SPIE.

ERNST ABBE'S
Theory of Image Formation in the Microscope

Written and published by Otto Lummer and Fritz Reiche under the title
Die Lehre von der Bildentstehung im Mikroskop von Ernst Abbe
Translated and annotated, with additional material,
by Anthony Yen and Martin Burkhardt

ERNST ABBE'S
Theory of Image Formation in the Microscope

ERNST ABBE'S
Theory of Image Formation in the Microscope

Written and published by Otto Lummer and Fritz Reiche under the title
Die Lehre von der Bildentstehung im Mikroskop von Ernst Abbe
Translated and annotated, with additional material,
by Anthony Yen and Martin Burkhardt

SPIE PRESS
Bellingham, Washington USA

Library of Congress Cataloging-in-Publication Data

Names: Abbe, Ernst, 1840–1905, author. | Lummer, O. (Otto), 1860–1925, editor. | Reiche, F. (Fritz), 1883–1969, editor. | Yen, Anthony, translator, annotator. | Burkhardt, Martin, translator, annotator.
Title: Ernst Abbe's theory of image formation in the microscope / written and published by Otto Lummer and Fritz Reiche; translated and annotated, with additional material, by Anthony Yen and Martin Burkhardt.
Other titles: Die Lehre von der Bildentstehung im Mikroskop von Ernst Abbe. English | Ernst Abbe's theory of image formation in the microscope
Description: First printing. | Bellingham, Washington : SPIE, [2022] | "This work was originally published in 1910, five years after Ernst Abbe's death. The original book, published by Friedrich Vieweg und Sohn, was compiled by Otto Lummer, professor of physics at the University of Breslau, and his then-assistant Fritz Reiche" – translators' foreword. | Includes bibliographical references and index.
Identifiers: LCCN 2022021996 | ISBN 9781510655232 (paperback) | ISBN 9781510655249 (pdf)
Subjects: LCSH: Microscopy.
Classification: LCC QH205 .A213 2022 | DDC 570.28/2–dc23/eng/20220720
LC record available at https://lccn.loc.gov/2022021996

Published by
SPIE
P.O. Box 10
Bellingham, Washington 98227-0010 USA
Phone: +1 360.676.3290
Fax: +1 360.647.1445
Email: books@spie.org
Web: http://spie.org

Cover image courtesy of ZEISS Archives.

Printed in the United States of America.
First printing 2023.
For updates to this book, visit http://spie.org and type "PM352" in the search field.

SPIE.

Ernst Abbe's Theory of Image Formation in the Microscope

Written and Published by Otto Lummer and Fritz Reiche under the title

Die Lehre von der Bildentstehung im Mikroskop von Ernst Abbe

Translated and annotated, with additional material,

by Anthony Yen and Martin Burkhardt

Translators' foreword

This work was originally published in 1910, five years after Ernst Abbe's death. The original book, published by Friedrich Vieweg und Sohn, was compiled by Otto Lummer, professor of physics at the University of Breslau, and his then-assistant Fritz Reiche. The book is an expanded version of notes taken by Lummer, who attended Abbe's lectures on the subject in Jena in 1887. It is the only detailed publication of Abbe's theory on image formation in the microscope. The entire book is based on classical optics; hence the arguments made in it are still valid today. In particular, the concept of a coherently illuminated image as two back-to-back diffraction processes of the object was beautifully described in it for the first time. In the section "Imaging of illuminated objects" in his 1933 classic textbook *Optik*, Max Born stated clearly that "this theory was developed by Abbe and was demonstrated by beautiful experiments. See E. Abbe, *Theory of Image Formation in the Microscope*." Frits Zernike began his famous 1934 article on phase-contrast imaging by stating "On the basis of Abbe's diffraction theory of optical imaging," and referred to this book in the article whose English translation can be found in the *Journal of Micro/Nanopatterning, Materials, and Metrology (JM^3)*, published by SPIE.

We believe the book is of more than historical value. Today, in partnership with ASML, ZEISS enables image formation at the finest resolution; it is the manufacturer of the imaging optics for extreme ultraviolet (EUV) lithography that employs the 13.5-nm wavelength radiation generated by laser-produced plasma. While these lithographic systems, used in the fabrication of the most advanced integrated semiconductor circuits, can resolve close to 20 nm in pitch, they are governed by the same underlying physics of image formation as microscopes of the late 19th century. The reader can therefore learn projection imaging directly from the master himself! Also, the book is inspirational, as it discusses several very innovative topics for that time. One example is off-axis illumination for improving the resolution in the imaging of two neighboring slits. Another example is a π-phase edge in the middle of an otherwise transparent slit, resulting in zero light intensity there in its image. And of course, there is the description of coherent imaging as two back-to-back diffraction processes. Abbe did not come up with the concept of partial coherence. Yet it states in the book that for incoherent imaging, the resulting intensity is obtained simply by summing the intensities generated by individual luminous points. We all know that one of the popular methods of calculating the aerial image in microlithography is done exactly this way, which is appropriately called the Abbe method.

Ernst Abbe received his doctorate from the University of Göttingen in 1861 under Wilhelm Weber. In 1866, at the invitation of Carl Zeiß, owner of Zeiss Works in Jena, Abbe became the research director there. He made numerous improvements to the performance of ZEISS microscopes based on physics rather than trial and error. In 1878, he built ZEISS' first immersion microscope. In 1873, he published the famous resolution formula $d = \lambda/(2n \sin \vartheta)$ (see below). After Carl Zeiß' death in 1888, Abbe placed the company under the Carl Zeiss Foundation that he established, with himself at the helm. Abbe also had held an academic position at the University of Jena since 1863. He died in 1905 at the age of 64.

We also want to mention the two distinguished scientists who compiled the original book. Otto Lummer received his doctorate under Hermann von Helmholtz at the University of Berlin in 1884 and was the latter's assistant for three years. From 1887 to 1904, he was a member of the scientific staff at the Imperial Physical Technical Institute [today's Physikalisch-Technische Bundesanstalt (PTB), similar in function to the National Institute of Standards and Technology (NIST) in the US]. He was appointed professor at the University of Breslau (in today's Wrocław, Poland) in 1905. Lummer worked mainly in the field of optics and thermal radiation. He developed a mercury vapor lamp and, together with Wilhelm Wien, constructed the first blackbody radiator and used it, together with Ernst Pringsheim, to conduct fundamental investigations of the spectral energy distribution of the blackbody radiation that led Max Planck to his quantum hypothesis. He died in Breslau in 1925, aged 64. At the time of the compilation of this book, Fritz Reiche was Lummer's assistant in Breslau. Reiche attended the University of Munich in 1901, but in the following year he transferred to the University of Berlin, where he received his doctorate under Max Planck in 1907. After a three-year stay with Otto Lummer in Breslau, he returned to the University of Berlin in 1911. In 1913 he became a lecturer at Berlin and worked and taught under Planck. Succeeding Erwin Schrödinger, he became a professor at the University of Breslau in 1921, the same year in which his book *The Quantum Theory* appeared. As a Jew, he was dismissed from his academic position by the national socialist government of Germany in 1933. With the help of many people, but mainly Rudolf Ladenburg, he was eventually (as late as 1941) able to leave Germany for the United States, where he held several academic positions and worked on supersonic flow and electromagnetic theory. He died in 1969 at the age of 85.

Annotations are given throughout this translation to make the points clearer, to provide relevant background information, and to correct certain errors, especially calculational errors; they are indexed with superscript Roman numerals throughout the text. Martin Burkhardt redrew all the figures in the text and typeset the translation with LaTeX. Anthony Yen furnished short write-ups on geometrical optics and on the resolution limit in the imaging of periodic patterns, adapted from his two *JM*³ articles on the subject of the resolution formula $d = \lambda/(2n \sin \vartheta)$; these articles attempt to clarify Abbe's original and independent discovery of this formula, valid for a periodic object. Alexandra MacWade of SPIE Press assisted with the copyediting of this book.

The completion of this project took a few years, as this was an after-hours effort by two practicing lithographers working in the semiconductor industry. To give the best possible service to our readers, we often went through multiple iterations on a particular topic, trying to figure out the true intention of the original authors, or playing detectives to figure out the right formula for plotting out a particular graph. We welcome readers' comments and suggestions.

It is our hope that this translation can serve as a self-study book for a wider circle and younger generations of readers who wish to learn optics and optical image formation. It is also a tribute to the original authors for their scientific achievements and devotion to the teaching and dissemination of precious knowledge. For practicing lithographers who try to extend the resolution limit one nanometer at a time, may a read through this book stimulate more innovative ideas down the road.

We dedicate this translation to Professor Henry I. Smith of MIT, our Ph.D. advisor (Doktorvater in German). Hank led us into the fascinating world of nanolithography in which we pursued careers.

Anthony Yen and Martin Burkhardt, California and New York, 2023

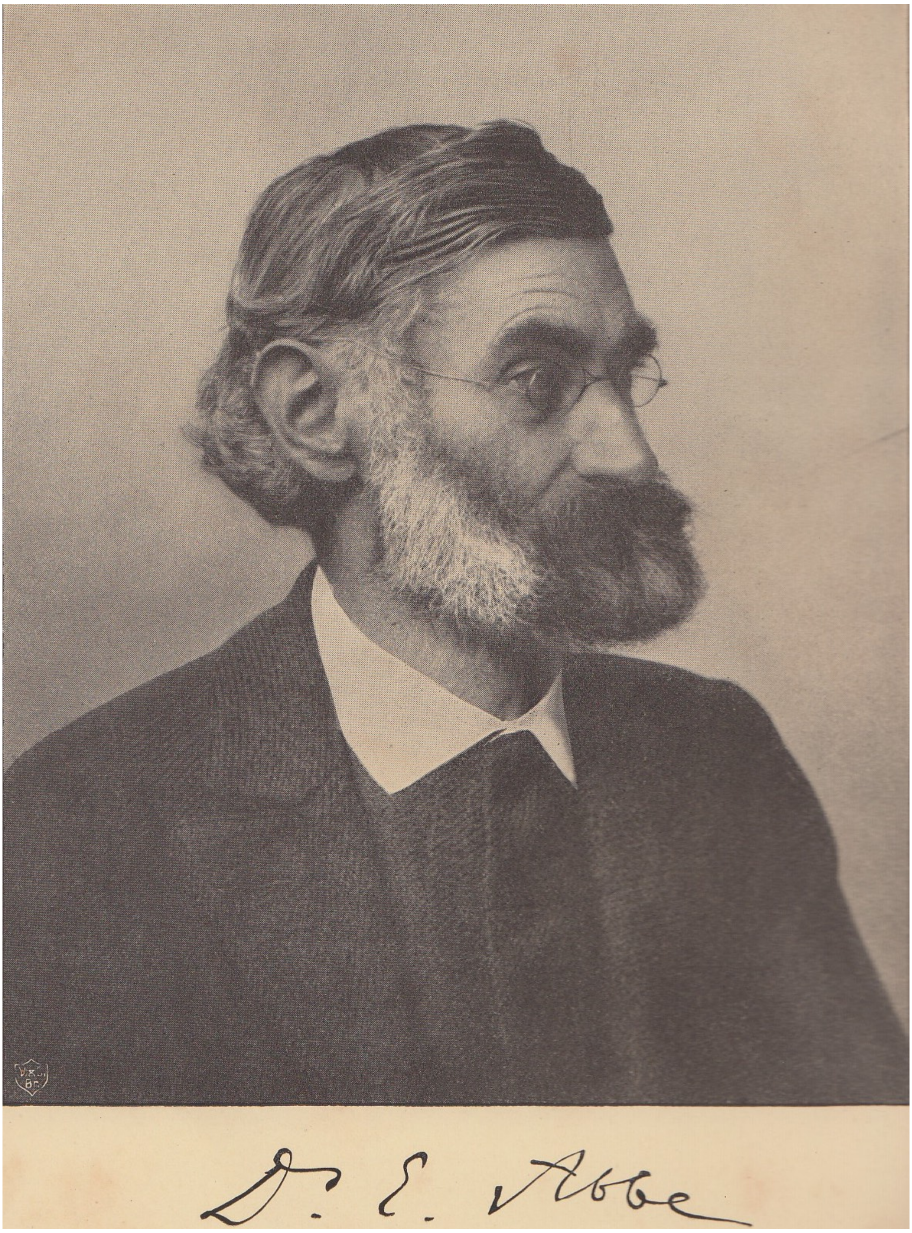
Special foreword

It is great to see the publication of an English edition of *Die Lehre von der Bildentstehung im Mikroskop von Ernst Abbe* 110 years after its original publication. The original book was based on lectures given by Ernst Abbe in 1887 on the theory of image formation in the microscope. It was edited and published by Otto Lummer and Fritz Reiche with contributions from Mieczysław Wolfke. In it, Abbe's theory was extended with the specific intent to link it with Kirchhoff's diffraction theory, and the calculated results were even experimentally verified using a ZEISS microscope fitted with a special objective prior to the book's publication. The book also offers many insights and was at the time the definitive book on the physics of imaging. The description of projection imaging as a double-diffraction process, as cited later by both Max Born and Frits Zernike in their respective works, is clearly presented here. It is no exaggeration to state that Abbe was the originator of Fourier optics. Today, besides its historical significance, the book can still serve as a textbook for self-study for anyone interested in learning the optics of projection imaging, especially for nanolithographers because imaging in lithography is closely related to imaging in microscopy; both make inherent use of partial coherence. With the advent of extreme ultraviolet lithography, the worldwide semiconductor industry has entered the single-digit nanometer era, bringing about never foreseen possibilities and benefits to our society. However, the imaging theory that gives all this progress its theoretical underpinning, derived from the principles of classical optics, remains valid today.

For more than 150 years, ZEISS has been at the forefront of optical engineering. Our latest success in the manufacturing of precision optics for extreme ultraviolet lithography, printing features of sub-30 nm in pitch with light of 13.5 nm in wavelength, epitomizes the long tradition of challenging the limits and accepting no second bests. This tradition, handed down from Ernst Abbe, is carried on by everyone working at ZEISS.

I also want to take this opportunity to thank the translators, Dr. Anthony Yen and Dr. Martin Burkhardt, for their tireless after-hours effort to bring this book to a new and hopefully wider audience of today, and for their generous agreement, at the suggestion of SPIE Press, to make this edition an Open Access item in the SPIE Digital Library so that generations to come may be inspired by Abbe's teaching.

Winfried Kaiser
ZEISS Fellow
Carl Zeiss SMT GmbH
Oberkochen, May 2020


Dr. E. Abbe

To the Memory of
Ernst Abbe

Preface[i]

More than 20 years ago, as a member of the Imperial Physical Technical Institute, I was sent to Prof. Ernst Abbe in Jena to attend his lectures on theoretical optics and to familiarize myself with the calculational methods of practical optics.

Abbe rarely completed his theoretical course. So it pleased him even more this time to be able to present his theories in front of a more educated circle. Besides myself, Prof. Winkelmann, Dr. Czapski, Dr. Rudolph, and doctoral candidate Straubel, the present successor of Abbe, took part in the course.

This winter season of 1887 is one of my fondest memories. We were granted the opportunity to witness the thought process of one of our greatest masters of theoretical and practical physics, and to see his work taking shape right in front of our eyes. Even though Abbe's theory of microscopic image formation had been developed by him long before this, and its conclusions had already brought great success to Zeiss Works, the Abbe lectures really only came into being at that time.

It was therefore not easy to follow Abbe, and he often corrected himself, by discarding an existing proof and replacing it with a more rigorous one. But it was precisely this that constituted the charm of those lectures, which were enhanced by the discussions during the Sunday walks in the lovely environs of Jena. Is there a back diffraction, i.e., can the energy that falls on a very narrow slit be diffracted back toward the light source? Such and similar questions were enthusiastically discussed, and the understanding was only deepened by the process of verbal articulation during the walk.

The introduction to practical optics and experimental validation of Abbe's teaching on the imaging of the illuminated objects went hand in hand with the purely theoretical lecture. Dr. Czapski introduced us to geometrical optics, Abbe's theory of ray limitation, and the calculation of the objectives aided by Abbe's approximation formulae. Abbe himself demonstrated to us dissimilarity in the imaging of microscopic objects by artificial clipping of diffraction orders. It was a wonderful time!

Decades have since passed and we always hoped to be able to read Abbe's lectures formulated by his own hand in print, for Abbe intended to publish his theory of microscopic imaging after that winter. We waited in vain! In the long time that has since passed, death claimed not only the life of our master, but soon afterwards also his pupil Dr. Czapski.

Thus it appeared that Abbe's own derivation of his theory would lie buried forever, for what is hitherto publicly available on Abbe's teaching are only its drawn conclusions and its popular derivations in Dippel's *Theory of the Microscope* and in "Optics" written by me at Abbe's urging in the Pfaundler-edited textbook *Müller-Pouillet*. Also, publications listed in the appendix, partly of a theoretical nature, contain no systematic and analytic development of the teaching according to Abbe.

But no one who, like myself, had worked out Abbe's theory and knew the treasures of astute thought contained therein could rest until they were brought to light. Since the two authorities were prevented from doing so, I was faced with the duty of honor to make up for what had been missing.

Only one question troubled me. Is Abbe's theory, built on the Fresnel–Huygens principle of interference of elementary waves, still up to date according to the contemporary standpoint of Kirchhoff's principle and Maxwell's theory? Would Abbe's theory have to be rebuilt on an entirely new foundation?

For this, first, a thorough immersion once more into the decades-long dormant course notes, which I had kept sacredly, was needed. In addition, I needed a theoretically trained coworker who had especially worked on the subject at hand. And when I had won over Dr. Reiche, who had worked in my institute for years, as my coworker, I turned to Mrs. Abbe in Jena during the Easter of 1909 to ask for permission to publish those lectures. At the same time, I asked the management of Zeiss Works whether an intention existed to accomplish this.

We went to work after receiving the notification, with only the course notes serving as the foundation. We present the result of our joint work below.

For a better understanding of the theory, we added Chapter 1 in which these concepts are explained based on geometrical optics, which will be needed later. For this, we leaned heavily on the presentation in my "Optics." With the derivation of the general expression for the light disturbance in the secondary images we went beyond Abbe, where we were guided by the desire to see to what extent the Abbe expression is valid on account of Kirchhoff's principle and Maxwell's theory.

Abbe derived this expression based on the Fresnel–Huygens principle and attached initially undetermined functions to take into account the influence of the angle of the emitting ray, the change of the amplitude due to passage through the optical system, and the tilt of the interfering elemental rays with respect to the optical axis. The sine condition, the Lambert cosine law, and the energy principle are then used to determine these functions.

If one starts from Kirchhoff's principle, one is bound by the function resulting from Kirchhoff's integral expression and this must naturally be an integral from the wave equation. The equations of Maxwell's light theory enter the derivation of the intensity expression, if one views the radiation as caused by dipoles. It can be shown that the radiation from a rotating dipole follows the Lambert cosine law on average, at least with the allowed restriction based on small convergence angles in the image space, as in the present case. It is therefore possible to find a function that depicts essentially the electric force of the dipole that replaces the luminous surface element. Since this function is an integral of the wave equation, Kirchhoff's integral theorem can be applied. Under the small convergence angle assumption in the image space, one obtains Abbe's expression for the light disturbance at the observation point in this rigorous way as well.

For further development of this expression and the derivation of general laws for the imaging of illuminated objects, we essentially followed Abbe. The phenomena treated in § 22 and § 24 were expanded, and the example worked out in § 23 was added. These calculations were carried out by Dr. Reiche.

In addition, the underlying mathematical problem for the "similarity law" of microscopic imaging was more precisely grasped by the distinction between a physical and an imaginary region of integration.

Chapter 4 was newly added. It is a hitherto not yet arithmetically carried out determination of the microscopic image of a grating with artificial clipping of its primary diffraction phenomenon. This calculation by Mr. Wolfke at our urging provides a touchstone for the exactness, with which the Abbe theory dipicts the experience.

We fulfill our obligation with joy to express our warmest thanks to Mrs. Abbe for her kind willingness with which she agreed to the publication of this book and the printing of a portrait of Ernst Abbe.

We would like to thank the publisher for the accommodation they showed us in every respect. Special thanks are due to them for the artistic reproduction of the portrait of Ernst Abbe that should be a welcoming gift to all readers.

Otto Lummer
Breslau, 1910

Contents

Introduction

Geometrical optics assigns reality to light rays and assumes that where light rays intersect is also where light concentration actually occurs. This arithmetic optics seeks accordingly to evaluate optical systems in such a way that two spaces are imaged onto each other point for point; i.e., outgoing rays from one point in one space (object space) reunite at one point in another space (image space). If an optical system meets this condition, it then transforms the outgoing convex spherical wavefront from the object point to a concave spherical wavefront whose center is the image point. Arithmetic optics does not have to deliver any more than this.

In order to understand the actual light distribution in the center of the concave spherical wavefront, i.e., the image point, image formation must be handled based on wave theory as a diffraction problem. One usually expresses the result of this approach by overlaying on the point of convergence of the homocentric ray bundle (image point in geometrical optics) the diffraction phenomenon that is uniquely determined by the type of blocking to the spherical wavefront in the image space. In reality, the process is reversed: the diffraction phenomenon is the primary image-forming process, and the image point is secondary. In fact, the image where the imaging ray bundle is limited by a circular aperture is at best a diffraction disk with alternating dark and bright diffraction rings of rapidly decreasing intensity. The greater the image angle whose sine is given by the ratio of the radius of the circular aperture in the image space to the radius of the accompanying spherical wavefront, the more the diffraction phenomenon shrinks to a point-like pattern. A true point-like concentration of light therefore never exists in an actual imaging process.

This is already valid for the imaging of *self*-luminous objects, where wave trains go out from individual elements of the object *incoherently*, i.e., not capable of interference. Next, we deal with the imaging of *illuminated* objects, whose individual surface elements send out wave trains that are *coherent*, i.e., capable of interference. Here, geometrical optics lets us down completely.

In the imaging of self-luminous objects, the conformation of wave theory to geometrical optics, with respect to similarity and pointwise convergence, becomes better and better with an increasing opening angle of the incoming ray bundle. So the two theories lead to different results only with regard to clarity in imaging, while they both preserve resemblance between the image and its object. With illuminated imaging, it is a different story. As Abbe first showed, quite dissimilar images appear from the object in certain cases. Moreover, one and the same optical system can provide images, from one and the same object, completely different from each other and dissimilar to the object, depending on the clipping of imaging ray bundles. For these abnormal phenomena, geometrical optics obviously cannot give any account.

However, as the following will show, wave theory can represent all phenomena with a good approximation to reality. To present Abbe's "theory of the illuminated objects," especially in microscopic imaging, we must first start from the fundamental laws of geometrical optics.

Chapter 1

Imaging laws of geometrical optics[ii]

§1. Construction of a ray refracted by a spherical surface
Let M (Fig. 1) be the center of the refracting sphere of radius r and refractive index n', and the ambient medium have the refractive index n. To find the refracted ray from the incident ray LE, we insert, according to the elegant method of construction of Weyerstraß, two auxiliary circles 1 and 2 with radii

$$r_1 = \frac{n'}{n} r$$

and

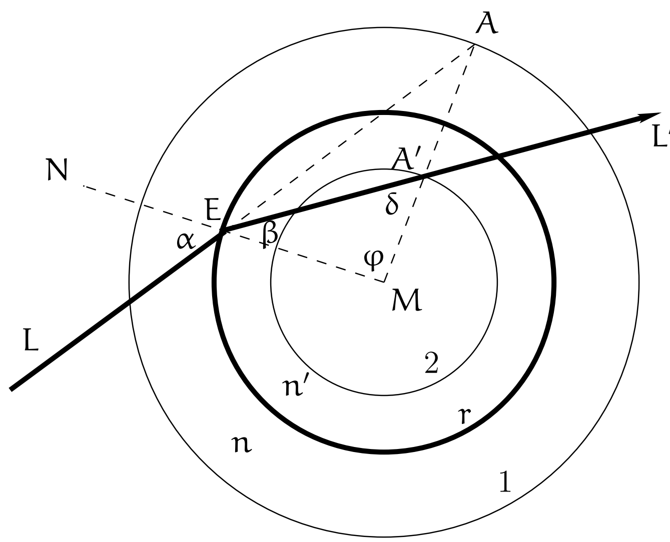
$$r_2 = \frac{n}{n'} r \, ,$$

extend ray LE until it intersects auxiliary circle 1 at A, and connect E with point A' where line AM and auxiliary circle 2 intersect. Line $EA'L'$ is the refracted ray associated with LE.

From the similarity of triangles EAM and $EA'M$, it follows that

$$\angle MEA = \angle EA'M$$ [iii]

Figure 1



or

$$\alpha = \delta\ ;$$

further,

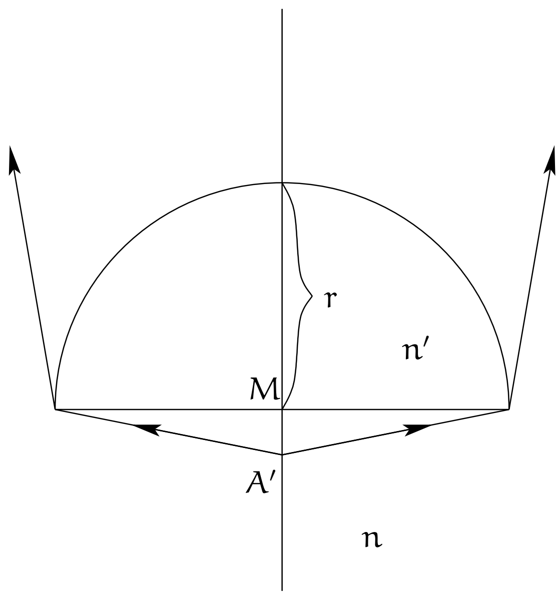
$$\frac{\sin\delta}{\sin\beta} = \frac{EM}{A'M} = \frac{n'}{n}\ ;$$

therefore,

$$\frac{\sin\alpha}{\sin\beta} = \frac{n'}{n}\ .$$

It follows immediately from this construction that all incident rays aiming toward A go through point A' after refraction. It follows therefore from the law of reciprocity that all outgoing rays from point A' in medium n' go through A after refraction, if they are extended backward. We want to designate these outstanding points A and A' as "aberration-free" points of a refracting spherical surface because the spherical aberration for them is zero. This "aberration-free" pair of points plays a major role in the construction of microscope objectives. One employs, e.g., a semispherical glass lens as the front lens in the apochromat (see Fig. 2).[iv] If one uses homogeneous immersion and brings the object to be imaged to distance $A'M = \frac{n}{n'}r$ of aberration-free point A', the divergence of near 180° is considerably reduced, without the occurrence of spherical aberration.

Figure 2

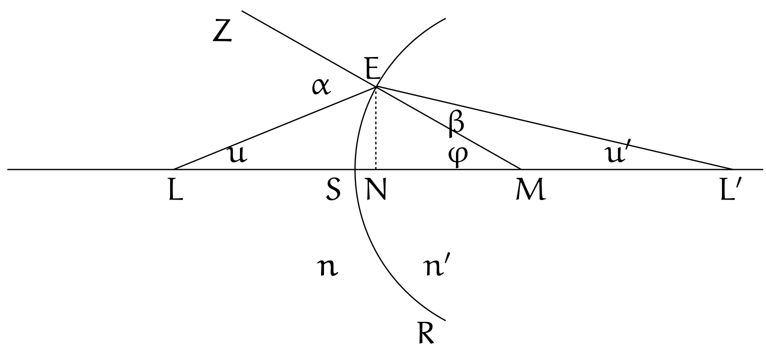


In order to learn more about the path of a ray bundle that is not coming from aberration-free points, we follow its path in analytical ways.

§2. Imaging of an arbitrary luminous axial point

Let M (Fig. 3) be the center of refracting spherical surface RSE that separates media n and n', onto which luminous point L sends rays. The unrefracted ray LSM passing through the spherical surface is

Figure 3



designated as the central ray and taken as the axis of the refractive system. If EL' is the refracted ray associated with LE, then

$$n \sin \alpha = n' \sin \beta \ .$$

According to the figure, the following holds:

$$\frac{\sin \alpha}{\sin u} = \frac{LM}{ME} \quad \text{and} \quad \frac{\sin \beta}{\sin u'} = \frac{L'M}{ME} \ ;$$

therefore,

$$\frac{LM}{L'M} = \frac{n' \sin u'}{n \sin u} \ .$$

Further, we have

$$\frac{\sin u'}{\sin u} = \frac{LE}{L'E} \ ,$$

and therefore,

$$\frac{LM}{L'M} \cdot \frac{L'E}{LE} = \frac{n'}{n} ;$$

the ratio $LM//L'M$ *is in general dependent on* u. We shall show that it becomes independent of u only if u and u' are small, that is, if we image using paraxial pencils (null rays).

Let us drop a vertical line EN onto the axis. Then,

$$LE = \frac{LN}{\cos u} = \frac{LS + SN}{\cos u} \text{ or } LE = \frac{LS + EM(1 - \cos\varphi)}{\cos u} .$$

Analogously,

$$L'E = \frac{L'S - EM(1 - \cos\varphi)}{\cos u'} .$$

If u, u', and therefore φ are so small that one can set $\cos u$, $\cos u'$, and $\cos\varphi = 1$, thus $LE = LS$ and $L'E = L'S$; then,

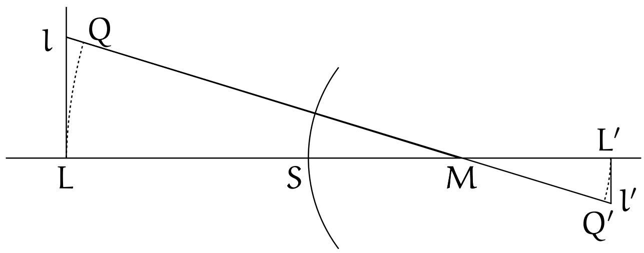
$$\frac{LM}{L'M} \cdot \frac{L'S}{LS} = \frac{n'}{n} . \tag{1}$$

Since $\frac{L'S}{LS}$ is completely independent of u, one therefore obtains the following theorem: *homocentric null rays remain homocentric after refraction.*[v]

§3. Imaging of luminous objects

If a second point Q (Fig. 4) is present besides the luminous axial point L, then what is valid for L in relation to LM is also valid for Q in relation to the neighboring axis QM. If one restricts oneself to point Q very close to axis LM, one sees that all object points lying on arc LQ with radius LM are imaged point-to-point onto the arc $L'Q'$ with radius ML'. Since one can, with the introduced restrictions, use instead of arcs LQ and $L'Q'$ their projections Ll and $L'l'$, we have the following theorem: *small surfaces perpendicular to the axis are imaged point-to-point as surfaces perpendicular to this axis.*

Figure 4



Since the conjugate points lie on the line going through the center of the sphere, we have

$$\frac{lL}{l'L'} = \frac{LM}{L'M} \quad \text{or} \quad \frac{y}{y'} = \frac{s-r}{s'-r}\,,$$

where lines y and y' are taken to be positive or negative depending on whether they lie above or below the axis. Since it was shown that[vi]

$$\frac{s-r}{s'-r} = \frac{n'}{n}\cdot\frac{s}{s'}\,,$$

therefore

$$\frac{y'}{y} = \frac{n}{n'}\cdot\frac{s'}{s}\,. \tag{2}$$

If one designates $\frac{y'}{y} = \beta$ as the "*lateral magnification,*" the following theorem is valid: *the lateral magnification is constant for conjugate planar pairs, but varies from pair to pair.*

From the figure,[vii] it is clear that

$$\frac{\tan u}{\tan(-u')} = \frac{+s'}{-s} \quad \text{or} \quad \frac{\tan u}{\tan u'} = \frac{s'}{s}\,,$$

where u and u' are to be evaluated as positive if the associated ray or its extension is rotated about L and L' in a clockwise fashion in order to reach the axis. By combining the last equation with Eq. 2, one gets

$$y'n' \tan u' = yn \tan u \,. \tag{3}$$

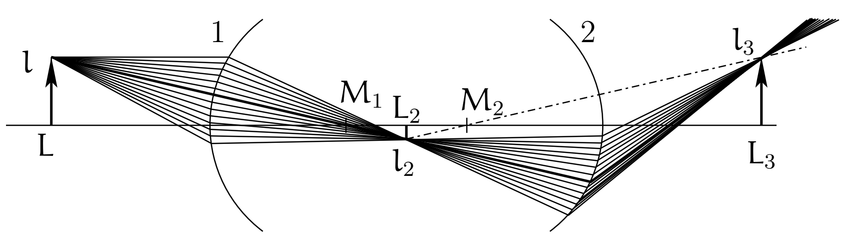
If one designates $\frac{\tan u'}{\tan u} = \gamma$ as "*angular magnification*," then

$$\beta\gamma = \frac{n}{n'} \,. \tag{4}$$

That is, "*The product of the lateral magnification and angular magnification is constant*" (law of Lagrange).

§4. Imaging by a centered system of refracting spherical surfaces
In a centered system, the centers of the refracting spherical surfaces all lie on a straight line, which we choose as the axis. Image L_2l_2 of object Ll (Fig. 5) produced by the first spherical surface can itself be interpreted as the object that generates image L_3l_3. Image point L_2 distinguishes itself from a self-luminous object in the same location in that its outgoing ray bundle does not fill completely the aperture of spherical surface 2. Nevertheless, l_2 will be imaged as a point that is l_3. Since this is also valid for every refracting spherical surface

Figure 5



that follows, we have the theorem: *the object space is imaged point-to-point in the image space. Planes perpendicular to the axis in the object space correspond point-to-point to the planes perpendicular to the axis in the image space.*

If one applies the Lagrange relation to each refracting surface in the system successively, one obtains the Lagrange–Helmholtz relation

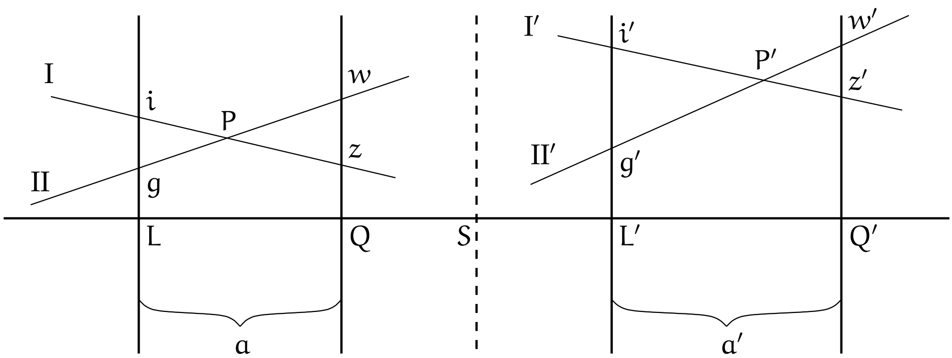
$$\left.\begin{aligned} \beta \cdot \gamma &= \frac{n}{n'} \\ \text{or} \qquad y'n' \tan u' &= yn \tan u \end{aligned}\right\}, \tag{5}$$

where β and γ now denote the lateral magnification and angular magnification with respect to the *entire system*, and n and n' are refractive indices of the front (object) and back (image) media.

§5. Imaging equations according to Abbe

In Fig. 6, let there be conjugate pairs of planes L and L' as well as Q and Q', and the associated lateral magnifications be given by

Figure 6

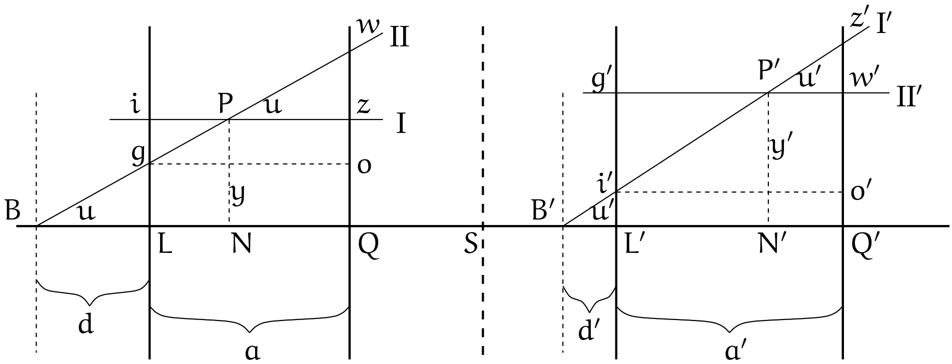


$$\frac{y_1'}{y_1} = \nu_1 \text{ and } \frac{y_2'}{y_2} = \nu_2\,;$$

overall imaging is thus determined. Let, e.g., a general ray I intersect object planes at i and z; one finds then, using values ν_1 and ν_2, the conjugate points i' and z' and with them the conjugate ray I′.[viii] Construct ray II analogously. Since every point P of the object space can be considered as an intersection of two rays that cut through planes L and Q, one can therefore, for every object point, find its conjugate image point P′.[ix]

To derive the imaging equations, we consider the special case (Fig. 7) of letting I run parallel to the axis while so directing II that its conjugate ray II′ runs parallel to the axis in the *image space*. In this case, for ray I, we have

Figure 7



$$y_1 = y_2 = y$$
$$\frac{y_2'}{y_1'} = \frac{\nu_2}{\nu_1}\,,$$

and for ray II,

$$y'_1 = y'_2 = y'$$
$$\frac{y_2}{y_1} = \frac{\nu_1}{\nu_2} \; ;$$

therefore,

$$\frac{B'L'}{i'L'} = \frac{L'Q'}{Q'z' - i'L'} \quad \text{or} \quad \frac{d'}{y'_1} = \frac{a'}{y'_2 - y'_1} \; ,$$

and so

$$d' = a' \frac{\nu_1}{\nu_2 - \nu_1} \; .$$

Analogously, we have

$$d = -a \frac{\nu_2}{\nu_2 - \nu_1} \; .$$

These equations tell us that distance d' is independent of y and distance d is independent of y'. We therefore have the theorem: *all rays parallel to the axis in the object space meet at an axial point in the image space (back focal point* B'*), and all outgoing rays from a definite axial point in the object space (front focal point* B*) travel parallel to the axis in the image space.*

From Fig. 7, we have

$$\tan u' = \frac{y'_2 - y'_1}{a'} = \frac{y}{a'} \left(\frac{y'_2}{y} - \frac{y'_1}{y} \right) \; ,$$

and therefore

$$\frac{y}{\tan u'} = \frac{a'}{\nu_2 - \nu_1} = F' \; , \tag{6}$$

where F' is a constant of the optical system and is defined (after Gauß) as the "*focal length*" of the image space. Analogously,

$$\frac{y'}{\tan u} = \frac{a\nu_1\nu_2}{(\nu 1 - \nu 2)} = F \; , \tag{7}$$

which is defined as the focal length of the object space.

We determine the location of P using a coordinate system whose z-axis coincides with the axis of the optical system and whose origin coincides with the front focal point B. We obtain the position of P′ using B′ as the origin. Let the positive sense of coordinates z and z' follow the direction of light propagation. Then,

$$\frac{y}{z} = \tan u \quad \text{and} \quad \frac{y'}{z'} = \tan u' .$$

If one combines these with the defining equations of focal lengths, one gets, finally,

$$\left.\begin{aligned} z \cdot z' &= F \cdot F' \\ \frac{y'}{y} &= \frac{F}{z} = \frac{z'}{F'} \end{aligned}\right\} . \tag{8}$$

The imaging equations in this form were first established by Abbe.[x]

§6. Imaging by wide-angle ray bundles (sine condition)

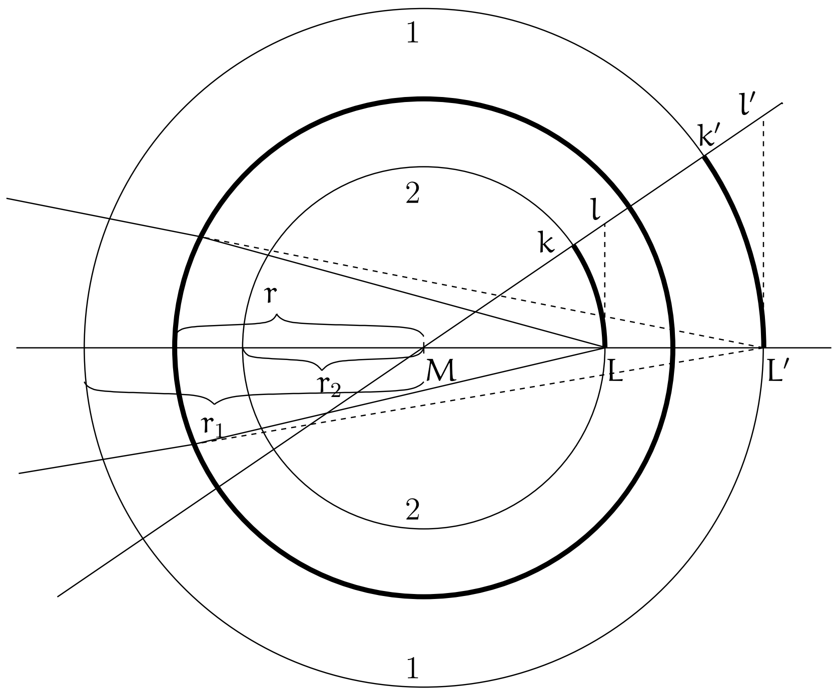
(a) *A refracting spherical surface.* Point-to-point imaging using null rays has no meaning in microscopic imaging, since it is necessary here, for reasons to be explained later, to bring *wide-angle* bundles of rays to union. The question is whether and under what conditions point-to-point imaging is possible by wide-angle bundles of rays in general.

As we have seen, this goal can be reached by a single refracting spherical surface for only a single pair of conjugate axial points.[xi] For this aberration-free pair of points, the relationship derived in § 2 is strictly valid:

$$\frac{LM}{L'M} = \frac{n' \sin u'}{n \sin u} = \text{const};$$

i.e., the length of convergence[xii] $L'M$ is independent of the opening angle u of the ray bundle.[xiii]

Figure 8



As one can see from Fig. 8, this is also valid for points k and k' with respect to the neighboring axis Mkk' for wide-angle ray bundles in general:

$$\frac{kM}{k'M} = \frac{n' \sin u'}{n \sin u} = \text{ const,}$$

where the constant has the same value as above. Therefore, arbitrarily large arc Lk of circle 2 with radius $r_2 = \frac{n}{n'} r$ can be imaged by general wide-angle ray bundles point-to-point and similar in perspective with respect to M, so that Lk is associated with the arc situated on the circle with radius $r_1 = \frac{n'}{n} r$ by

$$L'k' = Lk \cdot \frac{L'M}{LM} \,.$$

If we limit ourselves to very small objects Lk, we can set

$$\frac{L'k'}{Lk} = \frac{L'l'}{Ll} = \beta$$

and get

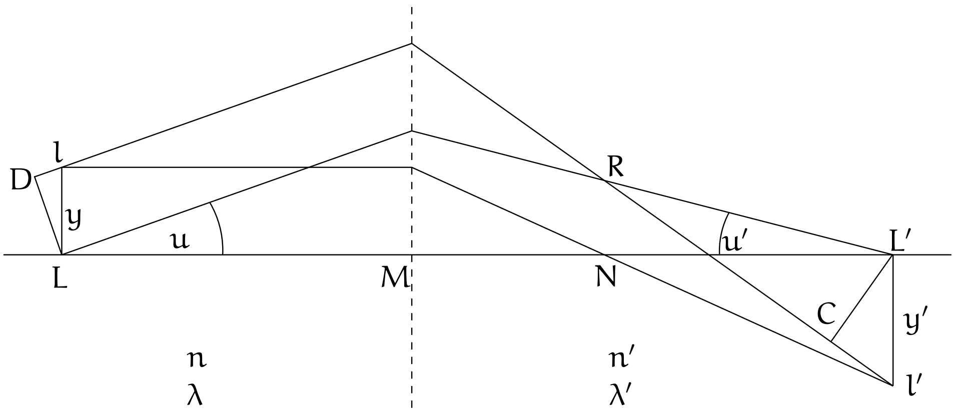
$$\frac{\sin u'}{\sin u} = \frac{n}{n'} \cdot \frac{1}{\beta} \,. \tag{9}$$

This is the condition under which a perpendicular-to-axis surface element at aberration-free point L is imaged as another perpendicular-to-axis surface element at the conjugate point L' point-to-point and in similarity by arbitrarily wide-angle ray bundles. It is called the "*sine condition*," and the conjugate aberration-free pair of points, for which this condition is satisfied, are called the "*aplanatic points*" of the refracting surface.

(b) *A centered system.* We now ask ourselves, which condition must be satisfied so that in a centered system of refracting spherical surfaces, a perpendicular-to-axis surface element Ll (Fig. 9) can be imaged to form another perpendicular-to-axis surface element $L'l'$ point-to-point and in similarity by wide-angle ray bundles in general. *All* rays coming from L should be refracted toward L', and rays from point l should be refracted toward the conjugate point l'.

The condition that all rays coming from axial point L are reunited at L' is identical therefore to stating that the system be

Figure 9



free of spherical aberration. In addition, should the rays coming from point l be reunited at the conjugate point l', a further condition must be satisfied such that the system for conjugate points l and l' with respect to the neighboring axis lMl' is free of spherical aberration. To find this condition, we track, according to the simple derivation by John Hockins,[1] two parallel rays originating from L and l in the object space, whose intersection in the image space is at R. The parallel-to-axis ray from l intersects the axis in the image space at N. We now draw perpendicular lines $L'C$ and LD. As a result of the absence of spherical aberration for the pair of points L and L', the following is valid for the optical lengths:

$$\overline{LRL'} = \overline{LNL'}\,.$$

But since

$$\overline{lRl'} = \overline{lNl'}\,,$$

[1] *Journ. Roy. Microscop. Soc.* 1884, Ser. 2, **4**, 337.

the following must be valid as well:

$$\overline{\mathfrak{l}R\mathfrak{l}'} - \overline{LRL'} = \overline{\mathfrak{l}N\mathfrak{l}'} - \overline{LNL'} = \overline{\mathfrak{l}N} + \overline{N\mathfrak{l}'} - \overline{LN} - \overline{NL'}\,;$$

since $L\mathfrak{l}$ is the wave front of the parallel-to-axis rays that meet at N, it is moreover valid that

$$\overline{LN} = \overline{\mathfrak{l}N}\,,$$

and therefore

$$\overline{\mathfrak{l}R} + \overline{R\mathfrak{l}'} - \overline{LR} - \overline{RL'} = \overline{N\mathfrak{l}'} - \overline{NL'}$$

or

$$\begin{aligned}(\overline{\mathfrak{l}R} - \overline{LR}) + (\overline{R\mathfrak{l}'} - \overline{RL'}) &= \overline{N\mathfrak{l}'} - \overline{NL'}\\ -\overline{D\mathfrak{l}} + \overline{C\mathfrak{l}'} &= \overline{N\mathfrak{l}'} - \overline{NL'}\\ \overline{C\mathfrak{l}'} &= (\overline{N\mathfrak{l}'} - \overline{NL'}) + \overline{D\mathfrak{l}}\,.\end{aligned}$$

If line segment $L'\mathfrak{l}'$ is small to the first order, the difference $\overline{N\mathfrak{l}'} - \overline{NL'}$ is small to the second order and is therefore negligible compared to line segment $\overline{D\mathfrak{l}}$.[xiv] We therefore obtain

$$\overline{C\mathfrak{l}'} = \overline{D\mathfrak{l}}\,;$$

or, if we transition to the equivalent line segments in vacuum, we have

$$n'L'\mathfrak{l}' \cdot \sin(-u') = nL\mathfrak{l} \cdot \sin u\,,$$

for u' is negative according to the prior agreement. Now,

$$\frac{L\mathfrak{l}}{L'\mathfrak{l}'} = \frac{y}{-y'} = -\frac{1}{\beta}\,;$$

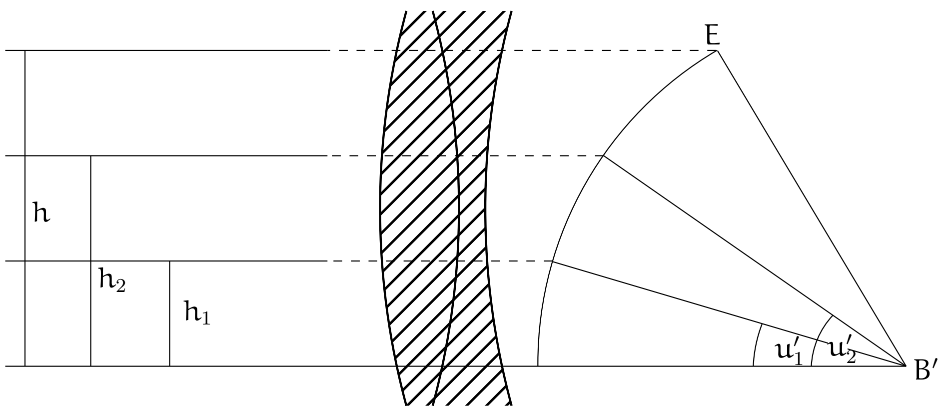
therefore we have, finally,

$$\frac{\sin u'}{\sin u} = \frac{n}{n'}\frac{1}{\beta} \,. \tag{10}$$

This *sine condition* is identical purely dioptrically in that *the various zones of the system of the object project a magnified image of the same ratio* β *at the same position* (the point of convergence of the null zone).

The sine condition takes on a very simple form if either the object or the image point lies at infinity. Then the sine condition goes from (see Fig. 10)[xv]

Figure 10



$$\frac{\sin u_1}{\sin u_1'} = \frac{\sin u_2}{\sin u_2'} = \text{const}$$

over to

$$\frac{h_1}{\sin u_1'} = \frac{h_2}{\sin u_2'} = \text{const}\,,$$

since the quotient $\frac{\sin u_1}{\sin u_2}$ under unbounded growth of the object distance approaches the value h_1/h_2 in the limit. Therefore,

$$\frac{h}{\sin u'} = \text{const} \, ;$$

since for very small values of u' we have

$$\frac{h}{\sin u'} = \frac{h}{\tan u'} = F' \, ,$$

the sine condition in this special case reads

$$\frac{h}{\sin u'} = F' \tag{11a}$$

or, if the image is at infinity,

$$\frac{h'}{\sin u} = F \, . \tag{11b}$$

One can see from Fig. 10 that

$$\sin u' = \frac{h}{EB'} \, ;$$

therefore, the following must be valid:

$$EB' = F' \, ,$$

that is, *the intersections of the extended parallel-to-axis incoming rays with their conjugate image rays must lie on a spherical surface having the back focal point* B' *as the center and the focal length of the system* F' *as the radius.*

Chapter 2

Imaging of self-luminous objects in terms of wave theory

§7. Diffraction problems solved on the basis of Maxwell's theory
We have seen that a centered system (microscope objective) images a surface element point-to-point and in similarity, using arbitrarily wide-angled ray bundles, only if the sine condition

$$\frac{\sin u'}{\sin u} = \frac{n}{n'} \cdot \frac{1}{\beta}$$

is fulfilled. If the system is so designed that this condition is satisfied, then all incoming rays to any point of the image remain perpendicular to a spherical surface centered on this point.[xvi] The lens designer[xvii] cannot offer anything more than this. We wonder whether and under what conditions this purely geometrical, pointwise concentration of rays is also physically present. Let us for the moment remain on the fiction of geometrical optics, that there were actually luminous points, so only the spherical wave emanating from this point would be a reality. Only with free, absolutely unhindered propagation, as would be the case in an arbitrarily extended, homogeneous medium, will the energy propagate along the radii exactly, as the ray theory assumes. If, however, as is always the case in reality, obstacles of any kind stand in the way of light propagation, i.e., if the medium exhibits inhomogeneities abruptly, light propagation can no longer be covered by ray-theoretic calculations; the wave fronts are no longer concentric spheres, but are somewhat deformed in a way (diffraction). The actually occurring propagation and distribution of the energy has been calculated based on Maxwell's electromagnetic theory of light only for very special cases.

The diffraction phenomenon appearing at the straight edge of an otherwise infinitely extended screen was treated by Sommerfeld.[1] Schwarzschild[2] succeeded in calculating the diffraction phenomenon associated with an infinitely extended slit of arbitrary width. Naturally, the numerical calculation becomes more difficult the smaller the slit width is in comparison to the wavelength. In addition, it must be emphasized that in both cases the material of the screen had to be assumed to have infinite conductivity. Under the same restriction, J. J. Thomson[3] could calculate the diffraction phenomenon of a sphere, whereas G. Mie[4] and P. Debye[5] carried out this case for spheres of arbitrary material. W. Seitz[6] and W. v. Ignatowsky[7] calculated the diffraction phenomenon of an infinitely long metallic cylinder of circular cross section and arbitrary conductivity, whereas Cl. Schaefer[8] carried out this calculation on cylinders of dielectric material and had it confirmed experimentally with the help of elec-

[1] *Mathem. Ann.* **47**, 317 (1896).
[2] ibid. **55** 177 (1902).
[3] J. J. Thomson, *Recent Researches in Electricity and Magnetism*, p. 361.
[4] *Ann. d. Phys.* **25**, 377 (1908).
[5] P. Debye, Dissertation. Munich 1908.
[6] *Ann. d. Phys.* **16**, 746 (1905); **19**, 554 (1906).
[7] *Ann. d. Phys.* **18**, 495 (1905).
[8] *Phys. Zeitschr.* X, **8**, 261.

trical waves (Großmann[9]). Finally, the diffraction phenomenon on metallic cylinders of elliptical cross section was treated (B. Sieger[10] and K. Aichi[11]), if only for material of infinitely large conductivity.

§8. The Kirchhoff principle

In general, the treatment of diffraction phenomena according to the Kirchhoff principle gives a far simpler form, allowing then the calculation of cases of our interest. Applying Green's theorems[xviii] to a function φ, which satisfies the wave equation[xix]

$$\frac{\partial^2\varphi}{\partial t^2} = a^2\Delta\varphi\,, \tag{12}$$

Kirchhoff[12] obtained the value of the function φ at an observation point P (Fig. 11) as a function of time t in terms of values of φ, $\partial\varphi/\partial t$, and $\partial\varphi/\partial\nu$ on the observation point–enclosing surface Σ with inward normal ν; here one must, for the magnitudes of φ, $\partial\varphi/\partial t$, and $\partial\varphi/\partial\nu$, insert the values that they possess at position $d\sigma$ at time $t' = t - r/a$, where r denotes the radius vector P $d\sigma$ and a the velocity of light in space V. It is[xx]

$$\varphi_P(t) = \frac{1}{4\pi}\int_\Sigma d\sigma \left[\varphi\frac{\partial(1/r)}{\partial\nu} - \frac{1}{ar}\frac{\partial\varphi}{\partial t}\cdot\frac{\partial r}{\partial\nu} - \frac{1}{r}\frac{\partial\varphi}{\partial\nu}\right]_{t'=t-\frac{r}{a}}\,. \tag{13}$$

Kirchhoff used this theorem to derive an approximation of the light intensity at observation point P (Fig. 12), if waves originating from L are disturbed by some obstacles. We want to carry out the calculation for the special case of an obstacle that is an *opaque* screen with aperture Σ_1. For this we place the surface of integration around

[9]Dissertation, Breslau 1909.
[10]*Ann. d. Phys.* **23**, 626 (1908).
[11]*Proc. Tokyo Mathem. Physical Soc.* (2) **4**, 966 (1908).
[12]Kirchhoff, *Lectures on Mathematical Physics, Vol. II, Optics*, 1891 (in German).

Figure 11

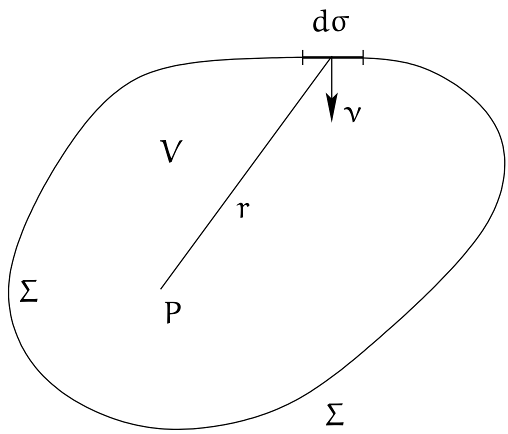
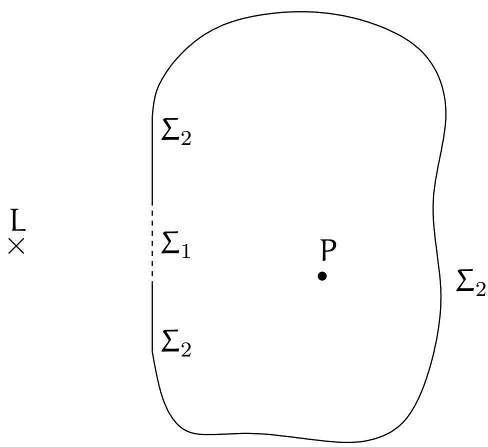


Figure 12

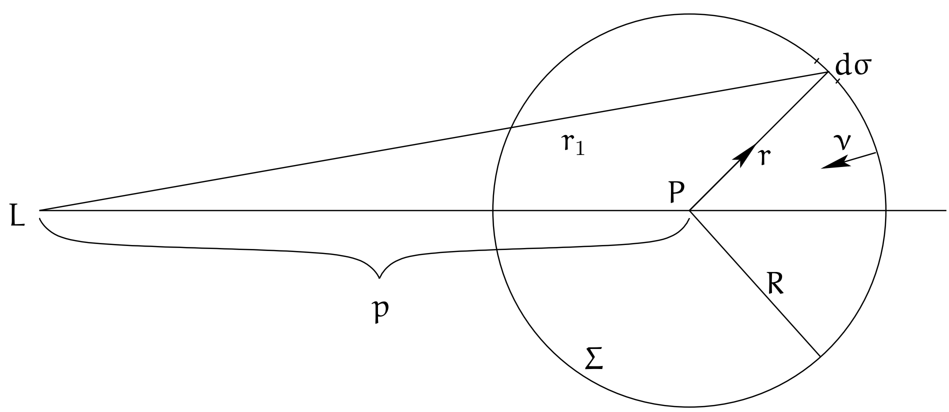


point P so that it is completely separated from L and let this surface consist of two parts, Σ_1 and Σ_2. Part Σ_2 wraps itself around the side of the screen facing the observation point and is thought of as closed at infinity. Let part Σ_1 be bordered by edges of the aperture.

The calculation of $\varphi_P(t)$ can only be carried out if one knows the values of φ, $\partial\varphi/\partial t$, and $\partial\varphi/\partial\nu$ at all points of the surface of integration; if one makes the natural hypothesis, *that the values on surface Σ_1 are the same as those of the undisturbed propagation*, and are zero on all points of surface Σ_2, then this assumption corresponds to the empirical knowledge that the bigger the aperture relative to the wavelength of the light, the closer it comes to the truth. In this case, the integral extends only over surface Σ_1.

The hypotheses made are strictly satisfied only for the *undisturbed* propagation. Here one knows the values of φ at P. We want to show that the calculation of φ by means of the Kirchhoff principle leads to this known value. For this we choose a sphere of radius R centered on P (Fig. 13) as the surface of integration and set, for points on surface Σ, as

Figure 13



$$\varphi = \frac{A}{r_1} \cos 2\pi \left(\frac{t}{T} - \frac{r_1}{\lambda} \right) .$$

Then we get

$$\frac{\partial \varphi}{\partial t} = -\frac{A}{r_1} \frac{2\pi}{T} \sin 2\pi \left(\frac{t}{T} - \frac{r_1}{\lambda} \right) ,$$

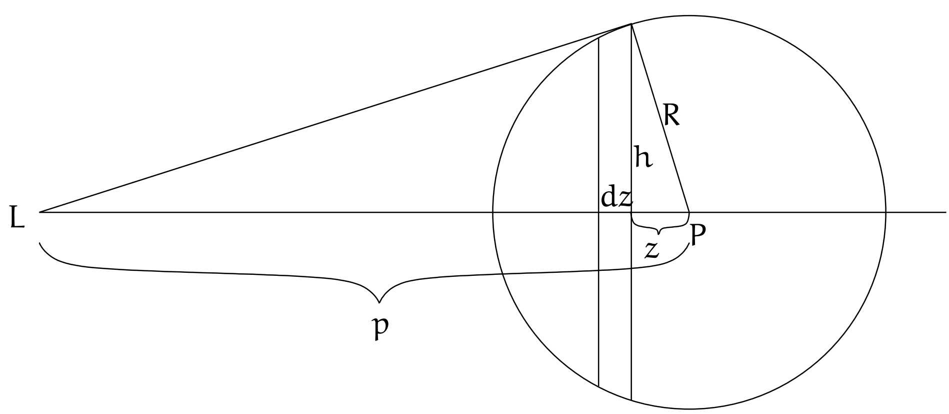
$$\begin{aligned} \frac{\partial \varphi}{\partial \nu} = \frac{\partial \varphi}{\partial r_1} \cos(r_1, \nu) = \cos(r_1, \nu) \Bigg\{ & -\frac{A}{r_1^2} \cos 2\pi \left(\frac{t}{T} - \frac{r_1}{\lambda} \right) \\ & + \frac{A}{r_1} \frac{2\pi}{\lambda} \cdot \sin 2\pi \left(\frac{t}{t} - \frac{r_1}{\lambda} \right) \Bigg\} , \end{aligned}$$

$$\frac{\partial (1/r)}{\partial \nu} = -\frac{1}{r^2} \cos(r, \nu) = +\frac{1}{r^2} ,$$

$$\cos(r_1, \nu) = \frac{p^2 - r^2 - r_1^2}{2 r r_1} .$$

We take, as element $d\sigma$ (Fig. 14) of the surface of integration, the piece of surface that is sliced from the spherical surface by two planes

Figure 14



perpendicular to PL and separated from each other by a distance dz. We then have

$$d\sigma = 2\pi R\, dz \ .$$

Since according to Fig. 14 we have

$$\begin{aligned} r_1^2 &= (p-z)^2 + h^2 \\ R^2 &= z^2 + h^2 \ , \end{aligned}$$

it follows then

$$r_1^2 = R^2 + p^2 - 2pz \, .$$

Differentiating this equation gives

$$dz = -\frac{r_1\, dr_1}{p} \ ,$$

where the limits of integration with respect to r_1 are $p - R$ and $p + R$. Inserting all these values, we have

$$\begin{aligned} \varphi_P(t) = -\frac{1}{4\pi} \int\limits_{p+R}^{p-R} \frac{2\pi R r_1\, dr_1}{p} \Bigg\{ \frac{A}{r_1 R^2} \cos\vartheta - \frac{A 2\pi}{a R r_1 T} \sin\vartheta \\ - \frac{A(p^2 - R^2 - r_1^2)}{R r_1 \cdot 2 R r_1} \left(-\frac{\cos\vartheta}{r_1} + \frac{2\pi}{\lambda} \sin\vartheta \right) \Bigg\} \ , \end{aligned}$$

where

$$\vartheta = 2\pi \left(\frac{t}{T} - \frac{R + r_1}{\lambda} \right) \ ;$$

or recast,

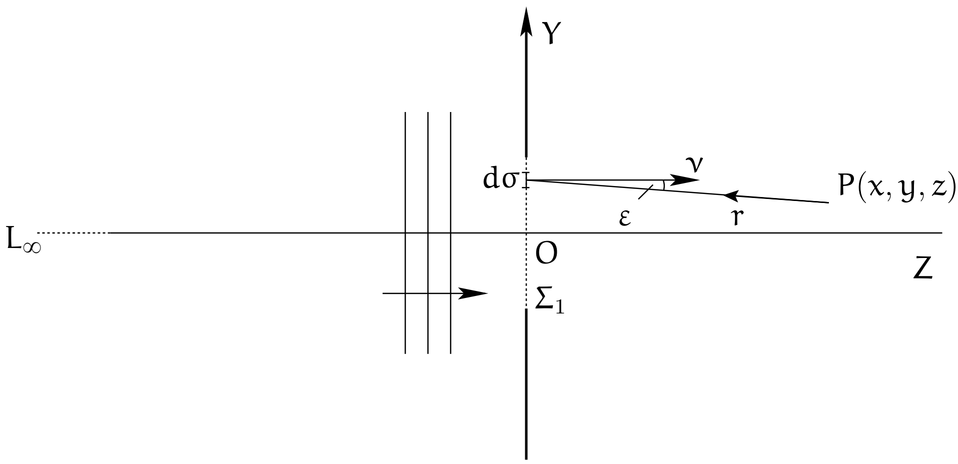
$$
\begin{aligned}
\varphi_P(t) &= -\frac{A}{2p}\int_{p+R}^{p-R} dr_1 \left\{ \frac{\cos\vartheta}{R} - \frac{2\pi\sin\vartheta}{\lambda} + \frac{p^2-R^2-r_1^2}{2Rr_1} \right. \\
&\qquad \left. \times \left(\frac{\cos\vartheta}{r_1} - \frac{2\pi}{\lambda}\sin\vartheta \right) \right\} \\
&= -\frac{A}{2p}\int_{p+R}^{p-R} dr_1 \left\{ \frac{\cos\vartheta}{R}\left(1+\frac{p^2-R^2-r_1^2}{2r_1^2}\right) \right. \\
&\qquad \left. -\frac{2\pi\sin\vartheta}{\lambda}\left(1+\frac{p^2-R^2-r_1^2}{2Rr_1}\right) \right\} \\
&= +\frac{AR}{2p}\int_{p+R}^{p-R} dr_1 \, \frac{d\left[\frac{\cos\vartheta}{R}\left(1+\frac{p^2-R^2-r_1^2}{2r_1R}\right)\right]}{dr_1} \\
&= \frac{AR}{2p}\left[\frac{\cos\vartheta}{R}\left(1+\frac{p^2-R^2-r_1^2}{2r_1R}\right)\right]_{p+R}^{p-R} = \frac{A}{p}\cos 2\pi\left(\frac{t}{T}-\frac{p}{\lambda}\right) ,
\end{aligned}
$$

i.e., the light disturbance taking place at P for the undisturbed propagation.

We now want to calculate the diffraction phenomenon caused by an arbitrary aperture in a *planar* screen for the case in which the point of light L is situated infinitely far from the diffraction aperture, that is, a *plane* wave is perpendicularly incident on the screen. The xy-plane (Fig. 15) is to lie in the plane of the screen, and the piece let go from the screen (diffracting aperture) is chosen as the surface of integration Σ_1. As the expression of the light disturbance φ, we set

$$
\varphi = A\cos 2\pi\left(\frac{t}{T}-\frac{z}{\lambda}\right) .
$$

Figure 15



We then have

$$\frac{1}{a}\frac{\partial\varphi}{\partial t} = -\frac{2\pi A}{\lambda}\sin 2\pi\left(\frac{t}{T}-\frac{z}{\lambda}\right) ,$$
$$\frac{\partial\varphi}{\partial\nu} = \frac{\partial\varphi}{\partial z} = \frac{2\pi A}{\lambda}\sin 2\pi\left(\frac{t}{T}-\frac{z}{\lambda}\right) ,$$
$$\frac{\partial r}{\partial\nu} = \cos(r,z) = -\cos\varepsilon ,$$

and therefore

$$\varphi_P(t) = \frac{1}{4\pi}\int d\sigma\left\{\frac{A}{r^2}\cos\varepsilon\cos\vartheta - \frac{2\pi A\cos\varepsilon}{r\lambda}\sin\vartheta - \frac{2\pi A}{r\lambda}\sin\vartheta\right\} ,$$

where

$$\vartheta = 2\pi\left(\frac{t}{T}-\frac{r}{\lambda}\right) .$$

If the distance r from the aperture to point P is large compared to the wavelength λ, the first term in the braces is negligible compared to the other two terms, and we obtain

$$\varphi_P(t) = -\frac{A}{\lambda}\int \frac{d\sigma}{r}\frac{1+\cos\varepsilon}{2}\sin 2\pi\left(\frac{t}{T}-\frac{r}{\lambda}\right) . \tag{14}$$

§9. Discussion of expression for the intensity at the observation point

From here, if one forms the average value $\overline{\varphi_P^2(t)}$,[xxi] it is then a direct measure of the observed intensity at observation point P; this is a consequence of the fact that we have used the ansatz of φ being a *plane* wave. For clarification, we note the following: according to the electromagnetic theory of light, the intensity of the field at every position is given by $\overline{\mathfrak{E}^2}$, where $\mathfrak{E}$ is simply the electric vector at the place of observation. For illustration, the following useful solution of Maxwell's equations is well known for spherical waves as well as for plane waves:[xxii]

$$\mathfrak{E}_x = \frac{\partial^2\varphi}{\partial x^2} - \frac{1}{a^2}\frac{\partial^2\varphi}{\partial t^2} , \qquad \mathfrak{H}_x = 0 ,$$

$$\mathfrak{E}_y = \frac{\partial^2\varphi}{\partial y\partial x} , \qquad \mathfrak{H}_y = +\frac{1}{a}\frac{\partial^2\varphi}{\partial z\partial t} ,$$

$$\mathfrak{E}_z = \frac{\partial^2\varphi}{\partial z\partial x} , \qquad \mathfrak{H}_z = -\frac{1}{a}\frac{\partial^2\varphi}{\partial y\partial t} ,$$

where φ must satisfy the equation[xxiii]

$$\frac{\partial^2\varphi}{\partial t^2} = a^2\Delta\varphi .$$

Here, $\mathfrak{E}$ and $\mathfrak{H}$ designate electric and magnetic vectors of the field. Let us start with a plane wave

$$\varphi = A\cos 2\pi\left(\frac{t}{T}-\frac{z}{\lambda}\right) ,$$

and get

$$\mathfrak{E}_x = \frac{4\pi^2 A}{\lambda^2}\cos 2\pi\left(\frac{t}{T}-\frac{z}{\lambda}\right), \qquad \mathfrak{H}_x = 0\,,$$
$$\mathfrak{E}_y = 0\,, \qquad \mathfrak{H}_y = \frac{4\pi^2 A}{\lambda^2}\cos 2\pi\left(\frac{t}{T}-\frac{z}{\lambda}\right),$$
$$\mathfrak{E}_z = 0\,, \qquad \mathfrak{H}_z = 0\,.$$

Therefore,

$$\overline{\mathfrak{E}^2} = \overline{\mathfrak{E}_x^2} = \frac{1}{2}\left(\frac{4\pi^2 A}{\lambda^2}\right)^2 = \frac{8\pi^4}{\lambda^4}A^2\,.$$

On the other hand, $\overline{\varphi^2} = \frac{1}{2}A^2$, which illustrates that, in the case of *plane* waves, $\overline{\varphi^2}$ differs from $\overline{\mathfrak{E}^2}$, which is relevant for the intensity, by only a constant factor, and that $\overline{\varphi^2}$ may be seen as a measure of the intensity.

The case of *spherical waves* is different, for which we have to start with[xxiv]

$$\varphi = \frac{A}{r}\cos 2\pi\left(\frac{t}{T}-\frac{r}{\lambda}\right),$$

where $r = \sqrt{x^2+y^2+z^2}$. If r is large compared to λ, then we get

$$|\mathfrak{E}| = \frac{4\pi^2 A\sin\vartheta}{\lambda^2 r}\cos 2\pi\left(\frac{t}{T}-\frac{r}{\lambda}\right), \tag{15}$$

where ϑ is the angle formed by the radial vector r with the x-axis.[xxv] From this it follows then

$$\overline{\mathfrak{E}^2} = \frac{8\pi^4 A^2\sin^2\vartheta}{\lambda^4 r^2},$$

whereas

$$\overline{\varphi^2} = \frac{A^2}{2r^2};$$

one therefore sees that, with spherical waves, one may not regard the last expression of $\overline{\varphi^2}$ as a measure of the intensity, because the true intensity $\overline{\mathfrak{E}^2}$ still varies with the direction ϑ at a constant r.

We may add that the field determined by Eq. 15 can be viewed as originating from an electric dipole or Hertzian oscillator whose axis of oscillation coincides with the x-axis.

In reality, one deals with the radiation of spatial objects that can be thought of as filled with radiating dipoles. In order to give a concept of the number of such dipoles, we must know the ratio of the number of radiating to the number of overall available molecules per unit volume. If we take luminous hydrogen as a basis and make the assumption that every molecule possesses one electron, then in every cubic centimeter, according to Ladenburg-Loria,[13] only 4×10^{12} are so-called radiating "dispersion electrons," compared to 2×10^{17} overall available electrons (molecules). In a cube of luminous hydrogen with an edge length of $0.001\,\mathrm{mm} = 1\,\mu\mathrm{m}$, there would then still be about four dispersion electrons present. In luminous vapors, however, even more dispersion electrons are present in such a volume element; in sodium vapor, e.g., there are about 1000. In reality, in radiating gases or vapors, we are not even dealing with individual undisturbed oscillating dipoles. On the other hand, we know that in radiating black bodies every surface element radiates according to Lambert's cosine law,[14] so that in free radiation the intensity at observation point P (Fig. 16) has the value

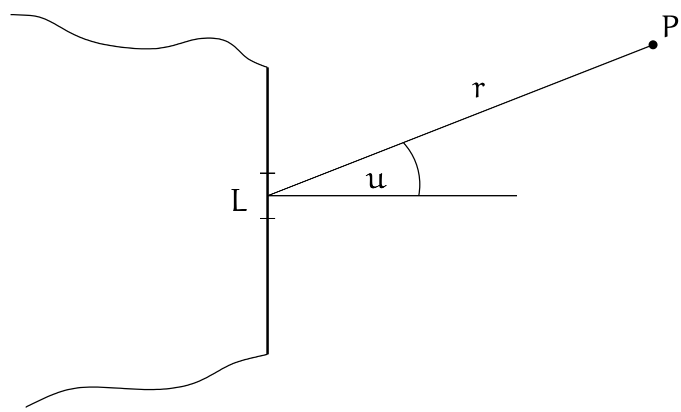
$$\frac{A^2}{r^2}\cos u\,;$$

here, too, the intensity depends on the direction of radiation r. Therefore, $\overline{\varphi^2}$ is a measure of intensity in neither free nor disturbed light propagation. Only when the luminous surface element is situated

[13] *Phys. Zeitschr.* (9) **24**, 875.

[14] O. Lummer and F. Reiche, *Dependence of radiation from a "Bunsen plate" (Bec Méker) on the radiating angle,* Verh. d. Schles. Ges. f. V. K. (1910) (in German).

Figure 16



so far from the diffracting aperture that we can consider the incident waves as planar, may we regard $\overline{\varphi_P^2(t)}$ as a measure of intensity.

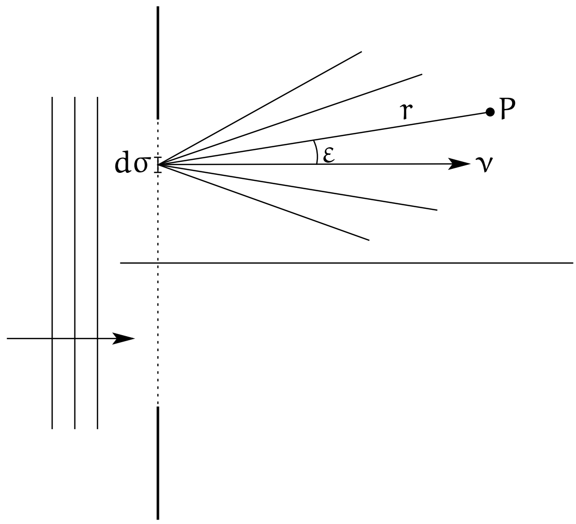
§10. Comparison of the Kirchhoff principle with the Fresnel–Huygens principle

We return to our expression (Eq. 14) for the light disturbance occurring at observation point P behind the diffraction aperture. It is

$$\varphi_P(t) = -\frac{A}{\lambda}\int \frac{d\sigma}{r}\frac{1+\cos\varepsilon}{2}\sin 2\pi\left(\frac{t}{T}-\frac{r}{\lambda}\right) .$$

In this version, we can interpret our formula as an expression of the Fresnel–Huygens principle, according to which one obtains the resulting light disturbance at observation point P due to the interference of imaginary coherent elemental waves leaving from all elements of the diffraction aperture. In our experience, the formula leading to correct results shows which factors to use when one takes into account the contribution of individual elemental waves; we can

Figure 17



write the contribution of each surface element dσ (Fig. 17) of the diffracting aperture as

$$-\frac{A'}{r}\sin 2\pi\left(\frac{t}{T}-\frac{r}{\lambda}\right)=\frac{A'}{r}\cos\left[2\pi\left(\frac{t}{T}-\frac{r}{\lambda}\right)+\frac{\pi}{2}\right],$$

where

$$A'=\frac{A\,d\sigma}{\lambda}\left(\frac{1+\cos\varepsilon}{2}\right).$$

Therefore, it is as if every element dσ sends out a spherical wave whose amplitude is A' at the unit distance, and whose phase with respect to that of the incident wave has been shifted by $\pi/2$. The amplitude, which one must enclose in the elemental waves in the direction of r, is to be set proportional to $\frac{1+\cos\varepsilon}{2}$, where ε is the angle between r and the incident direction of the impinging radiation.

In other words, it means that every surface element $d\sigma$ should not radiate according to Lambert's cosine law, according to which the amplitude would be proportional to $\sqrt{\cos\varepsilon}$; instead, the amplitude should vary proportional to $\frac{1+\cos\varepsilon}{2}$. One may easily see that both laws agree with each other up to the terms of order ε^2.[xxvi] However, there is absolutely no reason for assuming that these so defined elemental waves represent any kind of reality.

Fresnel made qualitatively similar assumptions in order to calculate the diffraction effect of an aperture. According to him, different surface elements contribute to the light disturbance at the observation point (1) proportional to its size; (2) inversely proportional to the distance from the observation point; and (3) proportional to a factor dependent on the direction with respect to the normal, with the normal direction being the maximum. Except for the phase of the oscillation, the Fresnel–Huygens principle also describes correctly the intensity distribution at least at a relatively large distance from the diffraction screen.

§11. Fraunhofer diffraction

One becomes independent of this proportionality factor, which is $\left(\frac{1+\cos\varepsilon}{2}\right)$ according to the Kirchhoff principle, if one lets the observation point go to *infinity*. To find the form that the phase takes in this case, we start from the relationship

$$r^2 = (x-\xi)^2 + (y-\eta)^2 + z^2 \,,$$

where x, y, z are the coordinates of the observation point and ξ, η, 0 are those of element $d\sigma$. If we set

$$x^2 + y^2 + z^2 = r_0^2 \,,$$

it follows then

$$r = r_0\sqrt{1 + \frac{\xi^2 + \eta^2 - 2(x\xi + y\eta)}{r_0^2}} \;;$$

if we let r_0 grow without restraint, ξ and η will always be small compared to r_0 and one can expand the square root in the following manner:[xxvii]

$$r = r_0 \left\{ 1 + \frac{\xi^2 + \eta^2}{2r_0^2} - \frac{x\xi + y\eta}{r_0^2} - \frac{(x\xi + y\eta)^2}{2r_0^4} \right\} .$$

If we set for the moment $x/r_0 = \alpha$, $y/r_0 = \beta$, we get

$$r = r_0 - \left((\xi\alpha + \eta\beta) + \frac{\xi^2 + \eta^2 - (\xi\alpha + \eta\beta)^2}{2r_0} \right) .$$

And so for infinitely large r_0,

$$r = r_0 - (\xi\alpha + \eta\beta) = r_0 - \frac{x\xi + y\eta}{r_0} .$$

Therefore,

$$\varphi_P(t) = -\frac{A}{\lambda} \int \frac{d\sigma}{r} \sin 2\pi \left(\frac{t'}{T} + \frac{x\xi + y\eta}{r_0 \lambda} \right) , \tag{16}$$

where we set $t' = t - r_0/c$.

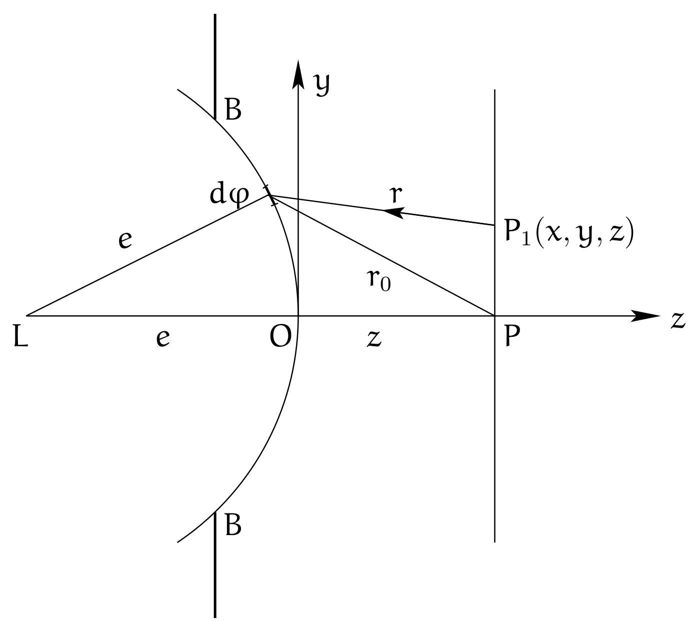
The phenomenon given by this expression is called Fraunhofer diffraction; it is exceptional in both formal and physical respects. Whereas with *finite* distance, be it of the luminous point or the observation point (Fresnel diffraction), *quadratic* terms in ξ and η appear in the expression for the phase, they *disappear* in Fraunhofer diffraction in which *the luminous point and the observation point lie at infinity*. This is realized if one brings the luminous point to the focal plane of a convex lens and observes the phenomenon in the focal plane of a second convex lens. Light source and observation point therefore lie in the planes that are, with respect to the imaging system (the two convex lenses), *conjugate* to each other. We want to show that we always get Fraunhofer diffraction; i.e., we always retain only *linear* terms in ξ and η in the expression for the phase if we make the luminous point and the observation point an *arbitrary conjugate* pair of points with respect to the imaging system. For this we investigate an auxiliary consideration.

§12. Auxiliary consideration

Let the diaphragm BB (Fig. 18) cut out, from the spherical wave coming from L, a piece of surface BOB that we choose as the surface of integration. If $d\varphi$ is an element of that surface and r is the distance between this element and observation point P_1, then we can depict the light disturbance at P_1 using the expression

$$s = \int \frac{A}{e}\frac{\mu}{\lambda}\frac{1}{r}\, d\varphi \sin 2\pi \left(\frac{t}{T} - \frac{e+r}{\lambda}\right) ,$$

Figure 18



where A/e is the amplitude of the light disturbance at $d\varphi$ and the factor μ takes into account the inclination of the elemental ray r with respect to $d\varphi$.

We choose O as the origin of a rectangular Cartesian coordinate system, LOP as the z-axis, the line through O pointing upward and perpendicular to LOP as the y-axis, and the line perpendicular to the drawing going into the paper as the x-axis.

If $\xi\eta\zeta$ are the coordinates of $d\varphi$, xyz are those of P_1, and we designate line segment $P\,d\varphi$ as r_0, then

$$\begin{aligned} r^2 &= (x-\xi)^2 + (y-\eta)^2 + (z-\zeta)^2 \\ &= (x^2+y^2) - 2(x\xi + y\eta) + r_0^2\,. \end{aligned}$$

The equation of the sphere is valid for the coordinates of $d\varphi$:

$$\xi^2 + \eta^2 + (e+\zeta)^2 = e^2 \text{ or } \xi^2 + \eta^2 = -\zeta^2 - 2e\zeta\,.$$

Therefore,

$$r_0^2 = \xi^2 + \eta^2 + (z-\zeta)^2 = (z-\zeta)^2 - \zeta^2 - 2e\zeta = z^2 - 2\zeta(z+e)\,.$$

r_0^2 takes on a particularly simple value if

$$z = -e\,.$$

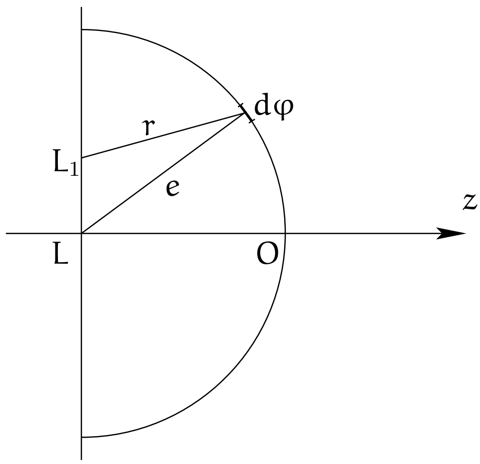
Then (Fig. 19),

$$r_0^2 = e^2 \text{ and } r^2 - r_0^2 = (r+e)(r-e) = x^2 + y^2 - 2(x\xi + y\eta)\,.$$

If we set

$$r - e = \rho \text{ and therefore } r + e = \rho + 2e\,,$$

Figure 19



then the following equation is valid:

$$\rho^2 + 2e\rho + [2(x\xi + y\eta) - (x^2 + y^2)] = 0\,.$$

It follows that

$$\rho = -e + \sqrt{e^2 - [2(x\xi + y\eta) - (x^2 + y^2)]}$$

or

$$\rho = -e + e\sqrt{1 - 2\frac{x\xi + y\eta - \frac{x^2+y^2}{2}}{e^2}}\,.$$

If x and y are small compared to e, i.e., if one limits oneself to observation points close to the line LOP, then

$$\rho = -e + e\left(1 - \frac{x\xi + y\eta}{e^2}\right)\,,$$

or finally,

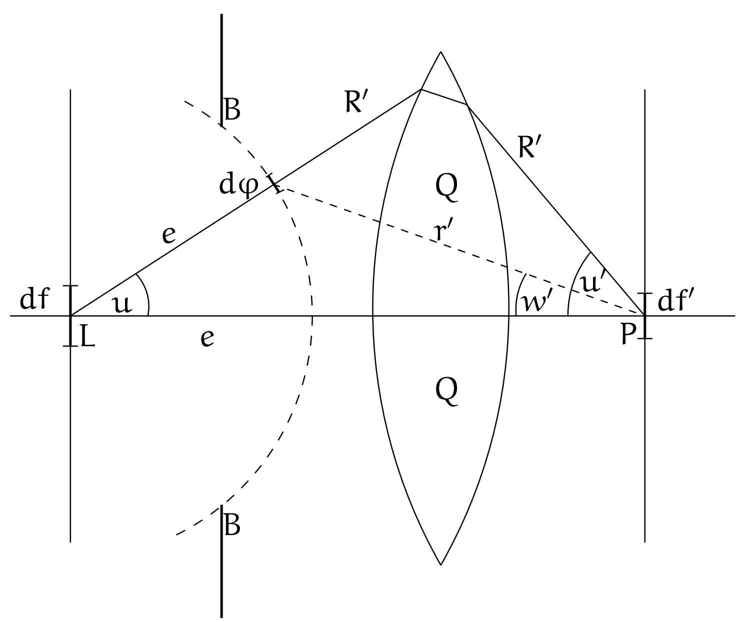
$$\rho = -\frac{x\xi + y\eta}{e} \; . \tag{17}$$

This simplification of the value ρ for $z = -e$, i.e., for the observation points *that lie in the object plane itself*, acquires a *physical* meaning with the introduction of imaging systems.

§13. Diffraction phenomena occurring in pairs of conjugate planes of optical systems

In Fig. 20, let the surface element df lying at L glow and its image df', projected by system Q, lie at P. Let diaphragm BB act as the entrance pupil that cuts an effective piece of the surface out of a

Figure 20



sphere centered at L with radius e. Let $d\varphi$ be an element of the surface; then the "amplitude" of the outgoing wave from df at $d\varphi$ is $\frac{A}{e} = \alpha$, where the designation "amplitude" is so understood that the intensity at the location of $d\varphi$ is given by the expression

$$J_{d\varphi} = \alpha^2 \, df = \frac{A^2}{e^2} \, df \ .$$

If we designate the rectilinear distance (dotted) from $d\varphi$ to P as r', then according to the Huygens principle, the *without-the-lens* light disturbance at P due to $d\varphi$ would have the amplitude

$$\frac{1}{\lambda} \alpha \, d\varphi \, \frac{1}{r'} \psi(w') \ ,$$

where $\psi(w')$ should take into account, with the interference of elemental waves, the influence of the inclination of the various elemental rays r' with respect to the direction of the axis LP and the inclination of the element $d\varphi$ to the associated elemental ray r'.

In the presence of the lens, from each element $d\varphi$ come the elemental rays that run in the immediate vicinity of chief ray R' associated with $d\varphi$, where R' also denotes the path length from $d\varphi$ toward P. With the lens we can therefore set the amplitude of the light disturbance at P originating from $d\varphi$ as

$$\frac{1}{\lambda} \alpha \, d\varphi \, f(R') \psi(u') \ ,$$

where $\psi(u')$ takes into account the various inclinations of the interfering elemental waves with respect to the axis and $f(R')$ their various geometrical lengths. The inclination of $d\varphi$ with respect to the effective elemental waves going out from $d\varphi$ is the same for all $d\varphi$. Since the geometrical length R' depends only on the accompanying angle of divergence u,[xxviii] we can then set

$$f(R') = \sigma(u) \ ,$$

and the resulting disturbance at P becomes

$$s = \frac{1}{\lambda} \int \alpha \, d\varphi \, \sigma(u)\psi(u') \sin 2\pi \left(\frac{t}{T} - \delta_P \right) \, ,$$

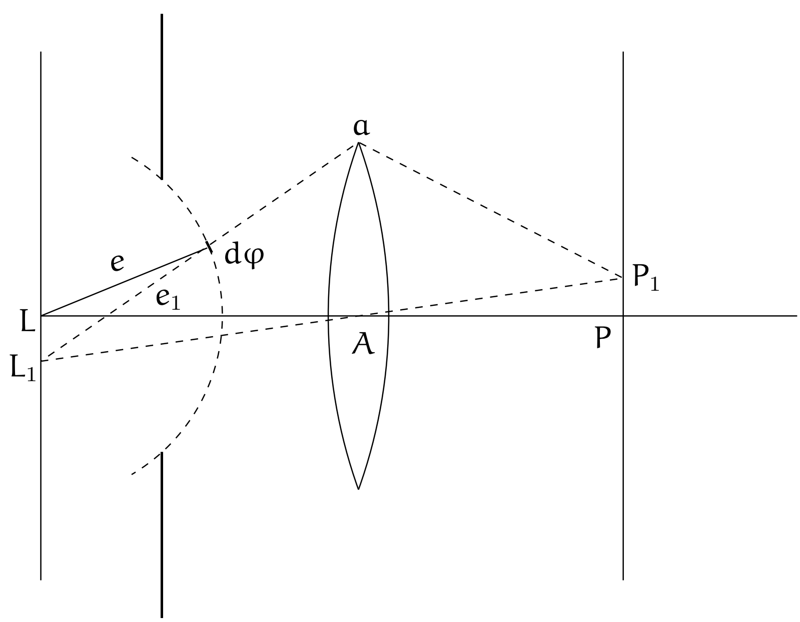
if δ_P designates the equal *optical* path length for all elemental rays between L and P.

The intensity at P is then given by

$$J_P = \overline{s^2} \, df \ .$$

Toward a point P_1 (Fig. 21) in the image plane come elemental pencils from $d\varphi$ that are seemingly coming from L_1, which is the

Figure 21



conjugate point of P_1. If the observation points are limited to be very close to the axis, one can express the resulting disturbance at P_1 as

$$s = \frac{1}{\lambda} \int \alpha \, d\varphi \, \sigma(u)\psi(u') \sin 2\pi \left(\frac{t}{T} - \delta_{P_1} \right) ,$$

where δ_{P_1} is the sum of the optical path length $\overline{Ld\varphi}$ and the optical path length $\overline{d\varphi a P_1}$. Now

$$\overline{d\varphi a P_1} = \overline{L_1 A P_1} - \overline{L_1 d\varphi} ,$$

where $\overline{L_1 A P_1}$ is a constant for the fixed location of P_1 and varies with the location of P_1.

Therefore,[xxix]

$$\overline{Ld\varphi P_1} = \text{const} - (\overline{L_1 d\varphi} - \overline{Ld\varphi})$$

and with that

$$\delta_{P_1} = -\frac{e_1 - e}{\lambda} + \text{const} = -\frac{\rho}{\lambda} + \text{const},$$

if one designates the segment $L_1 \, d\varphi$ by e_1. If P_1 moves toward P, $e_1 = e$ and the above constant becomes equal to δ_P. The phase difference between P and P_1 is exactly the same as that between their conjugate points L and L_1. If we designate the coordinates of L_1 by x, y, z and those of $d\varphi$ by ξ, η, ζ, then, as described earlier,[xxx]

$$\delta_{P_1} = \text{const} - \frac{\rho}{\lambda} = \text{const} + \frac{y\eta + x\xi}{e\lambda} .$$

With this, we obtain the resulting disturbance at P_1:

$$s = \frac{1}{\lambda} \int \alpha \, d\varphi \, \sigma(u)\psi(u') \sin 2\pi \left(\frac{t}{T} - \frac{x\xi + y\eta}{e\lambda} \right) , \qquad (18)$$

where the constant phase difference is lumped into t.

It should be pointed out here once and for all that in the expression for the light disturbance at observation point P_1 in the image plane found according to rules of geometrical optics, the coordinates of the observation point itself do not appear. Rather, the coordinates xy of the P_1-conjugate point L_1 in the object plane appear. Actually, we would have to substitute x and y with the expression

$$x = x'/\beta\,, \quad y = y'/\beta\,,$$

where $x'y'$ designate the coordinates of P_1 and β designates the lateral magnification. We do not, however, want to carry out this substitution because it only complicates the discussion of the expression of s and does not change the essence of the matter. The intensities calculated using pairs $x'y'$ and xy are exactly the same. If one depicts the diffraction phenomenon calculated in the image plane according to the rules of geometrical optics in the object plane, this depicted phenomenon is identical with the phenomenon calculated using the object points xy according to Eq. 18. One would see this phenomenon by replacing the optical system Q with the eye and accommodating on the object plane. In this respect, we are entitled to designate the phenomenon depicted by the expression s the "diffraction phenomenon in the object plane."

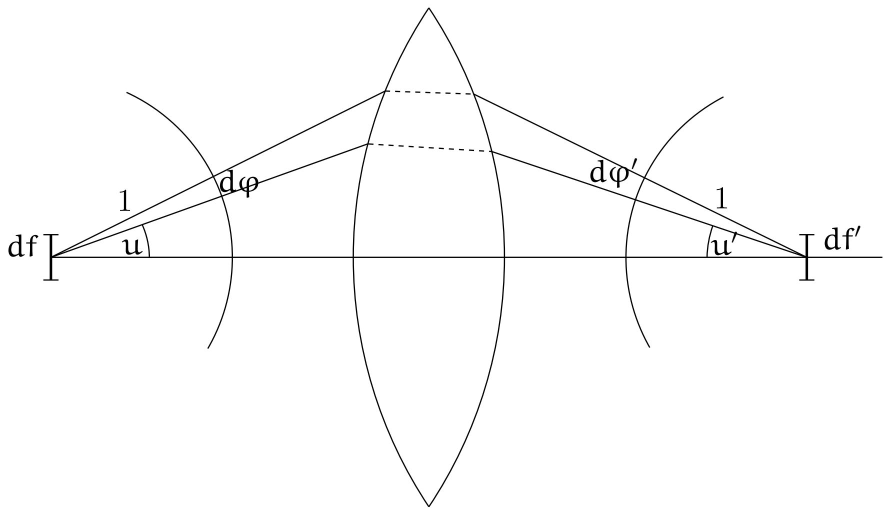
§14. Determination of factors α, $\sigma(u)$, and $\psi(u')$ based on energy considerations

To determine $\sigma(u)$, we presuppose that the sine condition is fulfilled. The energy principle says that in this case, the entire energy striking the system from object element df (Fig. 22) must flow through the point-to-point conjugate and similar image element df'. Since the same amount of energy must flow into conjugate elemental cones, we have

$$df \cdot d\varphi \cdot A^2 = df'\, d\varphi'\, A'^2\,,$$

if $d\varphi$ and $d\varphi'$ denote those surface elements that the elemental cones cut out of unit spheres about df and df', and A and A' denote

Figure 22



amplitudes present at $d\varphi$ and $d\varphi'$. If β denotes the lateral magnification of the system, then

$$\beta^2 = \frac{A^2\, d\varphi}{A'^2\, d\varphi'} \ .$$

If one introduces polar coordinates in a known manner,[xxxi] then

$$\begin{aligned} d\varphi &= \sin u\, du\, dv \\ d\varphi' &= \sin u'\, du'\, dv \ . \end{aligned}$$

Therefore,

$$\frac{d\varphi}{d\varphi'} = \frac{\sin u\, du}{\sin u'\, du'} \ .$$

One obtains a relationship between u and u' using the sine condition[xxxii]

$$\sin u' = \frac{\lambda'}{\lambda} \cdot \frac{1}{\beta} \sin u \ ,$$

where β denotes lateral magnification. Differentiation of the above expression yields

$$\cos u' \, du' = \frac{\lambda'}{\lambda} \cdot \frac{1}{\beta} \cdot \cos u \, du$$

and therefore

$$d\varphi' = d\varphi \cdot \left(\frac{\lambda'}{\lambda}\right)^2 \cdot \left(\frac{1}{\beta}\right)^2 \frac{\cos u}{\cos u'} \, .$$

If one inserts this value of $d\varphi'$ into the energy equation, it follows then

$$\frac{A'^2}{A^2} = \frac{\lambda^2 \cos u'}{\lambda'^2 \cos u} = \frac{n'^2 \cos u'}{n^2 \cos u} \, . \tag{19}$$

If $u' = 0$, i.e., the image moves to infinity, then

$$A^2 = \frac{n^2 \cos u}{n'^2} \cdot A'^2 \, .$$

Only when A'^2 is a constant for all elemental cones, i.e., when *the plane wave front has the same intensity everywhere in the image space*, does the above relationship transition to the law

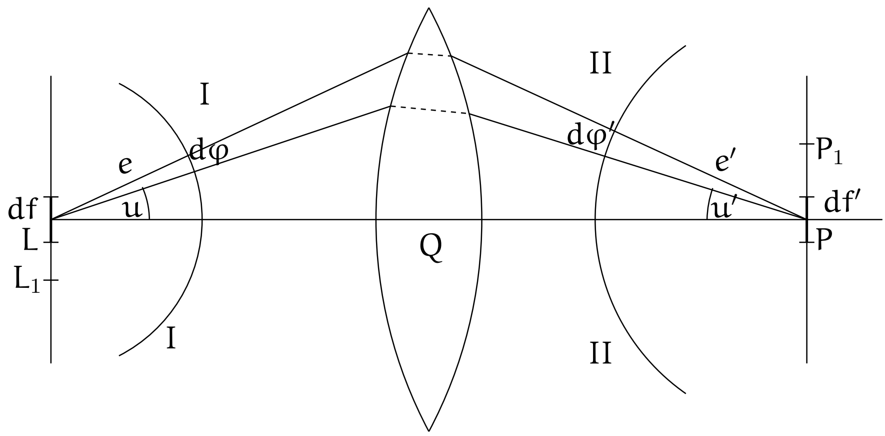
$$A^2 = \text{const}\, n^2 \cdot \cos u \, , \tag{20}$$

which represents the combination of the Lambert cosine law with the Kirchhoff–Clausius law of radiation.

We now construct the resulting light disturbance at P_1 while we consider, as boundary surfaces, one surface I situated at the distance e (Fig. 23) with elements $d\varphi$ and the other surface II located in the image space with elements $d\varphi'$. Let us denote the light disturbance at P_1 based on the boundary surface I as s_1; then, as before, we get

$$s_1 = \frac{1}{\lambda} \int\limits_{\mathrm{I}} \alpha \, d\varphi \, \sigma(u)\psi(u') \sin 2\pi \left(\frac{t}{T} - \frac{x\xi + y\eta}{e\lambda}\right) \, ,$$

Figure 23



where x, y, $-e$ are the coordinates of P_1's conjugate point L_1, and ξ, η are the coordinates of $d\varphi$. On the basis of boundary surface II,

$$s_2 = \frac{1}{\lambda'} \int\limits_{II} \frac{\alpha'}{e'} \, d\varphi' \, \psi(u') \sin 2\pi \left(\frac{t}{T} + \frac{x'\xi' + y'\eta'}{e'\lambda'} \right) \; ;$$

here, $\sigma(u)$ is replaced by $1/e'$ since our surface of integration, in the sense of light propagation, is located *after* the system Q; x', y' are the coordinates of P_1 and ξ', η' are those of $d\varphi'$.

If we introduce polar coordinates by making the substitution

$$\begin{aligned} \xi &= e \sin u \cos v \\ \eta &= e \sin u \sin v \\ d\varphi &= e^2 \sin u \, du \, dv \; , \end{aligned}$$

we get

$$s_1 = \frac{1}{\lambda}\int_0^{2\pi} dv \int_0^{U} du\, \alpha\sigma(u)\psi(u')e^2 \sin u \sin 2\pi\left(\frac{t}{T} - \sin u \frac{x\cos v + y\sin v}{\lambda}\right),$$

where U denotes the half angle of the aperture in the object space. Let us also introduce polar coordinates in s_2 and set in addition

$$x' = x\beta\,, \quad y' = y\beta\,.$$

If one bears in mind that for $\beta < 0$, ξ and ξ' as well as η and η' have the same sign, but x and x' as well as y and y' have opposite signs, whereas the reverse occurs for $\beta > 0$; by considering the sine condition,[xxxiii] one obtains

$$s_2 = \frac{1}{\lambda'}\int_0^{2\pi} dv \int_0^{U} du\, \alpha' e' \left(\frac{\lambda'}{\lambda}\right)^2 \left(\frac{1}{\beta}\right)^2 \frac{\cos u}{\cos u'}\psi(u')\sin u \cdot$$
$$\sin 2\pi\left(\frac{t}{T} - \sin u\frac{x\cos v + y\sin v}{\lambda}\right)\,.$$

By equating s_1 and s_2, we obtain the relation

$$\alpha' e' \frac{\lambda'}{\lambda}\left(\frac{1}{\beta}\right)^2 \frac{\cos u}{\cos u'} = \alpha e(e\sigma)$$

or

$$A'\frac{\lambda'}{\lambda}\cdot\frac{1}{\beta^2}\frac{\cos u}{\cos u'} = Ae\sigma\,.$$

By using the value of $\frac{A'}{A}$ $\left(= \frac{\alpha' e'}{\alpha e}\right)$ obtained from the energy principle,[xxxiv] we finally obtain

$$e\sigma = \frac{1}{\beta^2}\sqrt{\frac{\cos u}{\cos u'}}\,. \tag{21}$$

To determine $\psi(u')$, we construct for one time the resulting light disturbance at P under the premise that L glows and using surface II as the boundary surface. For another time, the resulting light disturbance is at L under the premise that P, the image of L, glows and using surface I as the intermediate surface. One can think of realizing this assumption by setting up a perfect mirror perpendicular to the axis at the location of P. The resulting light disturbance is given by the expression s_2 in the first case, if one sets $x = y = 0$ in it; for the case that P, the image of L, glows and I is used as the boundary surface, we obtain, for the light disturbance at L,

$$s_1' = \frac{1}{\lambda} \int d\varphi \, \frac{\alpha}{e} \psi(u) \sin 2\pi \frac{t}{T}$$

or, in polar coordinates,

$$s_1' = \frac{1}{\lambda} \int_0^{2\pi} d\nu \int_0^{U} du \, \alpha e \sin u \, \psi(u) \sin 2\pi \frac{t}{T} \, .$$

The amplitudes of the light disturbance at P (if L glows) and at L (if P glows) follow a known relationship. To determine this relationship, let us consider the following.

The contribution that the element $d\varphi'$ provides to the light disturbance is

$$ds_2 = B' \sin 2\pi \frac{t}{T} \, ,$$

where

$$B' = \frac{1}{\lambda'} \, d\nu \, du \, \alpha' e' \left(\frac{\lambda'}{\lambda} \right)^2 \frac{1}{\beta^2} \frac{\cos u}{\cos u'} \psi(u') \sin u \, .$$

We ask ourselves how large the resulting intensity caused by this contribution at P is. It is just as large as if df' itself radiated. That is,

$$J_P = \overline{ds_2^2} \, df' = \frac{1}{2} B'^2 \, df' \, ,$$

and therefore the energy that flows through df' in time dt is

$$E_P = J_P \, df' \, dt = \frac{1}{2} B'^2 (df')^2 \, dt \ .$$

Analogously, if df' radiates, the energy flowing through df that comes from $d\varphi$ is

$$E_L = \frac{1}{2} B^2 (df)^2 \, dt \ ,$$

where we define

$$B = \frac{1}{\lambda} \, d\nu \, du \, \alpha \, e \, \psi(u) \sin u \, .$$

According to the energy principle we must have

$$E_P = E_L \ ,$$

and it follows that

$$B' \, df' = B \, df$$

or

$$\frac{1}{\lambda'} \, d\nu \, du \, \alpha' e' \left(\frac{\lambda'}{\lambda}\right)^2 \frac{1}{\beta^2} \frac{\cos u}{\cos u'} \psi(u') \sin u \cdot \beta^2 = \frac{1}{\lambda} \, d\nu \, du \, \alpha \, e \, \psi(u) \sin u \, ;$$

or if one inserts here the previously obtained value of $\alpha' e' / \alpha e$,

$$\frac{\psi(u')}{\psi(u)} = \sqrt{\frac{\cos u'}{\cos u}}$$

or

$$\frac{\psi(u')}{\sqrt{\cos u'}} = \frac{\psi(u)}{\sqrt{\cos u}} \, .$$

Indeed, u and u' are dependent on each other in this special case; however, one can assign, by varying β (changing the system), every arbitrary value of u to the same u', so it is valid that

$$\frac{\psi(u')}{\sqrt{\cos u'}} = \frac{\psi(u_1)}{\sqrt{\cos u_1}} = \frac{\psi(u_2)}{\sqrt{\cos u_2}} \, ;$$

therefore, we must have

$$\psi(u') = \sqrt{\cos u'}\ . \tag{22}$$

§15. Expression of light disturbance at the observation point

If the radiating surface element radiates according to Lambert's law,

$$\alpha = \frac{\text{const}}{e}\sqrt{\cos u}\,,$$

considering the derived relationships (Eqs. 21 and 22)

$$\sigma(u) = \frac{\text{const}}{e}\sqrt{\frac{\cos u}{\cos u'}}$$
$$\psi(u') = \sqrt{\cos u'}\,,$$

Eq. 18 for the light disturbance at P_1 finally takes the form

$$s = \frac{k}{\lambda}\int\limits_{I} \frac{\cos u}{e^2}\, d\varphi \sin 2\pi \left(\frac{t}{T} - \frac{x\xi + y\eta}{e\lambda}\right)$$

or, since $d\varphi \cos u = d\xi\, d\eta$,

$$s = \frac{k}{\lambda}\int\limits_{I} \frac{d\xi\, d\eta}{e^2} \sin 2\pi \left(\frac{t}{T} - \frac{x\xi + y\eta}{e\lambda}\right)\ , \tag{23}$$

in which the integration extends over the projection of the boundary surface on the $\xi\eta$-plane.

x and y are the coordinates of L_1, the point, with respect to the system, conjugate to the observation point P_1. The intensity at P_1 is given by

$$J_{P_1} = \overline{s^2}\, df\ . \tag{23a}$$

One can of course, in the calculation of the light disturbance at P_1, also use integral s', which extends over surface II behind the system. Then,

$$s' = \frac{k'}{\lambda'}\int\limits_{II} \frac{d\xi'\, d\eta'}{e'^2} \sin 2\pi \left(\frac{t}{T} - \frac{x'\xi' + y'\eta'}{e'\lambda'}\right)\ ; \tag{24}$$

x' and y' are the coordinates of the observation point P_1. The intensity at P_1 is then

$$J_{P_1} = \overline{s'^2}\, df' = \beta^2 \overline{s'^2}\, df \ . \tag{24a}$$

Whereas one reaches the final expression of s or s' via a somewhat laborious determination of factors σ and ψ, which of course allows a deeper insight into the energy relationships, one obtains an expression in a shorter way by means of the Kirchhoff principle, which, for u' not too large, agrees with s' found above.

§16. Determination of light disturbance at the observation point using the Kirchhoff principle

Again let the intensity at $d\varphi$ of the radiation originating from element df (Fig. 23) be

$$J_{d\varphi} = \text{const} \frac{\cos u \cdot df}{e^2} \ .$$

According to the electromagnetic theory of light, up to a constant, this intensity must be identical with the time average of the governing electric field at the location of $d\varphi$; that is,

$$J_{d\varphi} = \overline{\mathfrak{E}^2} = \text{const} \frac{\cos u \cdot df}{e^2} \ . \tag{25}$$

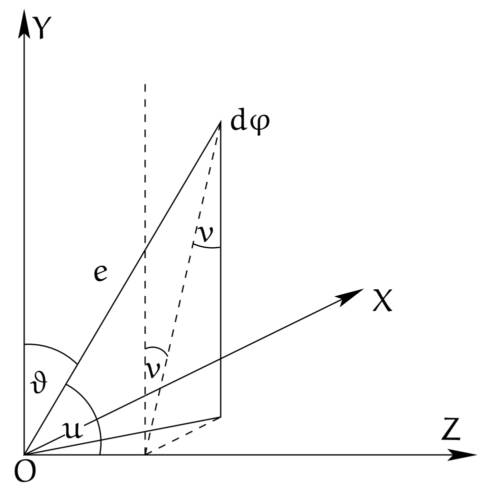
One can replace this unpolarized radiation of the surface element df with the radiation of a dipole whose axis stands perpendicularly to the axis of the system and rotates in the plane of element df about the system axis.

Proof: it is generally known that the electric field at $d\varphi$ generated by a *stationary* dipole at df is[xxxv]

$$\mathfrak{e} = \frac{A}{e} \sin\vartheta \cos 2\pi \left(\frac{t}{T} - \frac{e}{\lambda} \right) ,$$

provided that e is large compared to λ. ϑ is the angle that the radius vector e (Fig. 24) forms with the axis OY of the dipole at O. If one

Figure 24



introduces polar coordinates e, u, ν around the system axis OZ, then

$$\cos\vartheta = \sin u \cdot \cos\nu$$

or

$$\sin\vartheta = \sqrt{1 - \sin^2 u \cos^2\nu}\,,$$

in which ν, as the dipole *rotates*, varies between 0 and 2π. The average value of the electric field is therefore[xxxvi]

$$\begin{aligned}\mathfrak{E} &= \frac{1}{2\pi}\int_0^{2\pi} \mathfrak{e}\, d\nu = \frac{A}{e}\cos 2\pi\left(\frac{t}{T} - \frac{e}{\lambda}\right)\cdot\frac{1}{2\pi}\int_0^{2\pi}\sqrt{1 - \sin^2 u\cos^2\nu}\, d\nu \\ &= \frac{A}{e}\cos 2\pi\left(\frac{t}{T} - \frac{e}{\lambda}\right)\cos^2\frac{u}{2} \\ &\quad\times\left\{1 + \left(\frac{1}{2}\right)^2\tan^4\frac{u}{2} + \left(\frac{1}{2\cdot 4}\right)^2\tan^8\frac{u}{2} + \cdots\right\}.\end{aligned}$$[xxxvii]

If $\mathfrak{u}$ is not too large, we can restrict ourselves to the first term in the series, because even for $\mathfrak{u} = 20°$, the value of the second term is only 0.00024. We therefore obtain

$$\begin{aligned}\mathfrak{E} &= \frac{A}{e}\cos 2\pi\left(\frac{t}{T}-\frac{e}{\lambda}\right)\cos^2\frac{\mathfrak{u}}{2}\\ &= \frac{A}{e}\cos 2\pi\left(\frac{t}{T}-\frac{e}{\lambda}\right)\cdot\frac{1+\cos\mathfrak{u}}{2}\,.\end{aligned}$$

For not-too-large $\mathfrak{u}$ we can replace the factor $\frac{1+\cos\mathfrak{u}}{2}$ by $\sqrt{\cos\mathfrak{u}}$; *even with* $\mathfrak{u} = 20°$ *these two values agree to the third decimal place.*[xxxviii] Therefore, we finally obtain

$$\begin{aligned}\mathfrak{E} &= \frac{A}{e}\cos 2\pi\left(\frac{t}{T}-\frac{e}{\lambda}\right)\cdot\sqrt{\cos\mathfrak{u}}\,,\\ \mathfrak{E}^2 &= \frac{1}{2}\frac{A^2}{e^2}\cos\mathfrak{u}\,.\end{aligned}$$

Therefore, if we set according to Eq. 25

$$A^2 = 2\cdot\mathrm{const}\cdot df\ ,$$

we have proved that *one can replace the radiating surface element* df *according to the cosine law with the radiation of a rotating dipole.*

If the convergence angle $\mathfrak{u}'$ in the image space is not too large, as we assume, then we are justified to set at the location of $d\varphi'$,

$$\mathfrak{e}' = \frac{A'}{e'}\sin\vartheta'\cos 2\pi\left(\frac{t}{T}+\frac{e'}{\lambda'}\right)$$

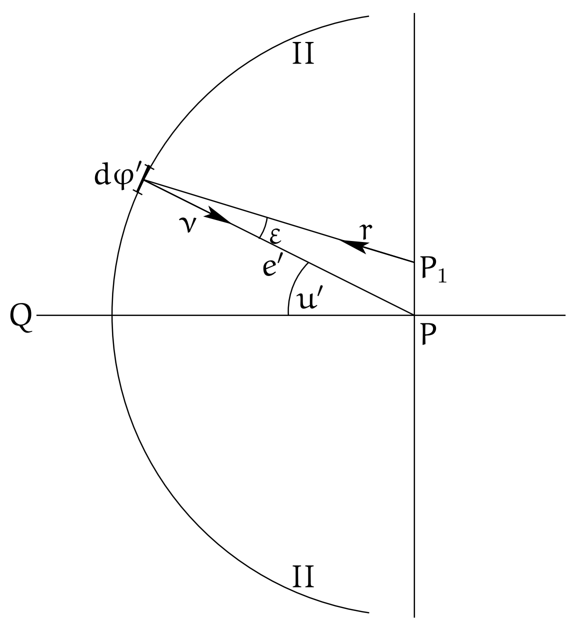
or

$$\mathfrak{e}' = \frac{A'}{e'}\sqrt{1-\sin^2\mathfrak{u}'\cos^2\nu}\cos 2\pi\left(\frac{t}{T}+\frac{e'}{\lambda'}\right)\ ,$$

where ϑ' for the image space has the analogous meaning as ϑ for the object space, and denotes the angle between e' and the axis of the dipole perpendicular to QP. To obtain *unpolarized* surface radiation, we must subsequently still form the average value of this expression over all ν, from 0 to 2π.

To apply the Kirchhoff principle to a vector, we must insert, as surface values, the values of those *vector components* and their derivatives with respect to the normal of the integration surface that are *parallel to the resulting vector at the observation point*. If we assume the bounding aperture to be *symmetrical* with respect to axis QP, the resulting vector $\mathfrak{e}'$ of the field at P (Fig. 25) and at paraxial point P_1, generated by the

Figure 25



stationary dipole, has necessarily the direction parallel to the dipole axis and perpendicular to axis QP. At $d\varphi'$, however, $\mathfrak{e}'$ is tangential to spherical surface II and therefore forms the angle $\frac{\pi}{2} - \vartheta'$ with the direction of the resulting vector at P_1.

Thus, as surface values, we take

$$\mathfrak{e}' \cos\left(\frac{\pi}{2} - \vartheta'\right) = \mathfrak{e}' \sin\vartheta' = \mathfrak{e}'\sqrt{1 - \sin^2 u' \cos^2 \nu}$$

and their derivatives with respect to ν.

If the dipole *rotates*, we form the average value of these magnitudes with respect to ν and obtain

$$\begin{aligned}
\mathfrak{E}' &= \frac{1}{2\pi}\int_0^{2\pi} \mathfrak{e}'\sqrt{1 - \sin^2 u' \cos^2 \nu}\, d\nu \\
&= \frac{1}{2\pi}\frac{A'}{e'} \cos 2\pi\left(\frac{t}{T} + \frac{e'}{\lambda'}\right) \int_0^{2\pi} (1 - \sin^2 u' \cos^2 \nu)\, d\nu \\
&= \frac{A'}{e'} \cos 2\pi\left(\frac{t}{T} + \frac{e'}{\lambda'}\right)\left(1 - \frac{1}{2}\sin^2 u'\right) ;
\end{aligned}$$

and since u' is assumed to be small,

$$\mathfrak{E}' = \frac{A'}{e'} \cos 2\pi\left(\frac{t}{T} + \frac{e'}{\lambda'}\right) \cdot \cos u' . \tag{26}$$

To apply Kirchhoff's law on $\mathfrak{E}'$, we still have to show that $\mathfrak{E}'$ is a solution of the wave equation (Eq. 12), which takes on, with the introduction of polar coordinates and especially for the present case, the following form:[xxxix]

$$\frac{1}{a'^2}\frac{\partial^2 \mathfrak{E}'}{\partial t^2} = \frac{1}{e'}\frac{\partial^2 (e'\mathfrak{E}')}{\partial e'^2} + \frac{1}{e'^2 \sin u'}\frac{\partial\left(\sin u' \frac{\partial \mathfrak{E}'}{\partial u'}\right)}{\partial u'} .$$

Here, a' is the velocity of propagation of the waves in the image space.

A solution of this equation is[xl]

$$\mathfrak{E}' = \frac{\text{const}}{e'} \cos u' \left\{\cos 2\pi\left(\frac{t}{T} + \frac{e'}{\lambda'}\right) - \frac{\lambda'}{2\pi e'} \sin 2\pi\left(\frac{t}{T} + \frac{e'}{\lambda'}\right)\right\} ,$$

which, since e' is large compared to λ', reduces to the expression identical to Eq. 26,

$$\mathfrak{E}' = \frac{\text{const}}{e'} \cos u' \cdot \cos 2\pi \left(\frac{t}{T} + \frac{e'}{\lambda'} \right) .$$

With this it has been shown that $\mathfrak{E}'$ is a solution of the wave equation for the case treated here and therefore can be inserted in place of φ in Eq. 13 of the Kirchhoff principle.

If one introduces once again s' via Eq. 24a,

$$J_{P_1} = \overline{\mathfrak{E}'^2} = \overline{s'^2} \cdot \overline{df'} ,$$

after easy calculation,[xli] if one replaces r with e' in the amplitude and $\frac{1+\cos u}{2}$ with 1, one obtains

$$\begin{aligned} s' &= \frac{k'}{\lambda'} \int\limits_{II} \frac{d\varphi' \cos u'}{e'^2} \sin 2\pi \left(\frac{t}{T} + \frac{x'\xi' + y'\eta'}{e'\lambda'} \right) \\ &= \frac{k'}{\lambda'} \int\limits_{II} \frac{d\xi'\, d\eta'}{e'^2} \sin 2\pi \left(\frac{t}{T} + \frac{x'\xi' + y'\eta'}{e'\lambda'} \right) , \end{aligned}$$

which is exactly the above derived expression (Eq. 24).

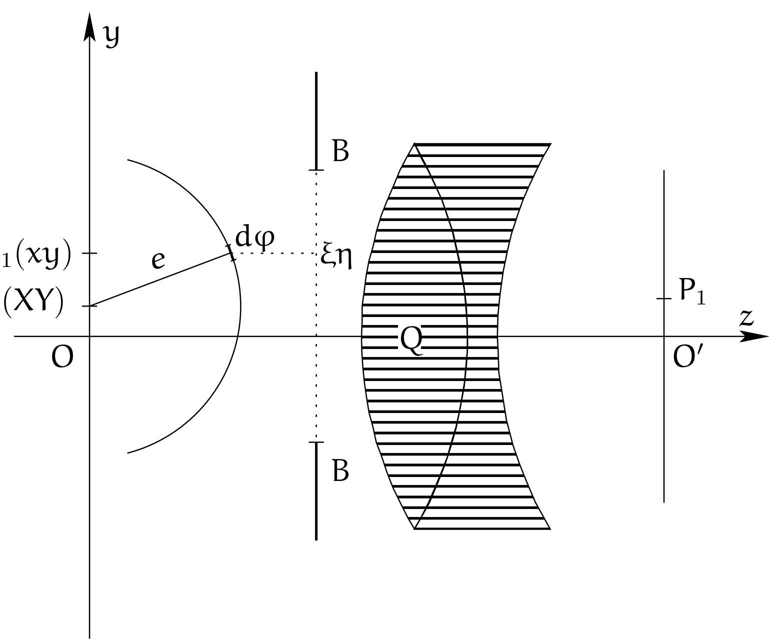
It should be pointed out once more that one obtains the "effective piece of boundary surface I" as one draws from the luminous point or surface element all possible rays toward the boundary points on the entrance pupil. The entirety of the intersections of these rays with the spherical surface I form the boundary of the "effective piece." Integration in the expression of s is extended over the projection of this "effective piece" onto the $\xi\eta$-plane.

§17. Calculation of diffraction on an aperture of specific form for points in the plane conjugate to the object plane in the presence of a luminous surface element

We choose the form of the diffracting aperture in such a way that *the projection of the effective piece of the boundary surface onto the $\xi\eta$-plane is* *a rectangle*. The diffracting aperture in this case, as easily calculated, is bordered by four hyperbolae and approximates better the form of a rectangle the smaller the dimensions of the aperture.

Let OO' (Fig. 26) be the optical axis of the imaging system Q, and O be the origin of the rectangular coordinate system whose z-axis coincides with the optical axis; let the y-axis be pointed toward the top, and the x-axis toward the back. Let the xy-plane be the object plane containing a luminous surface element df at L with coordinates XY. Let the plane perpendicular to OO' and containing O' be the image plane conjugate to the object plane, and the observation point lie at P_1. Let the ray-limiting aperture be represented by the physical

Figure 26



and perpendicular-to-the-z-axis standing diaphragm BB in front of the imaging system Q. Let the radius of the L-centered sphere that we choose as the boundary surface be e; let the luminous element be always so close to the axis that the quadratic terms in x and X, and y and Y can be neglected. Let $d\varphi$ be an element of the boundary surface and its projection on the plane of the diaphragm have the coordinates $\xi\eta$. Then the light disturbance at point P_1 situated close to the z-axis is given by the expression

$$s = \frac{k}{\lambda} \int\limits_{\xi_1}^{\xi_2} \int\limits_{\eta_1}^{\eta_2} \frac{d\xi \, d\eta}{e^2} \sin 2\pi \left(\frac{t}{T} - \frac{x'\xi + y'\eta}{e\lambda} \right) , \tag{27}$$

where

$$\begin{aligned} x' &= x - X \\ y' &= y - Y \end{aligned} \tag{27a}$$

are the coordinates of point L_1, which is conjugate to the observation point P_1, if one refers to them[15] using the luminous element at L as the starting point, and the integration is extended over the rectangular projection of the effective pieces of the boundary surface. One sets

$$\begin{aligned} \xi' &= \xi/e \\ \eta' &= \eta/e \end{aligned} \tag{28}$$

and Eq. 27 becomes

$$s = \frac{k}{\lambda} \int\limits_{\xi'_1}^{\xi'_2} \int\limits_{\eta'_1}^{\eta'_2} d\xi' \, d\eta' \sin 2\pi \left(\frac{t}{T} - \frac{x'\xi' + y'\eta'}{\lambda} \right) . \tag{29}$$

[15] It should be emphasized that these *relative* coordinates $x'y'$ are not identical with the absolute coordinates $x'y'$ of P_1 used in previous paragraphs.

If one decomposes the sine function into its components,

$$\sin 2\pi\left(\frac{t}{T}-\frac{x'\xi'}{\lambda}\right)\cos 2\pi\frac{y'\eta'}{\lambda}-\cos 2\pi\left(\frac{t}{T}-\frac{x'\xi'}{\lambda}\right)\sin 2\pi\frac{y'\eta'}{\lambda}\,,$$

one can carry out the integrations with respect to ξ' and η' separately and obtain[xlii]

$$s=\frac{k}{\lambda}\frac{\sin 2\pi x'\frac{\xi_2'-\xi_1'}{2\lambda}}{\pi\frac{x'}{\lambda}}\cdot\frac{\sin 2\pi y'\frac{\eta_2'-\eta_1'}{2\lambda}}{\pi\frac{y'}{\lambda}}$$
$$\cdot\sin 2\pi\left(\frac{t}{T}-\frac{x'(\xi_2'+\xi_1')+y'(\eta_2'+\eta_1')}{2\lambda}\right)\,.$$

The two integrations will no longer be independent of each other if the projection of the effective boundary surface deviates from the shape of the rectangle.

A simplification occurs if the aperture lies symmetrically with respect to the *z*-axis. In this case,

$$\frac{\xi_1'+\xi_2'}{2}=0 \text{ and } \frac{\eta_1'+\eta_2'}{2}=0\,.$$

If one further sets

$$\xi_2-\xi_1=2\alpha \text{ and } \eta_2-\eta_1=2\beta\,,$$

where α and β denote the half width and height of the projection of the boundary surface, we have

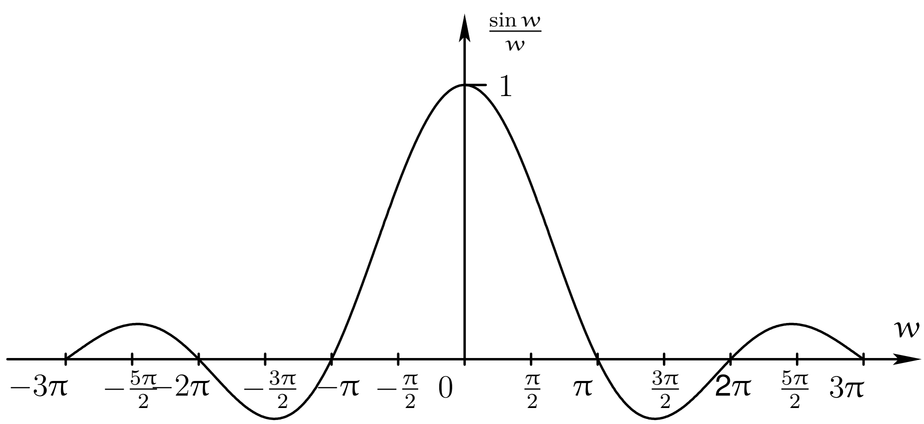
$$\frac{\xi_2'-\xi_1'}{2}=\frac{\alpha}{e}=\alpha' \text{ and } \frac{\eta_2'-\eta_1'}{2}=\frac{\beta}{e}=\beta'\,,$$

where α' and β' are the sines of the aperture angle of the half width and height of the projection of the symmetrical diaphragm. We then have

$$s=\frac{k}{\lambda}4\alpha'\beta'\frac{\sin 2\pi\frac{x'\alpha'}{\lambda}}{2\pi\frac{x'\alpha'}{\lambda}}\cdot\frac{\sin 2\pi\frac{y'\beta'}{\lambda}}{2\pi\frac{y'\beta'}{\lambda}}\sin 2\pi\frac{t}{T}\,. \tag{30}$$

The amplitude of the oscillation s, whose phase is given by $\sin 2\pi\frac{t}{T}$, consists of, apart from a constant, the product of two factors of the form $F(w) = \frac{\sin w}{w}$. The graph of this function of w is indicated in Fig. 27. For $w = \pm\mathfrak{a}\pi$ ($\mathfrak{a} = 1, 2, 3\ldots$), $F(w) = 0$; for $w = 0$, $F(w)$ takes on the undetermined expression $0/0$, whose true value is one.

Figure 27



Without further ado, one can see from the form of the function that the amplitude has its maximum at $w = 0$ and decreases gradually from there toward both sides symmetrically with increasing $|w|$. Whereas the first factor

$$\frac{\sin 2\pi\frac{x'\alpha'}{\lambda}}{2\pi\frac{x'\alpha'}{\lambda}}$$

depicts the amplitude in directions parallel to the x-axis, the second factor,

$$\frac{\sin 2\pi\frac{y'\beta'}{\lambda}}{2\pi\frac{y'\beta'}{\lambda}},$$

independent from the first, reproduces the course of the amplitude in directions parallel to the y-axis. One thus sees that the amplitude of the oscillation is arranged in a checkered way and symmetrically with respect to the lines $x' = 0$ and $y' = 0$ (or $x = X$ and $y = Y$). The amplitude is zero (minimum) on lines

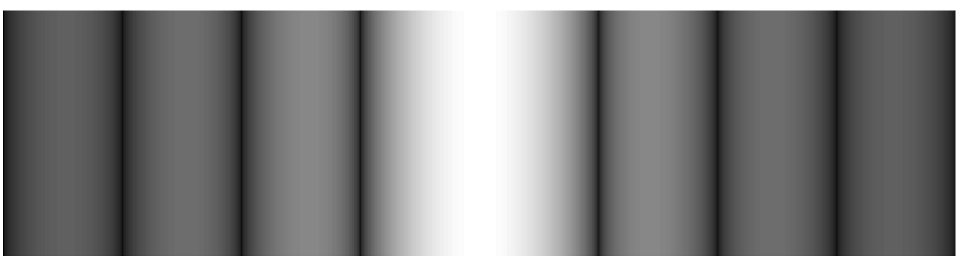
$$x' = \pm a \frac{\lambda}{2\alpha'} \; (a = 1, 2, 3 \ldots)$$

and

$$y' = \pm a \frac{\lambda}{2\beta'} \; (a = 1, 2, 3 \ldots) \,.$$

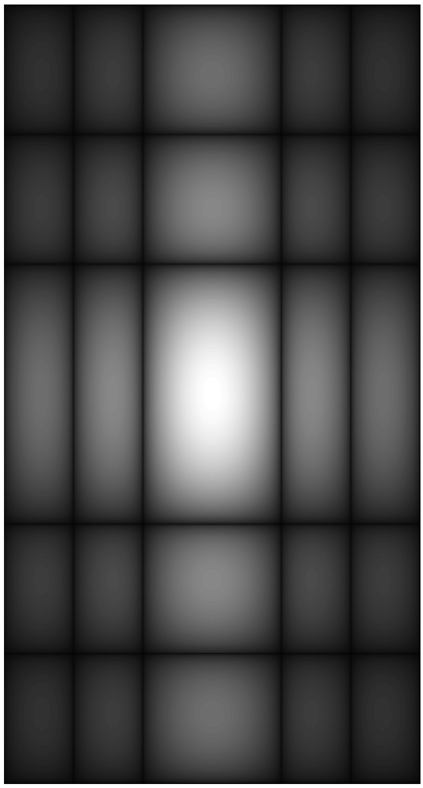
These lines form a system of rectangles in which the amplitude increases gradually from the sides to the middle and has its maximum there (the cross point of the diagonals). The closer the rectangle is situated to the center of the pattern, the greater the maximum. In the central rectangle, the amplitude reaches its absolute maximum (Fig. 28) at the position of the luminous element ($x' = 0$, $y' = 0$).

Figure 29



One can see from the equations for the lines of minima that the smaller the dimension of α', defined for the angular "width" of the diffracting aperture, the farther the lines parallel to the y-axis move

Figure 28



away from each other, and the distance of the lines parallel to the x-axis depends on β' (angular "height") in the same way.

If, for example, the width (α) is negligible compared to the height (β), i.e., the diffracting aperture is formed by a vertical narrow slit, the distribution of the amplitude then takes on the appearance sketched in Fig. 29.[xliii] The intensity distribution of the actually observed diffraction phenomenon emerges from the obtained amplitude distribution if one squares the amplitude at every location, for in general, $J = \overline{s^2}\, df$.

Chapter 3

Imaging of illuminated objects

§18. Presence of several luminous points

In the presence of *one* luminous surface element, the diffraction pattern is symmetrical with respect to the location of that element. This applies to an arbitrarily located surface element, as long as one limits oneself to points close to the axis of the system. *The diffraction pattern always remains stationary and moves with the luminous surface element.*

With the simultaneous presence of several luminous elements, the observed diffraction pattern depends on whether the individual elements emit independent *incoherent* waves from each other, or whether the waves emitted from individual elements are *coherent, i.e., capable of interference*.

The following laws hold, assuming that we are dealing with several luminous "*points*": *If different wave trains are incoherent, one obtains the resulting intensity at each location by simply summing the squares of the amplitudes, i.e., the intensities, that are generated by individual luminous points.*

If n luminous "points" contribute to the light disturbance at the observation point, and if the disturbance generated by their wave trains are represented by the value of the electric field (of the light vector),

$$\begin{aligned}
\mathfrak{E}_1 &= a_1 \cos\left(2\pi\frac{t}{T} + \delta_1\right) \\
\mathfrak{E}_2 &= a_2 \cos\left(2\pi\frac{t}{T} + \delta_2\right) \\
&\ldots\ldots \\
\mathfrak{E}_n &= a_n \cos\left(2\pi\frac{t}{T} + \delta_n\right) \ ,
\end{aligned}$$

then the resulting intensity in the case of *incoherent* wave trains is

$$J_{\mathrm{inc}} = \overline{\mathfrak{E}_1^2} + \overline{\mathfrak{E}_2^2} + \cdots + \overline{\mathfrak{E}_n^2} \ ,$$

and is within an insignificant proportionality factor $1/2$ given by

$$J_{\mathrm{inc}} = a_1^2 + a_2^2 + \cdots + a_n^2 \ .$$

On the other hand, if the wave trains are *coherent* and their electric field vectors $\mathfrak{E}$ have almost the same direction, which we assume for the sake of simplicity, then one has to first add the individual fields at the observation point to yield

$$\mathfrak{E} = \mathfrak{E}_1 + \mathfrak{E}_2 + \cdots + \mathfrak{E}_n \ .$$

The intensity is then given by

$$J_{\mathrm{coh}} = \overline{\mathfrak{E}^2} \ .$$

If we bring $\mathfrak{E}$ after summation into the form

$$\mathfrak{E} = A\cos 2\pi\frac{t}{T} + B\sin 2\pi\frac{t}{T} \ ,$$

the intensity is therefore, to within a factor of $1/2$,[xliv]

$$J_{\mathrm{coh}} = A^2 + B^2 \ .$$

The difference in intensity calculation for both cases is most striking for the observation point that is reached by all wave trains with the same phase. Then we have

$$J_{inc} = a_1^2 + a_2^2 + \cdots + a_n^2 \tag{31}$$

in contrast to

$$J_{coh} = (a_1 + a_2 + \cdots + a_n)^2 \,. \tag{32}$$

If additionally the amplitudes of the individual waves are of equal magnitude (a), we have

$$\begin{aligned} J_{inc} &= n \cdot a^2 \,, \\ J_{coh} &= n^2 a^2 = n \cdot J_{inc} \,. \end{aligned}$$

If $J_{coh} > J_{inc}$ for one observation point, then there must necessarily be another point for which the wave trains do *not* arrive with the same phase, and we have $J_{coh} < J_{inc}$. *This, however, is simply the nature of interference.*

§19. Presence of several luminous surface elements

In reality, we do not deal with luminous points but surface elements. We want to represent the disturbance caused by a luminous surface element df at observation point P by the previously used auxiliary vector s that is proportional to the electric field, giving us the intensity via form $\overline{s^2}\, df$. Let

$$s_P = a \cos\left(2\pi \frac{t}{T} + \delta\right) \,,$$

where we assume that all wave trains originating from the surface element have combined physically at the location of the observation point to a single wave train with amplitude a and phase $\left(2\pi\frac{t}{T} + \delta\right)$.

An extended luminous *surface* consists of many surface elements. The calculation of intensity at the observation point must therefore also be executed differently in the presence of a luminous surface, depending on whether the constituent surface elements emit coherent or incoherent wave trains. In the case of incoherence, the intensity is simply

$$J_{inc} = \int \overline{s^2}\, df$$

or to within a factor

$$J_{inc} = \int a^2\, df \;, \tag{33}$$

where the integration extends over the luminous surface. If a is equal for all surface elements, then we have

$$J_{inc} = a^2 \int df = a^2 F\,, \tag{33a}$$

where F is the size of the surface. In the case of coherence, on the other hand, one has to first calculate according to Huygens' principle the induced disturbance over the entire luminous surface at the observation point, that is, to form

$$S = \int \cos\left(2\pi\frac{t}{T} + \delta\right) df \;, \tag{34}$$

where again the integration extends over the luminous surface. Hereupon, one has to bring S into the canonical form

$$S = A\cos 2\pi\frac{t}{T} + B\sin 2\pi\frac{t}{T}\,. \tag{35}$$

The intensity at the observation point is then

$$J_{coh} = A^2 + B^2\,. \tag{36}$$

If there exists an observation point at which all wave trains arrive with equal phase and amplitude, then we get

$$S = a\cos\left(2\pi\frac{t}{T} + \delta\right) \int df = a\cos\left(2\pi\frac{t}{T} + \delta\right) F\,,$$

where $\delta = \text{const}$. If we bring S to the form of Eq. 35, we have

$$A = aF\cos\delta$$
$$B = aF\sin\delta$$

and the intensity is

$$J_{coh} = a^2 \cdot F^2 \,. \tag{36a}$$

§20. Single luminous slit

Let the slit run parallel to the y-axis (vertically) and extend from $Y = -b$ to $Y = +b$; let its width be small compared to its height and therefore be designated as dX.

I. If the slit is covered with *self-luminous* surface elements, we are dealing with incoherent wave trains. The intensity at the location of the resulting diffraction pattern is to be calculated according to Eq. 33 and becomes, if one substitutes x' with $x - X$ and y' with $y - Y$, according to Eq. 30,

$$J_{inc} = \left(\frac{k}{\lambda}4\alpha'\beta'\right)^2 dX \left(\frac{\sin 2\pi\frac{(x-X)\alpha'}{\lambda}}{2\pi\frac{(x-X)\alpha'}{\lambda}}\right)^2 \int\limits_{Y=-b}^{Y=+b} dY \left(\frac{\sin 2\pi\frac{(y-Y)\beta'}{\lambda}}{2\pi\frac{(y-Y)\beta'}{\lambda}}\right)^2 . \tag{37}$$

If we set

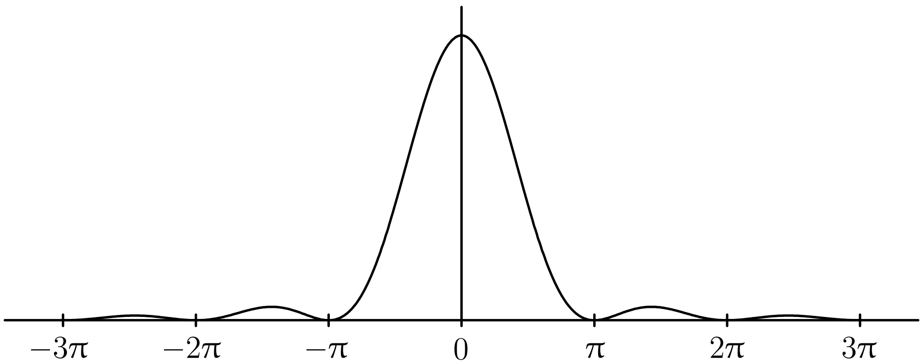
$$\frac{2\pi(y-Y)\beta'}{\lambda} = w\,,$$

then the integral appearing in Eq. 37 becomes

$$-\frac{\lambda}{2\pi\beta'} \int\limits_{2\pi(y+b)\frac{\beta'}{\lambda}}^{2\pi(y-b)\frac{\beta'}{\lambda}} \left(\frac{\sin w}{w}\right)^2 dw = +\frac{\lambda}{2\pi\beta'} \int\limits_{2\pi(y-b)\frac{\beta'}{\lambda}}^{2\pi(y+b)\frac{\beta'}{\lambda}} \left(\frac{\sin w}{w}\right)^2 dw\,.$$

The graph of the function $\left(\frac{\sin w}{w}\right)^2$ is shown schematically in Fig. 30. The function becomes zero at the same locations as

Figure 30



the function $\frac{\sin w}{w}$ that was previously discussed in more detail; the greatest maximum of w, having a value of one, also lies at $w = 0$, whereas the secondary maxima are consistently smaller than those of the function $\frac{\sin w}{w}$, and the entire curve lies above the w-axis because of its quadratic character.

The integral is represented by the areal content between the w-axis and the segment of the curve that is cut out by lines

$$w_1 = 2\pi(y - b)\beta'/\lambda$$

and

$$w_2 = 2\pi(y + b)\beta'/\lambda \, .$$

The limits of the integral are different depending on the location of the observation point xy relative to the luminous slit. If we define as "*slit zone*" the areal strip formed by moving the slit parallel to itself in both directions of the x-axis, we can distinguish three cases: the observation point lies outside the slit zone, in the immediate vicinity of its borders, or within the slit zone.

1. $y > +b$ or $y < -b$ and $|y - b|$ is large compared to $\lambda' = \lambda/\beta'$; i.e., the observation point lies a considerable number of wavelengths away from the edges to the outside. Then we can set both limits of the integral to infinity, in fact both positive if $y > b$ and both negative if $y < b$. The integral here becomes negligibly small.
2. $y = \pm b$: In this case, the limits of the integral become 0 and ∞ or ∞ and 0, and the integral itself takes on the value $\pi/2$ since we know[xlv]

$$\int_0^\infty \left(\frac{\sin w}{w}\right)^2 dw = \pi/2\,. \tag{38}$$

3. $y < b$ and $y > -b$ and further $|b - y|$ large compared to $\lambda' = \lambda/\beta'$; i.e., the observation point lies within the slit zone, but a considerable number of wavelengths away from the edges. In this case we can replace the limits of the integral by $-\infty$ and $+\infty$, and the integral takes on the value of π.

For the intensity in Eq. 37, the integral under consideration is multiplied by a function of x; accordingly, the intensity of light is zero for all points outside the slit zone (case 1). For points in the slit zone and near the borders (cases 3 and 2), the intensity depends only on x and drops suddenly to half the value if the observation point moves for constant x into one of the edges of the slit zone.

The intensity in the direction along the x-axis is given by the expression

$$J_{\text{inc}} = C \cdot \left(\frac{k}{\lambda} 4\alpha'\beta'\right)^2 dX \cdot \frac{\lambda}{2\pi\beta'} \left(\frac{\sin 2\pi\frac{(x-X)\alpha'}{\lambda}}{2\pi\frac{(x-X)\alpha'}{\lambda}}\right)^2 , \tag{39}$$

where $C = 0$ for case 1, $C = \pi/2$ for case 2, and $C = \pi$ for case 3. This functional dependence is, apart from a constant factor, the one schematically drawn in Fig. 30.

II. If the slit is covered with *illuminated* (i.e., not self-luminous) surface elements, then we are dealing with *coherent* wave trains. We therefore have to calculate the intensity according to Eqs. 34, 35, and 36, so that we obtain

$$J_{\text{coh}} = \left[\frac{k}{\lambda} 4\alpha'\beta' \, dX \, \frac{\sin 2\pi\frac{(x-X)\alpha'}{\lambda}}{2\pi\frac{(x-X)\alpha'}{\lambda}} \int\limits_{-b}^{+b} dY \cdot \frac{\sin 2\pi\frac{(y-Y)\beta'}{\lambda}}{2\pi\frac{(y-Y)\beta'}{\lambda}}\right]^2 ; \tag{40}$$

if we set $\frac{2\pi(y-Y)\beta'}{\lambda} = w$, the integral becomes

$$+\frac{\lambda}{2\pi\beta'} \int\limits_{2\pi\frac{y-b}{\lambda'}}^{2\pi\frac{y+b}{\lambda'}} \frac{\sin w}{w} \, dw \; .$$

The function $\frac{\sin w}{w}$ has the graph drawn in Fig. 27. Since the curve lies partly below the w-axis, the sign of the areal patches represented by the integral changes, and the value of the integral therefore approaches a finite limit as w increases, faster than the integral in case I, all else being equal.

To find the intensity versus position, we have to consider as well the three cases separately, where the observation point is inside, outside, and on the edges of the slit zone. Since this integral is

$$\int_{-\infty}^{+\infty} \frac{\sin w}{w} \, \mathrm{d}w = \pi \tag{41}$$

again, the resulting diffraction pattern has exactly the same appearance as in the case of the self-luminous slit. For homologous points the intensity differs only by a constant factor, and at the edges of the slit zone it goes to zero via the half-value even faster for the illuminated slit than in the case of the self-luminous slit.

§21. Two parallel and neighboring slits

Each of the two slits shall again be assumed to be *infinitely narrow*. Let their distance Δ be finite but of arbitrary value. As before, we would like to treat the case of two self-luminous slits separately from the case in which the slits receive their light from an external source. In the latter case, we also need to discuss the influence on the diffraction pattern exerted by the position of the light source on the illuminated slits. This is because only with oblique illumination do noticeable differences between diffraction patterns of self-luminous and illuminated double slits become evident.

I. *Self-luminous slits.* Each slit generates the diffraction pattern that was discussed in § 20 under I, whose appearance is completely identical for both slits. The center of each individual diffraction pattern coincides with the center of the slit that generates it. Thus, we are dealing with the superposition of two identical diffraction patterns whose principal maxima are separated from each other in the direction of the x-axis by the distance Δ of the two light slits. Since we are dealing with a self-luminous double slit, the resulting intensity at each location is therefore the sum of the intensities caused by each luminous slit.

This expression is given by the formula

$$J_{\text{inc}} = \text{const.}\left[\frac{\sin 2\pi\frac{(x-X_1)\alpha'}{\lambda}}{2\pi\frac{(x-X_1)\alpha'}{\lambda}}\right]^2 + \text{const.}\left[\frac{\sin 2\pi\frac{[x-(X_1+\Delta)]\alpha'}{\lambda}}{2\pi\frac{[x-(X_1+\Delta)]\alpha'}{\lambda}}\right]^2, \tag{42}$$

where X_1 is the abscissa of the first luminous slit and $X_1 + \Delta$ is the abscissa of the second luminous slit.

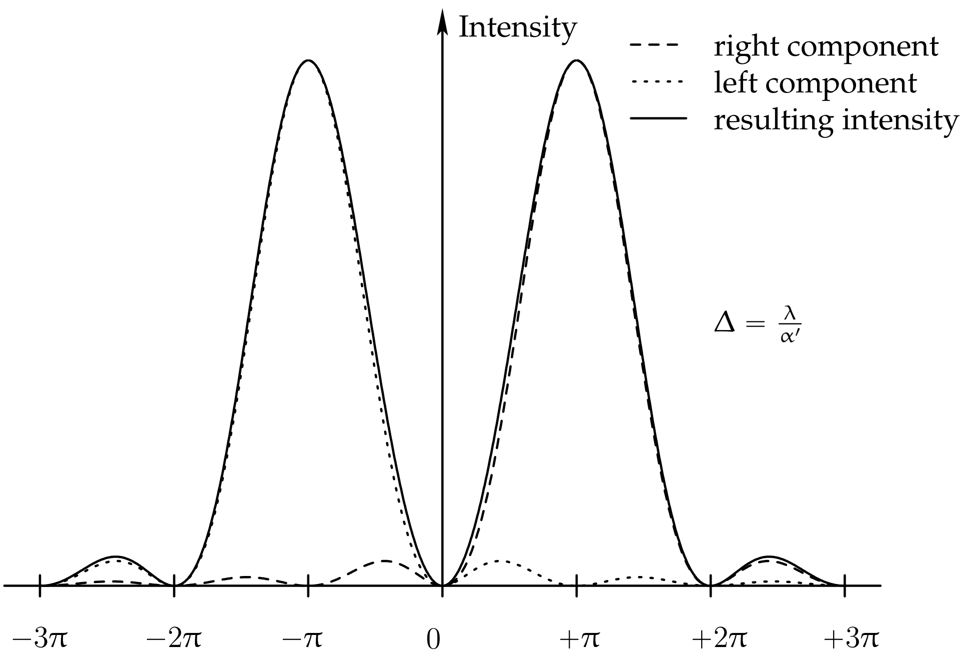
We want to carry out the discussion of this expression only for the two special cases $\Delta = \lambda/\alpha'$ and $\Delta = \lambda/2\alpha'$.

1. $\Delta = \lambda/\alpha'$. Then the expression for the resulting intensity becomes

$$J_{\text{inc}} = \text{const.}\left[\frac{\sin 2\pi\frac{x-X_1}{\Delta}}{2\pi\frac{x-X_1}{\Delta}}\right]^2 + \text{const.}\left[\frac{\sin\left(2\pi\frac{x-X_1}{\Delta} - 2\pi\right)}{2\pi\frac{x-X_1}{\Delta} - 2\pi}\right]^2.$$

We recognize easily that the two intensity curves are simply shifted along the x-axis by a distance 2π (Fig. 31). Each of the principal maxima coincides with the second minimum of the other curve, while the first minima coincide and bisect the distance $\Delta = \lambda/\alpha'$. By summing the ordinates we obtain the resulting intensity curve, which is shown as the solid line in the figure. This curve exhibits two principal maxima separated by the distance of the two luminous slits ($\Delta = \lambda/\alpha'$), and a steady and symmetrical decrease in brightness that reaches the value zero in the middle between the principal maxima. Going outward on both sides there is a series of secondary maxima that are separated from each other by complete minima.

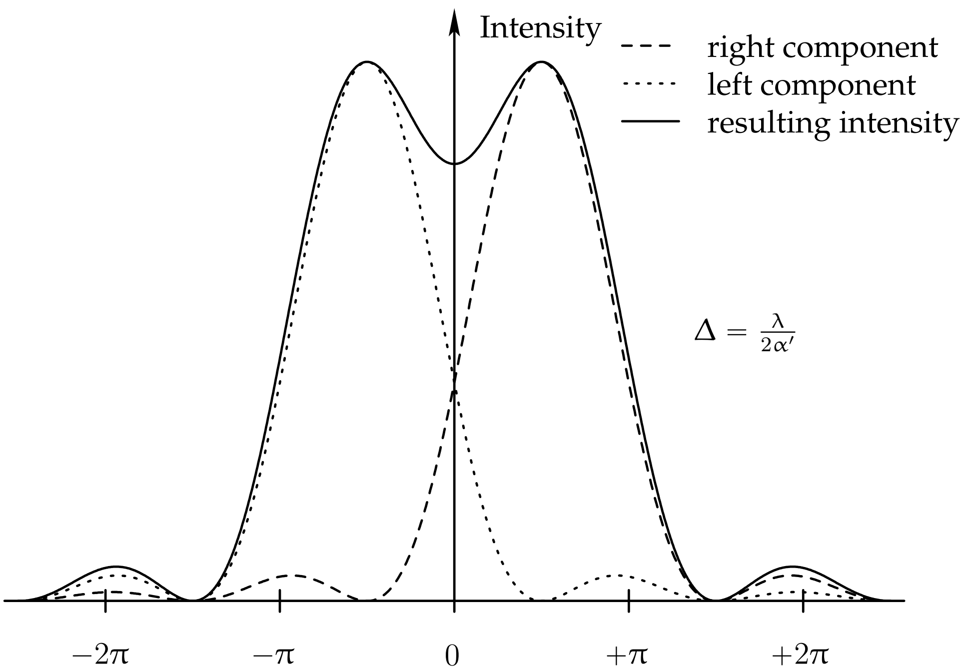
Figure 31



2. $\Delta = \lambda/2\alpha'$. In this case, an analogous observation shows that by superposing the two intensity curves, both principal maxima merge into a single, correspondingly wider central strip that exhibits a small intensity attenuation in the center. The first secondary maxima are still clearly noticeable (Fig. 32).

II. *Illuminated slits.* In this case, we are dealing with two infinitely narrow slits of finite separation that receive their light from an external source. As such, we would like to consider the intensely bright filament of a light bulb that is located in the focal plane of an objective lens, so that *plane* waves are emitted.

Figure 32



Let the filament be parallel to the direction of the slits. If the axis of this collimator is perpendicular to the plane of the slits, then coherent wave trains are emitted from there with zero phase difference. Their phase difference deviates from zero, however, if the axis of the collimator is tilted with respect to the plane of the slits.

We first consider the case of normal incidence. If we designate the angle of incidence of the light rays by u, then this case is characterized by $u = 0$.

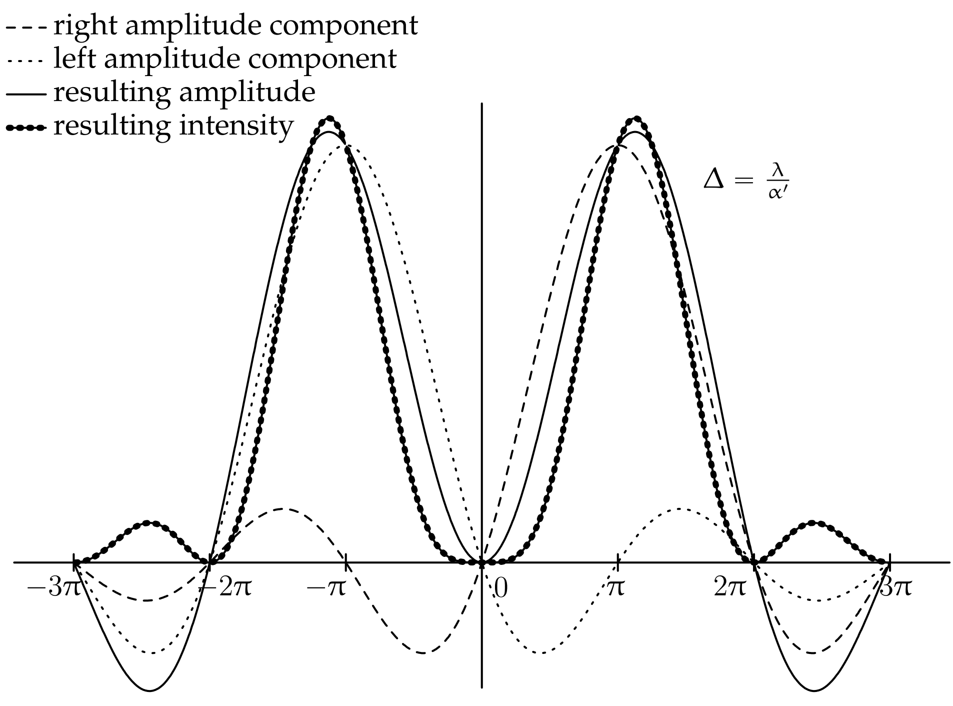
A. $u = 0$. The resulting intensity in the case of coherent wave trains is given by the expression

$$J_{coh} = \left[\text{const.} \frac{\sin 2\pi \frac{(x-X_1)\alpha'}{\lambda}}{2\pi \frac{(x-X_1)\alpha'}{\lambda}} + \text{const.} \frac{\sin 2\pi \frac{[x-(X_1+\Delta)]\alpha'}{\lambda}}{2\pi \frac{[x-(X_1+\Delta)]\alpha'}{\lambda}} \right]^2, \tag{43}$$

where the previous designations are kept. For this case, too, we would like to discuss in more detail this expression for the two special cases $\Delta = \lambda/\alpha'$ and $\Delta = \lambda/2\alpha'$.

1. $\Delta = \lambda/\alpha'$. In this case, both amplitude curves are shifted from each other by 2π in the direction of the x-axis and drawn in Fig. 33.

Figure 33



By summing the ordinates algebraically one obtains the resulting amplitude, and by squaring it one obtains the intensity of the resulting diffraction pattern. One can easily see that the two principal maxima are separated by a perfect minimum. The decrease of intensity toward this minimum is happening here more rapidly than in the analogous case of self-luminous slits. Going outward, the principal maxima are followed once again by secondary maxima, which in turn are separated from each other by perfect minima. The intensities of the corresponding secondary maxima are of greater magnitude than in the former case.

2. $\Delta = \lambda/2\alpha'$. For this case, Fig. 34 shows the respective position of the two amplitude curves. A consideration analogous to the above teaches us that the two principal maxima again merge into a single bright central strip that, in contrast to the analogous case of self-luminous slits, is brighter and drops faster, whereas, conversely, the secondary maxima are evidently much weaker than the former.

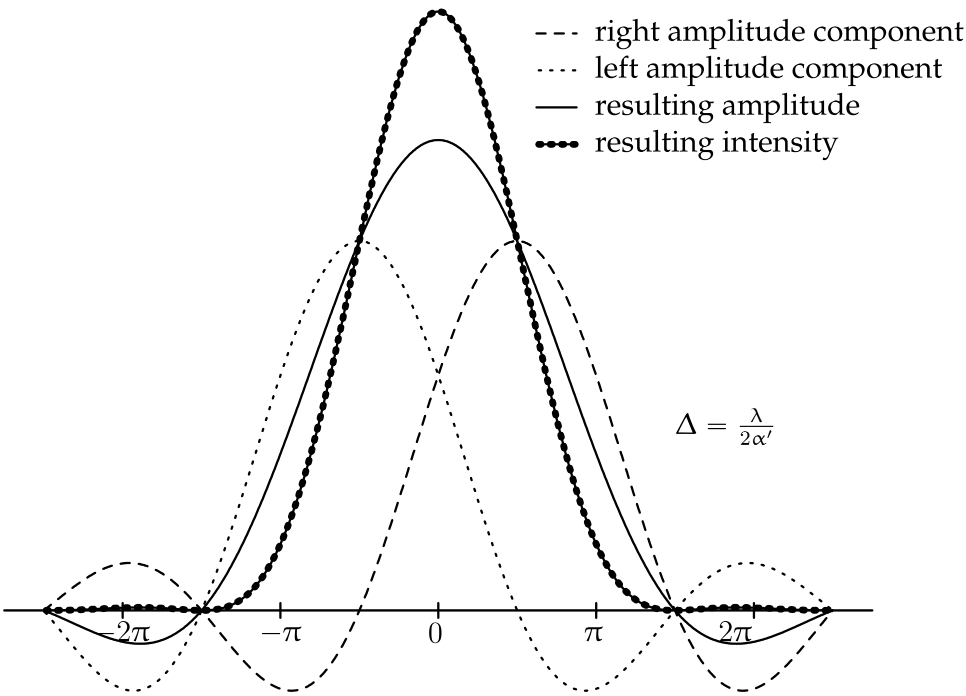
B. Angle of incidence $u > 0$. In Fig. 35, let Sl_1 and Sl_2 be the locations of the two slits of separation Δ, which are met by light at an angle u. As before, let the slits be so narrow that the phase can be considered constant even under oblique incidence of light. The path difference for them is therefore

$$\Delta \sin u \,,$$

so that the coherent disturbances emanating from Sl_1 and Sl_2 can be represented by

$$\alpha \sin 2\pi \left(\frac{t}{T} + \frac{1}{2} \frac{\Delta \sin u}{\lambda} \right)$$

Figure 34



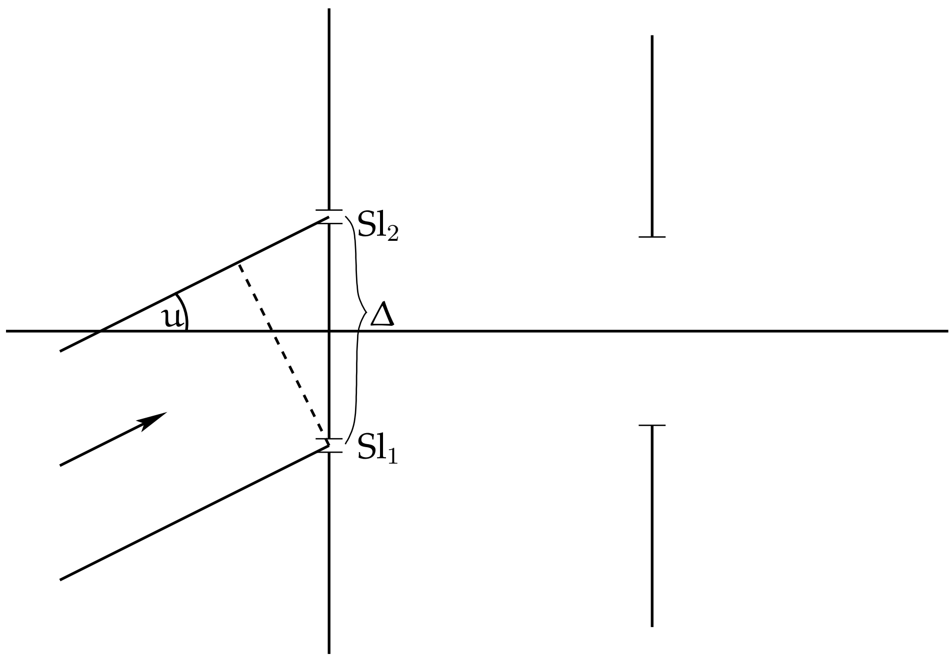
and

$$\alpha \sin 2\pi \left(\frac{t}{T} - \frac{1}{2} \frac{\Delta \sin u}{\lambda} \right) .$$

If only slit Sl_1 is present, then, according to earlier explanations, the disturbance at the observation point is

$$s_1 = \text{const} \frac{\sin w_1}{w_1} \sin 2\pi \left(\frac{t}{T} + \frac{1}{2} \frac{\Delta \sin u}{\lambda} \right) .$$

Figure 35



If only the slit Sl$_2$ is present, then the light disturbance at the same observation point is

$$s_2 = \text{const}\frac{\sin w_2}{w_2}\sin 2\pi\left(\frac{t}{T} - \frac{1}{2}\frac{\Delta\sin u}{\lambda}\right) .$$

The values of w_1 and w_2 are the same as those in the previously treated case of perpendicular incident light, into which our present case transitions when $u = 0$. Thus

$$\begin{cases} w_1 = \frac{2\pi(x-X_1)\alpha'}{\lambda} \\ w_2 = \frac{2\pi[x-(X_1+\Delta)]\alpha'}{\lambda} \end{cases} .$$

If both slits act simultaneously, the light disturbance at the observation point is then given by

$$\begin{aligned} S = s_1 + s_2 &= \text{const}\left(\frac{\sin w_1}{w_1} + \frac{\sin w_2}{w_2}\right)\cos\pi\frac{\Delta\sin u}{\lambda}\sin 2\pi\frac{t}{T} \\ &+ \text{const}\left(\frac{\sin w_1}{w_1} - \frac{\sin w_2}{w_2}\right)\sin\pi\frac{\Delta\sin u}{\lambda}\cos 2\pi\frac{t}{T} \\ &= A\sin 2\pi\frac{t}{T} + B\cos 2\pi\frac{t}{T}\,, \end{aligned}$$

so that the intensity becomes

$$\begin{aligned} J_{coh} = A^2 + B^2 = \text{const}^2 &\left[\left(\frac{\sin w_1}{w_1}\right)^2 + \left(\frac{\sin w_2}{w_2}\right)^2\right. \\ &\left. +2\frac{\sin w_1}{w_1}\frac{\sin w_2}{w_2}\cos 2\pi\frac{\Delta\sin u}{\lambda}\right]. \end{aligned} \tag{44}$$

It is readily apparent that this expression becomes identical with that for two *self*-luminous slits of equal separation Δ (see § 21, I) in case the cosine disappears. This is the case for

$$\frac{2\pi\Delta\sin u}{\lambda} = \pm(2a+1)\frac{\pi}{2}\,, \qquad a = 0, 1, 2,$$

i.e., for

$$\sin u = \pm\frac{(2a+1)\lambda}{4\Delta}\,.$$

We further see that the expression assumes likewise a very simple form if the cosine becomes $+1$ or -1. The former occurs for

$$\frac{2\pi\Delta\sin u}{\lambda} = \pm 2a\pi\,, \qquad a = 0, 1, 2,$$

i.e., for

$$\sin u = \pm \frac{a\lambda}{\Delta} .$$

We then have

$$J_{coh} = \text{const}^2 \left[\frac{\sin w_1}{w_1} + \frac{\sin w_2}{w_2} \right]^2 . \tag{45}$$

This intensity distribution, occurring periodically with variation of only u, is thus identical to that for two illuminated slits of the same separation Δ for normal incidence ($u = 0$).

The cosine becomes -1 for

$$\frac{2\pi\Delta \sin u}{\lambda} = \pm(2a + 1)\pi , \; a = 0, 1, 2,$$

i.e., for

$$\sin u = \pm \frac{(2a + 1)\lambda}{2\Delta} .$$

In this case, we have

$$J_{coh} = \text{const}^2 \left[\frac{\sin w_1}{w_1} - \frac{\sin w_2}{w_2} \right]^2 . \tag{46}$$

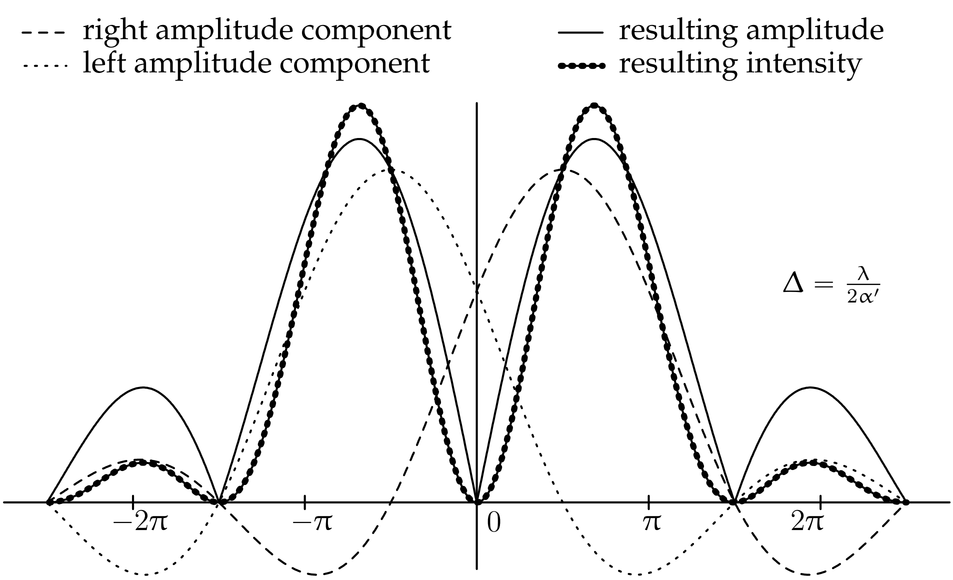
Whereas in the previous cases of coherent waves the resulting disturbance was obtained by adding the amplitudes, here the interesting case arises that the amplitudes of the individual fields are to be *subtracted* in order to obtain the resulting disturbance.

One consequence of this is the particularly noticeable difference, at these angles of incidence of light, between the diffraction pattern of self-luminous and illuminated slits of equal separation. This difference appears particularly striking for the special case $\Delta = \frac{\lambda}{2\alpha'}$ in which the angle of incidence must be

$$\sin u = (2\mathfrak{a} + 1)\alpha' \, .$$

For this case, Fig. 36 shows the respective position of the two amplitude curves. The resulting amplitude of the diffraction pattern is represented by the solidly drawn curve.[xlvi] It can be seen that the two principal maxima are separated by a perfect minimum, whereas in the self-luminous slits and also in the illuminated slit with normal incidence, the principal maxima are merged into a single and correspondingly broader bright central strip.

Figure 36



§22. An illuminated slit of finite width

If the slit is self-luminous, the result is easy to assess. The slit of finite width can be thought of as the result of shifting an infinitely narrow slit parallel to the x-axis. One therefore only needs to construct the diffraction pattern corresponding to the infinitely narrow slit situated at different positions and then *add the individual intensities* at each location. With the broadening of the self-luminous slit, the diffraction pattern of an infinitely narrow slit must become more and more unclear.[1]

Much more diverse are the phenomena of an *illuminated* slit of finite width. The term that gives the light disturbance at the observation point, in the case of an illuminated slit, is

$$s = \frac{k}{\lambda} \int_{-a}^{+a} \int_{-b}^{+b} 4\alpha'\beta' \, dX \, dY \, \frac{\sin 2\pi\alpha' \frac{x-X}{\lambda}}{2\pi\alpha' \frac{x-X}{\lambda}} \cdot \frac{\sin 2\pi\beta' \frac{y-Y}{\lambda}}{2\pi\beta' \frac{y-Y}{\lambda}} \cdot \sin 2\pi \frac{t}{T} , \tag{47}$$

where $2b$ and $2a$ denote the height and width of the illuminated slit. If the slit is infinitely narrow, the integration over dX becomes unnecessary and the integrand moves as a constant to the front of the integral, a case that has already been dealt with in § 20. For an infinitely narrow slit, the position of the light source, i.e., the direction of the angle of incidence of light, is of no influence on the diffraction pattern. In the case of a finite width of the slit, on the other hand, the

[1]This is the typical difference between a diffraction phenomenon and a pure interference phenomenon with a self-luminous slit (Lummer–Haidinger interference curves of equal inclination), in which only the angular magnitude of the visual field grows with the broadening of the light source (slit).

oblique incidence of the rays brings about phase differences along dX, so that in this more general case the light disturbance becomes

$$\left.\begin{aligned} s = \frac{k}{\lambda} \int\limits_{-a}^{+a} \int\limits_{-b}^{+b} 4\alpha'\beta'\, dX\, dY \, \frac{\sin 2\pi\alpha' \frac{x-X}{\lambda}}{2\pi\alpha' \frac{x-X}{\lambda}} \\ \cdot \frac{\sin 2\pi\beta' \frac{y-Y}{\lambda}}{2\pi\beta' \frac{y-Y}{\lambda}} \cdot \sin 2\pi \left(\frac{t}{T} - \frac{X \sin u}{\lambda} \right) \end{aligned}\right\} , \tag{48}$$

where u is the angle of incidence of the incoming plane wave.

This expression can be written in the following form:

$$\left.\begin{aligned} s = \frac{k}{\lambda} \int\limits_{-b}^{+b} dY \, 2\beta' \frac{\sin 2\pi\beta' \frac{y-Y}{\lambda}}{2\pi\beta' \frac{y-Y}{\lambda}} \\ \cdot \int\limits_{-a}^{+a} dX \, 2\alpha' \frac{\sin 2\pi\alpha' \frac{x-X}{\lambda}}{2\pi\alpha' \frac{x-X}{\lambda}} \cdot \sin 2\pi \left(\frac{t}{T} - \frac{X \sin u}{\lambda} \right) \end{aligned}\right\} . \tag{49}$$

This form reflects the formation of the resulting disturbance at the observation point. Consider the slit as a checkered pattern consisting of individual surface elements of size $dX\, dY$; the above form, when calculating the disturbance at the observation point, initially takes into account only the influence of surface elements located on a strip parallel to the y-axis with width dX and height $2b$, so that the first integral in itself represents the already treated case of an infinitely narrow illuminated slit. As we know, the value of this integral is, to within a constant, equal to π for observation points within the "slit zone" (see § 20).

The slit of finite width may be assembled purely from such strips whose effect at the observation point is a function of the location of the single strip and the prevailing phase there; i.e., it is a function of

X. This influence of the width is taken into account by the second integral.

In the calculation of s, we first restrict ourselves to the case in which the phase is the same for all individual strips, i.e., we assume normal incidence of light ($u = 0$). Then we have to consider the following integral:

$$J = \int_{-a}^{+a} dX \, \frac{\sin 2\pi\alpha' \frac{(x-X)}{\lambda}}{2\pi\alpha' \frac{(x-X)}{\lambda}} \ .$$

To solve this integral, we employ an artifice. It is known that

$$\frac{\sin 2\pi\alpha'\mu}{\pi\mu} = \int_{-\alpha'}^{+\alpha'} \cos(2\pi\mu\nu) \, d\nu \ .$$

So if we set

$$\mu = \frac{x - X}{\lambda} \ ,$$

the integral becomes

$$J = \frac{1}{2\alpha'} \int_{-a}^{+a} dX \int_{-\alpha'}^{+\alpha'} \cos\left(2\pi\nu\frac{x - X}{\lambda}\right) d\nu \ ,$$

and by switching the order of integration,

$$J = \frac{1}{2\alpha'} \int_{-\alpha'}^{+\alpha'} d\nu \int_{-a}^{+a} dX \cos\left(2\pi\nu\frac{x - X}{\lambda}\right) \ .$$

Now we can carry out the integration over X and get

$$J = \frac{1}{2\alpha'} \int\limits_{-\alpha'}^{+\alpha'} d\nu \, \frac{\sin 2\pi\nu\frac{x+a}{\lambda} - \sin 2\pi\nu\frac{x-a}{\lambda}}{2\pi\frac{\nu}{\lambda}}$$

$$= \int\limits_{-\alpha'}^{+\alpha'} d\nu \, \frac{\cos 2\pi\nu\frac{x}{\lambda} \cdot \sin 2\pi\nu\frac{a}{\lambda}}{2\pi\frac{\nu}{\lambda}\alpha'} \, .$$

If we set

$$2\pi\nu\frac{a}{\lambda} = w \, ,$$

then we obtain

$$J = \frac{\lambda}{2\pi\alpha'} \int\limits_{-2\pi a\frac{\alpha'}{\lambda}}^{+2\pi a\frac{\alpha'}{\lambda}} dw \, \frac{\cos\left(\frac{x}{a}w\right)\sin w}{w} \, . \tag{50}$$

We can see that this integral is a function of x; we would like to compare it with the integral

$$J_0 = \frac{1}{\pi} \int\limits_{-\infty}^{+\infty} dw \, \frac{\cos\left(\frac{x}{a}w\right)\sin w}{w} \, . \tag{51}$$

To find the value of the integral in Eq. 51, we start with the task of determining a function of x such that it takes on the value of 1 between $x = -a$ and $x = +a$, and the value 0 everywhere else.

In general, according to the Fourier integral theorem,[xlvii]

$$f(x) = \frac{1}{\pi} \int\limits_{0}^{\infty} dz \int\limits_{-\infty}^{+\infty} f(u) \cos z(u-x) \, du \, . \tag{52}$$

The function that we seek is therefore

$$f(x) = \frac{1}{\pi}\int_{0}^{\infty} dz \int_{-a}^{+a} \cos z(u-x)\,du$$
$$= \frac{1}{\pi}\int_{0}^{\infty} dz\,\frac{2}{z}\sin(az)\cos(zx)\,,$$

or, if we set additionally $az = w$,

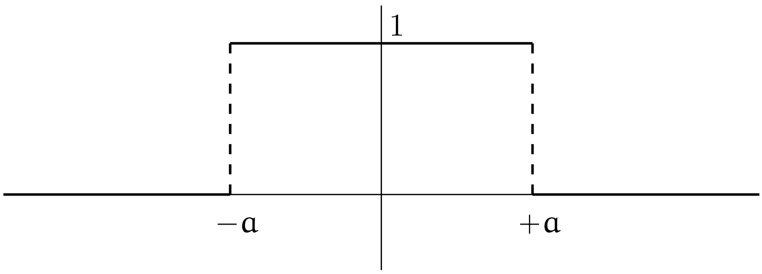
$$f(x) = \frac{2}{\pi}\int_{0}^{\infty} dw\,\frac{\sin w \cos\left(\frac{x}{a}w\right)}{w}$$
$$= \frac{1}{\pi}\int_{-\infty}^{+\infty} dw\,\frac{\sin w \cos\left(\frac{x}{a}w\right)}{w} = J_0\,.$$

The value of J_0 as a function of x is therefore

$$\left.\begin{array}{l} J_0 = 0 \text{ for } \begin{cases} x > -\infty \text{ and } < -a \\ x > +a \end{cases} \\ J_0 = 1 \text{ for } \begin{cases} x > -a \text{ and} \\ x < +a \end{cases} \\ J_0 = \frac{1}{2} \text{ for } \begin{cases} x = +a \\ x = -a \end{cases} \end{array}\right\}\,. \tag{53}$$

Its graph is represented by solid lines in Fig. 37. If $a\alpha'$ is much greater than λ, then $J = \frac{\lambda}{2\alpha'}J_0$ and the light distribution in the resulting diffraction pattern is a *uniformly* bright strip of width $2a$, outside of which there is complete darkness. This light distribution in the image becomes all the more congruent to that of the object (the illuminated

Figure 37



slit), the greater the width a for a given opening angle α' of the diffracting aperture or the larger the opening angle for a given slit width.

To gain an overview as to what values of the limits permit the use of the integral J_0 instead of the integral J, we consider the following:

$$\int\limits_{-2\pi\frac{a\alpha'}{\lambda}}^{+2\pi\frac{a\alpha'}{\lambda}} = \int\limits_{-\infty}^{+\infty} - \int\limits_{-\infty}^{-2\pi\frac{a\alpha'}{\lambda}} - \int\limits_{+2\pi\frac{a\alpha'}{\lambda}}^{+\infty} .$$

Since the function to be integrated is an even function, the last two integrals on the right are the same and we can write

$$J = \frac{\lambda}{2\alpha'} J_0 - \frac{\lambda}{\pi\alpha'} \int\limits_{\frac{2\pi a\alpha'}{\lambda}}^{\infty} dw \, \frac{\sin w \cos\left(\frac{x}{a} w\right)}{w} , \tag{54}$$

so that the amplitude of the resulting disturbance becomes

$$\text{const} \left\{ J_0 - \frac{2}{\pi} \int\limits_{\frac{2\pi a\alpha'}{\lambda}}^{\infty} dw \, \frac{\sin w}{w} \cos\left(\frac{x}{a} w\right) \right\} .$$

The integrand of the residual integral differs from the previously discussed $\left(\frac{\sin w}{w}\right)$ only by a factor $\cos\left(\frac{x}{a}w\right)$, which takes on the maximum value one. The residual integral is therefore, for all values of x, less than the integral

$$\mathfrak{P} = \frac{\lambda}{\alpha'\pi} \int\limits_{\frac{2\pi a\alpha'}{\lambda}}^{+\infty} dw\, \frac{\sin w}{w}\ .$$

If, for certain values of $2\pi\frac{a\alpha'}{\lambda}$, this integral is negligible with respect to the same integral between $-\infty$ and $+\infty$, then we have a stronger reason to neglect our residual integral in comparison to J_0. The following table shows the values of the integral as a function of its lower limit $2\pi\frac{a\alpha'}{\lambda}$:

$2\pi\frac{a\alpha'}{\lambda}$	$\frac{\alpha'\pi}{\lambda}\mathfrak{P}$	$2\pi\frac{a\alpha'}{\lambda}$	$\frac{\alpha'\pi}{\lambda}\mathfrak{P}$
0	1.5708	20	0.0226
1	0.6247	50	0.0192
2	−0.0346	100	0.0086
5	0.0209	200	0.0024
10	−0.0875	500	−0.0018

It can be seen from the table that $\mathfrak{P}$ decreases very rapidly and is practically zero for a value of $2\pi\frac{a\alpha'}{\lambda} = 2$.

If, for example, half the opening angle is equal to 3°, so that α' becomes approximately equal to $1/20$, then the lower limit of $\mathfrak{P}$ equals $\pi a/10\lambda$; further, if $a = 666\lambda$, or equal to $4\,\text{mm}$ for a wavelength of $\lambda = 0.6\,\mu\text{m}$, then $\mathfrak{P} = 0.0024 \cdot \frac{\lambda}{\alpha'\pi}$ and therefore $J = \frac{\lambda}{2\alpha'}\{J_0 - 0.0016\}$ according to Eq. 54.

We can also write the amplitude of the resulting disturbance as

$$A(x) = \text{const}\frac{2}{\pi} \int\limits_{0}^{\frac{2\pi a\alpha'}{\lambda}} dw\, \frac{\sin w}{w} \cos\left(\frac{x}{a}w\right)\ . \tag{55}$$

For $x = 0$, i.e., in the middle of the slit, the value of the amplitude is then

$$A(0) = \text{const}\frac{2}{\pi}\int\limits_0^{\frac{2\pi a\alpha'}{\lambda}} dw\,\frac{\sin w}{w}\,,$$

which transitions to "const" for large $\frac{2\pi a\alpha'}{\lambda}$. At the edge of the slit, for $x = a$, we get

$$A(a) = \text{const}\frac{2}{\pi}\int\limits_0^{\frac{2\pi a\alpha'}{\lambda}} dw\,\frac{\sin w\cos w}{w} = \text{const}\frac{2}{\pi}\cdot\frac{1}{2}\int\limits_0^{\frac{4\pi a\alpha'}{\lambda}} \frac{\sin w'}{w'}\,dw'$$

if we set $2w = w'$. For large values of $\frac{2\pi a\alpha'}{\lambda}$ this value $= 1/2$ const, or half the value at the center. In general, this simple relationship between $A(0)$ and $A(a)$ does not exist, and the values of $A(0)$ and $2\cdot A(a)$, respectively, are apparent from Figs. 38a and b (hatched).

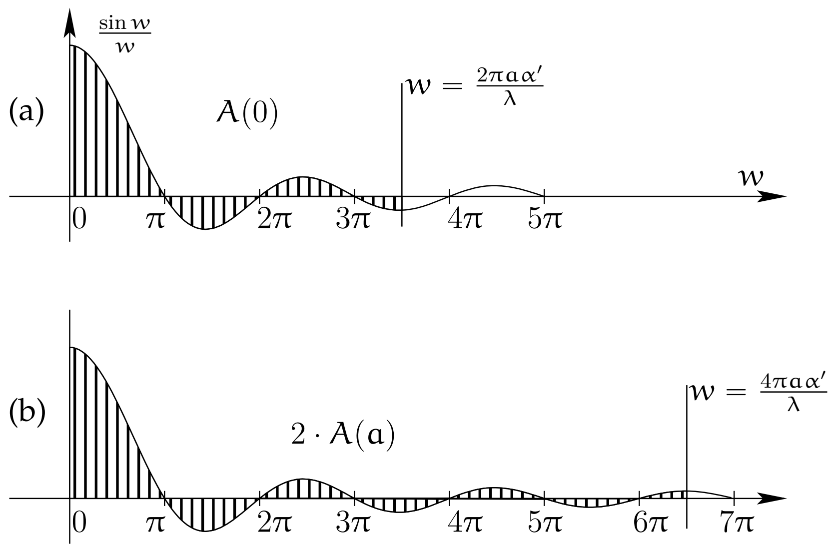
It is easy to see that in the general case, for which we cannot set $\frac{2\pi a\alpha'}{\lambda} = \infty$, the amplitude A for x inside and outside the slit fluctuates. To recognize this, we set[xlviii]

$$\begin{aligned}\frac{dA(x)}{dx} &= \frac{d}{dx}\left\{\frac{2}{\pi}\int\limits_0^{\frac{2\pi a\alpha'}{\lambda}} dw\,\frac{\sin w}{w}\cos\left(\frac{x}{a}w\right)\right\}\\ &= -\frac{2}{\pi}\int\limits_0^{\frac{2\pi a\alpha'}{\lambda}} dw\sin w\sin\left(w\frac{x}{a}\right)\\ &= +\text{const}\left\{\frac{\sin u}{u} - \frac{\sin v}{v}\right\}\,,\end{aligned}$$

where

$$u = \frac{2\pi\alpha'(a+x)}{\lambda}\,,\, v = \frac{2\pi\alpha'(a-x)}{\lambda}\,.$$

Figure 38



Let us fix, for given values of a and α', the point $\frac{2\pi\alpha' a}{\lambda}$ on the abscissa (Fig. 39), which corresponds to the point $x = 0$ (the middle of the slit), and let us go from this point to the right and left of the axis a distance $\frac{2\pi\alpha' x}{\lambda}$. Then we have in the ordinates the values of $\frac{\sin u}{u}$ and $\frac{\sin v}{v}$, whose difference is to be formed.

To fix this idea, let us choose, for example

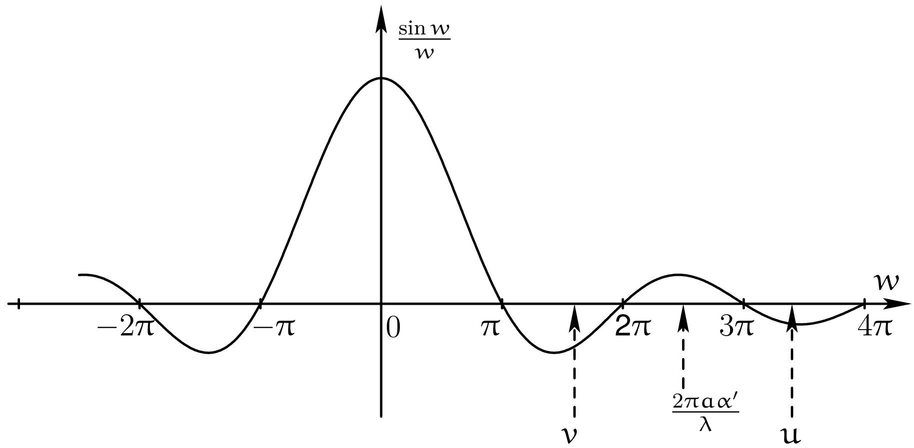
$$\frac{2\pi a \alpha'}{\lambda} = 2\pi\,,$$

so it is easy to see that if we let x grow from zero, first

$$\frac{\sin u}{u} - \frac{\sin v}{v}\,, \text{ i.e., } \frac{dA(x)}{dx}$$

is positive until it grows to a maximum value, then decays, and for $u = 3\pi$, $v = \pi$, i.e., for $x = a/2$, it is again zero. From there,

Figure 39



$\frac{dA(x)}{dx}$ becomes negative and reaches its largest negative value for $u = 4\pi$, $v = 0$, i.e., for $x = a$ at the edge of the slit. If x is allowed to grow beyond the edge of the slit, $\frac{dA(x)}{dx}$ increases again from its minimum value and reaches the value 0 for $u = 5\pi$, $v = -\pi$, i.e., $x = 3/2a$; in this way, the fluctuations of $\frac{dA(x)}{dx}$ continue and gradually die down.

Accordingly, the amplitude distribution will look somewhat like what is shown in Fig. 40.

If we choose $\frac{2\pi a\alpha'}{\lambda} = \pi$, the graph of the amplitude $A(x)$ in the interior of the slit is somewhat different; the maximum is then at $x = 0$ (see Fig. 41).

If $\frac{2\pi a\alpha'}{\lambda}$ is very small compared to π, then we can place in the expression for $A(x)$ the nearly constant factor $\frac{\sin w}{w} = 1$ in front of the integral and get

Figure 40

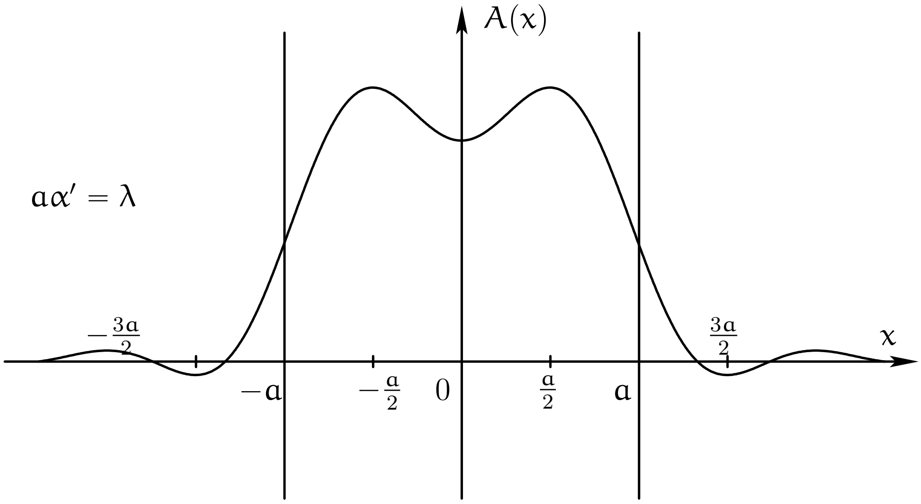
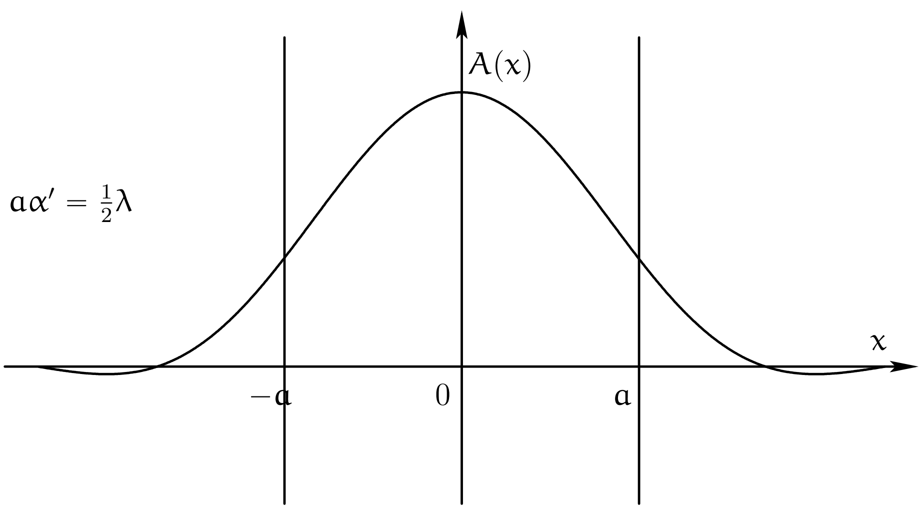


Figure 41



$$
\begin{aligned}
A(x) &= \frac{2}{\pi} \int\limits_{0}^{\frac{2\pi a \alpha'}{\lambda}} dw \cos\left(\frac{xw}{a}\right) \\
&= \frac{2a}{\pi x} \cdot \sin\left(\frac{2\pi\alpha' x}{\lambda}\right) \\
&= \frac{4a\alpha'}{\lambda} \cdot \frac{\sin\left(\frac{2\pi\alpha' x}{\lambda}\right)}{\frac{2\pi\alpha' x}{\lambda}} \,. \qquad (56)
\end{aligned}
$$

$A(x)$ has in this case the already discussed form $\frac{\sin w}{w}$. If $\frac{2\pi a\alpha'}{\lambda}$ is very large compared to π, then, as can be seen from the consideration of the form of $\frac{dA(x)}{dx}$, the fluctuations of the amplitude inside the slit are very small, and the value of the amplitude is therefore almost constant; only *at the edges* of the slit do fluctuations take place; namely (if we consider only positive values of x, since the phenomenon is symmetrical with respect to the J-axis), since $\frac{2\pi a\alpha'}{\lambda}$ was already assumed to be large, u is a fortiori large and therefore:

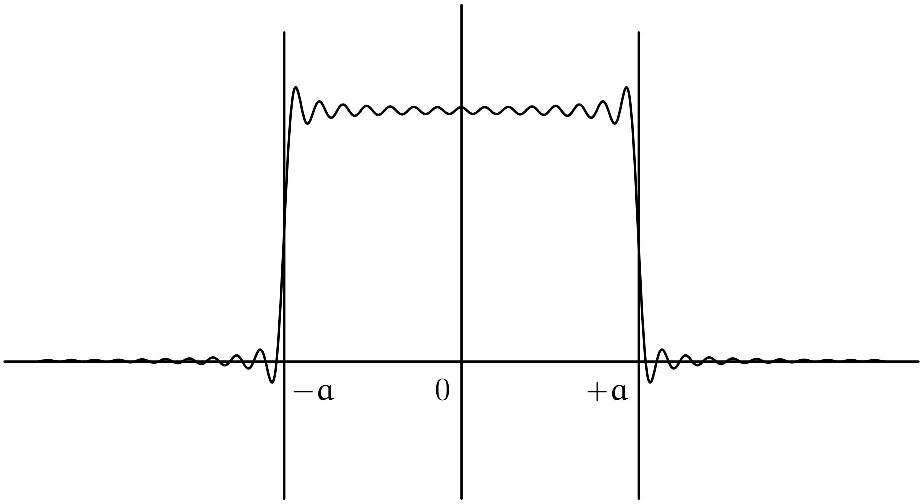
$$\frac{dA(x)}{dx} = -\text{const} \cdot \frac{\sin\nu}{\nu} \,.$$

Therefore, as ν gets closer and closer to the value $\nu = 0$ (as x increases), i.e., $x = a$ (edge of the slit), the fluctuations of $\frac{\sin\nu}{\nu}$ begin to become more and more noticeable. We therefore obtain the image of the amplitude indicated in Fig. 42:[xlix] the larger $\frac{a\alpha'}{\lambda}$ becomes, the more the variations at the edges converge, so that in the limit, for infinitely large $\frac{a\alpha'}{\lambda}$, we obtain the amplitude graph already shown in Fig. 37 above.

§23. Finite slit whose two halves possess a constant difference in phase

Let the slit have width $2a$ and height $2b$; let the phase in the half slit of height $2b$ and width a ($x = -a$ to $x = 0$) be equal to $2\pi\frac{t}{T}$, while

Figure 42



in the other half slit ($x = 0$ to $x = +a$) let it be $2\pi\frac{t}{T} + \delta$. Then the resulting light disturbance at the observation point is

$$s = \frac{k}{\lambda}\int\limits_{-b}^{+b} dY\, 2\beta' \frac{\sin 2\pi\beta' \frac{y-Y}{\lambda}}{2\pi\beta' \frac{y-Y}{\lambda}} \left\{ \int\limits_{-a}^{0} dX\, 2\alpha' \frac{\sin 2\pi\alpha' \frac{x-X}{\lambda}}{2\pi\alpha' \frac{x-X}{\lambda}} \sin 2\pi\frac{t}{T} \right.$$
$$\left. + \int\limits_{0}^{+a} dX\, 2\alpha' \frac{\sin 2\pi\alpha' \frac{x-X}{\lambda}}{2\pi\alpha' \frac{x-X}{\lambda}} \sin\left(2\pi\frac{t}{T} + \delta\right) \right\} . \tag{57}$$

If the observation point lies within the slit zone, then, as was shown previously (§ 20), the integral stretched out over dY becomes equal to λ;[1] if we split up $\sin\left(2\pi\frac{t}{T} + \delta\right)$, we get

$$s = A \sin 2\pi\frac{t}{T} + B \cos 2\pi\frac{t}{T} ,$$

where A and B are given by

$$A = k \cdot \left\{ \int_{-a}^{0} dX\, 2\alpha' \frac{\sin 2\pi\alpha' \frac{x-X}{\lambda}}{2\pi\alpha' \frac{x-X}{\lambda}} + \cos\delta \int_{0}^{+a} dX\, 2\alpha' \frac{\sin 2\pi\alpha' \frac{x-X}{\lambda}}{2\pi\alpha' \frac{x-X}{\lambda}} \right\}$$

$$B = k \sin\delta \int_{0}^{+a} dX\, 2\alpha' \frac{\sin 2\pi\alpha' \frac{x-X}{\lambda}}{2\pi\alpha' \frac{x-X}{\lambda}} \,.$$

To obtain the intensity, we have to form

$$\left.\begin{aligned} I &= A^2 + B^2 \\ \text{or } I &= J_1^2 + J_2^2 + 2J_1 J_2 \cos\delta \,, \\ \text{where } J_1 &= 2\alpha' k \int_{-a}^{0} dX \frac{\sin 2\pi\alpha' \frac{x-X}{\lambda}}{2\pi\alpha' \frac{x-X}{\lambda}} \\ J_2 &= 2\alpha' k \int_{0}^{a} dX \frac{\sin 2\pi\alpha' \frac{x-X}{\lambda}}{2\pi\alpha' \frac{x-X}{\lambda}} \end{aligned}\right\} . \tag{58}$$

We want to treat more special cases.

1. If $\delta = 0, 2\pi, 4\pi$, etc., i.e., the phase difference $0, \lambda, 2\lambda$, etc., then we have $I = (J_1 + J_2)^2 = J^2$, where

$$J = 2\alpha' k \int_{-a}^{+a} dX \frac{\sin 2\pi\alpha' \frac{x-X}{\lambda}}{2\pi\alpha' \frac{x-X}{\lambda}} \,;$$

   i.e., the intensity is of the same value as if the slit had *no* phase difference.

2. If $\delta = \pi, 3\pi, 5\pi \ldots$, i.e., the phase difference is $= \frac{\lambda}{2}, 3\frac{\lambda}{2}, 5\frac{\lambda}{2}$, etc., then we obtain

$$I = (J_1 - J_2)^2 \,. \tag{59}$$

If we set

$$2\pi\alpha' \cdot \frac{x - X}{\lambda} = w\,,$$

then J_1 and J_2 take on the following values:

$$\left.\begin{aligned} J_1 &= \frac{\lambda k}{\pi} \int\limits_{\frac{2\pi\alpha' x}{\lambda}}^{2\pi\alpha' \frac{x+a}{\lambda}} \frac{\sin w}{w}\, dw \\ J_2 &= \frac{\lambda k}{\pi} \int\limits_{2\pi\alpha' \frac{x-a}{\lambda}}^{\frac{2\pi\alpha' x}{\lambda}} \frac{\sin w}{w}\, dw \end{aligned}\right\} . \tag{60}$$

We see immediately that for $x = 0$, i.e., the center of the slit, $J_1 = J_2$ and therefore $I = 0$; in the middle of the slit there is a minimum, independent of the size of a.

To discuss further, we distinguish the two cases for which $\frac{2\pi\alpha' a}{\lambda}$ is small or large compared to π.

I. If $\frac{2\pi\alpha' a}{\lambda}$ is small, then we can expand according to Taylor's theorem as follows:[li]

$$\begin{aligned} J_1 =& \frac{\lambda k}{\pi} \left\{ \frac{\sin \frac{2\pi\alpha' x}{\lambda}}{\frac{2\pi\alpha' x}{\lambda}} \cdot \frac{2\pi\alpha' a}{\lambda} \right. \\ &\left. + \frac{\frac{2\pi\alpha' x}{\lambda} \cos \frac{2\pi\alpha' x}{\lambda} - \sin \frac{2\pi\alpha' x}{\lambda}}{\left(\frac{2\pi\alpha' x}{\lambda}\right)^2} \cdot \frac{\left(\frac{2\pi\alpha' a}{\lambda}\right)^2}{2!} \right\} \\ J_2 =& -\frac{\lambda k}{\pi} \left\{ \frac{\sin \frac{2\pi\alpha' x}{\lambda}}{\frac{2\pi\alpha' x}{\lambda}} \left(-\frac{2\pi\alpha' a}{\lambda}\right) \right. \\ &\left. + \frac{\frac{2\pi\alpha' x}{\lambda} \cos \frac{2\pi\alpha' x}{\lambda} - \sin \frac{2\pi\alpha' x}{\lambda}}{\left(\frac{2\pi\alpha' x}{\lambda}\right)^2} \cdot \frac{\left(-\frac{2\pi\alpha' a}{\lambda}\right)^2}{2!} \right\} , \end{aligned}$$

and get

$$J_1 - J_2 = \frac{\lambda k}{\pi} \frac{\frac{2\pi\alpha' x}{\lambda} \cos \frac{2\pi\alpha' x}{\lambda} - \sin \frac{2\pi\alpha' x}{\lambda}}{\left(\frac{2\pi\alpha' x}{\lambda}\right)^2} \cdot \left(\frac{2\pi\alpha' a}{\lambda}\right)^2 .$$

$$\text{If we set} \begin{cases} \dfrac{\lambda k}{\pi}\left(\dfrac{2\pi\alpha' a}{\lambda}\right)^2 = c \,, \\ \qquad\qquad \dfrac{2\pi\alpha' x}{\lambda} = \xi \,, \end{cases}$$

then we get

$$J_1 - J_2 = c\frac{\xi \cos \xi - \sin \xi}{\xi^2} = cf(\xi) \,.$$

To discuss the curve represented by the odd function

$$f(\xi) = \frac{\xi \cos \xi - \sin \xi}{\xi^2} \,,$$

we first determine its *zeros*. It turns out that

$$f(\xi) = 0 \text{ for } \xi = \tan \xi \,;$$

i.e., the zeros of the curve $f(\xi)$ lie at the intersections of the curves

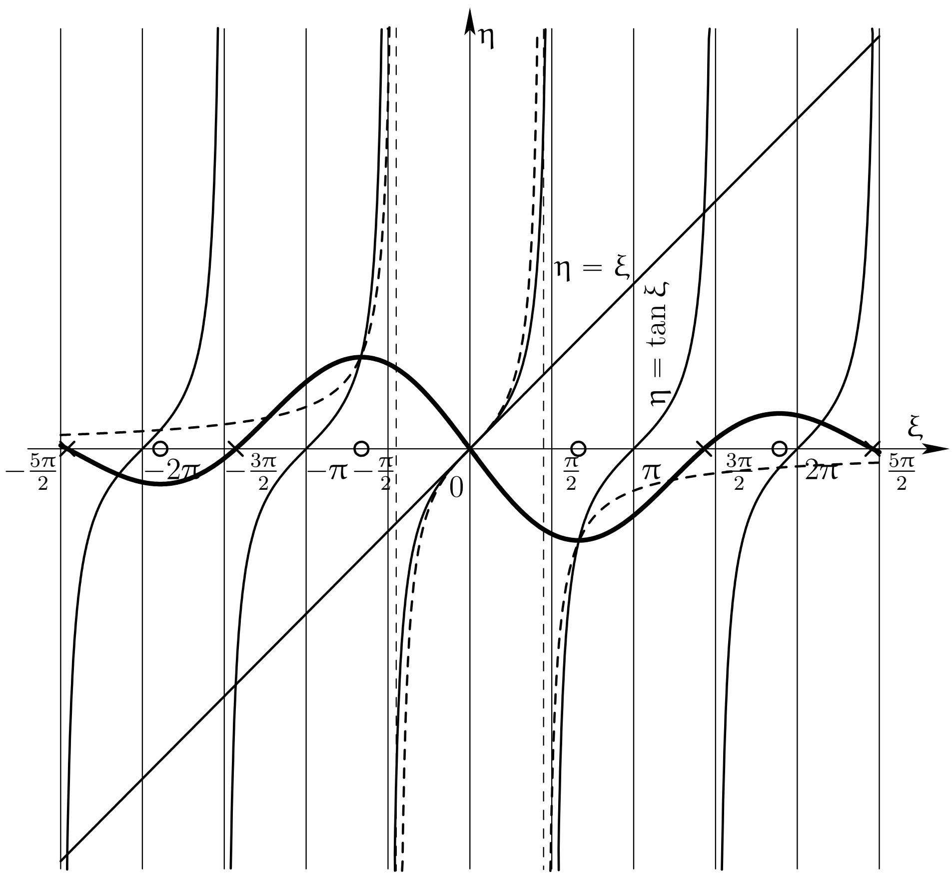
$$\eta = \xi \ \text{ and } \ \eta = \tan \xi \,.$$

The locations indicated by $\times$ in Fig. 43 are the zeros of the function $f(\xi)$; with growing $|\xi|$, the zeros thus approach the values $\pm(2a+1)\frac{\pi}{2}$ more and more closely.

We now determine the positions of the maxima and minima of $f(\xi)$. Its derivative is

$$f'(\xi) = \frac{-\xi^2 \sin \xi - 2\xi \cos \xi + 2 \sin \xi}{\xi^3} \,.$$

Figure 43



The maxima and minima of $f(\xi)$ are therefore at the locations for which

$$\tan \xi = \frac{2\xi}{2 - \xi^2} \, ,$$

i.e., at the intersections of the curves

$$\eta = \tan\xi \text{ and } \eta = \frac{2\xi}{2-\xi^2}\,.$$

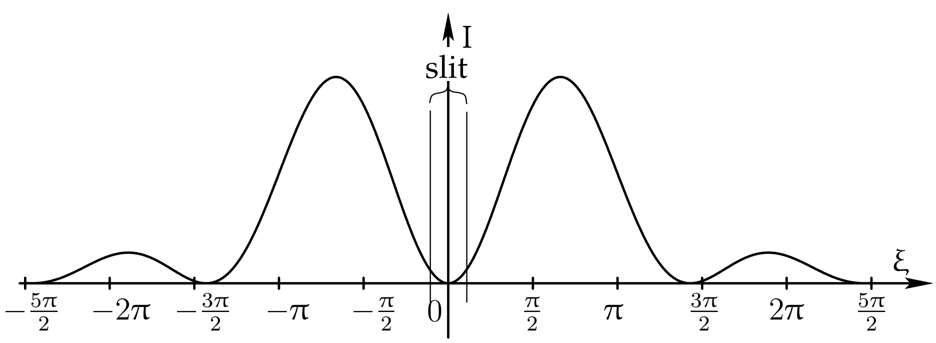
Curve $\eta = \frac{2\xi}{2-\xi^2}$ has the form that is represented by the dashed lines in Fig. 43. The points marked by $\circ$ are therefore the locations of the maxima and minima of $f(\xi)$, and $f(\xi)$ itself is approximately represented by the bold curve. $f(\xi)$ has for negative ξ opposite but equal values as for positive ξ.

If the intensity $I = c^2[f(\xi)]^2$ is formed, the intensity distribution shown in Fig. 44 is obtained. As we can see, two principal maxima appear, separated by a complete minimum and followed by secondary maxima and minima.

By assumption, $\frac{2\pi\alpha' a}{\lambda}$ is small compared to π. It is therefore *a fortiori for points of the object slit that*

$$\xi = \frac{2\pi\alpha' x}{\lambda} \text{ is small compared to } \pi$$

Figure 44



and one obtains the surprising result that the slit itself appears almost completely dark, and that the maxima and minima lie symmetrically on both sides of it.

II. If $\frac{2\pi\alpha' a}{\lambda}$ is large, we need to consider only positive x because the quantity $J_1 - J_2$ changes only its sign for the corresponding negative x, and I thus takes on the same value.

First, suppose

$$\xi = \frac{2\pi\alpha' x}{\lambda} \text{ is small.}$$

Then we can set[lii]

$$J_1 = \frac{\lambda k}{\pi} \int_{\xi}^{\xi + \frac{2\pi\alpha' a}{\lambda}} \frac{\sin w}{w}\, dw = \frac{\lambda k}{\pi} \left\{ \frac{\pi}{2} - \int_0^{\xi} dw \, \frac{w}{w} \right\}$$
$$= \frac{\lambda k}{2} - \frac{\lambda k}{\pi}\xi = \frac{\lambda k}{2} - 2k\alpha' x \,.$$

Likewise,

$$J_2 = \frac{\lambda k}{2} + 2k\alpha' x \,;$$

therefore,

$$J_1 - J_2 = -4k\alpha' x$$

and

$$I = 16k^2\alpha'^2 x^2 \,.$$

Therefore, the lowest minimum is found at $x = 0$; on both sides the intensity grows in a steep, parabolic rise. Since we can put

$$\frac{2\pi a\alpha'}{\lambda} + \xi = \infty \,,$$

we can in general write[liii]

$$J_1 - J_2 = -\frac{2\lambda k}{\pi}\int\limits_0^{\xi}\frac{\sin w}{w}\,dw + \frac{\lambda k}{\pi}\int\limits_{-\infty}^{\xi-\frac{2\pi\alpha' a}{\lambda}}\frac{\sin w}{w}\,dw\ . \quad (61)$$

Therefore, if ξ is large compared to π and the observation point is so far from the edge ($x = a$) of the slit that

$$\left|\xi - \frac{2\pi\alpha' a}{\lambda}\right| = \frac{2\pi\alpha'}{\lambda}|x - a|$$

is still large compared to π, then we can set

$$\xi - \frac{2\pi\alpha' a}{\lambda} = -\infty$$

and get

$$J_1 - J_2 = -\frac{2\lambda k}{\pi}\frac{\pi}{2} = -\lambda k\ .$$

We have therefore inside the slit, except in the immediate vicinity of its center and its edges, a nearly *constant brightness*.[liv]

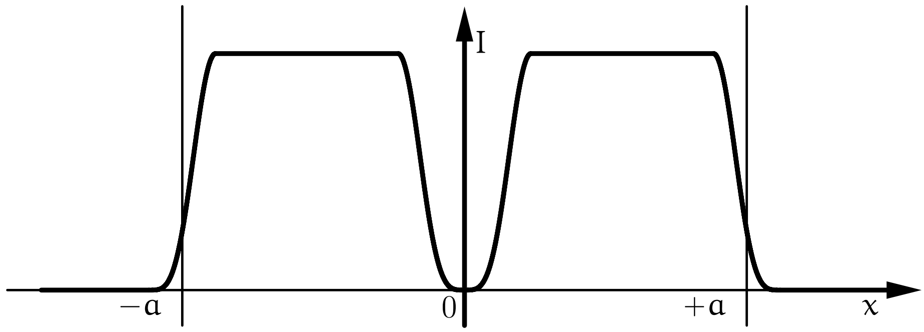
At the edge ($x = a$) we have

$$J_1 - J_2 = -\frac{2\lambda k}{\pi}\frac{\pi}{2} + \frac{\lambda k}{\pi}\frac{\pi}{2} = -\frac{\lambda k}{2}\ .$$

At the edge, therefore, there is only $1/4$ of the intensity that prevails in the slit. If one is *outside* the slit and far enough from its edges, then

$$\xi - \frac{2\pi\alpha' a}{\lambda}$$

Figure 45



is large compared to π and can be set to $+\infty$. We then get

$$J_1 - J_2 = 0 \text{ and therefore also } I = 0 .$$

The graph of the intensity is therefore largely represented by Fig. 45. This is not entirely correct. In fact, fluctuations of I still appear near the center of the slit $x = 0$ and the edges $x = a$. This can easily be recognized as follows.

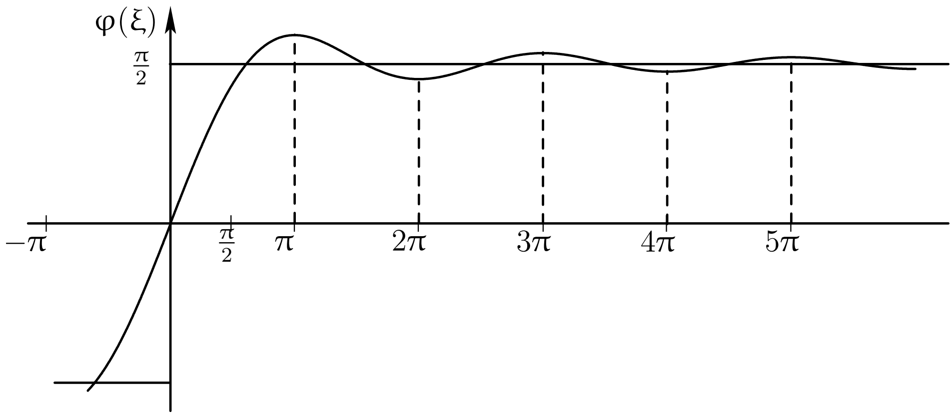
For the sake of simplicity, we base the consideration on the following numerical example:

$$\begin{cases} \dfrac{2\pi\alpha' a}{\lambda} = 50\pi , \\ \alpha' = 1 \text{ minute } = \dfrac{1}{60^2} , \\ \lambda = 6 \cdot 10^{-4}\,\text{mm} , \\ \text{therefore } a = 54\,\text{mm} . \end{cases}$$

Since the graph of

$$\varphi(\xi) = \int_0^{\xi} \frac{\sin w}{w}\, dw$$

Figure 46



is the one sketched in Fig. 46, we see that

$$\begin{aligned} \varphi(\xi) \text{ has a maximum for } \quad & \xi = \pi, 3\pi, 5\pi \cdots \\ \text{has a minimum for } \quad & \xi = 2\pi, 4\pi, 6\pi, \text{ etc.} \end{aligned}$$

Now, even for $\xi = \pi$ or $\xi = 2\pi$,

$$\xi - \frac{2\pi\alpha' a}{\lambda} \begin{cases} \text{equal to} & -49\pi \\ \text{or} & -48\pi \end{cases}$$

is still deeply *negative*, so that in $J_1 - J_2$ the second integral is small. The first, on the other hand, is $= \varphi(\xi)$, and therefore we have, according to Eq. 61,

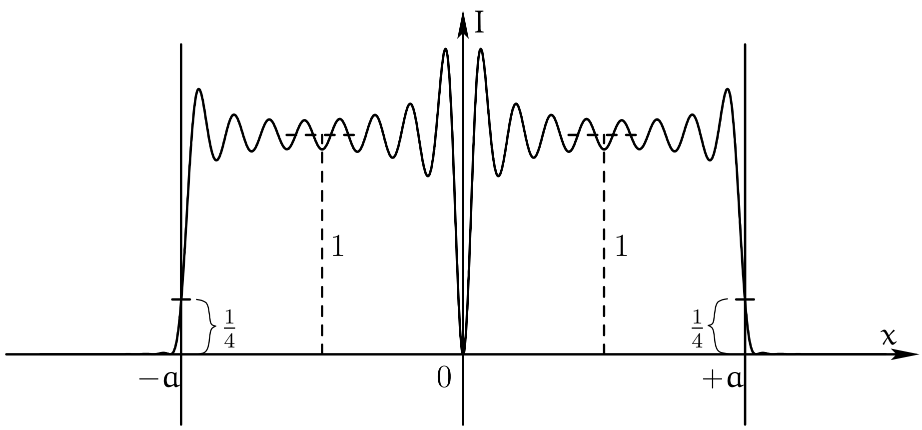
$$J_1 - J_2 = -\frac{2\lambda k}{\pi}\varphi(\xi) \ .$$

I therefore exhibits fluctuations of functional form $[\varphi(\xi)]^2$, so that I assumes a maximum for $\xi = \pi$ and a minimum for $\xi = 2\pi$. The values $\xi = \pi$ and $\xi = 2\pi$, however, correspond to the values

$$x = \frac{\lambda}{2\alpha'} = 1\,\mathrm{mm}$$
$$\text{and} \quad x = \frac{\lambda}{\alpha'} = 2\,\mathrm{mm}, \text{ respectively.}$$

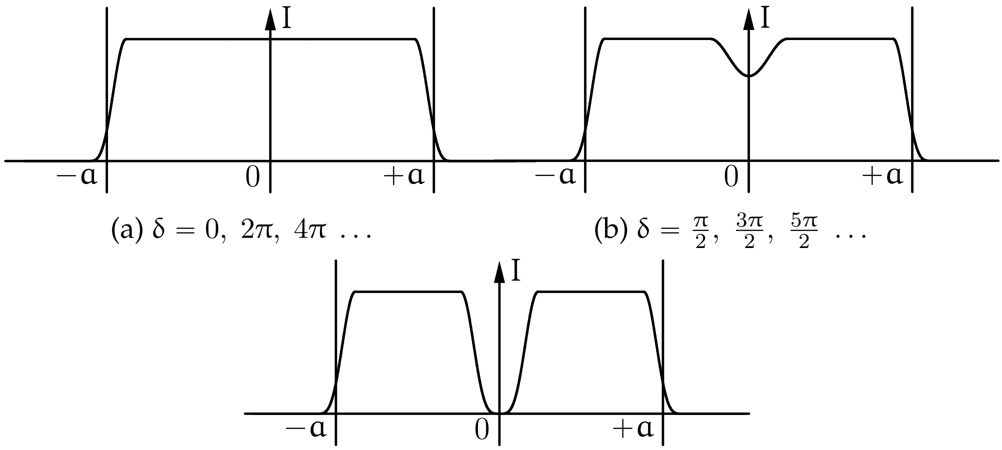
Thus, these "diffraction fringes" close to the center of the slit are still clearly visible. Something quite analogous also occurs at the edges of the slit ($x = \pm a$), as we saw in the previous section. The exact intensity curve will therefore have the form shown in Fig. 47.[lv]

Figure 47



When the phase difference δ of the two halves of the gap increases from 0 to π, the deep minimum in the center only gradually forms (see Figs. 48a, b, and c).

Figure 48



(a) $\delta = 0,\ 2\pi,\ 4\pi\ \ldots$

(b) $\delta = \frac{\pi}{2},\ \frac{3\pi}{2},\ \frac{5\pi}{2}\ \ldots$

(c) $\delta = \pi,\ 3\pi,\ 5\pi\ \ldots$

§24. Slit of finite width with oblique incidence of light

If u is the angle of incidence of the light rays, then the light disturbance at the observation point is

$$s = \frac{k}{\lambda} \left. \begin{aligned} &\int\limits_{-b}^{+b} dY\, 2\beta' \frac{\sin 2\pi\beta' \frac{y-Y}{\lambda}}{2\pi\beta' \frac{y-Y}{\lambda}} \int\limits_{-a}^{+a} dX\, 2\alpha' \frac{\sin 2\pi\alpha' \frac{x-X}{\lambda}}{2\pi\alpha' \frac{x-X}{\lambda}} \\ &\qquad\qquad\qquad\qquad \sin 2\pi \left(\frac{t}{T} - \frac{X \sin u}{\lambda} \right) \end{aligned} \right\} . \tag{62}$$

Therefore, for points within the slit zone, we have[lvi]

$$s = k \int\limits_{-a}^{+a} dX\, 2\alpha' \frac{\sin 2\pi\alpha' \frac{x-X}{\lambda}}{2\pi\alpha' \frac{x-X}{\lambda}} \sin 2\pi \left[\frac{t}{T} - \frac{x \sin u}{\lambda} + \frac{(x-X)\sin u}{\lambda} \right]$$

$$= k\int_{-a}^{+a} dX\, 2\alpha' \frac{\sin 2\pi\alpha' \frac{x-X}{\lambda}}{2\pi\alpha' \frac{x-X}{\lambda}} \cos\left(2\pi \sin u \frac{x-X}{\lambda}\right) \sin 2\pi\left(\frac{t}{T} - \frac{x \sin u}{\lambda}\right)$$

$$+ k\int_{-a}^{+a} dX\, 2\alpha' \frac{\sin 2\pi\alpha' \frac{x-X}{\lambda}}{2\pi\alpha' \frac{x-X}{\lambda}} \sin\left(2\pi \sin u \frac{x-X}{\lambda}\right) \cos 2\pi\left(\frac{t}{T} - \frac{x \sin u}{\lambda}\right) .$$

Because the last factors in both integrals do not contain X, we can write

$$\left.\begin{aligned}
s &= A \sin 2\pi\frac{t'}{T} + B\cos 2\pi\frac{t'}{T} , \\
\text{where } t' &= t - \frac{x\sin u}{c} \\
A &= k\int_{-a}^{+a} dX\, 2\alpha' \frac{\sin 2\pi\alpha' \frac{x-X}{\lambda}}{2\pi\alpha' \frac{x-X}{\lambda}} \cos\left(2\pi \sin u \frac{x-X}{\lambda}\right) \\
B &= k\int_{-a}^{+a} dX\, 2\alpha' \frac{\sin 2\pi\alpha' \frac{x-X}{\lambda}}{2\pi\alpha' \frac{x-X}{\lambda}} \sin\left(2\pi \sin u \frac{x-X}{\lambda}\right)
\end{aligned}\right\} . \tag{63}$$

If we set

$$2\pi\alpha' \frac{x-X}{\lambda} = w ,$$

then we have

$$\left.\begin{aligned}
A &= \frac{k\lambda}{\pi} \int_{2\pi\alpha'\frac{x-a}{\lambda}}^{2\pi\alpha'\frac{x+a}{\lambda}} dw \cdot \frac{\sin w}{w} \cdot \cos\left(\frac{\sin u}{\alpha'} w\right) \\
B &= \frac{k\lambda}{\pi} \int_{2\pi\alpha'\frac{x-a}{\lambda}}^{2\pi\alpha'\frac{x+a}{\lambda}} dw \cdot \frac{\sin w}{w} \cdot \sin\left(\frac{\sin u}{\alpha'} w\right)
\end{aligned}\right\} \tag{64}$$

and the intensity is $I = A^2 + B^2$. We need to consider only positive values of x since for a switch of x with $-x$, the value of A is unchanged and the sign of B changes, and therefore I remains the same.

To discuss the expression for B, we consider the integral

$$\begin{aligned} B_\infty &= \frac{1}{\pi}\int_{-\infty}^{+\infty} dw\,\frac{\sin w}{w}\,\sin\left(\frac{\sin u}{\alpha'}w\right) \\ &= \frac{1}{\pi}\int_{-\infty}^{0} dw\,\frac{\sin w}{w}\,\sin\left(\frac{\sin u}{\alpha'}w\right) + \frac{1}{\pi}\int_{0}^{+\infty} dw\,\frac{\sin w}{w}\,\sin\left(\frac{\sin u}{\alpha'}w\right) . \end{aligned}$$

It can readily be seen that the curve represented by the integrand of the first integral is the symmetrical mirror image of the curve represented by the integrand of the second integral with respect to the w-axis. Therefore, $B_\infty = 0$.[lvii]

To discuss A, we consider the integral

$$A_\infty = \frac{1}{\pi}\int_{-\infty}^{+\infty} dw\,\frac{\sin w}{w}\cdot\cos\left(\frac{\sin u}{\alpha'}w\right) . \tag{65}$$

According to earlier developments,[lviii] we have

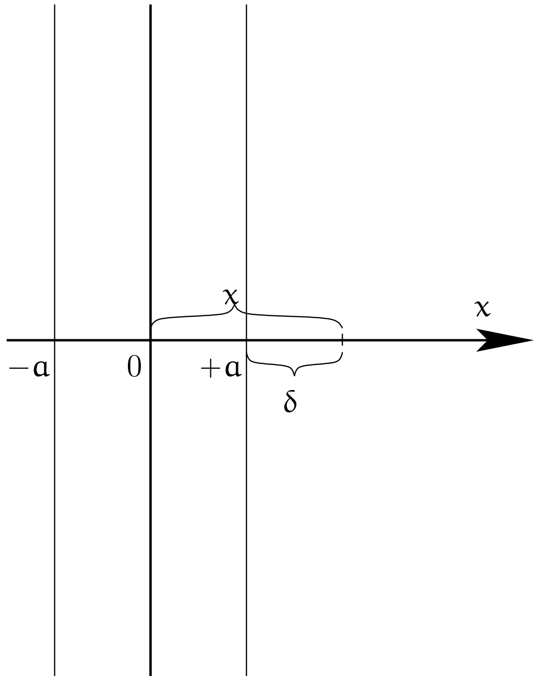
$$\left.\begin{aligned} &A_\infty = 0 \text{ for } \sin u < -\alpha' \text{ and } \sin u > +\alpha' , \\ &A_\infty = 1 \text{ for } \sin u > -\alpha' \text{ and simultaneously } \sin u < +\alpha' \\ &A_\infty = \frac{1}{2} \text{ for } \sin u = \pm\alpha' . \end{aligned}\right\} \tag{66}$$

To compare our integrals A and B with A_∞ and B_∞, we set

$$x = a + \delta ,$$

where δ is the distance of the observation point from the edge of the slit and is to be taken as positive if the observation point varies from

Figure 49



the edge with growing x (Fig. 49). Therefore, δ varies, for positive x, between $-a$ and $+\infty$. Then we have

$$A = \frac{k\lambda}{\pi} \int_{2\pi\alpha' \frac{\delta}{\lambda}}^{2\pi\alpha' \frac{2a+\delta}{\lambda}} dw \, \frac{\sin w}{w} \cdot \cos\left(\frac{\sin u}{\alpha'} w\right)$$

$$B = \frac{k\lambda}{\pi} \int_{2\pi\alpha' \frac{\delta}{\lambda}}^{2\pi\alpha' \frac{2a+\delta}{\lambda}} dw \, \frac{\sin w}{w} \cdot \sin\left(\frac{\sin u}{\alpha'} w\right) \, .$$

For a slit of finite width (a somewhat >1 mm) and a not too small opening angle of the diffracting aperture (α' somewhat $>1°$), $2\pi a\alpha'/\lambda$ is so large compared to π that the upper limit in A and B can be set to ∞. With respect to the lower limit, we again differentiate four cases.

1. The observation point is outside the slit and so far from the edge that one can replace $2\pi\alpha'\delta/\lambda$ by $+\infty$: $A = 0$ and $B = 0$; i.e., the light disturbance is zero regardless of the angle of incidence u. For points far away from the edge, therefore, there is no difference between the phenomena of normal incidence of light and those of oblique incidence of light.

2. The observation point lies within the slit and so far from the edge that $2\pi\alpha'\delta/\lambda$ can be replaced by $-\infty$; then we get

$$\begin{aligned} B &= B_\infty = 0 \\ A &= k\lambda A_\infty \,. \end{aligned}$$

   The value of A_∞ still depends on the angle of incidence; in fact, $A = k\lambda$ if $\sin u$ lies between $-\alpha'$ and $+\alpha'$, i.e., if the incident light rays extend through the slit into the diffracting aperture. The total intensity here is then equal to $k^2\lambda^2$. On the other hand, we have $A = 0$ if $\sin u < -\alpha'$ or $\sin u > +\alpha'$, i.e., if the extended light rays no longer hit the diffracing aperture. In this case, the total intensity is therefore equal to zero for all points within the slit but sufficiently far away from the edge.

   If the marginal ray of the incident light beam just hits the edges of the diffracting aperture, then $\sin u = \pm\alpha'$ and $A = \frac{1}{2}k\lambda$; i.e., the total intensity is equal to $k^2\lambda^2/4$.

3. The observation point lies on the edge of the slit. In this case, we have $2\pi\alpha'\delta/\lambda = 0$, and for each incidence angle the values of A and B are half of what they take on in case 2, i.e., when the observation point is located within the slit.[lix]

4. If the observation point lies in the immediate vicinity of the slit edge, we must decompose the integrals A and B in the following way:

$$A = \frac{k\lambda}{\pi}\int_0^\infty - \frac{k\lambda}{\pi}\int_0^{2\pi\alpha'\delta/\lambda} = \frac{k\lambda}{2}A_\infty - \frac{k\lambda}{\pi}\int_0^{2\pi\alpha'\delta/\lambda}$$

$$B = \frac{k\lambda}{\pi}\int_0^\infty - \frac{k\lambda}{\pi}\int_0^{2\pi\alpha'\delta/\lambda} = -\frac{k\lambda}{\pi}\int_0^{2\pi\alpha'\delta/\lambda} .$$

Of interest is the case where $A_\infty = 0$, i.e., when the extended light rays do *not* hit the diffracting aperture, or when

$$\begin{aligned} &\sin u < -\alpha' \\ \text{or} \qquad &\sin u > +\alpha' . \end{aligned}$$

While, as we have seen, in this case the inside of the slit and the slit edges become completely dark, the intensity for points infinitely close to the slit edges retains finite values.

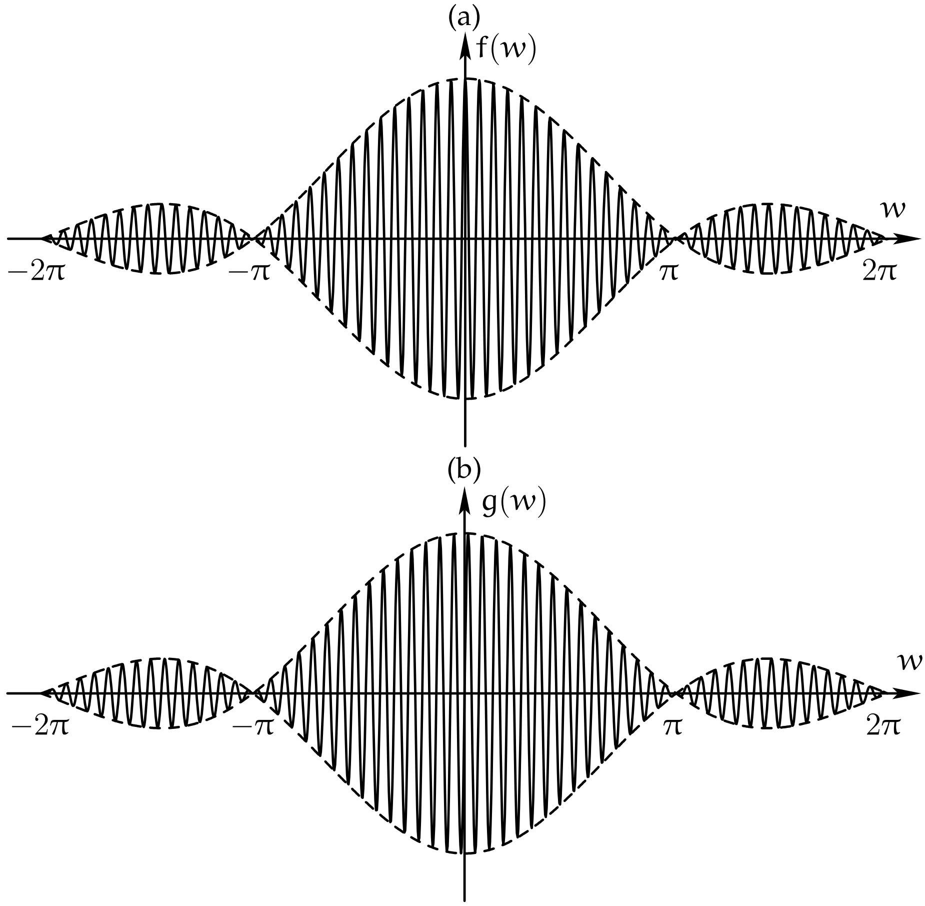
To calculate the intensity distribution close to the edges for various u, we consider the following:

If the value of $\rho = \frac{\sin u}{\alpha'}$ is large, e.g., the magnitude of α' has the value $\sin 1^\circ \simeq \frac{1}{60}$, while u, e.g., $= 30^\circ$, so that $\sin u = \frac{1}{2}$, then the graphs of the functions

$$\begin{aligned} f(w) &= \frac{\sin w}{w}\cos(\rho w) \\ \text{and} \qquad g(w) &= \frac{\sin w}{w}\sin(\rho w) \end{aligned}$$

are the ones plotted approximately in Figs. 50a and b.

Figure 50



We observe that the curves $f(w)$ and $g(w)$ intersect the axis $\rho - 1$ times between $w = 0$ and $w = \pi$ at distances $\frac{\pi}{\rho}$. The first intersection after the point $w = 0$ happens for the $f(w)$ curve at $w = \frac{1}{2}\frac{\pi}{\rho}$, and for the $g(w)$ curve at $w = \frac{\pi}{\rho}$. Now the intensity is

$$I = A^2 + B^2 ,$$

where
$$A = \text{const} \int_0^{\frac{2\pi\alpha'\delta}{\lambda}} dw \, \frac{\sin w}{w} \cos(\rho w)$$
$$B = \text{const} \int_0^{\frac{2\pi\alpha'\delta}{\lambda}} dw \, \frac{\sin w}{w} \sin(\rho w) .$$

If one then moves the boundary $\frac{2\pi\alpha'\delta}{\lambda}$ along the axis of the curves $f(w)$ and $g(w)$ and forms the corresponding areal content represented by A or B, one can easily recognize the following:

With growing $|\delta|$, I executes a series of fluctuations with decreasing amplitude. The minima of the fluctuations lie at locations

$$\frac{2\pi\alpha'\delta}{\lambda} = \pm\frac{2\mathfrak{a}\pi}{\rho} \; (\mathfrak{a} = 0, 1, 2, 3 \ldots) ,$$

i.e.,
$$\delta = \pm\frac{\mathfrak{a}\lambda}{\sin u} .$$

They maintain a distance $\frac{\lambda}{\sin u}$ from each other. The intensity of the maxima is extremely low.

If, on the other hand, $\sin u$ is only slightly different from α', that is to say $\sin u = \alpha' + \varepsilon$, where ε is small, then, according to simple calculation,[lx]

$$A = \frac{1}{2} \int\limits_0^{\frac{4\pi\alpha'\delta}{\lambda}} \frac{\sin w}{w}\, dw + (\varepsilon)$$

$$B = \int\limits_0^{\frac{2\pi\alpha'\delta}{\lambda}} \frac{\sin^2 w}{w}\, dw + (\varepsilon)\,,$$

where (ε) denotes quantities of the order of ε. Since the curve $\frac{\sin^2 w}{w}$ always runs above the abscissa, B grows everywhere with increasing δ, whereas A simultaneously experiences the known fluctuations. The minima of the intensity thus occur at intervals

$$\delta = \frac{\lambda}{2\alpha'}\,.$$

The maxima of intensity here have finite values (Fig. 51).[lxi]

So far, we have always assumed that the slit is so wide that $\frac{2\pi a\alpha'}{\lambda}$ is large compared to π.

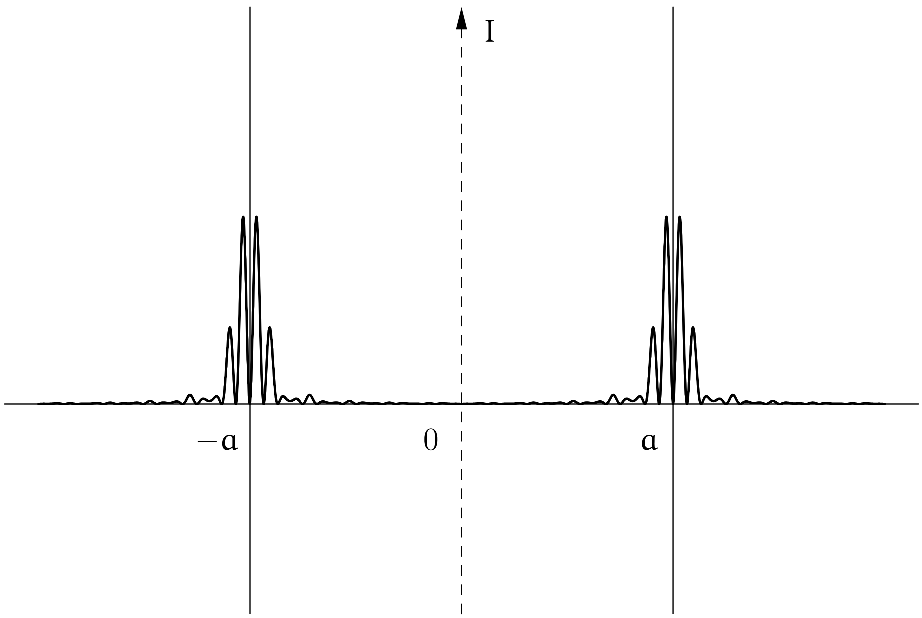
We now proceed to the consideration of a *finite but very narrow* slit by assuming that $\frac{2\pi a\alpha'}{\lambda}$ is small compared to π, thereby gaining a supplement and extension of the already discussed theory of the infinitely narrow slit. In practice, in order to make $\frac{2\pi a\alpha'}{\lambda}$ small compared to π, one must duly reduce α', since, e.g., even for

$$\alpha' = 1^{\circ}$$
$$a = \frac{1}{100}\,\mathrm{mm}$$
$$\lambda = 6 \cdot 10^{-4}\,\mathrm{mm}\,,$$

$\frac{2\pi a\alpha'}{\lambda}$ is about $\frac{\pi}{2}$ and still not small compared to π. If we set

$$\begin{cases} \dfrac{2\pi a\alpha'}{\lambda} = \varepsilon \text{ (small)} \\[2ex] \dfrac{2\pi\alpha' x}{\lambda} = \xi\,, \end{cases}$$

Figure 51



then the expression for the intensity is

$$I = A^2 + B^2$$

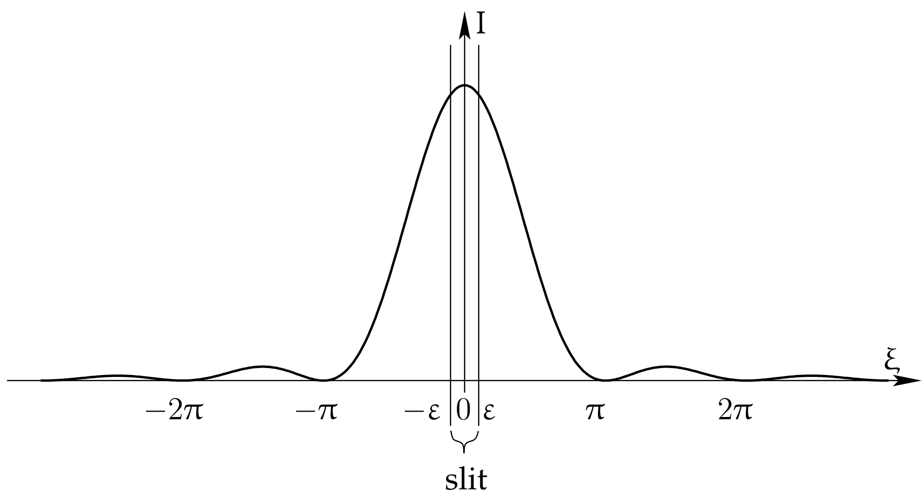
$$\begin{cases} A = \dfrac{k\lambda}{\pi} \displaystyle\int_{\xi-\varepsilon}^{\xi+\varepsilon} dw \, \frac{\sin w}{w} \cos\left(\frac{\sin u}{\alpha'} w\right) \\ B = \dfrac{k\lambda}{\pi} \displaystyle\int_{\xi-\varepsilon}^{\xi+\varepsilon} dw \, \frac{\sin w}{w} \sin\left(\frac{\sin u}{\alpha'} w\right) \end{cases} .$$

If $\frac{\sin u}{\alpha'}$ is not too large, so that $\varepsilon \frac{\sin u}{\alpha'}$ is small compared to 1, i.e., if we have almost normal incidence, we can expand A and B according to the Taylor series and obtain in a first approximation[lxii]

$$I = 4\varepsilon^2 \frac{k^2\lambda^2}{\pi^2} \left(\frac{\sin \xi}{\xi} \right)^2 .$$

The same value applies to normal incidence, $u = 0$. Thus, in the slit itself ($\xi \simeq 0$) there is an almost constant, strongest brightness; maxima and minima line up symmetrically on both sides of the slit (see Fig. 52).

Figure 52



If the incidence is tilted, i.e., if $\sin u$ has a finite value, then, since α' is very small, the magnitude

$$\rho = \frac{\sin u}{\alpha'}$$

is very large. The graphs of the integrands A and B, i.e., the functions

$$f(w) = \frac{\sin w}{w}\cos(\rho w)$$
$$g(w) = \frac{\sin w}{w}\sin(\rho w)\,,$$

are in this case already represented in Figs. 50a and b.

The graphs of A and B as functions of ξ therefore depend on the ratio of the small interval of integration

$$2\varepsilon = \frac{4\pi a\alpha'}{\lambda}$$

to the likewise small quantity π/ρ, which represents the distance between two successive zero points of the curves $f(w)$ and $g(w)$. We want to distinguish two main cases.

1. Let

$$2\varepsilon = 2\mathfrak{a}\frac{\pi}{\rho}\;(\mathfrak{a} = 1, 2, 3\ldots)\,.$$

Then we have

$$2a\sin u = \mathfrak{a}\lambda\;(\mathfrak{a} = 1, 2, 3\ldots)\,;$$

i.e., the path difference of the rays striking the edges of the object slit is an integer multiple of the wavelength. It is then for all ξ, as can be easily seen, A and B almost $= 0$, since in the formation of the integrals the adjacent pieces always cancel each other out. *The entire field of vision is therefore dark.* This is natural: the incident light experiences diffraction at the object slit. The principal maximum lies in the extension of the incident rays, that is, below the "diffraction angle" u. The minima lie in the directions

$$\sin u = \frac{\mathfrak{a}\lambda}{2a}\;(\mathfrak{a} = 1, 2, 3\ldots)\,,$$

and the secondary maxima lie in the directions

$$\sin u = \frac{(2a+1)\lambda}{2 \cdot 2a} \; (a = 0, 1, 2, 3 \ldots) \,.$$

In the considered case, $\sin u = \frac{a\lambda}{2a}$; therefore, a minimum generated by the object slit falls on the diffracting slit (α, β), and the field of view is therefore dark, as deduced above.

2. Let

$$2\varepsilon = (2a+1)\frac{\pi}{\rho} \; (a = 0, 1, 2, 3 \ldots) \,;$$

then we have

$$2a \sin u = \frac{(2a+1)}{2} \cdot \lambda \; (a = 0, 1, 2, 3 \ldots) \,.$$

In this case, one of the secondary maxima of the diffraction image generated by the object slit falls on the diffracting aperture.

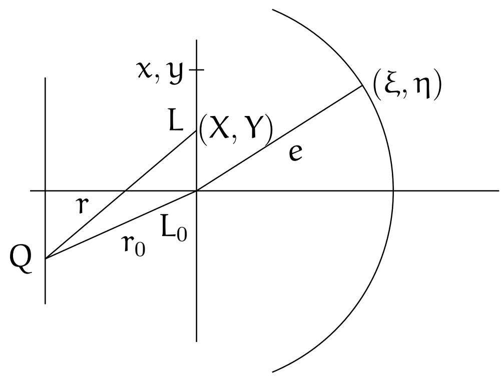
One sees immediately that for $\xi = 0$, that is, in the middle of the slit, A has a value different from zero, which becomes smaller the larger the 2ε, i.e., the more oblique the incidence of light and therefore the higher the order of the maximum that falls on the diffracting aperture. B, on the other hand, is always 0 for $\xi = 0$.

If ξ now grows, A and B periodically assume maxima and minima in rapid succession in such a way that whenever A becomes near 0, B reaches its maximum value and vice versa. At the same time, however, these maximum values decrease from $\xi = 0$ to $\xi = \pi$, and then increase again, thus causing periodic fluctuations in the "wide" intervals of π. Therefore, similar to normal incidence of light, the well-known diffraction pattern will appear, with the principal maximum at the place of the object slit and its secondary maxima and minima symmetrically on both sides, as shown in Fig. 52.

§25. Switching of the order of integration in the calculation of the resulting light disturbance

In what follows, we deal with the general problem: A point source Q (Fig. 53) illuminates the object whose center L_0 lies on the axis of the imaging system. An arbitrary point L of the object has the

Figure 53



coordinates X, Y. The image of the small object is sought using an arbitrary aperture of the imaging system. As before, we introduce as an "intermediate surface" a spherical surface whose points ξ, η have the nearly constant distance e from the individual object points X, Y. Then the light disturbance at a point X, Y of the object on the side facing the intermediate surface can be represented by

$$K\varphi(X, Y) \sin 2\pi \left[\frac{t}{T} - \Psi(X, Y) \right] , \tag{67}$$

where $\varphi(X, Y)$ is the transmission coefficient of the object element $dX\, dY$, and $K\varphi(X, Y)$ is the amplitude of the disturbance at the location of the element $dX\, dY$. $\Psi(X, Y)$ can be divided into two parts:

$$\Psi(X, Y) = \frac{r - r_0}{\lambda} + \psi(X, Y) .$$

In this case, the factor $\frac{r-r_0}{\lambda}$ takes into account the oblique incidence of the light and $\psi(X, Y)$ the delay of the waves as a result of passing through the object element.

According to earlier results,[lxiii] the sought resulting disturbance at the observation point x, y is then

$$\left.\begin{aligned} S = &\frac{K}{\lambda^2} \iint\limits_{\text{object}} dX\, dY\, \varphi(X, Y) \iint d\xi'\, d\eta' \sin 2\pi \left[\frac{t}{T} - \frac{\xi'(x - X)}{\lambda}\right. \\ &\qquad\qquad\qquad\qquad \left. - \frac{\eta'(y - Y)}{\lambda} - \Psi(X, Y)\right], \\ &\text{where we set } \xi' = \frac{\xi}{e}, \eta' = \frac{\eta}{e}. \end{aligned}\right\} \tag{68}$$

The integration with respect to X, Y extends over the illuminated object, the integration with respect to ξ', η' over the projection of the "effective patch" of the intermediate surface.

In carrying out the integration, one can proceed as before. One integrates first over the intermediate surface (ξ', η') and then over the object (X, Y). The first integration provides, in the object plane,[lxiv] the effect of diffraction of the extent-limiting aperture due to the presence of one object element; the second integration takes into account the extent of the object.

The formation of the image becomes physically clearer if one reverses the order of the integrations and carries out the integration with respect to X, Y first. This immediately provides the effect of diffraction of the illuminated object at the location of the intermediate surface. If the object is, e.g., a grating, then the well-known diffraction spectra occur on the intermediate surface, the positions of which depend on the grating constant and the angle of incidence of the light. After performing the first integration, one can therefore abstract both the light source and the object since both have been replaced by the diffraction spectra appearing on the intermediate surface.

The second integration over ξ', η' has therefore only the role of calculating the interference effect of these diffraction spectra at a point x, y in the object plane.

The resulting phenomenon ("image") is thus the interference effect of a diffraction phenomenon: the primary one being the diffraction phenomenon on the intermediate surface created by the light source and the object, and the secondary one being the effect of interference in the object plane. Only then can one recognize clearly the difference between the image of a self-luminous and an illuminated object.

In the presence of an object of a complicated structure, the evaluation of S is hardly feasible. On the other hand, general rules can be derived that specify under what conditions an "image" similar to the existing object appears, or to which fictitious object instead of the existing one the appearing phenomenon is similar.

To derive these rules, we decompose the expression S into two parts, S_1 and S_2. The first part, S_1, emerges from S if the integration is extended over the entire intermediate surface (hemisphere), i.e., if ξ' and η' take on all values from -1 to $+1$. S_2, however, extends over the entire intermediate surface with the exclusion of the "effective part."

For simplification, we set

$$\left.\begin{array}{c} \dfrac{X}{\lambda} = X'; \;\; \dfrac{Y}{\lambda} = Y'; \;\; \dfrac{x}{\lambda} = x'; \;\; \dfrac{y}{\lambda} = y' \\ \varphi(\lambda X', \lambda Y') = \varphi_1(X', Y'); \qquad \Psi(\lambda X', \lambda Y') = \Psi_1(X', Y') \end{array}\right\} . \tag{69}$$

We then get

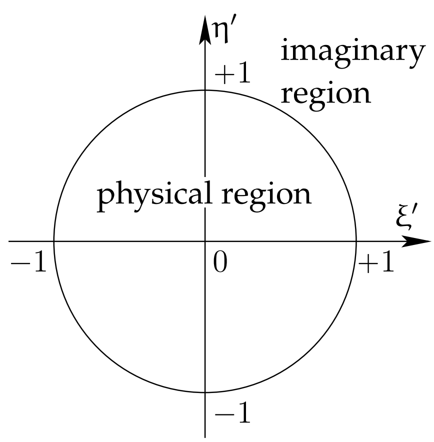
$$\left.\begin{aligned} S_1 &= K \iint_{-1}^{+1} d\xi'\, d\eta' \iint_{\text{object}} dX'\, dY'\, \varphi_1(X', Y') \sin 2\pi \left[\frac{t}{T} - \Psi_1(X', Y') \right. \\ &\qquad\qquad\qquad\qquad \left. - \xi'(x' - X') - \eta'(y' - Y') \right] \\ S_2 &= S_1 - S\,. \end{aligned}\right\} \tag{70}$$

The variables $\xi' = \frac{\xi}{e}$ and $\eta' = \frac{\eta}{e}$ are sines of the angles, and therefore the following relations are valid:

$$-1 \leqslant \begin{cases} \xi' \\ \eta' \end{cases} \leqslant +1\,. \tag{71}$$

If we represent ξ' and η' as orthogonal coordinates in the $\xi'\eta'$-plane (Fig. 54), then ξ', η' have physical meaning only in the unit

Figure 54



circle around the origin. Outside this circle, the angles to which ξ' and η' belong as sines become imaginary.

Only in the interior of this unit circle does the function contained in S_1,

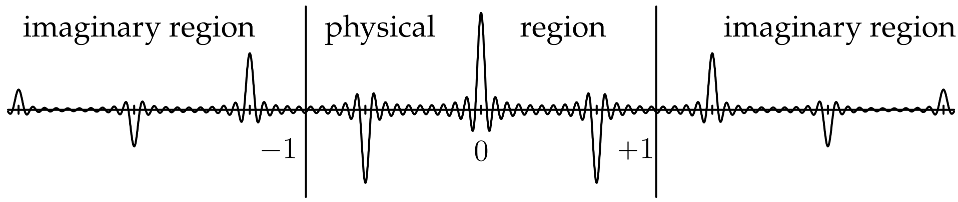
$$f(\xi', \eta') = K \iint\limits_{\text{object}} dX'\, dY'\, \varphi_1(X', Y') \sin 2\pi \left[\frac{t}{T} - \Psi_1(X', Y') + \xi' X' + \eta' Y' \right], \tag{72}$$

which represents the light disturbance in points ξ, η of the intermediate surface, have *physical significance*. Therefore, we want to call the unit circle in the $\xi'\eta'$-plane the *physical region* and the exterior of the unit circle the *imaginary region*.

Only in the physical region does $f(\xi', \eta')$ have a physical, real meaning. On the other hand, in purely mathematical terms, of course, one can continue the function $f(\xi', \eta')$ into the imaginary region. It is as if one were unaware of the meaning of the variables ξ', η' and treated them as infinitely variable.

For example, if the object is a grating, then part of the function $f(\xi', \eta')$ would be the known grating-generated diffraction image that extends across the hemisphere (intermediate surface) and *breaks off* at its boundaries $\xi' = \pm 1$ and $\eta' = \pm 1$. Mathematically, on the other hand, we can continue the diffraction image with its sharp, gradually extinguishing maxima up to $\xi' = \pm\infty$ and $\eta' = \pm\infty$. The number of maxima that are in the physical region depends on the grating constant and is greater, the larger the grating constant. (See Fig. 55, in which the *amplitudes* of the diffraction maxima are plotted.)[lxv]

Figure 55



If we form the integral

$$S_1^* = K \iint\limits_{-\infty}^{+\infty} d\xi'\, d\eta' \iint\limits_{\text{object}} dX'\, dY'\, \varphi_1(X', Y') \sin 2\pi \left\{ \begin{array}{l} \left[\dfrac{t}{T} - \Psi_1(X', Y') \right. \\ \qquad \left. -\xi'(x' - X') - \eta'(y' - Y') \right] \end{array} \right\}, \tag{73}$$

which extends over all real *and* imaginary maxima, we shall be able to identify this integral more closely with S_1, the smaller the contribution of $f(\xi', \eta')$ in the imaginary region, and in the example of a grating, the smaller the number of maxima lying in the imaginary region, i.e., the larger the grating constant. Strictly speaking, S_1^* is never equal to S_1. However, if the diffration effect of the object represented by $f(\xi', \eta')$ in the *imaginary region is vanishingly small*, so that almost the entire image of the function $f(\xi', \eta')$ has expanded in the physical region, the equation

$$S_1 = S_1^*$$

represents in praxis a well usable approximation.

We now prove that the expression S_1^* transitions into the expression

$$K\varphi(x,y)\sin 2\pi\left[\frac{t}{T}-\Psi(x,y)\right]\,,$$

if the observation point x, y coincides with the object point X, Y, i.e., that S_1^* represents the light disturbance present at the object points x, y on the side of the object that faces the intermediate surface.[lxvi]

For this purpose, we decompose the sine in the integral and write

$$\begin{aligned}
S_1^* =& K\iint\limits_{-\infty}^{+\infty} d\xi'\,d\eta' \iint\limits_{\text{object}} dX'\,dY'\,\varphi_1(X',Y')\sin 2\pi\left[\frac{t}{T}-\Psi_1\right] \\
&\qquad\cdot\cos 2\pi[\xi'(x'-X')+\eta'(y'-Y')] \\
&-K\iint\limits_{-\infty}^{+\infty} d\xi'\,d\eta' \iint\limits_{\text{object}} dX'\,dY'\,\varphi_1(X',Y')\cos 2\pi\left[\frac{t}{T}-\Psi_1\right] \\
&\qquad\cdot\sin 2\pi[\xi'(x'-X')+\eta'(y'-Y')]\,.
\end{aligned}$$

If we set $\begin{cases} K\varphi_1(X',Y')\sin 2\pi\left[\frac{t}{T}-\Psi_1\right] = F(X',Y') \\ K\varphi_1(X',Y')\cos 2\pi\left[\frac{t}{T}-\Psi_1\right] = G(X',Y')\,, \end{cases}$

we get

$$\left.\begin{aligned}
S_1^* =& \iint\limits_{-\infty}^{+\infty} d\xi'\,d\eta' \iint\limits_{\text{object}} F(X',Y')\,dX'\,dY'\cos 2\pi[\xi'(x'-X')+\eta'(y'-Y')] \\
&-\iint\limits_{-\infty}^{+\infty} d\xi'\,d\eta' \iint\limits_{\text{object}} G(X',Y')\,dX'\,dY'\sin 2\pi[\xi'(x'-X')+\eta'(y'-Y')]
\end{aligned}\right\}. \tag{74}$$

We can easily show that

$$\iint\limits_{-\infty}^{+\infty} d\xi'\, d\eta' \iint\limits_{\text{object}} F(X',Y')\, dX'\, dY' \cos 2\pi[\xi'(x'-X') + \eta'(y'-Y')]$$

$$= F(x',y')\text{ , if point } x', y' \text{ lies inside the object,}$$
$$= 0\text{ , if point } x', y' \text{ lies outside the object,}$$

and $$\iint\limits_{-\infty}^{+\infty} d\xi'\, d\eta' \iint\limits_{\text{object}} G(X',Y')\, dX'\, dY' \sin 2\pi[\xi'(x'-X') + \eta'(y'-Y')]$$

$$= 0 \text{ for all locations of point } x', y' \text{ .}$$

This is because the two Fourier theorems apply:[lxvii]

$$\int\limits_{-\infty}^{+\infty} d\xi \int\limits_{A_1}^{A_2} dX\, F(X) \cos 2\pi\xi(x-X) = \begin{cases} F(x)\text{ , if } x \text{ is inside } A_1 \dots A_2 \\ 0\text{ , if } x \text{ is outside } A_1 \dots A_2 \end{cases}$$

and

$$\int\limits_{-\infty}^{+\infty} d\xi \int\limits_{A_1}^{A_2} dX\, F(X) \sin 2\pi\xi(x-A) = 0\text{ , for all values of } x.$$

From this, it follows that

$$\int\limits_{-\infty}^{+\infty} d\xi \int\limits_{A_1}^{A_2} dX\, F(X,y) \cos 2\pi\xi(x-X) = F(x,y)\text{ , if } x \text{ is between } A_1 \text{ and } A_2,$$

and

$$\int\limits_{-\infty}^{+\infty} d\eta \int\limits_{B_1}^{B_2} dY\, F(X,Y) \cos 2\pi\eta(y-Y) = F(X,y)\text{ , if } y \text{ is between } B_1 \text{ and } B_2.$$

Therefore, by substitution,

$$\iint\limits_{-\infty}^{+\infty} d\xi\, d\eta \int\limits_{A_1}^{A_2}\int\limits_{B_1}^{B_2} dX\, dY\, F(X,Y) \cos 2\pi\xi(x-X) \cos 2\pi\eta(y-Y)$$

$= F(x,y)$, if x, y lie between $A_1 \ldots A_2$ and $B_1 \ldots B_2$, respectively.

By analogy, we have

$$\iint\limits_{-\infty}^{+\infty} d\xi\, d\eta \int\limits_{A_1}^{A_2}\int\limits_{B_1}^{B_2} dX\, dY\, F(X,Y) \sin 2\pi\xi(x-X) \sin 2\pi\eta(y-Y) = 0$$

for all values of x, y.

By subtracting the last formula from the one before, we get, finally,

$$\left.\begin{aligned} &\iint\limits_{-\infty}^{+\infty} d\xi\, d\eta \int\limits_{A_1}^{A_2}\int\limits_{B_1}^{B_2} dX\, dY\, F(X,Y) \cos 2\pi[\xi(x-X)+\eta(y-Y)] \\ &\qquad = F(x,y) \text{ , when } x \text{ and } y \text{ lie between } A_1 \text{ and } A_2 \\ &\qquad\qquad \text{and between } B_1 \text{ and } B_2\text{, respectively,} \\ &\qquad = 0 \text{ for all other locations of } x, y. \end{aligned}\right\} \tag{75}$$

It can easily be shown in an analogous fashion that we have, additionally,

$$\left.\begin{aligned} &\iint\limits_{-\infty}^{+\infty} d\xi\, d\eta \int\limits_{A_1}^{A_2}\int\limits_{B_1}^{B_2} dX\, dY\, G(X,Y) \sin 2\pi[\xi(x-X)+\eta(y-Y)] \\ &\qquad = 0 \text{ for all locations of } x, y. \end{aligned}\right\} \tag{76}$$

This proves what was already anticipated above that

1.

$$\left.\begin{aligned} S_1^* = F(x', y') &= K\varphi_1(x', y') \sin 2\pi \left[\frac{t}{T} - \Psi_1(x', y')\right] \\ &= K\varphi(x, y) \sin 2\pi \left[\frac{t}{T} - \Psi(x, y)\right] \end{aligned}\right\} \tag{77}$$

if the point x, y lies within the object.

2. $S_1^* = 0$ for all points x, y outside the object.

Therefore, S_1^* represents the light distribution present on that side of the object (X, Y) to be imaged, facing the intermediate surface.

§26. Pointwise and similar imaging of the object

Referring to the previous paragraph, a pointwise and similar imaging takes place when S can be completely replaced by S_1^*. This is always the case if all the diffraction maxima down to negligible intensity contribute to image formation, i.e., if the aperture of the imaging system (the "effective part" of the intermediate surface) collects all the rays diffracted from the object down to negligible intensity. Thus, there is always an absolute similarity between image and object if the *entire* image of the function $f(\xi', \eta')$ can be expanded within the aperture, but there is dissimilarity if the aperture *does not* collect all diffraction maxima of $f(\xi', \eta')$, i.e., if only parts of the image of the function lie within the aperture.

We shall discuss on which physical quantities the capacity of the system and thus its performance depends. For this, we consider the imaging of a grating. For a given wavelength λ_0 of the incident light, the position of the hth peak is given by the relation

$$\sin u_h = \lambda_0 \frac{h}{n\gamma} ,$$

where u_h denotes the diffraction angle of the hth maximum, n the index of refraction of the front medium that contains the intermediate surface (immersion fluid), and γ the grating constant.

The number of maxima within the aperture angle U of our system is therefore

$$h = n \sin U \frac{\gamma}{\lambda_0} \,. \tag{78}$$

As we know, the larger the h, the greater the similarity of the image, and we reach ideal similarity for $h = \infty$. For a given grating (γ) and wavelength (λ_0) of the incident light, the number (h) of the image-contributing diffraction maxima that are accepted within the aperture angle is proportional to the product: *index of refraction times sine of the aperture angle.* This product $A = n \cdot \sin U$ has been designated by Abbe as the *numerical aperture* of the system.

Thus, the important theorem follows: *If two systems have the same numerical aperture,*

$$n_1 \sin U_1 = n_2 \sin U_2 \,,$$

they image the same object grating with the same degree of similarity. Only in this way does one actually recognize the meaning of the term numerical aperture introduced by Abbe, that only the product $A = n \cdot \sin U$ determines the similarity of the image, not the aperture angle U of the system. As is well known, for the imaging of self-luminous objects, the numerical aperture is the quantity that alone determines the luminous intensity of the system.

If the aperture angle U of the system for a given λ_0 and γ, as with a dry system ($n = 1$), does not include all the diffraction maxima to vanishing intensity, then the image is a dissimilar one; it can then be transformed into a more similar one if one uses the same system as an immersion system ($n > 1$). As the equation

$$h = n \cdot \sin U \cdot \gamma/\lambda_0$$

shows, the similarity of the image can be increased even more by reducing λ_0.

For a given numerical aperture $A = n \sin U$ of a system with a given wavelength λ_0, the similarity of the grating image is solely due to the grating constant γ. The larger γ becomes, the more diffraction maxima can contribute to image formation, and the greater the similarity. The maximum numerical aperture of a system is reached when $U = 90^{\circ}$ and is then

$$A = n\,.$$

Therefore, in this case of *maximum possible performance*,

$$h = n\frac{\gamma}{\lambda_0}\,. \tag{79}$$

If we denote with h_l the last diffraction spectrum of intensity or brightness to be considered in the overall image of the function $f(\xi', \eta')$, the system with $A = n$ will image all gratings with absolute similarity, if

$$\gamma \geqslant \frac{h_l \cdot \lambda_0}{n}\,.$$

§27. Dissimilar imaging of the object

We shall base this investigation on a system with maximum aperture $A = n$, which still images a grating with constant γ with absolute similarity, meaning the satisfaction of the inequality

$$\gamma \geqslant h_l \lambda_0 / n\,,$$

where h_l is the last diffraction spectrum of intensity still to be considered in the overall image of the function $f(\xi', \eta')$. A grating with a smaller grating constant ($\gamma' < \gamma$) is therefore no longer imaged by the system similarly. If λ_0 has the smallest possible value (photographic waves) and n has the highest possible value (homogeneous immersion), then the grating $\gamma = h_l \lambda_0 / n$ is imaged in an absolutely similar way (a fortiori all gratings with *larger* grating constants), whereas it is physically impossible to image gratings with smaller grating constants ($\gamma' < \gamma$) similarly.

As an example, let us suppose that $\lambda_0 = 350\,\text{nm}$, $n = 1.65$, and $h_l = 10$, assuming that maxima with an intensity less than 1 % of the mean do not contribute to the image. Then the constant of the grating that can still be imaged with absolute similarity ("limit grating") is $\gamma \simeq 2\,\mu\text{m}$.

If we let γ decrease continuously from this limit, more and more maxima of the function $f(\xi', \eta')$ move from the physical region ($\xi', \eta' = -1$ to $+1$) into the imaginary region ($\xi', \eta' < -1$ and $> +1$); i.e., the number of maxima contributing to the image becomes ever smaller and the image becomes more dissimilar. If the grating constant has become so small that only the very center diffraction maximum (principal maximum) lies in the physical region, the dissimilarity reaches its highest degree. We shall denote this maximum dissimilarity as "absolute dissimilarity." It is evidenced by the fact that the image of the structure of the object grating does not show anything, but appears as an almost uniformly luminous area. Only if, in addition to the principal maximum, one of the two adjacent maxima comes into action does the lowest degree of similarity occur; i.e., the image shows interference maxima and minima (structure), and indeed possesses the same number of strokes as the grating.

The lowest degree of similarity is achieved with *central* illumination for

$$\gamma = \frac{\lambda_0}{A}\,, \tag{80}$$

where besides the principal maximum *both* adjacent maxima are contributing. But the same lowest degree of similarity is attained when, apart from the principal maximum, only one of the two adjacent maxima contributes. This can be realized by applying *oblique* illumination, where the grating constant may decrease down to a value of

$$\gamma_m = \frac{\lambda_0}{2A}\,. \tag{81}$$

With this value, the limit of the resolving power of a microscope system is reached.

As is well known, Helmholtz[2] came almost at the same time, albeit in a different way, to the same limit of resolving power.

If one starts by using the full aperture $A = n$ for the grating

$$\gamma < \frac{\lambda_0}{2A}$$

(absolute dissimilarity), with a continuously growing grating constant, new secondary maxima appear continuously and seamlessly in addition to the principal maximum, according to their ordinal number. Here the image always shows just as many interference maxima and minima as the respective grating has "strokes," whereas the intensity decrease from maximum to minimum becomes more and more similar to the intensity distribution in the object grating given by the function $\varphi(X, Y)$. In this way, one finally reaches the "limit grating," which is just about imaged with absolute similarity.

However, with the series of dissimilarities just considered, the variety of dissimilarities is not exhausted. Rather, a large number of variations of dissimilar images of one and the same object grating can be achieved by artificially restricting the aperture or by clipping individual arbitrary and arbitrarily located diffraction maxima. In all these cases, and more generally in the imaging of any microscopic object, a theorem can be derived from our earlier observations, which determines the kind of dissimilarity in each case.

For this, we create a fictitious object (O_f), whose *natural* and *complete* diffraction pattern $[\psi(\xi', \eta')]$ coincides with the diffraction pattern $f(\xi', \eta')$ of the real object (O_r), which was rendered artificially incomplete by stopping down the diaphragm, etc. It is therefore

$$\psi(\xi', \eta')_{\text{complete}} = f(\xi', \eta')_{\text{incomplete}}$$

[2]H. Helmholtz, "The theoretical limit of the resolving power of microscopes," Pogg. Ann, Jubelband 1874,[lxviii] pp. 557–584; Wissenschaftl. Abhandl. Bd. II, pp. 185–212, 1883.

and thus, finally,

$$S[f(\xi',\eta')_{\text{incomplete}}] = S[\psi(\xi',\eta')_{\text{complete}}] = S_1^*[\psi(\xi',\eta')_{\text{complete}}] \,. \quad (82)$$

Thus, the image of the given object O_r in the case of artificial clipping $S[f(\xi',\eta')_{\text{incomplete}}]$ is equal to the absolutely similar image of the fictitious object O_f of the form $S_1^*[\psi(\xi',\eta')_{\text{complete}}]$. For this kind of dissimilarity, we obtain the following general theorem: *The image of the given object* O_r *is identical to the absolutely similar image of that fictitious object* O_f *which would just produce a complete diffraction pattern equal to part of the diffraction pattern of* O_r *accepted by the aperture of the system.*

Chapter 4

Imaging of a grating with artificial clipping of diffraction orders[1]

§28. General intensity equation

Finally, as a typical example, we want to treat the imaging of a grating. Let the grating extend along the X-axis from $X = -A$ to $X = +A$, and along the Y-axis from $Y = -B$ to $Y = +B$, so that it lies symmetrically with respect to the X- and Y-axis and let it consist of N slits of width $2a$, which are separated by "bars" of width 2Δ. Therefore, $\gamma = 2(a + \Delta)$ is the grating constant. Let N be a large number. Let α' and β' be the angular height and width of the diffracting aperture (boundary), which lies as a whole or in its parts symmetrically to the X- and Y-axis.

[1]The results given in this chapter are taken, at our urging, from the doctoral dissertation of M. Wolfke (Breslau 1910), which will soon appear in *Annalen der Physik*.

Then, with normal light incidence, the resulting disturbance is given by

$$s_2 = \frac{K}{\lambda^2} \int_{-\alpha'}^{+\alpha'} \int_{-\beta'}^{+\beta'} d\xi' \, d\eta' \int_{-B}^{+B} dY \sum_{i=1}^{i=N} \int_{-p_i}^{+q_i} dX \sin 2\pi \left[\frac{t}{T} - \frac{\xi'(x-X)}{\lambda} - \frac{\eta'(y-Y)}{\lambda} \right] \Bigg\} , \tag{83}$$

according to Eq. 68, where p_i and q_i are the X-coordinates of the ith slit, so that $q_i - p_i = 2a$. If one carries out the integration on Y and η' and defines a "grating zone" analogously to that in § 20, the resulting intensity within this zone is represented by the expression

$$I = \text{const} \left[\int_{-\alpha'}^{+\alpha'} d\xi' \sum_{i=1}^{i=N} \int_{-p_i}^{+q_i} dX \cos \frac{2\pi\xi'(x-X)}{\lambda} \right]^2 .$$

Performing the integration and summation gives, after an easy calculation,

$$\left. \begin{aligned} I &= \text{const} \left[\int_{-\frac{2\pi a \alpha'}{\lambda}}^{+\frac{2\pi a \alpha'}{\lambda}} dw \, \frac{\sin w}{\lambda} \cdot \frac{\sin \frac{N\gamma w}{2a}}{\sin \frac{\gamma w}{2a}} \cos \frac{x}{a} w \right]^2 \\ \text{where we set} \qquad w &= \frac{2\pi a \xi'}{\gamma} . \end{aligned} \right\} , \tag{84}$$

The function

$$f(w) = \frac{\sin w}{w} \cdot \frac{\sin \frac{N\gamma w}{2a}}{\sin \frac{\gamma w}{2a}} \tag{85}$$

has at positions

$$w = \frac{2\mathfrak{a}\pi a}{\gamma} \quad (\mathfrak{a} = 0, 1, 2 \ldots) \tag{86}$$

principal extrema, of which two successive ones are separated by $N - 1$ zeros lying at positions

$$w = \frac{2\mathfrak{a}\pi a}{N\gamma} \quad (\mathfrak{a} = 1, 2, 3 \ldots)\,.$$

Between two successive zeros lies a secondary extremum of the function, so that between two principal extrema there are $(N - 2)$ secondary extrema. In addition, the function $f(w)$ has zeros at $w = \pm\mathfrak{a}\pi$ $(\mathfrak{a} = 1, 2, 3 \ldots)$.

Thus, the function $[f(w)]^2$ represents the known intensity distribution in the diffraction image of the grating.

In the following, we consider several special cases that are produced by varying the integration limits; i.e., we exclude certain parts (spectra) from the diffraction pattern of the grating and allow only the remaining parts to interfere.

§29. Case I: Only the central image (the 0th order) goes through

In this case, the expression for the intensity becomes

$$I = \text{const}\left[\int\limits_{-\frac{2\pi a}{N\gamma}}^{+\frac{2\pi a}{N\gamma}} dw\,\frac{\sin w}{w}\,\frac{\sin\frac{N\gamma w}{2a}}{\sin\frac{\gamma w}{2a}}\cos\frac{xw}{a}\right]^2 . \tag{87}$$

If we set

$$J_0 = \int\limits_{0}^{\frac{2\pi a}{N\gamma}} dw\,\frac{\sin w}{w}\cdot\frac{\sin\frac{N\gamma w}{2a}}{\sin\frac{\gamma w}{2a}}\cdot\cos\frac{xw}{a}\,, \tag{88}$$

then we get

$$I = 4\,\mathrm{const} \cdot J_0^2\,. \tag{89}$$

We now discuss the graph of J_0.

Since the integration interval of the integral J_0 is small compared to π and, if $\frac{2a}{\gamma}$ is not too small, also against $\frac{2a\pi}{\gamma}$, we can write

$$J_0 = \frac{2a}{\gamma} \int\limits_0^{\frac{2\pi a}{N\gamma}} dw\, \frac{\sin \frac{N\gamma w}{2a}}{w} \cos \frac{xw}{a}\,. \tag{90}$$

It follows then[lxix]

$$\frac{\partial J_0}{\partial x} = -4aN \frac{\sin \frac{2\pi x}{N\gamma}}{N^2\gamma^2 - 4x^2}\,. \tag{91}$$

As we can see, J_0 has a maximum for $x = 0$. If one lets x vary from the value $x = 0$ to the value $x = \frac{N\gamma}{2}$, then $\frac{\partial J_0}{\partial x}$ always remains negative, i.e., J_0 steadily decreases from the maximum.

For $x = \frac{N\gamma}{2}$, we have

$$\frac{\partial J_0}{\partial x} = -\frac{2a\pi}{N\gamma^2}\,.$$

At $x = \frac{N\gamma}{2} + \Delta$, i.e., at the very edge of the grating, we have

$$\begin{aligned}
\frac{\partial J_0}{\partial x} &= -4aN \frac{\sin\left(\pi + \frac{2\pi\Delta}{N\gamma}\right)}{N^2\gamma^2 - 4\left(\frac{N\gamma}{2} + \Delta\right)^2} \\
&= -aN \frac{\sin \frac{2\pi\Delta}{N\gamma}}{\Delta^2 + N\gamma\Delta}\,,
\end{aligned}$$

or, since $\frac{\Delta}{N\gamma}$ is very small,

$$\begin{aligned}\frac{\partial J_0}{\partial x} &= -aN\frac{\frac{2\pi\Delta}{N\gamma}}{N\gamma\Delta}\left(1-\frac{\Delta}{N\gamma}\right)\\ &= -\frac{2a\pi}{N\gamma^2}\left(1-\frac{\Delta}{N\gamma}\right),\end{aligned}$$

and is therefore negative.

Only *outside* the grating, for $x = \pm a\frac{N\gamma}{2}$ ($a = 2, 3, 4, \ldots$), do we have

$$\frac{\partial J_0}{\partial x} = 0\,,$$

and indeed J_0 has minima at the positions $x = \pm aN\gamma$ ($a = 1, 2, 3, \ldots$), whereas the maxima of J_0 lie at the positions

$$x = \pm\frac{2a+1}{2}N\gamma\ (a = 1, 2, 3, 4, \ldots)\,.$$

For large x, $\frac{\partial J_0}{\partial x}$ will get closer and closer to zero, i.e., J_0 itself is increasingly approaching a constant, which is actually zero.

Some special values of J_0 are as follows:

$$\begin{aligned}J_0(x=0) &= \frac{2a}{\gamma}\int_0^{\frac{2\pi a}{N\gamma}} dw\,\frac{\sin\frac{N\gamma w}{2a}}{w}\\ &= \frac{2a}{\gamma}\int_0^{\pi}\frac{dw'}{w'}\sin w'\\ &= \frac{2a}{\gamma}\cdot 1.85 = 3.7\frac{a}{\gamma}\,.\end{aligned}$$

Further, with good approximation,

$$
\begin{aligned}
J_0\left(x = \frac{N\gamma}{2} + \Delta\right) = J_0\left(x = \frac{N\gamma}{2}\right)
&= \frac{2a}{\gamma} \int\limits_0^{\frac{2\pi a}{N\gamma}} dw \, \frac{\sin \frac{N\gamma w}{2a}}{w} \cos \frac{N\gamma w}{2a} \\
&= \frac{a}{\gamma} \int\limits_0^{\frac{2\pi a}{N\gamma}} \frac{dw}{w} \sin \frac{N\gamma w}{a} \\
&= \frac{a}{\gamma} \int\limits_0^{2\pi} \frac{dw'}{w'} \sin w' = \frac{a}{\gamma} \cdot 1.43 \, .
\end{aligned}
$$

Finally, we have[lxx]

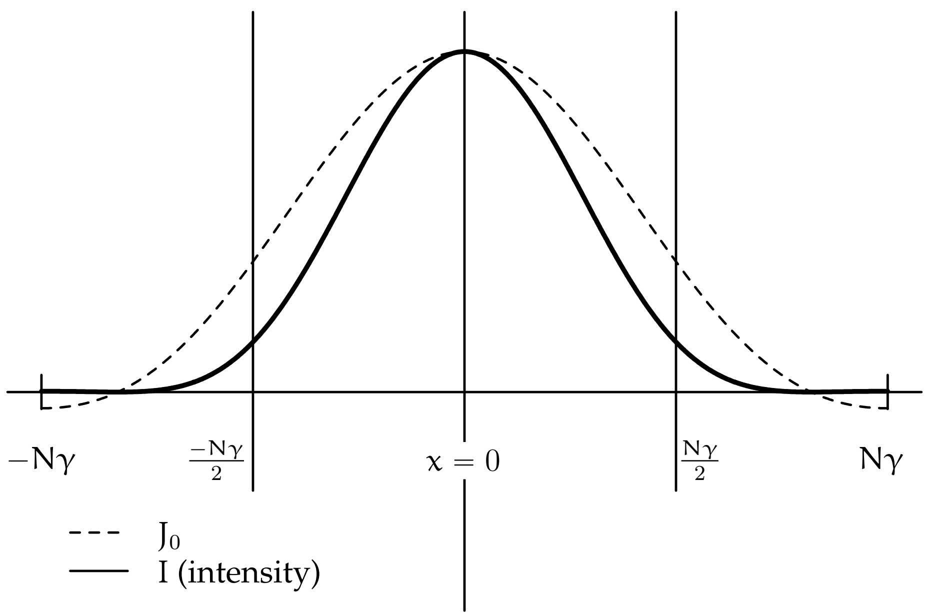
$$
\begin{aligned}
J_0(x = N\gamma) &= \frac{2a}{\gamma} \int\limits_0^{\frac{2\pi a}{N\gamma}} dw \, \frac{\sin \frac{N\gamma w}{2a}}{w} \cos \frac{N\gamma w}{a} \\
&= \frac{a}{\gamma} \left\{ \int\limits_0^{\frac{2\pi a}{N\gamma}} \frac{dw}{w} \sin \frac{3N\gamma w}{2a} - \int\limits_0^{\frac{2\pi a}{N\gamma}} \frac{dw}{w} \sin \frac{N\gamma w}{2a} \right\} \\
&= \frac{a}{\gamma} \left\{ \int\limits_0^{3\pi} \frac{dw'}{w'} \sin w' - \int\limits_0^{\pi} \frac{dw'}{w'} \sin w' \right\} \\
&= \frac{a}{\gamma} \{1.66 - 1.85\} = -0.19 \frac{a}{\gamma} \, .
\end{aligned}
$$

Hence,

$$J_0\left(x = \frac{N\gamma}{2} + \Delta\right) = \frac{1.4}{3.7}\,J_0(x = 0) = 0.4\,J_0(x = 0)$$
$$J_0\,(x = N\gamma) = -\frac{0.19}{3.7}\,J_0(x = 0) = -0.05\,J_0(x = 0)\,.$$ [lxxi]

The graph of the integral J_0 as a function of x is therefore the one shown in Fig. 56 by the dashed curve. The curve J_0^2 gives the intensity I

Figure 56



apart from a constant. Its graph is shown in the figure by the solid line.

We therefore obtain the following result:

If one blocks out all maxima in the primary diffraction pattern of the grating and allows only the undeflected central image (zeroth maximum) to *be used for image formation, then the secondary image of the grating shows a somewhat broadened structureless area whose brightness decreases from the center to the edges. On the two sides of the structureless area, secondary maxima of very low brightness* $(1/25)$ *occur.*

§30. Case II: Besides the central image, the left and right first maxima go through

In this case, the intensity is given by the expression

$$I = \text{const} \left\{ \left[\int_{-\frac{2\pi a}{N\gamma}}^{+\frac{2\pi a}{N\gamma}} dw \, \frac{\sin w}{w} \frac{\sin \frac{N\gamma w}{2a}}{\sin \frac{\gamma w}{2a}} \cos \frac{x}{a} w \right. \right. \\ + \int_{\frac{2\pi a(N-1)}{N\gamma}}^{\frac{2\pi a(N+1)}{N\gamma}} dw \, \frac{\sin w}{w} \frac{\sin \frac{N\gamma w}{2a}}{\sin \frac{\gamma w}{2a}} \cos \frac{x}{a} w \\ \left. \left. + \int_{-\frac{2\pi a(N+1)}{N\gamma}}^{-\frac{2\pi a(N-1)}{N\gamma}} dw \, \frac{\sin w}{w} \frac{\sin \frac{N\gamma w}{2a}}{\sin \frac{\gamma w}{2a}} \cos \frac{x}{a} w \right]^2 \right\} . \tag{92}$$

We can write this as

$$I = 4\text{const}[J_0 + J_1]^2 , \tag{93}$$

where J_0 is the integral discussed in detail above (case I), whereas J_1 is defined by

$$J_1 = \int_{\frac{2\pi a(N-1)}{N\gamma}}^{\frac{2\pi a(N+1)}{N\gamma}} dw \, \frac{\sin w}{w} \frac{\sin \frac{N\gamma w}{2a}}{\sin \frac{\gamma w}{2a}} \cos \frac{x}{a} w . \tag{94}$$

Since the integration interval of integral J_1 is very small compared to the period of $\sin w$, we can write

$$J_1 = \frac{\sin \frac{2\pi a}{\gamma}}{\frac{2\pi a}{\gamma}} \int\limits_{\frac{2\pi a(N-1)}{N\gamma}}^{\frac{2\pi a(N+1)}{N\gamma}} dw \, \frac{\sin \frac{N\gamma w}{2a}}{\sin \frac{\gamma w}{2a}} \cos \frac{x}{a} w \, .$$

If we introduce a new integration variable w_1 via the relationship

$$w_1 = \pi - \frac{\gamma w}{2a} \, ,$$

then we get

$$\begin{aligned}
J_1 &= \frac{(-1)^{N-1}}{\pi} \sin \frac{2\pi a}{\gamma} \int\limits_{-\frac{\pi}{N}}^{+\frac{\pi}{N}} dw_1 \, \frac{\sin N w_1}{\sin w_1} \cos \frac{2x}{\gamma} (\pi - w_1) \\
&= \frac{(-1)^{N-1}}{\pi} \sin \frac{2\pi a}{\gamma} \left\{ \cos \frac{2\pi x}{\gamma} \int\limits_{-\frac{\pi}{N}}^{+\frac{\pi}{N}} dw_1 \, \frac{\sin N w_1}{\sin w_1} \cos \frac{2 x w_1}{\gamma} \right. \\
&\left. + \sin \frac{2\pi x}{\gamma} \int\limits_{-\frac{\pi}{N}}^{+\frac{\pi}{N}} dw_1 \, \frac{\sin N w_1}{\sin w_1} \sin \frac{2 x w_1}{\gamma} \right\} .
\end{aligned}$$

Since the function under the integral is odd and the limits are symmetrical with respect to the origin, the second integral is identically equal to zero. If we set $\sin w_1 = w_1$, since the integration interval is small compared to π, and if we introduce the variable $w_2 = \frac{2 a w_1}{\gamma}$, we get

$$J_1 = \frac{2}{\pi} (-1)^{N-1} \sin \frac{2\pi a}{\gamma} \cos \frac{2\pi x}{\gamma} \int\limits_{0}^{\frac{2\pi a}{N\gamma}} dw_2 \, \frac{\sin \frac{N\gamma w_2}{2a}}{w_2} \cos \frac{x}{a} w_2$$

or

$$J_1 = \frac{\gamma}{\pi a}(-1)^{N-1} \sin\frac{2\pi a}{\gamma} \cos\frac{2\pi x}{\gamma} \cdot J_0 \,.$$

Therefore, the intensity becomes

$$I = 4 \text{ const} \left[1 + (-1)^{N-1} \cdot 2 \cdot \frac{\sin\frac{2\pi a}{\gamma}}{\frac{2\pi a}{\gamma}} \cos\frac{2\pi x}{\gamma} \right]^2 \cdot J_0^2 \,. \tag{95}$$

To discuss this expression, we consider the factor

$$A = 1 + (-1)^{N-1} \cdot 2 \cdot \frac{\sin\frac{2\pi a}{\gamma}}{\frac{2\pi a}{\gamma}} \cos\frac{2\pi x}{\gamma} \,.$$

If N is even, the maxima of A lie at locations

$$x = \pm\frac{2\mathfrak{a}+1}{2}\gamma \;(\mathfrak{a} = 0, 1, 2, \ldots) \,,$$

i.e., *in the middle of the slits* (since the grating is symmetrical with respect to the Y-axis), and the minima lie at locations $x = \pm\mathfrak{a}\gamma$. If N is odd, the maxima of A lie at locations $x = \pm\mathfrak{a}\gamma$ ($\mathfrak{a} = 0, 1, 2, \ldots$), so again in the middle of the slits, and the minima of A lie at $x \pm \frac{2\mathfrak{a}+1}{2}\gamma$. But since the intensity is given by the *square* of A, we have to consider that if the minima of A are *negative*, they give rise to *secondary maxima in intensity*. This happens when

$$1 - 2\frac{\sin\frac{2\pi a}{\gamma}}{\frac{2\pi a}{\gamma}} < 0$$

or

$$\frac{\sin\frac{2\pi a}{\gamma}}{\frac{2\pi a}{\gamma}} < \frac{1}{2} \,,$$

or if the condition

$$0 < \frac{2a}{\gamma} < 0.6$$

is met.

However, if $\frac{2a}{\gamma} > 0.6$, then the minima of A are positive and yield, after squaring, minima in intensity.

The decrease in intensity from the maximum to the minimum is in the form of a cosine, for it follows the law $I = (1 + C\cos u)^2$, where $u = \frac{2\pi x}{\gamma}$; maxima and minima have equal width (see Figs. 57a and b). We therefore obtain the following result:

If, in addition to the central order, the first two side maxima also contribute to the secondary image, the image shows a structure. The number of grating lines is reproduced correctly in the image, but the intensity drop from the maximum to the minimum is gradual, and the maxima and minima appear equally wide. In addition, under certain circumstances, secondary maxima still occur in the middle of the minima.

§31. Case III: Only the ith maxima on both sides contribute to imaging; the central image is blocked

The expression for intensity now becomes

$$\left.\begin{aligned} I &= \text{const} \cdot 4 \left[\int\limits_{\frac{2\pi a(Ni-1)}{N\gamma}}^{\frac{2\pi a(Ni+1)}{N\gamma}} dw \, \frac{\sin w}{w} \frac{\sin \frac{N\gamma w}{2a}}{\sin \frac{\gamma w}{2a}} \cos \frac{x}{a} w \right]^2 \\ &= \text{const} \cdot J_i^2 \end{aligned}\right\} . \tag{96}$$

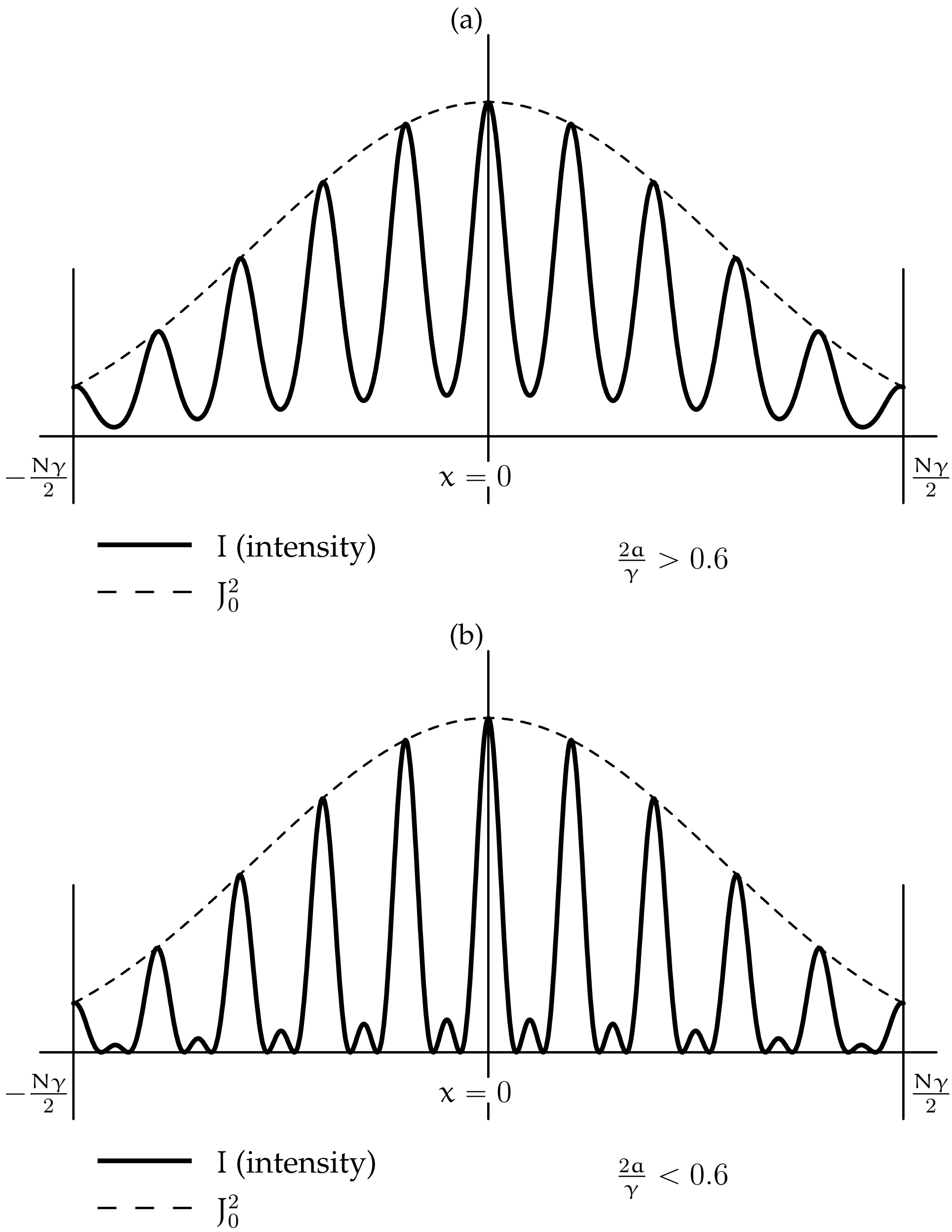
If we introduce a new variable,

$$w_1 = \pi i - \frac{\gamma w}{2a} ,$$

for the transformation of J_i, the integration limits will become symmetrical with respect to the origin; we can then again omit, as in case II above, the integral over the odd function and finally obtain, after introducing

$$w_2 = \frac{2a}{\gamma} w_1$$

Figure 57


(a)
$-\frac{N\gamma}{2}$
$x = 0$
$\frac{N\gamma}{2}$
I (intensity)
J_0^2
$\frac{2a}{\gamma} > 0.6$
(b)
$-\frac{N\gamma}{2}$
$x = 0$
$\frac{N\gamma}{2}$
I (intensity)
J_0^2
$\frac{2a}{\gamma} < 0.6$

for the intensity,

$$J_i = (-1)^{i(N-1)} \cdot 2 \cdot \frac{\sin \frac{2\pi a i}{\gamma}}{\frac{2\pi a i}{\gamma}} \cos \frac{2\pi i x}{\gamma} \cdot J_0 \,. \tag{97}$$

Therefore,

$$I = \text{const} \left(\frac{\sin \frac{2\pi a i}{\gamma}}{\frac{2\pi a i}{\gamma}} \right)^2 \cos^2 \frac{2\pi i x}{\gamma} \cdot J_0^2 \,. \tag{98}$$

The maxima of intensity lie obviously at locations

$$x = \pm \frac{a\gamma}{2i} \; (a = 0, 1, 2, 3, \ldots) \,. \tag{99}$$

The distance between two maxima is $\frac{\gamma}{2i}$, and the number of grating lines is $2Ni$. In addition, the location of the maxima is independent of whether N is even or odd.

We obtain therefore the following result:

If the two i*th maxima contribute to image formation without the central image, we get a dissimilar image of the object because* $2Ni$ *grating lines appear instead of the* N *actually existing ones. The intensity decrease from the maximum to the minimum follows the law* $\cos^2 u$*.*

Bibliography on the theory of imaging of illuminated objects

This is an overview of the related works, as far as they are known to us.

Rayleigh, "On the theory of optical images, with special reference to the microscope," *Phil. Mag.* **42** (1896).

J. Stoney, "Microscopic vision," *Phil. Mag.* **42** (1896).

L. Wright, "Microscopic vision and images," *Phil. Mag.* **45** (1898).

K. Strehl, "Theorie des Mikroskops (Theory of the microscope)," *Zeitschr. f. Instrkde.* **18** (1898).

K. Strehl, "Das Pleurosigmabild (Image of the pleurosigma)," ibid. **19** (1899).

K. Strehl, "Theorie der allemeinen mikroskopischen Abbidung (Theory of general microscopic imaging)," *Meeting reports of Phys.-Med. Soc. Erlangen* **32** (1900).

J. W. Gordon, "An examination of the Abbe's diffraction theory of the microscope," *Journ. Roy. Microscop. Soc.* 1901.

J. Rheinberg, "The common basis of the theories of microscopic vision treated without the aid of mathematical formulae," Leipzig, Hirzel, 1902.

A. E. Conrady, "Theories of microscopical vision. I," *Roy. Microscop. Soc.* 1904.

J. Rheinberg, "Influence on images of gratings of phase-differences amongst their spectra," *Roy. Microscop. Soc.* 1904.

J. D. Everett, "A direct proof of Abbe's theories on the microscopic resolution of gratings," *Roy. Microscop. Soc.* 1904.

A. E. Conrady, "Theories of microscopical vision. II," *Roy. Microscop. Soc.* 1905.

J. Rheinberg, "Doubling of the lines in the Abbe's experiments," *Quekett Club Journal* **IX** (1905).

J. Rheinberg, "Influence on images of gratings of phase-differences amongst their spectra," *Roy. Microscop. Soc.* 1905.

A. E. Conrady, "An experimental proof of phase-reversal in diffraction spectra," *Roy. Microscop. Soc.* 1905.

A. B. Porter, "On the diffraction theory of microscopic vision," *Phil. Mag.* **11** (1906).

A. B. Porter, "On the nature of optical images," *Phys. Rev.* **24** (1907).

Many of the mentioned works contain valuable contributions to the theory of imaging of illuminated objects, with the object illuminated from various sides, and some contain special cases treated by us. None of these treatments, however, give a systematic construction of the theory of imaging of illuminated objects and a full analytical development of Abbe's theory.

Translators' notes

i. The language used in this preface is in early 20th century style and is sometimes pompous and awkward compared to today's German. To render this old style in English is difficult and not the aim of the translators.

ii. Chapter 1 of this book is more briefly written for two main reasons. First, geometrical optics is not the main focus of the book; it is included in order to provide the background material needed for certain topics discussed in Chapter 2. Second, as Lummer stated in the Preface, material for this chapter came mainly out of his "Optics," which was part of *Müller-Pouillet* (a standard German textbook on physics consisting of several volumes, originally written by Johann Heinrich Jakob Müller, based on a successful French textbook written by Claude Pouillet; the edition mentioned by Lummer in the Preface should be either the 9th or 10th edition, edited by Leopold Pfaundler). In the Appendix, we provide a brief introduction to geometrical optics starting from Fermat's principle.

iii. Triangles EAM and $EA'M$ are similar because angle φ is common to both of them and $AM/EM = EM/A'M = n'/n$.

iv. To better convey the meaning in the text, we have rotated the original sketch 90 deg counterclockwise.

v. Equation 1 intends to show that, once the positions of L, S, and M are fixed, the position of L' does not change with u (and u'), and this is what "homocentric null rays (i.e., paraxial ray pencils originating from the object) remain homocentric after refraction" means. This is because LM and LS are fixed. So is therefore the ratio $L'S/L'M$, according to Eq. 1. Now let $L'S/L'M = \eta$. Since $L'S = L'M + MS$, $L'M = MS/(\eta - 1)$. Since MS is fixed, so is $L'M$. That is, L' does not move with u.

vi. The equation below is simply a recast of Eq. 1, with $s = LS$, $s' = L'S$, $s - r = LM$, and $s' - r = L'M$.

vii. Figure 3.

viii. I' has to be the conjugate ray of I. That is, I' is the continuation of I in the image space. Because i' is the image of i, any ray that starts at i must pass through i'. Likewise, any ray that starts at z must pass through z'. I possesses both of these properties. Therefore, I' must pass through both i' and z'.

ix. Since P lies on both rays I and II, P' must also lie on both rays I' and II'. Hence, it must lie at the intersection of I' and II'.

x. $zz' = FF'$ is commonly known as Newton's equation.

xi. That is, points A' and A in Fig. 1 and points L and L' in Fig. 8.

xii. The original German word is Vereinigungsweite. We have not found the corresponding technical term in English.

xiii. The above equation is incorrect. The ratio $\frac{n'}{n}$ should be $\frac{n}{n'}$, for the situation here is different from that associated with Fig. 3; here, the object point L lies inside the refracting sphere whereas the (virtual) image point L' lies in the ambient medium outside the sphere. The same mistake was made in the equation immediately below, and the ratio in Eq. 9 should be $\frac{n'}{n}$.

xiv. If we let $\overline{NL'} = a$ and $\overline{Ll'} = b$, then $\overline{Nl'} - \overline{NL'} = \sqrt{a^2 + b^2} - a = a(\sqrt{1 + (b/a)^2} - 1)$ and $(\overline{Nl'} - \overline{NL'})/\overline{NL'} = \sqrt{1 + (b/a)^2} - 1 = (b/a)^2 - (b/a)^4 + \cdots$. But $\overline{Ll'}/\overline{NL'} = b/a$. This is what is meant by "$\overline{Ll'}$ is small to the first order and $\overline{Nl'} - \overline{NL'}$ is small to the second order." They are both compared against $\overline{NL'}$.

xv. Figure 10 is drawn incorrectly. The straight line leading to B' at the bottom of the figure ought to go through the center of the lens.

xvi. The literal translation of this sentence is "If the system is so calculated that this condition is satisfied." However, per Bernd Geh (formerly of ZEISS), there is sometimes no differentiation between calculating and designing in German, since the person designing the lens is also the one calculating it, or at least telling their (human) calculators what to calculate.

xvii. The original German word is Dioptriker. Dioptrics is the branch of optics that deals with refractive systems. So a Dioptriker is someone who is an expert in this field, per Bernd Geh. Today we call such a person a lens designer.

xviii. The text mentions Green's theorems, but the only one obvious to us for solving the wave equation with boundary conditions is

$$\int_V (\varphi\nabla^2 G - G\nabla^2\varphi)\, d\nu = -\int_\Sigma (\varphi\frac{\partial G}{\partial \nu} - G\frac{\partial \varphi}{\partial \nu})\, d\sigma\ ,$$

where ν is the inward unit normal vector of surface Σ that encloses volume V.

xix. The Laplace operator is defined as $\Delta = \nabla^2$.

xx. Equation (13) may be obtained as follows. Tackle the problem first in the frequency domain. That is, solve first $\varphi(P, \omega)$. From Green's theorem, aided by the so-called free-space Green's function $G(r, \omega) = e^{i(\omega/a)r}/r$, the following expression results (see J. W. Goodman, *Introduction to Fourier Optics*, 4th ed., W. H. Freeman and Company, New York, 2017, Section 3.3):

$$\varphi(P,\omega) = \frac{1}{4\pi}\int\limits_\Sigma d\sigma \left[\varphi(P',\omega)\frac{\partial}{\partial\nu}\left(\frac{e^{i(\omega/a)r}}{r}\right) - \frac{e^{i(\omega/a)r}}{r}\frac{\partial\varphi(P',\omega)}{\partial\nu}\right]\ ,$$

where P' is on Σ. To get $\varphi(P, t)$, perform the inverse Fourier transform: $\varphi(P, t) = \frac{1}{2\pi}\int \varphi(P, \omega)e^{-i\omega t}\, d\omega$.

$$
\begin{aligned}
\varphi(P,t) &= \frac{1}{4\pi}\int\limits_{\Sigma} d\sigma\left[\frac{1}{2\pi}\int \varphi\frac{\partial}{\partial \nu}\left(\frac{e^{i(\omega/a)r}}{r}\right)e^{-i\omega t}\,d\omega\right.\\
&\quad\left.-\frac{1}{2\pi}\int\frac{e^{i(\omega/a)r}}{r}\frac{\partial\varphi}{\partial\nu}e^{-i\omega t}\,d\omega\right]\\
&= \frac{1}{4\pi}\int\limits_{\Sigma} d\sigma\left[\frac{1}{2\pi}\int i(\omega/a)\varphi\frac{e^{i(\omega/a)r}}{r}e^{-i\omega t}\frac{\partial r}{\partial\nu}\,d\omega\right.\\
&\quad\left.+\frac{1}{2\pi}\int\varphi\frac{\partial(1/r)}{\partial\nu}e^{i(\omega/a)r}e^{-i\omega t}\,d\omega-\frac{1}{2\pi}\int\frac{1}{r}\frac{\partial\varphi}{\partial\nu}e^{i(\omega/a)r}e^{-i\omega t}\,d\omega\right]\\
&= \frac{1}{4\pi}\int\limits_{\Sigma} d\sigma\left[\varphi(P',t')\frac{\partial(1/r)}{\partial\nu}-\frac{1}{ar}\frac{\partial\varphi(P',t')}{\partial t'}\cdot\frac{\partial r}{\partial\nu}-\frac{1}{r}\frac{\partial\varphi(P',t')}{\partial\nu}\right]_{t'=t-\frac{r}{a}}
\end{aligned}
$$

xxi. $\overline{\varphi_P^2} = \frac{1}{T}\int_0^T \varphi_P^2\,dt$

xxii. Although not directly mentioned here, such expressions for $\mathfrak{E}$ and $\mathfrak{H}$ indicate to us that the authors meant for φ to be a component (perpendicular to the direction of wave propagation) of the electric Hertz vector Π_e, which is defined by the relationships $\phi = -\nabla\cdot\Pi_e$ and $A = \frac{1}{a}\frac{\partial\Pi_e}{\partial t}$, where ϕ and A are scalar and vector potentials, respectively. Since $\mathfrak{E} = -\frac{1}{a}\frac{\partial A}{\partial t} - \nabla\phi$, we have $\mathfrak{E} = \nabla(\nabla\cdot\Pi_e) - \frac{1}{a^2}\frac{\partial^2\Pi_e}{\partial t^2}$. Also, $\mathfrak{H} = \nabla\times A = \frac{1}{a}\frac{\partial(\nabla\times\Pi_e)}{\partial t}$. This identification is confirmed later.

xxiii. In the Lorentz gauge, $\nabla^2\phi - \frac{1}{a^2}\frac{\partial^2\phi}{\partial t^2} = -4\pi\rho = 0$ and $\nabla^2 A - \frac{1}{a^2}\frac{\partial^2 A}{\partial t^2} = -\frac{4\pi}{a}j = 0$ in free space. Hence, it is also true that $\nabla^2\Pi_e - \frac{1}{a^2}\frac{\partial^2\Pi_e}{\partial t^2} = 0$. That is, Π_e satisfies the wave equation. So does any of its components φ.

xxiv. To be mentioned a few lines down in the text, the φ below can be viewed as the x-component of the electric Hertz vector $\Pi_e = \varphi\hat{x}$ of an x-oscillating electric dipole $A\cos(2\pi/T)t\hat{x}$ situated at the origin of the coordinate system. See, e.g., M. Born and E. Wolf, *Principles of Optics*, 7th ed., Cambridge University Press, Cambridge, 1999, Section 2.2.3.

xxv. Although the absolute-value sign appears in Eq. 15, the expression is actually the far-field component ($1/r$ dependence only) of the electric field produced by the electric dipole used in this case. This far field points in the $\hat{\vartheta}$-direction. See M. Born and E. Wolf, op. cit.

xxvi. The value for both cases is $1 - \varepsilon^2/4$.

xxvii. In the expression of r below, terms inside the braces come from Taylor expanding $\sqrt{1+X}$ as $1+\frac{1}{2}X-\frac{1}{4}X^2$, where $X=\frac{\xi^2+\eta^2-2(x\xi+y\eta)}{r_0^2}$ and disregarding terms containing higher than the second power of (ξ/r_0) and (η/r_0). Such an expression of r leads to the formula for Fresnel diffraction.

xxviii. Unlike the optical path length, the geometrical path length R' is not the same for all rays from L to P.

xxix. $\overline{Ld\varphi P_1}=\overline{d\varphi a P_1}+\overline{Ld\varphi}=(\overline{L_1AP_1}-\overline{L_1d\varphi})+\overline{Ld\varphi}=\text{const.}-(\overline{L_1d\varphi}-\overline{Ld\varphi})$.

xxx. See Eq. 17.

xxxi. Here ν is the polar angle in the xy-plane, with the x-axis pointing into the paper and $e=1$. The same expression for $d\varphi$ with an arbitrary e is given a few lines down.

xxxii. The refractive index n of a material is the ratio between the speed of light in vacuum c and that in the material v, i.e., $n=c/v$. Hence, n is inversely proportional to the wavelength of light in that material λ, as $v=\lambda f$, where f is the frequency of vibration of the light source. Therefore, n is inversely proportional to λ. For two different media, we then have $n/n'=\lambda'/\lambda$.

xxxiii. The task here is to change the integration over $d\varphi'$ to that over $d\varphi$ so that the integrands of s_1 and s_2 can be equated. The relationship between $d\varphi'$ and $d\varphi$ is given by the equation directly above Eq. 19.

xxxiv. The value of the ratio A'/A can be obtained by taking the square root of Eq. 19.

xxxv. The expression for the electric field here is the same as the one given by Eq. 15, with e and A replacing r and $4\pi^2A/\lambda^2$ in that equation, except that the field here is y-oscillating instead of x-oscillating.

xxxvi. It is difficult to understand why the authors carried out the steps below, as the calculation is logically flawed because the vector of the radiation field, which is orthogonal to the vector $\mathfrak{e}$, also rotates and cannot be summed algebraically. A more logical way may be to go after the average intensity. Since the source is incoherent, the average intensity of the unpolarized radiation from the surface element perpendicular to the axis of the optical system is simply given by

$$\frac{1}{2\pi}\int_0^{2\pi} \mathfrak{e}^2\, dv \propto \frac{1}{2\pi}\int_0^{2\pi} \sin^2\vartheta\, dv = \frac{1}{2\pi}\int_0^{2\pi} (1-\sin^2 u\cos^2 v)\, dv$$
$$= 1 - \frac{\sin^2 u}{2} = \frac{1+\cos^2 u}{2}\,.$$

For small u, $\cos u = 1 - \frac{u^2}{2}$. Disregarding the term $\frac{u^4}{4}$, $\frac{1+\cos^2 u}{2} = \frac{2-u^2}{2} = 1-\frac{u^2}{2} = \cos u$. Therefore, $\frac{1}{2\pi}\int_0^{2\pi} \mathfrak{e}^2\, dv \propto \cos u$, in agreement with Lambert's cosine law.

xxxvii. $I = \frac{1}{2\pi}\int_0^{2\pi}\sqrt{1-\sin^2 u\cos^2 v}\, dv$ is related to the complete elliptic integral of the second kind and can be evaluated as follows. Let $\sin u = k$ and expand $\sqrt{1-\sin^2 u\cos^2 v}$ in Taylor's series as

$$\sqrt{1-\sin^2 u\cos^2 v} = 1 - \frac{1}{2}k^2\cos^2 v - \frac{1}{8}k^4\cos^4 v - \cdots$$

$$\int_0^{2\pi} 1\cdot dv = 2\pi$$

$$\int_0^{2\pi} \cos^2 v\, dv = \int_0^{2\pi} \frac{1+\cos 2v}{2}\, dv = \pi$$

$$\int_0^{2\pi} \cos^4 v dv = \int_0^{2\pi} \cos^2 v(1-\sin^2 v)\, dv = \int_0^{2\pi} \cos^2 v\, dv$$
$$- \int_0^{2\pi} \frac{1+\cos 2v}{2}\cdot\frac{1-\cos 2v}{2}\, dv = \frac{3\pi}{4}\,.$$

Hence,

$$I = \frac{1}{2\pi}\left(2\pi - \frac{1}{2}k^2\pi - \frac{1}{8}k^4\frac{3\pi}{4} - \cdots\right) = 1 - \frac{1}{4}k^2 - \frac{3}{64}k^4 - \cdots$$
$$= 1 - \frac{1}{4}\sin^2 u - \frac{3}{64}\sin^4 u - \cdots$$

The above result differs, however, from the series expansion for I given in the text, which is

$$\cos^2\frac{u}{2}\left\{1+\frac{1}{4}\tan^4\frac{u}{2}+\frac{1}{64}\tan^8\frac{u}{2}+\cdots\right\}.$$

The reason the authors sought such a form was the subsequent approximation: if u is small, then only the term $\cos^2\frac{u}{2}=\frac{1+\cos u}{2}$ has to be kept. However, with our expansion using the sine function, if u is small, then

$$\begin{aligned}
I\approx 1-\frac{1}{4}\sin^2 u=\frac{3+\cos^2 u}{4}&=\frac{2+1+\cos^2 u}{4}\\
&=\frac{2+2\cos u+(1-\cos u)^2}{4}\\
&=\frac{1+\cos u}{2}+\left(\frac{1-\cos u}{2}\right)^2\approx\frac{1+\cos u}{2}.
\end{aligned}$$

Therefore, the conclusions are the same. We could not figure out how the authors arrived at their series expansion. They could have looked up the result from a published book of mathematical tables available at the time.

xxxviii. Again, both expressions approximate to $1-u^2/4$ for $u\ll 1$.

xxxix. In spherical coordinates,

$$\nabla^2\mathfrak{E}'=\frac{1}{e'}\frac{\partial^2(e'\mathfrak{E}')}{\partial e'^2}+\frac{1}{e'^2\sin u'}\frac{\partial\left(\sin u'\frac{\partial\mathfrak{E}'}{\partial u'}\right)}{\partial u'}+\frac{1}{e^2\sin^2 u'}\frac{\partial^2\mathfrak{E}'}{\partial v'^2}.$$

The last term is absent in the text because, after averaging, $\mathfrak{E}'$ is no longer a function of v'.

xl. This solution may be checked in the following way. Let $\mathfrak{F}=e'\mathfrak{E}'$ and rewrite the above equation as follows:

$$\frac{1}{a'^2}\frac{\partial^2\mathfrak{F}}{\partial t^2}-\frac{\partial^2\mathfrak{F}}{\partial e'^2}=\frac{1}{e'^2\sin u'}\frac{\partial\left(\sin u'\frac{\partial\mathfrak{F}}{\partial u'}\right)}{\partial u'}.$$

Plugging the proposed solution

$$\begin{aligned}
\mathfrak{F}&=\mathrm{const}\cdot\cos u'\left\{\cos 2\pi\left(\frac{t}{T}+\frac{e'}{\lambda'}\right)-\frac{\lambda'}{2\pi e'}\sin 2\pi\left(\frac{t}{T}+\frac{e'}{\lambda'}\right)\right\}\\
&=\mathfrak{A}+\mathfrak{B}
\end{aligned}$$

into the right-hand side (RHS) of the equation returns $-2\mathfrak{F}/e'^2$. When plugged into the left-hand side (LHS) of the equation, the first term, designated as $\mathfrak{A}$, which is a solution to the wave equation $\frac{1}{a'^2}\frac{\partial^2\mathfrak{F}}{\partial t^2} - \frac{\partial^2\mathfrak{F}}{\partial e'^2} = 0$, returns zero. The second term $\mathfrak{B}$ is also a solution to the wave equation, but an equation for spherical waves. That is, $\frac{1}{a'^2}\frac{\partial^2\mathfrak{B}}{\partial t^2} - \frac{1}{e'}\frac{\partial^2(e'\mathfrak{B})}{\partial e'^2} = 0$. Hence, we only have to check whether $\frac{1}{e'}\frac{\partial^2(e'\mathfrak{B})}{\partial e'^2} - \frac{\partial^2\mathfrak{B}}{\partial e'^2}$ is equal to $-2\mathfrak{F}/e'^2$. For this we proceed as follows:

$$\begin{aligned}
\frac{1}{e'}\frac{\partial^2(e'\mathfrak{B})}{\partial e'^2} - \frac{\partial^2\mathfrak{B}}{\partial e'^2} &= \frac{\partial}{\partial e'}\left(\frac{1}{e'}\frac{\partial(e'\mathfrak{B})}{\partial e'} - \frac{\partial\mathfrak{B}}{\partial e'}\right) + \frac{1}{e'^2}\frac{\partial(e'\mathfrak{B})}{\partial e'} \\
&= \frac{\partial^2}{\partial e'^2}\left(\frac{1}{e'}e'\mathfrak{B} - \mathfrak{B}\right) + \frac{\partial}{\partial e'}\left(\frac{1}{e'^2}e'\mathfrak{B}\right) + \frac{1}{e'^2}\frac{\partial(e'\mathfrak{B})}{\partial e'} \\
&= \frac{\partial}{\partial e'}\left(\frac{\mathfrak{B}}{e'}\right) + \frac{1}{e'^2}\frac{\partial(e'\mathfrak{B})}{\partial e'} \\
&= \frac{2}{e'}\frac{\partial\mathfrak{B}}{\partial e'}\,;
\end{aligned}$$

if we plug in

$$\mathfrak{B} = -\mathrm{const}\cdot u'\cdot\frac{\lambda'}{2\pi e'}\sin 2\pi\left(\frac{t}{T} + \frac{e'}{\lambda'}\right),$$

then

$$\frac{2}{e'}\frac{\partial\mathfrak{B}}{\partial e'} = \frac{-2\mathfrak{F}}{e'^2}\,.$$

Hence, LHS = RHS and $\mathfrak{F}$ is indeed a solution.

xli. For obtaining the subsequent result, one essentially makes use of the result obtained in § 12.

xlii. The expression below benefitted from the use of the following trigonometric identities:

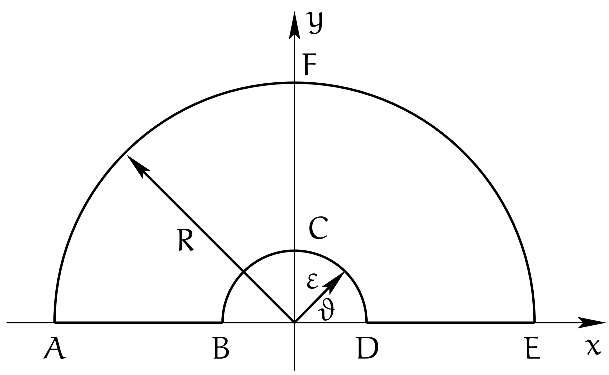
$$\begin{aligned}
\cos A - \cos B &= -2\sin\frac{A+B}{2}\sin\frac{A-B}{2} \\
\sin A - \sin B &= 2\cos\frac{A+B}{2}\sin\frac{A-B}{2}\,.
\end{aligned}$$

xliii. The original book contained an image page for Fig. 28 that was likely copied from a photographic plate. To produce the plots for Figs. 28 and 29, we calculated the diffraction pattern and produced a gray level image with a contrast function of $\sqrt[6]{I}$. Without this contrast function, only the central maximum would be visible.

xliv. Following the authors, the intensity is defined as the time average, over N cycles (N being large), of the square of the electric field. That is,

$$
\begin{aligned}
\mathfrak{E} \equiv \frac{1}{NT}\int_0^{NT} \mathfrak{E}^2\,dt &= \frac{1}{NT}\int_0^{NT}\left(A\cos 2\pi\frac{t}{T} + B\sin 2\pi\frac{t}{T}\right)^2 \\
&= \frac{1}{NT}\int_0^{NT}\left(A^2\cos^2 2\pi\frac{t}{T} + AB\sin 4\pi\frac{t}{T} + B^2\sin^2 2\pi\frac{t}{T}\right)dt \\
&= \frac{1}{NT}\int_0^{NT}\left(A^2\frac{1+\cos 4\pi\frac{t}{T}}{2} + AB\sin 4\pi\frac{t}{T} + B^2\frac{1-\cos 4\pi\frac{t}{T}}{2}\right)dt \\
&= \frac{A^2+B^2}{2}\,.
\end{aligned}
$$

xlv. The integral in Eq. 38 below can be obtained by contour integration, using the contour in the following diagram:



Let us first evaluate the integral in Eq. 41 since it is the simpler one. Since there are no singularities in the closed contour ABCDEFA, we have by

Cauchy's integral theorem,

$$\oint \frac{e^{iz}}{z}\,dz = \int\limits_{\overline{AB}} + \int\limits_{\overline{BCD}} + \int\limits_{\overline{DE}} + \int\limits_{\overline{EFA}} = 0$$

$$\int\limits_{\overline{AB}} + \int\limits_{\overline{DE}} = \int\limits_{-R}^{-\varepsilon} \frac{e^{ix}}{x}\,dx + \int\limits_{\varepsilon}^{R} \frac{e^{ix}}{x}\,dx = \int\limits_{\varepsilon}^{R} \frac{e^{ix}-e^{-ix}}{x}\,dx = 2i\int\limits_{\varepsilon}^{R} \frac{\sin x}{x}\,dx$$

$$\int\limits_{\overline{BCD}} = \int\limits_{\pi}^{0} \frac{e^{i\varepsilon e^{i\vartheta}}}{\varepsilon e^{i\vartheta}}\,d(\varepsilon e^{i\vartheta}) = i\int\limits_{\pi}^{0} e^{i\varepsilon e^{i\vartheta}}\,d\vartheta$$

$$\int\limits_{\overline{EFA}} = i\int\limits_{0}^{\pi} e^{iRe^{i\vartheta}}\,d\vartheta = i\int\limits_{0}^{\pi} e^{-R\sin\vartheta}e^{-iR\cos\vartheta}\,d\vartheta\ .$$

Now let $R \to \infty$ and $\varepsilon \to 0$. We then have

$$\oint \frac{e^{iz}}{z}\,dz = 2i\int\limits_{0}^{\infty} \frac{\sin x}{x}\,dx + i\int\limits_{\pi}^{0} 1\cdot d\vartheta + i\int\limits_{0}^{\pi} 0\cdot e^{iR\cos\vartheta}\,d\vartheta = 0\,.$$

Therefore,

$$\int\limits_{0}^{\infty} \frac{\sin x}{x}\,dx = \frac{\pi}{2}\,.$$

For the integral in Eq. 38, we evaluate the following contour integral:

$$\oint \frac{e^{2iz}-1}{2z^2}\,dz = \int\limits_{\overline{AB}} + \int\limits_{\overline{BCD}} + \int\limits_{\overline{DE}} + \int\limits_{\overline{EFA}} = 0$$

$$\int\limits_{\overline{AB}} + \int\limits_{\overline{DE}} = \int\limits_{-R}^{-\varepsilon} \frac{e^{2ix}-1}{2x^2}\,dx + \int\limits_{\varepsilon}^{R} \frac{e^{2ix}-1}{2x^2}\,dx = \int\limits_{\varepsilon}^{R} \frac{e^{2ix}+e^{-2ix}-2}{2x^2}\,dx = \int\limits_{\varepsilon}^{R} \frac{\cos 2x - 1}{x^2}\,dx$$

$$= -2\int\limits_{\varepsilon}^{R} \frac{\sin^2 x}{x^2}\,dx\ .$$

Because $R \to \infty$ and $\varepsilon \to 0$, $\int_{EFA}$ can again be shown to be zero. Using $e^{iz} \approx 1 + iz$ for small $|z|$, we have

$$\int\limits_{BCD} = \int\limits_{\pi}^{0} \frac{(1 + 2i\varepsilon e^{i\vartheta}) - 1}{2\varepsilon^2 e^{2i\vartheta}}\, d(\varepsilon e^{i\vartheta}) = \int\limits_{\pi}^{0} \frac{2i\varepsilon e^{i\vartheta}}{2\varepsilon^2 e^{2i\vartheta}} \varepsilon e^{i\vartheta} i\, d\vartheta = \pi\,.$$

Therefore,

$$\oint \frac{e^{2iz} - 1}{2z^2}\, dz = -2 \int\limits_{0}^{\infty} \frac{\sin^2 x}{x^2}\, dx + \pi + 0 = 0$$

and

$$\int\limits_{0}^{\infty} \frac{\sin^2 x}{x^2}\, dx = \frac{\pi}{2}\,.$$

xlvi. The solidly drawn curve in Fig. 36 is actually the absolute value of the amplitude.

xlvii. The (more familiar) Fourier integral theorem in complex form is

$$f(x) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} dz\, e^{-ixz} \int\limits_{-\infty}^{\infty} du\, f(u) e^{izu}\,.$$

It basically says that the inverse Fourier transform of a Fourier transform gives us the original function back. The above integral can be re-expressed as

$$\begin{aligned} f(x) &= \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} dz \int\limits_{-\infty}^{\infty} du\, f(u) e^{iz(u-x)} \\ &= \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} dz \int\limits_{-\infty}^{\infty} du\, f(u) \cos z(u-x) + \frac{i}{2\pi} \int\limits_{-\infty}^{\infty} dz \int\limits_{-\infty}^{\infty} du\, f(u) \sin z(u-x) \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} du\, f(u) \int_{-\infty}^{\infty} dz \cos z(u-x) + \frac{i}{2\pi}\int_{-\infty}^{\infty} du\, f(u)\int_{-\infty}^{\infty} dz \sin z(u-x) \\
&= \frac{1}{\pi}\int_{-\infty}^{\infty} du\, f(u) \int_{0}^{\infty} dz \cos z(u-x) + \frac{i}{2\pi}\cdot 0 = \frac{1}{\pi}\int_{0}^{\infty} dz \int_{-\infty}^{\infty} du\, f(u)\cos z(u-x)\,.
\end{aligned}
$$

The last form above is Eq. 52.

xlviii. The coefficient in front of the last integral below should be $-\frac{2}{\pi a}$.

xlix. Wiggles in the center of the slit are absent in the original hand-drawn graph.

l. If we let $\nu = 2\pi\beta' \frac{y-Y}{\lambda}$, then

$$
\int_{-\infty}^{\infty} dY\, 2\beta' \frac{\sin 2\pi\beta' \frac{y-Y}{\lambda}}{2\pi\beta' \frac{y-Y}{\lambda}} = \frac{\lambda}{\pi}\int_{-\infty}^{\infty} \frac{\sin\nu}{\nu}\, d\nu = \lambda\,.
$$

li. Taylor's expansion of J_1 goes as follows. First, let $J_1 = \frac{\lambda k}{\pi}\int_{z_0}^{z} \frac{\sin w}{w}\, dw$, where $z_0 = \frac{2\pi\alpha' x}{\lambda}$ and $z = \frac{2\pi\alpha'(x+a)}{\lambda}$. Since z is very close to z_0, we have

$$
\begin{aligned}
J_1 \approx \frac{\lambda k}{\pi}\Bigg(&\int_{z_0}^{z} \frac{\sin w}{w}\, dw\Big|_{z=z_0} + \frac{d}{dz}\int_{z_0}^{z}\frac{\sin w}{w}\, dw\Big|_{z=z_0}(z-z_0) \\
&+ \frac{1}{2!}\frac{d^2}{dz^2}\int_{z_0}^{z}\frac{\sin w}{w}\, dw\Big|_{z=z_0}(z-z_0)^2\Bigg) \\
= \frac{\lambda k}{\pi}\Bigg(& 0 + \frac{\sin z}{z}\Big|_{z=z_0}(z-z_0) + \frac{1}{2!}\frac{z\cos z - \sin z}{z^2}\Big|_{z=z_0}(z-z_0)^2\Bigg)\,.
\end{aligned}
$$

Inserting the values for z and z_0 gives us the expression for J_1 shown in the book. For J_2, first reverse the limits of integration and then let $z = \frac{2\pi\alpha'(x-a)}{\lambda}$.

lii. Note that for large apertures a/λ,

$$\int_{\xi}^{\xi+\frac{2\pi\alpha' a}{\lambda}} \frac{\sin w}{w}\,dw \simeq \int_{0}^{\infty} \frac{\sin w}{w}\,dw - \int_{0}^{\xi} \frac{\sin w}{w}\,dw \simeq \int_{\xi}^{\infty} \frac{\sin w}{w}\,dw$$
$$= \frac{\pi}{2} - \int_{0}^{\xi} \frac{w}{w}\,dw\ ,$$

where $\int_0^\infty \frac{\sin w}{w}\,dw = \frac{\pi}{2}$, and the final approximation is due to $\sin w \simeq w$ for small w.

liii. Derivation of Eq. 61:

$$J_1 - J_2 = \frac{\lambda k}{\pi}\left(\int_{\xi}^{\xi+\frac{2\pi\alpha' a}{\lambda}} - \int_{\xi-\frac{2\pi\alpha' a}{\lambda}}^{\xi}\right) \approx \frac{\lambda k}{\pi}\left(\int_{\xi}^{\infty} - \int_{\xi-\frac{2\pi\alpha' a}{\lambda}}^{0} - \int_{0}^{\xi}\right)$$
$$= \frac{\lambda k}{\pi}\left(\int_{0}^{\infty} - \int_{0}^{\xi} - \int_{\xi-\frac{2\pi\alpha' a}{\lambda}}^{0} - \int_{0}^{\xi}\right) = \frac{\lambda k}{\pi}\left(\int_{-\infty}^{0} - 2\int_{0}^{\xi} - \int_{\xi-\frac{2\pi\alpha' a}{\lambda}}^{0}\right)$$
$$= \frac{\lambda k}{\pi}\left(-2\int_{0}^{\xi} + \int_{-\infty}^{\xi-\frac{2\pi\alpha' a}{\lambda}}\right)$$

liv. Please note that, for this figure, we did an approximate curve fit to the graph in the original book using constant values of 0 and 1, and $\cos^2$ in the transition regions. We could not find information in the text that would allow us to reproduce the original graph exactly. The text itself mentions that this graph is not entirely correct.

lv. Wiggles in the centers of the two halves of the slit are absent in the original hand-drawn graph.

lvi. As before, the integration over Y gives λ.

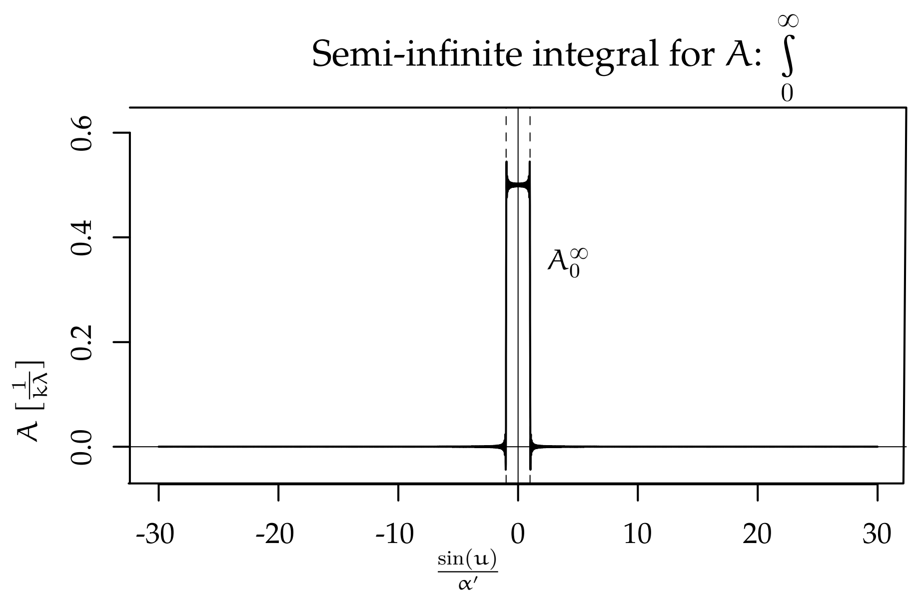
lvii. The integral is zero because the integrand is simply an odd function.

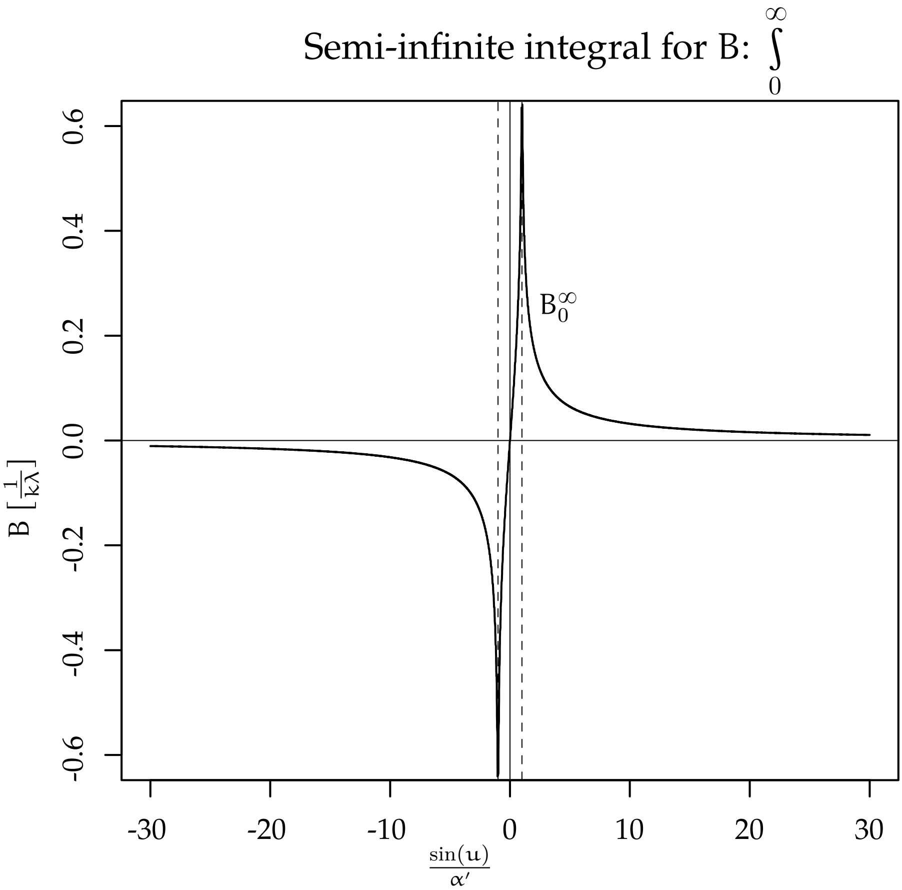
lviii. Cf. Eq. 53.

lix. Note that it is not clear how the authors arrived at the conclusion that B, which in this case is effectively equal to

$$B_0^\infty = \frac{k\lambda}{\pi} \int_0^\infty dw \, \frac{\sin w}{w} \cdot \sin\left(\frac{\sin u}{\alpha'} w\right),$$

is equal to $\frac{1}{2} B_\infty = 0$. In fact, B_0^∞ approaches zero only gradually as $\sin(u)$ becomes greater and greater than α' and the authors' claim of $B_0^\infty = 0$ is incorrect. For A, their claim that the integral is half the value of A_∞ is true, because its integrand is an even function with respect to w. The graphs below show the plots of both A_0^∞ and B_0^∞ using numerical integration.





lx. If $\sin u = \alpha' + \varepsilon$, then $\rho = \frac{\sin u}{\alpha'} = 1 + \frac{\varepsilon}{\alpha'}$, and A and B become

$$A = \int_{\xi}^{\frac{2\pi\alpha'\delta}{\lambda}} dw \, \frac{\sin w}{w} \cos(\rho w) = \int_{\xi}^{\frac{2\pi\alpha'\delta}{\lambda}} dw \, \frac{\sin w}{w} \cos(w + \frac{\varepsilon}{\alpha'} w)$$

$$= \int_{\xi}^{\frac{2\pi\alpha'\delta}{\lambda}} dw \, \frac{\sin w}{w} \left[\cos w \cos\left(\frac{\varepsilon}{\alpha'} w\right) - \sin w \sin\left(\frac{\varepsilon}{\alpha'} w\right)\right]$$

$$\approx \int_{\xi}^{\frac{2\pi\alpha'\delta}{\lambda}} dw \frac{\sin w}{w} \left(\cos w \cdot 1 - \sin w \cdot \frac{\varepsilon}{\alpha'} w\right)$$
$$= \int_{\xi}^{\frac{2\pi\alpha'\delta}{\lambda}} \frac{\sin w \cos w}{w}\, dw - \frac{\varepsilon}{\alpha'} \int_{\xi}^{\frac{2\pi\alpha'\delta}{\lambda}} \sin^2 w\, dw = \frac{1}{2} \int_{\xi}^{\frac{4\pi\alpha'\delta}{\lambda}} \frac{\sin w}{w}\, dw + (\varepsilon)$$

$$B = \int_{\xi}^{\frac{2\pi\alpha'\delta}{\lambda}} dw \frac{\sin w}{w} \sin(\rho w) = \int_{\xi}^{\frac{2\pi\alpha'\delta}{\lambda}} dw \frac{\sin w}{w} \sin(w + \frac{\varepsilon}{\alpha'} w)$$
$$= \int_{\xi}^{\frac{2\pi\alpha'\delta}{\lambda}} dw \frac{\sin w}{w} \left[\sin w \cos\left(\frac{\varepsilon}{\alpha'} w\right) + \cos w \sin\left(\frac{\varepsilon}{\alpha'} w\right)\right]$$
$$\approx \int_{\xi}^{\frac{2\pi\alpha'\delta}{\lambda}} dw \frac{\sin w}{w} \left(\sin w \cdot 1 + \cos w \cdot \frac{\varepsilon}{\alpha'} w\right) = \int_{\xi}^{\frac{2\pi\alpha'\delta}{\lambda}} \frac{\sin^2 w}{w}\, dw + (\varepsilon)\,.$$

lxi. The plot in Fig. 51 requires an explanation. First, since the intensity is plotted, the ordinate should be marked I instead of J used in the original text. Now, $I = A^2 + B^2$. Starting from Eq. 63, which has not been simplified with various approximations, we have

$$A = k \int_{-a}^{+a} dX\, 2\alpha' \frac{\sin 2\pi\alpha' \frac{x-X}{\lambda}}{2\pi\alpha' \frac{x-X}{\lambda}} \cos\left(2\pi \sin u \frac{x-X}{\lambda}\right)$$
$$B = k \int_{-a}^{+a} dX\, 2\alpha' \frac{\sin 2\pi\alpha' \frac{x-X}{\lambda}}{2\pi\alpha' \frac{x-X}{\lambda}} \sin\left(2\pi \sin u \frac{x-X}{\lambda}\right)\,.$$

Following the authors by setting $2\pi\alpha' \frac{x-X}{\lambda} = w$, we get

$$A = \frac{k\lambda}{\pi} \int_{2\pi\alpha' \frac{x-a}{\lambda}}^{2\pi\alpha' \frac{x+a}{\lambda}} dw \frac{\sin w}{w} \cos\left(\frac{w}{\alpha'} \sin u\right) \text{ and } B = \frac{k\lambda}{\pi} \int_{2\pi\alpha' \frac{x-a}{\lambda}}^{2\pi\alpha' \frac{x+a}{\lambda}} dw \frac{\sin w}{w} \sin\left(\frac{w}{\alpha'} \sin u\right).$$

Following the authors again by setting $x = a + \delta$, then

$$A = \frac{k\lambda}{\pi} \int\limits_{2\pi\alpha' \frac{\delta}{\lambda}}^{2\pi\alpha' \frac{2a+\delta}{\lambda}} dw \, \frac{\sin w}{w} \cos\left(\frac{w}{\alpha'} \sin u\right) \text{ and } B = \frac{k\lambda}{\pi} \int\limits_{2\pi\alpha' \frac{\delta}{\lambda}}^{2\pi\alpha' \frac{2a+\delta}{\lambda}} dw \, \frac{\sin w}{w} \sin\left(\frac{w}{\alpha'} \sin u\right).$$

Let us now take $\lambda = 0.6\,\mu\text{m}$. a and α' are suggested by the authors to be somewhat larger than 1 mm and 1 deg, respectively. So let us take $a = 1000\,\mu\text{m}$ and $\alpha' = \pi/180 \simeq \frac{1}{60}$. Therefore,

$$A = \frac{k\lambda}{\pi} \int\limits_{\frac{\delta}{6}}^{\frac{2000+\delta}{6}} dw \, \frac{\sin w}{w} \cos\left(\frac{w}{\alpha'} \sin u\right) \text{ and } B = \frac{k\lambda}{\pi} \int\limits_{\frac{\delta}{6}}^{\frac{2000+\delta}{6}} dw \, \frac{\sin w}{w} \sin\left(\frac{w}{\alpha'} \sin u\right) .$$

Since the upper limit is so large, we set it to infinity, following the authors. We can then approximate the integrals as

$$A = \frac{k\lambda}{\pi} \int\limits_0^\infty dw \, \frac{\sin w}{w} \cos\left(\frac{w}{\alpha'} \sin u\right) - \frac{k\lambda}{\pi} \int\limits_0^{\frac{\delta}{6}} dw \, \frac{\sin w}{w} \cos\left(\frac{w}{\alpha'} \sin u\right)$$

and

$$B = \frac{k\lambda}{\pi} \int\limits_0^\infty dw \, \frac{\sin w}{w} \sin\left(\frac{w}{\alpha'} \sin u\right) - \frac{k\lambda}{\pi} \int\limits_0^{\frac{\delta}{6}} dw \, \frac{\sin w}{w} \sin\left(\frac{w}{\alpha'} \sin u\right) .$$

We now let $\sin u > \alpha'$ or $\sin u < -\alpha'$. The first integral in A is equal to zero (half the value of A_∞; see Eq. 66). Hence,

$$A = -\frac{k\lambda}{\pi} \int\limits_0^{\frac{\delta}{6}} dw \, \frac{\sin w}{w} \cos\left(\frac{w}{\alpha'} \sin u\right) .$$

The first integral in B is not zero for $\sin u > \alpha'$ or $\sin u < -\alpha'$, as we have already remarked in a previous translators' note. From the graph in that note, we can see that if $\sin u \gg \alpha'$ or $\sin u \ll \alpha'$, then the value of the first integral is small and is negligible. Then,

$$B = -\frac{k\lambda}{\pi} \int\limits_0^{\frac{\delta}{6}} dw \, \frac{\sin w}{w} \sin\left(\frac{w}{\alpha'} \sin u\right) .$$

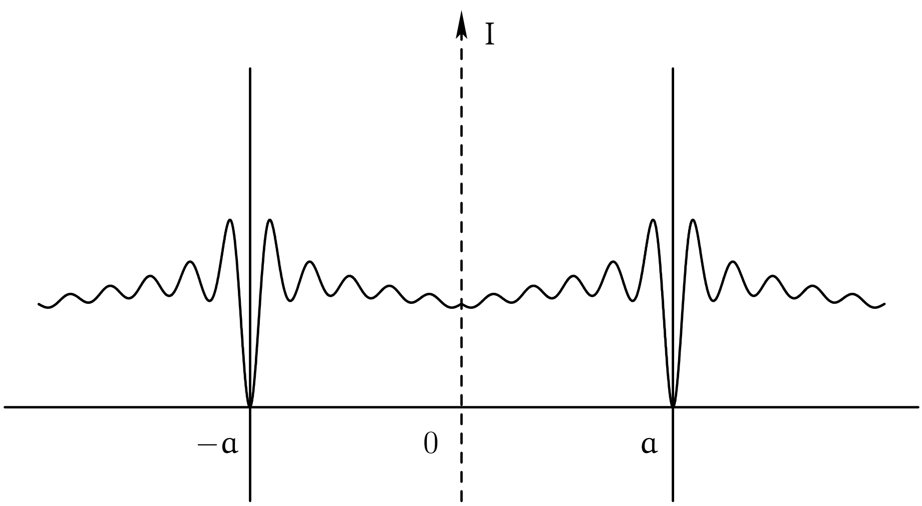
The numerical example mentioned in the book uses $\sin u = \sin 30^o = \frac{1}{2}$. Then, $\frac{\sin u}{\alpha'} = 30$ and

$$A = \text{const} \int_0^{\frac{\delta}{6}} dw\, \frac{\sin w}{w} \cos(30w) \text{ and } B = \text{const} \int_0^{\frac{\delta}{6}} dw\, \frac{\sin w}{w} \sin(30w)\ .$$

But it appears that Fig. 51 plots only the intensity associated with A. That is, the plot is of

$$\left[\text{const} \int_0^{\frac{\delta}{6}} dw\, \frac{\sin w}{w} \cos(30w) \right]^2 .$$

The book also mentions situations for which $\sin u = \alpha' + \varepsilon$, where ε is small. Such cases will not result in the intensity of A as shown in Fig. 51. For example, if $\frac{\sin u}{\alpha'} = 10$, then A^2 will look like



In fact, if $\frac{\sin u}{\alpha'} = 1$, then the integral becomes

$$\int_0^{\frac{\delta}{6}} dw\, \frac{\sin w}{w} \cos w = \frac{1}{2} \int_0^{\frac{\delta}{6}} d2w\, \frac{\sin 2w}{2w} = \frac{\pi}{4}\ , \text{as } \delta \to \infty\ .$$

That is, A^2 does not die out.

lxii. To obtain the expression shown below in the text, we Taylor expand A and retain only the first two terms of the series:

$$A(\varepsilon) = A(\varepsilon = 0) \left.\frac{dA}{d\varepsilon}\right|_{\varepsilon=0} \varepsilon = 0 + \frac{k\lambda}{\pi} \lim_{\varepsilon \to 0} \left(\frac{\int_{\xi-\varepsilon}^{\xi+\varepsilon} dw \cdots - \int_{\xi}^{\xi} dw \cdots}{\varepsilon} \right) \varepsilon$$
$$= \frac{k\lambda}{\pi} \frac{\sin\xi}{\xi} \cos\left(\frac{\sin u}{\alpha'}\xi\right)(2\varepsilon)\,.$$

Similarly,

$$B(\varepsilon) = \frac{k\lambda}{\pi} \frac{\sin\xi}{\xi} \sin\left(\frac{\sin u}{\alpha'}\xi\right)(2\varepsilon)\,.$$

Then,

$$I = A^2 + B^2 = 4\varepsilon^2 \frac{k^2\lambda^2}{\pi^2}\left(\frac{\sin\xi}{\xi}\right)^2 \left[\cos^2\left(\frac{\sin u}{\alpha'}\xi\right) + \sin^2\left(\frac{\sin u}{\alpha'}\xi\right)\right]$$
$$= 4\varepsilon^2 \frac{k^2\lambda^2}{\pi^2}\left(\frac{\sin\xi}{\xi}\right)^2 .$$

lxiii. See Eq. 29.

lxiv. Physically, the effect of diffraction occurs, of course, in the image plane. But it can just as well be described by the conjugate points in the object plane. See § 13. This sentence in the text simply means that the first integration gives the point-spread function of the system.

lxv. One can see from Fig. 53 that all orders of diffraction from the grating (corresponding to those in the physical region in Fig. 55 below) are captured by the intermediate surface, which is the entire hemisphere, corresponding to the boundaries $\xi' = \pm 1$ and $\eta' = \pm 1$. The diffraction orders in the imaginary regions in Fig. 55 can only appear on the intermediate surface if one increases the pitch of the grating or reduces the wavelength of the illuminating light.

lxvi. This statement simply means that the final image is the same as the object (i.e. 100% similarity), because the expression above is the same as Expression 67. This has to be the case because S_1^* (or S_1) includes all the diffracted orders in its expression.

lxvii. Since the transmission coefficient of the object $\varphi(X, Y) \neq 0$ only for $A_1 \leqslant X \leqslant A_2$ and $B_1 \leqslant Y \leqslant B_2$, $F(X) \neq 0$ also only for $A_1 \leqslant X \leqslant A_2$. This means that

$$\int_{-\infty}^{\infty} d\xi \int_{A_1}^{A_2} dX\, F(X) \cos 2\pi\xi(x - X) = \int_{-\infty}^{\infty} d\xi \int_{-\infty}^{\infty} dX\, F(X) \cos 2\pi\xi(x - X)\,,$$

which is equal to $F(x)$. Analogously,

$$\int_{-\infty}^{\infty} d\xi \int_{A_1}^{A_2} dX\, F(X) \sin 2\pi\xi(x - X) = \int_{-\infty}^{\infty} d\xi \int_{-\infty}^{\infty} dX\, F(X) \sin 2\pi\xi(x - X) = 0\,,$$

which can be shown by integrating the variable ξ first, since $\sin 2\pi\xi(x - X)$ is an odd function.

lxviii. This citation refers to a particular volume of the journal *"Annalen der Physik und Chemie,"* the citation of which includes the name of the editor-in-chief of that volume, Johann Christian Poggendorff, who held that position from 1824 to 1876. Such a citation scheme was useful because for each new editor-in-chief, the volume number was reset to 1. "Jubelband" refers to the fact that this was published as a jubilee volume for the editor-in-chief's 50th year of editing the journal, as a celebration. In a postscript of his paper, Helmholtz acknowledged the fact that Abbe's work had preceded Helmholtz's, but also pointed out that Abbe had not yet published the proofs of his findings. Abbe's paper on this subject is "Beiträge zur Theorie des Mikrosops und der mikroskopischen Wahrnehmung," in *Archiv für mikroskopische Anatomie 9*, pp. 413–468 (1873).

lxix. Eq. 91 is obtained from Eq. 90 as follows:

$$\frac{\partial J_0}{\partial x} = \frac{2a}{\gamma} \int_0^{\frac{2\pi a}{N\gamma}} dw\, \frac{\sin \frac{N\gamma w}{2a}}{w} \left(-\frac{w}{a} \sin \frac{xw}{a}\right) = \frac{1}{\gamma} \int_0^{\frac{2\pi a}{N\gamma}} dw \left(-2 \sin \frac{N\gamma w}{2a} \sin \frac{xw}{a}\right)$$

$$= \frac{1}{\gamma} \int_0^{\frac{2\pi a}{N\gamma}} dw \left[\cos \frac{(N\gamma + 2x)w}{2a} - \cos \frac{(N\gamma - 2x)w}{2a}\right]$$

$$
\begin{aligned}
&= \frac{1}{\gamma}\left[\frac{2a}{N\gamma+2x}\sin\frac{(N\gamma+2x)w}{2a}-\frac{2a}{N\gamma-2x}\sin\frac{(N\gamma-2x)w}{2a}\right]\Bigg|_0^{\frac{2\pi a}{N\gamma}} \\
&= \frac{1}{\gamma}\left[\frac{2a}{N\gamma+2x}\sin\frac{(N\gamma+2x)\frac{2\pi a}{N\gamma}}{2a}-\frac{2a}{N\gamma-2x}\sin\frac{(N\gamma-2x)\frac{2\pi a}{N\gamma}}{2a}\right] \\
&= \frac{1}{\gamma}\left[\frac{2a}{N\gamma+2x}\sin\left(\pi+\frac{2\pi x}{N\gamma}\right)-\frac{2a}{N\gamma-2x}\sin\left(\pi-\frac{2\pi x}{N\gamma}\right)\right] \\
&= \frac{1}{\gamma}\left(-\frac{2a}{N\gamma+2x}\sin\frac{2\pi x}{N\gamma}-\frac{2a}{N\gamma-2x}\sin\frac{2\pi x}{N\gamma}\right) = -\frac{4aN}{N^2\gamma^2-4x^2}\sin\frac{2\pi x}{N\gamma}\,,
\end{aligned}
$$

which is Eq. 91.

lxx. The final answer in the original text is erroneously stated as $-0.79\frac{a}{\gamma}$.

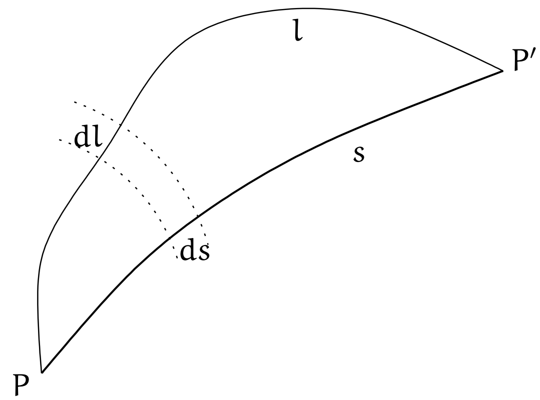
lxxi. The original text has $J_0(x = N\gamma) = -0.2\, J_0(x = 0)$, a consequence of taking $J_0(x = N\gamma) = -0.79\frac{a}{\gamma}$ above.

A brief introduction to geometrical optics

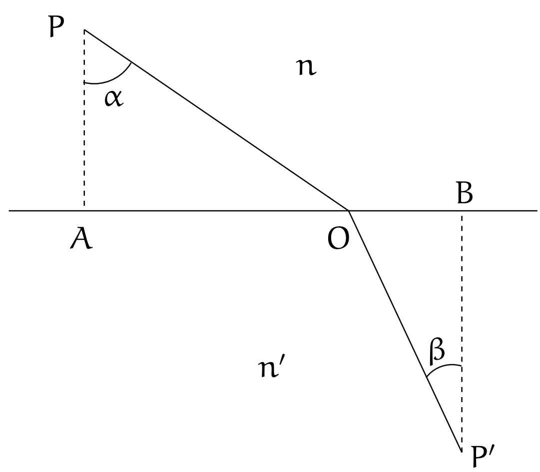
Fermat's Principle: Geometrical optics deals with the (artificial) concept of light rays. A light ray from point P to point P' is a P- and P'-containing path s that is always perpendicular to the successive wavefronts of the light as it propagates from P and P'. The optical length, which is the cumulative phase, is then equal to $\int_P^{P'} n\, ds$, where n is the index of refraction along the path. Being perpendicular to the two neighboring wavefronts, ds is the shortest distance between them. Therefore, if l is any other path connecting points P and P', it must be that

$$\int_P^{P'} n\, ds \leqslant \int_P^{P'} n\, dl \ .$$

This is the same as stating that the optical length $\int_P^{P'} n\, ds$ from points P to P' is a stationary one. This is called Fermat's principle. To find the actual ray path, we start by considering an arbitrary path from P to P', vary it while holding the two ends fixed, and set the variation $\delta\left(\int_P^{P'} n\, dl\right)$ to zero.



Snell's law: As a simple demonstration of Fermat's principle, let us derive from it Snell's law of refraction as light travels from a medium with an index of refraction n to one with an index of refraction n', as follows:



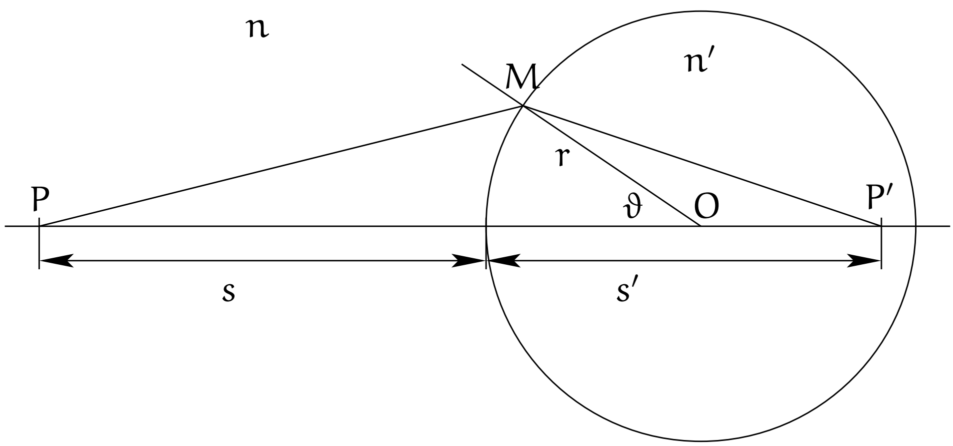
$$\int_P^{P'} n\,dl = n\overline{PO} + n'\overline{OP'} = n\sqrt{\overline{PA}^2 + \overline{AO}^2} + n'\sqrt{\overline{P'B}^2 + (\overline{AB} - \overline{AO})^2}\,.$$

Since the only variable in the above expression is $\overline{AO}$, differentiation with respect to $\overline{AO}$ gives

$$n\frac{2\overline{AO}}{2\sqrt{\overline{PA}^2 + \overline{AO}^2}} + n'\frac{-2(\overline{AB} - \overline{AO})}{2\sqrt{\overline{P'B}^2 + (\overline{AB} - \overline{AO})^2}}\,.$$

Setting it equal to zero results in Snell's law: $n \sin\alpha = n' \sin\beta$.

Paraxial imaging of a point by a refracting sphere:



Consider a sphere with an index of refraction n' surrounded by a medium with an index of refraction n. Light from point P is imaged at point P'. The optical length from P to P' is

$$L = n\overline{PM} + n'\overline{MP'}\,,$$

where

$$\overline{PM} = \sqrt{r^2 + (s+r)^2 - 2r(s+r)\cos\vartheta}$$

and
$$\overline{MP'} = \sqrt{r^2 + (s' - r)^2 + 2r(s' - r)\cos\vartheta} \, .$$

In order for P' to be a stigmatic image of P, all optical paths from P to P' must have the same optical length. That is, it must be that

$$\frac{dL}{d\vartheta} = \left(n\frac{s + r}{\overline{PM}} - n'\frac{s' - r}{\overline{MP'}} \right) r \sin\vartheta = 0 \, .$$

This leads to
$$n\frac{s + r}{\overline{PM}} = n'\frac{s' - r}{\overline{MP'}} \, .$$

Without approximation, this is equivalent to stating Snell's law at M. However, if M is close enough to axis $\overline{PP'}$, then we can let $\cos\vartheta \approx 1$ and the above condition becomes ϑ independent. We then have

$$\frac{n(s + r)}{s} = \frac{n'(s' - r)}{s'},$$

or

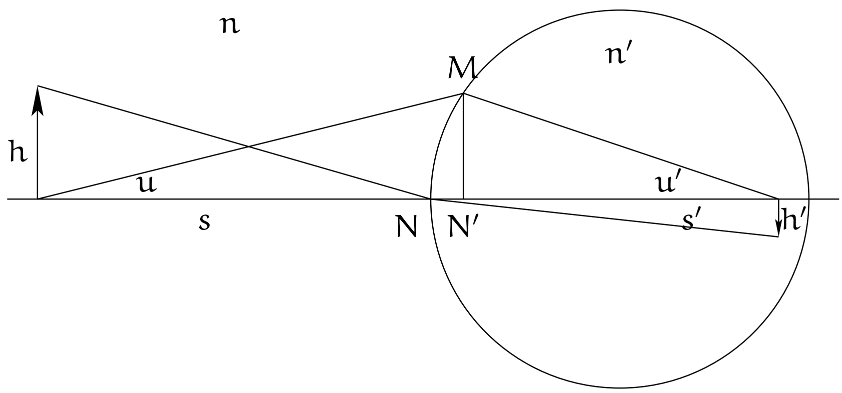
$$\frac{n}{s} + \frac{n'}{s'} = \frac{n' - n}{r} \, .$$

If s is situated at infinity, then $s' = \frac{n'r}{n'-n} = f'$, where f' is called the back focal length. If s' is at infinity (within an infinitely large refracting sphere), then $s = \frac{nr}{n'-n} = f$, where f is called the front focal length. In terms of f and f', the above equation becomes

$$\frac{f}{s} + \frac{f'}{s'} = 1 \, .$$

This is called Gauß' equation. If we let $z = s - f$ and $z' = s' - f'$, we then have, from Gauß' equation, $zz' = ff'$, which is called Newton's equation.

Lagrange–Helmholtz invariant:



We now consider an object of finite size h and its image h'. If the tips of h and h' lie close to the optical axis, i.e., $h \ll s$ and $h' \ll s'$, our previous analysis applies to these tips as well. In fact, the whole lengths of h and h' are imaged point by point. We now apply Snell's law at point N and get

$$n\left(\frac{h}{s}\right) = n'\left(\frac{h'}{s'}\right) .$$

In paraxial imaging, N and N′ nearly coincide. Further, $\overline{MN'} = s \tan u = s' \tan u' = su = s'u'$. Replacing s in the above equation with $s'u'/u$ leads to the so-called Lagrange–Helmholtz invariant:

$$nuh = n'u'h' .$$

Thin-lens formula: If, immediately after entering the first spherical surface of radius r_1 and index n from air, light rays encounter a second spherical surface of radius r_2 and exit back to air, the imaging process can be regarded as two back-to-back imaging processes by a single spherical surface, with the image of the first process serving as the (imaginary) object of the second process. Then,

$$\frac{1}{s} + \frac{n}{s'} = \frac{n-1}{r_1} \quad \text{and} \quad \frac{1}{s'} + \frac{1}{s''} = \frac{1-n}{r_2} \; .$$

Combining, we get

$$\frac{1}{s} + \frac{1}{s''} = (n-1)\left(\frac{1}{r_1} - \frac{1}{r_2}\right) = \frac{1}{f} \; .$$

This is the thin-lens formula, and f is the (front or back) focal length.

General relationships between an object and its image: Practical imaging[1] in microscopy or microlithography goes beyond paraxial imaging to include wide angles. Hence we need to introduce the following more general approach. Let us start by assuming that the object lies in plane

$$Ax + By + Cz + D = 0$$

and its image in plane

$$A'x' + B'y' + C'z' + D' = 0 \; ,$$

with

$$\begin{aligned} x' &= f(x, y, z) \\ y' &= g(x, y, z) \\ z' &= h(x, y, z) \; . \end{aligned}$$

The reason that this is possible can be found in Born and Wolf's *Principles of Optics* (7th ed., Cambridge University Press, p. 159, 1999). The equation for the image plane can then be written as

$$\Psi(x, y, z) = A'f(x, y, z) + B'g(x, y, z) + C'h(x, y, z) + D' = 0 \; .$$

[1] S. Czapski and O. Eppenstein, *Grundzüge der Theorie der optischen Instrumente nach Abbe*, Leipzig, Verlag von Johann Ambrosius Barth (1904).

Since $\Psi(x, y, z)$ vanishes for *all* points in the object plane that satisfy $Ax + By + Cz + D = 0$, it must be of the form

$$\Psi(x, y, z) = (Ax + By + Cz + D)\Phi(x, y, z) .$$

We then have

$$A'\frac{f}{\Phi} + B'\frac{g}{\Phi} + C'\frac{h}{\Phi} + D'\frac{1}{\Phi} = Ax + By + Cz + D .$$

Since LHS = RHS, terms on the LHS have to be of the form

$$\begin{aligned}
\frac{f}{\Phi} &= a_1x + b_1y + c_1z + d_1 \\
\frac{g}{\Phi} &= a_2x + b_2y + c_2z + d_2 \\
\frac{h}{\Phi} &= a_3x + b_3y + c_3z + d_3 \\
\frac{1}{\Phi} &= ax + by + cz + d .
\end{aligned}$$

Hence,

$$\begin{aligned}
x' &= f = \frac{a_1x + b_1y + c_1z + d_1}{ax + by + cz + d} \\
y' &= g = \frac{a_2x + b_2y + c_2z + d_2}{ax + by + cz + d} \\
z' &= h = \frac{a_3x + b_3y + c_3z + d_3}{ax + by + cz + d} .
\end{aligned}$$

Above is the general relationship between (x', y', z') and (x, y, z).

In a centered optical system, without loss of generality, we can always let $x = x' = 0$. Further, if we let $y \to -y$, then $y' \to -y'$, and the value of z' should not change. All these restrictions mean that the above relationships have to assume the forms

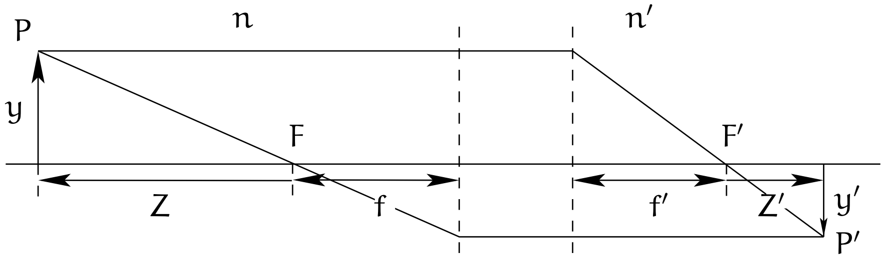
$$y' = \frac{b_2y}{cz + d} \text{ and } z' = \frac{c_3z + d_3}{cz + d} .$$

Now, if z' lies at infinity, then $z = -d/c$. If z is at (negative) infinity, then $z' = c_3/c$. These are the coordinates of foci (indicated by F and F′ in the diagram below) in the object and image spaces, respectively. If we define $Z = -d/c - z$ and $Z' = z' - c_3/c$, then we have

$$y' = \frac{b_2 y}{-cZ} \text{ and } Z' = \frac{-c_3 d + c d_3}{-c^2 Z} .$$

If we define $f = -b_2/c$ and $f' = \frac{-c_3 d + c d_3}{b_2 c}$, then we have the formula $ZZ' = ff'$. Also, $y'/y = f/Z = Z'/f'$. This also means that if $Z = f$, then $Z' = f'$. We draw planes perpendicular to the optical axis at $Z = -f$ and $Z' = -f'$ and call them the principal planes for the object and image spaces.

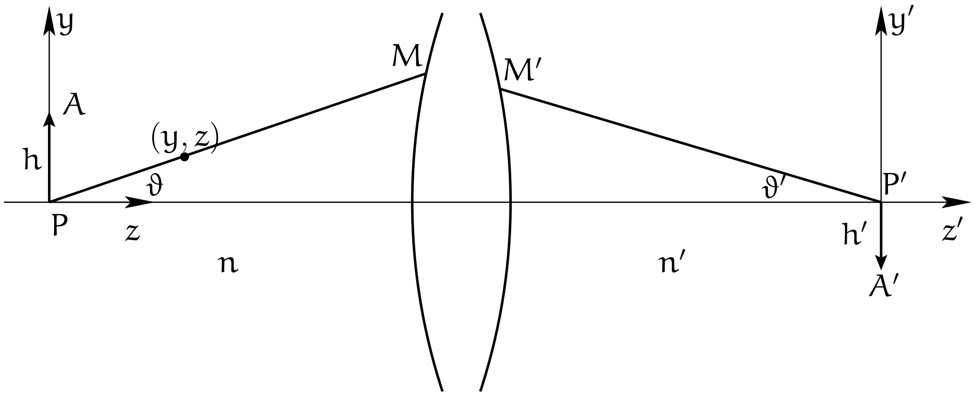
With the two principal planes and foci defined, the above relationships can be easily constructed geometrically as shown in the following diagram.



It is possible that Newton's formula in this context was first written down by Abbe, as stated in the text.

Sine condition: This is a requirement for obtaining aplanatic[2] images in any optical system. The only restriction is for the object and its image to lie very close to the optical axis. In paraxial optics, the sine condition is always satisfied. In fact, there it reduces to the Lagrange–Helmholtz invariant.

[2]Aplanatic means free of spherical aberration and coma. The Greek root of this word seems to mean not wandering around. That is, the image is point-like.



Here is a proof. Let $L(P; P')$ be the optical length from P to P' and $L(A; A')$ be the optical length from A to A'. For aplanatic imaging, the lengths of all optical paths from P to P' are equal, as are those from A to A'. L is therefore a function of only the locations of the end points. Let $F = L(A; A') - L(P; P')$. Assume that A lies very close to P, A' lies very close to P', and L is a continuous function (forget for a moment its physical meaning). Then, to the first order,

$$F = L(A; A') - L(P; P') = L(0, h; 0', h') - L(0, 0; 0', 0') = \left.\frac{\partial L}{\partial y}\right|_P h + \left.\frac{\partial L}{\partial y'}\right|_{P'} h' .$$

If the chosen optical path from P to P' includes $\overline{PM}$ and $\overline{M'P'}$, then

$$\left.\frac{\partial L}{\partial y}\right|_P = \frac{\partial}{\partial y}\left(n\sqrt{y^2 + z^2}\right) = n \sin\vartheta \, .$$

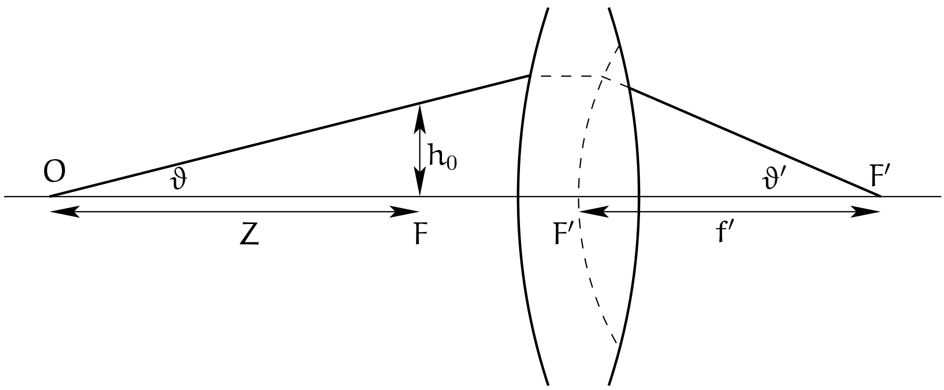
Likewise, $\left.\frac{\partial L}{\partial y'}\right|_{P'} = -n' \sin\vartheta'$. However, we may also choose the optical path to coincide with the optical axis; then, $\vartheta = \vartheta' = 0$, which implies $F = 0$. That is: if A lies very close to P and A' to P', the optical lengths from A to A' and from P to P' to the first order are the same! The sine condition then follows:

$$n \sin\vartheta \, h - n' \sin\vartheta' \, h' = 0 \quad \text{or} \quad n h \sin\vartheta = n' h' \sin\vartheta' \, .$$

Sine condition when object is at infinity: If the object lies at infinity, then we have $\sin\vartheta = h_0/Z$. Further, we have, in general, $h/h' = Z/f$. Therefore, we have

$$h_0 = \frac{n'f}{n}\sin\vartheta' = f'\sin\vartheta' ,$$

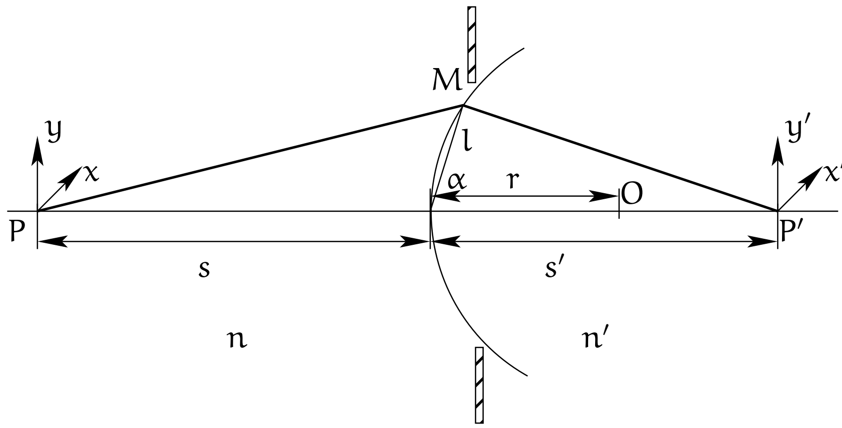
where $f' = n'f/n$ is the definition of the back focal length.



This is the form we see in Chapter 1. To satisfy the sine condition, principal planes are actually not planes but spherical surfaces.

A brief introduction to geometrical theory of aberrations: No one optical system is entirely free of aberrations. Those for which the sine condition is satisfied are well corrected for spherical and coma aberrations. Here we illustrate the essence of the aberration theory via a very simple example.[3] We consider again a single sphere with refractive index n', surrounded by a medium with refractive index n, as shown below. The object point is at P and its paraxial image at P'. We further assume that a circular aperture, centered on the optical axis, is placed over the sphere so that length l indicated in the figure has a maximum value.

[3]See also M. V. Klein, T. E. Furtak, *Optics*, 2nd ed., John Wiley & Sons, Sec. 4.3 (1986).

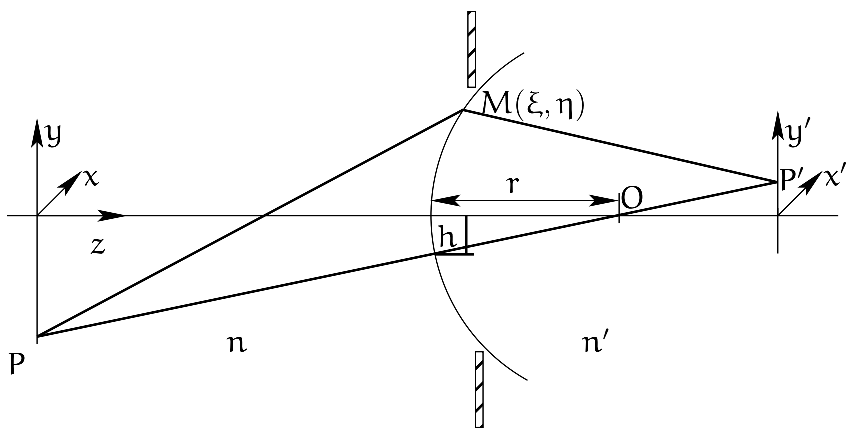


The optical length from P to P' via M is $n\overline{\mathrm{PM}} + n'\overline{\mathrm{MP}'}$. That of an axial ray is simply $ns + n's'$. Their difference is $W = n\overline{\mathrm{PM}} + n'\overline{\mathrm{MP}'} - (ns + n's')$. For small x, we have $\sqrt{1+x} \simeq 1 + x/2 + x^2/4$. In addition, we have $\cos\alpha = l/(2r)$. Therefore,

$$
\begin{aligned}
\overline{\mathrm{PM}} &= \sqrt{s^2 + l^2 + 2sl\,\cos\alpha} = s\sqrt{1 + \frac{l^2}{s^2} + \frac{l^2}{rs}} \\
&\approx s\left[1 + \frac{1}{2}\left(\frac{1}{s^2} + \frac{1}{rs}\right)l^2 - \frac{1}{4}\left(\frac{1}{s^2} + \frac{1}{rs}\right)^2 l^4\right] \\
\overline{\mathrm{PM}'} &= \sqrt{s'^2 + l^2 - 2s'l\,\cos\alpha} = s\sqrt{1 + \frac{l^2}{s^2} - \frac{l^2}{rs}} \\
&\approx s\left[1 + \frac{1}{2}\left(\frac{1}{s^2} - \frac{1}{rs}\right)l^2 - \frac{1}{4}\left(\frac{1}{s^2} - \frac{1}{rs}\right)^2 l^4\right]
\end{aligned}
$$

$$
\begin{aligned}
W &= n\overline{\mathrm{PM}} + n'\overline{\mathrm{MP}'} - (ns + n's') \\
&= \left(\frac{n}{s} + \frac{n'}{s'} + \frac{n - n'}{r}\right)\frac{l^2}{2} - \left[ns\left(\frac{1}{s^2} + \frac{1}{rs}\right)^2 + n's'\left(\frac{1}{s^2} - \frac{1}{rs}\right)^2\right]\frac{l^4}{4}.
\end{aligned}
$$

W being non-zero means that wavefronts converging on P' are not spherical, or all optical paths from P to P' would have the same optical length. In fact, W is simply the lumped-together wavefront error, called the wave aberration. As a good approximation, we may replace l with ρ, the distance from M to the optical axis. The quantity inside the parentheses in front of l^2 in the above equation is equal to zero because P' is the paraxial image point. More generally, the l^2 term is called defocus because it contributes to an additional circular curvature in the wavefront.[4] Excluding this term, the lowest order of aberrations is only the second term and can be written as $W = k\rho^4$.



For off-axis object points, without loss of generality, we can let P and P' be on the auxiliary axis lying in the yz-plane passing through the center of the sphere O, with h being the distance from the auxiliary axis to the optical axis in the plane of the aperture. We can then take

[4]A spherical refractive surface certainly does not result in spherical wavefronts. If it did, W would be zero. Spherical wavefronts can be approximated by parabolas, so in this case, the spherical wavefront at the refractive surface can be expressed as $z = ay^2 + s$. If the image point lies at a different location on the optical axis, the corresponding wavefront there will be $z = by^2 + s$. Hence, the difference in their optical lengths is approximately the difference of their z-coordinates, which is $(a - b)y^2 \propto \rho^2$. Therefore, such a term appearing in the expression of W is not considered an aberration; it simply means that the image is "defocused," i.e., it lies at a different location.

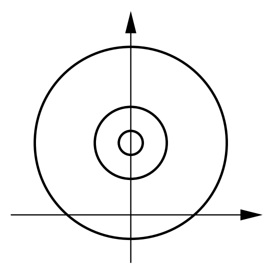
over our previous arguments if we replace ρ^4 with $[\xi^2 + (\eta + h)^2]^2 = [(\xi^2 + \eta^2) + 2\eta h + h^2]^2 = (\xi^2 + \eta^2)^2 + 4(\xi^2 + \eta^2)\eta h + 2(\xi^2 + 3\eta^2)h^2 + 4\eta h^3 + h^4$, for a general point $M(\xi, \eta)$ on the sphere. We may express W and name the various terms as follows:

$$W = \underbrace{A(\xi^2 + \eta^2)^2}_{\text{spherical}} + \underbrace{B(\xi^2 + \eta^2)\eta h)}_{\text{coma}} + \underbrace{C\eta^2 h^2}_{\text{astigmatism}} + \underbrace{D(\xi^2 + \eta^2)h^2}_{\text{field curvature}} + \underbrace{E\eta h^3}_{\text{distortion}} .$$

Since W is not zero, the wavefront is not spherical, and therefore the light ray from P via M does not go through P′. Instead, it will end up in the neighborhood of P′, the paraxial or Gaussian image point. Exactly where it lands in the Gaussian image plane is determined by the following set of formulae:[5]

$$x' - x^{*\prime} = \frac{s'}{n'}\frac{\partial W}{\partial \xi} \quad \text{and} \quad y' - y^{*\prime} = \frac{s'}{n'}\frac{\partial W}{\partial \eta} \,,$$

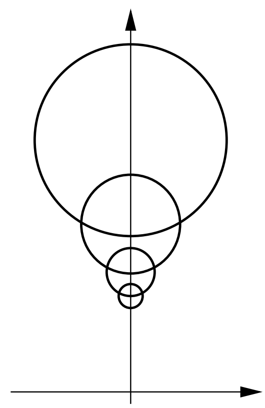
where $(x^{*\prime}, y^{*\prime})$ is the location of the paraxial image point. Therefore, in the presence of spherical aberration, the loci of the light rays in the $x'y'$-plane are $(x' - x^{*\prime})^2 + (y' - y^{*\prime})^2 = A'^2(\xi^2 + \eta^2)^2$.



The image of a point is not another point but a round spot of finite size centered on the paraxial image point.

[5]See M. Born and E. Wolf, *Principles of Optics*, 7th ed., Cambridge University Press, Sec. 5.1 (1999).

In the presence of coma, the loci are $(x' - x^{*\prime})^2 + [y' - y^{*\prime} - 2B'(\xi^2 + \eta^2)h]^2 = B'^2(\xi^2 + \eta^2)^2 h^2$.

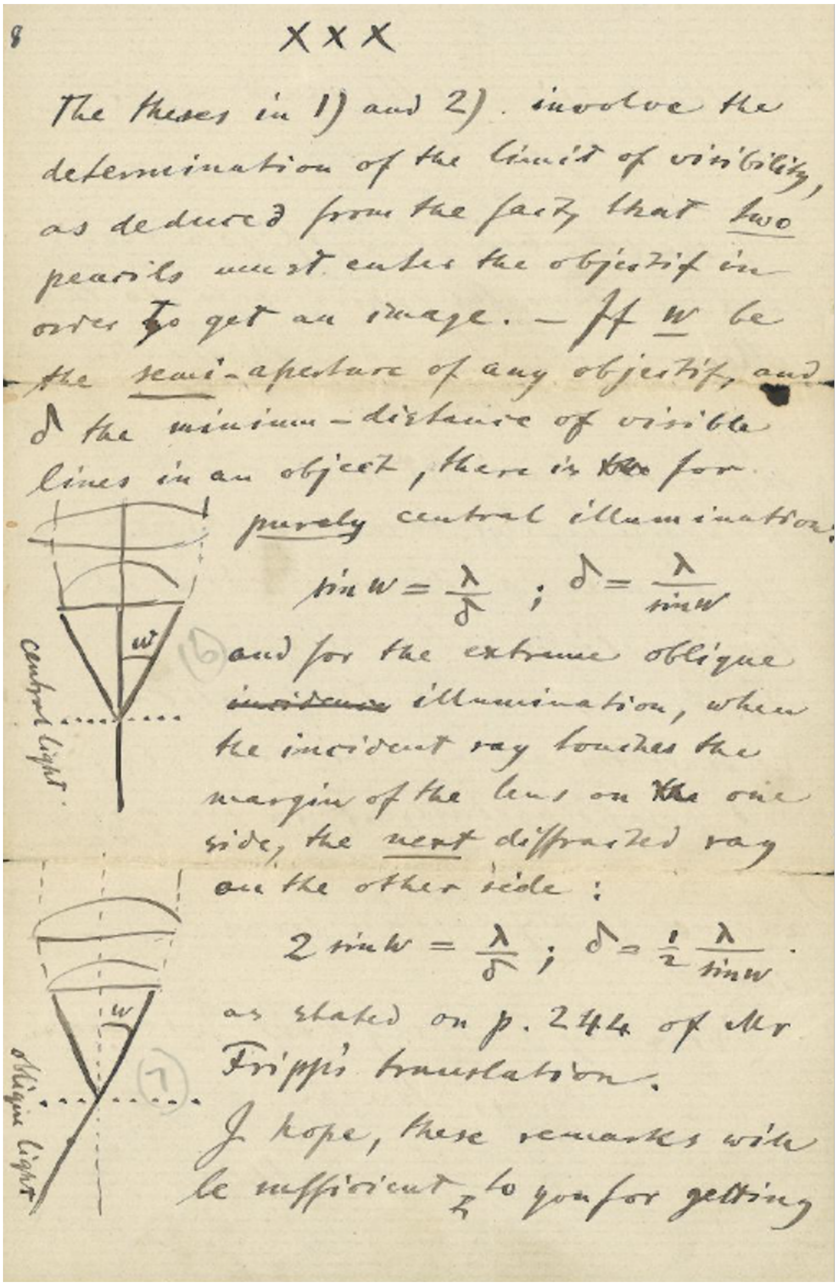


In this case, the image of point P looks like a comet-shaped spot with its tip at the paraxial image point. Hence the name coma.

On the 0.5 λ/NA resolution limit in the imaging of periodic patterns[1]

Ernst Abbe was the first person to state the limit of resolution of a periodic pattern in projection optical imaging. The formulae, $\delta = \frac{\lambda}{\sin w}$ for on-axis illumination and $\delta = \frac{1}{2}\frac{\lambda}{\sin w}$ for oblique illumination, where δ is the minimum pitch in an object, λ is the wavelength of the illuminating light, and w is the half-angle of the aperture, are stated in words in Abbe's 1873 article [1] and its English translation by H. E. Fripp [2]. And the question of how Abbe arrived at these formulae is answered in an 1876 letter to J. W. Stephenson, then treasurer of the Royal Microscopical Society (a facsimile of this letter and a transcript of which are reproduced at the end of this write-up). This time, Abbe wrote out the formulae explicitly, accompanied by two sketches showing the attainment of these resolutions by on-axis and oblique illuminations of the grating, respectively (see Fig. 1). In April of 1882, Abbe submitted a paper written in English to the Royal

[1]Adapted and abridged from A. Yen, "Rayleigh or Abbe? Origin and naming of the resolution formula of microlithography," *J. Micro/Nanolithogr. MEMS MOEMS.* **19**(4), 040501 (2020) https://doi.org/10.1117/1.JMM.19.4.040501, and A. Yen, "Clarifications on the 0.5 λ/NA resolution limit," *JM*^{3} **20**(1), 010501 (2021) https://doi.org/10.1117/1.JMM.20.1.010501.



8

XXX

The theses in 1) and 2). involve the determination of the limit of visibility, as deduced from the facts that two pencils must enter the objectif in order to get an image. — If w be the semi-aperture of any objectif, and δ the minimum-distance of visible lines in an object, there is ~~the~~ for purely central illumination:

$$\sin w = \frac{\lambda}{\delta}\ ; \quad \delta = \frac{\lambda}{\sin w}$$

and for the extreme oblique ~~incidence~~ illumination, when the incident ray touches the margin of the lens on ~~the~~ one side, the next diffracted ray on the other side:

$$2 \sin w = \frac{\lambda}{\delta}\ ; \quad \delta = \frac{1}{2}\frac{\lambda}{\sin w}.$$

as stated on p. 244 of Mr Fripp's translation.

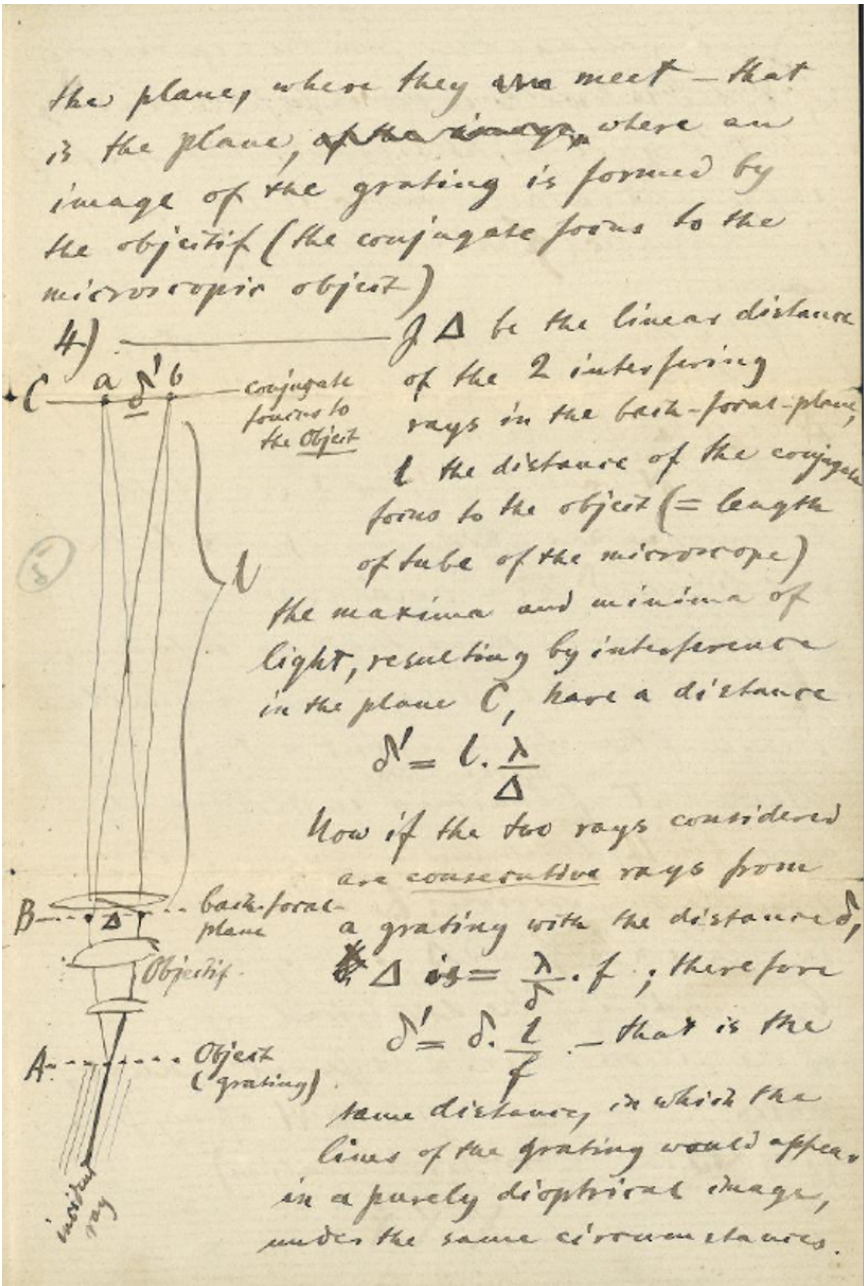
I hope, these remarks will be sufficient to you for getting

Figure 1. Page 13 of Abbe's 15 December 1876 letter to J. W. Stephenson, in which he illustrated the illumination method for obtaining the minimum resolution $\delta = \frac{1}{2}\frac{\lambda}{\sin w}$. (ZEISS Archives: Ernst Abbe Estate No. BACZ27167.)

Microscopical Society in which he further stated that, for periodic and regular features, "the minimum distance apart at which given elements can be delineated separately with the feature in question" was $\delta = \frac{1}{2}\frac{\lambda}{a}$, "where a denotes the numerical aperture and λ the wave-length light" [3].

To show how Abbe demonstrated that the two-beam case in Fig. 1 could actually lead to an image of the grating of pitch δ shown in the sketch (as dots), let us start from $2 \sin w = \frac{\lambda}{\delta}$, which can be obtained from the grating equation, as stated by Abbe on page 8 of the letter, for the oblique illumination case shown in Fig. 1. He then essentially stated that $\Delta = 2 \sin w \cdot f$, where Δ is the distance of separation of the two diffractions in the back focal plane of the lens with focal length f, by what he mentioned on page 9 of the letter as "a theorem enounced by me and by Mr. Helmholtz"; Abbe was certainly referring to the sine condition that he [1, 2] and Helmholtz independently discovered (see § 6 and Fig. 10 of this book; Helmholtz came to the same general form in [4]). By combining the grating equation and the sine condition, which must be fulfilled for aplanatic imaging, he obtained $\Delta = \frac{\lambda}{\delta} \cdot f$. Waves from these two diffraction spots then propagate and give rise to an interference pattern in the geometrical image plane (see Fig. 2). The pitch of such an interference pattern is $\frac{\lambda}{2 \sin \vartheta}$, where ϑ is half the angle formed by the two beams. In Abbe's case, $\sin \vartheta = \frac{\Delta/2}{l}$ because l, the distance from the back focal plane to the image plane, is large (as he assumed, on page 12 of the letter). Therefore, the pitch of the image is $\delta' = l \cdot \frac{\lambda}{\Delta}$, as he wrote on page 11 of the letter. Thus $\delta' = \delta \cdot \frac{l}{f}$, which means the pitch of the interference pattern in the image plane is of the correct magnification according to geometrical optics. It is therefore an image of the grating. And the minimum imageable pitch is $\delta = \frac{1}{2}\frac{\lambda}{\sin w}$. All this physics is discussed in more detail in § 25–27 of this book.

The $0.5\lambda/NA$ resolution limit is often called Rayleigh's criterion for resolution. It was derived explicitly in Rayleigh's 1879 article [5]



the planes where they ~~are~~ meet — that is the plane, ~~of the image~~ where an image of the grating is formed by the objectif (the conjugate focus to the microscopic object)

4) If Δ be the linear distance of the 2 interfering rays in the back-focal-plane, l the distance of the conjugate focus to the object (= length of tube of the microscope) the maxima and minima of light, resulting by interference in the plane C, have a distance

$$\delta' = l \cdot \frac{\lambda}{\Delta}$$

Now if the two rays considered are consecutive rays from a grating with the distance δ, Δ is $= \frac{\lambda}{\delta} \cdot f$; therefore

$$\delta' = \delta \cdot \frac{l}{f}$$

— that is the same distance, in which the lines of the grating would appear in a purely dioptrical image, under the same circumstances.

Figure 2. Page 11 of Abbe's 15 December 1876 letter to J. W. Stephenson, in which he illustrated and explained that plane C contains an image of the grating in plane A. (ZEISS Archives: Ernst Abbe Estate No. BACZ27167.)

(see Fig. 3), although the reasoning behind it — the minimum discernable separation of two neighboring lines is the distance between the principal maximum to the first minimum of the diffraction pattern in the focal plane — had already been given in two of his writings in 1874 ([6, 7]; in [6] Rayleigh stated the radius of the first dark ring of the point-spread function as $0.61\frac{f\lambda}{R}$, where R and f are the radius and the focal length of the lens, respectively). In the beginning part of the 1879 article, he put forward the formula obtained by G. B. Airy in 1834,

$$\vartheta = 1.2197\frac{\lambda}{2R}\,,$$

where ϑ is the angular radius of the bright central disk, λ represents the wavelength of the light, and $2R$ is the diameter of the circular aperture in front of a perfect lens, and went on to state that "in estimating theoretically the resolving-power of a telescope on a double star, we have to consider the illumination of the field due to the superposition of the two independent images. If the angular interval between the components of the star were equal to 2ϑ, the central disks would be just in contact. Under these conditions there can be no doubt that the star would appear to be fairly resolved, since the brightness of the external ring-systems is too small to produce any material confusion, unless indeed the components are of very unequal magnitude." He then went on to discuss two neighboring luminous lines and proposed his resolution criterion that is more lenient than the above requirement. Such luminous lines were generated in prism or grating spectroscopes by light sources with two spectral lines very close in wavelength. Rayleigh first stated, quoting Airy and Verdet, that the intensity (which he called brightness) of a luminous spectral line was proportional to the square of the sinc function,

$$\left(\frac{\sin\frac{\pi a\xi}{\lambda f}}{\frac{\pi a\xi}{\lambda f}}\right)^2 \equiv \mathrm{sinc}^2\left(\frac{a\xi}{\lambda f}\right)\,,$$

THE

LONDON, EDINBURGH, AND DUBLIN

PHILOSOPHICAL MAGAZINE

AND

JOURNAL OF SCIENCE.

[FIFTH SERIES.]

OCTOBER 1879.

XXXI. *Investigations in Optics, with special reference to the Spectroscope.* *By* LORD RAYLEIGH, *F.R.S.**

[Plate VII.]

§ 1. *Resolving, or Separating, Power of Optical Instruments.*

ACCORDING to the principles of common optics, there is no limit to resolving-power, nor any reason why an object, sufficiently well lighted, should be better seen with a large telescope than with a small one. In order to explain the peculiar advantage of large instruments, it is necessary to discard what may be looked upon as the fundamental principle of common optics, viz. the assumed infinitesimal character of the wave-length of light. It is probably for this reason that the subject of the present section is so little understood outside the circles of practical astronomers and mathematical physicists.

It is a simple consequence of Huyghens's principle, that the direction of a beam of limited width is to a certain extent indefinite. Consider the case of parallel light incident perpendicularly upon an infinite screen, in which is cut a circular aperture. According to the principle, the various points of the aperture may be regarded as secondary sources emitting synchronous vibrations. In the direction of original propagation the secondary vibrations are all in the same phase, and hence the intensity is as great as possible. In other direc-

* Communicated by the Author.

tions the intensity is less; but there will be no sensible discrepancy of phase, and therefore no sensible diminution of intensity, until the obliquity is such that the (greatest) projection of the diameter of the aperture upon the direction in question amounts to a sensible fraction of the wave-length of the light. So long as the extreme difference of phase is less than a quarter of a period, the resultant cannot differ much from the maximum; and thus there is little to choose between directions making with the principal direction less angles than that expressed in circular measure by dividing the quarter wave-length by the diameter of the aperture. Direct antagonism of phase commences when the projection amounts to half a wave-length. When the projection is twice as great, the phases range over a complete period, and it might be supposed at first sight that the secondary waves would neutralize one another. In consequence, however, of the preponderance of the middle parts of the aperture, complete neutralization does not occur until a higher obliquity is reached.

This indefiniteness of direction is sometimes said to be due to "diffraction" by the edge of the aperture—a mode of expression which I think misleading. From the point of view of the wave-theory, it is not the indefiniteness that requires explanation, but rather the smallness of its amount.

If the circular beam be received upon a perfect lens, an image is formed in the focal plane, in which *directions* are represented by *points*. The image accordingly consists of a central disk of light, surrounded by luminous rings of rapidly diminishing brightness. It was under this form that the problem was originally investigated by Airy*. The angular radius θ of the central disk is given by

$$\theta = 1{\cdot}2197\,\frac{\lambda}{2R}, \quad \dots\dots\dots \quad (1)$$

in which λ represents the wave-length of light, and $2R$ the (diameter of the) aperture.

In estimating theoretically the resolving-power of a telescope on a double star, we have to consider the illumination of the field due to the superposition of the two independent images. If the angular interval between the components of the star were equal to 2θ, the central disks would be just in contact. Under these conditions there can be no doubt that the star would appear to be fairly resolved, since the brightness of the external ring-systems is too small to produce any

* Camb. Phil. Trans. 1834.

Figure 3. Lord Rayleigh's 1879 article on the resolution of two neighboring features.

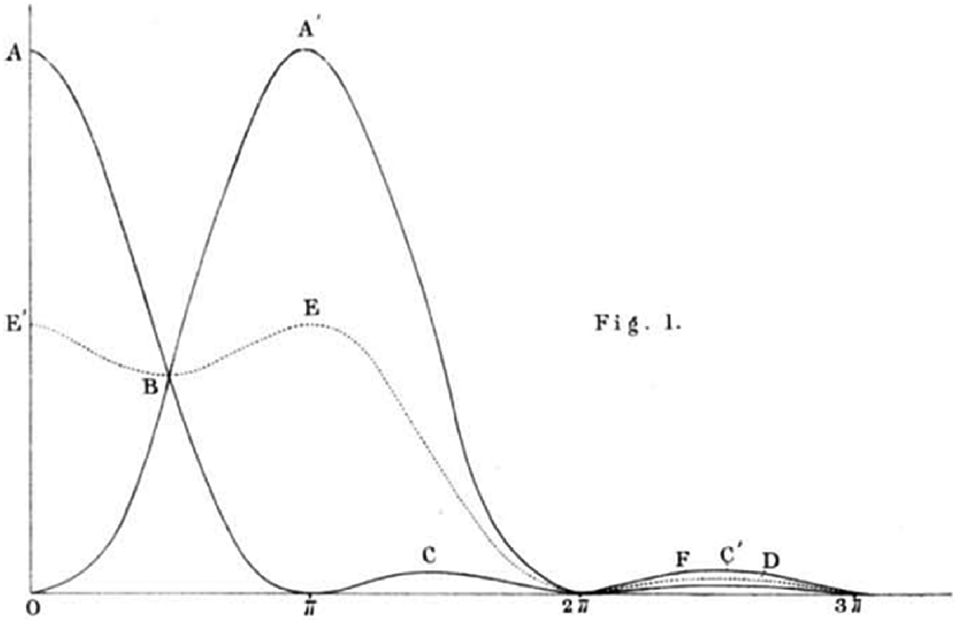


Figure 4. Rayleigh's plots in his article. $ABCD$ is $\left(\frac{\sin u}{u}\right)^2$; $OA'C'$ is $\left(\frac{\sin(u-\pi)}{u-\pi}\right)^2$; and $E'BEF$ is half of $\left[\left(\frac{\sin u}{u}\right)^2+\left(\frac{\sin(u-\pi)}{u-\pi}\right)^2\right]$.

where ξ is the horizontal axis, a is the horizontal dimension of the rectangular aperture (placed after the prism but before the focusing lens), and f is the focal length of the lens. He then tabulated the values of the above function and pronounced that if the two neighboring lines were so separated that the maximum intensity of one line fell onto the first minimum of that of the other line, then the two lines could be discerned because the combined brightness in the middle of the two peaks (which have the brightness of 1) dipped down to 0.8106 (see Fig. 4). Hence, the smallest discernable separation d of the two lines was

$$d = \frac{\lambda}{a/f} .$$

If we translate this to our language, a/f is twice the numerical aperture NA of the lens (in air). Hence, the Rayleigh criterion simply implies that the discernable separation of two neighboring lines is $0.5\lambda/\text{NA}$. In fact, the same criterion can also be applied to the Airy patterns. If we allow the maximum of the first Airy pattern to coincide with the edge of the bright central disk of the second pattern, then the light intensity at the saddle point in the middle of the two intensity peaks is 0.7348 times the intensity at either peak, and the minimum discernable distance in this case is $0.61\lambda/\text{NA}$, as has been stated in many textbooks.

What Rayleigh stated in his article can be easily explained. Light disturbance in the image plane, produced by a distant star, is simply the point-spread function of the optical system (of the telescope), since the distant star can be regarded as a δ-function object. In a number of textbooks [8], one can find that the light intensity of the Airy disk, which is the square of the point-spread function (the Fraunhofer diffraction of a circular aperture), is proportional to

$$\left[\frac{J_1\left(\frac{2\pi}{\lambda}\text{NA}\cdot r\right)}{\frac{2\pi}{\lambda}\text{NA}\cdot r}\right]^2 ,$$

where J_1 is the Bessel function of the first kind, order 1, whose first zero occurs at the argument of 1.22π, and NA, the numerical aperture of the optical system, equals Rayleigh's R/d, where R is the radius of the aperture and here d is the distance from the aperture to the image plane. Setting the argument of the above Bessel function to 1.22π, the diameter of the Airy disk is then

$$2r = 1.22\frac{\lambda}{\text{NA}}$$

or, as Rayleigh stated, the angular radius is

$$\vartheta = \frac{r}{d} = 1.22\frac{\lambda}{2R}\,.$$

Now let the object be an infinitely thin and infinitely long vertical line, represented by $\delta(\xi) \cdot 1$. Its image is simply its convolution with the point-spread function $U(\xi, \eta)$, i.e. $[\delta(\xi) \cdot 1] \otimes U(\xi, \eta)$. The result, omitting again the pre-factor, is $\mathrm{sinc}\left(\frac{a\xi}{\lambda f}\right)$, as stated by Rayleigh. See also § 20 of this book.

Rayleigh understood, however, that what he put forward was not the absolute resolution limit. He stated in the article that "this rule is convenient on account of its simplicity." Born and Wolf also stated in their book [9] that "no special physical significance is to be attached to the Rayleigh criterion, and from time to time other criteria of resolution have been proposed." Rayleigh dealt with incoherent illumination. Under incoherent illumination, light intensity of the final image is the sum of the intensities produced by each individual point or line. For two neighboring lines, we may argue that their minimum distance can even be 0.45 λ/NA, as the intensity in the mid-point between the two peaks dips down to 0.954 times the intensity at either peak. To be extreme, one can even argue that a one-percent intensity dip at the mid-point should be considered discernment of the two features. In fact, more than a century ago, C. M. Sparrow stated that he was able to discern the two lines, by direct vision and in positive and negative film, all the way down to where the second derivative of the combined intensity curve at point B in Fig. 4 reached zero, meaning no intensity dip at all, and 0.83 times the Rayleigh separation [10]. Therefore, a criterion based on a two-point or two-line structure is ambiguous. Also, the separation of the two peaks in the image intensity for the 0.45 λ/NA case is not 0.45 λ/NA but 0.365 λ/NA, resulting in a condition called "pitch walking" in microlithography. The root cause of all this ambiguity lies in the continuous nature of the spatial frequencies of a two-point (or two-line) object not having sharp peaks (δ-functions or near δ-functions in the spatial frequency domain) associated with periodic or regular structures that are either passed or eliminated without ambiguity by the pupil aperture.[2] Abbe's resolution criteria deal exactly with such periodic or regular structures and hence his resolution limit is black and white.

Finally, we want to mention Helmholtz's contribution to the expression of the resolution limit. Hermann von Helmholtz wrote down the expression $\varepsilon = \frac{\lambda}{2\sin\alpha}$ explicitly in a lecture on the resolution limit of the microscope to the Royal Prussian Academy of Sciences on 20 October 1873 [12], and explained it in detail in [4]. Here, ε is the smallest discernable distance in an object, λ is the wavelength of light in the medium, and α is the angle formed by the outermost rays emanating from the axial point of an object and going through the system, and the optical axis. Hence, the expression is exactly the same as Abbe's. Like Rayleigh, Helmholtz derived this criterion based on two neighboring bright lines and hence suffers from the same ambiguity (mentioned by Helmholtz himself in [4]), especially as he derived the expression in the context of microscopy. As to whether Abbe or Helmholtz came up with the expression first, Lummer and Reiche (see the footnote after Eq. 81) stated that Helmholtz came up with the expression "almost at the same time, though in another way." This claim is substantiated by a postscript at the end of Helmholtz's 1874 article ([4], p. 584), in which he described seeing Abbe's 1873 publication and noticing a large overlap of subjects discussed in the two works, at the last moment before dispatching his own manuscript. Helmholtz wrote, "The special festive occasion for which this volume[3] of the annuals is published forbids

[2]The fact that the illumination is incoherent does not affect this conclusion. The intensity is the same as if the two slits are coherently illuminated but one of them has a phase shift of $\pi/2$ [11]. So we can apply the spectral analysis for coherent imaging.

[3]The jubilee volume of the journal in which Helmholtz's 1874 article appears.

me to withhold or completely withdraw my work. Because it contains the proofs, which Mr. Abbe still withheld, of the theorems needed by both of us and a few simple attempts on the explanation of the theoretical considerations, may its publication be pardoned from the scientific standpoint." ([4], p. 584). What modesty from a great scientist!

References

[1] E. Abbe, "Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung," *Arch. Mikrosk. Anat.* 9, 413–468 (1873).

[2] E. Abbe, "A contribution to the theory of the microscope, and the nature of microscopic vision," *Proceedings of the Bristol Naturalists' Society,* New Series, Vol. 1 Part II, 200–261 (1874). This is a translation of [1] by H. E. Fripp. Description of the relevant formulae is found in the last paragraph of p. 244.

[3] E. Abbe, "The Relation of aperture and power in the microscope (continued)," *Journal of Royal Microscopical Society* Ser. II–Vol. II, Part 2, pp. 460–473 (1882).

[4] H. Helmholtz, "Die theoretische Grenze für die Leistungsfähigkeit der Mikroskope," *Pogg. Ann.,* Jubelband (jubilee volume), pp. 557–584 (1874).

[5] Lord Rayleigh, "Investigations in optics, with special reference to the spectroscope," *Phil. Mag.* Ser. 5, Vol. 8, 261–274 (1879).

[6] Lord Rayleigh, "On the manufacture and theory of diffraction-gratings," *Phil. Mag.* Ser. 4, Vol. 47, pp. 81–93 (1874).

[7] Lord Rayleigh, "On the manufacture and theory of diffraction-gratings," *Phil. Mag.* Ser. 4, Vol. 47, pp. 193–205 (1874).

[8] See, e.g., J. W. Goodman, *Introduction to Fourier Optics,* 4th ed., W. H. Freeman and Company, New York, pp. 91–93 (2017).

[9] M. Born and E. Wolf, *Principles of Optics,* 7th ed., Cambridge University Press, Cambridge, footnote on p. 371 (1999).

[10] C. M. Sparrow, "On spectroscopic resolving power," *The Astrophysical Journal,* Vol. XLIV, pp. 76–86 (1916).

[11] J. W. Goodman, ibid., pp. 216–219.

[12] H. Helmholtz gave a lecture on the resolution limit of the microscope on 20 October 1873, reported in *Monatsberichte der Königlich Preußischen Akademie der Wissenschaften zu Berlin*, 625–626 (1873).

Abbe's 15 December 1876 Letter to J. W. Stephenson[4]

Jena, Dec. 15, 76

Dear Sir

You are too obliging a correspondent, and I am rather ashamed of your praise. For I myself take pleasure in writing about my microscopical interests to a gentleman, whom I know to have a perfect understanding of those things by his mathematical training—since microscopists, in general, have no, or little, understanding.

I add a few remarks, which, I hope, will remove a difficulty, you have found in my explanations, perhaps.

In some passages of my letter I distinguish: the pencil of direct rays and the diffraction-pencils. But this distinction does not mean any principal difference in the function, or action, of these rays. From a general point of view, the pencil of direct rays, transmitted by a lined, or marked, object to the microscope, is one among the diffraction pencils; it is different from the others only by its greater intensity of light; but in its action, in the formation of the images of structured objects, it is quite on the same range with the others. (The outlines of any object, it is true, are delineated by the direct rays alone, in bright field; by diffracted rays only in dark field.)

[4] Transcript of the letter provided by ZEISS Archives (Ernst Abbe Estate No. BACZ 27167), with additional editing.

Therefore in the production of the image of Pleurosigma, with oblique light, you have at least three active pencils, not two; the direct pencil, a [*sketch*], being the third. The three sets of lines arise from the 3 combinations: $\widehat{a,b}$; $\widehat{a,c}$, and $\widehat{b,c}$—every pair producing <u>one</u> set, rectangular to the connecting line of corresponding points. The image of Pl. angul. with straight light, by an immersion lens, is formed by <u>7</u> pencils.

As to Pleuros., whether there are 3 sets of lines or two, the following will state my opinion.

The diffraction-phenomena, produced by 3 sets, crossing at 60°, or by 2 sets, or by <u>isolated</u> apertures of <u>any</u> form, arranged: [*sketch*], are <u>not</u> different one from another in the <u>first</u> row of spectra around the direct pencil. Those diffraction-phenomena are different only in the more distant spectra, which differ in position and relative brightness.

Now, with Pleuros. those more distant spectra are <u>not</u> visible by any objective, from the great angular dispersion of the diffracted rays off the incident ray—owing to the smallness of the structure. (Those more distant pencils could be visible only in a medium of considerably higher refractive index, than air, or water has.)

The microscopic image, depending on the distance and relative position of the diffraction pencils, which are <u>effective</u> in the microscope, must be the same for all the different structures named above, as far as the first row of spectra is admitted only; what you see in that case, either in Pleurosigma or on the [*sketch*] gratings, is the <u>typical</u> image belonging to the inner part of a diffraction-phenomenon of this kind: [*sketch*]

Therefore <u>nothing</u> can be inferred from the microscopic image of Pl. ang. relating to the <u>detail</u> of the structure.

There may be 2 sets of lines, or three sets, or isolated apertures in the scale etc.—in every case the known images will result.

That rhombical apertures, as on the [*sketch*] grating look as hexagonal fields, is not surprising, if you consider my theoretical explanation: that the microscopic images result from the interference of the different diffraction pencils, which enter the microscope (the direct pencil there included). From this point of view, the real forms of the structure have no direct relation to the image—only an indirect relation, by determining the diffraction-phenomenon partially.

The microscopic image, which any structure will show, is the more similar to the structure, the more all the diffracted light is admitted to the microscope. The interference of *all* the diffraction-pencils, which come from the object, produces a copy of the real structure, alike to a dioptrical image. This is the key-stone of my theory. From this is to be inferred: The smaller a structure is (the more dispersed therefore the diffraction-pencils) the less similar the microscopic image will be, for any aperture of the objective applied; and those objects as the fine diatoms, give, with any lens, only typical images (not copies of the real forms), because any lens will admit only a few pencils of the diffraction-phenomenon.

I am sorry not to have in my possession one single copy neither German nor English of my paper, in which I have stated more precisely—though very briefly—the consequences of this theory touching the interpretation of microscopical images. Perhaps it will be possible to you to lend the 1st vol. of the "Bristol Naturalists' Society's Proceedings," Part 2, in which you will find Mr. Fripp's translation in extenso; but you should observe the table of "Errata," which Mr. Fripp has given some time afterwards, because many sentences in the translation are quite unintelligible by errata. The abstract, which appeared in the M. M. Journal, is useless.

I add a few remarks about the mathematical side of the theory, of which I have stated only the point of view in my paper. I think you will quite understand the principle of my mathematical deduction by considering the simplest case,—one set of equidistant lines—and observing the following notices:

1) be δ the distance of the lines in a microscopical object; λ the wavelength for one definite colour; [*sketch*] be u_0 the angle of incidence, in which a ray meets the grating; $\ldots u_{-2}$, u_{-1}, u_0, u_{+1}, u_{+2} ... the angles formed by: the several diffracted rays $-2, -1, 0, +1, +2$, the direct ray (0) included, the perpendicular line taken as zero-direction; there is

$$\sin u_{+2} - \sin u_{+1} = \sin u_{+1} - \sin u_0 = \sin u_0 - \sin u_{-1} = \cdots = \frac{\lambda}{\delta}$$

in which formula the case of normal incidence is included, of course. If now the diffracted rays enter a microscope, the sines of the angles of any two consecutive rays with the axis of the microscope have the same difference $= \frac{\lambda}{\delta}$.

2) [*sketch*] If an objective is focussed to the grating, and if this objective is perfectly* aplanatic for its focal point any ray forming an angle u with the axis below the objective, is refracted in such a way, that it will pass the upper (the back) focal-plane of the lens in a linear distance from the axis

$$\Delta = f \cdot \sin u$$

if f is the focal-length of the objective (by a theorem enounced by me and by Mr. Helmholtz).

From this theorem 2), in connection with 1) is to be inferred: the linear distance of the diffraction-spectra, which appear in the back-focal-plane of the objective is always $= \frac{\lambda}{\delta} \cdot f$, if corresponding points in every two consecutive spectra are considered—independent of the inclination of the incident rays to the grating. If you go from central light to oblique light, all the spectra

*"Perfectly aplanatic" means: without spherical aberration not only for one point on the axis, but for the points aside the axis too.

move within the back-plane of the system, without changing their relative position.

3) All the rays, which result by diffraction, from one incident ray have their oscillations in equal phase, if points are compared on these rays which are situated in the back-focal-plane, where the spectra are formed as images of the illuminating object; all those rays therefore must interfere within the plane, where they meet—that is the plane, where an image of the grating is formed by the objective (the conjugate focus of the microscopic object).

4) [*sketch*] If Δ be the linear distance of the 2 interfering rays in the back-focal-plane, l the distance of the conjugate focus to the object (= length of tube of the microscope), the maxima and minima of light, resulting by interference in the plane C, have a distance

$$\delta' = l \cdot \frac{\lambda}{\Delta}.$$

Now if the two rays considered are consecutive rays from a grating with the distance δ, Δ is $\frac{\lambda}{\delta} \cdot f$; therefore $\delta' = \delta \cdot \frac{l}{f}$—that is the same distance, in which the lines of the grating would appear in a purely dioptrical image, under the same circumstances.

λ being eliminated from the expression of δ', the intervals in the interference-image must be equal for the different colours; this image must be achromatic, if the objective is achromatic (constant for different colours).

If the two rays considered were not consecutive (as in the experiment with the 3-holes-stop) Δ would have double (or triple...) the value taken above; therefore δ' would be $\frac{1}{2}$, or $\frac{1}{3}$... of the distance, which corresponds to the real distance δ in a similar image.

This reasoning shows, that the interference of the diffracted rays can give a similar image of the structure, but not must.

The want of mathematical exactness in the deduction above (in 4), arising from the supposition: l infinitely great in relation to f and Δ, is perfectly removed by considering the dioptrical effect of the microscope in a different manner, which I have stated in No. VI of my paper (page 213 in Mr. Fripp's translation).

The theses in 1) and 2) involve the determination of the limit of visibility, as deduced from the fact, that two pencils must enter the objective in order to get an image. If w be the semi-aperture of any objective, and δ the minimum distance of visible lines in an object, there is for purely central illumination: [*sketch*]

$$\sin w = \frac{\lambda}{\delta} \quad ; \quad \delta = \frac{\lambda}{\sin w}$$

and for the extreme oblique illumination, where the incident ray touches the margin of the lens on one side, the next diffracted ray on the other side [*sketch*]:

$$2 \sin w = \frac{\lambda}{\delta} \quad ; \quad \delta = \frac{1}{2} \frac{\lambda}{\sin w}$$

as stated on p. 244 of Mr. Fripp's translation.

I hope these remarks will be sufficient to you for getting a clear notion of the mathematical principles of the theory.

I shall be very glad, if you should like to show the experiments to the Microscopical Society—especially if you should think it convenient to produce them not as paradox phenomena, but rather as phenomena illustrating a distinct idea of the functions of the microscope. For there is no want of optical curiosities among microscopists; and I take no interest in bringing forth more of that. Please, make any use of my explanations, you like.

With my best regards I remain
Yours truly
E. Abbe

4

Jena, Dec. 15, 76.

Dear Sir.

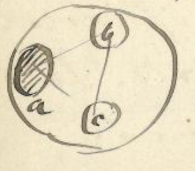
You are too obliging a correspondent, and I am rather ashamed of your praise. For I myself take pleasure in writing about my microscopical interests to a gentleman, whom I know to have a perfect understanding of those things ~~by~~ his mathematical training — since microscopists, in general, have no, or little, understanding.

I add a few remarks, which, I hope, will remove a difficulty, you have found in my explanation, perhaps.

In some passages of my letter I distinguish: the pencil of direct rays and the diffraction-pencils. But this distinction does <u>not</u> mean

any* principal difference in the
function, or action, of these rays.
From a general point of view, the
pencil of direct rays, transmitted
by a lined, or marked object to the
microscope, is <u>one</u> among the diffraction
pencils; it is different from the others
only by its greater intensity of light;
but in its action, in the formation
of the images of structured objects, it is quite on the same
range with the others. (The <u>outlines</u>
of any object, it is true, are delineated
by the direct rays alone, in bright
field; by diffracted rays only in
dark field.)

Therefore in the production of
the image of Pleurosigma, with ob-
lique light, you have at least <u>three</u> active
pencils, not two; the direct pencil, <u>a</u>,



being the third. — The
three sets of lines arise

1

from the 3 combinations: a, b; a, c, and b, c — every pair producing one set, rectangular to the connecting line of corresponding points. The image of Pl. angul. with straight light, by an immersion lens, is formed by 7 pencils.

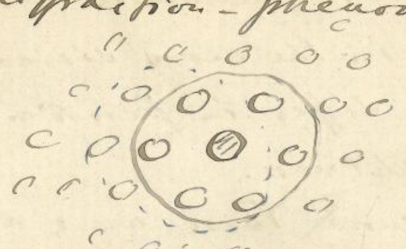
As to Pleuros., whether there are 3 sets of lines or two, the following will state my opinion.

The diffraction-phenomena, produced by 3 sets, crossing at 60°, or by 2 sets, or by isolated apertures of any form, arranged: ⁘, are not different from one another in the first row of spectra around the direct pencil. Those diffraction-phenomena are different only in the more distant spectra, which differ in position and ~~in~~ relative brightness.

Now, with Pleuros. those more distant spectra ~~are~~ are not visible by any objectif, from the great angular dispersion of the diffracted rays off from the incident ray – owing to the smallness of the structure. (Those more distant pencils could be visible only in a medium of considerably higher refractive index, than air, or water has)

The microscopic image, depending on the distance and relative position of the diffraction pencils, which are *effective* in the microscope, must be the same for all the different structures, named above, as far as the first row of spectra ~~above only~~ is admitted only; what you see in that case, either on Pleurosigma or on the # gratings, is the *typical* image belonging to the inner part of a diffraction-phenomenon of this kind:

2



Therefore *nothing* can be ~~concluded~~ inferred from the microscopic image of Pl. ang. relating to the *detail* of the structure.

X

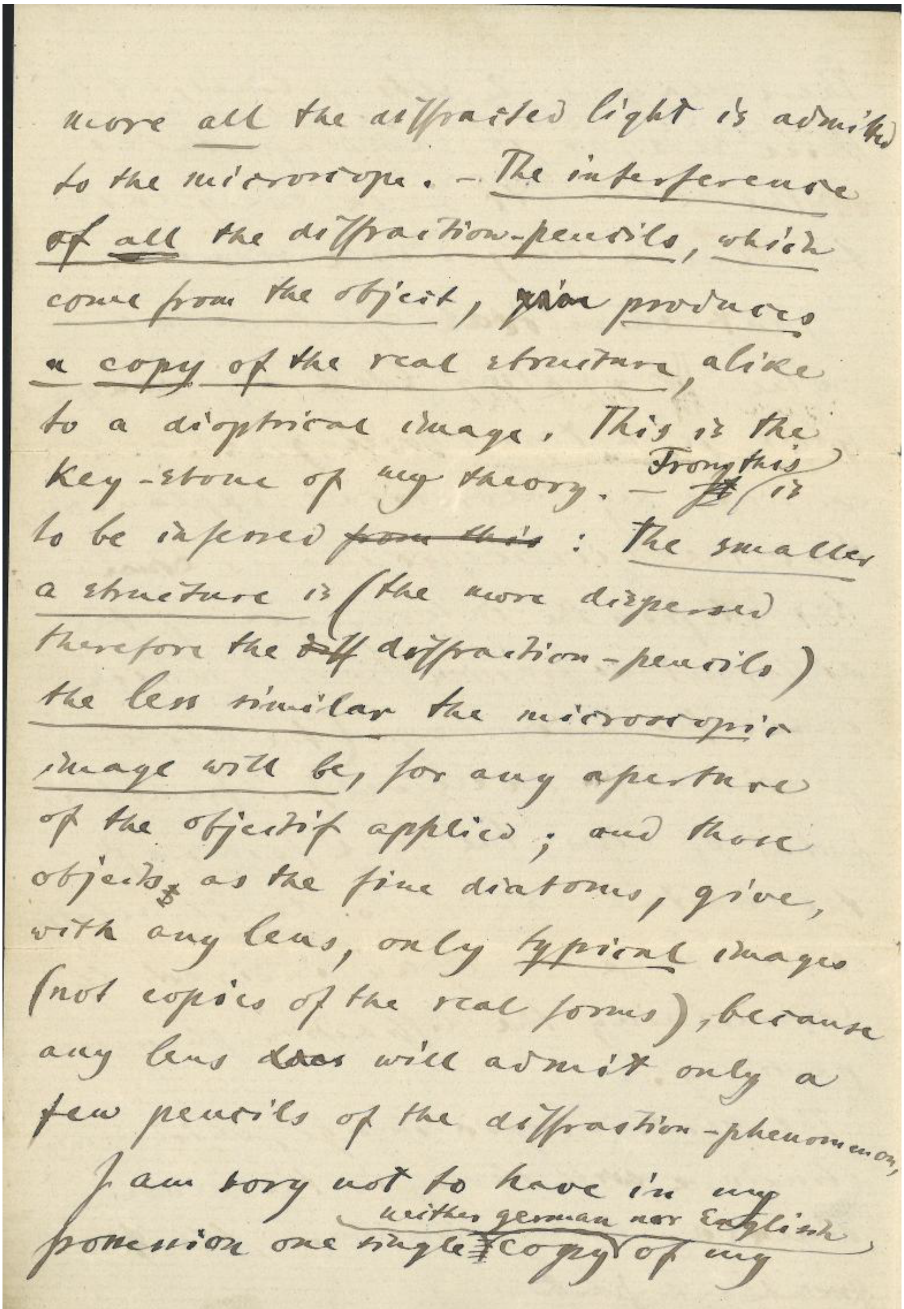
8

X

There may be 2 sets of lines, or three sets, or isolated apertures in the scale etc. — in every case the known images will result.

That rhomboidal apertures, as on the # -grating look as hexagonal fields, is not surprising, if you consider my theoretical explanation: that the microscopic images ~~are~~ result from the <u>interference</u> of the different diffraction pencils, which enter the microscope (the direct pencil there included). From this point of view, the <u>real</u> forms of the structure have no <u>direct</u> relation to the image — only an <u>indirect</u> relation, by determining the diffraction-phenomenon partially.

The microscopic image, which any structure will show, is the more <u>similar</u> to the structure, the ~~[illegible]~~ ~~greater a part~~

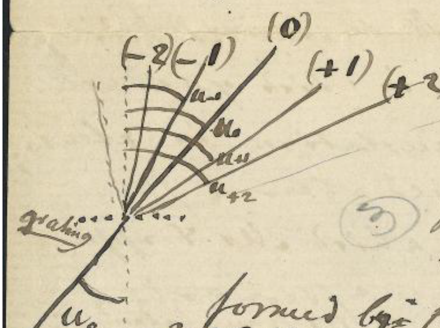


more all the diffracted light is admitted
to the microscope. — The interference
of all the diffraction-pencils, which
come from the object, ~~gives~~ produces
a copy of the real structure, alike
to a dioptrical image. This is the
key-stone of my theory. — From this ~~it~~ is
to be inferred ~~from this~~: The smaller
a structure is (the more dispersed
therefore the ~~diff~~ diffraction-pencils)
the less similar the microscopic
image will be, for any aperture
of the objectif applied; and those
objects, as the fine diatoms, give,
with any lens, only typical images
(not copies of the real forms), because
any lens ~~does~~ will admit only a
few pencils of the diffraction-phenomenon.
I am sorry not to have in my
possession one single neither german nor English copy of my

paper, in which I have stated more precisely — though very briefly — the consequences of this theory touching the interpretation of microscopical images. — Perhaps it will be possible to you to lend the I^t vol. of the Bristol ~~Microscop.~~ Naturalists' Society's "Proceedings", Part 2, in which you will find Mr Fripp's translation in extenso; but you ~~must~~ should observe the table of "Errata", which Mr Fripp has given some time afterwards, because many sentences (in the translation) are quite unintelligible by errata. — The abstract which appeared in the M. M. Journal, is useless.

I add a few remarks about the mathematical side of the theory, of which I have stated only the point of view in my paper. — I think you will quite understand the principle of my mathematical deduction by considering the simplest case, — one set of aequidistant lines — and observing the following notices:

1) – be δ the distance of the lines in a microscopical object; λ the wave length for one definite colour; be u_0 the angle of incidence, in which a ray meets the grating; ... u_{-2}, u_{-1}, u_0, u_{+1}, u_{+2} ... the angles



3

formed by $-2, -1, 0, +1, +2$ ... the several diffracted rays, the direct ray (0) included, ~~those~~ ~~[illegible]~~ the perpendicular line taken as zero-direction; there is

$$\sin u_{+2} - \sin u_{+1} = \sin u_{+1} - \sin u_0 = \sin u_0 - \sin u_{-1} = \ldots\ldots = \frac{\lambda}{\delta}.$$

in which formula the case of normal incidence is included, of course. — ~~The piece~~ — If now the diffracted rays enter

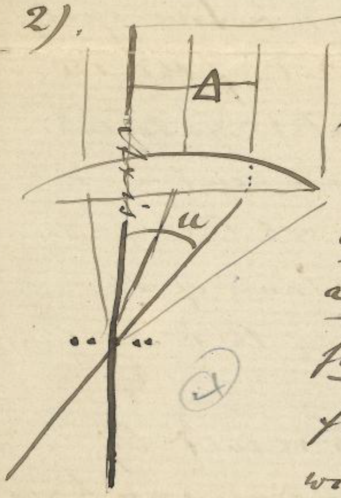
XX

8

XX the angles of

a microscope, the sines of any two consecutive rays ~~have~~ ~~the~~ ~~...~~ with the axis of the microscope have the same difference $= \frac{\lambda}{\delta}$.

2). ~~If said rays enter a lens~~ If an objectif is focused to the ~~the~~ grating, and if this objectif is perfectly x) aplanatic for its focal point any ray forming an angle u below the objectif with the axis, is refracted



in such a way, that it will pass the upper (the back) focal-plane of the lens in a linear distance from the axis

$$\Delta = f \cdot \sin u$$

if f is the focal-length of the objective (by a theorem enounced by me and by Mr Helmholtz)

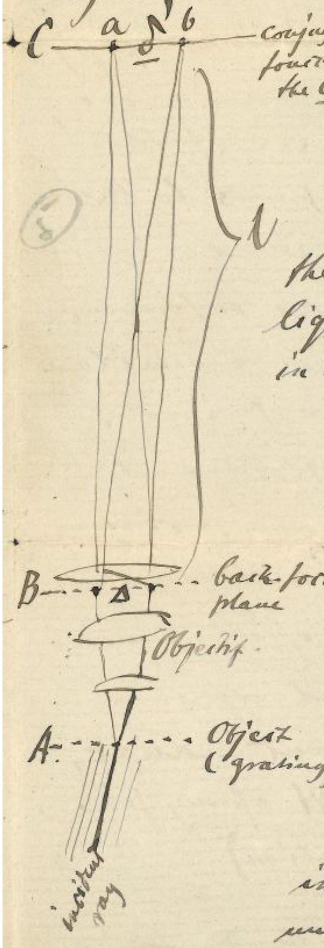
x) "perfectly aplanatic" means: ~~for~~ without spherical aberration not only for ~~the~~ one point on the axis, but for the points ~~besides~~ aside the axis too.

From this theorem 2) in connection with 1)
is to be inferred: the linear distance
of the diffraction-spectra, which appear
in the back-focal-plane of the objective
is always $= \frac{\lambda}{\delta} \cdot f$, if corresponding
points in every two *consecutive* spectra
are considered — *independent* of the in-
clination of the incident rays to the
grating. — If you go from central
light to oblique light, all the spectra
move within the back-plane of the
system, without changing their relative
position.

3). All the rays, which result by
diffraction, from *one* incident ray have
their oscillations in *equal phase*, if
points are compared on these rays which
are situated in the back-focal-plane,
where the spectra are formed as images
of the illuminating object; all those
rays therefore *must* interfere ~~in~~ within

the plane, where they ~~are~~ meet — that is the place, ~~of the image~~ where an image of the grating is formed by the objectif (the conjugate focus to the microscopic object)

4) ——— If Δ be the linear distance of the 2 interfering rays in the back-focal-plane, l the distance of the conjugate focus to the object (= length of tube of the microscope)



the maxima and minima of light, resulting by interference in the plane C, have a distance

$$\delta' = l . \frac{\lambda}{\Delta}$$

Now if the two rays considered are consecutive rays from a grating with the distance δ, ~~&~~ Δ ~~is~~ $= \frac{\lambda}{\delta} . f$; therefore

$$\delta' = \delta . \frac{l}{f}$$

— that is the same distance, in which the lines of the grating would appear in a purely dioptrical image, under the same circumstances.

λ being eliminated from the expression
of δ', the intervalls in the interference-image
must be equal for the different coulours; this
image must be achromatic, if the objectif
is achromatic (f is constant for different
colours)

If the two rays considered were not consecutive
(as in the experiment with the 3-holes-stop)
Δ would ~~be~~ have double (or triple...) the value taken above;
therefore δ' would be $\frac{1}{2}$, or $\frac{1}{3}$ ... of the
distance, ~~in an~~ which corresponds to the
real distance δ in a similar image.

This reasoning shows, that the interference
of the diffracted rays can give a similar
image of the structure, but not must.

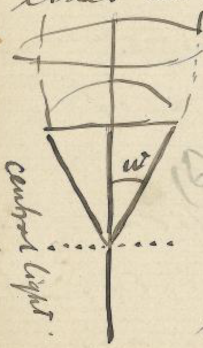
The want of mathematical exactness in the deduction
above (in 4) ~~is perfectly remove~~, arising
from the supposition: L infinitely great
in relation to f and Δ, is perfectly removed
by considering the dioptrical effect
of the microscope in a different manner,
which I have stated in No VI of my paper
(pag 213 in Mr Fripp's translation)

XXX

8

XXX

The theses in 1) and 2). involve the determination of the limit of visibility, as deduced from the facts that <u>two</u> pencils must enter the objectif in order to get an image. — If <u>w</u> be the <u>semi</u>-aperture of any objectif, and δ the minimum-distance of visible lines in an object, there is ~~the~~ for <u>purely</u> central illumination:



$$\sin w = \frac{\lambda}{\delta}\ ;\ \delta = \frac{\lambda}{\sin w}$$

6 and for the extreme oblique ~~incidence~~ illumination, when the incident ray touches the margin of the lens on ~~the~~ one side, the <u>next</u> diffracted ray on the other side:

$$2 \sin w = \frac{\lambda}{\delta}\ ;\ \delta = \frac{1}{2}\frac{\lambda}{\sin w}.$$

7 as stated on p. 244 of Mr Fripp's translation.

I hope, these remarks will be sufficient to you for getting a clear notion of the mathematical principles of the theory.

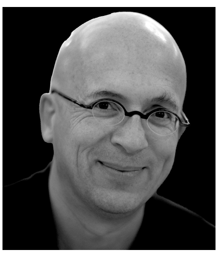
I shall be very glad, if you should like to show the experiments to the Microscopical Society – especially if you should think it convenient to produce them not as paradox phenomena, but rather as phenomena illustrating a distinct idea of the functions of the microscope. For there is no want of optical curiosities among microscopists; and I take no interest in bringing forth more of such. – Please, make any use of my explanations, you like.

With my best regards I remain

Yours truely

E. Abbe.

Anthony Yen is Vice President and Head of the Technology Development Center at ASML, responsible for providing the company with mid- and long-term directions in semiconductors and working with customers, peers, universities, and research centers to develop enabling technologies. Prior to joining ASML, he headed the Nanopatterning Technology Infrastructure Division at TSMC and played a key role in developing EUV lithography for high-volume production. He received his undergraduate degree in electrical engineering from Purdue University and his master's, engineer's, doctoral, and MBA degrees from MIT. He is a Fellow of SPIE, a Fellow of IEEE, a recipient of the Frits Zernike Award for Microlithography from SPIE, and a recipient of the Outstanding Electrical and Computer Engineer Award from Purdue.



Martin Burkhardt is a Research Staff Member at IBM Research in Yorktown Heights, NY. He has worked in lithography since leaving graduate school, first at Texas Instruments, then at ASML, before joining IBM in 2001. Starting with i-line lithography, he now works on imaging and mask technologies for EUV lithography, following the shrinking wavelengths that were used in state-of-the-art lithography for integrated circuits over the years. He received a Dipl.-Ing. degree from the Technical University of Berlin, and a Ph.D. from MIT, both in electrical engineering. He is a Fellow of SPIE.

ERNST ABBE'S
Theory of Image Formation in the Microscope

Written and published by Otto Lummer and Fritz Reiche under the title *Die Lehre von der Bildentstehung im Mikroskop von Ernst Abbe*
Translated and annotated, with additional material, by Anthony Yen and Martin Burkhardt

This book is an English translation of *Die Lehre von der Bildentstehung im Mikroskop von Ernst Abbe,* the only published detailed account of Abbe's theory of image formation in the microscope. The original German edition, written and published by Otto Lummer and Fritz Reiche in 1910, was an expanded version of the lectures given by Abbe in 1887. The book presents an introduction to geometrical optics, discusses image formation theory based on optical diffraction, and deals with optical images of several kinds of objects being illuminated on and off axis. It also introduces coherent imaging as two back-to-back diffraction processes and discusses the resolution limit of an imaging system. The book concludes with a discussion of the effect of artificial blocking of certain diffraction orders on the final image.

This translation, which includes annotations and other added material, can serve as a self-study book for readers who wish to learn optics and optical image formation. It is also a tribute to the original authors' scientific achievements and devotion to the teaching and dissemination of precious knowledge.

ISBN 978-1-5106-5523-2
90000
9 781510 655232

SPIE.

P.O. Box 10
Bellingham, WA 98227-0010

ISBN: 9781510655232
SPIE Vol. No.: PM352