Phase Contrast, A New Method for the Microscopic Observation of Transparent Object

F. Zernike, Groningen

July 1942

PART 1

Every microscopist knows that transparent objects show light or dark contours under the microscope in different ways varying with change of focus and depending on the kind of illumination used. Curiously enough the wave theory of light has never been explicitly applied to the case of absolutely transparent objects, the details of which only differ in thickness or refractive index. This is carried through in the first part of the present paper, in which it is found that the theory is capable of explaining every detail of the microscopic aspect of transparent objects. At the same time a new method of observation is at once suggested by the results of our discussion. In Part II the effect of this method, which we call Phase Contrast, will be explained on a different line, which unlike the first treatment is not restricted to objects of simple structure such as gratings. Finally the construction of the requisite optical parts will be described and the appearance of a few objects by the new method as compared to the older methods shown by photomicrographs.

1 The Rayleigh and the Abbe treatment.

The wave theory of microscopic image formation leads to the conclusion that the microscopic image of illuminated objects is the result of a twofold diffraction [1]. Indeed the incident light first suffers diffraction on passing through the object and is diffracted again by the aperture of the objective. This means that every point of the object as well as every point in the aperture of the objective contributes to the vibration in a selected point of the image. In the calculation we find accordingly a twofold integration extending over the plane of the object as well as over the aperture. The practical bearing of this is usually made clear by considering each diffraction process (or integration) separately. If the second diffraction, by the objective, is considered in the first place, each point of the object is found to give rise to the well known Airy disc in the image plane. The peculiarities of nonluminous objects are then found at the second stage, the integration with respect to the object. Physically this means that the overlapping and interference of the diffraction patterns from different object points are taken into account. This is the well known way in which Lord Rayleigh [2] has presented the diffraction theory of the microscope. The other order in which the diffraction processes may be taken leads to the older treatment of Abbe [3]. He considered the diffraction by the object and the resulting diffraction images of the light source in the back focal plane of the objective as constituting the “primary” phenomenon. At the second stage the cones of light emerging from these diffraction images then overlap in the image plane and form the final image by their interference. This corresponds to the diffraction by the aperture of the objective. The methods of Lord Rayleigh and of Abbe are thus by no means controversial theories, they only give different physical pictures of the same general diffraction formula. Both methods must therefore always lead to the same final result. We shall follow the Abbe method as in our case it is found to be the simpler one, except
in a certain detail where the Rayleigh method will be preferred. The essentials of the well known Abbe theory will be reviewed briefly with the aid of fig. 1. The transparent object $P$ is illuminated by parallel light (source $S$ at infinity) and the lens system $L$ produces a magnified image at $P$ according to the laws of geometrical optics. If the object is a grating, as supposed by Abbe, the incident beam is split up by diffraction into several pencils which the lens focuses into as many diffraction images of the light source, $S_0$, $S_1$, etc. in its back focal plane $F$. According to Huygens’ principle these diffraction images may be taken as new sources from which cones of light emerge and cause interference fringes where they overlap. A calculation to be given presently shows that in the image plane $P$ the fringes form a pattern exactly similar to the object in structure and phase, which represents the image (Lummer, 1912). The exact similarity between image and object is attained only in the ideal case that all diffracted pencils enter into the microscope. As is well known, Abbe explained the limited resolving power by the limited number of diffracted beams in reality admitted. As we are not primarily concerned here with the limits of resolving power, but rather with the observation of the smallest contrasts, we shall not give special attention to the number of diffraction images, and shall even suppose the ideal case mentioned to be realized.

2 Calculations.

Our first step will be to calculate the amplitudes and phases of the beams diffracted by a transparent grating in parallel light. Let $x$ be a coordinate in the plane of the grating measured perpendicular to the bars from an arbitrary fixed point 0 as origin and let the unit of $x$ be chosen so as to make the period of the grating extend from $-\pi$ to $\pi$. In the most general case each surface element of the grating will change the light passing through it in amplitude and phase, which change may be characterized by a single complex number $f$. If a perpendicularly incident plane light wave of unit amplitude is represented by $e^{i\omega t}$, the emerging vibration will be given by $fe^{i\omega t}$, $|f|$ being the amplitude and arg$(f)$ the phase. The structure of the grating will be determined by the course of $f$ as a periodic function of $x$.

Now in order to calculate the direct light wave which is transmitted in the direction of the incident wave, we have to take the average of the vibrations emerging from the grating. We get exactly the same result by taking the average over one period of the grating

$$c_0 = \frac{1}{2\pi} \int_{-\pi}^{+\pi} f(x)dx$$

(1)

In order to calculate the phases and amplitudes of the spectra of different order, we notice that diffracted wavefronts are propagated only in such directions that the phases of the waves from contiguous periods of the grating show phase differences of $m\pi$ on the wave front. There for the diffracted light wave which gives rise to the $m^{th}$ order spectrum is represented by:

$$c_m = \frac{1}{2\pi} \int_{-\pi}^{+\pi} f(x)e^{imx}dx$$

(2)

Here $m$ may have all positive and negative integral values, corresponding to the different diffracted waves on the right (positive $m$) and on the left (negative $m$) of the direct wave. As is indicated in fig. 1 the
phases of the different waves are here compared for wave fronts going through the origin \(O\). The above mentioned grating is now placed under the microscope according to fig. 1. In the conjugate plane \(P\) we define a coordinate \(x\), expressed in such units that conjugate points in \(P\) and \(O\) have equal coordinates. We have now to find the interference effect of the diffracted waves in the plane \(P\), e.g. in \(O\). Now the optical paths from \(O\) to \(O\) through the objective are all equal. Therefore the relative phases of the different beams are the same in \(O\) as they were in \(O\), so that the resulting vibration in \(O\) is simply found by summing the complex numbers \(c_m\). To find the result for an arbitrary point of the image for which \(x = a\) we shift the origin \(O\) to \(x = a\) and find the new values \(c'_m\) from (2).

\[
c'_m = \frac{1}{2\pi} \int_{-\pi}^{+\pi} f(x) e^{im(x-a)} dx = e^{-ima} c_m
\]

The vibration in \(x' = a\) is then found by summing the \(c_m\)

\[
V = \sum_{-\infty}^{+\infty} e_m e^{-imx'}
\]

This expression together with (2) is nothing but the complex form of the Fourier series for \(f(x)\). This proves our former statement that the image is exactly similar to the object in structure and phase, at least in the ideal case that the sum in (3) is extended from \(-\infty\) to \(+\infty\).

3 Amplitude grating and phase grating.

The practical bearing of these formulae will become clear from a few particular cases. One extreme case is that of a grating which only alters the phases, leaving the amplitudes and therefore the intensities of the passing light unchanged. Let this be called a phase grating. It comprises all gratings ruled on transparent plates and in a more general way all structures of unstained microscopic objects (e.g. diatoms). For a phase grating \(f(x)\) has unit modulus but this does not correspond to any simple criterion for the \(c_m\). It is different, however, in case the changes of phase are only small. The real part of \(f(x)\) has then approximately unit value, independent of \(x\), and therefore cancels in \(c_m\) except in \(c_0\), while the imaginary part of \(f(x)\) clearly gives rise to values of \(c_m\) and \(c_{-m}\) with real parts of opposite sign, the imaginary parts being equal. The intensities of both \(m^{th}\) order spectra are again equal. In the following we shall restrict the term “phase-grating” to gratings causing small changes of phase. Fig. 2 illustrates these results. The phases are represented by the angles of the resp. vectors with the horizontal direction.

The most obvious difference between an amplitude grating and a phase grating, however, is not to be seen in the plane of the spectra, but in the image plane. Indeed the phase grating must be invisible in that plane, i.e. its image must show uniform intensity, being similar to the object. This means that the interference of the diffraction spectra, though these may be very conspicuous and very similar to those of an amplitude grating, only gives rise to differences of phase in the image.

[Diagram Figure 2:]

\(m = 2\)
For different reasons the invisibility is seldom as absolute as theory predicts. Indeed every deviation from the ideal conditions will disturb the exact balance of the different compounding vibrations. Certain larger deviations are used intentionally to get a clearly visible image of a phase grating. We shall treat successively: the “Schlieren” method, where the spectra on one side are intercepted, the oblique dark ground illumination, where the central image is intercepted also, central dark ground illumination, which intercepts the central image only, illumination by a narrow pencil, where slight changes of focus cause phase differences between the spectra, and finally the newly proposed method of phase contrast, where a phase difference is artificially introduced between the central image and the spectra.

4 Schlierenmethod.

If the left hand spectra in fig. 1 are intercepted, the vibration $V$, in the image plane will be given by

$$V = \sum_{-\infty}^{0} c_m e^{-imx'}$$

instead of by

$$V = \sum_{-\infty}^{+\infty} c_m e^{-imx'}$$

In order to take into account that we are observing a phase grating, we put according to the above results

$$c_0 = 1, c_m = -1_m + ib_m, c_{-m} = a_m + i b_m$$

and combine the terms with $c_m$ and $c_{-m}$

$$V = 1 + 2i \sum_{1}^{\infty} \left( a_m \sin mx + b_m \cos mx \right)$$

$$V_s = 1 + \sum_{1}^{\infty} \left( a_m \cos mx - b_m \sin mx \right)$$

+ $\sum_{1}^{\infty} \left( a_m \sin mx - b_m \cos mx \right)$  \hspace{1cm} (4)

In squaring these expressions we must remember that they are approximations for small values of the $a$s and $b$s, so that squares of these quantities should be neglected compared with first powers. Therefore we find for the intensities

$$|V|^2 = 1 \quad |V_s|^2 = 1 + 2 \sum_{1}^{\infty} (a_m \cos mx - b_m \sin mx)$$

Though it cannot be seen in the image, the imaginary part of $V$ in (4) clearly specifies the structure of the phase grating. The result is therefore that instead of the real structure represented by the series

$$\Sigma_1 = \sum_{1}^{\infty} \left( a_m \sin mx + b_m \cos mx \right)$$  \hspace{1cm} (5)

the schlierenmethod shows us

$$\Sigma_2 = \sum_{1}^{\infty} \left( a_m \cos mx - b_m \sin mx \right)$$  \hspace{1cm} (6)

Here each term equals the $m$th part of the derivative of the corresponding term of (5).

The appearance of the schlierenimage $V$, may be illustrated in the following way. Suppose we have a plaster relief of the real structure, differences of refractive index being rendered by small differences of height. The relief will become visible when illuminated obliquely, the slopes turned towards the light source showing brighter, those turned away from it showing darker. The differences of intensity will be proportional to the sloping angles, when these are only small. Under ordinary circumstances the structure of the plaster relief will be at once apparent to the observer, which means that he concludes from the observed brightness function to its integral, the height of the relief. In the same way the observer of the schlierenimage (6) will get the impression that he sees an obliquely illuminated relief, the height of which is given by the integral of (6):

$$\sum_{1}^{\infty} \frac{1}{m} \left( a_m \sin mx + b_m \cos mx \right)$$  \hspace{1cm} (7)

This is indeed rather similar to the real structure (5), with the difference, however, that because of the
factor $1/m$ the higher terms are relatively too small. Especially sharp edges will therefore be seen rounded off, as is clearly shown in fig. 3a,b in which the calculation of a simple case has been plotted.

\[
V_r = 1 + 2 \sum_{1}^{\infty} (a_m \cos mx - b_m \sin mx)
\]  

(8)

This improved form of the schlierenmethod therefore doubles the sensitivity. It may be remarked also that it has a different effect on an amplitude grating. As is easily found from our formulae, the original schlierenmethod does not materially alter the image of an amplitude grating, while the improved form will, to the same approximation, make it invisible. Our improved method will thus not emphasize differences of refraction in the object, but also reduce differences of absorption.

5 a. Oblique Dark Ground Illumination.

In this method only the spectra on one side contribute to the image, either because only these enter the objective, or because the other side and the central image are screened off. From (4) it follows that the intensity of the image is given by

\[
\left[ \sum_{1}^{\infty} (a_m \sin mx + b_m \cos mx) \right]^2 + \left[ \sum_{1}^{\infty} (a_m \cos mx - b_m \sin mx) \right]^2
\]  

(9)

Both sums (5) and (6) appear in this result. The first term here is the square of the real-structure and does not represent it very clearly as differences of sign disappear in the square. The second term is the square of the “Schlieman” function (6), which specially emphasizes abrupt changes in the structure. This is responsible for the well known dark ground effect in which all contours appear bright and broadened.
Fig. 3c gives the calculated result for the same grating.

In this case it makes no difference on which side the spectra are intercepted. Therefore obliquely illuminating pencils may as well come from all sides, as in the usual dark ground condensers.

b. Central Dark Ground Illumination.

The schlieren function will be absent if the central image alone is screened off. From (4) the intensity in the image is found to be

\[ 4 \left( \sum_{m=1}^{\infty} (a_m \sin mx + b_m \cos mx) \right)^2 \] (10)

This gives an image very different from that obtained by the oblique dark ground illumination, as may be seen from fig. 3d. Indeed, (10) is nothing but the square of the real structure (5) and gives therefore a much better representation of the object than (9). The advantages of this method have not been remarked by earlier workers such as Siedentopf who especially mentions the method. They were discovered by Spierer [5] in 1926, who found the images so different from the usual dark ground images that he did not even recognize his method as a kind of dark-ground illumination. More recently American investigators [6] have continued the work of Spierer. From their statements and even more from the photomicrographs they reproduce the Spierer method must be identified with central dark ground illumination.

c. Narrow central pencil.

Of course it is the ordinary practice to use a narrow central pencil in observing transparent objects, as the image tends to disappear if the condenser diaphragm is opened. According to the above it is not the invisibility of the transparent object but on the contrary the appearance of an image with a narrow diaphragm that needs a special explanation. With most objects the observer will not readily find out that the image only appears when slightly out of focus, because different details lie in different planes. Artificial plane objects and also diatoms with faint markings, however, show clearly the disappearance when in exact focus, and the inversion of the image from positive to negative when passing from inside to outside focus.

Our theory explains this behavior as follows. According to it, the invisibility of the phase grating depends on the phase relations represented in fig. 2. Any changes of the relative phases of the spectra will cause the intensity in the image to vary, and will thus cause an image that gives some indication of the structure of the object. Now such phase differences are already brought about by changing the focus, which amounts to shifting the plane in which the interference is observed. If the plane is moved a small distance \( l \) towards the objective, the optical path is evidently diminished by \( l \cos \theta \) for a ray making an angle \( \theta \) with the axis. As \( \theta \) is only small, path differences with the central ray of \( \frac{1}{2} \theta l^2 \) are thus introduced. In other words, the phase of the \( m^{th} \) term in the series is retarded proportional to \( m^2 \). If the first order spectrum is retarded by \( \gamma \), the resulting vibration \( V_\gamma \) will be

\[
V_\gamma = \sum_{m=-\infty}^{+\infty} c_m e^{-im^2 \gamma} e^{-imx}
\]

\[ = 1 + 2 \sum_{m=1}^{\infty} \left( a_m \sin mx - b_m \cos mx \right) \sin m^2 \gamma \]

\[ + 2i \sum_{m=1}^{\infty} \left( a_m \sin mx + b_m \cos mx \right) \cos m^2 \gamma \]

and the intensity

\[ |V_\gamma|^2 = 1 + 4 \sum_{m=1}^{\infty} (a_m \sin mx + b_m \cos mx) \] (11)

This result has been applied to our test case for two values of \( \gamma \) (fig. 3e,f). As the coefficients \( \sin m^2 \gamma \) and \( \cos m^2 \gamma \), introduced by this method, change rapidly with \( m \), the image gives no clear idea of the detailed structure of the object.

An experimental proof of the necessity of out of focus observation may be found by using annular diaphragms of different size in the condenser. A
specimen of Pleurosigma angulatum in balsam was observed with an ordinary objective of 0.85 aperture. The hexagonal markings were beyond the resolving limit if there was a narrow central stop in the condenser, but began to appear with a stop of 9 mm diameter. With an annular stop of 10 mm mean diameter the markings appeared very clearly, positive with one adjustment, negative with a higher one. With annular stops of $12 \frac{1}{2}$-15 mm the markings were hardly visible, but they were again very clearly seen when the annulus had a diameter of 17 mm or more. In this case the positive and negative images were seen in reverse order. By removing the eyepiece, the direct light and the spectra were found to lie as in fig. 4. In the first case the spectra are farther from the axis than the direct light, just as in the case of a central illuminating pencil. In the second case, b, both are on the average at the same distance from the axis, so that a change of focus cannot introduce the difference of phase necessary according to the above theory. The observed effect would not appear with an amplitude grating, i.e. with markings consisting of absorbing matter.

The observed effect would not appear with an amplitude grating, i.e. with markings consisting of absorbing matter. The above different methods for observing phase gratings have all been introduced apart from their explanation by the wave theory. On the basis of the theory here given a more obvious method is to try to bring the spectra and central image of fig. 2b to the case of fig. 2a. This is effected in the figure by turning the vectors representing the spectra through $90^\circ$, as indicated by the dotted lines. The phase grating will therefore appear as an amplitude grating of exactly corresponding structure when all the spectra as a whole are changed in phase by $90^\circ$ with respect to the central image. Indeed our formulae (4) and (5)

$$V = 1 + 2\Sigma_1$$

will in this way be changed into

$$V_p = 1 \pm \Sigma_1$$

giving for the intensity

$$|V_p|^2 = 1 \pm 4\Sigma_1$$

which according to (5) is exactly what we want to see. Here the upper sign will hold when the spectra have been retarded. This may be called positive phase contrast. When the spectra are advanced in phase the negative sign in (12) must be taken (negative phase contrast). The results are given in fig. 2g,h.

It is easily possible to introduce this change of phase. Let a glass plate be placed in the back focal plane ($F$ in fig. 1) which carries a thin spot of varnish at the point $S_0$, so that the direct light from the light source goes through this varnish, while the diffracted light passes besides. In this way the optical path of the central image is clearly increased by ($n - 1)d$ and this can be made equal to $\frac{1}{4}\lambda$, its phase being thereby retarded by $90^\circ$. This is the same as the second case above (spectra advanced). To get the positive phase contrast the layer can be made three times thicker, retardation of 3. $90^\circ$ being equivalent to an advance of $90^\circ$. Alternatively the varnish can be made to cover the spectra only, a small hole being scratched out at $S_0$.

With the positive phase contrast thicker or higher refracting parts of the grating – corresponding to negative values of the sum $\Sigma_1$ – will appear darker, in the negative case brighter, than the mean illumination.

The phase contrast method as here described is clearly much better and more sensitive than the observation with narrow pencils or the schlierenmethod. It also gives a much better indication of the exact structure of the object than the dark ground methods. The latter however are ordinarily made more sensitive for showing very small differences in refraction, by greatly increasing the intensity of the incident light. Now the phase contrast method may also
be improved in this way, provided the central pencil be afterwards weakened. If the small spot of varnish at $S_0$ is made absorbing, the amplitude of the direct light only will be lowered to a fraction $a$, the result in the image being, instead of (12),

$$|V_{pa}|^2 = a^2 \pm 4a\Sigma_1 \quad (13)$$

so that the relative differences of brightness have been increased in ratio of $a$ to 1. Thus the sensitivity will be increased three times by weakening the direct light to one ninth of its intensity, while the spectra pass unabsorbed.

References
(5) C. Spierer, Arch. SC. phys. nat. 8, 121, 1926.
(7) F. Zernike, Physik. Z. 38, 994, 1937.