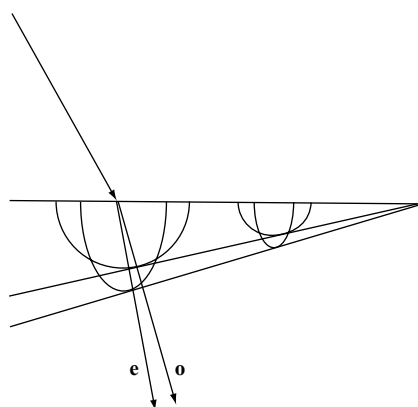


Fig. 2.2. Birefringence after Huygens. o = ordinary beam; e = extraordinary beam; the arrows indicate the beam direction.

refraction of light using the Huygens principle. The explanation of the strange effects of birefringence (the splitting of the incident beam into an ordinary and an extraordinary beam) can be viewed as an extraordinary success of the principle. It was based on the ingenious guess that the surfaces of propagation for the extraordinary beam are not spherical surfaces but the surfaces of a rotational ellipsoid (see Fig. 2.2), an insight that is fully verified by modern crystallography. However, Huygens had great difficulty understanding the experiments he performed with two crystals of calcite. We will discuss this problem in more detail in the next section.

2.2 Newton's light particles

As in mathematics, so in natural philosophy, the investigation of difficult things by the method of analysis, ought ever to precede the method of composition. This analysis consists in making experiments and observations, and in drawing general conclusions from them by induction, and admitting of no objections against the conclusions, but such as are taken from experiments, or other certain truths.

Isaac Newton
(*Opticks, 3rd Book*)

Isaac Newton (1643–1727) was the founder of the particle theory of light. Even though the light particles postulated by Newton do not have anything in common with what we now call photons, it is still exciting to trace back the considerations which led such a sharp mind to the conclusion that light of certain color is composed of identical, elementary particles. As an abiding supporter of the inductive method as the method of natural sciences, Newton was guided by a simple experience: the straight line propagation of light “rays,” recognizable on the sharp

contours of shadows of (nontransparent) objects placed into the beam path. This effect seemed to Newton to be easily explained by assuming that the light source emits tiny “bullets” propagating along straight lines until they interact with material objects. He believed a wave process to be incompatible with straight line propagation of the excitation. Water waves showed a completely different type of behavior: they obviously run around an obstacle!

Since the breakthroughs of Young and Fresnel it has been known that Newton’s conclusion was premature. What happens when a wavefront hits an obstacle depends crucially on the ratio between the size of the obstacle and the wavelength. When the ratio is very large, the wave character is not noticeable; in the limit of very small wavelength the propagation of light can be described by straight lines, i.e. light rays. On the other hand, wave properties become predominant when the dimensions of the obstacle are of the order of the wavelength, as in the above example of water waves.

Newton himself observed with impressive precision experimental phenomena where the outer edge of a body (for instance the cutting edge of a razor) deflects the light “rays” in its proximity a little from their original straight direction so that no ideally sharp shadows are observable. He did not take these phenomena, now called the diffraction of light, to be hints of a wave-like character of light; instead, he considered the bending as the result of a force applied onto the particles (in his opinion caused by the density of the ether increasing with increasing distance from the object), an idea which was completely in accord with the well established concepts of mechanics. Newton’s belief that light had a particle nature should be judged from the perspective of the seventeenth century atomism, which was, at the time, a deeply rooted concept. “True” physics – in contrast to scholasticism which categorized light and color phenomena into the class of “forms and qualities” – was imaginable only as a mechanical motion of particles under the influence of external forces.

The most important argument expounded by Newton against the wave theory of light advanced by Christian Huygens was, however, a very odd observation made and reported by his great opponent (who even honestly admitted to have “found no satisfactory explanation for it.”)

What was it? It is well known that a light beam is split by a calcite crystal into an ordinary beam and an extraordinary beam and – provided it is incident orthogonally to the rhombohedral plane – the latter beam is shifted to the side. Then the two beams lie in one plane, the so-called principal intersection of the incident beam.

Huygens arranged vertically two calcite crystals with different orientations and let a light beam impinge from above. He made the following observation: usually both the ordinary and the extraordinary beam leaving the first crystal were split again in the second crystal into two beams, an ordinary and an extraordinary

one. Only when the two crystals were oriented either so that the intersections were mutually parallel or mutually orthogonal did just two beams emerge from the second crystal. Whereas in the first case the ordinary beam remained ordinary in the second crystal (the same naturally applied also to the extraordinary beam), in the second case, in contrast, the ordinary beam of the first crystal was converted into the extraordinary beam of the second crystal, and correspondingly the extraordinary beam of the first crystal was converted into the ordinary.

These last two observations surprised Huygens. He wrote (Roditschew and Frankfurt, 1977, p. 43): "It is now amazing why the beams coming from the air and hitting the lower crystal do not split just as the first beam." In the framework of the wave theory of light – note that we are dealing with a scalar theory similar to the theory of sound, where the oscillations are characterized by alternating expansions and compressions of the medium; at that time no one thought of a possible transverse oscillation! – we face a real dilemma: a wave, using modern terminology, is rotationally symmetric with respect to its propagation direction, or, as Newton formulated it, "Pressions or motions, propagated from a shining body through an uniform medium, must be on all sides alike" (Newton, 1730; Roditschew and Frankfurt, 1977, p. 81). There does not seem to be a reason why the ordinary beam leaving the first crystal, for example, should in any way "take notice" of the orientation of the second crystal.

Newton saw a way of explaining the effect using his particle model of light. The rotation symmetry can be broken by assuming the particles not to be rotationally symmetric but rather to have a kind of "orientation mark." The simplest possible picture is the following: the light particles are not balls but cubes with physically distinguishable sides, and the experiment suggests that opposite sides should be considered equivalent. Newton himself is not explicit about the form of the particles, whether they are cubic or block-like, and is satisfied by ascribing to the light particle four sides, the opposites of which are physically equivalent. We thus deal with two pairs of sides, and Newton called one of these pairs the "sides of unusual refraction."

The orientation of the side pairs with respect to the principal intersection of the calcite crystal determined in Newton's opinion, the future of the light particle when it enters the crystal in the following sense: depending on whether one of the sides of the unusual refraction or one of the other sides is turned towards the "coast of unusual refraction" (this means the orientation of the side is orthogonal to the principal intersection plane), the particle undergoes an extraordinary or an ordinary refraction. Newton emphasizes that this property of a light particle is present from the beginning and is not changed by the refraction in the first crystal. The particles remain the same also, and do not alter their orientation in space.

In detail, the observations of Huygens can now be explained as follows (Fig. 2.3(a)): the original beam is a mixture of particles oriented one or the other

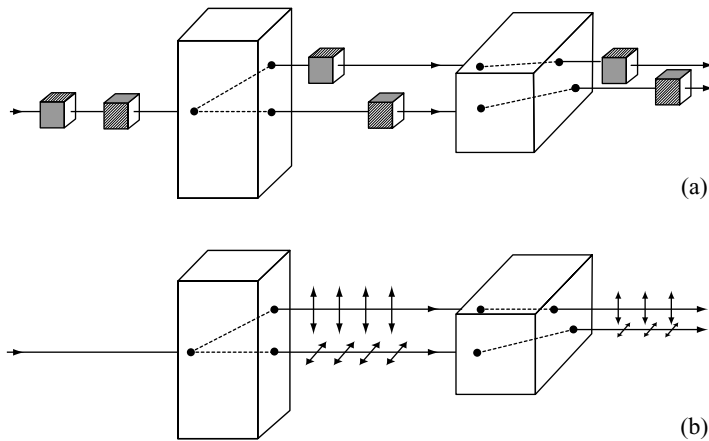


Fig. 2.3. Passage of a beam through two calcite crystals rotated by 90° . (a) Newton's interpretation; (b) modern description. (The arrows indicate the direction of the electric field strength.)

way with respect to the principal cut plane of the first crystal. The first crystal induces a separation of the particles, depending on their orientation, into an ordinary beam and an extraordinary beam. When the crystals are oriented with their principal intersection planes in parallel, the orientations of the particles with respect to the two crystals are identical. The ordinary beam of the first crystal is also the ordinary beam of the second, and the same applies to the extraordinary beam. However, when the principal intersections of the crystals are mutually orthogonal, the orientation of the particles leaving the first crystal with respect to the second changes so that the ordinary beam becomes the extraordinary beam, and vice versa.

With this penetrating interpretation of Huygens's experiment, in fact Newton succeeded in describing phenomenologically polarization properties of light. Even the name "polarization" was coined by Newton – a fact that is almost forgotten now. (He saw an analogy to the two poles of a magnet.) Today it is well known that the direction to which the "sides of unusual refraction" postulated by Newton points is physically nothing else than the direction of the electric field strength (see Fig. 2.3(b)).

Even though Newton's arguments in favor of a particle nature of light no longer convince us, and the modern concept of photons is supported by completely different experimental facts which Newton could not even divine with his atomic light concept, this ingenious researcher raised an issue which is topical even now. Newton analyzed the simple process of simultaneous reflection and refraction which is observable when a light beam is incident on the surface of a transparent medium. The particle picture describes the process in such a way that a certain percentage of the incident particles is reflected while the rest enters the medium

2.3 Young's interference experiment

9

as the refracted beam. In the spirit of deterministic mechanics founded by him, Newton asks what causes a randomly chosen particle to do the one or the other. In fact, the problem is much more acute for us than for Newton because we are now able to perform experiments with individual particles, i.e. photons. While quantum theory sees blind chance at work, Newton postulated the cause of the different behavior to be “fits of easy reflection” and “fits of easy transmission” into which the particles are probably already placed during the process of their emission. These “fits” show a remarkable similarity to the “hidden variables” of the twentieth century which were advocated (as it turned out, unsuccessfully) to overcome the indeterminism of the quantum mechanical description.

We would not be justified, however, in considering Newton, one of the founders of classical optics, to be a blind advocate of the particle theory of light. On the contrary, he was well aware that various observations are understandable only with the aid of a kind of wave concept. He formulated his thoughts in the form of queries with which he supplemented later editions of his *Opticks* (Newton, 1730; Roditschew and Frankfurt, 1977, p. 45), and all of them were characteristically formulated in the grammatical form of negations. It seemed to him that along with the light particles, waves propagating in the “ether” also take part in the game. Is it not so, he asks, that light particles emitted by a source must excite oscillations of the ether when they hit a refracting or reflecting surface, similar to stones being thrown into water? This idea helped him to understand the colors of thin layers which he studied very carefully, mainly in the form of rings formed on soap bubbles (which were named after him).

Altogether we find so many hints for a wave nature of light in Newton's *Opticks* that Thomas Young could cite Newton as “king's evidence” for the wave theory in his lecture “On the theory of light and colours” (Young, 1802; Roditschew and Frankfurt, 1977, p. 153). Even though Young definitely missed the point, it would be just to see Newton as one of the forerunners advocating the dualistic concept that light has a particle as well as a wave aspect (even though he gave emphasis to the first one). From this point of view, Newton seems to us much more modern and nearer in mentality to us than to most of the representatives of nineteenth century physics who were absolutely convinced of the validity of the wave picture.

2.3 Young's interference experiment

The theory of light and colours, though it did not occupy a large portion of my time, I conceived to be of more importance than all that I have ever done, or ever shall do besides.

Thomas Young

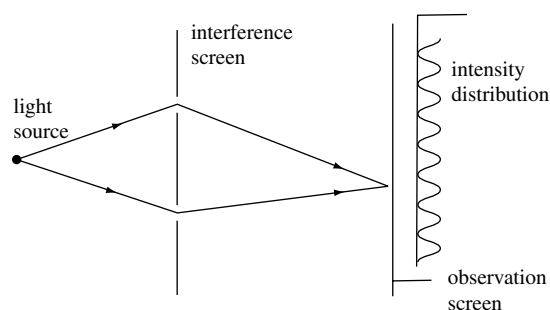


Fig. 2.4. Young's interference experiment.

Interference phenomena are viewed as the most convincing “proof” of the wave nature of light. The pioneering work in this field was carried out by Thomas Young (1773–1829) and, independently, by Augustin Fresnel (1788–1827). It was Young, despite his work going practically unnoticed by his contemporaries, who performed the first interference experiment which found its way into all the textbooks. The principle of the experiment is the spatial superposition of two (we would say coherent) light beams. This is achieved by using an almost point-like monochromatic source and allowing its light to fall onto an opaque screen with two small holes or slits (Fig. 2.4). The two holes themselves become secondary sources of radiation, but because they are excited by the same source they do not radiate independently. We can now position an observation screen at a convenient distance from and parallel to the first screen and we will observe, near the point where the normal (constructed precisely at the center of the straight line connecting the two holes) intersects the observation plane, a system of mutually parallel bright and dark stripes (orthogonal to the aforementioned connecting line), the so-called interference fringes.

The distance (which is always the same) between neighboring stripes depends primarily on the distance between the holes – it is larger when the holes are closer together – and secondly on the color of the light; when we work with sunlight we obtain colored stripes which become superposed at a certain distance from the center of the interference pattern, and the eye gets the impression of a surface uniformly illuminated by white light.

The surprising feature of this observation (let us consider again a monochromatic primary source) is the existence of positions that appear dark even though they are simultaneously illuminated by both holes. When one of the holes is blocked, the interference fringes vanish and the observation screen looks uniformly