2

Principles of Light Microscopy

Ulrich Kubitscheck

Friedrich-Wilhelms-Universität, Institut für Physik & Theoretische Chemie, Wegelerstr. 12, 53115 Bonn, Germany

2.1 Introduction

In this chapter, we use the basic knowledge of geometrical optics and the properties of light waves to understand the construction and functioning of light microscopes. We focus more on the underlying physical principles of microscope operation and less on the technical details of the optical construction. The central goal is to achieve an understanding of the capabilities and limitations of microscopic imaging, which will serve as a basis for the subsequent chapters. The chapter also includes an introduction to the (currently) most important optical contrast techniques in the biosciences: phase contrast and differential interference contrast (DIC). These techniques allow the creation of contrast-rich images of almost all transparent specimens that are typically encountered in biological and medical research. This chapter presents the basic material that is needed to understand fluorescence microscopy and its wide range of applications, as well as new technical developments as detailed in the following text.

2.2 Construction of Light Microscopes

2.2.1 Components of Light Microscopes

An optical microscope consists of two functionally distinct sets of components the illumination elements and the imaging elements. Newcomers usually consider only the optical elements of the imaging beam path to form the "real" microscope and tend to overlook the importance of the illumination components. Indeed, the two parts are deeply interconnected with each other and of equal importance for the achievable optical resolution and the creation of image contrast. The two parts comprise their own light paths, which are in a certain sense complementary to each other. To begin with, we consider them according to the rules of geometrical optics, which means that we consider light rays and beam paths while neglecting the wave nature of light and any diffraction processes. These are treated in the later sections of this chapter.

Fluorescence Microscopy: From Principles to Biological Applications, Second Edition.
Edited by Ulrich Kubitscheck.

2.2.2 Imaging Path

The basic element of the imaging part of a modern microscope is formed by two lenses. The specimen or object is located in the front focal plane of the first lens of the optical system, which is designated as the *objective lens* and produces an image of the object at infinity. This image is focused by a second lens – the *tube lens* – into the primary image plane, as shown in Figure 2.1.

The imaging process can be understood by the geometric imaging laws as follows: To this end, we follow the beams emitted from the two extreme ends of an object. The light emitted by the object point at the lens focus o1, is transformed by the objective lens into a set of light beams that travel parallel to the optical axis (Figure 2.1, full lines). This parallel light bundle enters the tube lens and is refracted into its focus. The fate of the light beams emitted from the off-axis point o2 can be determined by the following two special light beams: the ray rc, hitting the lens center, passes the objective unrefracted; the ray rp, emitted parallel to the optical axis, is focused into the back focal point of the objective lens. We note that both beams are parallel to each other in the space between the objective lens and the tube lens. This space is designated as *infinity space.* All beams emitted from o2 will also be parallel to each other in this space, because o2 lies in the front focal plane of the objective (Figure 2.1, dashed lines). The position of the tube lens in relation to the objective can be chosen quite arbitrarily. However, usually it is positioned such that its front focal point coincides with the back focal point of the objective lens, at least approximately. In order to find out to which point the tube lens focuses these beams, we can follow the known special beams of the tube lens. We follow the beam passing through the front focus of the tube lens. This is translated into a beam parallel to the optical axis behind the tube lens, rp′. The central beam rc′ passes the tube lens again unrefracted. Hence, the point o2 is projected to the intersection of these two beams, b2, in the back focal plane

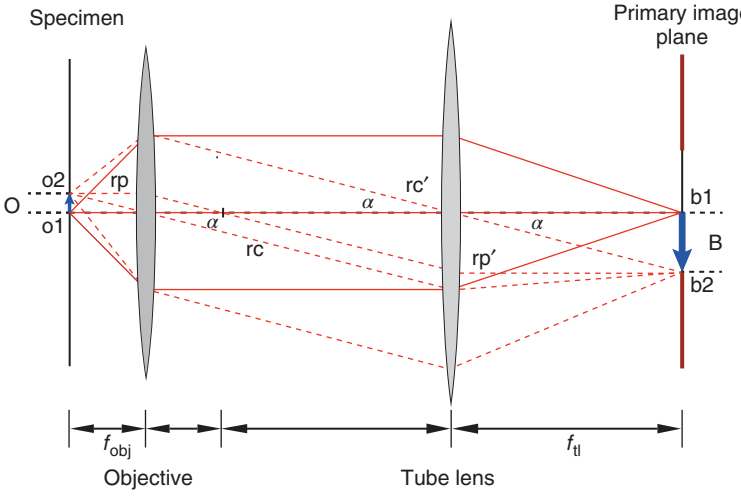


Figure 2.1 Infinity-corrected imaging process using a light-sensitive array image detector. f_{obj} and f_{tl} denote the focal length of the objective and tube lens, respectively.

of the tube lens. This plane contains a real image of the object and is designated as the *primary image plane.* The image size corresponds to the vertical distance between this point and the optical axis where the image point b1 of o1 was found. It can be seen that

$$\tan \alpha = \frac{O}{f_{\text{obj}}} \tag{2.1}$$

where α is the angle between the optical axis and beam rc in the object space, which lies to the left of the objective lens. O is the distance between o1 and o2 corresponding to the object size, and f_{obj} is the focal length of the objective lens. Examining the beams that form the image with the size B on the right-hand side of the tube lens, we observe also that

$$\tan \alpha = \frac{B}{f_{\text{tl}}} \tag{2.2}$$

where f_{tl} is the focal length of the tube lens. It follows that the magnification factor M of this two-lens system is

$$M = \frac{B}{O} = \frac{f_{\text{tl}}}{f_{\text{obj}}} \tag{2.3}$$

The ratio of the focal lengths of the tube lens and the objective lens determines the magnification factor. The tube lens is fixed within the microscope, and therefore the choice of the objective lens determines the respective magnification. The minimum focal length of the objective lens is limited for practical and technical reasons, but the focal length of the tube lens is relatively large. Magnification factors usually range between 1 and 100.

For visual observation of microscopic samples, two magnification steps are arranged in sequence (Figure 2.2). The objective lens and tube lens form the first magnification step and create together the *primary* or *intermediate image.* Its absolute size has a diameter of 2–3 cm depending on the type and quality of the objective lens. The second magnification step is also a two-lens infinity optical system working in a way equivalent to the first magnification step. It is formed by the eyepiece and the eye lens of the observer, which project the final image onto the retina. The retina in the eye of the observer is the *secondary image plane.*

In principle, there are two different ways in which a lens system with a focal length f may produce a real image: when the object is placed at a distance s_o from the lens with $2f > s_o > f$, a real, inverted, and magnified image is produced directly at a fixed distance from the lens in the primary image plane. This type of configuration is called the *finite optics setup*. The other possibility is to place the object at the focal plane of the lens and use a second lens in an infinity setup as discussed earlier. In microscopy, both magnification steps have been realized using finite and infinite setups.

Modern research microscopes feature infinite optics as shown in Figures 2.1 and 2.2. They are designated as infinity-corrected microscopes. The object is located exactly in the front focal plane of the objective lens, which therefore creates its image at an infinite distance. This requires an additional lens to form a real primary image, the tube lens, which may also be used to compensate

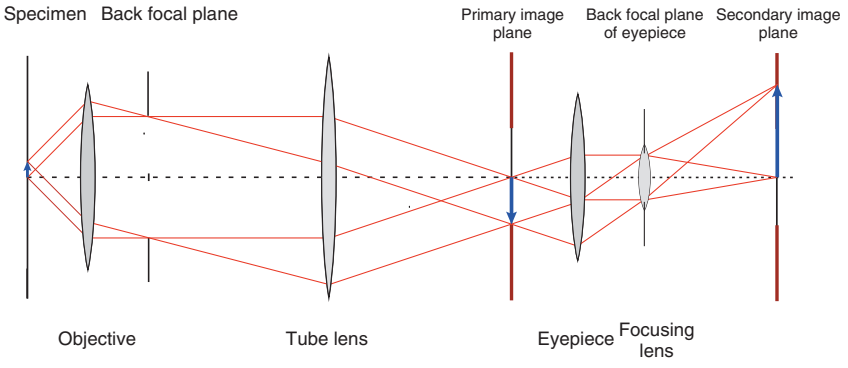


Figure 2.2 Microscope with two infinity-corrected magnification steps. The magnification factor of the sequential arrangement of two magnification steps is the product of the two single factors. Multiple creation of infinity spaces should be appreciated.

for residual image aberrations. The great advantage of this setup is that the light path behind the objective contains only sets of parallel rays that originate from the various object locations. There is ample space to insert additional flat optical elements without distorting the beam path, since parallel beams are only displaced by plane-parallel optical elements. Optical components such as the fluorescence filter cube, polarization beam splitter, analyzer, or additional filters are required for various contrast techniques. They are mostly inserted between the objective lens and the tube lens. Another advantage of infinity-corrected objective lenses is that one can focus by moving the objective rather than the specimen stage.

2.2.3 Magnification

In the finite optics setup, the objective lens produces directly a real and inverted primary image. For common transmitted-light microscopes of small to medium magnifications, there is a commercial advantage in employing finite optics. Its major handicap is that the insertion of optical components, even plane-parallel optical devices, always distorts the converging imaging beam path. To prevent this, an additional lens system must be inserted in the space between the objective lens and the primary image plane.

In many modern microscopic systems, a light detection device such as a charge-coupled device (CCD) camera is positioned already at the primary image plane, which means that the image is registered after a single magnification step. For visual observation, however, the primary image is magnified once more by the eyepiece system. Then, the primary image is located (almost) in the front focal plane of the eyepiece, which maps its image to infinity, and the human eye focuses the final image onto the retina.

Thereby, the eye becomes an intrinsic part of the imaging system (Figure 2.2). As the primary image is located in the front focal plane of the eyepiece, the emerging rays leave the eyepiece parallel to each other, such that the eye is looking at an object in infinity, which is physiologically optimal for relaxed viewing. The eyepiece is used as a magnifying glass by which the eye observes the primary

image. As shown in Figure 2.2, the eyepiece and eye together form a second infinity setup, which projects the final image onto the retina. In this figure and further drawings of this chapter, the eyepiece is sketched as a single lens. This is a simplification, as the eyepiece – similar to the objective – is a system comprising several internal lenses that are needed to correct imaging artifacts. Further details on the optical construction of various types of eyepieces can be found in the references to this chapter.

What is the magnification of this second stage alone? To answer this, we need the magnification factor of a magnifying glass. The effect of a magnifying glass is usually quantified by relating the image size of an object with size O when seen with (B') and without magnifying glass (B), $M_{mg} = B'/B$. To calculate that ratio, we can use the magnification of the magnifying glass–eye lens system when looking at it as infinity system, $M_{is} = f_{eye}/f_{mg}$. Here, f_{eye} and f_{mg} denote the focal lengths of the eye lens and the magnifying glass, respectively. The image size B' is then $B' = O \times f_{eye}/f_{mg}$. When we place the object in the normal viewing distance of 250 mm in front of the eye, and observe it without the magnifying glass, the image on the retina has a size of $B = O \times d_{eye}/250\,\text{mm}$, where d_{eye} is the distance between the eye lens and the retina, that is, the eye diameter. This value is close to the focal length of the eye, $f_{eye} \approx d_{eye}$. Relating the two image sizes yields therefore a magnification of $M_{mg} = B'/B \approx 250\,\text{mm}/f_{mg}$.

Consequently, we can write for the magnification factor of a microscope eyepiece, M_{ep}

$$M_{ep} = \frac{250\,\text{mm}}{f_{ep}} \tag{2.4}$$

here f_{ep} designates the focal length of the eyepiece. Finally, the product of the magnification factors of the two magnification stages yields the total magnification of the compound microscope, M_{total}:

$$M_{total} = M_{obj} \times M_{ep} \tag{2.5}$$

The total magnification can theoretically be infinite; however, the meaningful magnification that should be chosen can be well defined. The determination of the optimal magnification factor is discussed in Section 2.3.6.

2.2.4 Angular and Numerical Aperture

An important characteristic of an objective lens is the maximum angle at which the light emitted from the object can be collected by the objective lens. Only the light contained in the respective cone contributes to the final image. The total angular opening – also designated as angular aperture – of the objective lens is the maximum angle at which the beams emitted from the focus F finally contribute to the image (Figure 2.3). The angle α in the figure designates half of the angular aperture. The sine of this angle α plays a central role in the theory of microscopy, because the *numerical aperture* of an objective lens, the NA, is defined as

$$\text{NA} = n \sin \alpha \tag{2.6}$$

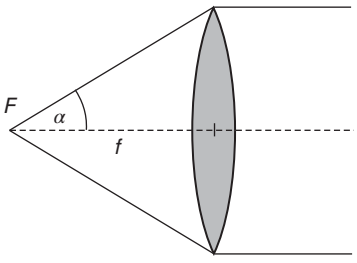


Figure 2.3 Angular aperture of an objective.

where n is the refractive index of the medium between the lens and the object. Biological samples are often observed in an aqueous buffer solution whose refractive index is $n_w = 1.33$. Often, the so-called oil immersion is used, for which $n_{oil} = 1.518$. Generally objectives that are designed and optimized for the respective media must be used. The crucial importance of the NA will be made clear in the course of this chapter.

2.2.5 Field of View

The observable object field is limited in the primary image plane by a field stop, which is often located inside the eyepiece. The diameter of this aperture in millimeters is designated as *field number* or *field-of-view number*, *s*. With this number and the magnification of the objective lens, M_{obj}, the diameter of the object field, D_{obj}, that can maximally be imaged is given by

$$D_{obj} = \frac{s}{M_{obj}} \tag{2.7}$$

Clearly, the quality of the image depends on the field performance of the objective lens, which refers to the image quality in the peripheral regions of the image field. In older microscopes, the usable field number was up to 18, whereas nowadays for plan achromats 20 is standard, and modern plan apochromats feature a corrected maximum field number of up to 28. Objectives with a poor field performance will produce a blur and chromatic aberrations in the peripheral regions of the image.

2.2.6 Illumination Beam Path

The illumination system of a light microscope does more than simply illuminate nonluminescent specimen. Rather, it defines the contrast mode, the resolution of the instrument, and the overall brightness in an equally important way as the imaging optics. For the illumination system, two principally different optical setups are in use. The simpler one is the *source focus* or *critical illumination* and the other, which is by far more prevalent, is called *Köhler illumination.*

2.2.6.1 Critical and Köhler Illumination

The illumination beam path for critical illumination is shown in Figure 2.4a. The light source is imaged directly by a lens, the so-called condenser lens, into the object plane. A variable aperture is located in the front focal plane of the condenser lens, whose variation allows the control of the aperture of the lens

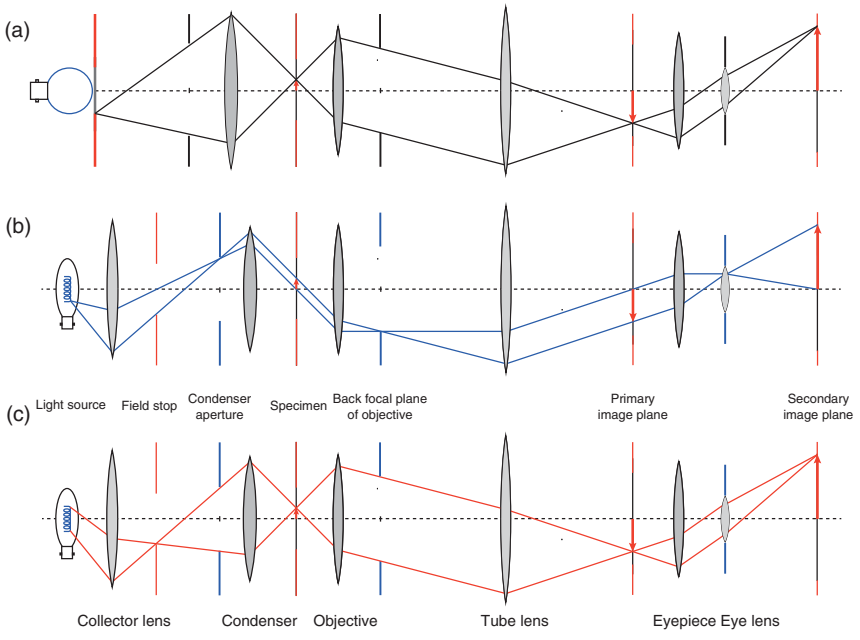


Figure 2.4 Illumination and imaging beam path. (a) Critical illumination. (b) Köhler illumination and (c) the corresponding imaging beam path.

and thereby the amount of light that illuminates the specimen. At the same time, however, the size of the illuminated field is affected. The illumination field size is fully controlled by a variable aperture located immediately in front of the light source. Images of the light source are located in the object plane, the primary image plane, and the retina. Therefore, the light source must be as large and featureless as possible; otherwise, the object brightness will be modulated by the uneven illumination. Usually, a light source with an opal glass bulb is used, or a pane of opal glass illuminated from the back. Critical illumination is used in routine instruments that are not equipped with special contrast techniques. It is simple, easy to set up, and requires few optical components.

Köhler illumination is the standard illumination mode for scientific microscopes and is named after its developer August Köhler who introduced it in 1893 [1]. The illumination components are

- the light source
- two diaphragms, namely the field stop and the condenser aperture stop
- two lenses, namely the collector and the condenser.

For transmitted-light illumination, usually a white light source is used, for example, a halogen lamp or a light-emitting diode (LED). The Köhler illumination beam path is shown in Figure 2.4b. The filament of the light source is imaged by the collector lens on the front focal plane of the condenser lens where the condenser aperture is located. This aperture should completely be filled with the image of the filament. Since the image of the filament is in the front

focal plane of the condenser, all illumination light rays leaving the condenser toward the object plane are parallel bundles and traverse the specimen in a parallel manner. The undiffracted, parallel beams are focused by the objective onto its back focal plane and form there an image of the light source. Planes that contain images of each other are designated as *conjugate planes.* Thus, the *back focal plane of the objective lens* and the *front focal plane of the condenser lens* are conjugate to each other. This is of utmost importance because it is used in the key contrast techniques as, for example, phase contrast and DIC. Behind the back focal plane of the objective lens, the illumination light diverges, but it is again collected by the tube lens. The distance between tube lens and back focal plane of the objective matches the focal length of the tube lens, at least approximately. Thereby, a second infinity space for the illumination beam path is generated that contains the primary image plane. The illumination light traverses this space in parallel rays. Finally, the eyepiece focuses the incoming parallel light beams onto its back focal plane, thus generating another image of the lamp filament, which is the so-called *exit pupil* or *Ramsden disk* of the microscope. Only light passing this pupil can contribute to image formation. Here, the pupil of the observing eye must be placed, as it represents the ultimate limiting pupil. Hence, the illumination rays cannot be focused onto the retina. Consequently, the illuminated object field appears homogeneously illuminated to the observer.

In summary, optically conjugated planes of the *illumination beam path* are located in the plane of the lamp filament, the condenser aperture, the back focal plane of the objective lens, and finally the exit pupil of the eyepiece. The condenser produces an image of the object back at a position close to the collector lens of the lamp. In this plane, a variable field stop is located that defines the extension of the illuminated object field. Therefore, we have a second set of conjugate planes comprising the field stop, the object plane, the primary image plane, and the retina, called the *secondary image plane*. This set of conjugate planes is called the *imaging optical path*. The two sets of conjugate planes are complementary to each other (Figure 2.4b,c).

Köhler illumination has a number of important features:

- The specimen is evenly illuminated, although a structured light source, for example, a lamp filament, is used. This is due to the fact that all illuminating beams traverse the specimen in a parallel manner.
- The size of the illuminated field is determined by the field stop alone, because it is located in a conjugate plane of the specimen.
- Changing the size of the condenser diaphragm changes the opening angle and the NA of the light incident on the specimen. Thereby, the specimen contrast and the optical resolution (see later) can be adjusted without affecting the size of the illumination field.
- Illumination field size and condenser aperture are completely independent of each other.
- The front focal plane of the condenser and the back focal plane of the objective are conjugate to each other. This principle is used in important light microscopic contrast techniques.

2.2.6.2 Bright-Field and Epi-Illumination

In the illumination beam paths discussed earlier, the specimen is placed between the light source and the objective lens. This configuration is called *dia-illumination*, *transmitted-light illumination*, or simply *bright-field illumination*. The illuminating light that has traversed the specimen and the light that is diffracted by the specimen are collected by the objective lens. Various light microscopic techniques use this type of illumination.

In many cases, however, it is advantageous to illuminate the specimen from the side of observation, for instance, when looking at the reflection of opaque or fluorescent samples. In that case, some optical components of illumination and imaging are identical, for example, the objective lens. The specimen is illuminated through the objective lens by means of a semitransparent mirror, as shown in Figure 2.5.

The semitransparent mirror directs the light through the objective lens, which is simultaneously taking over the role of the condenser. Such setups are

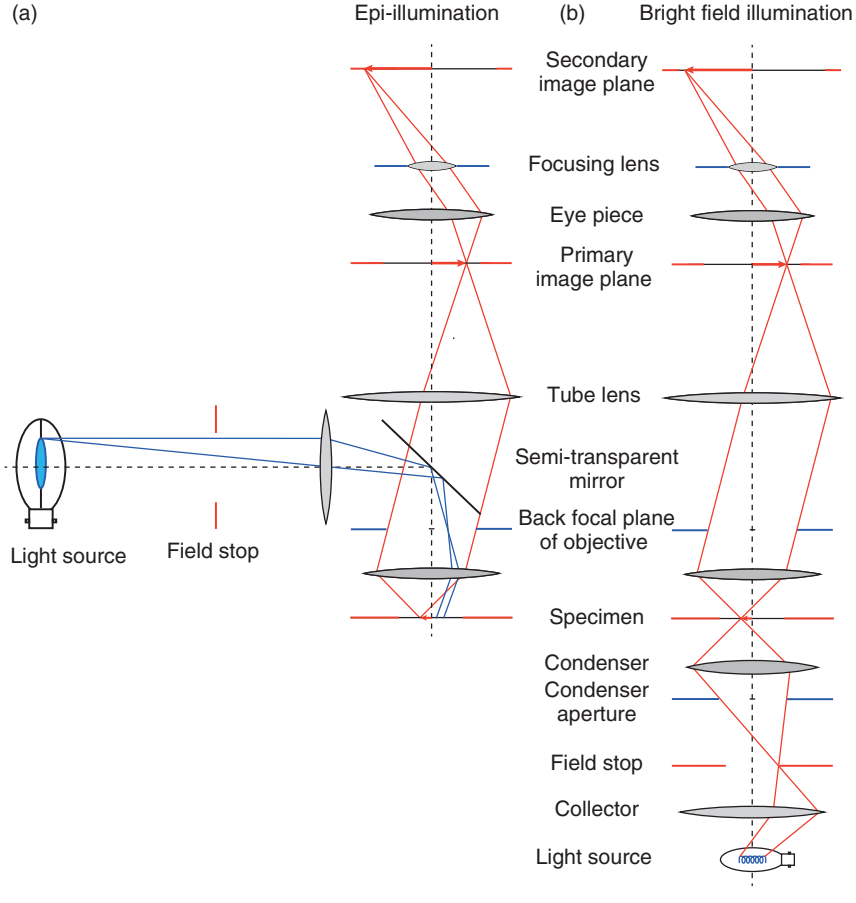


Figure 2.5 Illumination modes in a microscope. (a) Epi-illumination. (b) Transmitted-light mode.

designated as *incident-light* or *epi-illumination* configurations. Köhler illumination can also be realized in epi-illumination. To this end, it is only necessary to project an image of the lamp filament onto the back focal plane of the objective lens as shown in the figure.

In epi-illumination, the illumination field size is adjusted by a field stop that is located between the light source and the semitransparent mirror. The front focal plane of the condenser lens corresponds in epi-illumination to the back focal plane of the objective lens. However, the aperture diameter is not adjusted in this plane – as it is usually inaccessible – but in an additional conjugate plane, which is not shown in Figure 2.5 for the sake of simplicity.

The traditional microscope configuration is upright. The object is placed on an object glass that is illuminated from below and viewed from above. Upright microscopes have a direct optical path with great optical efficiency. A disadvantage is, however, that it is difficult to manipulate the object during observation.

On *inverted* microscopes, specimens are positioned above the objective lens and therefore observed from below. This is a convenient configuration when samples are examined in open chambers containing liquid media, for example, cell culture flasks or dishes. Specimen manipulation during microscopic examination such as microinjection can easily be performed. In addition to the advantage of better specimen accessibility, inverted microscopes are also generally more stable. Microscopes that feature a camera port at the bottom of the instrument have the highest possible optical efficiency for this base port.

2.3 Wave Optics and Resolution

The path of light rays in the microscope is determined by the law of refraction and the position and curvature of lens surfaces. Using simple geometrical optics, we are able to deduce many fundamental imaging characteristics of optical systems, for example, magnification factors and lens aberrations.

The origin of the limited microscopic resolution and image contrast, however, can be understood only by considering the wave nature of light. Before going into details, it is important to carefully discriminate between magnification and resolution. It is possible to increase the magnification by adding more and more magnification stages to a microscope. However, it is not possible to improve the resolution of an imaging system beyond a fundamental limit.

Resolution refers to the level of detail that can be recognized in an image, such as small and intricate structures or the distance between closely placed small objects. Indeed, such a distance is used to define and quantify the optical resolution. Using light microscopy, minute objects can be discriminated from each other when they are positioned at a minimum distance of ~0.25 µm from each other and green light and a high-quality oil-immersion objective lens are used. This obviously means that proteins and supramolecular complexes occurring in living cells cannot be recognized in detail. In later chapters, we discuss how the principal optical resolution limit can be overcome or circumvented by advanced optical techniques.

2.3.1 Wave Optical Description of the Imaging Process

There is a rigorous approach to the optical resolution limit that follows the experiments and argumentation of Ernst Abbe [2]. We indicate here the principal line of argument and examine what happens when light illuminates an object and it is imaged by a lens system [3]. We begin with a simple object, and then generalize the basic ideas.

Let us assume that the object is a simple rectangular grating formed by multiple lines within a distance d from each other, which is designated as grating constant and of the order of micrometers – somewhere near the wavelength of visible light. We illuminate the grating in transmission with coherent light of wavelength λ. For simplicity, we consider only illumination parallel to the optical axis. The grating diffracts the incoming light and forms a diffraction pattern due to the Huygens wavelets that emanate from the various grating lines. Such grating diffraction patterns are discussed in any textbook on optics [4]. It can be shown that the Huygens wavelets originating from the grating lines interfere constructively only in very distinct directions α_n.

Only in directions at which the angle α_n fulfills the so-called Bragg condition

$$d \sin \alpha_n = n\lambda \tag{2.8}$$

with n being an integer, interference is constructive. In all other directions, the waves cancel each other by destructive interference. Hence, behind the grating, we will observe diffracted wave fronts – or light beams – in specific directions α_n. If we place a screen at a great distance behind the grating, we will observe distinct spots of diffracted light corresponding to the various diffraction orders n. However, when we place a lens in the space behind the object, it will focus the diffraction pattern into its back focal plane. All diffraction orders are plane-parallel wave fronts that will be focused onto points at positions p_n in the focal plane of the lens. Positive and negative p_n denote positions above and below the optical axis, respectively (Figure 2.6). In order to create the conditions of infinity optics introduced earlier, we place the lens – thought to represent the objective lens – exactly a focal length f behind the object. If the lens satisfies the *von Bieren condition* (see Box 2.5 for details on the *sine* and *von Bieren conditions*), we get

$$\sin \alpha_n = \frac{p_n}{f} \tag{2.9}$$

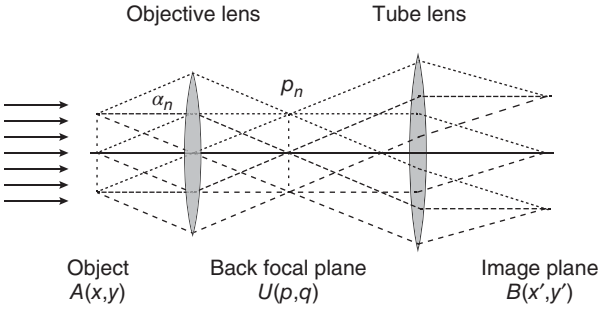


Figure 2.6 Imaging a grating structure. (Source: Modified from [3], with kind permission by Carl Zeiss Microscopy GmbH.)

Inserting the Bragg condition (Eq. (2.8)) in Eq. (2.9), we can calculate the positions p_n:

$$p_n = \frac{n\lambda f}{d} \tag{2.10}$$

Starting at the focus at $p = 0$, we find the diffraction orders above and below the optical axis corresponding to increasing orders n of the light diffracted by the grating. Interestingly, the diffraction image of finer gratings with smaller constants d displays points at positions p_n that have greater distances from the optical axis, because d is in the denominator in Eq. (2.10). In addition, p_n are linearly dependent on the wavelength λ. Using white light for illumination will thus produce points that have red edges away from the axis and blue edges toward the axis. It is a very general conclusion that a lens produces an image of the diffraction pattern of the object in its back focal plane, and it is not required that the object is in the front focal plane, as we chose to place it. For microscopy, the extremely important role of the back focal plane was noted by Ernst Abbe, and it was extensively discussed in his famous, classic article that was published in 1873 [2].

Starting from here, we can build a microscope by simply using a second lens – the tube lens – as discussed earlier. The diffraction pattern presents again a repetitive structure. We will find a second diffraction pattern in the back focal plane of the tube lens, which is the diffraction pattern of the first one. This is the plane that we introduced earlier as the primary image plane. The second diffraction pattern will be magnified according to the ratio of the focal distances of objective and tube lens as given by Eq. (2.3). Actually, Abbe quite logically designated the structure seen in the back focal plane of the objective lens as the "primary" image, and the final, projected image as the "secondary" image. However, we will stick to our earlier definition.

What happens if we image finer and finer gratings? Equations (2.8) and (2.10) reveal that α_n and the distances p_n from the optical axis, where we find the foci of the diffraction orders, increase more and more for decreasing d. However, the lens represents an aperture of finite size that limits the angle to a value α_{max}, for which the diffraction pattern can be collected. Soon, some diffraction orders will be missed even by large lenses because, for practical reasons, we cannot construct a lens with an aperture angle approaching 90° ($\pi/2$). This leads to an incomplete first diffraction pattern that is used to form the final image, which will not be any longer a true image of the initial structure. And when not even the first diffraction order is transferred by the lens, that is

$$d < \frac{\lambda}{\sin \alpha_{max}} \tag{2.11}$$

all information of the grating structure is lost. This is the underlying physical reason for the *resolution limit of microscopy.* Finally, when we consider that there may be an immersion medium with a refractive index n in the object space that reduces the vacuum wavelength of the light λ_0, we finally get the resolution limit for periodic objects such as gratings:

$$d < \frac{\lambda_0}{n \sin \alpha_{max}} \tag{2.12}$$

This expression is valid if the illumination direction is parallel to the optical axis. In case we illuminate the object under an oblique angle, the interference conditions are slightly different, and we obtain a further factor of 2 in the denominator of Eq. (2.12) for the maximum possible oblique illumination [2]. Thus, a high degree of oblique illumination improves the resolution of the microscope. A similar expression is given later for the resolution limit of a microscope when imaging a point object and not a repetitive grating structure. The essential point of the presented considerations is that even an ideal, aberration-free objective lens that does not modify phases or amplitudes of the first diffraction pattern in its back focal plane will nevertheless have severe consequences on the imaging process. Obviously, the outer points of the diffraction pattern represent the finest object details, because a reduction of the grating constant d will increase α_n and cause the points p_n to move away from the optical axis. Hence, the image that can be produced by the microscope will often miss the finest details present in the original structure (Box 2.1).

Box 2.1 General Theory of Image Formation

In a more general approach, we image a two-dimensional object of size $2a$ in the x-direction and $2b$ in the y-direction. It may be described by an amplitude transmission function $A(x, y)$, for which then $A(x, y) > 0$ only for $|x| \leq a$ and $|y| \leq b$. In general, the diffraction pattern U is given as a function of the angle α_x and α_y away from the optical axis by the Huygens–Fresnel integral equation:

$$U(\alpha_x, \alpha_y) = c \int_{-\infty}^{+\infty} A(x, y)\, \mathrm{e}^{i\frac{2\pi}{\lambda}(x \sin\alpha_x + y \sin\alpha_y)} \mathrm{d}x\, \mathrm{d}y \tag{2.13}$$

This expression is nothing but the mathematical representation of the Huygens wavelets that originate at the object structure, and describes the diffraction pattern far away from the diffracting object. Next, we place the object at the focal plane of the objective lens with a focal length f. If it satisfies the von Bieren condition (see Box 2.5 for details), we can write

$$\begin{aligned} \sin\ \alpha_x &= \frac{-p}{f} \\ \sin\ \alpha_y &= \frac{-q}{f} \end{aligned} \tag{2.14}$$

The negative sign was chosen for convenience; we are free to define the axis. We can substitute that into Eq. (2.13) for the diffraction pattern, which we can now write as a function of p and q:

$$U(p, q) = c \int_{-\infty}^{+\infty} A(x, y)\, \mathrm{e}^{-i\,\frac{2\pi}{\lambda f}(px+qy)} \mathrm{d}x\, \mathrm{d}y \tag{2.15}$$

This is the field of the diffraction pattern in the back focal plane of the objective lens. Indeed, this expression mathematically represents a Fourier integral with the

(Continued)

Box 2.1 (Continued)

Fourier frequencies ω_x and ω_y:

$$\omega_x = 2\pi R_x = \frac{2\pi p}{\lambda f} = -\frac{2\pi}{\lambda}\sin\alpha_x$$
$$\omega_y = 2\pi R_y = \frac{2\pi q}{\lambda f} = -\frac{2\pi}{\lambda}\sin\alpha_y \tag{2.16}$$

R_x and R_y are the spatial frequencies. If we identify the constant c with $1/2\pi$, we see that U represents the Fourier transform of $A(x)$ [5]:

$$U(\omega_x, \omega_y) = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\int A(x,y)\, \mathrm{e}^{-i(\omega_x x+\omega_y y)}\mathrm{d}x\, Ey = \tilde{F}[A] \tag{2.17}$$

We note that the Fourier frequencies are linearly related to the positions p and q in the back focal plane. This shows that the diffraction image in the back focal plane represents a principally undistorted Fourier spectrum of the object function. The inverse Fourier transformation leads back to the object function:

$$\tilde{F}^{-1}\{U(\omega_x,\omega_y)\} = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\int U(\omega_x,\omega_y)\, \mathrm{e}^{i(\omega_x x+\omega_y y)}\mathrm{d}\omega_x\, \mathrm{d}\omega_y = A(x,y) \tag{2.18}$$

We could imagine this back projection to occur by simply placing a mirror at the objective's back focal plane. Or, as indicated Figure 2.6, we could simply add a second lens – the tube lens – to regenerate an image out of the diffraction pattern.

The crucial point is that even an ideal objective lens, which does not modify any phases or amplitudes of the Fourier frequency components in the back focal plane, still *limits* the Fourier frequencies due to its NA. An alteration of the Fourier spectrum, for example, by modifying the Fourier frequencies by multiplication with a function $G(\omega_x, \omega_y)$, to $U'(\omega_x, \omega_y)$ would result in a modified object function $A'(x, y)$

$$\tilde{F}^{-1}\left\{U'(\omega_x,\omega_y)\right\} = \tilde{F}^{-1}\{U(\omega_x,\omega_y)\, G(\omega_x,\omega_y)\} = A'(x,y) \tag{2.19}$$

As discussed earlier, there is always a limiting collecting angle of the objective, that is, $\alpha_{\max}$, beyond which the diffracted light cannot be collected by the lens. The lens represents a circular aperture of finite size that naturally limits the angle for which the diffraction pattern can be collected to $\alpha_{\max}$. This leads to a cut-off frequency, an upper limit of Fourier frequencies that can be transmitted by the lens, $\omega_{\max}$. It is given by

$$\omega_{\max} = 2\pi R_{\max} = -\frac{2\pi}{\lambda}\sin\alpha_{\max} \tag{2.20}$$

This basically means that even an ideal lens represents a low-pass filter for the frequency information of an object. The image that can be produced by a lens will always miss the highest frequencies present in the original structure. The spatial frequency is directly related to the Fourier frequency by multiplication with 2π. The limiting spatial frequency can be inverted to yield the smallest distance that can be resolved.

This argument allows some very important insights into the imaging process. First, the back focal plane contains the diffraction image of our object structure. The spatial coordinates in that plane can be directly related to the spatial frequencies or Fourier frequencies of our object. Second, even the ideal lens will distort the spatial frequencies contained in our object function, namely, by restricting it to certain values. The exact limiting value depends on the collection efficiency of the lens, which is quantified by the NA.

2.3.2 The Airy Pattern

The earlier discussion of imaging grating structures illustrates the physical reason for the resolution limit of microscopy. Let us now place a point object instead of a grating in the focus in front of the objective lens, as depicted in Figure 2.7. Also, we will assume that the angles of the light beams with regard to the optical axis are small, and we will neglect the vectorial properties of light. The point object diffracts the incoming light into a spherical wave w_o. A segment of the spherical wave is transformed by the lens, which is assumed to be ideal and aberration free, into a parallel wave front w_i. As discussed earlier, this wave front of constant amplitude corresponds to the diffraction image of the point object. However, it is modified by the limited angular aperture of the objective lens. According to the Huygens principle, we can imagine each point of the parallel wave front w_i in the back focal plane of the lens to be the origin of a Huygens wavelet. All wavelets have identical phases and are therefore completely coherent. If we place a tube lens behind the objective lens to project the diffraction pattern, the wavelets interfere with each other constructively exactly at the focus in the image plane because only for this point the optical path length is identical for all wavelets. This point also corresponds to the position of the image of the point object, as it can be determined by geometrical optics or by ray tracing. However, because this image spot is the result of an interference process of a finite range of wavelets, it has a finite diameter. Around that spot in the image plane at a certain distance from the optical axis is a circular zone where the wavelets interfere destructively owing to their *optical path length differences* (OPDs). This is indicated as the dark

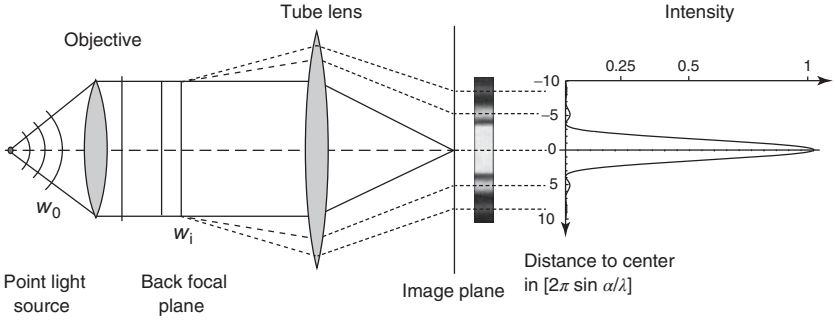


Figure 2.7 Interference of wavelets originating from the exit pupil of the objective lens.

stripe in Figure 2.7. At an even greater distance from the focus, we find a second circle where a majority of wavelets interferes constructively. Here appears a bright ring around the dark zone of destructive interference (indicated by the dotted line in Figure 2.7). This ring is followed again by a ring of destructive interference, then again of constructive interference, and so forth. Altogether, the interference process of the wavelets results in the intensity distribution sketched on the right-hand side of Figure 2.7. The sketch shows a central cut through the two-dimensional, rotationally symmetric intensity distribution. This dark–bright intensity profile is also designated as an Airy pattern, and the central bright spot is the Airy disk. As discussed earlier for the grating, the final image is formed by the interference of the Huygens wavelets originating in the diffraction structure in the back focal plane of the objective lens.

Hence, it is a consequence of the interference that the image of a point object is not a point but a smeared spot with a characteristic intensity profile. This profile is called the *point spread function* (PSF) of the imaging lens. Its analytical derivation is not difficult but lengthy and cannot be discussed here in detail. Thus, we only indicate the derivation.

Let us consider again the circular plane wave segment in the back focal plane of the objective lens, which is the lens's aperture-limited diffraction image of the point object. It corresponds to the field behind a circular aperture of radius a that is illuminated by a plane wave. This situation is discussed in optics textbooks under the topic "diffraction at a circular aperture" [4]. For the imaging process, only the field at a great distance from the aperture is of relevance. Then we can neglect the curvature of the outgoing waves and consider only the planar waves (Figure 2.8a).

In order to compute the far field, we have to integrate over all wavelet contributions originating at the aperture and calculate the total field. The situation is radially symmetric and the field strength is only a function of the angle θ to the optical axis. For $R \gg a$, the integration of the wavelets over the aperture surface results in the following expression for the field $E(\theta)$ [4]:

$$E(\theta) \propto E_0 \frac{J_1\left(2\pi\alpha \sin\theta / \lambda\right)}{2\pi\alpha \sin\theta / \lambda} \tag{2.21}$$

where $J_1(x)$ represents the Bessel function of the first kind of order 1. A tube lens would project this Fraunhofer diffraction pattern from the far field into its focal plane (Figure 2.8b). Parallel light falling on a lens with an angle θ is focused at the focal plane at a distance r from the optical axis, as shown in Box 2.2. There, it is also shown that for small values of angle θ

$$a \sin\theta \approx r \sin\beta \tag{2.22}$$

where β is half of the opening angle of the tube lens. The approximation in Eq. (2.22) is certainly justified for the tube lens when we consider that its focal length is, in practice, between 160 and 250 mm, the lens aperture up to 20 mm, and the maximum image size 28 mm.

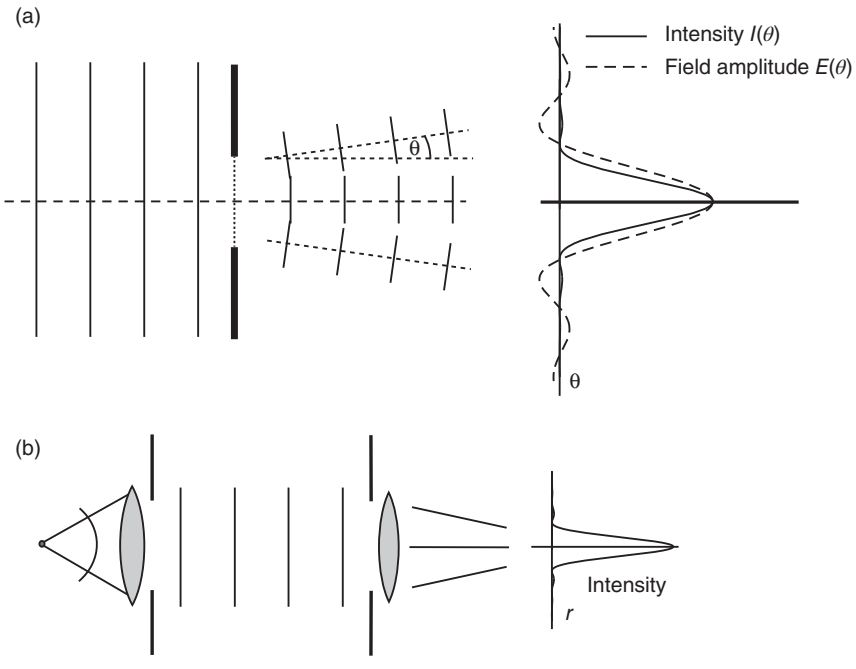
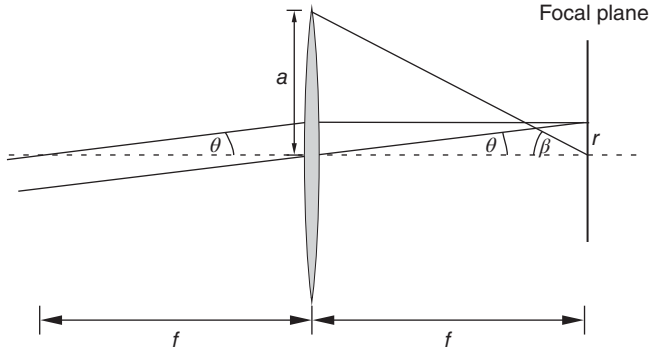


Figure 2.8 Fraunhofer diffraction (a) at a circular aperture and (b) with a lens positioned before and behind the aperture.

Box 2.2 Proof of Eq. (2.22)



$\tan\beta = a/f \approx \sin\beta$, and also $\tan\theta = r/f \approx \sin\theta$, as θ is small.

Solving both expressions for f and equating the results yields $r/\sin\theta \approx a/\sin\beta$, and after rearranging

$$r \sin\beta \approx a \sin\theta$$

In this context, see also Box 2.5.

When we use this expression in Eq. (2.21), the amplitude of the field in the focal plane of the tube lens is obtained as a function of the distance to the optical axis, r:

$$E(r) \propto E_0 \frac{J_1\left(2\pi r \sin\beta/\lambda\right)}{2\pi r \sin\beta/\lambda} \tag{2.23}$$

The light intensity $I(r)$ in the focal plane of the lens as a function of the distance to the optical axis is found by calculating $I(r) = [E(r)]^2$. This results in the following expression:

$$I(r) = I_0 \left[\frac{J_1\left(2\pi r \sin\beta/\lambda\right)}{2\pi r \sin\beta/\lambda}\right]^2 \tag{2.24}$$

where I_0 is the light intensity at the focus of the lens. The two functions $E(r)$ and $I(r)$ are plotted in Figure 2.9.

2.3.3 Point Spread Function and Optical Transfer Function

The earlier discussion leads directly to central concepts of imaging. Box 2.3 explains that an optical system such as a microscope is nothing but a low-pass

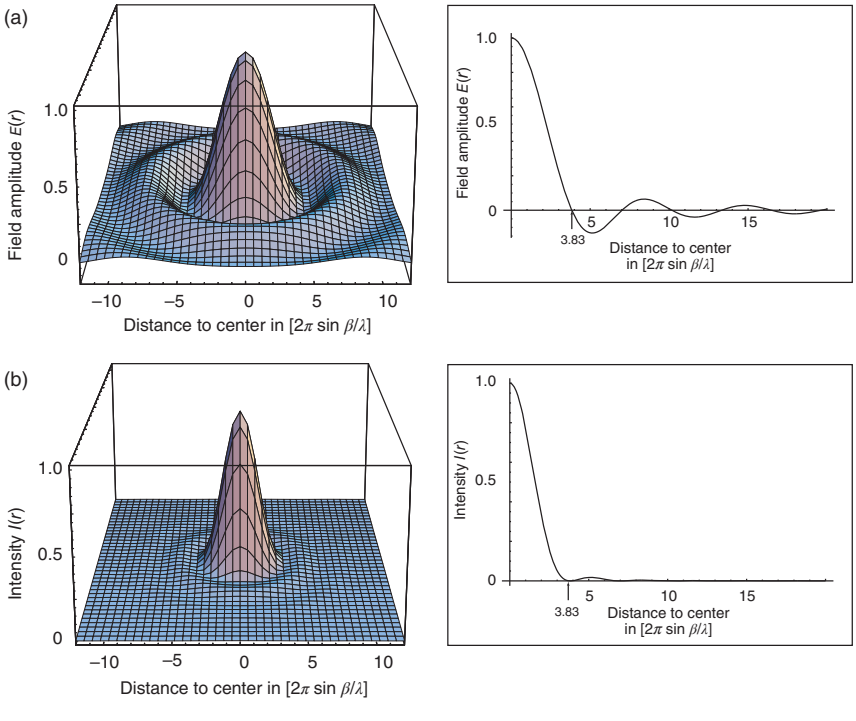


Figure 2.9 Intensity and field distribution at the focal plane of the tube lens when imaging a self-luminous point object. The surface plots illustrate the distribution of the (a) field amplitude $E(r)$ and (b) the light intensity $I(r)$ in the image plane as a function of the distance to the optical axis in optical units when a single point object is imaged. The latter function, $I(r)$, is called the *point spread function* (PSF) of the lens system. Notably, the intensity distribution in (b) is identical to the light distribution in the focal plane of a lens when illuminated with a plane wave.

filter for the spatial frequencies of the object. Fine object details are smeared because point objects are imaged as Airy patterns. Microscopic imaging means to smooth reality.

Box 2.3 PSF and OTF

Let us recapitulate the imaging of a point object. Illumination produces a Fraunhofer diffraction pattern. Mathematically, this corresponds to the Fourier transform of the point object. In Fourier theory, it is trivial to show that the Fourier transform of a point – or a δ-function – is a constant, because it contains all frequencies. The objective lens collects only a limited fraction of the diffraction pattern and projects it into its back focal plane. Therefore, the spatially extended, complete Fourier transform in the back focal plane is restricted to the set of frequencies below the limiting frequency. It contains only those frequencies that can be transmitted by the lens. The inverse Fourier transformation of this is done by the tube lens and corresponds to the image of a point. This image is the so-called PSF. Its Fourier transform in the back focal plane of the objective lens is the instrument function of the microscope, which is also called the *optical transfer function* (OTF). As indicated in Box 2.1, the imaging process of an arbitrary object $A(x, y)$ can be formulated as follows: The objective lens projects the object onto its back focal plane which corresponds to the "frequency space." Here, the complete transform of the object function, $\tilde{F}[A]$, is limited to the transmittable frequencies. Mathematically, this can be formulated as a multiplication with the OTF, which contains the information about the transmittable frequencies.

The multiplication yields a truncated and modified version of $\tilde{F}[A]$, which is designated as $\tilde{F}'[A]$:

$$\tilde{F}'[A] = \tilde{F}[A] \cdot \mathrm{OTF} \tag{2.25}$$

According to Fourier theory, a multiplication in frequency space corresponds to a convolution in object space. Thus, convolution of object function $A(x, y)$ and PSF yields the image $A'(x, y)$:

$$A'[x, y] = A[x, y] \otimes \mathrm{PSF}(x, y) \tag{2.26}$$

The multiplication of $\tilde{F}[A]$ and OTF leads to a deletion of the high frequencies in the Fourier spectrum of the object. Thus, the imaging process corresponds to a low-pass filtering of the object. An objective lens is a low-pass filter that blurs fine structures of the object. This is further illustrated in Appendix A.

2.3.4 *Lateral and Axial Resolution*

2.3.4.1 Lateral Resolution Using Incoherent Light Sources

Diffraction at the aperture of a lens system is the reason for the resolution limit of optical imaging. The image of a luminous point object corresponds to the intensity distribution $I(r)$ (Eq. (2.24) and Figure 2.9b), which is known as the *Airy pattern*. In that pattern, the circular area with a radius defined by the distance between the central maximum and the first zero is designated as the Airy disk.

The zero of the Bessel function $J_1(x)$ is at $x = 3.83$, therefore the Airy disk's radius in the image plane, $r_{im,0}$, is given by

$$r_{im,0} \approx \frac{0.61\lambda}{\sin\beta} \tag{2.27}$$

In the image plane, 84% of the total light of the complete distribution is found within the distance $0 < r < r_{im,0}$. To find out to what distance $r_{obj,0}$ corresponds in the object plane, Eq. (2.27) must be divided by the magnification factor M. The value of $r_{im,0}$ corresponds in object space to $r_{obj,0} = r_{im,0}/M$. When taking the sine condition of the objective/tube lens system into account (see box 2.5)

$$M \sin\beta = \sin\alpha \tag{2.28}$$

and considering that there often is an immersion medium with a refractive index n in object space, which requires the use of λ/n as wavelength, we obtain for $r_{obj,0}$

$$r_{obj,0} \approx \frac{0.61\lambda}{n \sin\alpha} \tag{2.29}$$

where α denotes the opening angle of the objective lens divided by 2. Note that $n \sin\alpha$ is the NA of the objective lens.

Now, we look at the image that is produced by two neighboring point objects at a distance d. First, we consider incoherent light sources, for example, two single fluorescent molecules. Their light cannot interfere, as the emitters are independent of each other. The two point objects are not imaged as dots but rather as "smeared" intensity distributions, as shown in Figure 2.10. The total resulting light distribution is obtained by adding up the two intensity distributions given

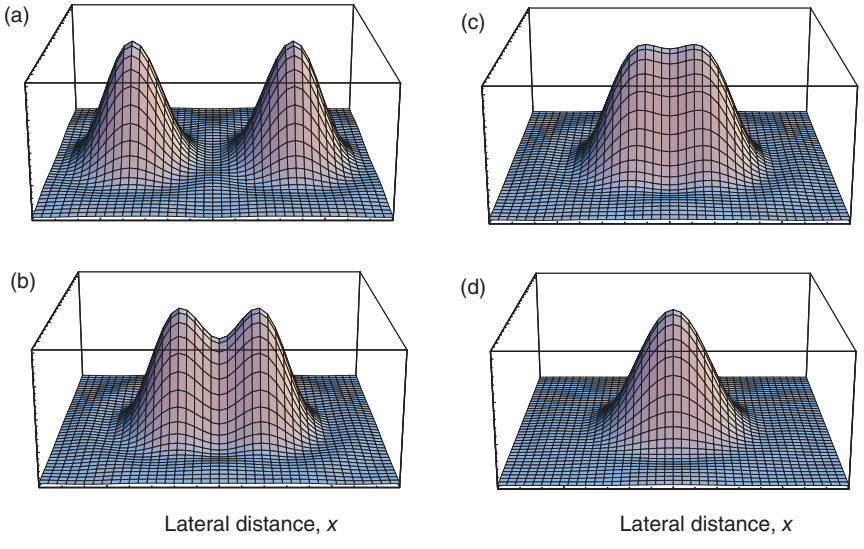


Figure 2.10 Surface plot of the intensity distribution in the image plane of two incoherent point light sources. Images of two point objects separated from each other by (a) a distance much greater than the Rayleigh distance d_R (Eq. 2.30), (b) a distance equal to d_R, (c) a distance corresponding to the full-width at half-maximum (FWHM) of the PSF, which is designated as the "Sparrow criterion," and (d) a distance of $d_R/10$, at which the images cannot be discriminated from each other.

by Eq. (2.24). Two well-separated maxima can still be observed when $d > r_{obj,0}$ (Figure 2.10a). If the point objects are situated close to each other with a separation d, and $d = r_{obj,0}$ in object space, their Airy disks clearly overlap in the image plane (Figure 2.10b). In this case, the signal intensity in the center between the two maxima is about 75% of the maximum intensity. We say that two objects are resolvable when they can still be discriminated from each other in the image. The exact distance at which the resolution is lost depends on the ability of the detector or observer to discriminate intensity levels and on the signal-to-noise ratio (Figure 2.10c). Therefore, the exact definition of a resolution distance is somewhat arbitrary. Lord Rayleigh suggested the distance $d = r_{obj,0}$ as a criterion for optical resolution, because in that case the human eye can still perceive two separate image points. This condition is called the *Rayleigh criterion*. If the distance between the objects is much smaller than that value, they can no longer be perceived as separate objects (Figure 2.10d). The optical resolution d_R according to the Rayleigh criterion is therefore

$$d_R = \frac{0.61\lambda}{n \sin \alpha} \tag{2.30}$$

The smaller the distance d_R, the higher the resolving power of an optical system. This formula explains why using light of short wavelengths – blue or even ultraviolet (UV) – gives a superior resolution compared to long wavelengths – red or infrared. Utilization of high NA objectives and immersion media with higher refractive indices such as water, glycerol, or oil in combination with suitable objectives enhances the resolution further.

2.3.4.2 Lateral Resolution of Coherent Light Sources

The situation is different for two coherent point light sources. Such light sources could be formed by two small structures in the same plane of a microscopic sample that are illuminated by a plane wave. Then, the diffracted light waves from the two sources have identical phases and may interfere with each other in the image plane, and the resulting field distribution is determined by the sum of the two field distributions according to Eq. (2.23), as shown in Figure 2.9a. The observed light intensity distribution is now obtained by squaring the sum of the field contributions, and not by summing the squares of the individual field contributions as in the case of self-luminous point objects as discussed earlier. This produces a very different image than for incoherent point objects, as shown in Figure 2.11.

In Figure 2.11, the image of two coherent point light sources is shown, which are separated in the object plane by a distance d from each other. The image is quite different from the incoherent case (Figure 2.10). In particular, for $d = r_{obj,0}$, as defined in Eq. (2.30), the two image points are clearly not resolvable. Rather, a larger distance d with

$$d = \frac{1.22\lambda}{NA_{obj}} \tag{2.31}$$

is required to observe a minimum of $\sim 0.75 I_0$ between the two maxima. This value might be used as a "Rayleigh-type" resolution limit for two coherent point light sources. Obviously, the optical resolution is reduced for coherent objects by a

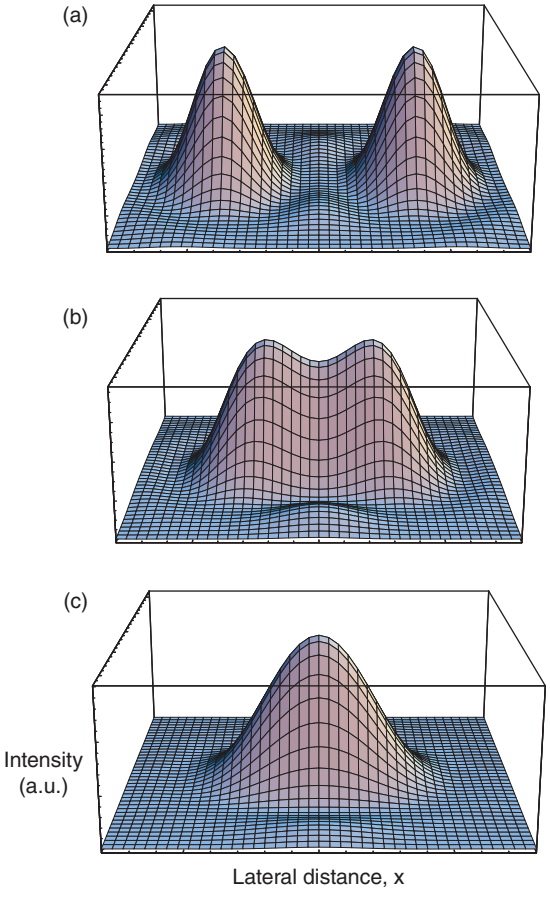


Figure 2.11 Surface plot presentation of the intensity distribution in the image plane of two coherent point light sources. Airy disks and lateral resolution: (a) large distance more than twice the Rayleigh distance ($2d_R$), (b) a well-resolved distance, $d_R = 1.3$, and (c) a distance that corresponds to the Rayleigh distance d_R (Eq. (2.30)). Coherent point objects at that distance cannot be resolved anymore.

factor of 2. This is an important fact because coherent objects are frequent in microscopic imaging. Closing the condenser aperture in bright-field illumination produces a pointlike light source. The light passing through the specimen is coherent, and so are the diffracted waves. The lateral resolution in this case would be given by Eq. (2.31).

So far, we assumed that the illuminating wave propagates along the optical axis and hits the structures in the object plane simultaneously. In that case, the diffracted waves have no phase shift with respect to each other. However, when the incoming wave propagates at a certain angle in relation to the optical axis, the situation is quite different. Then, the diffracted Huygens wavelets originating in the object plane have a specific phase shift with respect to each other because they are produced at different time points depending on their positions in the sample. This phase shift in the diffracted light is still present in the image plane where the light is focused and the image is formed by the interference of the coherent and phase-shifted waves. Now, the two field amplitudes must be summed as earlier, but their relative phase shift must be taken into account.

When a sample is illuminated by transmitted light with an extended light source, for example, by a completely opened condenser aperture, then we

have numerous distinct illumination point light sources. The illumination by these sources is of varying angles when approaching the limit of the aperture. Each source on its own illuminates the sample and creates coherent, diffracted light from the sample. With respect to each other, however, these sources are incoherent, because they correspond to different loci on the illumination lamp filament, and therefore the light distributions of all these produced in the image plane must be summed together. The sum must be performed in an appropriate way over the intensity distributions produced by the various oblique illumination angles. It must be taken into account that the contribution of the more oblique illumination directions is greater owing to their larger number in the outer part of the diaphragm.

The final image depends on the detailed mode of illumination or, in other words, on the degree of coherence in the illumination light, which can be regulated by the size of the condenser aperture. In exact terms, it is the NA of the condenser illumination, NA_{cond}, that determines the degree of coherence of the illumination light. The final resolution, according to the Rayleigh criterion, can be formulated as follows:

$$d = \frac{1.22\lambda}{NA_{obj} + NA_{cond}} \tag{2.32}$$

For bright-field illumination with a condenser, the achievable resolution depends on the properties of the illuminating lens and its NA. Closing the condenser aperture – which reduces light scattering in the sample – diminishes the attainable optical resolution because the overall coherence of the illumination light increases. One has to choose between reduced stray light and higher depth of focus on one hand, and a reduced lateral optical resolution on the other.

2.3.4.3 Axial Resolution

So far, we discussed the light intensity distribution exactly within the image plane. However, image formation is a 3D process. The light waves travel toward the image plane through the space in front of the objective lens – the object space – through the objective lens, the space between the objective and tube lens, and the tube lens. Then, the waves are focused within the image space into the image.

They not only interfere with each other exactly in the image plane but also fill all the space behind the tube lens producing a complex 3D intensity distribution of light. The Airy pattern describes the lateral intensity distribution exactly in the primary image plane. Its mathematical formulation (Eq. (2.24)) can be derived analytically in the so-called paraxial approximation. This means that it is valid only in a small region close to the optical axis of the object and image space and that the considered angles with regard to the optical axis must be small. Using paraxial theory, the complete 3D light intensity distribution near the image center of a self-luminous point can be analytically calculated [6]. A graphical representation is shown in Figure 2.12.

The distribution is given as a function of the distances to the image center in normalized optical units. These correspond to the expressions occurring in the functions describing the lateral and axial light distribution (for details,

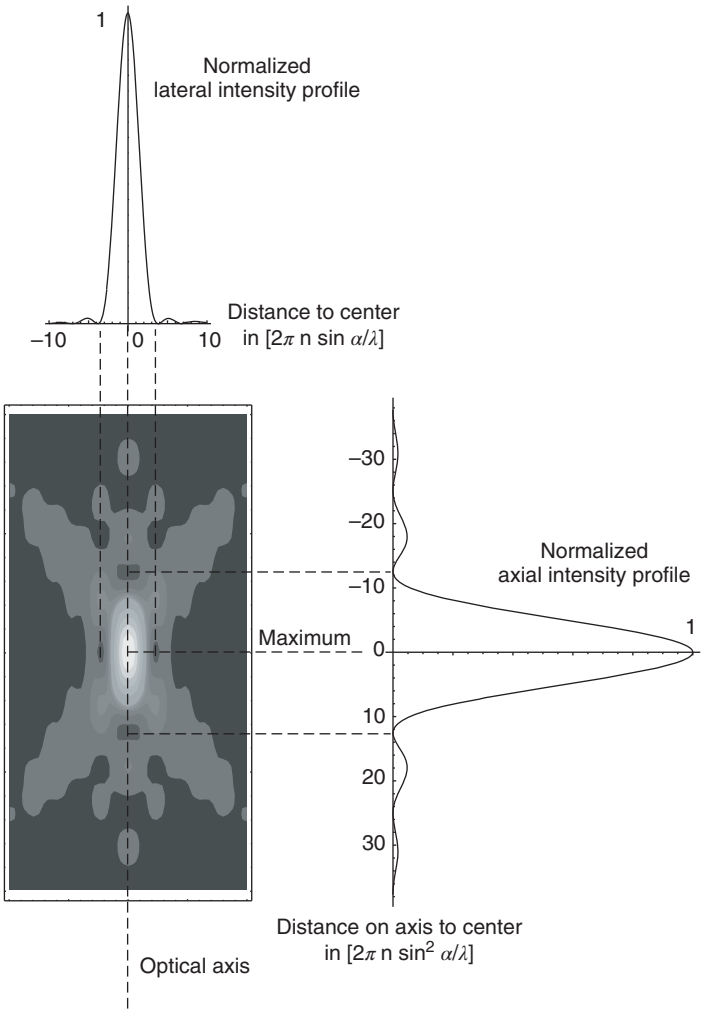


Figure 2.12 3D light intensity distribution according to paraxial theory.

see Box 2.4). The intensity distribution along the optical axis can be described analytically, and the distance between the intensity maximum and the first minimum on the optical axis, $z_{0,\mathrm{im}}$, can be used as an axial resolution criterion in analogy to the Rayleigh criterion. In image space, for low NA_{obj} $z_{0,\mathrm{im}}$ is given by

$$z_{0,\mathrm{im}} = \frac{2M^2\lambda}{\mathrm{NA}_{\mathrm{obj}}^2} \tag{2.33}$$

Because $z_{\mathrm{im}} = z_{\mathrm{obj}}M^2/n$, this corresponds to an axial distance $z_{0,\mathrm{obj}}$ in the object space of

$$z_{0,\mathrm{obj}} = \frac{2\lambda n}{\mathrm{NA}_{\mathrm{obj}}^2} \tag{2.34}$$

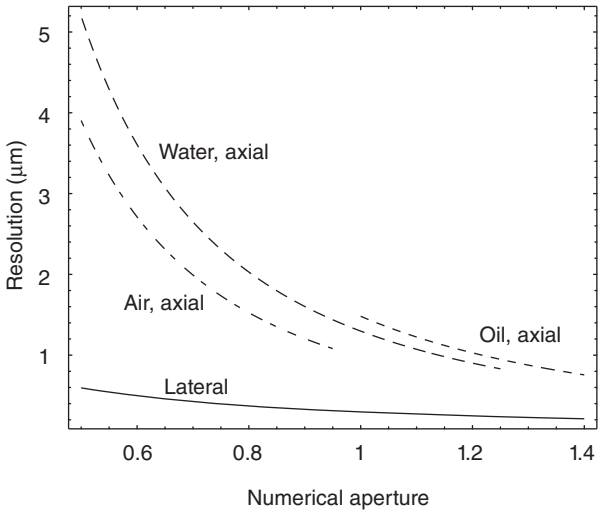


Figure 2.13 Axial and lateral resolution as a function of numerical aperture for air-, water-, and oil-immersion objectives. The curves were calculated according to Eqs (2.30) and (2.34) for indices of refraction of 1 for air, 1.33 for water, and 1.515 for oil, and for $\lambda = 488$ nm.

Practically, this is the vertical distance by which the object stage with a self-luminous point object must be moved so that the image changes from the central intensity maximum to its first minimum. While the lateral resolution depends on the reciprocal of the NA, the axial resolution is related to the reciprocal of NA^2. The dependence of the lateral and axial resolutions on the NA for different immersion media is shown in Figure 2.13. Indeed, water- and air-immersion objective lenses exhibit a higher axial resolution than oil-immersion lenses at identical NA, whereas their radial resolutions do not differ.

The given formula for the axial resolution is valid for lenses with a small NA because it was derived on the basis of paraxial optics. A more general description can be obtained by using the Fresnel–Huygens theory, which works with fewer approximations and gives a more realistic 3D distribution of light intensity in the focal region for high NA lenses.

Box 2.4 Lateral and Axial Optical Units

The definition of optical units results from the derivation of the intensity distribution at the focus of a lens as derived by paraxial theory in Born and Wolf [6]. The intensity distribution in the lateral plane was given by Eq. (2.24) and is plotted in Figure 2.9. The argument of the function is the dimensionless variable υ

$$\upsilon = r\frac{2\pi}{\lambda_0} n \sin\alpha \tag{2.35}$$

(Continued)

Box 2.4 (Continued)

The intensity distribution along the optical axis can be derived as

$$I(u) = I_0 \frac{\sin^2 u}{u^2} \tag{2.36}$$

The function $\sin u/u$ is also designated as the *sinc function*. The argument u of this function in the object space is [6]

$$u = z\frac{2\pi}{\lambda} n \sin^2 \alpha \tag{2.37}$$

It was shown, however, that the latter expression leads to erroneous results for deviations from paraxial theory, for example, in the case of a large NA. The large discrepancy can be corrected by a simple redefinition of u as follows [7]:

$$u = z\frac{8\pi}{\lambda} n \sin^2 \frac{\alpha}{2} \tag{2.38}$$

The coordinates u and v are designated as "optical units" and allow the description of the intensity distribution in the PSF independent of the NA of the lens and of the refractive index n of the medium.

2.3.5 Magnification and Resolution

Earlier we have seen that the lateral image of a point object corresponds to an Airy pattern. Often the image is recorded by a digital device with a certain detector element size d_d, for example, by a CCD camera (see Chapter 3). It can be shown that the overall magnification should be chosen such that the radius of the Airy disk, d_{Airy}, in the image plane exceeds the size of a single detector element d_d at least by a factor of 2 if the resolution of the microscope is to be fully exploited. As demonstrated in Figure 2.14, all available information in the image can be recorded sufficiently well by the image detector when $d_d \approx d_{Airy}/2$.

The optimal magnification factor, therefore, depends on the resolving power of the optical system and the type of the detection system, more exactly on the size of a single detector element, for example, the edge length d_d of the photodiodes of a CCD.

Thus the optimal magnification $M_{opt,d}$ should be approximately such that

$$M_{opt,d} \approx \frac{2d_d}{d_R} \tag{2.39}$$

where d_R is the optical resolution attainable by the lens system as defined in Eq. (2.30). Hence, using a microscope with a lateral optical resolution $d_R = 0.25\,\mu m$ and a detector comprising elements with a size of $6.25\,\mu m$, the optimal magnification would be $M_{opt,d} \geq 50$.

The situation is different for visual inspection of the microscopic image. The minimum visual opening angle for discriminating two object points by the eye should be ≥ 2 arc min or 10^{-3} rad. Placing an object with the size $d_R = 0.25\,\mu m$ at the comfortable viewing distance of 25 cm would produce only a viewing

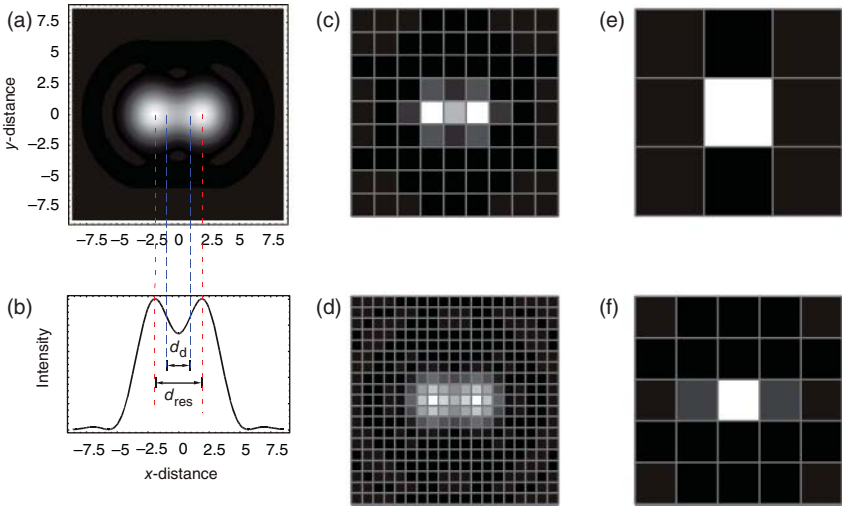


Figure 2.14 Optical resolution and detector element size. (a) Image of two point objects separated by the Rayleigh distance d_{res}. (b) Line profile through the horizontal center of the image. The points are optically well resolved. (c) Image created with a detector element size $d_d = M \times d_{res}/2$. M is the overall magnification factor. In the digitized image, the two objects are separated. (d) Detector size $d_d = M \times 0.24 \times d_{res}$. Details of the intensity distribution of the two object can be recognized. This situation is designated as "super resolution." If $d_d > M \times d_{res}/2$, the digitized image does not preserve the optical resolution as shown in (e) $d_d = 1.5 \times M \times d_{res}$ and (f) $d_d = 0.9 \times M \times d_{res}$.

angle of 10^{-6} rad. Therefore, the total magnification for visual observation M_{eye} should be

$$M_{eye} \approx \frac{10^{-3}}{10^{-6}}\text{rad} = 1000 \tag{2.40}$$

This magnification can be generated by a combination of two magnification stages, where the first stage comprising objective and tube lens magnifies 100-fold and the second comprising the eyepiece and the eye magnifies 10-fold. Magnifications above the values defined by Eqs (2.39) and (2.40) are designated as *empty magnifications* because no new information is obtained.

2.3.6 Depth of Field and Depth of Focus

The axial extension of the PSF represents the achievable resolution along the optical axis. It is also called the *depth of field*, D_{ob}, and is twice the value that is returned by Eq. (2.34). It corresponds to the axial distance by which a thin object can be moved axially without significantly losing its sharpness. The depth of field depends on the NA of the objective lens and the refractive index of the medium. The *depth of focus*, D_{of}, is the corresponding distance in the image space. It is obtained from D_{ob} by multiplication with the square of the magnification factor M and, therefore, is usually relatively large, namely, in the range of several millimeters for a magnification of 100.

2.3.7 Over- and Undersampling

There exist some special imaging applications where magnification factors well above M_{opt} are useful. This situation is called *oversampling.* Obviously, no structural information beyond the optical resolution can be obtained from optical imaging. It is possible, though, to locate objects with sizes far below the resolution limit such as single molecules with a precision much smaller than the optical resolution. For this purpose, it is necessary to have well-resolved images of the complete PSF. Usually, 3–7 pixels are used to sample the diffraction-limited signals in each direction with sufficient detail in these super-resolution approaches (Chapters 8 and 12).

Other imaging situations exist where it is not essential to acquire data visualizing all structures with the utmost possible resolution. For example, it may be sufficient to just count the number of cells within a given field of view for a specific experiment. Then, it is not required to image at high resolution. A relatively crude image would be sufficient, which is designated as *undersampled.*

2.4 Apertures, Pupils, and Telecentricity

In the last sections, we discussed the capability of optical microscopes to produce magnified images of specimen. This is only one part of the use of microscopes. In order to generate high-quality images, we have to ensure that the light waves forming the image finally reach the image plane. This is not a trivial task in such a complex optical system such as a microscope, which comprises several magnification stages and often additional relay stages.

The amount of light that is collected by a lens is obviously limited by the physical diameter of the lens. Such apertures define the amount of light that is available for image formation. Apertures are principally independent of the focal length of lenses, which defines the imaging magnification as we saw earlier.

In simple cases, it is the edge of a lens or its mounting that defines the "entrance" aperture which *a priori* limits the amount of light that can pass to the image plane. In complex imaging systems such as an objective lens constructed of multiple lenses, it is usually the diameter of the smallest lens or an extra, sometimes adjustable aperture stop that limits the lateral extension of the beam path. The image of that aperture as seen from the object defines the maximum amount of light that contributes to the image. The image of the aperture as seen from the on-axis point of the object is called the *entrance pupil.* Similarly, the limiting aperture as seen from the on-axis point of the image is called the *exit pupil*. In case the aperture is located in front of the lens in the object space, the virtual image of the aperture represents the exit pupil (Figure 2.15a). All light passing through the entrance pupil also passes through the exit pupil, because they are conjugate to each other and to the aperture stop in cases where it is not identical with one of the pupils.

Objective lenses are constructed in a special way: their beam-limiting aperture stops are located in their back focal plane, which therefore are simultaneously their exit pupils. Accordingly, the entrance pupil is located in infinity (Figure 2.15b). This construction has a number of very significant consequences.

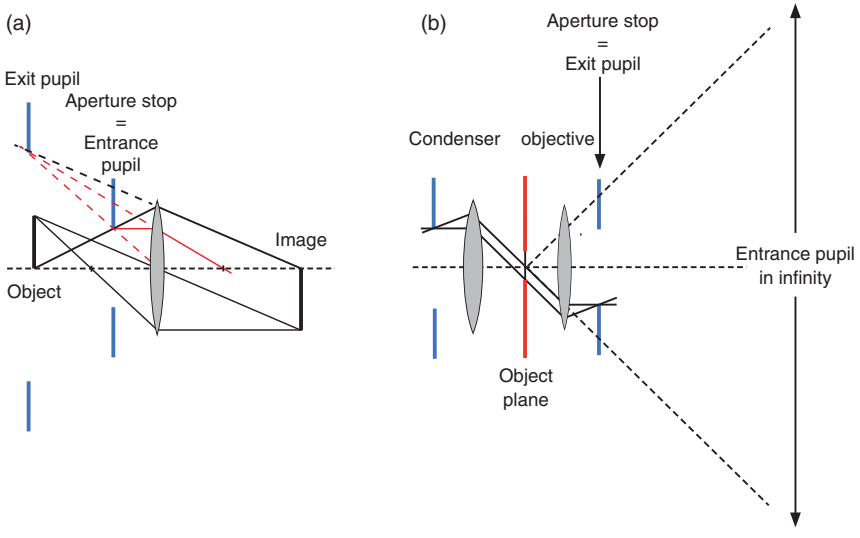


Figure 2.15 Aperture stop and pupils of the objective lens. (a) The aperture stop is in front of the lens, and hence identical to the entrance pupil. The exit pupil is the virtual image of the aperture as seen from the on-axis point of the image (note the black full dashed line from that point to the aperture image). (b) The aperture stop and exit pupil of an objective are located at its back focal plane. This causes the entrance pupil to be at infinity in the image space. In addition, the entrance pupil of the condenser lens and the exit pupil of the objective lens are conjugate to each other in Köhler illumination.

For all microscope systems, the position of the image plane is well defined and located in the front focal plane of the eyepiece. In order to understand the importance of the exit pupil's position, we first consider the case where the aperture stop is *not* in the back focal plane of lens but, for example, in front of the lens (Figure 2.16a). In that case, defocusing the specimen has two simultaneous effects: the image becomes blurred in the shape of the aperture stop, and the blurred image is reduced or magnified in size. This is demonstrated in Figure 2.16b. Clearly, such a magnification change due to defocusing would be inconvenient and very confusing for the observer.

Now we modify the optical setup and place the exit pupil at the back focal plane of the objective lens and thus the entrance pupil at infinity. Now the chief ray runs parallel to the optical axis in the object space. Such a lens configuration is called telecentric in the object space (Figure 2.16c). We note the important consequence that defocusing of the object results in a blurred image in the (fixed) image plane, but no magnification change occurs (Figure 2.16d). This effect is of great importance for the practical work with microscopes.

In transmission microscopy, usually a Köhler illumination mode is employed. For Köhler illumination, the entrance pupil of the illumination system is realized by the condenser aperture, which is located in the front focal plane of the condenser lens. We already noted that this aperture is imaged onto the back focal plane of the objective lens. This plane also contains the exit pupil of the imaging system. Obviously, it is optimal to match the sizes of these two pupils, because

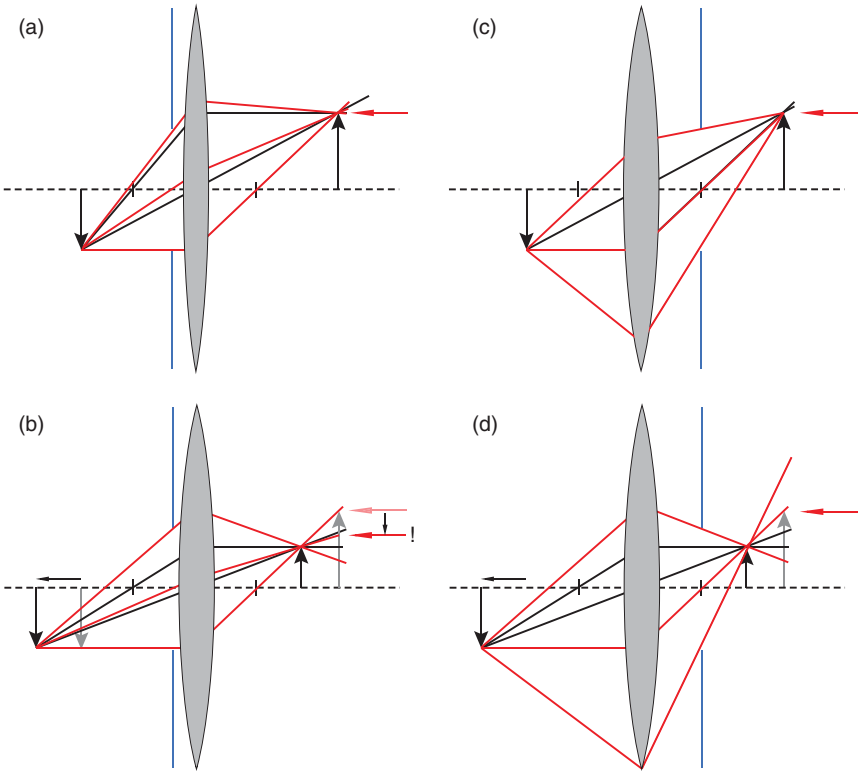


Figure 2.16 Telecentricity. (a) Imaging process of a lens with an aperture stop – the entrance pupil – in the object space. The chief ray is the one that passes from the outmost object position (tip of black arrow) through the center of the aperture. The chief ray represents the cone of light creating the image of that point, and its course in the image space marks the center of the defocused image. In the sketches, the chief ray is the red central ray. (b) When the object is defocused, the chief ray moves downward on the fixed image plane compared to its position in (a), as indicated by the arrows. (c) Imaging by a lens that is telecentric in the object space. According to the definition for such lenses, the aperture stop is located at the back focal plane. The chief ray now passes through the back focus and is parallel to the optical axis in the object space. (d) For telecentric lenses, defocusing the object does not alter the position of the chief ray, as it remains parallel to the optical axis. It hits the image plane at the same position as in (c) (see position of red arrow in (c) and (d)).

only then all incoming illumination light will pass the exit pupil and be available for image formation. Matching the size of these pupils, however, implies the existence of a defined position of the exit pupil of the objective lens and of its entrance pupil, which is in infinity. The position of these pupils should be identical for all objective lenses in order to avoid an axial repositioning of the condenser lens when switching objectives. Therefore, the location of the entrance pupil at infinity, or correspondingly of the exit pupil at the back focal plane of the respective objective lens, is very convenient because it provides the required fixed position. Switching from one objective lens to another with a different magnification does

not require a repositioning of the condenser. Only the size of the condenser aperture has to be readjusted to the new exit pupil diameter.

A further very important advantage of the exit pupil being at the back focal plane – and its image being in the front focal plane of the condenser – is its use in several important transmitted light contrasting techniques, as discussed in Section 2.2.6. In epi-illumination setups such as in fluorescence microscopes, the entrance pupil of the illumination light path and the exit pupil of the imaging path both fall together at the same position and are at the back focal plane of the objective lens. Therefore, their sizes are intrinsically matched.

2.5 Microscope Objectives

The *objective* is the central component of a microscope. It comprises up to 20 different single convex and concave lenses with different radii of curvature. For optimal performance, different glass substrates and cements are used within one objective. To avoid any aberrations on the resulting image, all these single lenses have to be mounted with very high precision. While this is already demanding in high-end objectives, it becomes decisive when designing and manufacturing objectives with movable lens groups as one can find in objectives with a correction collar. Therefore, objective design is closely associated with precision mechanics.

The objective is the most sophisticated optical component of the microscope. Its function is to collect light from the object and to form the primary real image in combination with the tube lens. The image should be as free from aberration and the resolution as high as possible. The objective achieves the major part of the magnification of the microscope and often creates a specific image contrast. Today, there exists a large variety of different objectives with widely differing properties for different purposes. This makes the choice of the appropriate objective for the specimen of interest essential for satisfactory imaging results.

2.5.1 Objective Lens Design

The objective lens determines the overall imaging quality of the complete microscope. The formation of an ideal image, however, is in practice impaired by monochromatic and chromatic aberrations. Monochromatic aberrations occur because most lenses have spherical surfaces. In addition, chromatic aberrations basically cannot be avoided because the refractive index of transparent media is a function of the wavelength, and therefore aberrations arise when light with different wavelengths is used. The type and extent of aberrations in imaging is dependent on the object field size, the focal length, and therefore the magnification, along with the width and inclination of the incident beams.

The naturally occurring aberrations can be reduced by combining different types of glass materials with different dispersion properties, radii of curvature, thicknesses, and spacing between the different lens elements. These are the parameters that define the imaging properties of the composite system. For each single aberration type, there exists a strategy to reduce its effect by combining

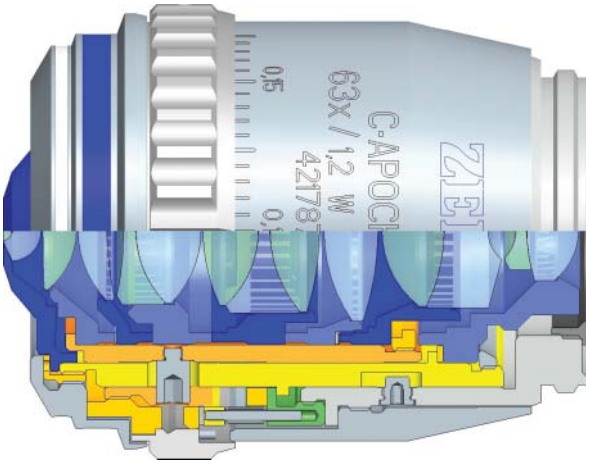


Figure 2.17 Objective lens system. Modern objectives are master pieces of optical and precision mechanical engineering. The graphics shows a 63 × /1.2 NA C-Apochromat objective. This objective is apochromatically corrected from the UV through the visible to the near-infrared spectral range. It features a correction collar for compensating differences in cover glass thickness and/or the refractive index within the specimen. The path of specific rays and further details are shown in Figure 2.18. (Reprinted by permission from the Carl Zeiss Microscopy GmbH.)

lenses with different parameters. For example, axial chromatic aberration can be corrected by combining two lenses made of different glasses with different color dispersions. This strategy brings two selected colors to exactly the same focus, but other colors will show slight deviations from this focal point. These residual aberrations are called the *secondary spectrum.*

The combined elimination or reduction of multiple aberrations is partly a science and partly a genuine art, which requires a lot of optical knowledge, intuition, and experience. For high-power objective lenses, this results in an arrangement of many different lenses into a complex lens system (Figure 2.17).

The first step of designing a lens is to define the desired properties such as magnification, field number, NA, wavelength region, immersion medium, working distance (WD), and the degree of aberration correction. Then the principal setup of the lens is selected. This is done by referring to existing lens designs with similar desired properties or by the creation of a completely new design. This step represents the most creative and important phase, but also very difficult one. The arrangement of each lens element, the corresponding allotted refractive powers, and the choices of the materials define not only the image quality but also the time required for the design, which might extend over weeks and months, and the costs of manufacturing.

The detailed imaging properties of the lens system are studied by computerized ray tracing. Objectives built from several thick lenses cannot be treated any more by the simple Gaussian imaging laws which are valid only for paraxial rays. In the design and analysis of objectives, these simplifying assumptions are not valid and must be replaced by considering the thickness of lenses and their aberrations. For

ray tracing, representative points of the object plane with a number of light ray bundles of appropriate colors emerging from them are chosen. Then, the rays are traced by calculating how they are refracted at each single lens surface according to the law of refraction. Here, the exact optical properties of the respective lens materials and the wavelength-dependent index of refraction are considered. The rays in the bundles are chosen such that the complete exit pupil of the system is scanned, and any ray can exactly be traced through the optical system. Therefore, ray tracing with dedicated computer programs is today the chief tool of optical designers.

By ray tracing, the imaging process can be simulated and evaluated. In the early stage, deviations from the desired imaging properties may be substantial, which must then be reduced by adjusting the lens parameters or by the addition of new elements. Sometimes, the start design may have to be dropped. If a design comes close to the desired one, several iterations of parameter optimization to achieve further aberration reductions are performed. For microscope objective lenses, the correction of most aberrations has to be done rigorously up to the resolution limit. Especially, the correction of spherical and chromatic aberrations is important. At this stage, the lens design requires detailed observation of the changes in the various aberrations on the computer until the design is finally optimized, which can be achieved by automatic lens design computer programs. The accumulation of many small aberration corrections completes the final design.

Example designs for five different 63× objectives are shown in Figure 2.18. The number "63×" means that the lens systems yield a 63-fold magnification when used with the corresponding tube lens. The magnification of all objectives shown is identical, but the level of aberration correction increases from left to right for the N-Achroplan, Plan-Neofluar, and the Plan-Apochromat. The last two objectives have special qualities. C-Apochromat has a high NA and is a high-performance lens designed for use with water immersion. The Alpha Plan-Apochromat has a lower correction degree, but it is a complex lens system featuring a very high NA and is designed for use in total internal reflection fluorescence microscopy (see Box 12.2).

To evaluate and compare the correction level of different objectives, quantitative measures are required. There are numerous different parameters for this purpose, for example, the mentioned secondary spectrum or the curvature of field that reports the deviations in the axial position of the images of point objects that are located in the object plane on axis and at the border of the field of view.

A comfortable general measure are the OPDs in the exit pupil. For a perfect objective, all monochromatic rays that originate at the focus have identical phases at the exit pupil. Their optical path lengths are identical, and they form an ideal parallel bundle of light rays when leaving the objective. As discussed earlier, such a performance cannot be achieved by real lenses. Rather, there occur OPDs between rays traversing the lens system in different regions of the lenses, for example, centrally or marginally. To quantify these differences, the OPDs for all traced rays in relation to a reference ray are squared and then averaged. The root of the average represents the root-mean-square-optical path length difference (rms-OPD) for a given wavelength. Obviously, residual chromatic aberrations will have an impact on the rms-OPD. Hence, rms-OPDs are normalized to the

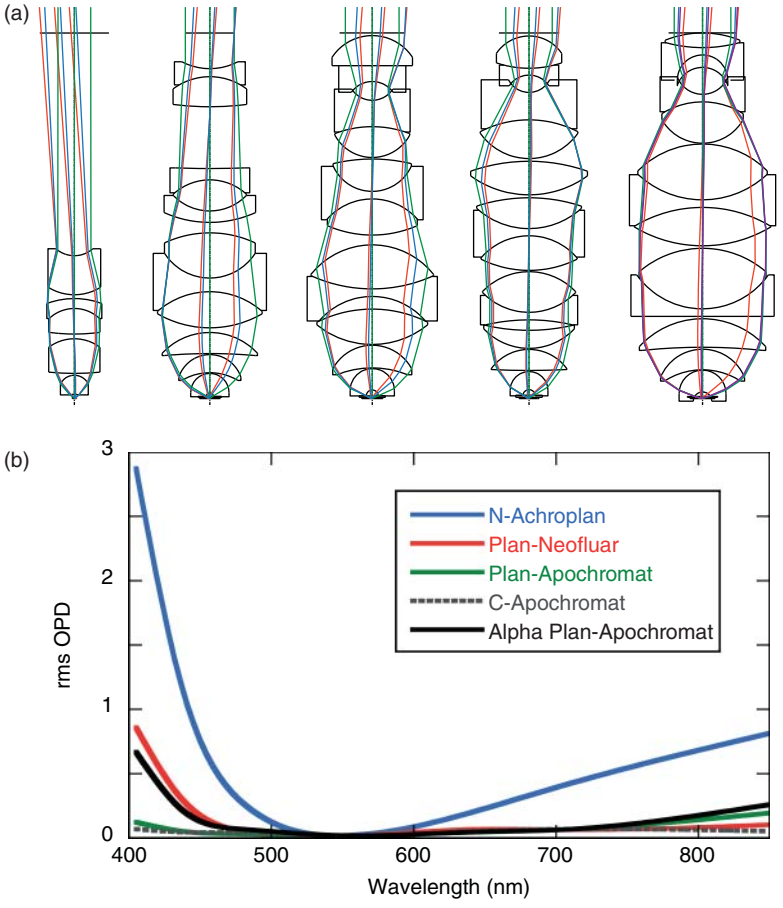


Figure 2.18 Correction of aberrations in objectives. (a) Optical design of five different 63× objective lens systems. From left to right: N-Achroplan with NA 0.85, air immersion; Plan-Neofluar with NA 1.25, oil immersion; Plan-Apochromat with NA 1.4, oil immersion; C-Apochromat with NA 1.2, water immersion; and Alpha Plan-Apochromat with NA 1.46, oil immersion. The different colors label ray bundles emerging from different positions in the focal plane. (b) Comparison of the degree of correction for the objectives shown in (a) in terms of the root-mean-square deviations of the optical path length difference (rms-OPD) in the exit pupil of the objective plotted as a function of the wavelength. N-Achroplan has the simplest design and is achromatic; Plan-Neofluar is optimized for fluorescence applications and shows good correction of chromatic aberrations in the visible wavelength region, whereas the Plan-Apochromat shows excellent chromatic aberration and a very low curvature of field. C-Apochromat has excellent chromatic correction from near-UV to the near-infrared spectral range, and shows an especially low curvature of field. Alpha Plan-Apochromat features an especially large NA and very good chromatic and planar correction. (Reprinted by permission from the Carl Zeiss Microscopy GmbH.)

wavelength and plotted for the wavelength range for which the objective was designed. Therefore, the rms-OPD curves represent a quantitative measure of the existing residual aberrations of an objective, such as chromatic aberrations, spherical aberration, and curvature of field, and are an excellent tool to compare the quality of different objectives.

Figure 2.18b shows the significant improvement of the corrections for the various lens systems sketched in Figure 2.18a. The increasingly complex design of the first three objectives (blue, red, and green curves) results in a significant reduction in the rms-OPD over the spectrum from near-ultraviolet to near-infrared light. C-Apochromat shows the smallest aberrations and appears to be the lens of the highest imaging quality over a wide range of wavelengths. Finally, it is obvious that the optical designers of the Alpha Plan-Apochromat had to sacrifice some imaging quality to create an objective with an NA of 1.46.

No real lens system is perfect, but an important question is whether the residual aberrations are below or above the intrinsic diffraction-limited effects. For pure visual observation, an aberration is not relevant when its effect is below the diffraction limit. However, such aberrations are significant for applications in the field of super-resolution microscopy. Usually, only a careful test and comparison of objectives reveals their suitability for the intended imaging tasks.

2.5.2 Light Collection Efficiency and Image Brightness

We saw earlier that any objective can capture only a fraction of the elementary waves emitted or diffracted by the object owing to its finite opening angle. This results in the limited resolution of optical systems. Furthermore, it limits the light collection efficiency of the objective, which is of course a central feature of every microscope. Ultimately, it determines the brightness of the final image.

Surprisingly, it is well known that image brightness is related to the NA of the objective – meaning to the sine of the opening angle α – and not to the opening solid angle of the objective. This would be the first intuitive guess. Hence, is this true, and why? A careful examination of this question reveals the reason and also leads to an important concept in microscopy, the so-called sine condition.

First, we consider a luminous point object that emits light into the surrounding space. The light flux that is collected by a lens is directly proportional to the surface of a spherical segment filling the opening of the lens or, in mathematical terms, to the opening solid angle as shown in Figure 2.19a. Considering the rotational symmetry around the optical axis, which may be labeled by an angle φ, we can calculate the total solid angle of the surface segment by integrating over all surface elements $\mathrm{d}S = \mathrm{d}\phi \sin\theta \, \mathrm{d}\theta$ according to

$$\Phi_{\text{point}} \propto \int_0^{\alpha} \left[\int_0^{2\pi} \mathrm{d}\varphi \right] \sin\theta \mathrm{d}\theta = 2\pi(1 - \cos\alpha) \tag{2.41}$$

The integration can be performed directly because the light of the point object falls perpendicular on all surface elements regardless of their position in space, which is characterized by the angles φ and θ. Hence, the light collection efficiency

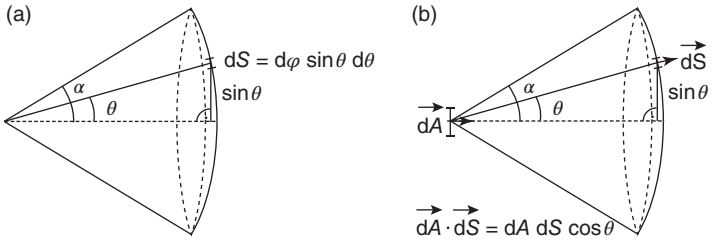


Figure 2.19 Light efficiency of objectives. (a) Imaging of a luminous point object. (b) Imaging of a luminous 2D object, a Lambert radiator.

of an objective when examining point objects is *not* related to the sine of the opening angle respectively the NA of the lens, but given by Eq. (2.41).

The situation is different, however, when a 2D extended luminous object is imaged. Now we must determine the amount of light emitted by a surface element dA and collected by the lens. Such a luminous surface is called a *Lambert radiator*. However, the effective surface of dA that is visible under different angles θ varies (Figure 2.19b). For increasing angles θ, the effectively perceivable surface becomes smaller until it is 0 for a hypothetical $\theta = 90°$. Mathematically speaking, we must now consider the scalar product $\overrightarrow{\mathrm{d}A} \cdot \overrightarrow{\mathrm{d}S}$.

This adds a factor of cos θ. If we take this into account and perform the integration over the total surface of the spherical segment, we find now

$$\Phi_{\mathrm{d}A} \propto \mathrm{d}A \int_0^{\alpha} \left[\int_0^{2\pi} \mathrm{d}\varphi \right] \sin\theta \cos\theta \mathrm{d}\theta = \mathrm{d}A \cdot \pi \sin^2\alpha \tag{2.42}$$

We note that the collected light flux is related to the size of the emitting surface element dA and the square of the sine of the opening angle of the objective. This expression is valid for air-immersion objective lenses. When a material with a higher index of refraction is located in front of the lens, such as water or oil, an appropriate immersion lens must be used. This situation is sketched in Figure 2.20.

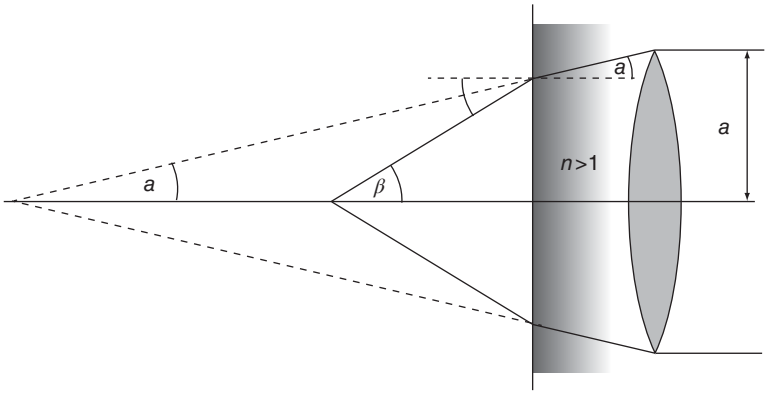


Figure 2.20 Light flux with and without an immersion medium. (Modified from [8].)

We see that the light flux entering the lens with a maximum angle α in the material with an index of refraction n would be identical to the light flux enclosed by the angle β when working in air. According to the law of refraction, $\sin\beta = n\sin\alpha$. The light flux, however, does not change. Thus, when using optically denser materials, we must multiply the sine of the opening angle by the index of refraction to get the correct expression for the collected light flux:

$$\Phi_{n,\mathrm{d}A} \propto \mathrm{d}A \cdot \pi \cdot n^2 \sin^2\alpha = \mathrm{d}A \cdot \pi \cdot \mathrm{NA}^2 \tag{2.43}$$

Hence, we see that the amount of light collected by an objective is proportional to the square of its NA if a flat 2D specimen is imaged. Of course, this is most often the case in microscopy.

Box 2.5 The Sine Condition

Considering the light flux through a lens allows further insights. To achieve a faultless image, all light collected from the surface element dA must obviously be projected onto the corresponding, magnified surface element dA' in the image plane. This is equivalent to stating that the surface element dA must be mapped exactly onto dA'. This requirement sounds trivial, but indeed it cannot be achieved by a simple spherical lens but rather requires elaborate lens systems. However, we can turn the tables: we can use this as a condition that must be fulfilled by an optimally working microscope. Considering that the image space is usually in air with a refractive index of 1, the condition can be formulated as follows:

$$\mathrm{d}A \cdot \pi \cdot n^2 \sin^2\alpha = \mathrm{d}A' \cdot \pi \cdot \sin^2\alpha' \tag{2.44}$$

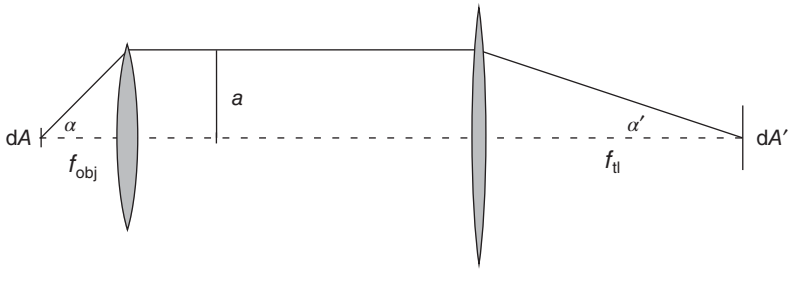
For the definition of α', see the following sketch. We can simplify this expression by considering that the size of the surface dA is related to its image by the square of the magnification factor M:

$$\mathrm{d}A' = M^2 \mathrm{d}A \tag{2.45}$$

Inserting this above, we obtain

$$n\sin\alpha = M\sin\alpha' \tag{2.46}$$

This expression is known as *sine condition* and was formulated by Ernst Abbe as a condition that a good imaging system must fulfill. Now we can also reverse the argument: the light flux for a Lambert radiator in an immersion medium with refractive index n must be given by Eq. (2.43).



(Continued)

Box 2.5 (Continued)

A further condition (used in Eqs (2.9) and (2.14)) is closely related to the sine condition. It can be illustrated as follows: For an air imaging system, as shown in Figure 2.1, M is given by the ratio of the focal lengths of tube lens and the objective (Eq. (2.3)). If we insert this into the sine condition Eq. (2.46), we have

$$\sin\alpha = \frac{f_{\mathrm{tl}} \sin\alpha'}{f_{\mathrm{obj}}} \tag{2.47}$$

Now let us increase f_{tl} more and more. The image size increases, and sin α' decreases accordingly. As α is not altered, the product f_{tl} sin α' must be a constant, and we can call it a. In a system fulfilling the sine condition, this is valid for all angles α, and if we consider small α, we can identify a. For small α, we have sin $\alpha \approx \tan\alpha = a/f_{\mathrm{obj}}$ (see sketch) and thus Eq. (2.47) yields in general

$$\sin\alpha = \frac{a}{f_{\mathrm{obj}}} \tag{2.48}$$

Hence, an incoming beam through the front focus with an inclination angle α proceeds behind the lens in the infinity space parallel to the optical axis at a distance a, where a is given by f_{obj} sin α (Eq. (2.47)). This insight is central for Fourier optics (Box 2.1), where it is the starting point for realizing that the back focal plane of a lens contains the Fourier transform of the object function [9]. It is called the *von Bieren condition*.

In summary, the light collection efficiency is normally proportional to $\sin^2\alpha$ and NA^2, and translates directly into image brightness. In the special case where single isotropic point emitters are examined, as may occur in single-molecule imaging (Chapters 8 and 12), the light collection efficiency is higher and related to $(1 - \cos\alpha)$. In Figure 2.21, the differing light collection efficiencies are illustrated

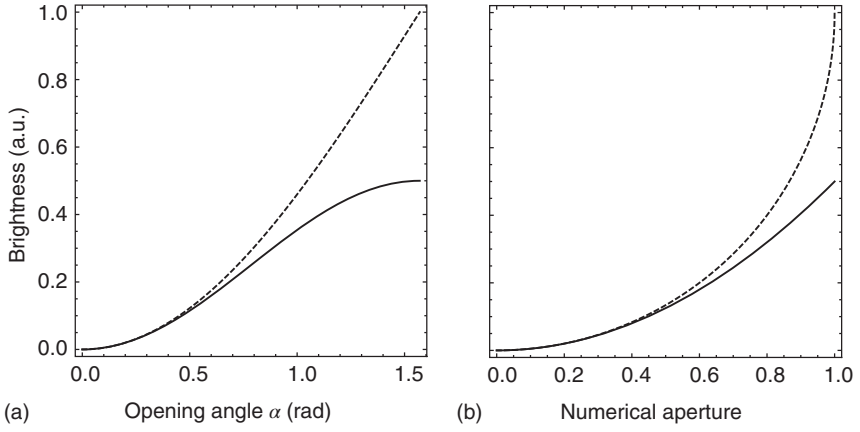


Figure 2.21 Light collection efficiency in air for a luminous point emitter (dashed line) and a flat luminous object (full line) as a function of (a) the opening angle and (b) the NA.

as a function of the opening angle α and of the NA. The importance of using high NA lenses is obvious in both cases, but especially when imaging point objects.

The collected light passes the objective's exit pupil and is projected onto the image plane. How is image brightness related to the magnification? This becomes obvious when considering the image of a 2D self-luminous object: increasing the magnification enlarges the image but reduces its luminous density proportional to the square of the magnification, because the total integrated luminosity of the image remains constant.

Combining the earlier considerations, we find that the image brightness B is proportional to the square of the NA and inversely proportional to the square of the magnification:

$$B \propto \frac{\mathrm{NA}^2}{M^2} \tag{2.49}$$

Obviously, these considerations neglect absorption and reflection effects of the lenses forming the objective, or special lens features such as, for example, the absorbing ring in phase contrast objectives. Thus, objectives with the same magnification and NA may produce images with quite different brightness values.

Finally, it should be noted that the impact of NA in fluorescence microscopy is even more significant. In this case, the illumination is achieved via the objective, which introduces a further factor of NA^2 in the irradiance in the sample plane. Consequently, the brightness of the final image is related to the used NA as

$$B \propto \mathrm{NA}^4 \tag{2.50}$$

We see that using a high NA is extremely important for achieving bright images. However, in fluorescence imaging, the dependence of the brightness on the magnification vanishes, as its effects on the illumination and imaging cancel each other.

2.5.3 Objective Lens Classes

We have seen that a very important feature of an objective is its degree of aberration correction. Two of the monochromatic aberrations, namely, spherical aberration and coma, have a diffraction-limited correction in modern objectives, which means that these aberrations are below the visibility threshold for all objective types. Curvature of field and chromatic aberrations are still the most significant aberrations, and objective lenses are classified according to the degree of correction of these aberrations. The first part of the objective class name refers to the degree of correction of the curvature of field, and the second part to the degree of correction of the chromatic aberrations. A "plan" designation is added to the lens class with a particular low curvature of field and distortion. Such objectives have a large field over which the image is well corrected. Lateral and longitudinal chromatic aberrations are of particular significance to achieve good bright-field contrast, because objects are imaged with colored fringes when these aberrations are present. With regard to chromatic correction, microscope objectives are classified into achromats, semi-apochromats or fluorites, and apochromats. These specifications are inscribed on the objective lens barrel. Only achromats are usually not designated as such.

Unfortunately, each manufacturer denotes their objective lenses in a different way, which makes distinctions and comparisons difficult. Nevertheless, the name indicates the degree of correction. Aside from the correction level, further important parameters are inscribed color-coded or abbreviated on the objectives (Table 2.1). These indicate the magnification, immersion medium, and suitability for special imaging applications. Furthermore, one or sometimes two colored rings indicate the achievable magnification and the required immersion medium (Tables 2.2 and 2.3).

2.5.4 Immersion Media

Objectives may be used in air or in combination with a special immersion medium such as oil or water. Air objectives are also designated as dry lenses. They are often used with or without a cover glass, but an NA of 0.25 is the upper threshold value for their universal use. Departure from the standard cover glass thickness of 0.17 mm is not critical for objectives with an NA below 0.4. The maximum NA of an air objective is 0.95, which corresponds to an aperture angle of ~72°. Such objectives are very difficult to use because the large aperture angle results in a very short free WD, which can be as short as 100 µm. Air objective lenses that are designed for use with cover glasses are even more critical in their handling.

Many objective lenses are designed for use with a medium of higher refractive index than air between the cover glass and the front lens of the objective. Mostly, media such as immersion oil, glycerol, or water are used. This increases the NA and thus the achievable brightness and resolution. A homogeneous immersion is beneficial, that is, a situation where the refractive indices of mounting medium, cover glass, and immersion fluid, and – ideally – the front optical element of the objective lens are matched as well as possible. This is favorable for avoiding spherical and chromatic aberrations at these interfaces. The construction of such objectives is more straightforward, because all media may be looked upon as an extension of the front lens element. At the same time, reflections at the interfaces are reduced, which improves image contrast. This implies the use of an immersion oil with a refractive index close to that of glass, the use of a glass cover slip, and the mounting of the specimen in a resin with its refractive index matched to glass and oil.

Imaging using high NA oil-immersion objectives is especially critical and requires experience and precautions. First, the refractive indices of the cover glass and oil are not perfectly matched. The objectives are optimized for the standard cover glass thickness. However, there are cover glasses with thicknesses between 120 and 200 µm. The most common is 170 µm. Even in a single batch, the thickness may vary considerably. The actual thickness should not deviate by more than ±5 µm from the indicated value on the objective for an NA > 1.3; otherwise, aberrations occur. In addition, immersion objectives are designed to be used at 23 °C. Deviation from this temperature causes minute thermal movements of the lens arrangement in the objective, which may affect the image quality. Some objectives have a mechanical correction collar that allows the adjustment of the internal position of a few lenses of the system to correct for

Table 2.1 Inscriptions on objective barrels.

Position/type	Abbreviation	Meaning
1. Correction of image curvature	—	Lower class of image curvature correction
	NPL, N plan	Normal field of view plan (20 mm)
	PL, plan, plano	Plan field of view (25 mm)
2. Longitudinal chromatic correction	—	Achromat (lower class of chromatic correction)
	FL, fluotar, neofluar, fluor, neofluor	Fluorite or semi-apochromat
	Apo	Apochromat
3. Magnification	2.5–160	Magnification in the primary image plane when used with appropriate tube lens
4. Numerical aperture	0.1–1.46	In case a range is given (e.g., 1.25–0.65), the objective has an adjustable aperture built in
5. Immersion	—	Air
	Water, W, W.I.	Water immersion
	GLY, Gly	Glycerol immersion
	Oil, HI, H	Oil immersion
	IMM	Can be used with air, water, or oil immersion (old)
	LCI	Can be used with several media, for example, glycerol, water, or oil immersion
6. Mechanical tube length	∞	Infinite mechanical tube length
	160–250	Tube length in millimeters
7. Cover glass thickness	—	May be used with or without cover glass
	0	To be used without cover glass
	0.17	Stipulated specimen cover glass thickness
8. Working distance	WD + number	Working distance in millimeters
9. Other (examples)	Phase, Ph, phaco, PC PH1, PH2, and so on	Phase-contrast objective using respective phase condenser annulus
	D	Dark-field objective
	I, IRIS, W/IRIS	Iris diaphragm built in for adjustment of NA
	UV, UV-A	UV-transmitting, down to 340 nm

(Continued)

Table 2.1 (Continued)

Position/type	Abbreviation	Meaning
	Ultrafluar	Fluorite objective for imaging down to 250 nm in the UV as well as in the visible region
	DIC	Differential (Nomarski) interference contrast
	L	Long working distance
	LL, LD, LWD	Very long working distance

Inscriptions on objectives usually indicate the degree of correction, magnification, numerical aperture, immersion medium, cover slip thickness, and often the working distance. Magnification and numerical aperture as well as tube length and cover glass thickness are usually separated by a slash, respectively. The table gives a selection of common abbreviations. These are not exhaustive, as every manufacturer uses their own designation code.

Table 2.2 Color coding for magnification.

Color code	Magnification
White	100–160×
Dark blue	60–63×
Light blue	40–50×
Green	16–32×
Yellow	10×
Orange	6.3×
Red	4–5×
Brown	2.5×
Gray	1.6×
Black	1–1.25×

If the objective shows a single colored ring, it refers to the magnification factor. Many objectives have two colored rings. In this case, the ring located near the mounting thread indicates the magnification factor, and the one near the front lens indicates the immersion medium. The magnification factor is also printed in numbers. Special features are sometimes indicated by a special color of the text inscriptions.

Table 2.3 Color coding for immersion media.

Color code	Immersion
White	Water
Orange	Glycerol
Red	Multi-immersion
Black	Oil

If two rings exist, the ring near the front lens indicates the immersion medium.

different temperatures or cover glass thicknesses. To keep the warming of the objective – and the specimen – minimal, the intensity of the illumination light should be kept as low as possible using neutral density filters or by reducing the lamp current. In case no correction collar is present, such deviations may be compensated by using specific immersion oils that counter the temperature effects inside the objective.

Several options are available for microscopic analysis of specimen in aqueous solution. The so-called water-dipping objectives are designed for direct use in aqueous media for observation of live biological specimen from above. However, an important and very widespread application of microscopy is the imaging of living cell samples using inverted microscopes. Obviously, this requires the use of cover glasses that are mounted on the microscope stage and separate specimen and immersion fluids. In this case, the inhomogeneous optical system between the front lens and specimen is unavoidable. The use of dedicated water-immersion lenses is highly recommended, which provide very good images from planes deep in the specimen. For these objectives, the significant and inevitable spherical aberration at the cover glass–water interface is corrected by the intrinsic objective design. For water-immersion objectives, distilled or at least demineralized water should be used, as it is troublesome to remove sediments from water that has dried on the objective front lens. When dedicated high-quality water-immersion objectives are not readily available, oil-immersion lenses are often used instead. This works surprisingly well for thin objects that are very close to the cover glass/buffer interface. However, image quality degrades very rapidly when focusing deeper into a water-mounted specimen with an oil-immersion objective lens. Figure 2.22 illustrates how the PSF is degraded in this case.

2.5.5 Special Applications

The WD of an objective is the distance between the objective front mount and the closest surface of the cover glass when the specimen on its other side is in focus. In case no cover glass is used, the WD is the distance to the specimen itself. As long as the immersion conditions are maintained, this also corresponds to the depth inside the specimen that can sharply be imaged.

For objectives with high resolution, the WD is quite short, because the NA and the opening angle must be as high as possible. For high-power objective lenses of 100× magnification and 1.4 NA, WD is shorter than 200 µm, whereas it goes up to 6 mm for a 10× objective with NA 0.2. For high-magnification objectives, the front part is usually spring-mounted such that it may slide into the objective barrel when the specimen is hit, in order to avoid damage to the front lens and the specimen. For some applications, very long WDs are required, for example, when the specimen is manipulated during microscopic observation, for instance, by microinjection of living cells. These objectives are called *long-distance* or LD objectives. For such purposes, dedicated objectives were developed despite the difficulty involved in achieving large NAs and a good degree of correction. As such objectives are most often built with a correction collar, they are usually relatively expensive.

Besides accomplishing pure magnification, objectives must have certain characteristic features for specific contrast techniques. Objectives for use in fluorescence microscopy must have low intrinsic auto-fluorescence, low intrinsic reflectivity, and high light transmittance down to the near-UV. These properties depend on the employed glass types and careful coating of the lens surfaces. However, today's objectives have transmission coefficients in the range of 90% to at least 360 nm. Objectives featuring high transmission below this wavelength, which are suited for work with UV light, are usually specially labeled.

Objectives that employ specific contrast techniques are specifically designed for this task. Usually this requires highly modified optical properties. These are discussed in the following section.

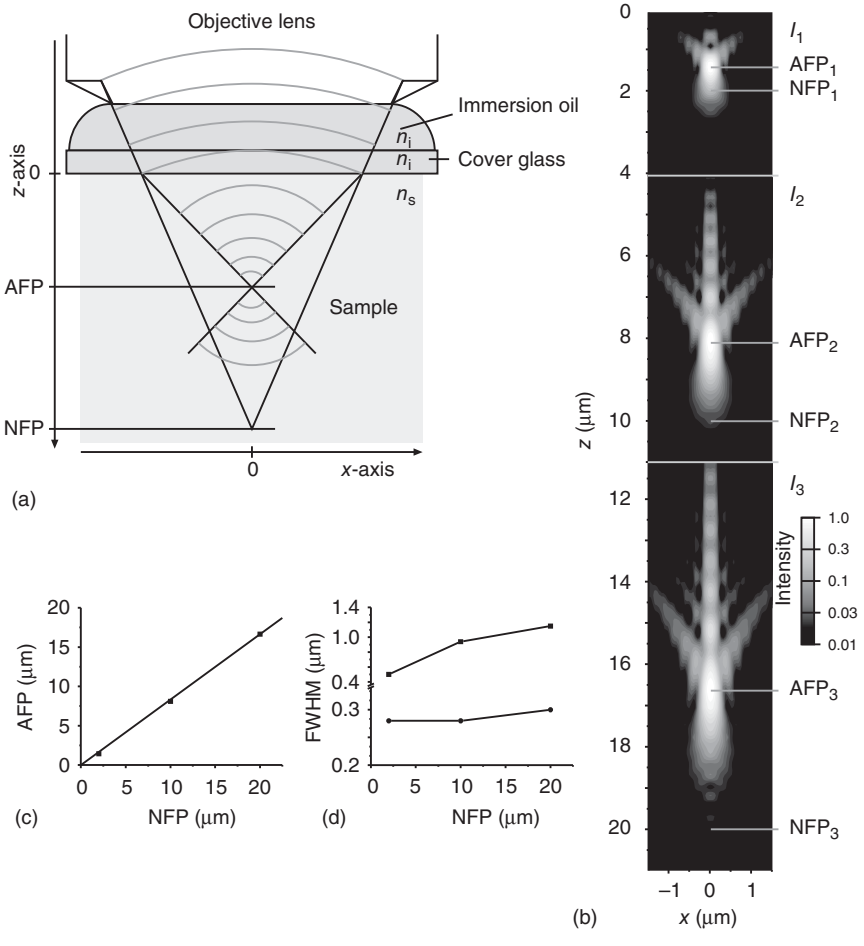


Figure 2.22 Spherical aberration at the cover glass when using an oil-immersion objective for imaging in water. The degradation of the PSF as a function of the distance to the cover glass was computed by Dr Christian Verbeek according to [10]. (a) The immersion oil and the cover glass have a matched refractive index of $n_i = 1.518$. Incoming parallel light waves are focused by the objective lens into water with a refractive index of $n_s = 1.33$. The resulting light intensity distribution represents the PSF. The distance of the intensity maximum from the interface as it would be for a medium with a matched refractive index is designated the nominal focus position (NFP). The sketch illustrates the problem that occurs in a mismatch situation: the incoming light is refracted at the cover glass/water interface if the refractive indices of both materials are not identical. Therefore, the light waves do not interfere constructively in the NFP, but rather above it in the so-called actual focus position (AFP). (b) The resulting intensity distribution in water was numerically computed for a 100× oil-immersion objective lens with an NA of 1.3 for different exemplary NFPs, namely $NFP_1 = 2\,\mu m$, $NFP_2 = 10\,\mu m$, and $NFP_3 = 20\,\mu m$, respectively. The computation showed that the refraction shifted the PSF maxima to $AFP_1 = 1.45\,\mu m$, $AFP_2 = 8.1\,\mu m$, and $AFP_3 = 16.6\,\mu m$, respectively. The resulting actual PSFs were calculated in regions around the AFP_i, namely, I_1 (0–4 μm), I_2 (4–11 μm), and I_3 (11–21 μm), respectively. The results are shown as contour plots with a logarithmic intensity scale (see inset). (c) The relationship between NFPs and AFPs is almost linear. Since $AFP_i < NFP_i$, the axial extension of thick samples appears elongated when it is determined by measuring the stage movement when focusing to the top and bottom of the sample. Therefore, spheres appear axially stretched when imaged by confocal microscopes. (d) Furthermore, the PSF is more and more smeared out for increasing NFPs, which is shown by the calculation of the FWHM values along and perpendicular to the optical axis. This leads to a depth-dependent reduction of lateral and axial resolution. A further effect is the reduction of the maximum of the distribution, which is not quite visible in (b) because the shown distributions were normalized to unity for a better representation of low intensity values. (Hell *et al*. 1993 [10]. Reproduced with permission of John Wiley & Sons.)

2.6 Contrast

Light is characterized by different properties – phase, polarization, wavelength (or color), and intensity (or brightness). Objects in the beam path usually alter one or more of these properties. Both the human eye and electronic light detectors are sensitive only to wavelength and intensity. Therefore, we can directly detect only objects that change one or both of these properties. Such objects are called *amplitude objects*, because they alter the amplitude of the light waves traversing them. If this occurs in a wavelength-dependent manner, the respective objects will appear in color. Objects that alter only the phase of light waves – possibly in a polarization-dependent manner – are called *phase objects*. Typical examples are transparent objects made of glass or crystals. Such objects effect a phase delay on the traversing light wave owing to a different index of refraction as compared to the surrounding medium. In addition, biological objects such as thin layers of cells or tissue influence the phase, but much less the amplitude of traversing light waves by absorption.

A phase object that is well resolved and magnified is not necessarily visible for the detection device. For an object to be visible, either its color or its light intensity must differ sufficiently from those of the surroundings. In other words, the

contrast between the object and the background must be detectable. Contrast, *C,* means here the degree to which the object differs from its background in terms of intensity or wavelength. For light intensity *I,* one of several possible definitions is

$$C_{\mathrm{I}} = \frac{I_{\mathrm{obj}} - I_{\mathrm{sur}}}{I_{\mathrm{obj}} + I_{\mathrm{sur}}} \tag{2.51}$$

where I_{obj} and I_{sur} denote the intensity of the object and of the surroundings, respectively. For electronic light detection devices, the detectable contrast depends on the overall signal-to-noise ratio and the number of discretization steps used when measuring the signal, which corresponds to the bit-depth of the detection device. These questions are dealt with in detail in the next chapter.

For human perception in bright light, the contrast in intensity may be as little as 0.02, but in poor light the contrast required for perception of the object is 0.05. For very small objects, it needs to be up to 0.2. Contrast difference between the object and its background is essential for seeing fine details and should be as large as possible for a clear perception of the object.

In the history of microscopy, numerous different techniques were developed to render phase objects visible, which are referred to as *optical contrast methods*. Obviously, such techniques play a major role in the microscopy of biological objects. In principle, there are two different types of contrast techniques: chemical and physical approaches. Chemical techniques introduce a color contrast by selective labeling of object structures with dyes. In modern microscopy, mostly fluorescent dyes are employed, which are the focus of Chapters 3 and 4. The physical techniques turn the initially unperceivable phase contrast into an amplitude contrast. Today, the most important optical contrast techniques for biomedical applications are phase contrast and DIC. Dark-field contrast represents an easy access to the physical contrast techniques.

2.6.1 Dark Field

The simplest of the physical contrast techniques is *dark-field contrast,* also called *dark-field illumination*. The principle is to retain only the object-diffracted light and eliminate all direct illumination light by means of a special illumination system. This can easily be achieved by illuminating the specimen from the side [11]. Alternatively, when illuminating the specimen from below in a transmitted-light dark field, the inclination of the incident radiation can be chosen so high that no direct light is captured by the objective lens. To achieve this, the angle of the incident light rays with respect to the optical axis must exceed the maximum objective aperture angle. Then only the light scattered by the specimen enters the objective, as shown in Figure 2.23, and creates an image with a high contrast. The object background is completely dark, and only light-scattering object structures appear bright.

When using low-power objectives with NA $\leq$ 0.4, dark-field illumination can easily be achieved by inserting a ring-shaped mask of an appropriate size into the front focal plane of the condenser. For this purpose, even phase-contrast condensers can be employed: as discussed later, a phase-contrast condenser contains

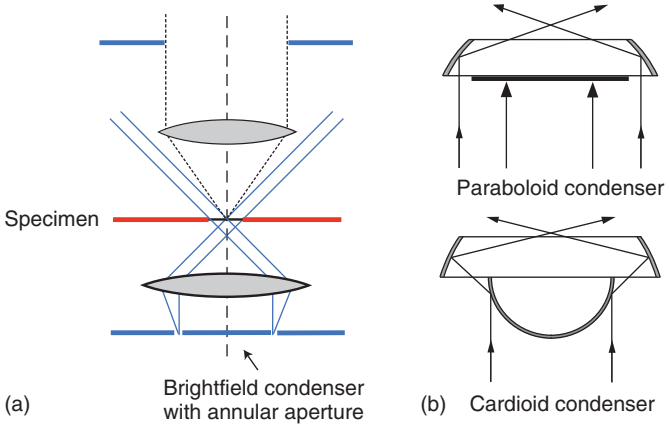


Figure 2.23 Dark-field contrast. (a) Principle of dark-field illumination and (b) dark-field condensers.

an annular aperture in its front focal plane. The central portion of the illumination light is blocked by a circular, axially centered mask (Figure 2.23a). The advantage is that the mask is positioned in the correct front focal plane of the condenser and can be axially centered by adjusting screws. A disadvantage, however, is that reflections at the lens surfaces of the condenser will diminish the dark-field effect. When the use of higher NA objectives is intended, dedicated dark-field condensers of a reflective type must be employed (Figure 2.23b). Such condensers have additional advantages. In comparison to ring condensers, they yield higher irradiance and result in identical inclinations for different colors, resulting in a sharper dark-field illumination cone. It is important to exactly position these dark-field condensers on the optical axis of the objective. When a cardioid condenser is used, an immersion fluid is added between the object slide and the condenser surface to avoid refraction into the object slide, which would lower the effective inclination angle. The use of high NAs for the condenser (e.g., NA 1.4) is especially advantageous because the illumination light is totally reflected at the upper surface of the cover glass when an air objective is used for observation. In that case, the elimination of the illumination light is particularly complete. The use of an objective lens whose NA can be adjusted by means of an internal iris diaphragm is practical because its effective NA may be adjusted until a completely dark background is created.

In principle, a dark-field effect is also obtained by blocking the illumination light in the back focal plane of the objective lens. This, however, would mean the loss of an appreciable amount of diffracted light and require the use of a specific objective for dark-field observation.

2.6.2 Phase Contrast

The principle of phase contrast and some of the interference contrast methods described later is the separation of light that is diffracted by the object from the light that is not altered by the object, that is, the background or undiffracted light. The amplitudes and phases of these two light components are modified in such a

way that finally an amplitude contrast due to constructive or destructive interference between the two light components occurs. Frits Zernike introduced in 1932 a versatile and useful microscopic technique to visualize phase objects that are almost invisible in bright-field microscopy [12]. He performed a series of experiments that elegantly illustrated the principle of his contrast technique – and simultaneously the wave theory of image formation in the microscope. His experiments are very instructive, and therefore we will follow his argument in detail.

2.6.2.1 Frits Zernike's Experiments

We begin the series of experiments by examining a tiny amplitude object using bright-field illumination. With a wide open condenser aperture, a dark spot on a bright ground is visible in the image plane – the standard bright-field image of a small amplitude object. Next, we close the condenser aperture to a small spot and thus illuminate the sample now by a point light source. Simultaneously, we place a small light-blocking mask in the back focal plane of the objective lens (Figure 2.24c). What happens? The condenser lens transforms the light emanating from the point source in its front focal plane into plane waves. These waves encounter the specimen, which they partly pass undiffracted and are partly diffracted. The undiffracted light is focused again by the objective lens into its back focal plane, which is conjugate to the front focal plane of the condenser. Here it is blocked by the above-mentioned mask. Its shape is complementary to the mask in the condenser front focal plane, and selectively prevents the undiffracted background light from passing the back focal plane of the objective lens. Blocking the undiffracted light passing by the object leads to a dark background in the image plane. The diffracted light, however, is not focused at the back focus of the objective because it is not a plane wave but rather a diverging spherical wave in the object space as discussed in Section 2.3 of this chapter. Therefore, it is not blocked in the back focal plane by the mask. It passes the back focal plane and is focused by the tube lens into the primary image plane, creating a bright spot on a dark background. Actually, this construction with the two complementary masks in conjugate planes represents a special type of dark-planefield illumination, which was discussed in Section 2.6.1.

Figure 2.24 shows the two different sets of conjugate planes: the condenser front focal and the objective back focal plane, as well as the object and image planes. In the first experiment, both undiffracted and diffracted light reach the image plane and produce a dark object image. When we block the background light in the back focal plane of the objective lens, the image is formed by the diffracted light alone, and we see a bright spot. The physical explanation for these two quite different image patterns is that the diffracted light is phase-shifted by 180° compared to the undiffracted light. In the first experiment, the image is the result of the destructive interference of the object-diffracted and the undiffracted light waves. In the second experiment, the image is the result of the object-diffracted light waves *alone.* No destructive interference can take place (Figure 2.25).

In the third experiment we image a small phase object. A small transparent corn is placed on the object slide, the condenser aperture is opened, and the mask is removed from the back focal plane of the objective, thus returning to bright-field microscopy. Now we see a bright background field and practically no object, as it

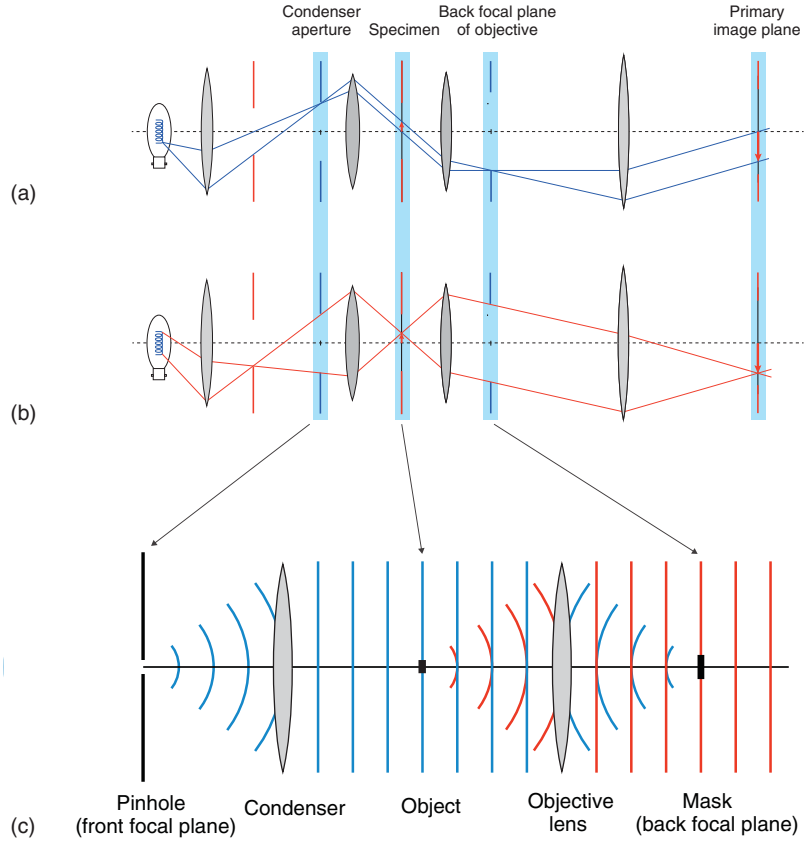


Figure 2.24 Zernike's second experiment. (a) Illumination beam path and (b) imaging beam path of a microscope. The arrows indicate the respective planes in the sketch underneath. (c) Magnification showing the space between the front focal plane of the condenser and the back focal plane of the objective lens. The undiffracted and diffracted waves are shown in blue and red, respectively. The four planes marked in light blue are further discussed in Figure 2.25.

is transparent. Next, we again close the condenser aperture and place the complementary mask back into the back focal plane of the objective lens. Again, we see a dark background because the undiffracted illumination light is blocked – and a bright spot from the object. Obviously, the transparent object does indeed diffract light similar to an amplitude object, but its phase is obviously not shifted by 180°. Otherwise, a dark object image on a bright background would be visible in the bright field.

Zernike assumed that the diffracted light of a phase object is shifted only by about 90°. A light wave with small amplitude that is phase-shifted by 90° when interfering with the background wave produces a wave of almost unaltered amplitude and phase compared to the undiffracted wave. This is demonstrated in the last column of Figure 2.25. It shows the amplitudes and phase relationships of the different wave components in all experiments. Amplitude and phase of the undiffracted wave are represented by the blue arrow pointing along the

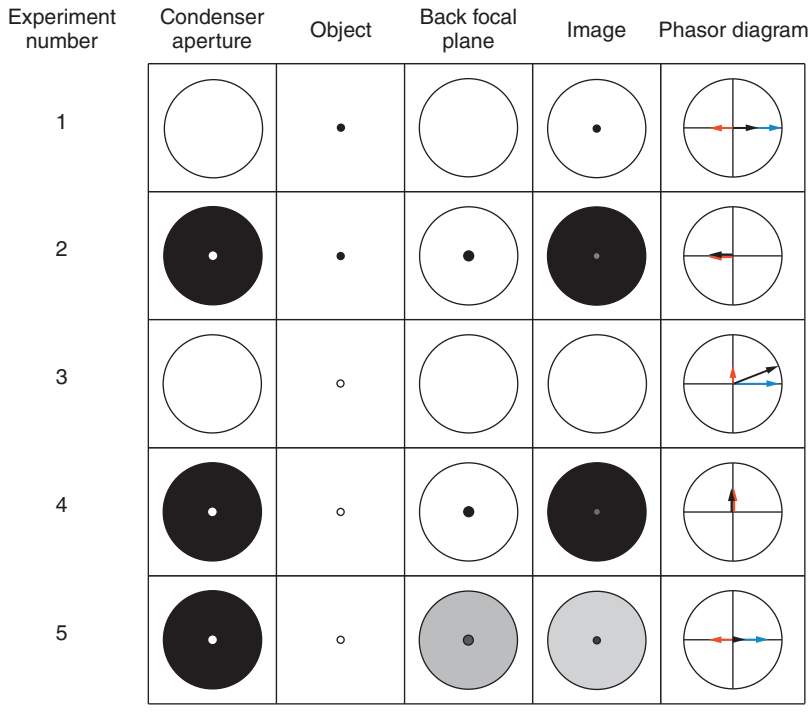


Figure 2.25 Masks, objects, images, phases, and amplitudes in Zernike's experiments. The first and third columns describe the masks in the front focal plane of the condenser and the back focal plane of the objective, respectively. In the upper two experiments, absorbing objects and in the lower three experiments transparent or phase objects were examined as indicated in the second column. The fourth column sketches the resulting image, and the fifth column indicates the relative phases and amplitudes of the involved waves in terms of so-called phasor diagrams (blue, undiffracted background light; red, diffracted object light; black, resulting wave at position of the object obtained by vector addition of the two other components).

x-direction. The amplitude and phase of the diffracted wave is indicated by the red arrow. For example, in the first experiment it is pointing into the $-x$-direction due to the phase shift of 180° with respect to the undiffracted wave. This arrow is also shorter, because the amplitude of the diffracted wave is smaller than that of the reference. The black arrow indicates amplitude and phase of the sum wave of the diffracted and undiffracted wave at the position of the spot image. It is obtained by vector addition of the first two arrows. Such diagrams are called *phasor diagrams.*

For each experiment, the produced contrast can be assessed by comparing the lengths of the blue and the black arrows. In the first experiment, they differ considerably, and in the second experiment there is no undiffracted wave in the image plane because it is blocked at the back focal plane of the objective. The small change in amplitude that occurs in the third experiment produces only very little contrast. Therefore, the transparent object is not perceptible on the background. In order to verify Zernike's assumption, we perform a final experiment that additionally demonstrates the principle of phase-contrast imaging. We insert

a special mask into the back focal plane of the objective lens. This mask comprises a light-absorbing center which reduces the amplitude of the undiffracted background light to almost that of the diffracted light. This center is surrounded by a phase plate that adds an additional phase delay of 90° to the diffracted wave. This modification to the previous experiment yields a dark spot on a gray background in the image plane. This result verifies Zernike's assumption. The gray background is due to the reduction of the background light amplitude. The diffracted wave from the object traversing the outer region of the back focal plane of the objective lens is phase-delayed by additional 90°, resulting in a total phase delay of 180° compared to the background. The diffracted light now interferes destructively in the image plane with the dimmed undiffracted light, and creates a dark image of the phase object on a gray background. The trick to reduce the amplitude of the undiffracted light leads to a high image contrast.

In summary, the observable image contrast is dependent on the physical characteristics of the object, the illumination mode, and, in addition, selective manipulations of background and diffracted light. These two light components are spatially separated from each other and therefore independently accessible in the back focal plane of the objective lens.

2.6.2.2 Setup of a Phase-Contrast Microscope

The pinhole that was used for object illumination in the above experiments is now replaced by an annular mask. By this modification, more illumination light reaches the object, resulting in a brighter phase-contrast image. At the same time, the effective NA of the condenser is enlarged, which improves optical resolution. Of course, the mask at the back focal plane of the objective lens must be replaced by an appropriate complementary annular mask. This ring carries an absorbing layer and has a different thickness than the rest of the mask. It is called the *phase ring* . The undiffracted illumination light from the front focal plane of the condenser passes through this ring. The diffracted light from a specimen will pass through the central and outer regions of the mask, which is thicker – or thinner for the so-called negative phase contrast – than the absorbing ring such that the required total phase difference of about 180° is achieved.

The exact phase delay produced by the mask depends on the wavelength of the illumination light. Hence, the mask creates a 90° phase delay only for one specific wavelength. This is usually designed to be 550 nm or green – the color for which the human eye is most sensitive. A color filter in the illumination beam path may be inserted to select only this wavelength of the illumination light for imaging.

The image of the illumination ring at the objective's back focal plane must perfectly match the absorbing ring. Therefore, for different objective lenses with varying magnifications and NA, illumination rings of different diameters and widths must be used. These are located on a rotating turntable, which is positioned in the condenser's front focal plane. One position on the turntable is open for normal bright-field illumination. The coincidence of the illumination ring image and phase ring in the back focal plane of the objective can be checked and adjusted using a special auxiliary microscope, which is inserted instead of the eyepiece. It comprises two lenses by which the back focal plane can directly be observed. An alternative to the auxiliary microscope is the "Bertrand" lens,

which can be switched into the beam path projecting an image of the back focal plane onto the principle image plane, and hence allows viewing the back focal plane using the existing eyepiece. In case the two ring structures are not perfectly matched, the lateral position of the illumination ring in the condenser turntable may be adjusted by two screws.

The back focal plane of the objective is located inside the objective housing. Therefore, objectives for phase-contrast imaging are specific products. Clearly, the absorbing phase ring also diminishes the image brightness when used in normal bright-field mode. Phase-contrast objectives are especially detrimental in epi-illumination techniques such as fluorescence microscopy.

2.6.2.3 Properties of Phase-Contrast Images

A phase-contrast image does not simply show a magnified view of the object. The generated contrast is due to the special manipulation of the phase of the undiffracted and the diffracted light. Therefore, its interpretation is not straightforward.

The phasor diagram in Figure 2.25 (experiment 5) shows that phase objects are seen dark on a bright background. However, if the phase delay due to the object becomes too large, for example, it exceeds 45° – in addition to the inherent 90° – the contrast is reduced again, and upon approaching 90° it almost vanishes. Then the phase relationships for bright-field observation of phase objects are recovered. Hence, the use of phase contrast is advisable only for thin objects producing a relatively small phase delay up to ~30°. This is the case for thin biological specimens such as cell monolayers or tissue sections. If the phase object produces only a small contrast because the phase shift is too small – or already too large – it may be improved by changing the refractive index of the mounting medium. The phase shift is proportional to the path length in the object and to the difference of the refractive indices of the object and the mounting medium. Therefore, the phase shift may be modified by choosing a suitable mounting medium.

The production of the phase contrast is based on the fact that the direct illumination light and the diffracted light are spatially separated from each other in the back focal plane of the objective lens. However, this separation is not complete owing to the finite width of the phase ring. Light diffracted from coarse specimen features may not be diffracted strongly enough to pass outside the phase ring, but rather traverse it. Likewise, some of the light diffracted by fine specimen features passes through the ring as well. The result of these effects is the bright halo that surrounds objects that appear dark, making the identification of object borders difficult.

2.6.3 Interference Contrast

The solution of the problem of phase objects is to translate the phase difference between the object and its surroundings into an amplitude difference. For this purpose, the phase of the diffracted wave is modified such that destructive interference between the diffracted and undiffracted waves occurs in the image plane. Zernike accomplished this by manipulating both wave components spatially separated in the back focal plane of the objective lens.

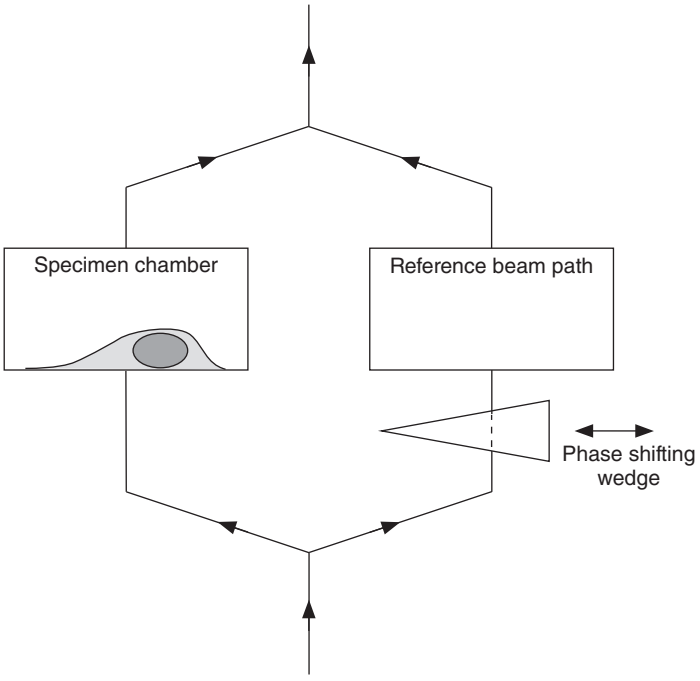


Figure 2.26 Principle of interference contrast.

There are further techniques that effectively reach the same goal. Actually, some of the so-called interference contrast techniques are even more versatile than phase contrast. The principal idea of interference contrast is shown in Figure 2.26: before encountering the object, the illuminating light is separated by a beam splitter into two coherent wave trains. One traverses the specimen chamber containing the object. The second wave – the *reference wave* – travels along an identical optical path, but without the object. Furthermore, a defined phase delay can be introduced in the reference beam path by inserting a glass wedge. After passing the specimen and the reference chamber, both wave trains are recombined and form the image by interference in the image plane.

This setup enables the microscopist to do the same as – or more than – what is achieved by Zernike's phase-contrast setup [13]. It is possible to manipulate the amplitudes and phases of the light diffracted by the object and the primary undiffracted light. The final image is created by the interference of three different waves: those diffracted by the object, the undiffracted light passing the object chamber, and the undiffracted reference waves.

The phase of the reference wave can be delayed in a defined manner by the glass wedge. As a result, the phase of the undiffracted waves can be shifted relative to the object-diffracted waves such that their final total phase difference is 180°, no matter how large the actual phase shift by the object is. To see how this is achieved, we consider first what happens without an object in the specimen chamber. To begin with, we move the glass wedge completely out of the beam path (Figure 2.27a). As both beams take identical optical paths, the two

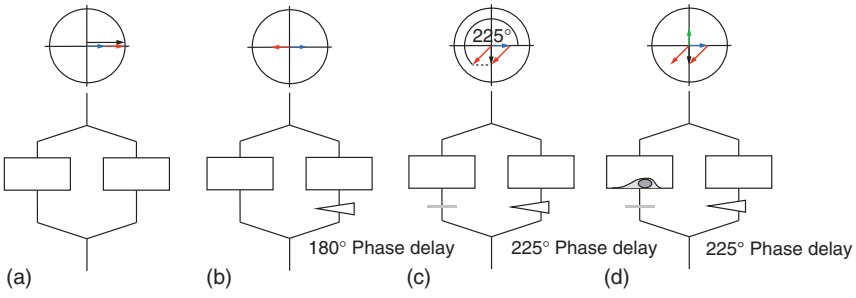


Figure 2.27 Interference contrast. The top sketches illustrate the phase relationships between the involved waves. The phasor of the background wave of the specimen chamber is shown in blue and the reference wave phasor in red. The phasor of the object-diffracted wave is represented by a green arrow, and the phasor of the sum background wave as a black one. (a) Constructive interference of the waves and (b) destructive interference due to 180° phase difference. (c) A phase shift of ~225° in the reference wave leads to an overall phase shift of ~270° in the sum of the two waves compared to the background wave in the specimen chamber. Note that the brightness was also slightly decreased in the specimen chamber. (d) A phase object is introduced in the specimen chamber. It diffracts a 90° phase-shifted object wave. The 270° phase-shifted background will interfere destructively with the object wave.

wave fronts will reach each point in the image plane simultaneously and interfere constructively.

This results in a homogeneous, bright background. Next, we move the glass wedge into the reference beam path and create a phase delay of 180° compared to the wave passing the specimen chamber (Figure 2.27b). Now, the two undiffracted waves interfere destructively everywhere in the image plane, which results in a completely dark background. The reference phase delay may also be adjusted such that the sum of the two waves in the image plane suffers effectively a 270° phase delay compared to the undiffracted wave from the specimen chamber (Figure 2.27c). When we finally add an object to the specimen chamber, the phases of the undiffracted reference and specimen chamber waves remain unchanged. The phase of the object-diffracted wave, however, shows the usual 90° phase delay. The phasor diagram shows that the phase difference between the sum of the undiffracted background waves and the object-diffracted wave is now 180° (Figure 2.27d). The undiffracted background light and the diffracted light will interfere destructively, resulting in a distinct phase contrast. We will see a dark object on a bright background. Even more, as we can choose the phase of the background sum wave with reference to the object wave arbitrarily, we can change from negative to positive contrast, or adjust the phase difference in such a way that we get a 180° phase difference also for objects whose diffraction wave phase is shifted by more or less than 90°. The physical separation of the two undiffracted waves allows free definition of their sum wave phase in relation to the object wave phase. Furthermore, by introducing a known phase shift into the background wave, the phase delay produced by the object may even be measured, and thus information gained on its refractive index or – if that is known – its physical thickness.

However, there is a price for this versatility. Interference microscopy requires a very special optical setup. There must be special elements to separate the specimen chamber and the reference beams, to recombine them again, and to adjust a defined phase delay between them. This requires very precise matching of the optical path length of the specimen and reference beam paths, which is not trivial to achieve. However, we will not focus here on the optical setups of interference microscopy, because it is actually another but related technique that is most important today for the observation of medical and biological samples.

2.6.4 Advanced Topic: Differential Interference Contrast

DIC microscopy is a very powerful method for contrast enhancement of phase objects. DIC can be used with transmitted or reflected light. For biomedical applications, epi-illumination DIC is of minor importance. How transmission DIC works and how it is realized can best be seen by considering the optical setup of a DIC microscope [14]. Its understanding requires a detailed knowledge of the wave properties of light such as interference, coherence, and polarization, as well as the function of polarizers and analyzers.

2.6.4.1 Optical Setup of a DIC Microscope

The optical setup of the DIC microscope is a further example of exploiting the optical conjugation of the front focal plane of the condenser and the back focal plane of the objective lens. The beam path is relatively complex (Figure 2.28). A polarizer is placed between the illumination field stop and the condenser. Let us assume that it produces light with a polarization direction of 45° (Figure 2.28b). This light encounters the *Wollaston prism* in the front focal plane of the condenser. A Wollaston prism is an optical device that splits the vertical and parallel polarization components of polarized light into two distinct waves that diverge by a small angle. We call these two coherent waves polarized in vertical and parallel direction (Figure 2.28a,c). In the object space, the two differently polarized wave fronts move parallel to each other, as the prism is located in the front focal plane of the condenser, but the two wave trains have a small lateral offset from each other (Figure 2.28b). The offset or *shear* is defined by the construction of the Wollaston prism, and chosen to be smaller than the optical resolution of the employed objective. The two spatially shifted or *sheared* beams are recombined again by a second Wollaston prism in the back focal plane of the objective. This Wollaston prism may be adjusted in its lateral position such that an additional phase delay Δ between the two beams with the perpendicular polarization to each other can be introduced. A lateral movement of the prism changes the optical path lengths inside the Wollaston prism for each of the two beams differently.

An analyzer is placed behind the second Wollaston prism. It transmits only a selected polarization component of the recombined beam. Altogether, the image is created by the interaction of four different wave components, namely, the diffracted and undiffracted waves of the two displaced wave fronts with their different polarization directions. The contrast generation depends on the OPDs between the two split wave fronts. It is relatively complex owing to the multiple parameters that modify these, which include the illumination light polarization,

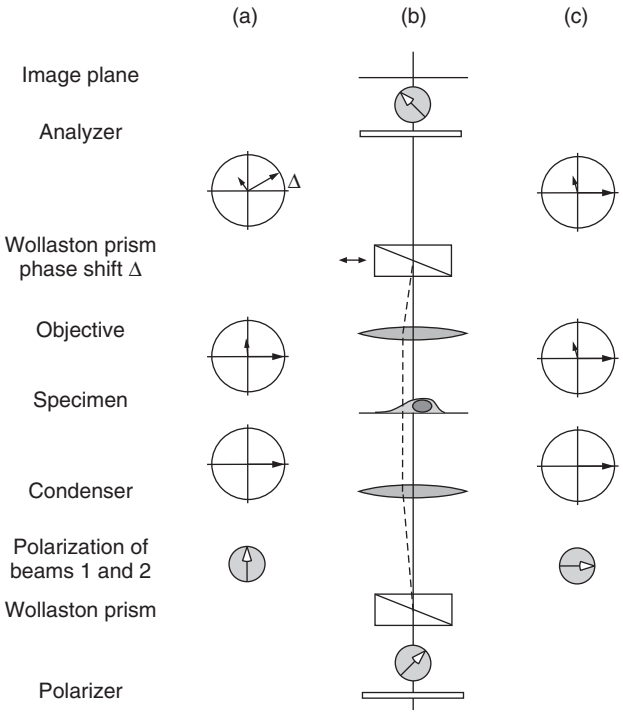


Figure 2.28 Optical setup of a DIC microscope. (a) Phases of undiffracted (long arrow) and diffracted waves (short arrow) of the offset beam (dashed lines in (b)) with vertical polarization, (b) sketch of beam path, and (c) phases of undiffracted (long arrow) and diffracted waves (short arrow) of the beam with parallel polarization. The circles with gray background indicate the *polarization* of the respective beam path – not the phase!

the lateral position of the second Wollaston prism (which determines the phase delay Δ between the two polarization directions), the setting of the analyzer, and finally the sample properties themselves.

An understanding of the contrast generation process is best achieved by discussing an example as given in Figure 2.29. The figure also directly relates to Figure 2.28. The upper panel of Figure 2.29 shows the specimen chamber picturing a cell with nucleus as an example of a phase object. This chamber is traversed by the two laterally shifted, perpendicularly polarized wave trains indicated by the two sheared vertical lines (dashed and full). To understand the contrast generation, we evaluate the wave components at four different positions in the sample (Figure 2.29a–d). The polarization state of the sum wave in the image at the indicated four positions is shown for two different values of the phase shift Δ (0° and 30°) in the middle and lower panels of Figure 2.29.

We proceed from Figure 2.29a–d, and assume that the analyzer above the second Wollaston prism is crossed to the polarizer in front of the first Wollaston prism (as indicated in Figure 2.28). At background positions, there are no diffracted waves from the object (Figure 2.29a). Hence, in the image, the background around the object is dark if the second Wollaston prism is adjusted

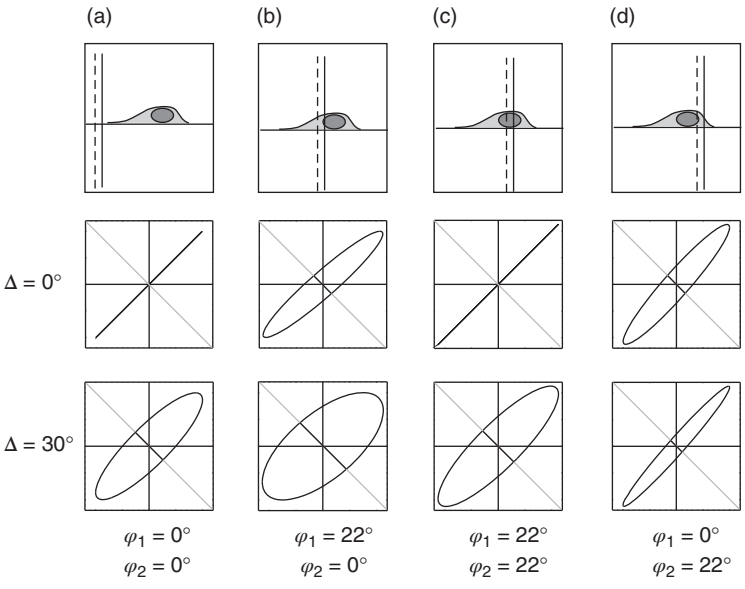


Figure 2.29 (a–d) Contrast generation, polarization, and phase relationships in DIC microscopy. For details, see the text.

such that the optical path lengths of the undiffracted background waves for both polarization directions are identical. For $\Delta = 0°$, this is the case. The polarization for this value of Δ is shown in the middle panel of Figure 2.29. Polarization and phase of the two shifted background waves are not altered at this sample point. Therefore, the background light is completely blocked by the analyzer, because it is crossed with the polarizer. The diagonal gray lines in the middle and lower panels of Figure 2.29 indicate the direction of the analyzer. A phase shift between the two polarization directions, which can be introduced by a lateral move of the upper Wollaston prism, results in an elliptical polarization of the combined background waves. The lower panel in the figure shows exemplarily the elliptical polarization, when $\Delta = 30°$. The analyzer lets pass only the polarization component along its orientation. The size of this component is indicated in the middle and lower panels of Figure 2.29 by the thick, black line in diagonal direction. Thus, we can modify the polarization state of the background light by moving the Wollaston prism in the back focal plane of the objective.

Let us move it back to $\Delta = 0°$. What happens at an object position where the parallel-polarized beam encounters the edge of the phase object, that is, the cell nucleus, but not the vertical-polarized beam (Figure 2.29b)? A wave with parallel polarization is diffracted. The light is diffracted by a phase object, which means that it gets an intrinsic phase shift of 90° in relation to the undiffracted wave plus a small additional phase delay φ_1, which is due to the difference in optical path length caused by the higher optical density of the nucleus. As an example, we assume $\varphi_1 = 22°$, but it could be any value. The crucial point now is

that the diffracted wave component with its phase shift of (90° + 22°) changes the polarization of the sum wave that meets the analyzer. It is no longer completely linearly polarized as in specimen regions without the object, but elliptically polarized. This has the important consequence that the amplitude of the sum wave when passing the analyzer is not zero anymore, which results in some light at this position in the image. A phase difference is turned into a visible brightness difference.

In regions of the object where the sheared beam with the vertical polarization encounters the same object feature – the nucleus – a second vertically polarized object-diffracted wave with a phase retardation $90° + \varphi_2 \approx 90° + \varphi_1$ is created, which reduces the ellipticity of the sum wave back to a linear polarization because the parallel- and vertical-polarized waves have again identical phases and amplitudes (Figure 2.29c). The parallel- and the vertical-polarized waves with zero phase shift together form a linearly polarized wave that is completely blocked by the analyzer, leading to zero intensity at this image position. Thus, a contrast was just created at the edge of the object.

At the other side of the nucleus (Figure 2.29d), only the wave with vertical polarization interacts with the object whereas the parallel-polarized beam passes already undiffracted. Therefore, again a certain ellipticity results and the analyzer lets pass some light to the corresponding image point.

In summary, the final contrast in a DIC image is determined by the OPD at laterally slightly displaced positions along a certain direction in object space. Therefore, it is named "differential." The direction along which the path length difference occurs is defined by the orientation of the first Wollaston prism. A difference in the optical path lengths of adjacent object positions results in a pronounced contrast. Even more importantly, the contrast can be adjusted by the choice of Δ in such a way that a positive contrast occurs at one edge of the object and a negative contrast at another edge. This is demonstrated in the lower panel of Figure 2.29, which shows the elliptical polarizations that occur if the vertical-polarized waves acquire an additional phase shift of 30°. Then, the left boundary of the nucleus is brighter than the background, the nuclear interior, and, notably, the opposing boundary of the nucleus. In this manner, a visual impression of a 3D object, which is illuminated from one side and casts a shadow on the other side, is created. This asymmetrical contrast is typical for DIC.

DIC is a complex technique and not easy to realize. The second Wollaston prism must be located in the back focal plane of the objective lens. As it has a considerable thickness, this calls for special objectives. As prisms are incorporated in the objectives, they are unsuitable for other tasks. To overcome this problem, specifically constructed prisms were introduced that place the plane of beam recombination outside the prism itself [15]. These so-called Nomarski prisms may therefore be located above the focal plane, which can be easily accomplished. In addition, they are removable. This modification of classical DIC is known as *Nomarski DIC*, and has become one of the most widespread interference contrast techniques in the biosciences.

2.6.4.2 Interpretation of DIC Images

The strength of DIC is defined by the lateral amount of the shear between the two beams, the respective orientations of the prisms, polarizer, and analyzer, and the sample features. DIC has the advantage that the way of contrast generation and image appearance can widely be adjusted. However, this also leads to ambiguities in image interpretation.

The main feature of DIC is its excellent edge contrast. This results in a strong depth dependence of the effect. DIC varies very sensitively upon the axial movements of the specimen and therefore allows acquiring very thin optical sections. It is well suited for thick specimens. The pronounced relief structure shown by phase objects is rather a side effect of the technique. One edge can be bright, and the opposing edge can be dark. The object appears to be illuminated from one side and cast shadows on the opposite side. Thereby the images have a 3D appearance. However, all contrast effects are solely due to differences in optical path length. They may be related to the object topology, but may also be caused by refractive index differences, while the true object thickness is constant. The true specimen topology is not visualized. Furthermore, the image of a circular object is not rotationally symmetrical because we see OPDs only along the direction of the beam shear. It is helpful to use a rotating stage in order to understand the full structure of the object. Notably, it is also possible to turn the bright–dark contrast into a color contrast, which makes it even more difficult to interpret DIC images.

2.6.4.3 Comparison between DIC and Phase Contrast

The applications of phase-contrast and transmission DIC overlap. Both techniques convert phase differences between the object-diffracted light or between light traversing different regions of the object into perceivable intensity changes. However, there are several important differences between both techniques.

In phase contrast, the object is imaged in a rotationally symmetrical way. DIC shows optical path length *gradients* that occur along one spatial orientation only. DIC images show a typical 3D surface-shading appearance, which is an artifact of contrast generation. DIC images display a higher resolution compared to phase-contrast images, not least because they do not show the halo encircling object features with large phase differences to the surroundings. Contrast in DIC images is pronounced for small objects, and it works with thin and thick objects. Importantly, DIC shows excellent axial resolution. Object features in planes other than the focal plane do not appear in the image, which allows one to make optical sections. This is in clear contrast to phase contrast, which is not very effective for thick objects. Phase contrast works best with very thin objects. The depth of field is very large, and object features in planes outside the focal plane are quite disturbing. Problems occur when the path difference becomes too large, particularly if it exceeds the limit of 30°, which is the upper limit for a good phase contrast. In summary, DIC is technically more demanding but has a wider range of applications than phase contrast.

2.7 Summary

The key points of this chapter are the following:

- The basic imaging elements of a microscope are the objective and tube lenses. Typically they form an infinity-corrected setup.
- Magnification steps may be arranged sequentially.
- A microscope contains two complementary beam paths, the illumination path and the imaging beam path. Each comprises specific components and a different set of conjugate planes.
- Köhler illumination provides even illumination, separates the illumination field from the illumination aperture, and creates conjugate optical planes in the front focal plane of the condenser and objective.
- The back focal plane of the objective contains the diffraction pattern of the object. It corresponds to the truncated Fourier transform of the object function.
- The image is the convolution of the object and the PSF.
- A lens is a low-pass filter for object structures.
- The lateral optical resolution is given by the Rayleigh criterion $0.61\lambda/n \sin \alpha$.
- Coherent and incoherent objects have different resolution limits.
- The optimal magnification is twice the ratio of the detector size and the optical resolution.
- Objectives should be telecentric.
- Objectives are of varying types and correction levels. They always exhibit residual aberrations.
- Image brightness is typically proportional to the square of the NA and inversely proportional to the square of the magnification.
- Image brightness in fluorescence microscopy increases with the fourth power of NA.
- The objective must be carefully selected for a given application. Only a proper objective will give satisfactory imaging results.
- The most important contrast techniques for biomedical applications are phase contrast and DIC.
- Optical contrast techniques are based on the selective manipulation of background and object-diffracted light.

Acknowledgments

The author thanks Dr Rolf Käthner, Dr Thorsten Kues, Manfred Matthä, and Dr Michael Zölffel from Carl Zeiss Microscopy GmbH, Göttingen, for helpful discussions and for critical reading of the text and numerous constructive suggestions. In addition, he thanks Manfred Matthä for providing Figures 2.17 and 2.18a and for the data illustrated in Figure 2.18b. Finally, he thanks Lisa Büttner, Jana Bürgers, Alexander Harder, Sahand Memarhosseini, Nicolai Pechstein, and Prof. Dr Rainer Heintzmann for critical comments on this text.

References

1 Köhler, A. (1893) Ein neues Beleuchtungsverfahren für mikrophotographische Zwecke. *Z. Wiss. Mikrosk. Mikrosk. Tech.*, **10**, 433–440.
2 Abbe, E. (1873) Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung. *Arch. Mikrosk. Anat.*, **9**, 413–468.
3 Köhler, H. (1981/1982) A modern presentation of Abbe's theory of image formation in the microscope, part I. *Zeiss Inf.*, **26** (93), 36–39.
4 Hecht, E. and Zajac, A. (2003) *Optics*, Addison-Wesley, San Francisco, CA.
5 Butz, T. (2005) *Fourier Transformation for Pedestrians*, 1st edn, Springer, Berlin, Heidelberg.
6 Born, M. and Wolf, E. (1999) *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*, 7th edn, Cambridge University Press, Cambridge.
7 Sheppard, G.J.R. and Matthews, H.J. (1987) Imaging in high-aperture optical systems. *J. Opt. Soc. Am. A*, **4** (8), 1354–1360.
8 Michel, K. (1950) *Die Grundlagen der Theorie des Mikroskops*, Wissenschaftliche Verlagsgesellschaft M.B.H., Stuttgart.
9 von Bieren, K. (1971) Lens design for optical Fourier transform systems. *Appl. Opt.*, **10**, 2739–2742.
10 Hell, S., Reiner, G., Cremer, C., and Stelzer, E.H.K. (1993) Aberrations in confocal fluorescence microscopy induced by mismatches in refractive index. *J. Microsc.*, **169**, 391–405.
11 Siedentopf, H. and Zsigmondy, R. (1902) Über Sichtbarmachung und Größenbestimmung ultramikoskopischer Teilchen, mit besonderer Anwendung auf Goldrubingläser. *Ann. Phys.*, **315**, 1–39.
12 Beyer, H. (1965) *Theorie und Praxis des Phasenkontrast-verfahrens*, Akademische Verlagsgesellschaft Geest und Portig KG, Frankfurt am Main, Leipzig.
13 Beyer, H. (1974) *Theorie und Praxis der Interferenzmikroskopie*, Akademische Verlagsgesellschaft Geest und Portig KG, Frankfurt am Main, Leipzig.
14 Lang, W. (1968–1970) Differential-Interferenzkontrast-Mikroskopie nach Nomarski. I. Grundlagen und experimentelle Ausführung. *Zeiss Inf.*, **16**, 114–120 (1968); II. Entstehung des Interferenzbildes. *Zeiss Inf.* (1969) **17**, 12–16; III. Vergleich Phasenkontrastverfahren und Anwendungen. *Zeiss Inf.* (1970) **18**, 88–93.
15 Allen, R., David, G., and Nomarski, G. (1969) The Zeiss-Nomarski differential interference equipment for transmitted-light microscopy. *Z. Wiss. Mikrosk. Mikrosk. Tech.*, **69** (4), 193–221.