

FRITS ZERNIKE

How I discovered phase contrast

Nobel Lecture, December 11, 1953

« Phase contrast » was not discovered while working with a microscope, but in a different part of optics. It started from my interest in diffraction gratings, about from 1920 on. Such a (reflecting) grating consists of a plane or concave mirror with a large number of equidistant grooves ruled on its surface. Small nearly unavoidable imperfections in the location of the grooves show clearly in the optical behaviour of the grating. The most conspicuous error is a periodic one which repeats itself after each revolution of the screw of the ruling engine. The regularly recurring displacement of the grooves causes corresponding changes of the optical path, just as if the mirror surface were wavy. Consequently the instrument behaves as if a coarse grating, with a constant of about 2 millimeters, were superimposed on it, so that each strong spectral line is accompanied to right and left by a number of weak spurious lines, the so-called « Rowland ghosts ». These have a remarkable effect if one looks down at the surface of the grating, placing the eye at the position of a spectral line. A perfect grating would in this case show an evenly illuminated surface, in the colour of the spectral line. In reality, however, one sees a strongly striped surface. At the end of a 1902 paper, H. S. Allen remarked that these stripes were nothing real, but simply the effect of the interference between the principal line and its ghosts. Indeed the stripes disappear when the ghosts are covered up. I remember strongly objecting against his conclusion of unreality. On the contrary I was convinced that the striped surface gave more information about the periodic ruling errors than that obtainable by photographing the ghosts, because in the first case the relative phases of the ghosts come into play, whereas these are lost in the second case. I kept the question in mind, planning to look further into it as soon as an opportunity would arrive.

About 1930 our laboratory had obtained a large concave grating and set it up in a Runge-Paschen mounting. The striped appearance of the surface was soon found, but as the grating was 6 meters from the eye, I tried a small telescope pointed at the grating. Then the unexpected happened. The stripes were seen very clearly, but disappeared as the telescope was exactly focussed

on the surface of the grating! By a succession of experiments and calculations I soon succeeded in explaining this. On looking back to this event, I am impressed by the great limitations of the human mind. How quick are we to learn, that is, to imitate what others have done or thought before. And how slow to understand, that is, to see the deeper connections. Slowest of all, however, are we in inventing new connections or even in applying old ideas in a new field. In my case the really new point was that the ghosts differed in phase from the principal line. Now it is common knowledge that in all interference phenomena differences of phase are all-important. Why then had phases never been considered before in this case, nor in the corresponding one in the microscope? Some excuse may be found in the difficulty to define them exactly. Let me explain this for a more simple case, the diffraction image of a slit. The way to observe this may be as follows. A telescope is pointed at a vertical line-source of light, such as the filament of an incandescent lamp. A vertical slit of say 2 mm width is placed closely behind the objective of the telescope. This causes the image of the source to be broadened out into a diffraction pattern: a bright central stripe (order zero) is accompanied on both sides by weaker and weaker secondary maxima (orders one, two, etc.). The formula for this diffraction pattern is given in the textbooks, the amplitude being determined by the function $\sin x/x$. In the few cases where the phases are mentioned in the literature, on the other hand, there is no consensus of opinion. Some say the phases are equal over the whole pattern - except for the obvious reversal of the odd orders, whereas others make them change proportional to x^2 . I find that it all depends on the, often tacitly assumed, surface to which the phases are referred. If this reference surface is the focal plane of the telescope objective, one comes to the second statement, if it is a cylindrical surface with the centre line of the slit as axis, the equality of phases results.

You may want to ask whether these phases can be observed. I find they can. All one has to do is to throw the diffraction image on a *coherent background*, obtained in the following way. The slit is covered by a glass plate with a thin metallic layer which transmits a few percent of the light. A fine scratch is made in this layer, forming a narrow slit which is adjusted so as to lie in the centre of the broad slit. The light through the scratch is broadened out by diffraction and thus forms the desired background, which interferes with the diffraction pattern. The phases of this pattern are thus compared with those of the auxiliary wave forming the background. In the experiment the auxiliary wavefront therefore plays

the role of the cylindrical reference surface in the theoretical treatment.

It is only by the introduction of an adequate reference surface that a definite statement about the phase differences involved can be made. In the case of the Rowland ghosts the result was: their phases differ by ninety degrees from the principal line. Now I happened to know of a simple method to change this. Lord Rayleigh described in 1900 how to make very shallow etchings in glass surfaces without spoiling their optical quality, by the slow action of very dilute hydrofluoric acid. By this process I made what I called *phase-strips*: glass plates with a straight groove, a millimeter or less wide and of a uniform depth of half a wavelength. Such a phase-plate was placed in the spectrum so that a bright spectral line fell on the strip, whereas its ghosts passed through the glass beside it. In a telescope behind the phase-plate the stripes on the grating surface then stood out clearly.

For a physicist interested in optics it was not a great step to change over from this subject to the microscope. Remember that in Abbe's remarkable theory of the microscope image the transparent object under the microscope is compared with a grating. To be precise, a transmission grating is considered as the test-object and the diffraction by this grating as the primary phenomenon. At first sight this has nothing to do with the magnified image of the object formed by the microscope objective. Instead, the objective forms an image of the light source, practically in its back focal plane, consisting of a central direct image accompanied by diffracted images on both sides. This, although on a very much smaller scale, is the analogue of the grating line with its ghosts. The light issuing from these images overlaps in the eye-piece of the microscope and by interference gives rise to stripes which, curiously enough, resemble a magnified image of the object! Abbe's theory has been summarized in this sentence: « The microscope image is the interference effect of a diffraction phenomenon. »

You will now readily understand that, acquainted with this theory, I soon tried my phase-strip in a microscope, throwing the direct image of a linear light-source on the strip placed closely above a low-power objective.

I must now explain why the unexpected discovery of the 90° phase shift applies to the microscope image as well. It all depends on the nature of the object under the microscope. In his theory Abbe and his followers always considered an object of alternate opaque and transparent strips. The diffraction images for such a grating calculated in the well-known way are in phase with the central image. On the other hand, if the object consists of alternate thicker and thinner strips, then the phase difference of 90° is found. In the

first case, the diffraction is caused by the unequal amplitudes of the light passing the strips, in the second case by the unequal light paths, i.e. by the unequal *phases*. I therefore distinguish the two by calling the first kind an *amplitude grating*, the second a *phase grating*. Or in the general case of an irregular structure, an « amplitude object », resp. a « phase object ». Nearly all objects of biological or medical interest belong naturally in the second group. The highly developed staining techniques evidently aim at changing them, or their special details one wants to see, into amplitude objects.

It will now be seen that for a phase object my phase-strip in the focal plane of the microscope objective brought the direct image of the light-source in phase with the diffracted images, making the whole comparable to the images caused by an amplitude object. Therefore the image in the eyepiece appears as that of an absorbing object, that is with black-and-white contrast as if the object had been stained. The full name of the new method of microscopy might be something like: « phase-strip method for observing phase objects in good contrast ». I shortened this into *phase-contrast method*. Before going into further details about the development of the method, a few general remarks should be made.

In a treatise on the Abbe theory, Lummer comes to the conclusion that « in the ideal case the microscope image is exactly similar to the object in structure and phase ». Now the absolutely transparent details of a phase object leave the intensity of the passing light unchanged. All they do is to impress phase differences on it. According to Lummer, then, the image will show the same phase differences - which however are invisible - and an equal intensity everywhere. In other words, the phase object is absolutely invisible « in the ideal case ». Of course the practical microscopist has never been content with this; as a matter of fact, he never found it out! Without realizing it, he had always turned the fine adjustment, that is, put the object a little out of focus, in order to see the tricky transparent details. Only a somewhat diffuse and watery image is obtained in this way. This will be explained further on.

With the phase-contrast method still in the first somewhat primitive stage, I went in 1932 to the Zeiss Works in Jena to demonstrate. It was not received with such enthusiasm as I had expected. Worst of all was one of the oldest scientific associates, who said: « If this had any practical value, we would ourselves have invented it long ago ». Long ago, indeed! The great achievements of the firm in practical and theoretical microscopy were all due to their famous leader Ernst Abbe and dated from before 1890, the year in which Abbe became sole proprietor of the Zeiss Works. From then on he

had been absorbed in administrative and social problems, partly also in other fields of optics. Indeed his last work on microscopy dates from that same year. In it he gave a simple reason for certain difficulties with transparent objects, but this was of no account. His increasing staff of scientific collaborators evidently under the impression of his inspiring personality, formed the tradition that everything worth knowing or trying in microscopy had been achieved already. For more than twenty-five years after Abbe's death in 1906, his great authority thus barred the way to further progress.

Here is one more remarkable historic point. Whereas all the other achievements of Abbe's were greatly appreciated by all practical microscope users, his theory of image formation was firmly rejected by most of them. To the physicist this may seem incredible, especially when he remembers Abbe's experiments which in his opinion confirm the theory in a convincing way. The opposing microscopists, however, said these experiments only showed how the microscope may be used, or rather misused, by the physicist for interference experiments which have nothing to do with the ordinary proper use of the instrument. A long story could be told about the violent controversies of this kind which occurred time and again through half a century. We can say now that Abbe and his followers were as much to blame as the other party. For one thing, their theory was too abstract, it only applied to the oversimplified cases of a point source of light and an object of regular structure. However, it could very well have been extended to meet the practical cases. But then it was also incomplete, it did not explain the peculiarities in the imaging of transparent objects, and what is worse, its defenders never recognized this incompleteness. Small wonder therefore that the microscopist rejected the theory as useless in practice.

Returning now to the phase-contrast method, I will first give an account of its working principle. Let the incident light for simplicity be a plane wave. Without an object, that is if there is only a clear glass-plate under the microscope, this wave passes unchanged, is brought to a focus closely above the objective (in its back focal plane), and spreads out again to an evenly illuminated field in the eyepiece. If there is a real object, every small detail of it will give rise to a slight perturbation of the wave. One may always consider this as resulting from a perturbed wave to be superimposed - in amplitude, not in energy - on the unchanged wave. This last one shall be called the *direct* light, it will clearly give the even background. The perturbed wave will spread out from the detail in all directions, will fill the whole aperture of the objective and will reunite in the corresponding image point in the eyepiece.

The perturbed waves from all the object points together will be called the *diffracted light*. The microscope image in the eyepiece now results from the interference of the diffracted light with the direct light. In order to obtain phase contrast the two must be treated differently, so as to change their relative phases. This is possible, because they are spatially separated in the back focal plane of the objective. The interplay of phases in this decomposing and reuniting of vibrations can best be visualized in a vector diagram (Fig. 1a). As is well known, a harmonic vibration is obtained from a vector MV rotating uniformly round M . The projection P on the horizontal axis performs the vibration. The vector MP' , projection of MV' which remains always perpendicular to MV , performs a similar vibration, one quarter period in advance of P . In accordance with general usage the projecting is understood, we speak of the vibrations MV , MV' , etc.

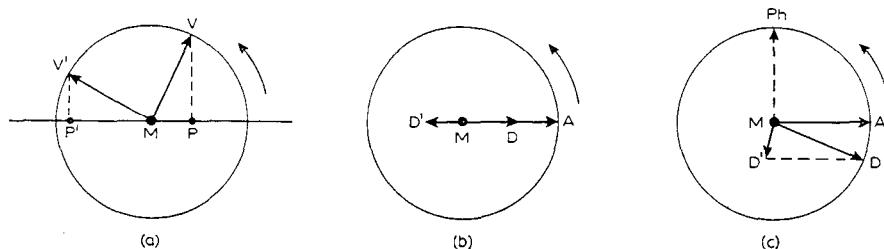


Fig. 1.

Now consider a microscopic object with slightly absorbing details on a transparent background (stained preparation). The vibration emerging from the background may be represented by MA (Fig. 1b). An absorbing detail weakens the light, it gets a smaller amplitude, such as MD . The vector MD results also from compounding MA with MD' , so that MD' represents the change caused by the details, i.e. the perturbed vibration. Now according to a well-known theorem the optical paths along all rays from an object point to its image are equal. Therefore the direct and the diffracted vibrations arrive at the image point in the same relative phases they had in the object, and the same Fig. 1b may thus serve for the reuniting of these vibrations. As a result the absorbing detail is seen darker than the background. Now compare this with the case of a transparent object (unstained preparation, Fig. 1c). Its details will ordinarily be somewhat stronger refracting than the embedding medium. This means that the light is propagated with less speed so that the emerging vibration MD will be retarded in phase compared to MA ,

but equal in amplitude. The change caused by the detail is now represented by MD' , nearly perpendicular to MA . The compounding of these in the image gives again MD , equal in intensity to the background MA , the detail remains invisible. It will appear, however, on slightly defocussing, as the light-paths are no longer equal in that case, resulting in some change of respective phases. At the same time the image becomes blurred, so that the observer has to find a compromise between its disappearance from the first cause, exact focus, and from the other, fading out by lack of focus. In the phase-contrast method however, the direct light has to pass the phase-strip, which is thinner than its surround through which the diffracted light passes. The direct light is thus advanced by 90° , being then represented by MP_h . This causes the detail to be represented by the vector-sum of MP_h and MD' , making it darker than the background. Clearly the relations are about the same as in Fig. 1b, the transparent detail may be said to be « optically stained ».

Two further improvements of phase contrast, which I made in the first years can now be explained. One is the absorbing phase-strip. Details in the object that are very thin will cause only very small phase differences, and in Fig. 1c this corresponds to a very short vector MD' , to be compounded with MP_h . The thin detail therefore appears only very little darker than its surround, i.e. with very little contrast. Now there is no simple way of increasing the amplitude MD' of the diffracted light, but the same result, increased contrast, may be attained by diminishing the amplitude of the direct light MP_h . To this end the phase-strip must not only accelerate the direct light but also partly absorb it. This is for instance obtained by a thin metallic deposit on the strip. An absorption of 75 % is often used; the strip then transmits 25 % of the energy, or one half of the amplitude of the direct light. The contrast is thus doubled, a quite marked effect. In my own experiments I could go down to 4% transmission, i.e. a five times enhanced contrast, the limit being set by the unavoidable stray light. It is only under specially favourable circumstances that a higher increase has been attained by the French astronomer Lyot. In his study of the minute ripples of polished lens surfaces he had independently rediscovered phase contrast and could use strips that diminished the amplitude to one thirtieth, so that ripples only one thousandth of a wavelength high showed in good contrast.

A last point to explain is the halo that is always observed surrounding objects which show great contrast. This must be ascribed to the action of the phase-strip on the diffracted light. As we saw before, the phase-strip is meant to act on the direct light only. However, the diffracted light, which

fills the whole aperture of the objective, will for a small part be intercepted by the phase-strip and this part remains inactive. To find the effect of this missing part, we consider the reverse case, that it would be the only active part. Because of the narrow strip, it would form an image of much less resolving power, i.e. blurred by diffraction. As this part is missing, the «strip image» must be subtracted, in amplitude, from the full image formed by the whole aperture. The interference with the direct light then results in a very diffuse and weak negative image, appearing as a bright halo round dark details, as a dark halo round bright details.

With the straight phase-strips used in the beginning, the halo may be disturbing because the strip image of a small detail is by diffraction only spread out in one direction, namely perpendicular to the strip. This makes small bright spots in the image appear as if marked by short crossing pencil streaks. To remedy this I soon introduced *annular strips*, which make the halo spread in all direction, so that it is much fainter and indeed quite harmless.

Zeiss in Jena, who had started with so little enthusiasm, slowly continued with the method. After some more of my visits, after some years of developing too complicated instruments and after further delay by the War, they brought out phase-contrast objectives and accessories in 1941.

I will end by showing some slides of photomicrographs, some I made with my own phase-strips, others from the Zeiss prospectus.