The invention of the phase contrast microscope by Frits Zernike

Joseph Braat
Emeritus professor of optics, Department of Applied Physics,
Delft University of Technology, Delft, The Netherlands

This text is a faithful English translation of the original publication (in the Dutch language):
Braat, J. J. M., De uitvinding van de fasecontrastmicroscoop door F. Zernike,
Ned. Tijdschr. Natuurk. **89**(11):32-37, 2023.

Introduction

Frits Zernike has left his mark in several important research areas of optics such as diffraction theory, coherence theory, aberration theory of optical imaging systems and optical metrology. He has received much recognition from his peers for these various research topics. But only one subject/instrument has really brought him international fame, the phase contrast microscope, invented by him in 1932. The 'conventional' microscope used until then only detected absorption changes or reflection variations of an object (*amplitude* objects). The phase contrast microscope, on the other hand, allows sharp and contrast-rich observation of weak density variations in transparent media or of very small height differences on reflective or transparent surfaces (*phase* objects). Physicians and biologists studying cellular structures, tissues, or bacteria could not believe their eyes at first when they replaced a conventional lens with its accompanying lighting unit in the microscope by the phase-contrast alternative and then looked through the microscope.

1. The road to the phase contrast microscope

Before delving into the principle of Zernike's phase contrast method, let us first describe the quality and possibilities of microscopic observation over a hundred years ago.

1.1 **The quality of the microscopic image around 1900.**

It is interesting to compare the optical quality of microscopes at the turn of the nineteenth century with the performance of a microscope today. Around 1900, the Zeiss company was a leader in the field of microscopy, thanks to the driving scientific and organizational strength of its director Ernst Abbe (1840-1905), the successor of the founder Carl Zeiss (1816-1888). As early as 1877 this firm produced a conventional microscope lens of superior quality for use in oil immersion and which had the very high numerical aperture (*NA*)[1] of 1.30. This high *NA*-value (the refractive index of immersion oil is about 1.55) is clearly comparable to the current maximum *NA* values in microscopy.

Microscope objectives from that period already had good image quality thanks to the advanced optical metrology developed by Abbe and his associates in the company. What was sadly lacking in a 19th-century lens were anti-reflective layers on the lens surfaces, which did not become commonplace until the 1950's. In a high-NA microscope, about twenty glass-to-air passages from the object to the eyepiece are not unusual. This leads to an overall loss of transmission of the lens and, in addition, by light reflected an equal number of times, to a significant background illumination in the formed microscopic image with loss of contrast as a result.

In addition, the conventional transmission microscopes of the time suffered from the limitation that they were only suitable for detecting light absorption variations in (partly) transparent materials

---

[1] The aperture value of a lens is given by the sine of the angle between the outermost light rays in the light beam captured by a lens and the optical axis of the lens. To obtain the *numerical* aperture, the aperture value is multiplied by the refractive index n of the medium in which the object is to be immersed. From optical imaging theory it then follows that the (vacuum) wavelength λ of the light, divided by the value of $2(NA)$, corresponds to the smallest period that can just be observed.

(*amplitude* objects). In reflection microscopy, only local variations of the reflection coefficient of the object's surface were made visible. One could also observe transitions in an object where strong light scattering occurs. Details in a phase-type object went largely unnoticed in the conventional microscope.

1.2 **Abbe's imaging theory for the microscope.**

A high-quality instrument such as a microscope also requires an imaging theory to reconstruct the object from the observed or measured microscope image. That is why we first give here a theoretical description of the resolution of the microscope as developed by the pioneer in this field, E. Abbe [1]. Fig. 1 shows the essence of his imaging theory. An incident parallel beam of light is split by a flat periodic object in B into three parallel sub-beams (diffraction orders) which are focused at the back focal plane of the lens L.

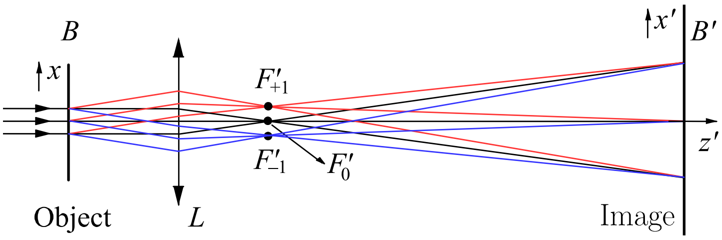

Fig. 1. Beam path through a microscope to the intermediate image in B' according to Ernst Abbe. The beams diffracted by the object are coloured red and blue to distinguish them from the zero-order beam.

The three spherical waves with their convergence points in F'_0, F'_{+1} and F'_{-1} give rise to overlapping interference patterns in the plane of the sharp intermediate image B', producing an intensity distribution that resembles the transmission function of the original object in plane B. The innovation in Abbe's theory was the introduction of a periodic grating structure as a building block for a more complex object. And a 'black-and white' optical transmission grating fits perfectly into his model.

1.3 **Zernike's research on a reflecting phase grating.**

Reflective grooved gratings have been made since the beginning of the 19th century (Fraunhofer). In a glass substrate, straight grooves were cut with a diamond cutter to a depth of $\ll 1$ µm. From cut to cut, the substrate was moved laterally with submicrometric precision. A very thin layer of silver was then applied to the engraved glass surface by a chemical deposition process. The final product was a uniformly reflective surface with a periodic relief of very shallow depth. A beam diffracted by such a gating could be used in high-precision spectroscopy. The reflection grating affects only the phase of the incident light and is an optical phase object.

In his Nobel Prize nomination [2], Zernike mentions that since 1920 his interest in reflecting phase gratings had been aroused and that all sorts of speculative claims were made about the behaviour of this type of gratings. Around 1930 he was given a carefully made spectroscopic phase grating, and he says in his lecture that, as a *physicist with an interest in optics*, he had made some calculations and experiments that gave him an understanding of the behaviour of a phase grating. In [2] he briefly discusses the two experiments with the phase grating that paved the way for him to optical imaging with *phase contrast*.

a) In a first test of the grating, he immediately went in search of the dreaded 'Rowland ghosts'. These manifest themselves optically as weak satellites of the first or higher diffraction orders of a grating and can thereby reduce the spectral resolving power of the grating. The Rowland ghosts or satellites can be attributed to periodic irregularities in the position of the drawn grooves. These are caused by imperfections in the reduction mechanism for the table movement of the cutting machine. Zernike first noted that the unwanted satellites were sharply imaged in a different focal plane than that of the main spectral line. He states that the difference in optimum focus, observed by him between a main spectral line and its ghosts, merits further investigation. But he's already assuming that there's a 90-degree phase shift between the diffracted light going to the desired spectral line and the unwanted diffracted light going to the satellites. In a later experiment, he succeeded in establishing a difference in optical path of λ/4 between the desired spectral main beam and its satellites. The result was that the main spectral line and the satellites were optimally imaged in the same focal plane, a phase contrast image 'avant-la-lettre'. The point- or slit-interferometer proposed by Zernike for the introduction of, for example, a phase difference of 90 degrees, has resurfaced in 1974 in the literature, now carrying the name of R.N. Smartt.

In retrospect, Zernike's theoretical description of the 'Rowland ghosts' must have been identical to the analysis related to frequency modulation of, say, a radio carrier wave to broadcast a sound signal. Strong frequency modulation results in so-called Bessel sidebands in which all uneven sidebands are in phase quadrature with the carrier wave.

b) In the second test of the spectroscopic grating, Zernike imaged the grating surface itself. This test with the grating proved essential for the later invention of phase contrast imaging. When focusing on the grating surface with a magnifying telescope, Zernike noticed that the focus setting for a scratch in the very thin silver layer was clearly different from the focusing required to optimally visualize the periodic structure with the shallow grooves. This topic, different behaviour in the imaging of amplitude and phase structures with respect to the optimum focus setting, was noted by him as important for later research. In the end, this 'later research' took place several years later, but then quickly led to the fundamental invention of microscopic imaging with phase contrast.

1.4 **Geometrical description of a phase grating and the corresponding diffraction orders**.

In his first publication on phase contrast imaging [3], Zernike gives a global description of the method. In what follows, we will focus as an example on a one-dimensional phase grating with two discrete levels whose surface shares can still be freely selected via the 'duty-cycle' β. In Fig. 2 the (uniform) phase of a planar light wave, incident on a phase grating, is outlined in the incident plane A, perpendicular to the propagation direction of the wave. The grating is used here in transmission.

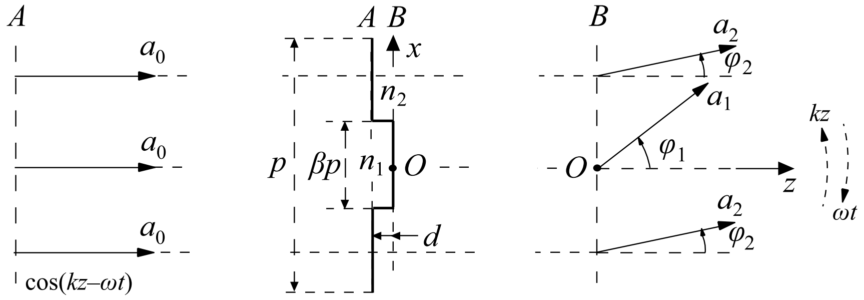

Fig. 2. Left: incident parallel beam A with uniform phase and phase vector a_0; the phase of the wave is given by $(kz - \omega t)$ and is set to zero in plane A.
Centre: Cross-section, perpendicular to the grating lines, of a period with length p of a pure phase grating with two heights (height difference d, greatly exaggerated in the figure if the phase grating is optically 'weak').
Right: the spatial phase change in the near field (exit plane B). The discrete phase changes are denoted by ϕ_1 and ϕ_2 $(n_1 > n_2)$; the (rotated) phase vectors in plane B by a_1 and a_2.

The phase difference between the two levels with height difference d within a period p of the grating is given by

$$\delta = \phi_1 - \phi_2 = (n_1 - n_2)\, kd, \qquad (1)$$

where the wave number k is equal to $2\pi/\lambda$. In the absence of absorption, the phase vectors have identical lengths in plane B.

The periodic phase object, illuminated by a parallel beam, gives at a sufficiently large distance (in optical jargon one speaks of the 'far field') a collection of separate diffracted beams, plus the original beam with modified amplitude and phase. In the back focal plane of the lens L of Fig. 1 the various focal points of the three considered parallel beams arise. For the understanding of the phase contrast image, it is important to know not only the amplitude of the diffracted beams, but also their phase.

The direction of propagation of a diffraction order complies with the 'grating law' in the case of perpendicular illumination,

$$\sin\alpha_m = m\,\lambda/p\ , \tag{2}$$

where α_m is the angle with the z-axis and m is the order number of a diffracted beam.

The amplitude and phase of an m^{th} order beam in the far field is obtained by multiplying the transmission function $t(x)$ of the grating by the amplitude function of a plane wave with a propagation direction given by α_m in the exit plane B of the phase grating (Fourier transform). Without giving the details of the optical Fourier transform, we present in Fig. 3 the value of the phase difference $\Delta\theta_{\pm1,0}/\pi$ between a first- and a zero-order beam of the phase grating in Fig. 2, as a function of the phase diff cycle' β of the central part of a period. For all the values of $|\delta/\pi|$ ‹‹ 1 we can approximate the value of $|\Delta\theta_{\pm1,0}|$ to $\pi/2$. For the value β=1/2 it's even true that $|\Delta\theta_{\pm1,0}| = \pi/2$, irrespective of the value of δ.

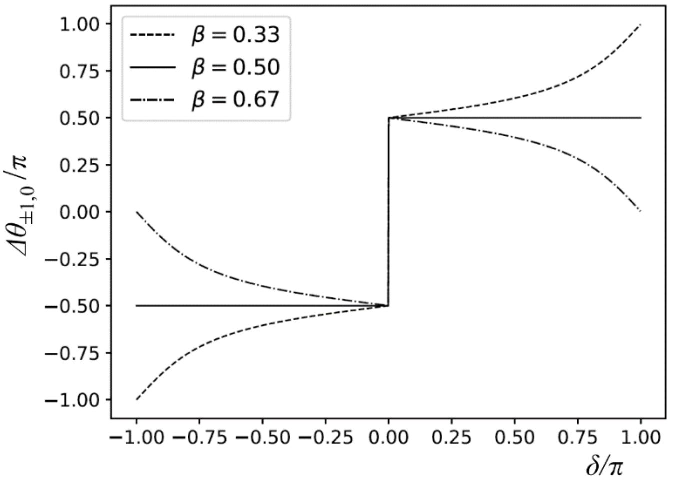

Fig. 3. The phase difference $\Delta\theta_{\pm1,0}$ between the zero-order beam and a first-order diffracted beam as a function of the phase difference in the 'near field' of the grating (plane B). The grating is symmetric with respect to point O.

1.5 **Periodic intensity distribution in the intermediate image (plane B').**

Zernike explained the poor contrast or even complete absence of (weak) phase gratings in the image at B' by the mutual shifting of the interference patterns between the 0th and +1st order beams on the one hand and the 0th and -1st order beams on the other hand. Their location-dependent intensity contributions are given in the intermediate image plane B' by respectively $\mathrm{I}_{+1,0} \propto \cos\left(\Delta\theta_{+1,0} + \frac{2\,\pi x'}{p'}\right)$ and $\mathrm{I}_{-1,0} \propto \cos\left(\Delta\theta_{-1,0} - \frac{2\,\pi x'}{p'}\right)$, where the ratio x'/x is given by the optical magnification between the intermediate image in B' and the object plane in B and p' is the (magnified) period in the plane B'. The two periodic intensity distributions are shifted relative to each other over a phase angle $\Delta\theta_{+1,0} + \Delta\theta_{-1,0}$. Since the detected intensity contains the sum of the two interference patterns and $\Delta\theta$ is identical for first diffraction orders of the chosen phase grating, we find for this sum the expression,

$$I_{+1.0} + I_{-1,0} \ \propto \cos\left(\Delta\theta_{\pm1,0}\right)\cos\left(\frac{2\pi x'}{p'}\right)\ . \tag{3}$$

The special case $\Delta\theta_{\pm1,0} = \pm\pi/2$ thus only produces a uniform background since the two periodic interference patterns are shifted relative to each other over exactly half a period. Thus, such a phase grating is not visible in a conventional microscope at optimum focus!

1.6 **Modification of the phase difference between zero and first order beams.**

From the above intensity calculations, Zernike found that the experimental observation of enhanced contrast in a defocused plane can be easily explained by the altered distance difference between 0th and 1st order. This effect is illustrated in Fig. 4. The point *E* is in a displaced plane B_d on which the lens *L* of Fig. 1 is focused. The object is unchanged in plane *B*.

For the phase difference $\Delta\theta_{+1,0,d}$ between the +1st and the 0th order beam in *E* at a defocus distance $OE{=}\Delta z$, we find the expression,

$$\Delta\theta_{+1,0,d} = \Delta\theta_{+1,0} + k(OG - OE) = \Delta\theta_{+1,0} + k\,(\Delta z)\,[\cos\alpha - 1\,]\,. \tag{4}$$

If the phase term caused by the defocus distance Δz is equal to $\pm\pi/2$, we find that

$$\Delta z = \mp\left(\frac{\lambda}{4}\right)\left\{\frac{1}{2\,\sin^2(\alpha/2)}\right\}. \tag{5}$$

An original phase difference between the diffraction orders in *O* of $\pi/2$ can now effectively be changed to 0 or π. The periodic intensity distribution in the adjusted microscope image plane is now visible again. Given the intensity contributions of the interference patterns between the 0th and +1st order and the 0th and −1st order, an increase or decrease of their mutual displacement by half a period *p* occurs by a the introduction of a defocusing Δz according to Eq.(5). With a further increased defocusing, the grating image disappears again, appears anew as a sharp image with opposite contrast, and disappears when $2\pi = k(\Delta z)\{\cos\alpha - 1\}$. This phenomenon continues with further increasing defocusing and immediately also explains the phenomenon of the multiple re-images of an amplitude grating that were already observed by Talbot [4] in 1836. Note that defocusing on a phase grating may be useful in making visible one specific grating frequency; but this does not work globally for a general object that is made up of a wide spectrum of grating frequencies.

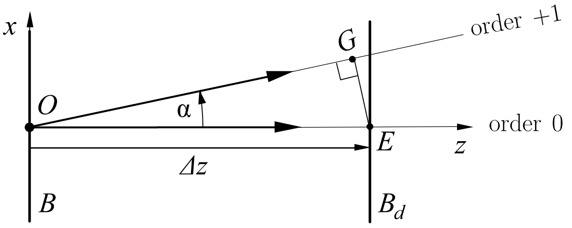

Fig. 4. Calculation of the modified phase difference between the zero and first diffraction orders at defocused observation of a grating situated in plane *B*. The microscope is focused at the point *E* in the plane B_d that is found at a distance Δz from the plane *B*.

1.7 **The final design of the phase contrast microscope.**

The ability to visualize a phase grating by defocusing was undoubtedly the guiding thought for Zernike when designing the phase contrast microscope in 1932.

For the visualization of phase structures, he has chosen the solution which leaves the phase of the diffracted light unchanged and changes the phase of the 0th order, thus producing the desired phase difference in a single step, irrespective of the diffraction angle α of the first order beams. His first technical implementation in 1932 was the introduction of a permeable glass substrate behind the objective lens *L*, where one of the glass surfaces coincides with the back focal plane at F'_0 in Fig. 1. By etching the glass surface in a small disc-shaped area at the focal point, a technique well mastered by Zernike, the required phase difference of $\pi/2$ is introduced (negative in this case) according to Fig. 1.

Zernike later introduced improvements to increase image contrast, such as,

- the replacement of the central phase-contrast 'disc' by a narrow phase-contrast annular region or 'phase ring' to minimise artificial disturbances at the edge of large continuous-phase areas,
- the partially absorbing phase-contrast ring in the back focal plane of the lens to increase the contrast in the image,
- the adjustable phase contrast using polarization optics,
- the 'coloured' phase contrast for a more precise measurement of the occurring phase differences.

The impressive improvement in the imaging of (weak) phase objects using the phase contrast method is illustrated by Fig. 5. Here I show 'through-focus' images of a phase object in which some (black) absorbing perturbations are also present in the microscopic light path, as reference objects. The two lenses used, one conventional, the other equipped for phase contrast imaging, both have an *NA* of 0.65. The positive and negative defocus distance Δz in the figure is equal to approximately one depth of field (equivalent to 0.55 µm at an average wavelength λ of 0.55 µm). For example, using the conventional objective lens, observe the change in contrast of a phase structure when going through focus, and also the virtually complete absence of the phase structures in optimum focus! Fig. 5 also shows that phase contrast imaging is not only useful for physicians or biologists but that the method is also proving useful in optical metrology in the nanometric range.

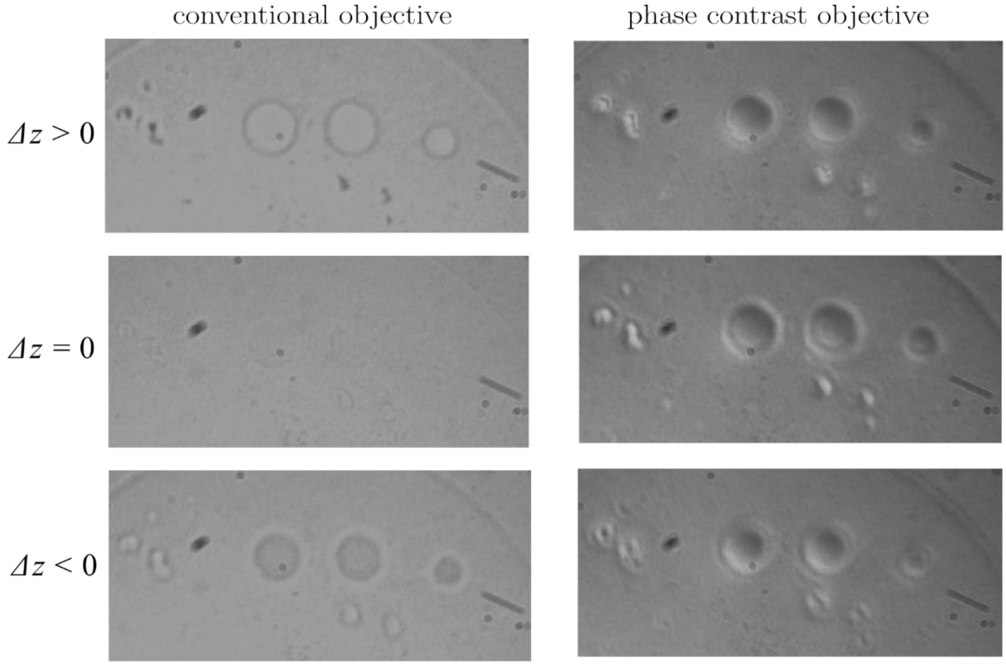

Fig. 5. Images of a phase object. Left: image in a conventional microscope. Right: image in a microscope with phase contrast. Δz is the defocus distance. The most conspicuous circular phase objects are tiny 'shrinkage dents' formed in the surface of an injection moulded plastic plate during the cooling process.

2. The international breakthrough of the phase contrast microscope.

Zernike was convinced that the further development of his invention required industrial participation and sought contact with the Zeiss company. Unfortunately, his phase contrast method was dismissed there as an academic curiosity. In his Nobel presentation, Zernike says that one of the older Zeiss employees literally said that "if the invention had any practical use, it would have been invented by us long ago". In the late 1930's, an interesting contact was made with Dr. Caroline Bleeker [5]. She had set up a fine-mechanical workshop for the manufacture of scientific instruments in Utrecht. But World War II prevented further developments in the Netherlands.

Immediately after World War II, Zernike tried to restart industrial production in the Netherlands through cooperation with Caroline Bleeker's company. Successfully this time, because until the end of the seventies of the last century, high-quality phase contrast microscopes were produced there, in the company now called Nedoptifa. Utrecht University recently honoured Bleeker's activities with a mural painting, see Fig. 6. Since the Utrecht scientific mural paintings are intended to illustrate a formula, the 'symbolic' expression $P=S+D$ has been used, which means that the obtained image wave P (Phase contrast) is the sum of the background wave S (Shifted wave), shifted in phase, and the diffraction waves D generated by the phase object.

Fig. 6. Wall painting in the Strosteeg of the city of Utrecht about the successful series production of phase contrast microscopes (from 1951 on) by the company Nedoptifa of Dr. C.E Bleeker. At the bottom right, at the phase contrast microscope, the adjustments in the illumination system and in the focal plane of the objective lens are shown schematically with the annular version of the phase plate.

Through licenses, optical companies such as Leitz and Nikon later began to produce phase contrast microscopes in addition to Zeiss. The award of the Nobel Prize has brought about a marked acceleration in this respect. The phase contrast microscope has acquired a permanent position in research areas and test environments where objects with small density variations or a very small surface relief are present (medical science, biology, bacteriology, study of living tissues and cells, optical metrology).

3. **Nobel prize winner**

Now it remains for me to give a possible explanation for the somewhat mysterious postponement of the Nobel Prize research which Zernike mentions a few times in his notes and in his Nobel Prize speech. What could have been so important that it delayed the phase contrast study? In the period 1920-1930 statistical mechanics of atomic and molecular systems formed his main research in Groningen. Optical subjects of his interest were thus in a sort of waiting queue. It was not until the end of that decade that they were fully implemented. The result of this can be read in several sublime optical publications by Zernike in the thirties and forties in which he successively presents:

- diffraction theory of the Foucault knife-edge test [3] of 1934, using the phase contrast method, to make the detection of polishing defects on telescope mirror surfaces much more sensitive,

- additions to and physical interpretation of the optical coherence analysis by P. H. of Cittert (1934) in a 1938 publication by Zernike, resulting in the van Cittert-Zernike theorem,
- the orthogonal circle polynomials of Zernike [3] that have become indispensable in applied optics [6],
- the still widely used and recently expanded Nijboer-Zernike diffraction theory [3][7][8], from the years 1934 to about 1950, together with the Ph.D. students Nijboer (1942) and Nienhuis (1948).

Finally, I would like to express my gratitude to Thim Zuidwijk of the Optics Section of the Delft University of Technology, who facilitated my access to the phase contrast microscope. He also pointed me to several literature references about Caroline Bleeker.

References

[1] Abbe, E., Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung, *Arch. Mikrosk. Anat.*, **9**(1):413-468, 1873. DOI: 10.1007/BF02956173.

[2] *Nobel Lectures, Physics* 1942-1962, Elsevier Publishing Company, Amsterdam, 1964. See: https://www.nobelprize.org/prizes/physics/1953/zernike/lecture/

[3] Zernike, F., Beugungstheorie des Schneidenverfahrens und seiner verbesserten Form, der Phasenkontrastmethode, *Physica*, **1**:689-704, 1934. DOI: 10.1093/mnras/94.5.377 (in English).

[4] H.F. Talbot Esq. F.R.S., LXXVI. *Facts relating to optical science*. No. IV, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, **9**:56, 401-407, 1836. DOI: 10.1080/14786443608649032.

[5] van Ginkel, G., *Dr. Caroline Emilie Bleeker en de Nederlandse Optiek- en Instrumentenfabriek Dr. C.E. Bleeker*, Fylakra Limited Editions, Faculteit Natuur- en Sterrenkunde, Universiteit Utrecht, Utrecht, 1997. PDF-file can be obtained via https://web.science.uu.nl/FylakraFiles/Bleeker_Biografie_Gijs_van_Ginkel.pdf
See also: van der Heijden, M., *Een eeuw natuurkundevrouwen: Caroline Emilie Bleeker en Marjolein Dijkstra, Ned. Tijdschr. Natuurkd*. **87**(2):32, 2021.

[6] Braat, J. and Török, P., *Imaging Optics*, Cambridge University Press, Cambridge, U.K., 2019. DOI: 10.1017/9781108552264.

[7] Nijboer, B. R. A., *The diffraction theory of aberration*s. Noordhoff Editors, Groningen, 1942.

[8] For the Extended Nijboer-Zernike diffraction theory, see https://www.nijboerzernike.nl.