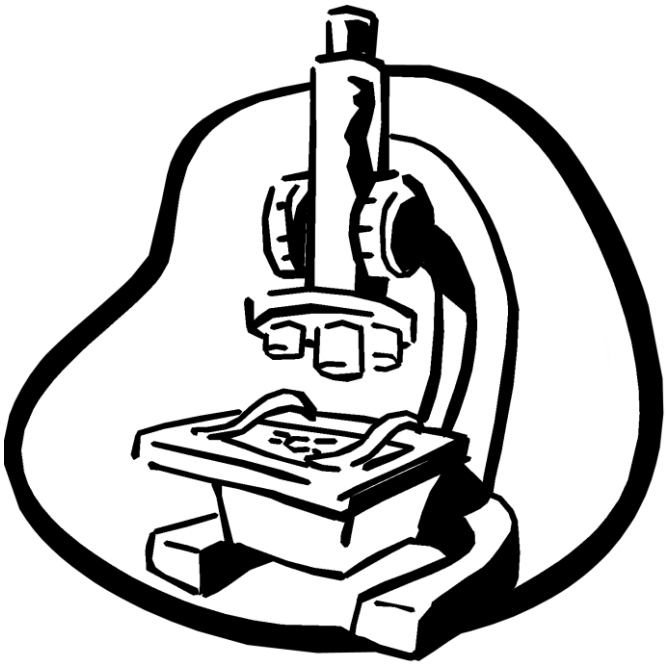
Biomedical & X-ray Physics
Kjell Carlsson



Light Microscopy

Compendium compiled for course SK2500,

Physics of Biomedical Microscopy

by

Kjell Carlsson

Contents

Introduction

This compendium is divided into three main parts: “Wide-field microscopy,” “Confocal microscopy,” and “Recent microscopy techniques.” Wide-field microscopy relates to the ordinary, classic microscope. The name derives from the fact that the entire field of view in the microscope is uniformly illuminated, and can be viewed through eyepieces or photographed using a camera. In confocal microscopy, on the other hand, a focused spot of light is scanned across the specimen to record an image. This is a more recent technique, which has found widespread use only within the last decades (the classic microscope has been around for centuries). Finally, in the last chapter, other new techniques are described, most of which have only recently found practical use. The purpose of the last chapter is to give a brief orientation of some new variations and developments in light microscopy, and the interesting results obtained.

1. Wide-field microscopy.

1.1 Basics of light microscopy

The basic principle of the light microscope is shown in Fig. 1. An image of the object (specimen) is formed by the objective lens, which typically provides a magnification in the range 10x to 100x. This magnified image is then viewed through the eyepiece (ocular), whose magnification is usually 10x.

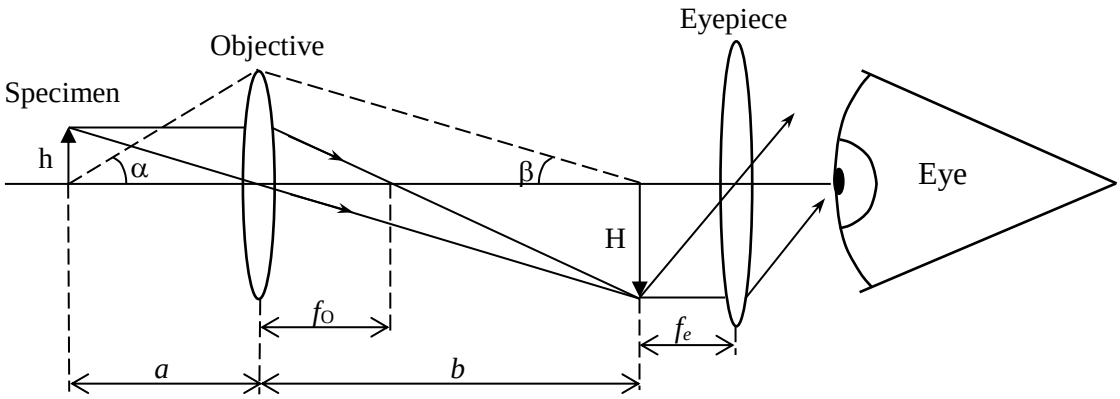


Fig. 1. The microscope consists of two lens systems, the objective and the eyepiece. The objective forms a real image of the specimen, and this image is viewed through the eyepiece forming a virtual image at infinity.

The total magnification of a microscope is obtained by multiplying the objective and eyepiece magnifications. Typically total magnifications are in the range 100X to 1000X. Compared to what? This is a relevant question, because the image is (usually) not presented as a real image, but rather as a virtual image viewed through the eyepiece. The somewhat arbitrary definition for microscope magnification is the following: Magnification = The angle subtended by a (small) object as seen through the microscope, divided by the angle when the same object is viewed by

the naked eye at a distance of 250 mm. Using the denotations in Fig. 1, we can express the magnifications of the objective, M_0, and eyepiece, M_e, as

$$M_o = \frac{H}{h} = \frac{b}{a} \qquad \text{(see footnote}^{\clubsuit}\text{)} \tag{1a}$$

$$M_e = \frac{250\,(\text{mm})}{f_e\,(\text{mm})} \tag{1b}$$

Microscope objectives are always labeled with their magnification and numerical aperture (*N.A.*). The numerical aperture is defined by

$$N.A. = n\sin\alpha \tag{2}$$

where n is the refractive index for the medium between specimen and objective, and α is defined in Fig. 1. A high *N.A.* means that the objective collects light efficiently (large α). As we shall see in later sections, a high *N.A.* also means high resolution. This is the reason why so-called immersion objectives are often used. The space between specimen and objective is then filled with a fluid (often oil) with a high refractive index to increase the *N.A.*

In recent years (since about 1990) most microscope manufacturers have abandoned the optical layout in Fig. 1, and instead adopted a so-called "infinite tube length" approach. This layout is illustrated in Fig. 2.

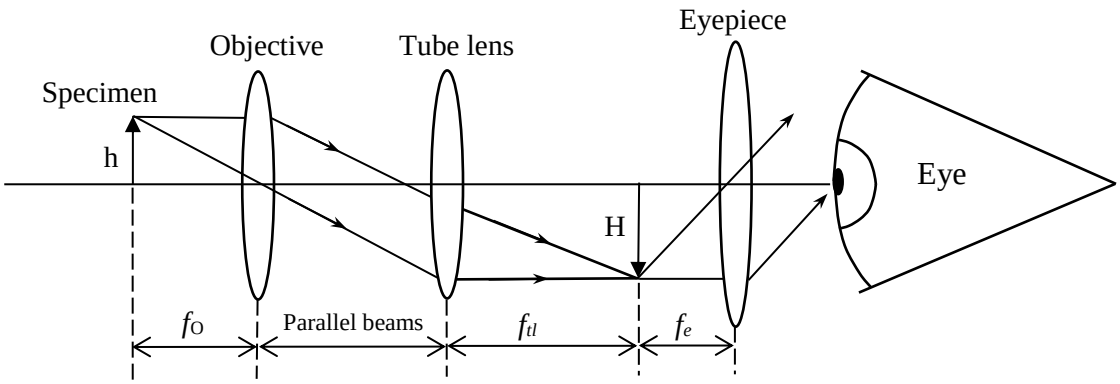


Fig. 2. Most modern microscopes use an optical layout that differs slightly from that of Fig. 1. The microscope objective forms an image of the specimen at infinite distance (this is often referred to as "infinite tube length"), and an additional tube lens (focal length f_{tl}) is needed to form a real image that can be viewed by the eyepiece. The advantage with this design is that a region of parallel beams is formed between the objective and tube lenses. In this region optical components like filters, beam splitters and other plane-parallel components can be inserted without affecting the size, location or quality of the specimen image. This is a great advantage in many practical applications.

$^{\clubsuit}$ Equation 1a is valid only if angles α and β in Fig. 1 are small, and the refractive index in specimen space, n, is 1. The more general equation is given by the so called sine condition (ref. [1]): $M_O = \frac{H}{h} = \frac{n\sin\propto}{\sin\beta}$.

When using a microscope, it is important to understand the information that is marked on objectives and eyepieces. Otherwise the equipment may be incorrectly used, resulting in poor image quality. Examples of such information can be seen in Fig. 3.

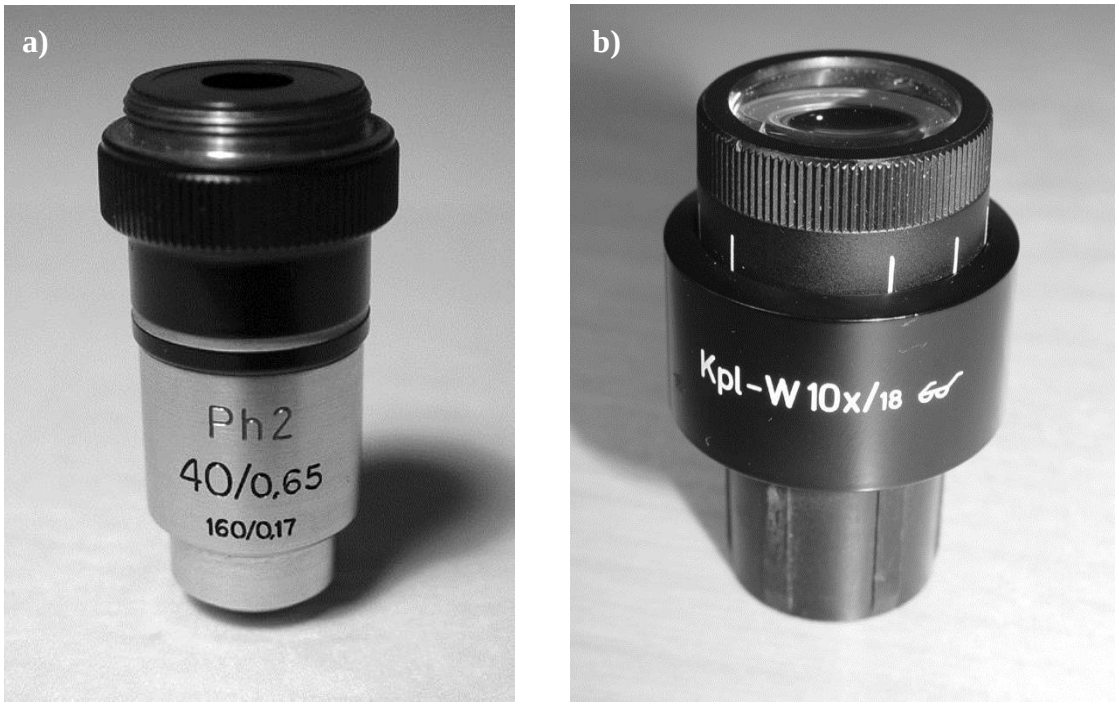


Fig. 3. Microscope objectives and eyepieces are marked with magnification and other data that are important for the user to know. See text for details.

Microscope objectives are always marked with (at least) magnification and numerical aperture. The marking 40/0.65, Fig. 3a, means that the magnification is 40 and the numerical aperture (*N.A.*) is 0.65. Oil immersion objectives are usually marked "Oil" (or Oel), and phase contrast objectives (see sect. 1.5.2) are marked "Ph" (the number 2 indicates the size of the phase ring). In addition "tube length"[♣] and cover glass thickness[♠] are often indicated (160 and 0.17 in Fig. 3a).

The objective is the most crucial optical component in a microscope. Many different types of objectives can be used depending on the application and, not least, the amount of money available. Depending on the aberration corrections[♦], objectives

[♣] Microscope objectives are designed either to form an image at a finite distance (Fig. 1) or at an infinite distance (Fig. 2). In the first case the mechanical tube length (total length of microscope tube) is usually specified as 160 mm and in the second case as ∞ (i.e. in the absence of a tube lens an infinitely long tube would have to be used). Performance will be degraded if objectives are not used with the proper image distance. Modern microscopes are usually designed for ∞ tube length.

[♠] Common values are 0.17 mm, 0 mm (i.e. no cover glass) and – (can be used with or without cover glass). Some objectives can be adjusted for different cover glass thicknesses by rotating a ring on the objective. If an objective is used with incorrect cover glass thickness the image quality will be degraded.

[♦] A summary of different types of aberrations in optical systems is found in Appendix I.

can be classified into the following main categories (in order of rising performance and cost):

-Achromats
-Flat-field achromats
-Fluorites
-Flat-field fluorites
-Apochromats
-Flat-field apochromats

Achromats are color-corrected for a relatively narrow wavelength range, whereas fluorites and, especially, apochromats are useful over a much larger range. "Flat-field" means that the curvature of field is very small. For objectives that are not of the flat-field type it may be necessary to refocus the microscope slightly depending on whether one is studying the center or periphery of the field-of-view. Some microscope manufacturers have selected to correct for some of the remaining objective aberrations by introducing an equal but opposite aberration in the eyepiece (or alternatively in the tube lens). This is then called a compensating eyepiece. In such cases it is important to use objectives and eyepieces that are matched.

The type of correction is usually marked on the objective. If it is not, you can assume that it is an achromat (i.e. the cheapest type). “Plan” means flat-field corrected. “Fluo” means fluorite objective, and “Apo” means apochromatic. Thus the marking “Plan-Apo” is found on objectives with the best aberration corrections (and highest price!).

Microscope eyepieces are marked with (at least) magnification and field-of-view♦. In Fig. 3b the marking 10x/18 means that the magnification is 10, and the field-of-view is 18 mm. In combination with a 100x objective, this means that a 180 μm diameter circle in the specimen plane is viewed. The spectacle symbol indicates “high eyepoint,” meaning that the exit pupil, see Fig. 4, is located so far from the eyepiece that spectacles can be worn. Additional information like Pl(an), i.e. flat-field correction, and K (compensating for residual objective aberrations) can sometimes be found. W usually means wide-field (unnecessary information if you have the field-of-view number).

A real microscope is considerably more complicated than those illustrated in Figs. 1 and 2. First, both objective and eyepiece are compound lenses, i.e. consist of many lens elements. Second, additional optical components are often inserted to distribute the light to multiple eyepieces, photographic equipment etc. Third, the illumination system is not shown in Figs. 1 and 2. Many microscopes are equipped both for epi- and trans-illumination (illumination from same and opposite side as viewing). Illumination systems will be the topic of the next section.

♦ The field-of-view number, e.g. 18 mm, is the diameter of the circular image you see in the eyepiece ***measured in the image plane of the objective***.

1.2. Illumination systems

So far we have assumed that the specimen emits light without describing the method of illumination. In practical microscopy, however, the choice of illumination technique is very important. Furthermore, the operator must be able to adjust the illumination system properly based on specimen type, objective used etc. In order to do this, it is helpful to understand the basics of how the illumination system works. Let us therefore start by describing the standard set-up used for trans-illumination of the specimen. This is called Köhler illumination, and is shown in Fig. 4 for a finite tube length microscope. The corresponding ray path when using an infinite tube length microscope is shown in Appendix II.

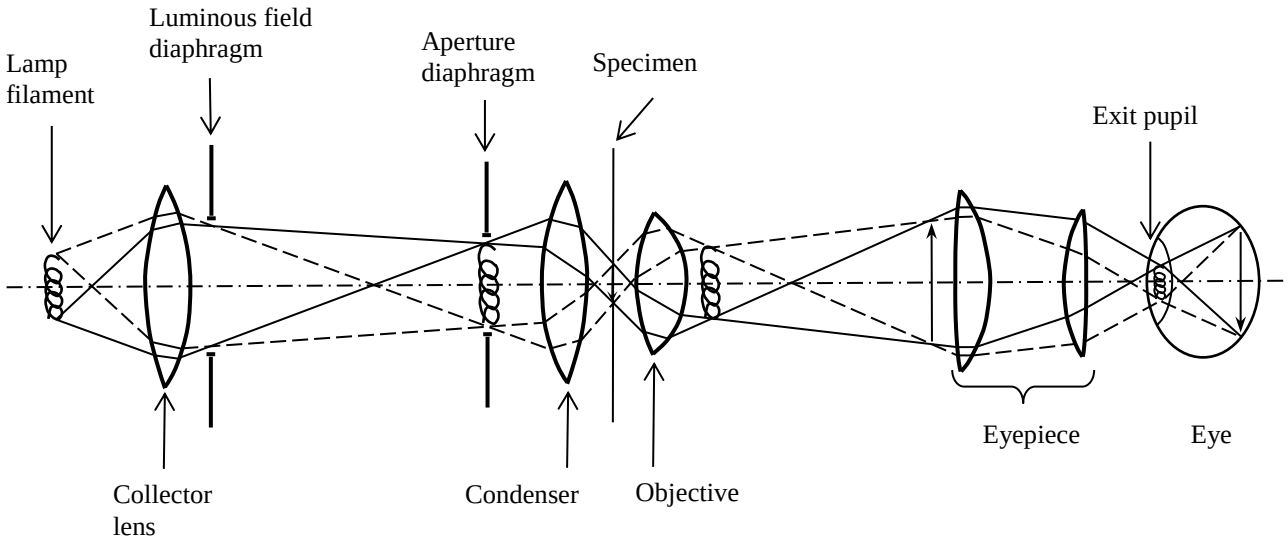


Fig. 4. Ray path in a microscope with Köhler illumination. The corresponding ray path for an infinite tube length microscope is given in Appendix II.

Compared with Fig. 1, the complexity of the ray path is now considerably higher. Still Fig. 4 shows only a simplified version of the optical layout, because in a real research microscope there are additional lenses, beam splitters, filters etc. The reason for the relatively complicated illumination ray path is that the operator needs several degrees of freedom when adjusting the microscope. The basic requirements for an illumination system in microscopy can be summarized in the following way:

- The imaged specimen area must be uniformly illuminated
- The light throughput should be reasonably efficient
- The intensity of the illumination must be adjustable
- The size of the illuminated specimen area must be adjustable
- The solid angle (see Appendix IV) under which the specimen is illuminated must be adjustable (or more popularly expressed: it must be possible to vary the diffuseness of the illumination)

Further requirements may include the possibility to spectrally filter or polarize the light, and to influence the color temperature of the illumination.

To be able to fulfill the above requirements, the illumination system must first of all be properly assembled and aligned. In practical use, the operator must then operate the different controls in such a way that the microscope can perform at its best. To perform these tasks it is helpful to understand how the Köhler system works, and we shall therefore start by describing this.

Looking at Fig. 4, we see that two different ray paths exist in the microscope. These are the *illumination* and *imaging* ray pathshs. These ray paths image the lamp filament and specimen respectively. The illumination and imaging ray paths are complementary in the sense that when one is focused to an image, the other is totally defocused. Starting with the light source, we see that light from the lamp is collected by the *collector* lens. Then come the *luminous field diaphragm* and the *aperture diaphragm*. To the right of the aperture diaphragm (at a distance of one focal length) is the condenser lens. These components together constitute the illumination system of the microscope. Both collector and condenser may consist of several lens elements in a real system. The collector and luminous field diaphragm are often built into the microscope stand, and controls for adjusting the lens position and diaphragm size are accessible from the outside. The condenser unit, comprising condenser lens(es) and aperture diaphragm, is mounted right under the specimen stage in the microscope. It can easily be exchanged, because different applications may require different condensers. The position of the entire condenser unit along the optical axis can be adjusted with a focusing knob, so that the luminous field diaphragm is sharply imaged in the specimen plane. The lamp filament, on the other hand, is sharply imaged in the plane of the aperture diaphragm by moving the collector lens. The sizes of the luminous field and aperture diaphragms are adjustable.

The microscope lamp is often a low-voltage halogen lamp (for example 12V/50W). The collector lens forms an image of the lamp filament in the plane of the aperture diaphragm. Because this diaphragm is located in the focal plane of the condenser lens, the lamp image will be totally defocused in the specimen plane. Note that each point of the lamp filament will illuminate the entire specimen plane, but light from different points on the filament will illuminate the specimen from different angles. Note also that the luminous field diaphragm is sharply imaged in the specimen plane by the condenser lens. This means that by adjusting the size of the luminous field diaphragm we can change the size of the illuminated field in the specimen (hence the name). The light intensity within the illuminated field does not change, however. Changing the size of the aperture diaphragm, on the other hand, will not affect the size of the illuminated field in the specimen. The reason for this is that the aperture (as well as the filament) is completely defocused in the specimen plane.

The size of the aperture diaphragm will affect two things. First, it will affect the illumination level in the specimen. The larger the aperture, the brighter the illumination will be. Second, it will affect the solid angle under which the specimen is illuminated. The larger the aperture, the larger the solid angle will be, and therefore the illumination will be more diffuse. What the aperture diaphragm does is to select what parts of the lamp filament that are used for illuminating the

specimen. If a small opening is used, only light from the central part of the filament is allowed to pass. The fact that changing the size of this aperture has two effects, varying light intensity as well as solid angle, is a bit unfortunate and often leads to incorrect use. *The primary role of the aperture diaphragm is to control the solid angle of the illumination*. It should not normally be used for changing the light intensity. Instead, the light intensity should be varied by gray filters or the lamp voltage (the latter will also change the color temperature of the illumination). The solid angle of the illumination should, in general, be adjusted so that it matches the numerical aperture of the objective used. Only then can we utilize the full resolution of the objective. This will be discussed further in section 1.4. In analogy with objectives, condensers are therefore labeled with their (maximum) numerical apertures. The condenser N.A. is equal to $\sin\varphi$, where φ is the maximum angle that an illumination ray forms with the optical axis. Restricting the aperture diaphragm size will reduce the effective N.A. of the condenser. When working with high N.A. immersion objectives one should, for optimal results, also use an immersion condenser where the light is coupled into the specimen via immersion fluid under the specimen glass (a bit messy to work with).

As mentioned, condenser N.A. will affect the resolution, but it will also affect the depth of field, as well as the image contrast when studying highly transparent and color-less specimens. The depth of field effect is the same as in photography, where stopping down to a small aperture will increase the depth of field. The influence on contrast depends on refraction effects in the specimen. Just to get some feeling for depth of field numbers, it can be mentioned that for $N.A._{obj} = 1.3$ (with matching $N.A._{cond}$) a specimen thickness of only about 0.3 μm can be sharply imaged. For numerical apertures of 0.6 and 0.3, the corresponding numbers will be about 1.5 and 7 μm respectively. Also in these cases we assume that $N.A._{cond}$ matches that of the objective. In some cases it may be useful to play around a little with the aperture diaphragm in order to see the effects of depth of field and image contrast, but in most cases one adjusts the aperture to match the objective N.A. to obtain good image resolution.

In addition to the just described trans-illumination technique, one often uses epi-illumination in microscopy. Epi-illumination means that the specimen is illuminated from the same side as it is being observed, Fig. 5. In this case the objective also functions as a condenser. Epi-illumination is used when studying reflecting or fluorescent specimens. Fluorescence microscopy, which is very common in biomedical applications, will be described in section 1.5.1. The same illumination controls are available to the operator both in trans- and epi-illumination microscopy. Thus the luminous field diaphragm is used to control the size of the specimen area illuminated. Only fine adjustments of this control are usually needed, because when using epi-illumination the size of the illuminated field is automatically scaled in a suitable way by the objective, which also functions as condenser. A difference compared with trans-illumination microscopy, is that the aperture diaphragm is often used for adjusting the light intensity of the illumination. The solid angle (numerical aperture) of the illumination is of no interest in epi-illumination microscopy of fluorescent or highly scattering specimens. The reason is that the illumination ray geometry is totally lost in the fluorescence/scattering process. No matter under what angles the specimen is illuminated, the fluorescence/scattered light is emitted more or less isotropically in all directions.

The illumination can be regarded as just a way of getting energy into the sample so that the fluorescence/scattering process is activated.

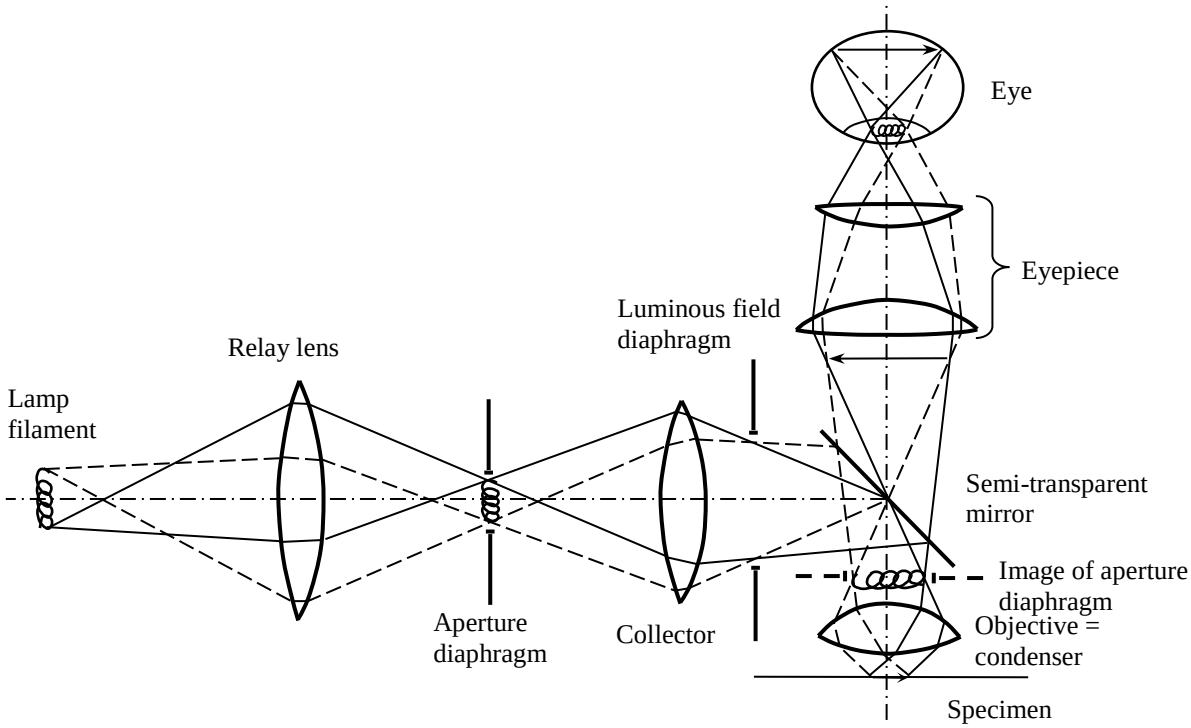


Fig. 5. Illustration of the epi-illumination principle (see Appendix II for infinite tube length case). Epi-illumination is used when studying reflecting or fluorescent specimens. Note that the microscope objective also acts as condenser. Most microscope objectives are not equipped with diaphragms in their back focal planes (such a diaphragm would affect both N.A.$_{obj}$ and N.A.$_{cond}$). Therefore, an aperture diaphragm is located to the left of the collector lens, and imaged by the collector lens so that a real optical image of the aperture is formed in the back focal plane of the objective. To make this arrangement possible, an extra imaging step of the lamp filament is introduced using a relay lens. In fluorescence microscopy the semi-transparent mirror is replaced by a dichroic beam splitter, and color filters are added in both the illumination and imaging ray paths. More on this in section 1.5.1.

1.3 Imaging properties (incoherent case)

As mentioned in section 1.1, the maximum magnification used in microscopy is approximately 1000X. Why stop at this, why not 1,000,000X? To answer this question we must investigate the resolution limit of the microscope, i.e. the smallest object details that can be studied without everything becoming a fuzzy blur. No matter how perfectly the lenses are manufactured such a limit will exist, and it is set by diffraction in the lens openings. Of fundamental importance in the study of resolution is the point spread function, *psf* (*x*,*y*). This is the intensity distribution in the image (x,y) plane when an infinitely small, point-like object is imaged. Such an object can be described by the two-dimensional intensity function $\delta(x,y)$. δ is also

called a Dirac (or delta) function, and is equal to zero except for x = 0 and y = 0, where it reaches infinity. Its integral value over the x,y plane is finite, however. In reality there are, of course, no objects that are described by such delta functions, but it is a good approximation for objects considerably smaller than the resolution limit of the microscope. There are, for example, colloidal gold particles with diameters of 20 nm, and fluorescently labeled microspheres 50 nm in diameter, that are suitable as test objects in microscopy. The image of a point object is derived in many textbooks on optics (see, e.g., ref [1]) resulting in a circularly symmetrical *psf*(*r*), where $r = \sqrt{x^2 + y^2}$. By introducing a normalized optical coordinate, v, see below, the *psf*(*r*) can be expressed as

$$psf(r) = \left(\frac{2J_1(\mathrm{v})}{\mathrm{v}} \right)^2 \qquad (3)$$

which is illustrated in Fig. 6. J_1 is a Bessel function of the first order, and v is given by $\mathrm{v} = \frac{2\pi}{\lambda} r \sin\beta$, where λ is the wavelength of the light, *r* a coordinate in the image plane of the microscope objective, and β is given in Fig. 1♣.

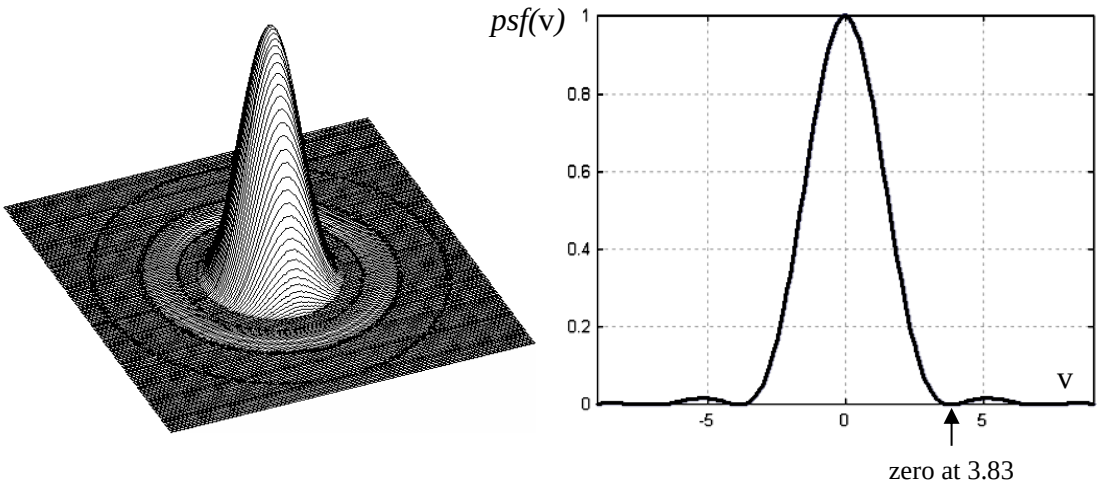


Fig. 6. The point spread function for a diffraction-limited, circular lens.

The *psf* contains all information necessary to completely characterize the imaging properties of an optical system. Classically these properties have been described by the Rayleigh resolution limit, which is the minimum distance between two equally bright point objects if these are to be seen as separate objects in the image plane. If

♣ $\sin\beta$ is called the numerical aperture on the image side. It is related to the numerical aperture on the specimen side, eq. 2, through the equation $\sin\beta = \frac{n\sin\alpha}{M}$, where *M* is the objective magnification. The numerical aperture value labeled on an objective is always related to the specimen side (eq. 2).

the light intensity drops by 26 % between the two image peaks this is considered fulfilled, Fig. 7.

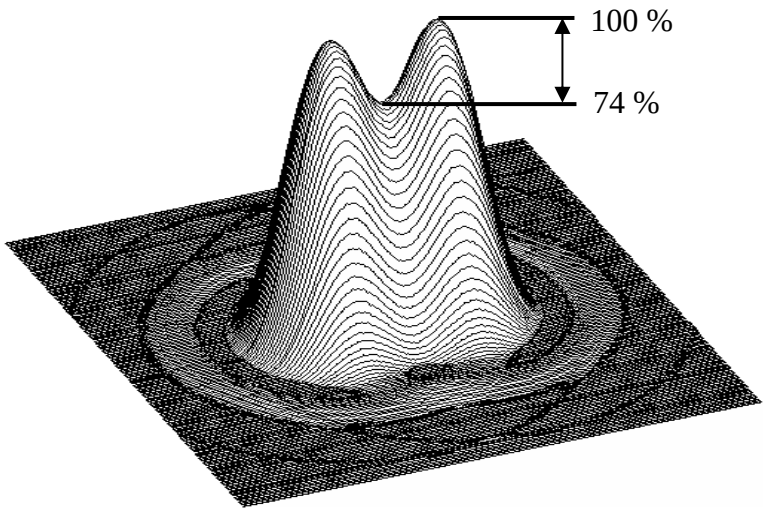


Fig. 7. Resolution limit according to the Rayleigh criterion. Two equally bright point objects close together are imaged. The image will then consist of two superimposed point spread functions (cf. Fig. 6). The resolution is defined as the distance between the point objects that produces an intensity dip of 26 % between the peaks.

Using the Rayleigh criterion and the *psf* of eq. 3, the minimum resolved distance in the microscope specimen, D, is given by

$$D = \frac{0.61\lambda}{N.A.} \qquad (4)$$

for a diffraction-limited objective (i.e. aberrations are negligible). *N.A.* = numerical aperture = n sinα. α is given in Fig. 1, and n is the refractive index of the medium between specimen and objective. Many microscope objectives are of the immersion type, which means that the space between objective and specimen is filled with a fluid (oil, glycerin or water) having n>1. This will increase the N.A., and therefore improve the resolution. Microscope objectives are labeled with their magnification and N.A. Eq. 4 is valid for incoherent imaging, i.e. when the phase relationship between light emitted from different specimen parts varies randomly. This is true when imaging fluorescent specimens♣, and at present we will limit ourselves to the study of incoherent imaging (see sect. 1.4 for coherent imaging). In addition to the objective, a microscope also contains an eyepiece that may affect the imaging properties. The eyepiece has a much simpler task to perform, because it operates on an image that has already been strongly magnified by the objective. In practice the resolution is therefore determined almost entirely by the objective. This means that eyepieces are relatively simple to manufacture, whereas objectives often consist of very complicated lens systems.

♣ λ in Eq. 4 is the fluorescence wavelength, not the illumination wavelength (see sect. 1.5.1)

From eq. 4 we can now calculate the smallest distance between resolved objects in the microscope. Assuming a wavelength of 500 nm and a N.A. of 1.4 (which is near the maximum that is commercially available) we arrive at D = 0.22 μm. This is a fundamental limit that we cannot improve on regardless of the magnification of the microscope. Actually, the magnification does not influence the resolution at all (cf. eq. 4). The only purpose of the magnification is to let the human operator see all the details present in the image formed by the objective. Investigations of the resolving power of the human eye have given results that vary considerably depending on the experimental conditions. It seems, however, that objects separated by less than one minute of arc cannot generally be separated by the eye. Using this figure we get a maximum useful magnification of 330X. By using magnifications of up to 1000X we can be assured that the microscope, and not the eye, sets the resolution limit. On the other hand, higher magnification would be useless, because the image blur caused by diffraction would become clearly visible. In practice the resolution is always lower than that given by eq. 4 because of imperfections in the objective (spherical and chromatic aberrations etc). However, microscope objectives of high quality often perform near the diffraction limit, especially near the center of the field-of-view.

Although the Rayleigh resolution limit gives useful information, a single number can never give reliable information about the imaging of real specimens, e.g. biological tissue. But, as mentioned previously, the *psf* contains all the information necessary to completely characterize the imaging. This information is obtained by taking the Fourier transform of the *psf*, which yields the modulation transfer function, MTF (ref. [9]). The MTF describes how the modulation of different spatial frequencies in the specimen is affected by the objective. Using the *psf* of eq. 3 and the convolution theorem in Fourier analysis we can easily calculate the MTF for the microscope objective. Denoting the Fourier transform FT we get

$$MTF = FT\{h(r)\cdot h(r)\} = H(\nu)\otimes H(\nu) \qquad (5)$$

where $\otimes$ denotes convolution and $h(r) = \dfrac{2J_1\left(\dfrac{2\pi}{\lambda} r\sin\beta\right)}{\dfrac{2\pi}{\lambda} r\sin\beta}$. $H(\nu)$ is the Fourier transform of $h(r)$, and ν represents spatial frequency in the ***specimen*** plane. $H(\nu)$ turns out to be a very simple function, having a constant value for $\nu < \dfrac{N.A.}{\lambda}$ and zero outside. Actually $H(\nu)$ is a scaled version of the pupil function of the objective, see Appendix III for details. According to eq. 5 the MTF is the autocorrelation function of $H(\nu)$, which is the common area of two identical circles as a function of their separation, Fig. 8a, and the resulting MTF curve is plotted in Fig. 8b. The bandwidth of the MTF is $\dfrac{2N.A.}{\lambda}$, corresponding to a period of $\dfrac{\lambda}{2N.A.}$. This is rather

close to the Rayleigh limit of $\frac{0.61\lambda}{N.A.}$. It is quite natural that the Rayleigh limit gives a somewhat more "pessimistic" value, because it corresponds to having a certain minimum modulation in order to resolve the objects.

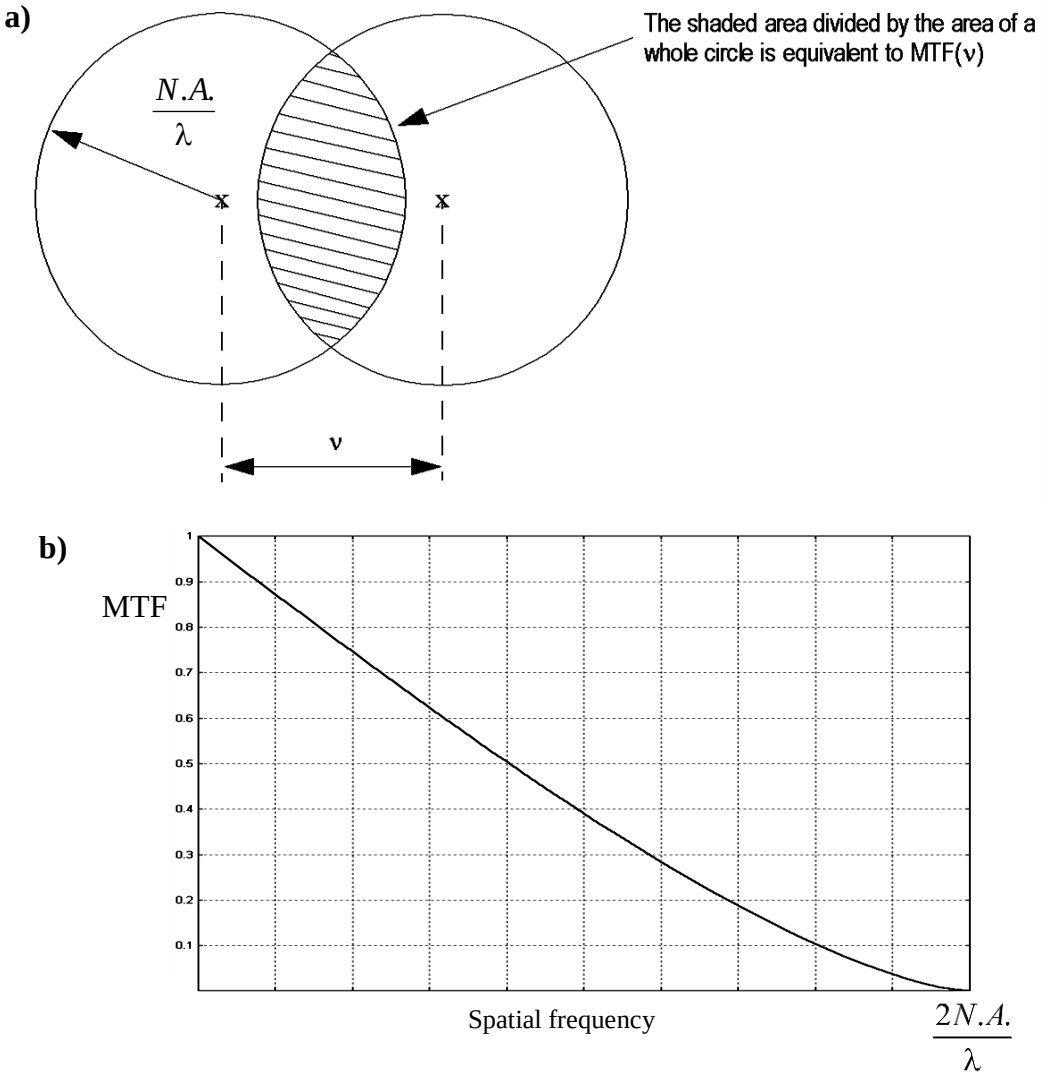


Fig. 8. Modulation transfer function for incoherent imaging in a microscope. a) To calculate the MTF as a function of spatial frequency, ν, the common area of two circles, displaced a distance ν relative to each other is calculated. The result is then normalized by dividing with the value for ν = 0, i.e. the area of one circle. b) The resulting MTF curve. The spatial frequency refers to the ***specimen*** *plane.*

The results we have obtained are valid for a diffraction-limited objective (i.e. an objective free from aberrations). In reality objectives are, of course, not entirely free from aberrations, and therefore the actual performance is expected to be somewhat worse than we have predicted. For well-corrected objectives properly used performance can, however, be rather close to the diffraction-limited case.

1.4. Imaging properties (coherent imaging)

In the previous section the imaging properties in terms of resolution and MTF were described for incoherent imaging. This closely corresponds to the situation in fluorescence microscopy, because all coherence is destroyed in the fluorescence process. In trans-illumination or reflected light microscopy, however, the imaging is usually (more or less) coherent. This is due to the fact that each point of the light source illuminates the entire specimen area that is imaged, see Fig. 4. Coherence complicates things, because we no longer have a system that has a linear intensity response. Coherent systems are linear in amplitude, but this is of little help as long as our detectors respond to the intensity, rather than the amplitude, of the light. As a consequence MTF is not applicable. We will not look at coherent imaging in any detail in this compendium (interested persons can find more information in, e.g., ref. 1). Nevertheless, we will review in a rather simple way what imaging properties to expect, and how they are affected by, for example, the illumination geometry. Just as for the incoherent case, it is possible to determine a resolution limit according to the Rayleigh criterion, see Fig. 7. The *psf* will be the same as for incoherent imaging, eq. 3, but, the wave fields from the two point sources must be added coherently. Furthermore, the numerical aperture of the condenser, $N.A._{cond}$, will slightly influence the result. Applying the Rayleigh criterion, the resolution, D, measured in the specimen plane will be approximately in the interval

$$\frac{0.6\lambda}{N.A._{obj}} \leq D \leq \frac{0.8\lambda}{N.A._{obj}} \qquad (6)$$

depending on $N.A._{cond}$. $N.A._{obj}$ is the numerical aperture of the objective. Comparing this with the incoherent case, eq. 4, we see that the resolution is practically the same.

Instead of point objects, we can, in analogy with MTF theory, look at the imaging of line patterns. If the specimen consists of a line grating illuminated by planar wavefronts perpendicular to the optical axis, we have the situation shown in Fig. 9. This illumination situation corresponds to using a condenser with $N.A._{cond} \approx 0$.

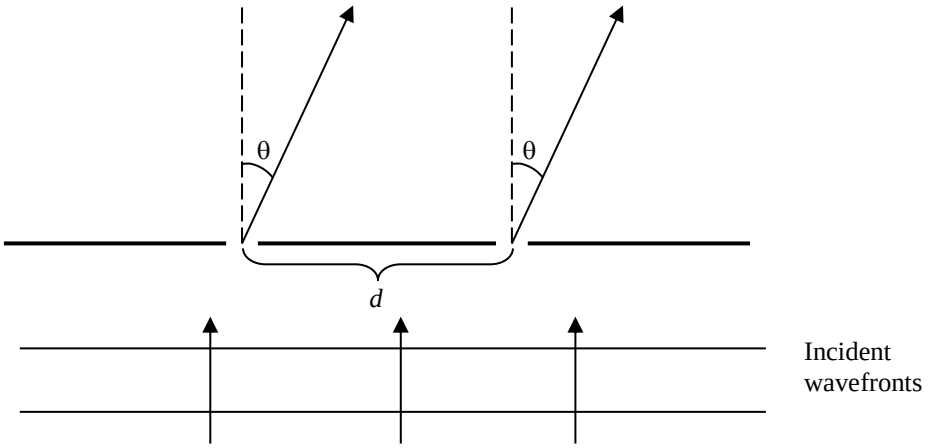


Fig. 9. Geometry when imaging a diffraction grating as described in the text. The rays forming an angle of θ *with respect to the optical axis represent the first diffraction order (an identical order is found to the left of the optical axis, but is omitted in this figure).*

If the grating period is *d*, the first diffracted order will form an angle of θ with respect to the optical axis, where θ is given by $\sin\theta = \frac{\lambda}{d} = \frac{\lambda_0}{nd}$, where *n* is the refractive index of the immersion medium (if any, otherwise use *n* = 1) and λ_0 is the wavelength in vacuum. As long as $n\sin\theta$ is less than the numerical aperture of the objective, the first diffracted order will be transmitted through the objective. According to the theory of Fourier optics, a line pattern will then be reproduced in the image plane, i.e. our grating has been resolved by the microscope. The period, $d_{\min}$, of the densest line pattern that can be resolved is then found from the equation $\frac{\lambda_0}{d_{\min}} = N.A._{obj.}$. $d_{\min}$ obtained from this equation corresponds to a spatial frequency of

$$\nu_{\max} = \frac{N.A._{obj.}}{\lambda_0} \qquad (7)$$

This is the highest spatial frequency that can be imaged by the objective with illumination according to Fig. 9. Now consider what happens if the illuminating wavefronts are planar, but not perpendicular to the optical axis, Fig. 10.

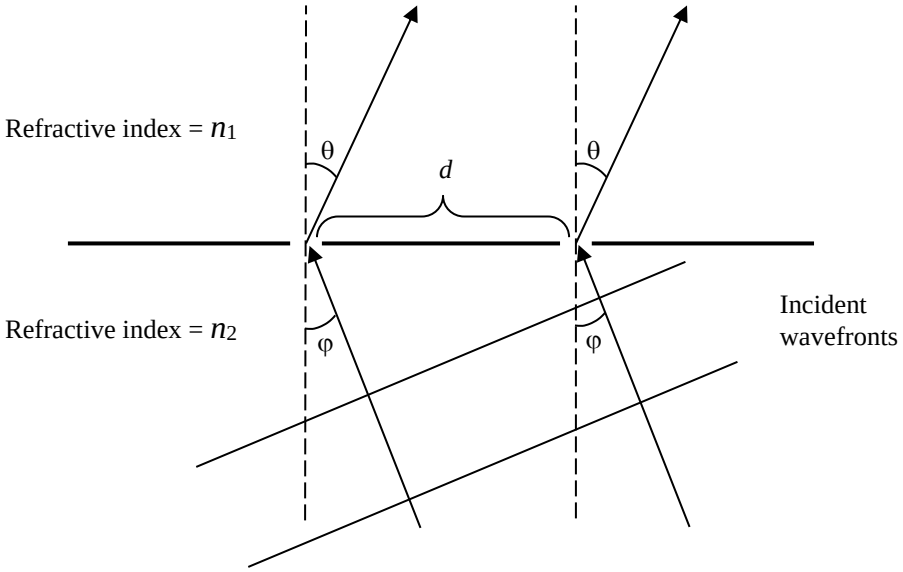


Fig. 10. Geometry when imaging a diffraction grating with incident wavefronts that are not perpendicular to the optical axis.

The first diffracted order now satisfies the condition $n_1 d \sin\theta + n_2 d \sin\varphi = \lambda_0$. There is also a first-order component on the other side of the central undiffracted beam, which satisfies $n_1 d \sin\theta - n_2 d \sin\varphi = \lambda_0$, but this component will not be transmitted by the objective if *d* is relatively small. As long as the objective transmits the undiffracted beam and one of the two first-order components, the line pattern will be visible in the image plane (this is somewhat analogous to the situation when using single-sideband transmission in radio communication). The period, $d_{\min}$, of the densest line pattern that can be resolved is then given by

$$d_{\min} = \frac{\lambda_0}{n_1 \sin\theta_{\max} + n_2 \sin\varphi_{\max}} = \frac{\lambda_0}{N.A._{obj} + N.A._{cond}} \quad (8)$$

Eq. 8 is valid for $N.A._{cond} \le N.A._{obj}$. For the case $N.A._{cond} = N.A._{obj}$ we get

$$\nu_{\max} = \frac{2N.A._{obj.}}{\lambda_0} \quad (9)$$

This is the highest spatial frequency that can be imaged in the case of coherent illumination. It is exactly the same frequency limit as we had in incoherent imaging. The situation is also similar in the sense that the highest frequencies will be imaged with very low contrast. In the case of coherent imaging this is due to the fact that, of all the rays passing through the condenser, only those that form the maximum angle with respect to the optical axis will contribute to the imaging of the highest frequencies♣. In practice, it may therefore be better to use a slightly lower numerical aperture for the condenser than for the objective ($N.A._{cond} \approx 0.7N.A._{obj} - 0.9N.A._{obj}$ is a common recommendation). This reduction, which is done by adjusting the

♣ For the case of coherent imaging where $N.A._{\text{cond}}$ is very small, on the other hand, all spatial frequencies up to the limiting frequency given by eq. 7 are imaged with high contrast.

aperture diaphragm, will only marginally affect the resolution, but it will often reduce the stray light level. For the same reason, stray light reduction, the luminous field diaphragm should not be larger than necessary. Because the aperture diaphragm is used for adjusting $N.A._{cond}$, it should normally not be used for adjusting the light intensity in the microscope (which is another effect of changing the aperture size). The light intensity should instead be adjusted by the lamp voltage or by using gray filters.

1.5. Contrast techniques

Two things are important in microscopy: resolution and contrast. Resolution has been discussed in previous sections. In this section we will look at some methods for creating contrast. The need for contrast techniques is obvious to everyone who has worked with biological samples in microscopy. These samples often consist of a number of cells in some liquid that mostly consists of water. The cells themselves also consist mostly of water, in which some highly transparent structures are embedded, surrounded by a very thin and highly transparent cell membrane. Looking at such a specimen in a microscope, is almost like looking at clear water. Very little can be seen regardless of the resolution capabilities of the microscope. There exist two main categories of techniques for creating contrast in a microscopic specimen. One category is to label the specimen with some chemical substance that changes the light transmission or reflection properties, or that is fluorescent when illuminated by a suitable wavelength. The other category is to visualize by optical means some characteristic of the specimen that is usually invisible. In this category we often find techniques that convert tiny refractive index variations into light intensity variations. In addition, only scattered or diffracted light may be viewed as is done in various dark-field techniques.

The oldest contrast technique is to dye the specimen with some chemical substance (often brightly colored) that binds preferentially to some parts of the specimen, for example cell nuclei. When looking at such a specimen in an ordinary trans-illumination microscope, such as the one shown in Fig. 4, good contrast is obtained through absorption of light in the labeled specimen parts. Much better sensitivity and specificity can, however, be obtained today by using immuno-fluorescence labeling. This type of labeling uses combinations of antibodies and fluorescent substances. Different types of antibodies bind specifically to different substances (called antigens) in a preparation. One type may bind to DNA, whereas another type may bind to a certain signal substance in nerve cells. Antibodies may be chemically coupled to a number of different fluorescent substances, which emit light of different colors. By labeling a specimen with one or more fluorescent antibodies, it is possible to see in great detail the distribution of different substances in a specimen. Immunofluorescence is very much used in biomedical microscopy, and therefore we will look a little closer at the properties of fluorescent dyes. An understanding of these properties is important in order to get good image results, especially if multiple fluorophores are used simultaneously in the same preparation.

1.5.1 Fluorescence labeling

A simplified energy level diagram of a fluorophore molecule is shown in Fig. 11a. Normally the molecule is in the lowest vibrational level of the lowest electronic state S_0. An incident photon of sufficiently high energy can be absorbed, thus raising the molecule to the excited state S_1. Higher states like S_2, S_3 etc may also be reached if the excitation photon has sufficient energy. Immediately after excitation, the molecule is usually in some higher vibrational level, but the vibrational energy is transferred to surrounding molecules within a few picoseconds, leaving the molecule in the lowest vibrational level of the excited state S_1. The average lifetime of the molecule in this excited state is typically a few nanoseconds, but this can vary a lot for different fluorescent substances. From level S_1 the molecule may return to the ground state (level S_0) through emission of a photon (fluorescence light), or the excess energy may just be converted into heat. Immediately after return to the ground state the molecule is usually in a higher vibrational level, but this vibrational energy is again quickly lost through collisions with other molecules. The process can then be repeated.

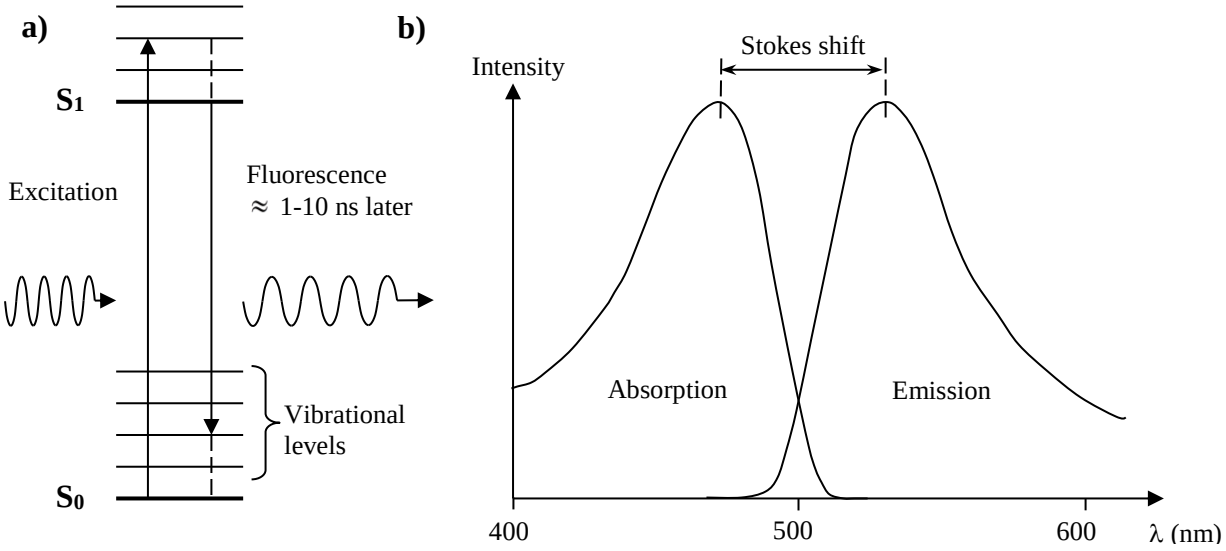


Fig. 11. Characteristics of a typical fluorophore. a) Energy level diagram. b) Absorption and emission spectra.
Superimposed on the vibrational levels in a) are closely spaced rotational levels. Furthermore, the levels are broadened and therefore overlapping, explaining the continuous spectra shown in b).

As a result of the vibrational energy losses, the emitted photon usually has a longer wavelength than the excitation photon, Fig. 11b. The conversion efficiency from excitation to fluorescence light is often very poor. The intensity of the illuminating light is usually many orders of magnitude higher than that of the fluorescence light. In addition to excitation energy being lost to heat, also other processes occur. One is called intersystem crossing, which means that the excited molecule is transformed into a semi-stable triplet state, where it can stay for a very long time (up to seconds) and therefore cannot contribute to the fluorescence process. Another source of energy loss is that many excitation photons travel right through the (usually thin) specimen without being absorbed. Finally, the fluorescence light is

emitted isotropically in all directions, whereas the microscope objective can only collect light under a limited solid angle, Ω, given by

$$\Omega = 2\pi\left(1-\sqrt{1-\frac{N.A._{obj}^2}{n^2}}\right) \Rightarrow \text{Collection efficiency} = \frac{\Omega}{4\pi} = 0.5\times\left(1-\sqrt{1-\frac{N.A._{obj}^2}{n^2}}\right) \qquad (10)$$

where *n* is the refractive index of the immersion medium. As a consequence of all these losses, the fluorescence light coming from the specimen is usually quite weak even if a powerful lamp is used for illumination. In principle the low intensity of the fluorescent light can be compensated for by increasing the exposure time when an image is recorded. This is also done in practice, but it doesn't always work well. First, a long exposure time is sometimes problematic because the specimen may move, either because it is alive or because it is not embedded in a solid medium. Second, a prolonged illumination of a fluorescent specimen often results in so-called photobleaching. This means that the fluorophore molecules will ultimately be destroyed by the excitation light. Photobleaching is a poorly understood process, but the root to the problem is that the fluorophore molecule is less chemically stable when it occupies an excited state, i.e. just before it emits a fluorescent photon. Instead of emitting a photon the molecule may undergo an irreversible chemical change, for example react with molecules in the surrounding medium. As a result of photobleaching, the fluorescence light intensity from the specimen may be dramatically reduced after a short period of exposure. This period may be as short as seconds in an ordinary microscope, and milliseconds or less in a confocal laser-scanning microscope (see sect. 2 of this compendium). On the scale of an individual molecule, this means that it may undergo on the order of ten thousand excitation/emission cycles before it is permanently destroyed. Photobleaching is a statistical process, however, so individual molecules may "live" much longer, or shorter, than this. Photobleaching is a major problem in fluorescence microscopy, and it is therefore important not to expose the specimen to higher excitation intensity, or for longer periods, than necessary. Some fluorophores bleach more easily than others, and this is one important aspect when selecting suitable fluorophores for labeling.

Bleaching is not all bad news, however. It can actually be utilized for extracting useful information about the specimen. One technique for doing this is called fluorescence recovery after photobleaching (FRAP). When using FRAP, part of the specimen is strongly photobleached, for example through illumination with an intense laser beam. Fluorescence in the bleached region will then be very low. It often happens, however, that the fluorescence will recover gradually over time as a result of an inflow of fluorophore-rich liquid. By studying the fluorescence recovery, valuable information about the diffusion properties within a specimen may be obtained.

Most fluorophores used in microscopy should be excited with near UV, blue, or green light. A mercury (Hg) arc lamp is therefore an ideal light source, providing high intensity at wavelengths of 365, 405, 436, and 546 nm, see Fig. 12.

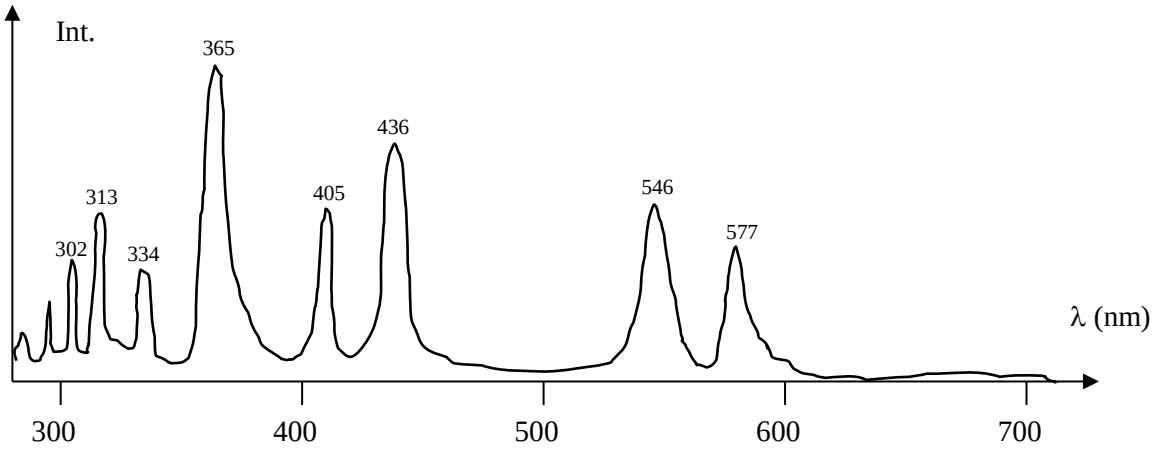


Fig. 12. Spectral distribution of light from a high-pressure mercury lamp. A 50 or 100W lamp of this type is commonly used in fluorescence microscopes.

In fig. 13 the filters used in an epi-fluorescence microscope are illustrated. Using a suitable excitation filter, E, one or several wavelengths can be selected for specimen illumination. The light from the Hg lamp is then reflected by a dichroic beam splitter D. This beam splitter reflects the short-wavelength Hg light (the 436 nm line in Fig. 12) whereas it transmits the longer fluorescence wavelengths. Therefore, when looking at transmission curve D in Fig. 13, one should keep in mind that low transmission means high reflection and vice versa. Dichroic beam splitters are not perfect, however, and therefore some of the Hg light reflected by the specimen can be transmitted and reach the eyepiece. This reflected Hg light can completely ruin the image contrast and, in the case of UV excitation, possibly damage the eyes of the viewer. It is therefore blocked by an additional barrier filter that only transmits wavelengths above a certain value. For good results, there must be absolutely no overlap between the transmission curves for the excitation and barrier filters.

Epi-illumination, rather than trans-illumination, is advantageous in fluorescence microscopy, because thick specimens that do not transmit much light can be studied. In addition, the illumination is very efficient because it is concentrated only to the area of the specimen that is viewed (because the objective is used as condenser as well).

In recent years fluorophores with longer excitation and emission wavelengths have been developed, for example Cy-5, which is best excited around 650 nm and emits in the near IR region (an IR-sensitive camera is used for recording the images). Differently colored fluorophores are needed mainly because specimens are often labeled with multiple fluorophores, coupled to different antibodies. If the fluorophores have different colors, it is possible to simultaneously see the distribution of several substances in the same preparation. The more fluorophores that are used in a single preparation, the more different colors must be utilized. A problem with this, is that the emission spectra of different fluorophores are rather broad and may overlap in some wavelength regions. This spectral overlap can cause serious problems, especially when doing quantitative evaluation. We will return to this subject in the section on confocal microscopy. In addition to the color of the emitted light, fluorophores also differ concerning the lifetime of the excited state of the molecule. This lifetime difference can also be utilized. It turns out that the lifetimes of some fluorophores vary depending on the chemical environment. For example, the pH value may influence the lifetime (the emission spectrum may vary

as well). This opens interesting possibilities to explore the chemical environment on a microscopic scale. In addition to pH, the concentration of ions such as Ca^{2+} and K^+ can be monitored in this way.

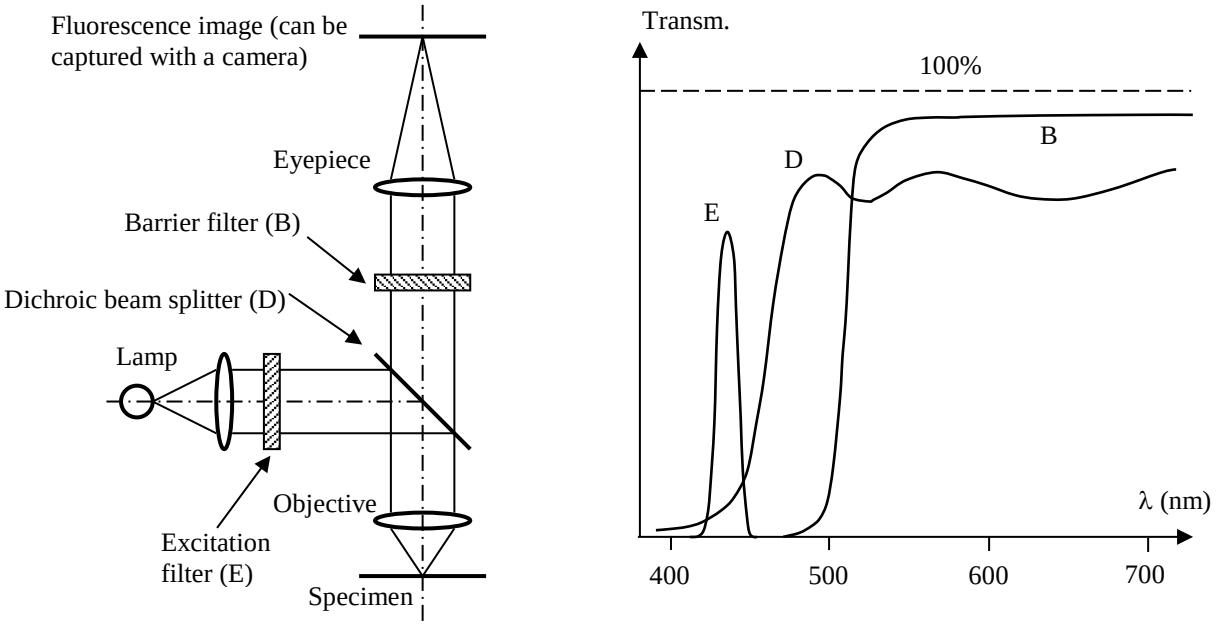


Fig. 13. Simplified schematic diagram of an epi-fluorescence microscope set-up. Several filters are needed for proper operation. The excitation filter selects which of the wavelengths emitted by the lamp that will be used for specimen illumination. In this example we assume that the 436 nm Hg line is to be used. In fluorescence microscopy a dichroic mirror is usually used instead of the semi-transparent mirror shown in Fig. 5. The dichroic mirror is a wavelength-selective mirror that reflects the excitation light, whereas the fluorescent light is transmitted. A barrier filter is needed to block all remnants of reflected excitation light (this filter would not be needed if the dichroic beam splitter were perfect, which, alas, it is not!)

In recent years an alternative to fluorescence labeling has appeared in the form of so-called quantum dots. These are small (a few nanometers) particles made of semiconductor material (cadmium selenide), coated with some material that makes them bio-compatible. Quantum dots have optical properties similar to fluorophores, but they they have several advantages such as narrower emission spectra, no photobleaching and bright light emission. On the negative side we have that quantum dots are more difficult to use for specific labeling (high background), and that they have a tendency to "blink", i.e. change between active (bright) and inactive (dark) states.

1.5.2. Phase imaging techniques

Many biological specimens are highly transparent and colorless, and they are therefore almost invisible. It is sometimes desirable to improve the image contrast without any chemical labeling of the specimen. Labeling takes time, the substances are often expensive and may furthermore be toxic to living specimens. In such cases phase imaging techniques may be useful. These techniques use interference to convert optical pathlength differences into intensity differences. In this way tiny

refractive index differences between cell components like nucleus, membrane and the surrounding medium can be visualized. Although it is possible to build interference microscopes that resemble Michelson interferometers, i.e. with two separate arms, this is not commonly done because of the complexity. Instead a common arm is usually used. Several different phase imaging techniques are used in microscopy. We will look at the two most common of these, phase contrast and differential interference contrast (DIC).

Phase contrast

Phase contrast involves manipulation of the Fourier plane (= back focal plane) of the objective. In Fourier optics a non-absorbing specimen can be described by the function $e^{i\Phi(x,y)}$, where $\Phi(x,y)$ describes optical pathlength variations over the specimen surface♣. If the pathlength variations are small, we can make the approximation $e^{i\Phi(x,y)} \approx 1 + i\Phi(x,y)$. In the back focal plane of the microscope objective we get the Fourier transform (FT) of this function. "1" represents a constant light level over the entire field of view, and its FT will be a dot on the optical axis. $\Phi(x,y)$ on the other hand represents specimen structure, and therefore its FT will be spread out over a large area in the Fourier plane. The trick now is to introduce a "phase plate" that alters the phase of the light that passes near the optical axis compared with light that passes other regions of the Fourier plane, see Fig. 14a). A phase change of $\frac{\pi}{2}$ is introduced between these two regions. As a result, the (amplitude) image function in the microscope is given by $1 + \Phi(x,y)$, i.e. the phase variations have been transformed to amplitude variations. The corresponding intensity function is given by the square of the amplitude function. Therefore, the intensity function will be approximately $1 + 2\Phi(x,y)$.

We have assumed that $\Phi(x,y) << 1$, and therefore the contrast of the intensity image thus created will be rather low. To increase the contrast one usually combines the phase change in the central part of the Fourier plane with an attenuation of this central light. This means that the constant light level represented by the number "1" will be reduced to a smaller number ε, resulting in an intensity function $\varepsilon + \Phi(x,y)$ which has a much higher contrast but lower overall light level.

For this simplified theory of phase contrast to be valid, we require that the specimen is illuminated by planar wavefronts that are perpendicular to the optical axis of the microscope, i.e. $N.A._{cond} \approx 0$. This can be achieved by stopping down the aperture diaphragm to a very small diameter, but as a consequence we cannot fully utilize the resolution of the microscope and the light throughput will also be quite low. In reality a ring-shaped aperture is used. This means that we get a cone-formed illumination of the specimen, which corresponds to the illumination situation in Fig. 10. We therefore get higher resolution. In addition, we also get better light throughput. When using a ring aperture, the phase plate in the back focal plane of the objective will also have to be ring-shaped, Fig. 14b, because in this case the undiffracted zero order will be located along a ring. The basic physical principle,

♣ For further details, see text books on optics, e.g. ref. [1].

however, is the same as previously described. In phase contrast microscopy special objectives are used that have the phase plate built into the back focal plane.

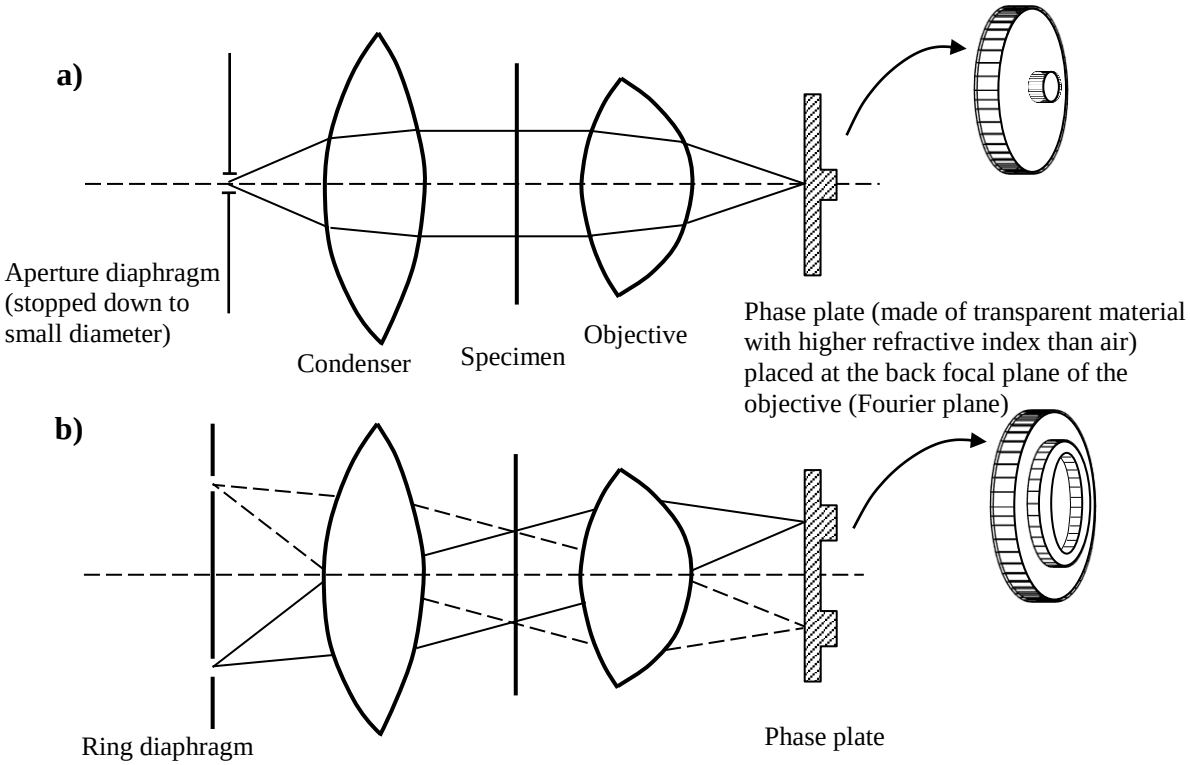


Fig. 14. Phase contrast a) Using low N.A. illumination and a phase plate that retards the rays passing through the central part of the back focal (= Fourier) plane. These central rays represent the undiffracted light (i.e. the constant background, not specimen features) b). In practice, a ring diaphragm is used in the condenser. This provides cone-shaped illumination. The undiffracted light then forms a circle in the back focal plane, and consequently the phase plate consists of an annular ridge.

A final note concerning the phase shift of $\frac{\pi}{2}$ introduced by the phase plate. Because of dispersion, this shift is strictly obtained only for one wavelength of light, commonly 550 nm (green light). Consequently a green filter should be used if proper phase contrast imaging is to be obtained. In practice, however, acceptable contrast can usually be obtained also when using other wavelength bands, or even white light.

In practical use of phase contrast, image artifacts often appear. One effect often encountered is a “halo” around objects such as cells. The reason we get such halos is (among other things) that the ring aperture and phase ring cannot be made infinitely narrow, because then no light would get through. As a result, not only light from the constant background is phase shifted, but also some of the coarser features of the specimen will suffer the same fate. Another source of image artifacts is that the specimen may be too thick, and therefore the assumption that pathlength variations are small may not be valid. Phase contrast therefore works best with thin and highly transparent specimens.

Differential interference contrast (DIC)

In differential interference contrast (DIC) microscopy, image contrast is also dependent on the optical pathlength through the specimen. In this case, however, the light intensity value in the final image is a measure of the rate of change of optical pathlength across the specimen, i.e. the derivative in a certain direction (hence the name "differential"). The difference in this respect between phase contrast and DIC is illustrated in Fig. 15.

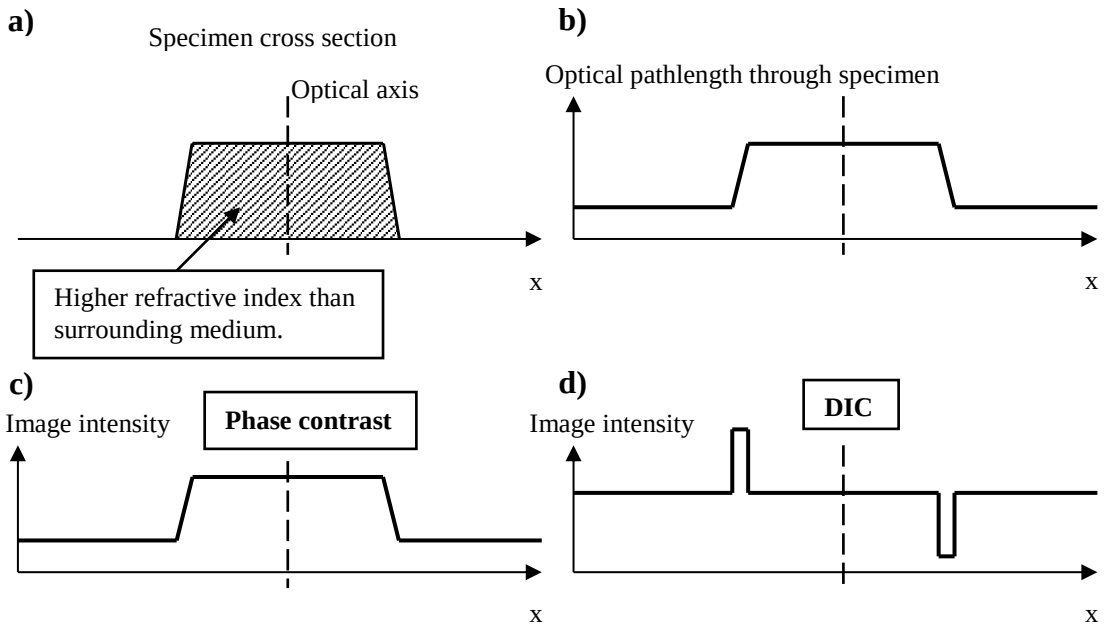


Fig. 15. Comparison of phase contrast and DIC imaging. a) The specimen (dashed area) is assumed to be highly transparent, but having a higher refractive index than the surrounding medium. b) Optical pathlength variation across specimen. c) Image intensity profile when using phase contrast imaging. d) Image intensity profile when using DIC imaging.

DIC images tend to be quite beautiful, often deceptively so because they can easily be misinterpreted. Even people familiar with microscopy tend to be misled by the beautiful "relief landscape" images produced by, for example, a cell. Such images often give a strong three-dimensional impression, with illusions of shadows and highlights that seem to display the topography of a "landscape." In reality, what is shown is the derivative of the optical pathlength through a transparent object in a transparent medium (yes, it is hard to interpret!). That being said, the images can be extremely useful because they clearly display extremely small, unstained, details in a specimen. DIC suffers no halo artifacts like phase contrast microscopy, and it can be used on thicker specimens. So, how does it work?

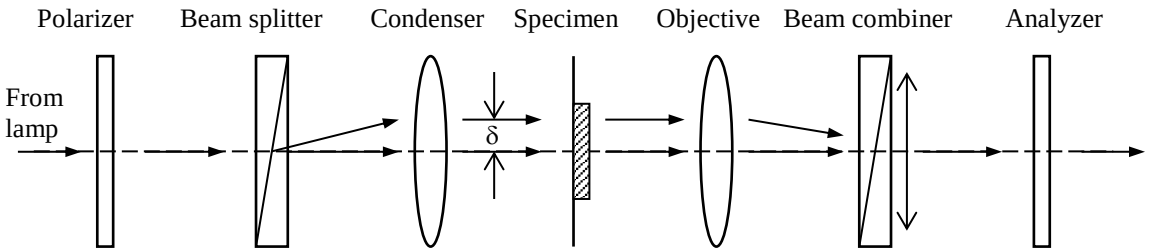


Fig. 16. Simplified schematic of a DIC microscope. Light from the lamp is plane-polarized and incident on a Wollaston prism (beam splitter), where it is split into two components of mutually perpendicular polarization. After the condenser the two components are parallel, but sheared by a small distance δ. They travel through the specimen at slightly different positions, and may therefore suffer different phase retardations. After the objective, the two beams are recombined, and obtain a common polarization direction in a second Wollaston prism and analyzer. At this point the two beams interfere. A constant phase retardation bias between the two beams can be introduced by moving the second Wollaston prism in a direction perpendicular to the optical axis.

In DIC (also called "Nomarski") microscopy the illuminating light is split up into two components with mutually perpendicular polarizations, Fig. 16. These components are sheared sideways by a very small amount, δ, with respect to each other (δ is less than the resolution limit of the microscope) before they pass through the specimen. After passing through the specimen, they are sheared back by the same amount, and their polarizations are made parallel so that they can interfere with each other. This produces a "double image" effect. Two identical images, where one is displaced sideways by a very small amount, are coherently added. Because the displacement is less than the resolution, no double image is observed, the only effect being that the intensity value in each image point is given by the superposition of two light beams passing through two adjacent points in the specimen. The operator can introduce a phase bias between the two interfering beams. This is usually done in such a way that, if the pathlengths through the specimen are equal, the phase difference between the beams is $\frac{\pi}{2}$. The net effect of this will be an image that represents a constant intensity value plus the derivative of the optical pathlength in the shear direction, see Fig. 15d. In addition to the phase bias control, the operator can also vary the shear direction, i.e. the direction of the derivative. The phase difference between the interfering waves will, in general, depend on the wavelength of the light. As a result, the "peak" and "valley" in light intensity shown in Fig. 15d may not be neutral gray in color but slightly tinted. This is not usually considered a problem, and color filters are not needed in DIC microscopy.

1.5.3 Dark-field imaging

A simple way of obtaining image contrast in highly scattering specimens is to use dark-field illumination. This illumination principle is illustrated in Fig. 17. No direct light from the lamp can enter the objective and therefore, in the absence of a specimen, the image is completely black. When a specimen is inserted, light rays may be deviated through scattering, refraction or diffraction so that they enter the objective. In this way specimen features appear bright against a dark background. In this arrangement the numerical aperture of the condenser must be higher than that of the objective. Therefore dark-field illumination cannot be arranged for objectives with very high numerical apertures (like 1.4). Even for objectives with slightly lower *N.A.* it may be necessary to use special condenser designs in order to get satisfactory results. For low *N.A.* objectives, however, dark-field illumination is easy to arrange. A great advantage with dark-field imaging, is that the black background gives images with very high contrast. Therefore very small light-scattering particles, far below the resolution limit of the microscope, can be detected. This does not mean that we have magically surpassed the resolution limits previously described. It only means that we can detect the presence of a small object, but we cannot measure its true size. (For the same reason we can see the stars in the sky, although they are much smaller than the resolution limit of the eye.)

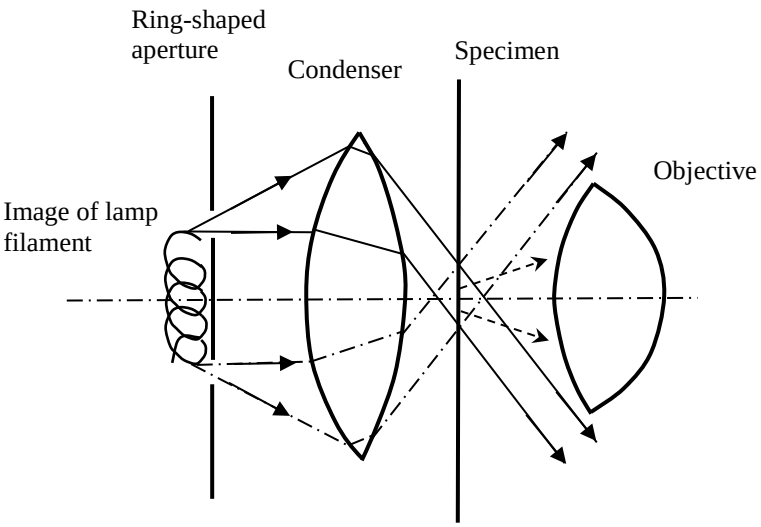


Fig. 17. Dark-field illumination. Only scattered light can enter the microscope objective.

1.6 Microscope photometry

So far we have basically assumed that the microscopic image has been viewed directly by eye through the eyepieces. In actual scientific use, it is quite common to record images. Formerly photographic recording was used for this purpose, but today electronic recording using CCD or CMOS matrix sensors is standard procedure. Such cameras offer several advantages compared with photographic

film. First, they have a much higher quantum efficiency, which is especially important in fluorescence microscopy where the light is weak. Second, no film development is needed. Third, image data from matrix sensors can be immediately downloaded to a computer for image processing and evaluation. Finally, as opposed to film, matrix sensors have a linear intensity response. This facilitates quantitative image evaluation♣. Quantitative evaluation can also be done by using a single detector, for example a photomultiplier tube (PMT), positioned in the image plane of the microscope. Using different masks, the operator can select the area of measurement for the PMT. This single-detector technique is not so common today, because CCD and CMOS sensors usually have sufficient performance for this task, and they offer better flexibility. (PMTs are often used in laser-scanning microscopy, however. More on that later.)

The optical arrangement in the microscope must be slightly different when the image is to be recorded by a camera rather than imaged by eye through the eyepieces. In the first case a real optical image must be projected onto the matrix sensor, whereas in the latter case a virtual image at infinite distance is to be produced. In principle the matrix sensor can be positioned directly in the image plane of the objective, where a real image is formed, Fig. 1. The size of this image may not fit the sensor, however. In such cases an extra imaging step is employed that magnifies or demagnifies the microscope image so that it fits the size of the sensor. To obtain a suitable exposure level, the illuminance level and exposure time are adjusted. As previously explained, the light intensity is often low when imaging fluorescent specimens. It is therefore important to collect the photons as efficiently as possible. Not surprisingly, the collection efficiency depends on the numerical aperture of the objective, but also other factors influence the illuminance level in the image plane. We will therefore present some equations that describe how the illuminance level depends on different parameters (for a fuller account, see ref.[2]).

In the absence of a specimen, the illuminance level♠, E, in the image plane of the objective in a trans-illumination microscope is given by

$$E = \frac{\pi L (N.A.)^2}{M^2} \qquad (11)$$

L is the luminance of the light source, *N.A.* represents the numerical aperture of the objective or condenser, whichever is smallest, and M is the magnification of the objective. In eq. 11 we have neglected light losses due to reflection and absorption in the microscope optics. If a specimen is inserted that does not scatter the incident light, the illuminance E in q. 11 is multiplied by a factor T, where T is the transmission of the specimen.

The situation will be different if the specimen is highly scattering or fluorescent. In this case light is diffusely emitted, regardless of the geometry of the incident light. The illuminance level in the image plane of the microscope objective is then given by

---

♣ Beware! Consumer-type cameras incorporate “gamma correction” resulting in a strongly non-linear response.

♠ See Appendix IV for a summary of quantities used in radiometry and photometry.

$$E \propto \frac{L(N.A._{cond})^2 (N.A._{obj})^2}{M^2} \qquad (12)$$

where L is the luminance of the light source, $N.A._{cond}$ and $N.A._{obj}$ are the numerical apertures of the condenser and objective respectively, and M is the magnification of the objective. The symbol $\propto$ means proportional to, and the proportionality constant depends on the efficiency with which illuminating light is turned into scattered or fluorescent light. This constant can vary a lot for different specimens. In epi-illumination microscopy the objective is also used as condenser. If the full numerical aperture of the objective is used for illumination (i.e. the aperture diaphragm is fully open), we get $N.A._{cond} = N.A._{obj}$ and E is maximized. We then get

$$E_{\max} \propto \frac{L(N.A._{obj})^4}{M^2} \qquad (13)$$

From eqs. 11 to 13 we can identify the factors that influence image illuminance. To get a bright image we should use a lamp with high luminance, the numerical apertures of condenser and objective should be high, and the magnification should be low. It is interesting to note that in all cases E is proportional to lamp *luminance*, i.e. luminous intensity per unit area, rather than total luminous intensity. This means that a light source that is very small and bright (for example a high-pressure mercury or xenon discharge lamp) can produce extremely bright images. Because of the small source area, however, the lamp image projected by the collector onto the aperture diaphragm plane must be strongly enlarged, cf. Fig. 4. If the lamp image is too small, the full numerical aperture of the condenser cannot be utilized (which in turn will lead to a lower E–value). Therefore a collector lens with sufficiently high magnification must be used to fully utilize the brightness potential of discharge lamps. In trans-illumination microscopy image brightness is usually more than sufficient, and often the light has to be attenuated. In fluorescence microscopy, on the other hand, the brightness is often very low and optimization becomes more important.

2. Confocal microscopy.

2.1 What is confocal microscopy?

Although the practical embodiment of a confocal microscope can take many different shapes, the principle can be illustrated by the simple schematic diagram in Fig. 18. Rather than illuminating an extended region of the specimen, as is done in a wide-field microscope, only a single specimen point is illuminated at a time. Although a laser is commonly used, other illumination sources, such as arc lamps, are possible. The size of the illuminated spot is determined by diffraction effects in the optics, and is typically on the order of 0.5 μm. Reflected or fluorescent light from the illuminated specimen spot is collected by the objective and focused onto a small aperture with a diameter in the range 10 to 100 μm. Light passing through this aperture will reach the detector, which is typically a photomultiplier tube

(PMT). Only a single specimen point is illuminated, and therefore the confocal arrangement does not by itself produce an image of the specimen. It is therefore necessary to include a scanning arrangement in all confocal systems.

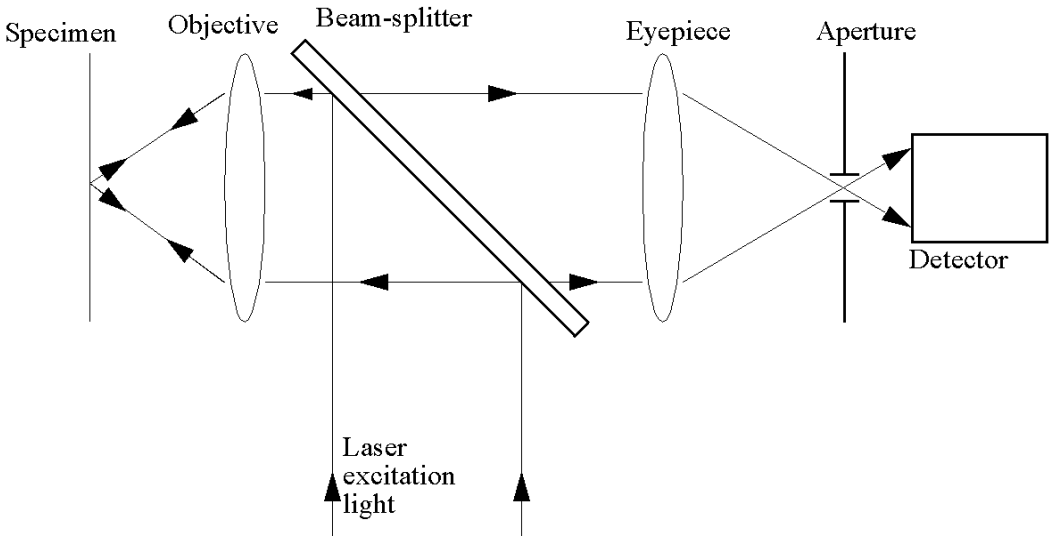


Fig. 18. The confocal principle.

Scanning can be performed in two different ways - either the specimen is moved and the laser beam is kept stationary or vice versa. Many early confocal microscopes used the specimen scanning method, whereas today beam scanning dominates strongly. However, both methods have certain advantages and disadvantages. The specimen scanning method gives uniform imaging properties over the entire image area, because off-axis aberrations and vignetting will not be present. However, specimen scanning requires precise movements of the specimen at high speed, which can be difficult to obtain, especially for heavy specimens. Beam scanning, on the other hand, is much less demanding concerning mechanical precision, because the scanning mechanism can be placed on the image side of the microscope objective, thus acting on a magnified image of the specimen. Also, specimen size and weight is of no concern when beam scanning is used. The disadvantage with beam scanning is that the imaging properties will not be uniform over the image area due to off-axis aberrations and vignetting.

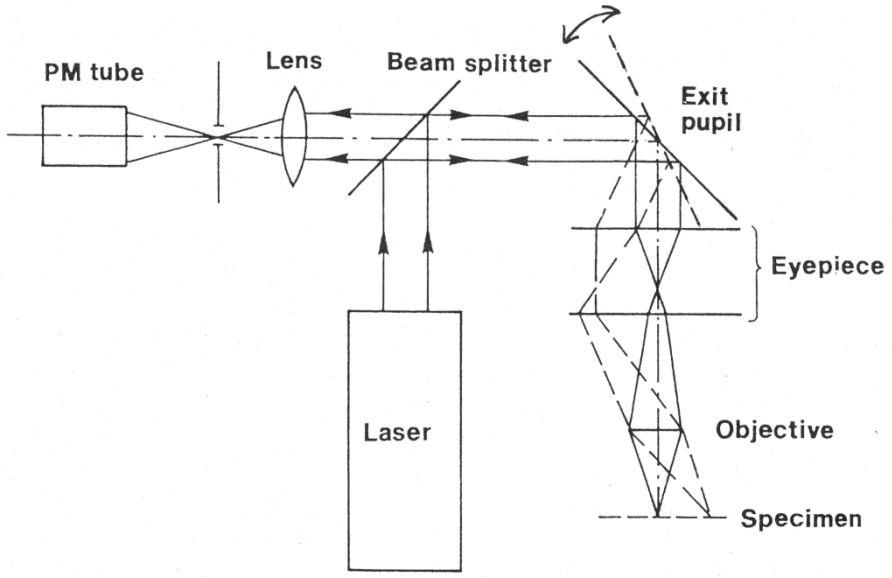


Fig. 19. Scanning of the laser beam across the specimen surface can be performed by a rotating mirror at the exit pupil of the microscope. Fluorescent light from the specimen travels an identical path through the microscope, but in the opposite direction, and is "descanned" by the mirror, forming a stationary beam that is focused onto the detector aperture.

Fig. 19 shows the beam scanning principle adopted by most manufacturers today. A scanning mirror is placed at the exit pupil of the microscope, i.e. at the position where the eye is placed for normal viewing of a specimen (cf. Fig. 4). By rotating the mirror back and forth the focused laser beam will scan along a line in the specimen. Light emanating from the illuminated point in the specimen will be "descanned" by the scanning mirror, forming a stationary beam coinciding with the incoming laser beam, but propagating in the opposite direction. This light is then focused by an additional lens onto the aperture in front of the detector. To obtain a two-dimensional (i.e. area) scanning of the specimen, the scan mirror shown in Fig. 19 is rotated around two perpendicular axes, or an additional scanning mirror is added. The latter method is the most common because it is mechanically simpler to implement.

A practical realization of a confocal microscope system is shown in Fig. 20. The movements of the scanning mirrors are controlled by a microprocessor, which also handles data collection and storage as well as focus control. The recorded images are transferred to a computer for processing and evaluation.

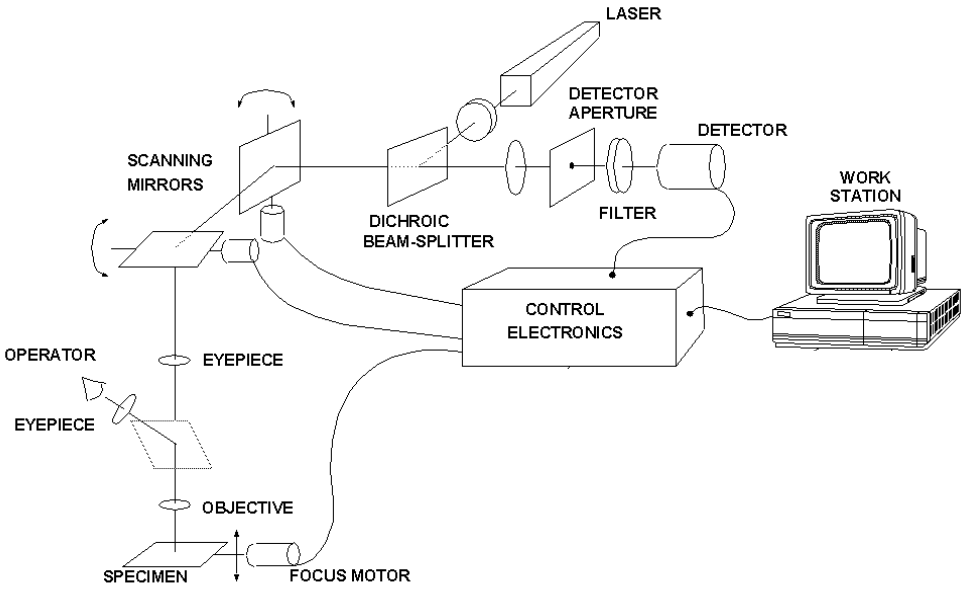


Fig. 20. Schematic diagram of a computer-controlled confocal scanning laser microscope.

From the description given above it is obvious that a confocal microscope is more complicated, and thus more expensive, than a conventional microscope. True, a confocal microscope produces digital images that can be conveniently stored and processed by computers, but this can be accomplished at a much lower cost by attaching a CCD or CMOS matrix camera to a wide-field microscope. The real reason for the interest in confocal microscopy is that it provides improved imaging properties. These properties will be described in the next section.

2.2. Imaging properties of confocal microscopy.

Confocal microscopy is superior to conventional light microscopy for two reasons. First, it produces images with higher resolution. Second, it is possible to record the three-dimensional (3-D) structure of a specimen. We will start by looking at the improved resolution.

As described in sect. 1, the imaging properties of a system can be described in different ways using, e.g., the point spread function (*psf*), the Rayleigh resolution limit, or the modulation transfer function (MTF). MTF provides much more information than the Rayleigh limit, and can be calculated from the *psf*. The *psf* of a diffraction-limited, non-confocal microscope is given by eq. 3. In a confocal microscope both the illuminating laser beam and the light from the specimen are focused to diffraction-limited spots. This double focusing results in a confocal point spread function, $psf_{conf}(r)$, that is the square of the *psf* for conventional microscopy, i.e.

$$psf_{conf}(r) = \left(\frac{2J_1(\mathrm{v})}{\mathrm{v}} \right)^4 \qquad (14)$$

Both *psf*s are shown in Fig. 21. The confocal *psf* is narrower and has smaller sidelobes, both of which are desirable features; a narrow *psf* means a sharper image, and smaller sidelobes means less imaging artifacts.

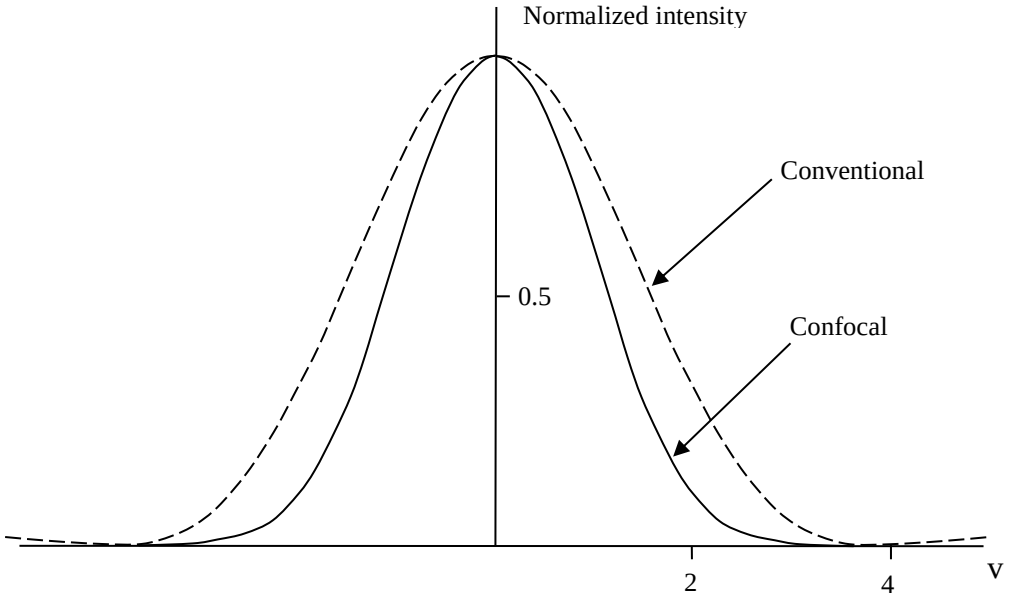


Fig. 21. Point spread function for conventional and confocal microscopy.

To simplify the following analysis we will mainly consider the case of incoherent imaging. We can then use MTF to completely describe the imaging properties. It may seem unrealistic to assume incoherent imaging when a laser is used to illuminate the specimen. However, the most important application of confocal microscopy, especially in biomedicine, is the imaging of fluorescently labeled specimens. The coherence properties of the laser beam are then lost when it is transformed into fluorescence light, making the imaging process incoherent. Occasionally coherent imaging is used in biomedical applications, e.g. for imaging Golgi-labeled specimens, where the laser light is reflected by the specimen rather than transformed into fluorescent light. Therefore, for comparison, we will also briefly mention how the results for coherent imaging differ from those of incoherent imaging.

As previously mentioned, for incoherent imaging the MTF is the Fourier transform of the *psf*. The fact that the confocal *psf* is the square of the non-confocal *psf* means that the confocal MTF is equal to the auto-correlation function of the non-confocal MTF, which was derived in sect. 1. These relationships are illustrated in Fig. 22, showing that the bandwidth for confocal microscopy is twice as large as for conventional microscopy.

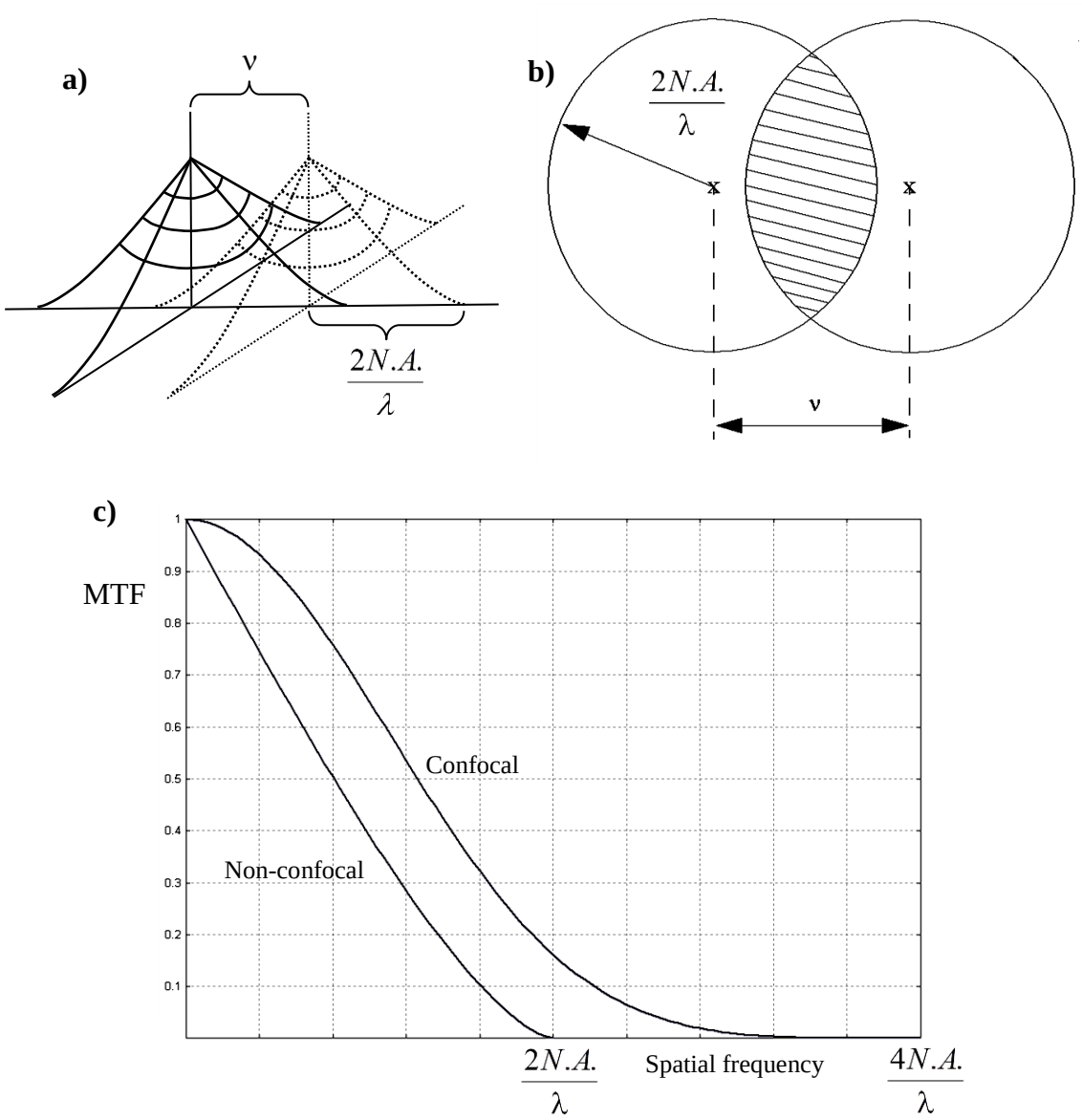


Fig. 22. MTF for confocal incoherent microscopy. a) The confocal MTF is the autocorrelation function of the non-confocal MTF. To calculate this function two rotationally symmetric versions of Fig. 8b (shaped as circus tents) are displaced a distance corresponding to ν, multiplied, and the resulting volume is calculated. b) The area of integration when determining the volume (The circus tents are seen from above). c) Comparison of the confocal and non-confocal MTF curves. All spatial frequencies refer to the specimen plane.

The bandwidths of the MTF curves depend on the wavelength, λ. It is then natural to ask whether this refers to the excitation wavelength or the longer fluorescence wavelength (which is not well defined because it is actually a distribution). Actually the confocal MTF in Fig. 22 presupposes that excitation and fluorescence wavelengths are equal. If this is not the case the bandwidth will be $\frac{2N.A.}{\lambda_1}+\frac{2N.A.}{\lambda_2}$ rather than $\frac{4N.A.}{\lambda}$ (λ_1 = exc. wavelength, λ_2 = fluorescence wavelength). For wide-field microscopy, Fig. 8b, the wavelength refers to that of the fluorescent light.

Although the bandwidth for confocal microscopy is twice as large as for conventional microscopy, the MTF values are very low in the region $\frac{3N.A.}{\lambda}$ to $\frac{4N.A.}{\lambda}$. Therefore it is not fair to say that the resolution is twice as high as in conventional microscopy. As we shall later see, resolution according to the Rayleigh criterion is 36 % higher for confocal imaging.

However, the main advantage of confocal microscopy is not that it provides slightly improved resolution, but that it opens the possibility for three-dimensional recording of specimen structures. The mechanism behind this property is illustrated in Fig. 23.

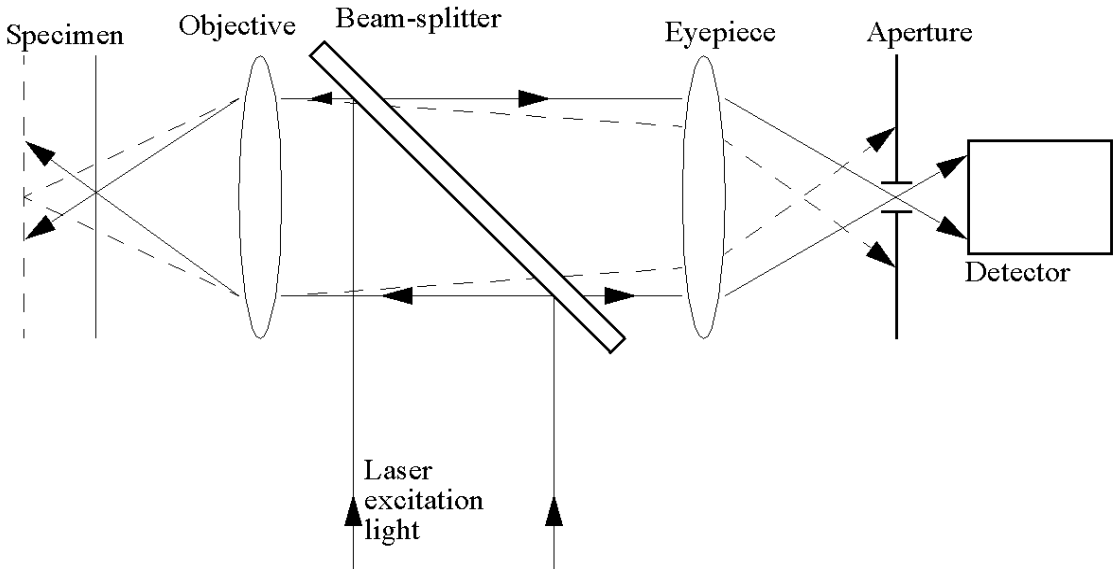


Fig. 23. Illustration of the optical sectioning property of confocal microscopy.

Light from specimen parts in focus can pass through the tiny aperture and reach the detector, whereas light from out-of-focus specimen parts is mostly blocked. The effect of this is that only light from a very thin specimen layer is detected. This property of confocal microscopy is called "optical sectioning," and opens the possibility to study very thin (less than 1 μm) sections of thick specimens. By recording a number of such optical sections, refocusing the microscope slightly between individual sections, it is possible to record a "stack" of images representing the three-dimensional specimen structure. This has several advantages compared with the traditional method of mechanically sectioning the specimen:

1) It is much faster.
2) The recorded sections will be in perfect register, simplifying 3-D reconstruction.
3) Living specimens can be studied.

In order to perform such optical sectioning, it is necessary that the specimen is semi-transparent, so that light can penetrate the uppermost layers to illuminate those

below. This is the case for many biological specimens, and specimen volumes with a depth of several hundred μm are routinely studied with confocal microscopy.
When using confocal microscopy for recording serial sections it is interesting to know what the imaging properties are in the depth direction. It is possible to calculate a 3-D *psf*, Fig. 24, from which a 3-D MTF can be obtained.

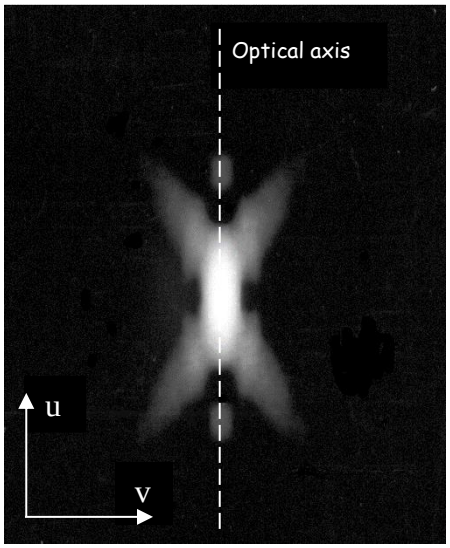


Fig. 24. Three-dimensional confocal point spread function. The psf is rotationally symmetric around the optical axis.

Three-dimensional MTFs are usually illustrated as surfaces, where the MTF value (the height of the surface) is plotted as a function of the spatial frequencies in the radial and depth directions. Examples of such a plots are given in Fig. 25.

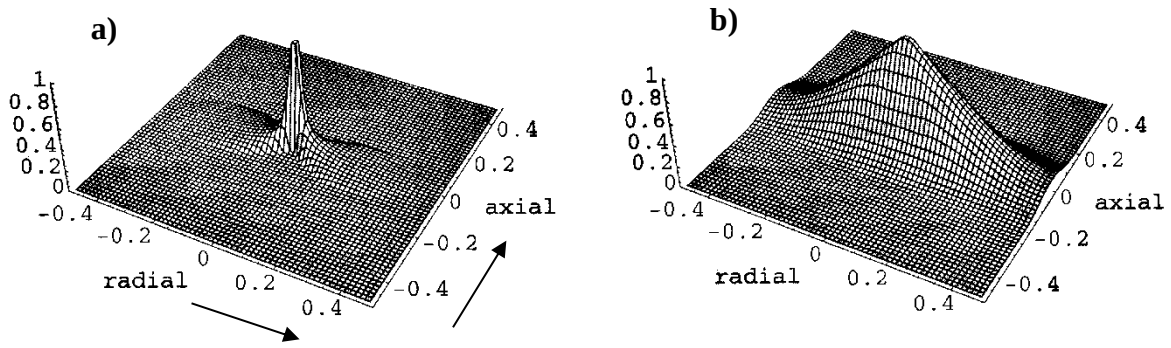


Fig. 25. Examples of how the three-dimensional MTF extends in the radial and axial spatial frequency directions. a) Non-confocal. b) Confocal. (After ref. [3])

These 3-D plots are conceptually identical to conventional MTF curves, but also contain information on how specimen structures in the depth dimension are recorded. Thus, a specimen consisting of alternating flat layers of high and low intensity, and where the layers are oriented perpendicular to the optical axis, is represented by a spatial frequency in the axial direction in Fig. 25 (radial spatial freq. = 0). If the same specimen is oriented so that the layers are parallel to the

optical axis, it can be represented by a frequency in the radial direction. Other orientations will correspond to spatial frequencies located in the radial-axial plane. The larger the volume covered by the MTF surface, the better the imaging properties will be. Comparing Fig. 25a and 25b, we see that the MTF for confocal imaging is clearly superior to the MTF for non-confocal microscopy. However, it is clear that the non-confocal microscope also has some 3-D imaging capability, because its MTF extends into regions of non-zero axial frequencies. Therefore, such microscopes can be used to extract 3-D information from a specimen volume. When used in combination with computer enhancement of the images, this technique can produce quite good results. A serious disadvantage with using non-confocal microscopy for 3-D imaging is that the behavior of the MTF for spatial frequencies in the axial direction is rather erratic. For example, the previously mentioned specimen consisting of alternating layers oriented perpendicular to the optical axis cannot be imaged by a non-confocal microscope. The reason for this is that such a specimen is represented by a spatial frequency in the axial direction only in Fig. 25 (i.e. radial spatial freq. = 0), and in this case the MTF for a non-confocal microscope is zero. This region of zero MTF is called the "missing cone" due to its shape. Because these data are missing in the recorded images from the non-confocal microscope, no amount of computer processing can produce a faithful rendering of the object. The confocal microscope, on the other hand, does not display any missing cone region, and can therefore be used for 3-D imaging of all objects. However, even for a confocal microscope the resolution in the axial direction cannot match that in the lateral direction (i.e. perpendicular to the optical axis). The lateral and vertical spatial frequency limits, $\nu_{r,\max}$ and $\nu_{z,\max}$, for the diffraction-limited case are given by

$$\nu_{r,\max} = \frac{4N.A.}{\lambda} \qquad (15)$$

$$\nu_{z,\max} = \frac{(N.A.)^2}{\lambda n} \qquad (16)$$

where n is the refractive index of the immersion medium. As λ one can use the average value of $\lambda_{excitation}$ and $\lambda_{fluorescence}$[♣]. Eqs. 15 and 16 give a ratio of lateral to vertical resolution that depends on the value of the N.A. At high numerical apertures, say 1.4, the lateral to vertical resolution is approximately four, whereas at N.A. = 0.3 it is 13. These examples show that the vertical resolution is always considerably lower than the lateral, the more so the lower the N.A. Good 3-D imaging properties therefore require that the numerical aperture be as large as possible. For a microscope objective with a numerical aperture of 1.4 (oil immersion), and using a wavelength of 500 nm, the spatial frequency limits are 11 μm^{-1} laterally and 2.6 μm^{-1} vertically.

Although MTF provides a comprehensive description of the imaging properties for all objects, a more intuitive way of describing the imaging performance is to use the classical resolution limit according to Rayleigh, or the halfwidth of the *psf*. Also, it is sometimes nice to be able to summarize the imaging performance using a single

♣ The exact expression for $\nu_{r,max}$ is given on page 36. A corresponding expression for $\nu_{z,max}$ is, however, not correct - things are more complicated in the z-direction.

number rather than using a curve. Therefore these classical resolution limits are often investigated and reported in the scientific literature even today.
We shall now look at the performance of confocal and conventional microscopy according to this classical method. One should keep in mind, however, that they provide rather limited information concerning the image quality for real objects, e.g. biological specimens.
Let us first consider the lateral resolution. As previously mentioned, the resolution according to the classical Rayleigh criterion is 36 % higher for confocal microscopy than for non-confocal microscopy (cf. eq. 4). Thus we obtain for incoherent confocal imaging ($\lambda \approx$ average value of $\lambda_{excitation}$ and $\lambda_{fluorescence}$)

$$D = \frac{0.45\lambda}{N.A.} \qquad (17)$$

It is interesting to notice that this resolution improvement is much smaller than the increase in MTF bandwidth, which is a factor of two. The reason for this difference is that the Rayleigh limit is reached before the MTF has fallen to zero, i.e. we require a certain minimum modulation before the object is considered resolved. Therefore not only the bandwidth but also the shape of the MTF curve will influence the Rayleigh limit. The confocal MTF falls to very low values long before it reaches the bandlimit, and therefore the Rayleigh limit will be considerably lower than the corresponding bandlimit.
Performance in the depth direction is often described by the response to an infinitely thin and infinitely extended horizontal layer, i.e. $I(x,y,z) = \delta(z)$. When imaging such a layer, the recorded intensity as a function of focus position is called the optical sectioning performance, and the halfwidth of the curve is often referred to as optical section thickness. In Fig. 26 this function is plotted for both non-confocal and confocal microscopy as a function of *u*, defined by $u = \frac{8\pi n}{\lambda} z \sin^2\left(\frac{\alpha}{2}\right)$ where *z* is the vertical focus position, *n* is the refractive index of the immersion medium, and α is obtained from N.A. = $n\sin\alpha$.

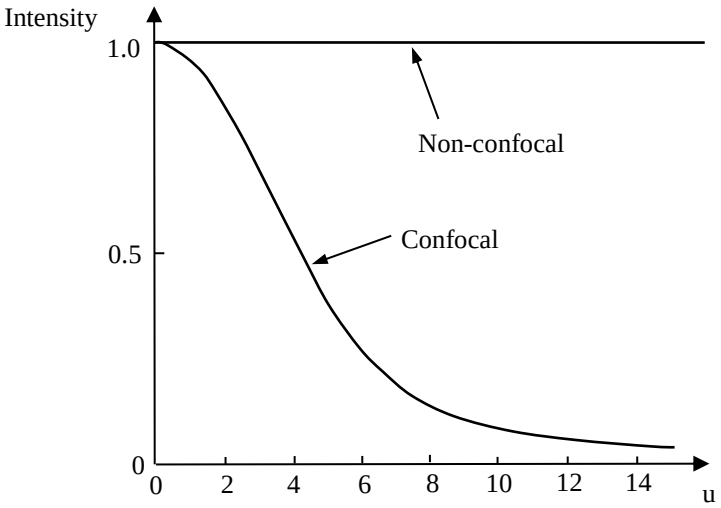


Fig. 26. Light intensity as a function of defocus for a specimen consisting of a planar, infinitely thin, fluorescent layer.

As seen in Fig. 26, the non-confocal microscope displays no depth discrimination whatsoever for this type of object (compare the MTF in Fig. 25) whereas the confocal microscope produces a curve with a full-width-half-maximum (*FWHM*) that can be expressed as

$$FWHM = \frac{8.5\lambda}{8\pi n \sin^2\left(\frac{\alpha}{2}\right)} \qquad (18)$$

where α is calculated from N.A. = $n\sin\alpha$, and $\lambda \approx$ average value of $\lambda_{excitation}$ and $\lambda_{fluorescence}$. This expression for optical section thickness is a bit more complicated than that for lateral resolution, because it cannot be described directly in terms of numerical aperture. However, the depth resolution is approximately inversely proportional to the *square* of the numerical aperture (this is a good approximation for small numerical apertures). This means that the depth resolution depends much more critically on the N.A. than does the lateral resolution. This agrees well with the result we obtained when studying the 3-D MTF. It is therefore important to use the maximum numerical aperture possible to obtain good depth resolution in confocal microscopy. Using an objective with N.A. 1.4 and a wavelength of 500 nm, we get a lateral resolution limit of 0.16 μm and an optical section thickness of 0.40 μm. These figures represent the best performance we can theoretically get with a confocal microscope, and the actual performance is worse due to, for example, optical aberrations and the size of the detector aperture.

The performance described so far relates to incoherent imaging, e.g. fluorescence microscopy. For coherent imaging (i.e. reflected-light microscopy) the properties are somewhat different, but the differences between confocal and non-confocal microscopy will be largely the same. Thus, optical sectioning is obtained in confocal coherent microscopy but not with non-confocal microscopy, Fig. 27. The depth response function, $I'(u)$, for the coherent confocal case can be expressed as

$$I'(u) = \left(\frac{\sin\left(\frac{u}{2}\right)}{\frac{u}{2}} \right)^2 \qquad (19)$$

Compared with incoherent imaging the optical sections are 35 % thinner (8.5 in eq. 18 is replaced by 5.6), but the curve displays secondary maxima (sidelobes), which are not present in incoherent imaging. Coherent imaging is not linear in intensity, and therefore MTF theory is not valid. Instead an amplitude transfer function can be defined to characterize the imaging, but this concept is more difficult to grasp than MTF. If the intensity variations in the object are small, however, imaging will be approximately linear in intensity also for coherent imaging, and in this case the MTF can be calculated. It turns out that the bandlimits for the coherent MTF are half of those for the incoherent case both for confocal and non-confocal microscopy (but the shapes of the curves are different). Therefore, in summary, coherent imaging gives worse lateral resolution, but better depth resolution compared with incoherent imaging.

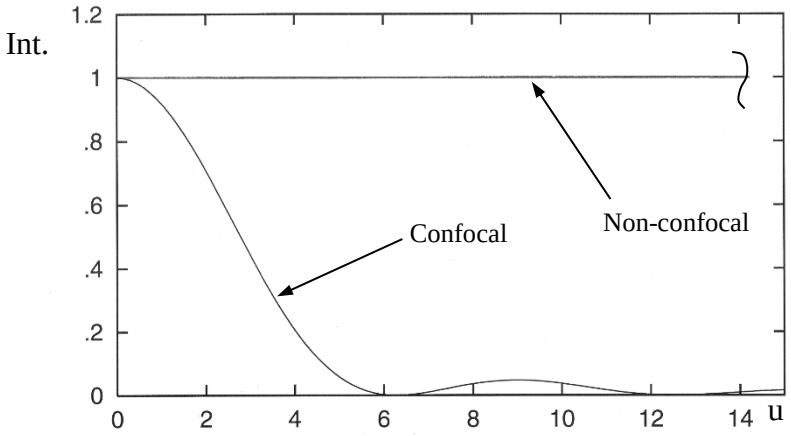


Fig. 27. Light intensity as a function of defocus for a planar reflecting specimen (i.e. coherent imaging)

2.3. Limitations and errors in confocal microscopy

As mentioned, optical aberrations will usually reduce the resolution to values lower than those given above. One serious factor in this context (and one that is easily overlooked) is the specimen itself. If the refractive index of the specimen volume differs from that of the medium above, light rays will be refracted upon entering the specimen, Fig. 28a. The specimen then acts as an unwanted optical component, introducing spherical aberration in the focused beam of light. This influence will become more serious the farther into the specimen volume the beam is focused. Fig. 28b shows how severe this effect can be in a practical case, using an oil immersion objective in combination with a specimen volume consisting of water. For a focusing depth of 300 μm the FWHM optical section thickness has increased by a factor of five. In addition to reduced resolution, the recorded light intensity will be strongly reduced, because aberrations will cause most of the light to miss the detector aperture (the curves in Fig. 28b are normalized).

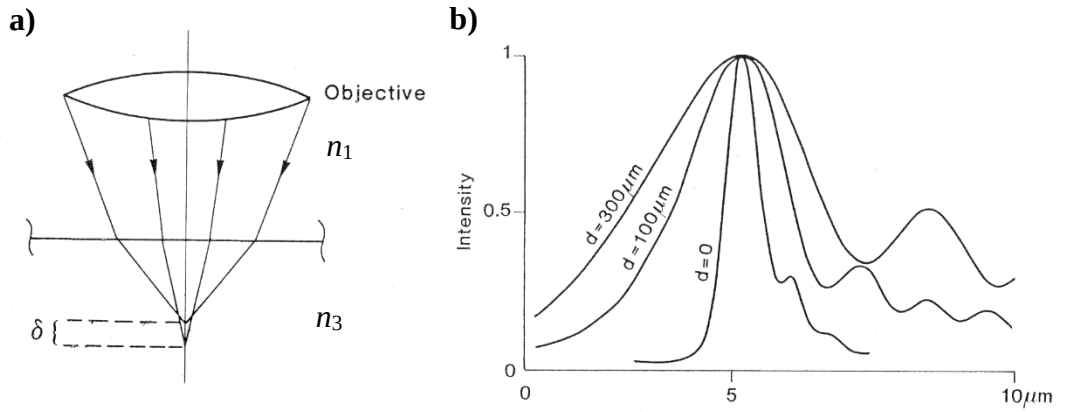


Fig. 28. Influence of specimen refractive index. a) Aberration introduced when light propagates into a specimen medium with a lower refractive index. b) Depth response using a flat mirror as specimen. An oil immersion 40/1.0 objective was used, and the depth response recorded when imaging through water layers of three different thicknesses, d.

Another effect associated with specimen refractive index should also be mentioned. The fact that light rays are refracted at the boundary between the specimen and the surrounding medium, as shown in Fig. 28a, means that geometric distortion will be introduced in the depth dimension. If the refractive index of the specimen is higher than that of the medium above the recorded depth scale becomes compressed, and if it is lower the scale will be expanded. This effect can be described approximately by the equation

$$z' = \frac{n_1}{n_3} z \qquad (20)$$

where z' is a vertical distance measured in the stack of recorded section images, and z is the true vertical distance in the specimen volume. n_1 and n_3 are the refractive indices of immersion medium and specimen, respectively. This effect is analogous to the well-known distortion effect obtained when looking at objects submerged in water (the swimming pool effect). Unless corrected for, it will produce an erroneous ratio of lateral to depth dimensions in the recorded stack of images. Fortunately the effect is easy to (approximately) compensate for if the refractive index of the specimen is known - one simply adjusts the depth scale by a factor equal to the ratio of the refractive indices in the specimen and the medium in contact with the objective. It doesn't matter if a cover glass is used on the specimen or not. See Appendix V for further details on depth distortion and how it affects resolution.

In all the previous discussions of confocal microscopy we have tacitly assumed that the detector aperture is infinitely small. This is, of course, not realistic because no light would be detected and we would not be able to see all the beautiful image results. Confocal theory predicts that as the aperture size is successively increased the imaging properties will approach those of a non-confocal microscope (but they will never get worse even for an infinitely large aperture). As the aperture is opened up the lateral resolution will be the first to suffer. This begins when the diameter reaches $\frac{0.15\lambda M}{N.A.}$, and for an aperture diameter of $\frac{\lambda M}{N.A.}$ the lateral resolution is approximately equal to that of a non-confocal microscope (M is the objective magnification). In fluorescence microscopy the light level is often quite low, and therefore aperture sizes near the upper limit are often used. It is therefore difficult to obtain improved lateral resolution by using confocal fluorescence microscopy. The situation is somewhat different in the depth direction. Here the resolution is not so sensitive to aperture size, and therefore the depth resolution can be almost fully utilized even for confocal fluorescence microscopy. If the maximum permissible deterioration in depth resolution (measured as FWHM) due to aperture size is set to 10%, the maximum diameter of the aperture is approximately

$$D = \frac{0.8\lambda M}{N.A.} \qquad (21)$$

Thus, using a 100X/1.4 objective and a wavelength of 500 nm, the aperture should not be larger than 30 μm. The results should only be taken as a guideline because the effect of the aperture size on depth resolution depends on a number of factors, such as the Stokes shift of the fluorescent light (the wavelength in eq. 21 is the

fluorescence wavelength). For reflected-light confocal microscopy, the maximum diameter of the detector aperture is given by

$$D = \frac{0.95\lambda M}{N.A.} \qquad (22)$$

again provided that a FWHM depth resolution 10% larger than for an infinitely small aperture can be accepted.

Eqs. 21 & 22 refer to the case where the aperture is located in the image plane of the objective. In many confocal microscopes there are relay optics between the image plane of the objective and the actual position of the aperture (cf. Fig. 19). In such cases *M* should be the total magnification from specimen to aperture.

To simplify the task of choosing an appropriate aperture size, so called Airy units are used in many confocal microscopes. One Airy unit (AU) corresponds to the diameter of the central maximum of the diffraction-limited *psf* in the image plane of the objective (Eq. 3 and Fig. 6). We get

$$1\ \mathrm{AU} = \frac{1.22\lambda M}{N.A.} \qquad (23)$$

An aperture with a diameter of 1 AU will transmit approximately 84 % of the light intensity (provided the objective is diffraction limited). The aperture sizes given in Eqs. 21 and 22 correspond to 0.66 and 0.78 AU respectively. This means that approximately 75 and 80 % of the light intensity will be transmitted to the detector for those aperture sizes. And to fully utilize the improved lateral resolution we would need an aperture size of 0.12 AU. This would mean that only about 5 % of the light intensity is recorded.

2.4. Illumination and filter choice in confocal microscopy

To efficiently excite fluorescent specimens, the laser should emit wavelengths in the same general region as a mercury lamp. A good candidate for this is the argon ion laser, which has been used a lot in confocal microscopes. Argon lasers are capable of emitting the wavelengths 364, 458, 488, and 514 nm. The 364 nm wavelength can only be obtained from relatively large and expensive lasers. For that reason, and the fact that aberrations in the microscope objective usually become quite severe in the UV, the most commonly used wavelengths are 458, 488, and 514 nm. In cases where relatively low excitation power is needed, He-Ne lasers can be used. In addition to the "classic" 633 nm wavelength, He-Ne lasers can deliver 543 and 594 nm light. These wavelengths can be used for exciting red- or IR-emitting fluorophores. Recently diode lasers have become more common in confocal microscopy, and this trend will probably continue. Diode lasers have many advantages, such as small size and low power consumption (no cooling needed). They can also be designed to deliver more or less any wavelength that is required.

The illumination method in confocal fluorescence microscopy resembles very much the technique used in non-confocal epi-fluorescence microscopy, cf. Figs 13 and

18. A dichroic beam-splitter reflects the laser beam toward the specimen, whereas the longer-wavelength fluorescent light traveling in the opposite direction will be transmitted and reaches the detector. A barrier filter in front of the detector prevents the small remainder of laser light reflected by the specimen, and transmitted through the beam splitter, from being recorded.

Biological specimens are often labeled with multiple fluorophores. It is therefore interesting to simultaneously record light from several fluorophores with a confocal laser-scanning microscope. For this purpose many confocal microscopes are equipped with multiple detectors and beam-splitters to direct different parts of the fluorescent spectrum to different detectors. We will here describe the principle of a dual-detector system in some detail, but the extension to more detectors is straightforward. Let us assume that we want to record the light from one green emitting and one red-emitting fluorophore, say FITC and TRITC. Excitation and emission spectra for these fluorophores are shown in Fig. 29. To efficiently excite both fluorophores simultaneously we need to use two different excitation wavelengths, in this case we assume 488 nm from an argon laser (FITC excitation) and 568 nm from a krypton laser (TRITC excitation).

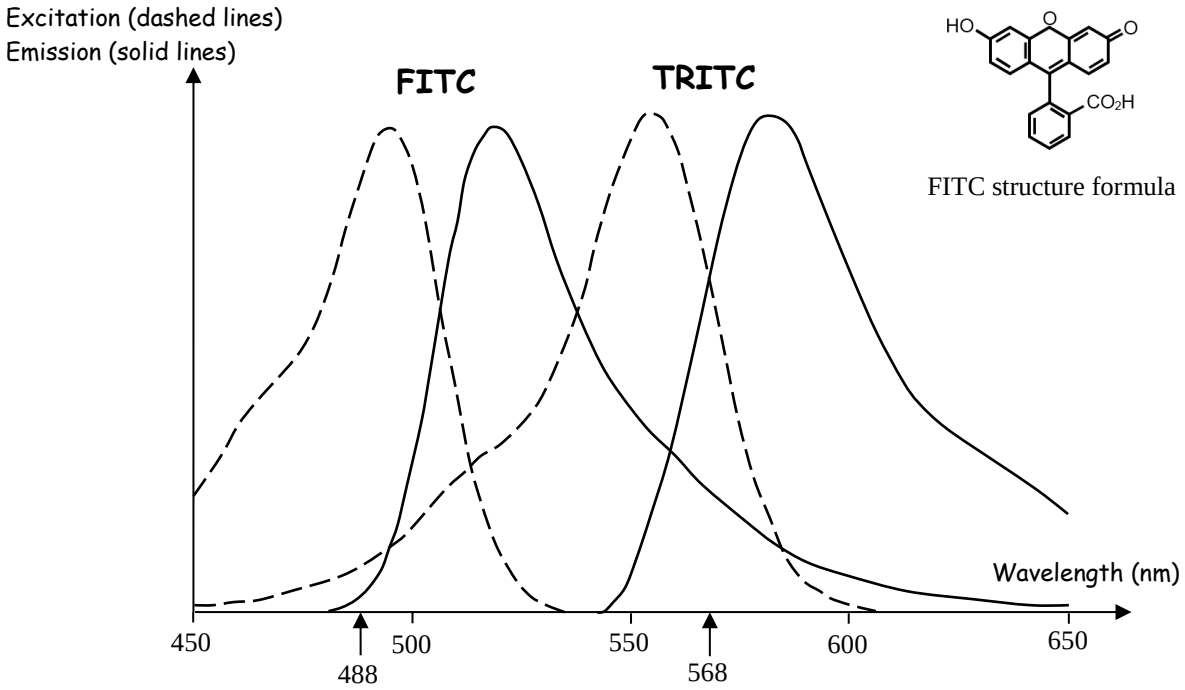


Fig. 29. Excitation and emission spectra for two commonly used fluorophores, FITC (fluorescein-isothiocyanate) and TRITC (tetramethylrhodamine-isothiocyanate). The excitation spectra show how efficiently different wavelengths can excite the fluorophores. Indicated with arrows are two commonly used excitation wavelengths, 488 and 568 nm.

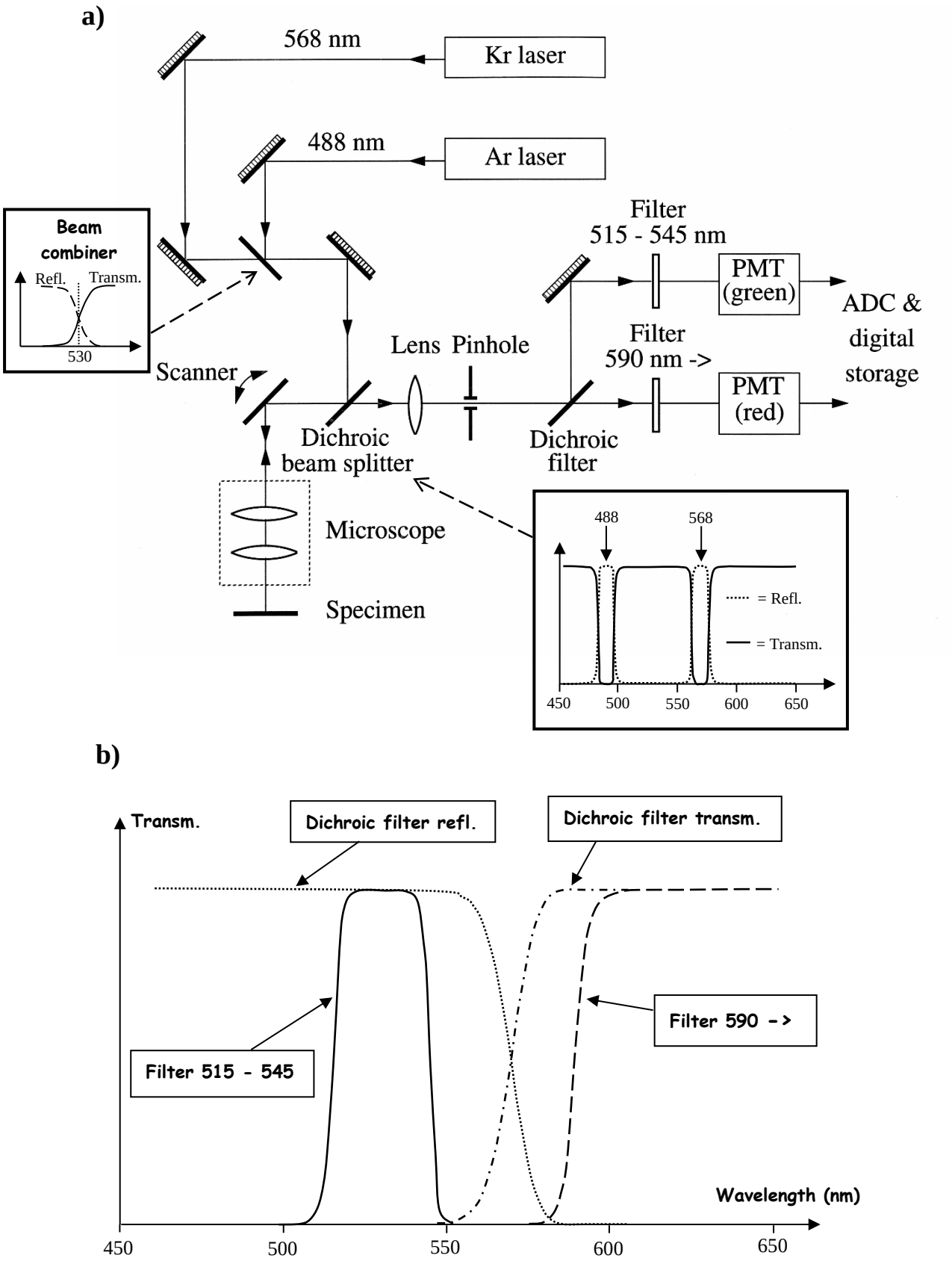


Fig. 30. a) Confocal microscope with two photomultiplier tube (PMT) detectors. b) Spectral transmission characteristics for filters in the detection unit, i.e. after the pinhole. See text for details.

A schematic diagram of the dual-detector system is shown in Fig. 30a. In the illumination ray path a beam combiner is used that reflects 488 nm, whereas 568 nm is transmitted. A dichroic beam splitter is then used that reflects both of these wavelengths, whereas other wavelengths are transmitted. Both 488 and 568 nm wavelengths simultaneously illuminate the specimen, exciting both fluorophores. The mixture of green and red fluorescent light travels up through the microscope, is transmitted through the dichroic beam splitter, and reaches the detector aperture (pinhole). After the pinhole a dichroic filter reflects the green part of the spectrum whereas the red part is transmitted. This means that light from FITC is detected by the "green" detector and light from TRITC is detected by the "red" detector.
Additional detector filters, Fig. 30b, immediately in front of the PMTs restrict the recorded spectrum, and filter out any remaining laser light. Most commercial confocal microscopes today have at least three detection channels. In addition to fluorescence, reflected or transmitted laser light may also be recorded. The more expensive confocals sometimes use more sophisticated spectral recording, incorporating prisms or gratings, and a large number of detection channels.

There is a problem, however, with the dual- (and more) detector recordings. This has to do with the broad and partially overlapping emission spectra of the fluorophores, shown in Fig. 31 for FITC and TRITC together with the spectral regions detected by the two PMTs. We see that the emission from FITC has a long red tail extending into the sensitivity region of the red detector. Likewise, but to a smaller extent, TRITC has a green tail that is detected by the green detector. This will produce "crosstalk" between the recording channels. Because of this we cannot tell from which fluorophore the light recorded in each channel originates. The situation is further complicated by the fact that both fluorophores can be co-localized, meaning that simply comparing signal strengths in the two channels is not sufficient to solve the problem. Nevertheless, it can be shown that forming the proper linear combination between the two channels, crosstalk can in principle be eliminated (ref. [5]). Such compensation methods are often incorporated in software in commercial systems. These methods often strongly reduce the crosstalk, but in many cases it is far from eliminated. For the compensation method to work, it is necessary that the emission spectra of the fluorophores are constant for different specimen parts. Unfortunately this is not the case in many specimens. Factors such as concentration of the fluorophore, or its chemical environment, will affect the emission spectra. Additional methods have therefore been developed to reduce crosstalk. These methods can utilize, for example, absorption spectra and fluorescence lifetimes. It is also possible to record the actual spectral distribution of the fluorescent light in each pixel, to gain more information that can be used to reduce crosstalk. The improved channel separation provided by these additional methods comes at a price, namely added complexity (ref. [6]).

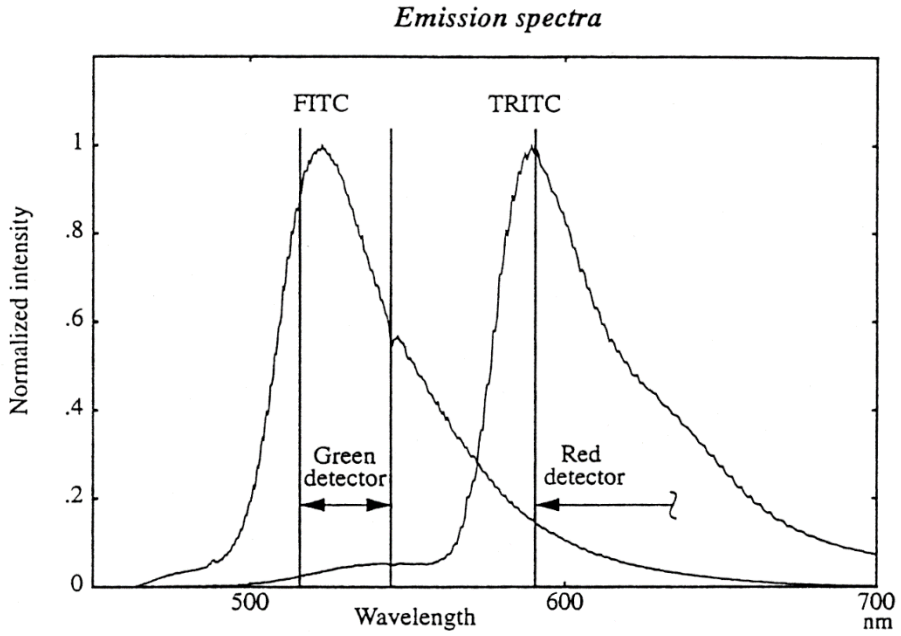


Fig. 31. Emission spectra for FITC and TRITC. The spectral passbands for the two detector filters are also indicated.

3. Recent microscopy techniques

3.1 Introduction

New microscopy techniques are constantly being developed to meet the needs of various users. In one case it may be an improved contrast technique for a specific application; in another case it may be a general method to improve the rendering of small details in the specimen. In this last section we will look at some rather recent techniques for improving microscopic imaging. They have been somewhat arbitrarily selected, and many more could have been added. The selection was made based on the criterion that the techniques should have a rather broad potential application, and/or they should fundamentally improve some aspect of microscopic imaging. Most of the techniques have in common that they can improve some aspect of resolution, and thereby potentially display small specimen structures with better clarity.

It may sound strange that resolution and MTF can be improved so that they exceed the theoretical limits given by eq. 4 and Fig. 8. After all, these results show the ultimate limits that can be reached with a diffraction-limited objective. But we have already seen that confocal microscopy is a way of improving the results by using point, rather than wide-field, illumination. As a matter of fact it has been known for a long time that, at least theoretically, there is no limit to the resolution or optical bandwidth in terms of spatial frequencies that can be obtained in a microscope. It

is “only” a matter of finding a good practical method for carrying out such a resolution improvement. One can use either linear methods, i.e. methods where the recorded light signal depends linearly on the illumination intensity, or one can use non-linear methods. Linear methods like confocal microscopy or structured illumination (sect. 3.5) can double the optical bandwidth (from $\frac{2N.A.}{\lambda}$ to $\frac{4N.A.}{\lambda}$). For larger improvements one can use nearfield microscopy (sect. 3.8), statistical methods (sect. 3.3) or non-linear methods, for example STED (sect. 3.4) or non-linear structured illumination (sect. 3.5).

It is also interesting to improve the depth resolution, so that thinner optical sections can be obtained than what is possible with a confocal microscope. This is possible with some of the recently developed techniques.

Another approach to improve the resolution is to use shorter wavelengths, such as ultraviolet and X-rays. Such techniques will not be covered in this compendium, which deals only with visible, or close to visible, wavelengths. It is very challenging to work at very short wavelengths, because both radiation sources and optics often have to be very advanced and use “exotic” technologies.

3.2 Two-photon microscopy

An interesting effect occurs when a fluorescent specimen is illuminated with excitation light of very high intensity. In this case there is a reasonable probability that two photons arrive at the fluorophore molecule more or less simultaneously (say within a femtosecond or so), and then the energies of the two photons can combine to excite the molecule from state S_0 state to S_1 (cf. Fig 11). This process is called two-photon excitation (TPE). Because the energies of the two photons are added, the wavelength of the excitation light can be twice as long as in ordinary (one-photon) excitation. Thus a fluorophore like FITC, Fig. 29, can be excited with wavelengths around 950 nm. The fluorescent light will still have the same color (green/yellow), so the fluorescence light has a ***shorter*** wavelength than the exciting light.

When using one-photon excitation (OPE), the intensity of the fluorescent light is (within limits) directly proportional to the intensity of the exciting light. When using TPE, the intensity of the fluorescent light is proportional to the ***square*** of the intensity of the excitation light. This means that TPE will be more efficient if the available excitation photons are concentrated in brief pulses, than if they are spread out uniformly in time. Therefore pulsed lasers are usually used for TPE. Typical pulse lengths are from 100 femtoseconds to 1 picosecond (which means rather expensive lasers). TPE will also be more efficient if the excitation light is focused to a small spot than if it is spread out over a larger area.

TPE is sometimes used as an alternative to confocal microscopy. Exactly as in confocal microscopy, the specimen is illuminated by a focused laser beam that is scanned over the specimen, and the fluorescent light is detected. TPE has an important advantage when studying thick and highly scattering specimens. Because of the longer wavelength, the excitation light will be scattered and absorbed to a

lesser extent than shorter wavelengths. Therefore we get better penetration depth. Furthermore, there is no need to collect the fluorescent light via the microscope objective and focus it on a detector aperture, as is done in confocal microscopy (Fig. 18). Instead one can collect as much as possible of the fluorescent light, either via the microscope objective, or with detectors placed near the specimen, and still get optical sectioning similar to that in confocal microscopy. Optical sectioning is obtained in TPE microscopy because the fluorescence intensity depends on the square of the excitation intensity. Therefore the total fluorescence intensity emitted from the specimen will come mainly from the depth layer where the excitation beam is tightly focused (i.e. in the focus plane of the microscope). Of course TPE can be combined with confocal detection of the fluorescent light to obtain even thinner optical sections. But this is usually not done, because the intensity of the fluorescent light is typically rather low. The light level would therefore be unacceptably low when using a detector aperture.

An important advantage with TPE is that photobleaching is less than when using confocal OPE microscopy. The reason is that for each recorded optical section image, all the depth layers are hit by the (more or less) defocused laser beam that scans over the specimen. If many section images are recorded this is repeated many times, often resulting in severe photobleaching of the fluorophore molecules in the OPE case. When using TPE, on the other hand, excitation will be negligible in out-of-focus planes due to the quadratic dependence on excitation intensity. Because photodamage occurs when the molecule is in the excited state, photobleaching will be restricted basically to the in-focus plane. Because each depth layer in the specimen is scanned only once, the accumulated damage is much lower.

Laser-scanning microscopes based on TPE have a number of disadvantages, however. One disadvantage is that the laser used to produce the extremely short pulses of excitation light is very expensive. There is also a risk that the pulses will be broadened as they travel through the optics and the specimen. Such pulse broadening can be compensated for, but this requires further equipment. A brute force method would be to use higher average power instead of extremely short pulses. This is cheaper, but there is a great risk that the higher average power will damage the specimen. Even when using ultra-short pulses one is often near the damage limit. Specimen damage, however, depends on the type of specimen and wavelength used, and it has even been demonstrated that completely unpulsed laser light can, in some cases, be used for TPE.

The resolution in TPE microscopy is lower than in confocal microscopy. This is a natural consequence of the longer wavelength used for excitation. (Remember that all the *imaging* is done with excitation light, the fluorescent light is never focused to an image, and hence its wavelength does not influence the resolution) The point spread function in TPE microscopy is given by

$$psf(r) = \left(\frac{2J_1(\mathrm{v})}{\mathrm{v}} \right)^4 \qquad (24)$$

This is exactly the same mathematical expression as in eq. 14, which gives the point spread function for confocal OPE microscopy. One must keep in mind, however, that $v = \frac{2\pi}{\lambda} r \sin\beta$ is not the same in these two cases, because λ is twice as long for TPE. This gives a *psf* with twice the width, yielding a resolution that is only half of what OPE microscopy can produce. In reality the difference may not be quite as large as that, because eq. 14 assumes that the fluorescence wavelength is the same as the excitation wavelength. In reality it is longer, which will reduce the confocal resolution somewhat. Also, it turns out that the excitation spectra of fluorophores tend to be much wider for TPE, and therefore an excitation wavelength of less than double the one-photon value is often used in practice. This also narrows the resolution gap between TPE and confocal microscopy. The fact remains, however, that we can expect TPE to produce images with lower resolution than confocal microscopy. The same is true for the depth resolution; the optical section thickness, measured as FWHM, is enlarged by the same factor as the *psf*. In many applications, however, the lower resolution is more than compensated for by the fact that much thicker samples can be studied. So, in summary, because of better penetration and reduced photobleaching TPE has become a useful tool in biomedical research.

3.3 Stochastic methods

Stochastic methods can produce images with much higher resolution than ordinary wide-field or confocal microscopes. The principle is that only a few widely separated fluorophore molecules emit light simultaneously. Each fluorophore molecule will produce a *psf* (which is diffraction limited) in the image plane, Fig 32. But the position of each molecule can be determined with an accuracy that is only limited by noise (i.e. there is no diffraction limit). If the molecules are widely

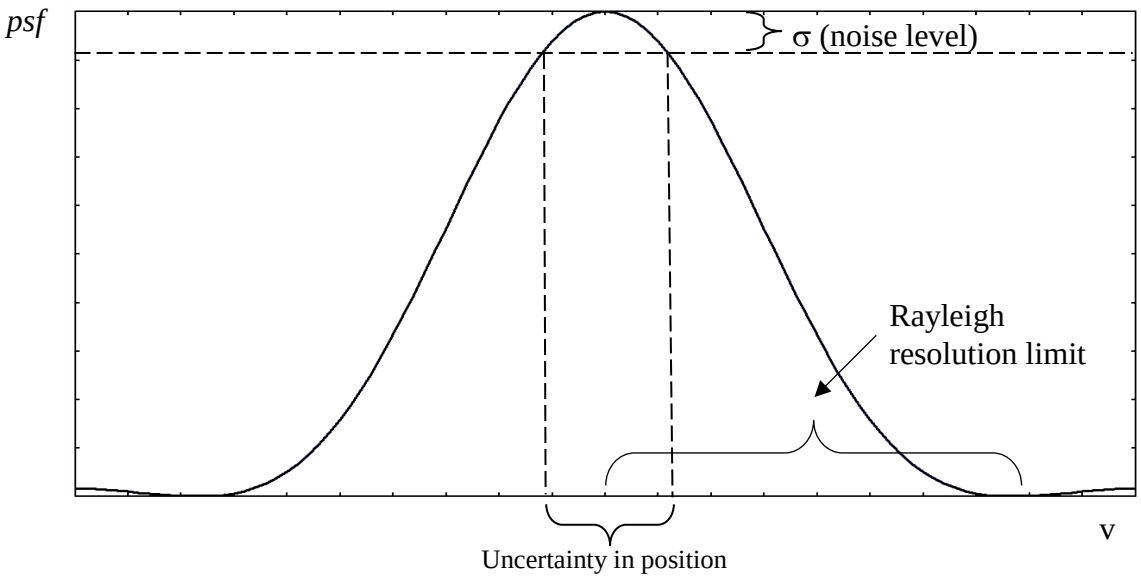


Fig. 32. The position of a fluorophore molecule can be determined with considerably higher accuracy than the Rayleigh resolution limit (Fig. 7 and eq. 4), provided that the noise level is sufficiently low.

separated, the *psf*s will overlap to a negligible amount, making it possible to determine the positions of the center-points for each *psf* with high accuracy (in practice something like 20 nm). Theory predicts that the positional accuracy, Δx, can be expressed approximately according to eq. 25.

$$\Delta x \approx \frac{0.2\lambda}{N.A.\sqrt{N}} \qquad (25)$$

where *N* is the number of detected photons. Eq. 25 describes a best case scenario, where we assume zero background and zero detector noise. It has also been assumed that the pixellation of the recorded image is extremely fine. If $\approx 10^4$ photons can be detected from a single fluorophore molecule before it is photobleached (which is very optimistic), we would be able to get a Δx on the order of 1 nm. In reality, as mentioned, we can expect something like $\Delta x \approx 20$ nm.

This method only works if the fluorophore molecules that emit light are sparsely distributed, so that we get little overlap between the individual *psf*s. In most microscopic specimens this requirement is not fulfilled. Therefore it is necessary to "switch on" only a subset of molecules when each image is recorded. This can be done, for example, by using so-called photoactivated fluorophores. A small random subset of the fluorophore molecules are then brought to the active state by a brief pulse of weak ultraviolet radiation. The activated molecules can then be excited like ordinary fluorphores and will emit fluorescent light. Their positions can be determined as described above, and finally the molcules are destroyed by photobleaching. The process of activation, recording and bleaching is then repeated many times for different subsets of molecules until all (or nearly all) molecules have been recorded. Each recorded subset of molecules will result in an image, where the positions of the molecules are indicated by, for example, white dots. Finally all these images are added together, so that all molecule positions are displayed simultaneously. Images produced in this way will look different from other microscopic images. They are not grayscale images, but only show lots of dots representing the positions of individual molecules.

3.4 Stimulated emission depletion

Stimulated emission depletion (STED) can be seen as a further development of confocal microscopy. Like in confocal microscopy, STED uses point illumination of the specimen and a small aperture in front of the detector. To record an image, the illumination point is scanned across the specimen, and the fluorescent light from the specimen is recorded. Due to diffraction, the light distribution in the illuminated spot will look like Fig. 6 and the diameter of the spot will be $\frac{1.22\lambda}{N.A.}$. The idea behind STED (ref. [7]) is to inhibit fluorescence emission from the outer parts of the illuminated spot, thereby "sharpening up" the psf in order to improve resolution. This is accomplished in the following manner.

1) A short pulse of light with a suitable wavelength for fluorescence excitation illuminates the specimen. The size of the spot is fairly large due to diffraction (equation above).

2) Shortly thereafter (before the excited fluorophore molecules have returned to the ground state) another short pulse illuminates the same specimen part. The second pulse produces a spot that has a dougnut shaped light distribution and it has a longer wavelength (approximately the same as the fluorescent light). This second pulse will cause stimulated emission of light from excited fluorophore molecules. Because of the doughnut distribution, most of the stimulated emission will take place at the periphery of the original pulse, and none in the exact center. As a result, the spontaneous fluorescence emission, which takes place a little later, will be less intense in the peripheral parts, because the fluorophore molecules have been more strongly de-excited in these parts. The net effect will be a narrower psf of the microscope.

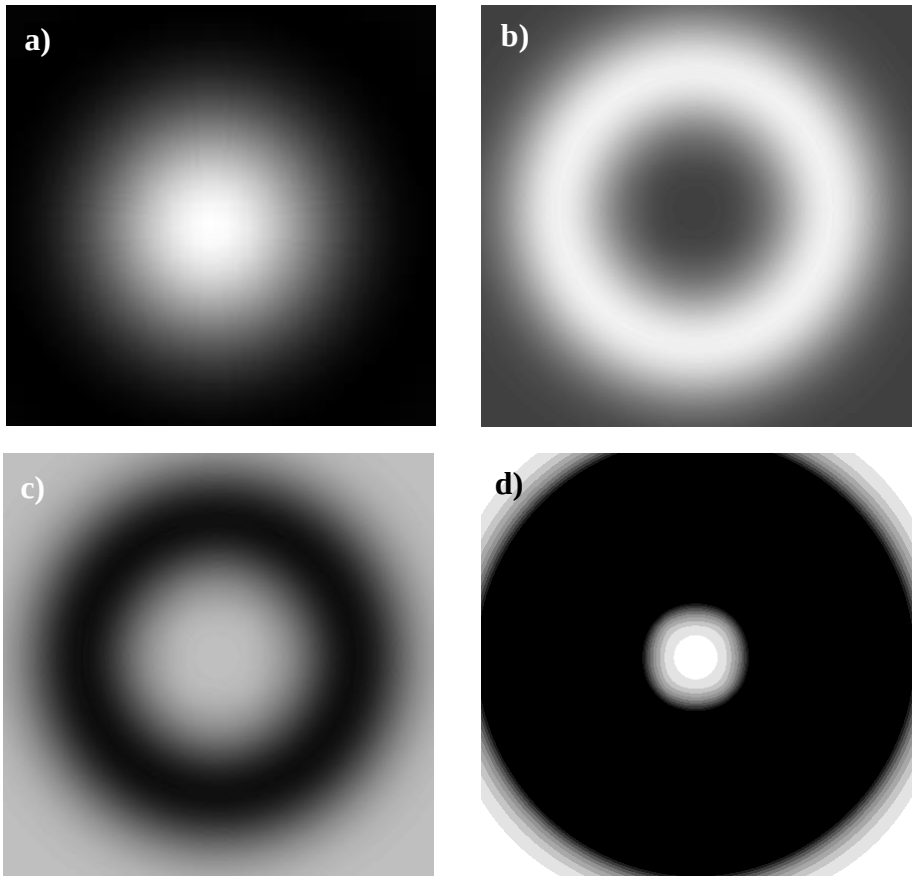


Fig. 33. Schematic illustration of how the STED technique can produce a narrower psf and thereby higher resolution. a) Light distribution in the excitation pulse (this is the psf of an ordinary microscope). b) Shortly after the excitation pulse a second pulse illuminates the specimen. The second pulse has a doughnut distribution and longer wavelength in order to produce stimulated emission. c) Due to stimulated emission fluorophore molecules are de-excited in the ring-shaped black area. d) By increasing the intensity of the doughnut pulse, fluorophore molecules will be almost completely de-excited except in a small central region. Only light from this central region will be detected when the spontaneous emission is recorded a little later. The net result will be a much narrower psf.

The second (doughnut) pulse will, of course, also be blurred due to diffraction. To get substantially improved resolution it is therefore necessary to saturate the inhibition of fluorescence emission caused by stimulated emission. Therefore a very intense doughnut beam is used so that the depletion region moves closer to the center of the spot. This mechanism is illustrated in Fig. 33.

Using the STED technique it has been demonstrated that resolution can be increased by an order of magnitude compared with ordinary microscopy. The practical resolution limit is set by signal levels and possibly specimen damage.

If the doughnut-shaped de-excitation beam is combined with a beam that is structured in the depth direction, it is possible to obtain optical sections with a thickness of about 0.1 μm (i.e. clearly superior to traditional confocal microscopy).

3.5 Structured illumination

So far we have looked at two extreme cases concerning illumination. Either we have uniformly illuminated a large area of the specimen (Figs. 4 & 5) or we have focused the light to an essentially diffraction-limited spot (confocal, two-photon and STED microscopy). There are, of course, many other possibilities. One case will be briefly mentioned here just to give an idea of how structured illumination can be used for improving resolution.

A structured illumination method to increase the resolution is to illuminate the specimen volume with a fringe pattern. This can be accomplished by using, for example, two counter-propagating coherent beams. In this way high-resolution images in both two and three dimensions can be produced. To illustrate the basic principle we will consider a two-dimensional case (further details can be obtained from ref. [8]). In fluorescence microscopy (using one-photon excitation), the distribution of the light emitted from the specimen plane, $I_f(x,y)$, is described by

$$I_f(x,y) = I_e(x,y)\rho(x,y) \qquad (26)$$

where $I_e(x,y)$ is the excitation light intensity distribution (structured light) and $\rho(x,y)$ is the fluorophore distribution. $\rho(x,y)$ is the information we want to record. If the illumination is completely uniform, i.e. $I_e(x,y)$ = constant, we get $I_f(x,y) = const. \cdot \rho(x,y)$, Fig. 34a. Thus the fluorescence light intensity distribution in the specimen is a true representation of the fluorophore distribution. This is the normal situation in a microscope with Köhler epi-illumination (Fig. 5), and the MTF is given by Fig. 8b. Thus spatial frequency components of $\rho(x,y)$ up to the limit $\frac{2N.A.}{\lambda}$ can be imaged. Let us now consider a one-dimensional example where $\rho(x)$ varies periodically with a spatial frequency of $\frac{2.5N.A.}{\lambda}$. This is beyond the resolution limit, and we would just see a uniform gray image in the microscope, Fig. 34a. Let us now insert a mask according to Fig. 34b in the plane of the luminous

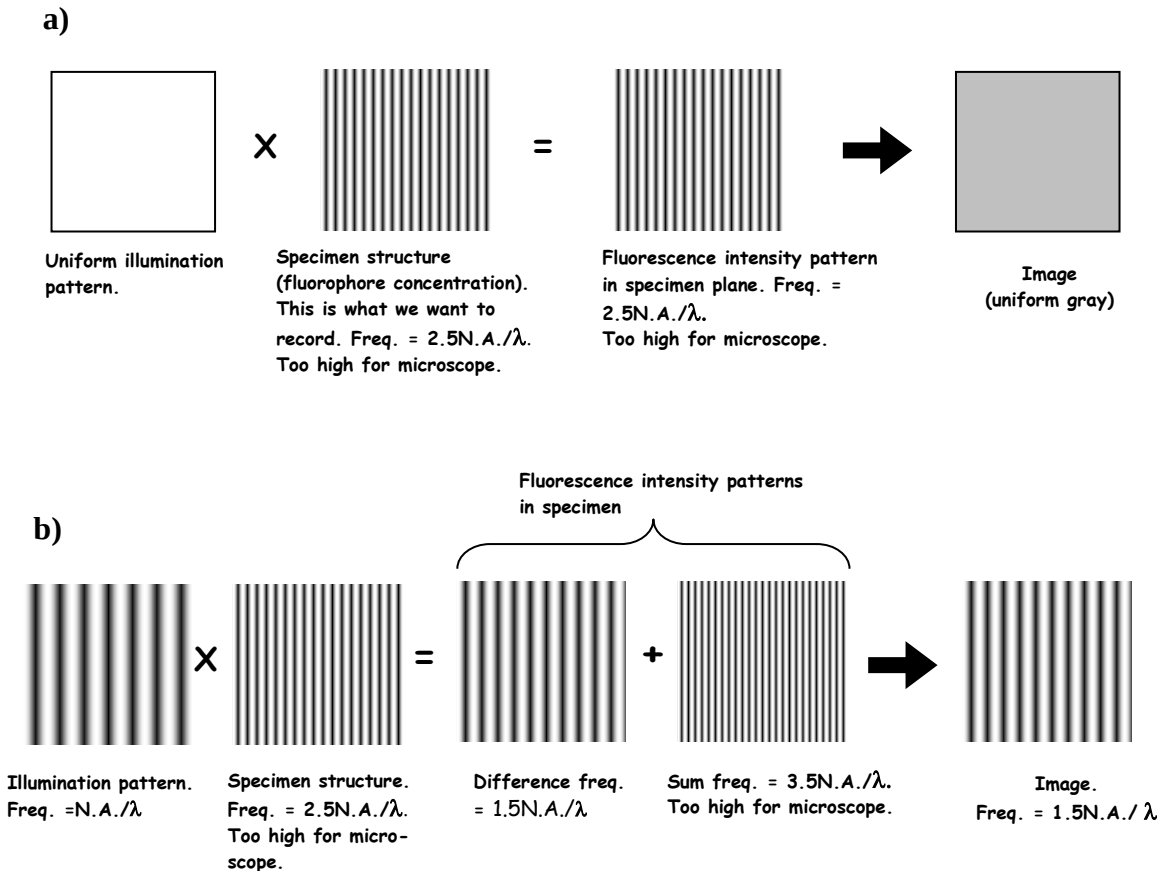


Fig. 34. Example of resolution improvement through structured illumination. a) Ordinary imaging. The specimen plane is uniformly illuminated. The spatial frequency of the specimen is too high to be resolved by the microscope. b) By illuminating the specimen with a periodic pattern, new frequencies are created. The difference frequency is low enough to be imaged by the microscope (albeit with incorrect period). From the image and the known illumination pattern the true specimen structure can be calculated.

field diaphragm. The mask will then be sharply imaged in the specimen plane. The mask consists of a sinusoidally varying transmission in the x direction. The density of the grid is such that its image in the specimen plane will be a pattern with spatial frequency $\frac{N.A.}{\lambda}$. Thus $I_e(x,y)$ is a periodic function with spatial frequency $\frac{N.A.}{\lambda}$, and $\rho(x)$ is a periodic function with spatial frequency $\frac{2.5N.A.}{\lambda}$. Multiplying these two functions to get $I_f(x,y)$ yields two frequency components: the sum frequency, $\frac{3.5N.A.}{\lambda}$, and the difference frequency, $\frac{1.5N.A.}{\lambda}$ (this is, by the way, analogous to the heterodyning performed in a radio receiver). The lower of these spatial frequencies, $\frac{1.5N.A.}{\lambda}$, is below the resolution limit of the microscope, and will therefore be visible. What this means is that a high spatial frequency has been

transposed to a lower one that can be imaged by the microscope. We can thus see things beyond the classical resolution limit of the microscope. In practice this "heterodyning" technique requires a bit of computer processing to restore all the spatial frequencies to their proper values in the final image, but it really works beautifully. One can show that this technique is useful up to a resolution limit of $\frac{4N.A.}{\lambda}$, an improvement by a factor of two. The limiting frequency is then the same as for confocal microscopy, Fig. 22c, but in practice it performs better because the MTF-value for confocal microscopy is practically nil for frequencies above $\frac{3N.A.}{\lambda}$

It is possible to extend the structured illumination technique further by using non-linear effects. This can be done by using a very high excitation light intensity so that the fluorophore saturates in the brightest parts of the illumination pattern. This saturation gives rise to harmonics in the illumination pattern, and as a result illumination patterns with spatial frequencies exceeding $\frac{2N.A.}{\lambda}$ can be produced. This means that one can record spatial frequencies in the specimen that are higher than $\frac{4N.A.}{\lambda}$. Using this non-linear type of structured illumination, resolution improvements of nearly an order of magnitude compared with traditional wide-field microscopy have been demonstrated.

It is also possible to arrange the structured illumination in such a way that optical sectioning is obtained. In this way sections with less than 0.1 μm FWHM have been demonstrated.

3.6 Total internal reflection fluorescence microscopy

It is often interesting to study the contact layer between a specimen and a glass substrate. An example of this is the study of fluorophores located within, or close to, the membrane of a cell growing on a glass substrate. To avoid recording fluorescence from the interior of the cell, it is necessary to record only an extremely thin section of the specimen close to the glass. Previously mentioned techniques, e.g. confocal imaging, can produce optical sections, but the section thickness is often too large compared with a cell membrane. An optically simple method that can be used in a case like this is total internal reflection (TIR) microscopy. If the specimen is illuminated from below through the glass substrate, and the angle of incidence is large, total internal reflection will occur, Fig. 35. An evanescent wave will then extend above the glass/cell interface. This means that the light intensity will decay exponentially in this region, with a practical penetration distance on the order of 100 nm. If the wavelength of the incident light is chosen so that it can excite the fluorophore, we will thus see fluorescence from an approximately 100 nm thick specimen layer only. Section images of this type can be easily recorded with a CCD or CMOS camera. Compared with confocal microscopy, TIR has the advantages that no mechanical scanning is needed, recording is faster, and the sections are thinner. The drawback, of course, is that only the contact layer between

specimen and glass substrate can be recorded. With confocal microscopy, sections anywhere within a cell can be recorded.

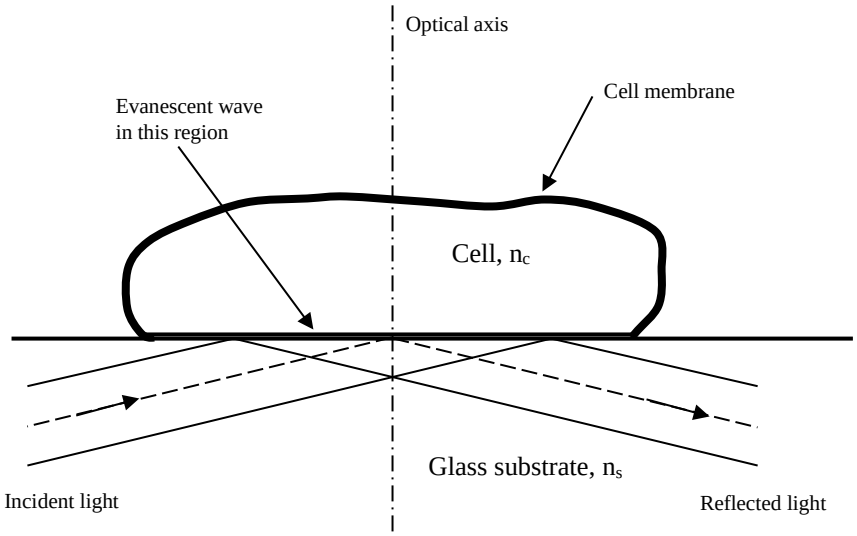


Fig. 35. Principle of total internal reflection excitation. A fluorescently labeled cell on a glass substrate is studied. If the refractive index of the glass, n_s, is higher than that of the cell, n_c, total internal reflection can be obtained. Then only the cell layer in close contact with the glass will be illuminated by the evanescent wave. As a result, only a thin ($\approx$ 100 nm) contact layer will be visible in the microscope.

3.7 Deconvolution

With cheap and powerful computers available, it is today possible to a certain extent to compensate for the influence of the point spread function (*psf*) on image data. Mathematically, this influence can be described by

$$I_B = I_O \otimes psf \qquad (27)$$

where I_B and I_O represent the image and object functions respectively, and $\otimes$ denotes convolution (ref. [9]). Taking the Fourier transform of eq. 27, and rearranging, we obtain

$$\hat{I}_O = \frac{\hat{I}_B}{\hat{psf}} \qquad (28)$$

where the symbol ^ denotes the Fourier transform. This means that if we have measured the *psf* of the microscope, and recorded an image function I_B, we can take their Fourier transforms and insert into eq. 28. This will give us the Fourier transform of the true object function I_O. By taking the inverse transform, we then get I_O. This procedure is called deconvolution, and it can be carried out on both

two- and three-dimensional image data. There is a catch, of course, otherwise we could get infinitely high resolution by using this procedure. The catch is that both $\hat{I}_B$ and $\hat{psf}$ become zero at high spatial frequencies, so that in eq. 28 we get zero divided by zero. However, up to the frequency where $\hat{psf}$ becomes zero (i.e. the limiting frequency of the MTF), it should in principle be possible to restore the amplitudes perfectly. What this means in practice, is that the contrast of small specimen details can be improved so that they can be seen more clearly. A practical problem with this restoration is that it also amplifies high-frequency noise. Therefore deconvolution methods often employ some technique for noise suppression (but there is always a trade-off between resolution and noise).

In recent years deconvolution has often been applied to 3-D image data. Typically an ordinary wide-field (i.e. non-confocal) microscope is used to record a number of images from different depths in the specimen. As can be seen in Fig. 25a the MTF of a non-confocal microscope extends to some extent also in the depth direction. This means that 3-D deconvolution can be applied on the stack of images, resulting in a volume representation of the specimen (ref. [10]). This procedure has to be carried with some restrictions, and it works best on specimens that are not too dense (small point-like objects spread out in a large volume is fine, but dense tissue is not).

3.8 Near-field microscopy

There is a possibility to reach a very high resolution in light microscopy without resorting to non-linear methods or short wavelengths. This requires that we circumvent the limit caused by diffraction that occurs when light passes through an aperture. A simple idea to avoid this limitation is to get rid of the microscope objective altogether, and instead illuminate the specimen through an extremely small hole in an opaque screen placed very close to the specimen, Fig. 36. Typical numbers in a real case would be a 20-100 nm hole placed about half the hole diameter (10-50 nm) from the specimen. The specimen will then be placed in what is called the near-field region of the aperture. In this region the light has not yet spread out due to diffraction, so it will form a spot of the same size as the aperture. We may regard this as a shadow image of the aperture. This is true only out to a distance of about one hole radius from the aperture. For longer distances the light will be strongly spread out due to diffraction, and we are in what is called the far-field region. But as long as the specimen is placed in the near-field region we can, in principle, make the illuminating spot arbitrarily small♣.

By scanning the small aperture over the specimen and detecting the fluorescent or reflected light from the illuminated spot, we can record an image with a resolution of approximately the spot size, i.e. $<< \lambda$. This technique is called NSOM (Near-field Scanning Optical Microscopy). Although light detection can be made via the small aperture, it is often done by an external detector which is more efficient. This is possible because there is no need to form an image of the light coming from the specimen. As a result, the resolution is determined only by the size of the

♣ In ordinary, i.e. far-field, microscopy, the minimum spot size to which we can focus light with a microscope objective is a diameter of approximately λ when using an objective with a high numerical aperture.

illuminating spot formed in the specimen. In practice, the aperture is usually made by heating and pulling an optical fiber so that it forms a narrow tip, Fig. 37.

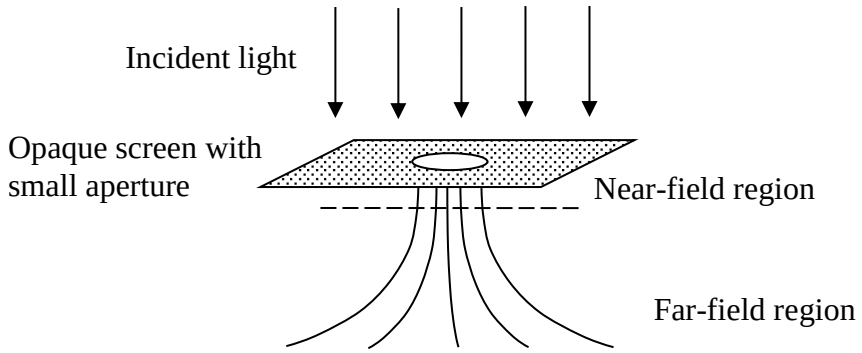


Fig. 36. In near-field microscopy a sub-wavelength aperture in an opaque screen is illuminated and placed close to the specimen surface. In the near-field region the size of the light spot is determined by the size of the hole regardless of the wavelength used. In the far-field region, on the other hand, the light is severely spread out due to diffraction.

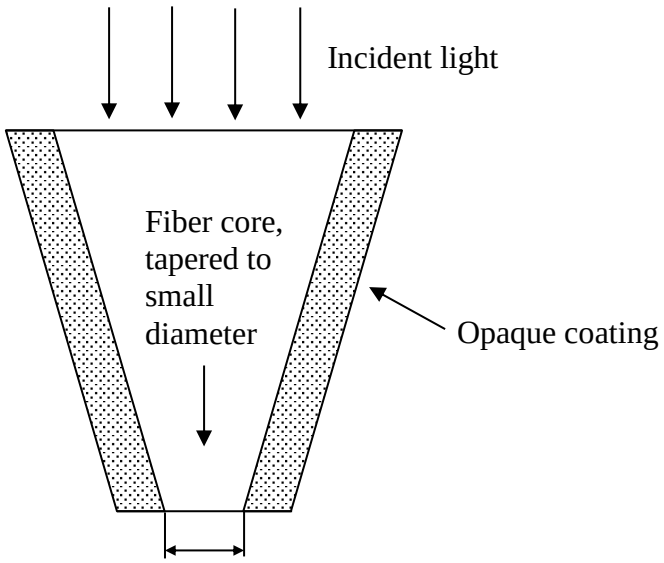


Fig. 37. Practical realization of an aperture. An optical fiber is heated and pulled so that it forms a tapered tip. Because of the small core dimensions at the tip light will leak out, and this has to be blocked with an opaque layer. Also, because of the small core diameter at the tip, most of the incident light will be reflected back through the fiber.

From what has been said so far it may seem that near-field microscopy can produce arbitrarily high resolution using visible light. So why isn't everybody using it all the time? The answer is that there are several practical difficulties. First, as the hole

becomes smaller less light will be transmitted. In this region of sub-wavelength aperture size, the light intensity will decrease much more rapidly than what would be expected from the decrease in aperture area (in fact it decreases as the inverse sixth power of the aperture radius). Increasing the incident light intensity to compensate for this light loss will ultimately damage the aperture due to overheating. Therefore, as the aperture becomes smaller, the detected signal from the specimen will become weaker resulting in a poor signal-to-noise ratio. Second, the aperture must be located no further from the specimen than about the hole radius, and it must be maintained at this distance during scanning. This calls for a rather sophisticated control system and precise scanning. Third, NSOM places rather stringent limitations on the specimen. It must be rather flat if the technique is to be successful.

Finally, it can be mentioned that instead of using a small aperture, it is possible to use a very small blocking object, for example a narrow metal tip, positioned near the specimen surface. If an intense beam of light, for example a laser beam focused by an ordinary lens, illuminates the specimen, there will be a strong local enhancement of the light intensity in a very small region close to the tip. By scanning the tip in the same manner as the small aperture described previously, a high-resolution image can be obtained.

Appendix I (Optical aberrations)

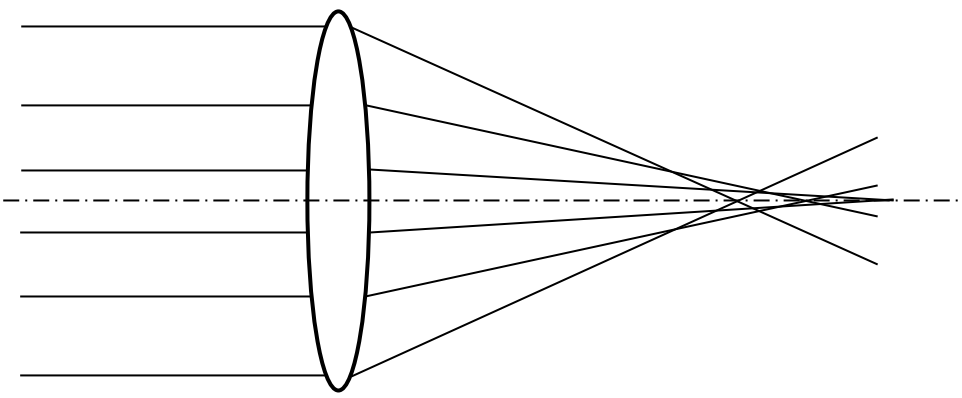
Two things limit the resolution in an optical system – diffraction and aberrations. In this appendix a brief description of some fundamental aberrations will be given. Aberrations mean refraction errors in lenses. Most aberrations result in a blurring of the image (the exception is distortion, which results in a geometrically distorted image). This blurring is caused by the fact that, in real optical systems, light rays emanating from an object point are not refracted perfectly so that they meet at a single point. The most fundamental aberrations are:

- Spherical aberration
- (Longitudinal) chromatic aberration
- Astigmatism
- Curvature of field
- Coma
- Distortion

The first two aberrations are present all over the image field, including the optical axis. All other aberrations are so-called off-axis aberrations, meaning that they are absent on the optical axis and then (usually) become more severe the further we move towards the image periphery. In the description below we will, for simplicity, assume that only one aberration is present at a time. In reality, of course, the image will be influenced by all aberrations simultaneously. Aberrations can be reduced by combining several lens elements, both positive and negative, made from different types of glass. The more lens elements that are used, the better the aberrations can be controlled (but the objective will also be more expensive). Not surprisingly, aberrations are more difficult to control in high numerical aperture objectives (these may consist of something like ten lens elements).

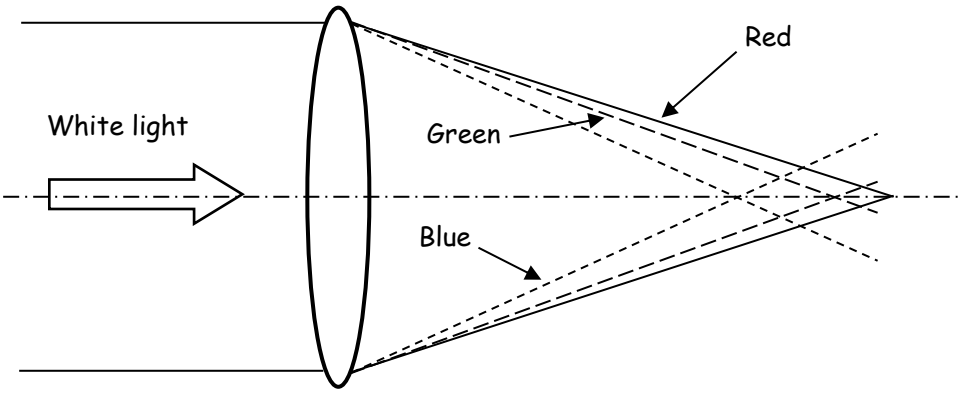
Spherical aberration:

The name "spherical aberration" can be a bit misleading. One might believe that it is a consequence of failure to obtain perfect spherical surfaces on the lenses. But this is not the case. The aberration is caused by the fact that the surfaces ***are*** spherical (aspherical surfaces can be used to avoid this aberration). Spherical aberration for a single lens element is illustrated below. As can be seen, different focus points are obtained for central and peripheral rays.



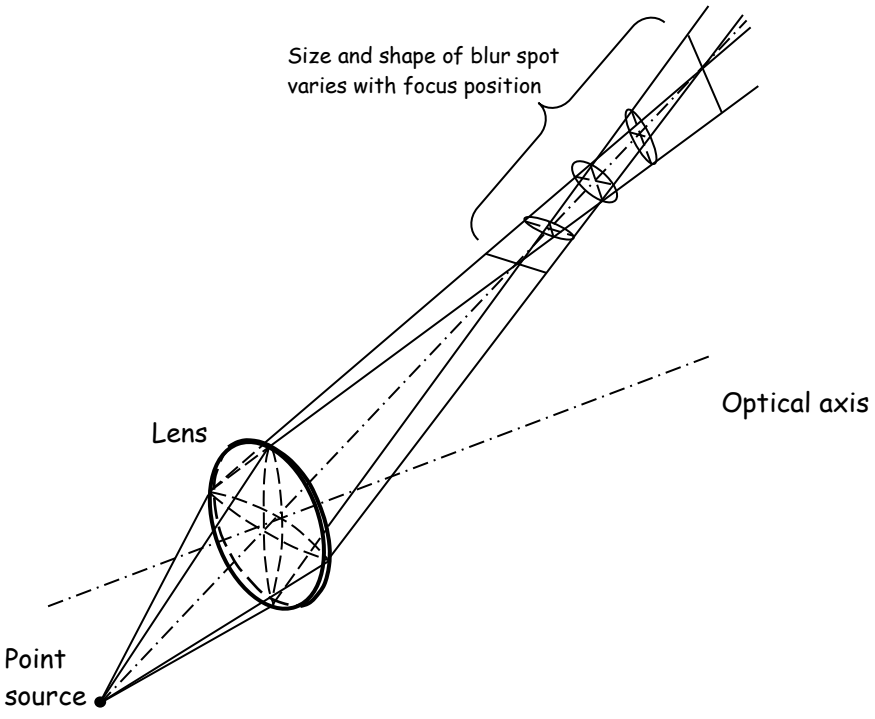
Chromatic aberration (longitudinal):

This aberration is caused by dispersion in the glass (i.e. the refractive index is different for different wavelengths). As a result, the focal length of a single lens is wavelength-dependent. (There is a second type of chromatic aberration called "lateral chromatic aberration," which will not be described here. Those interested can get information about this in, e.g., ref [1]).



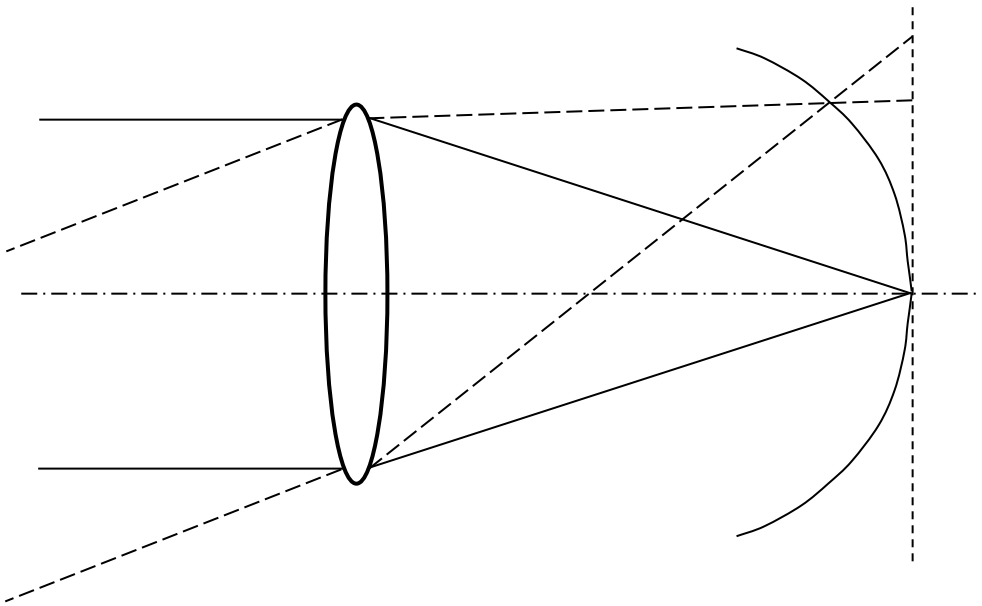
Astigmatism:

This is an off-axis aberration that gives, in general, a non-symmetric blurred spot when imaging a point object. Depending on the focus position, the blurred spot may take the shape of a line, an ellipse or a circle.



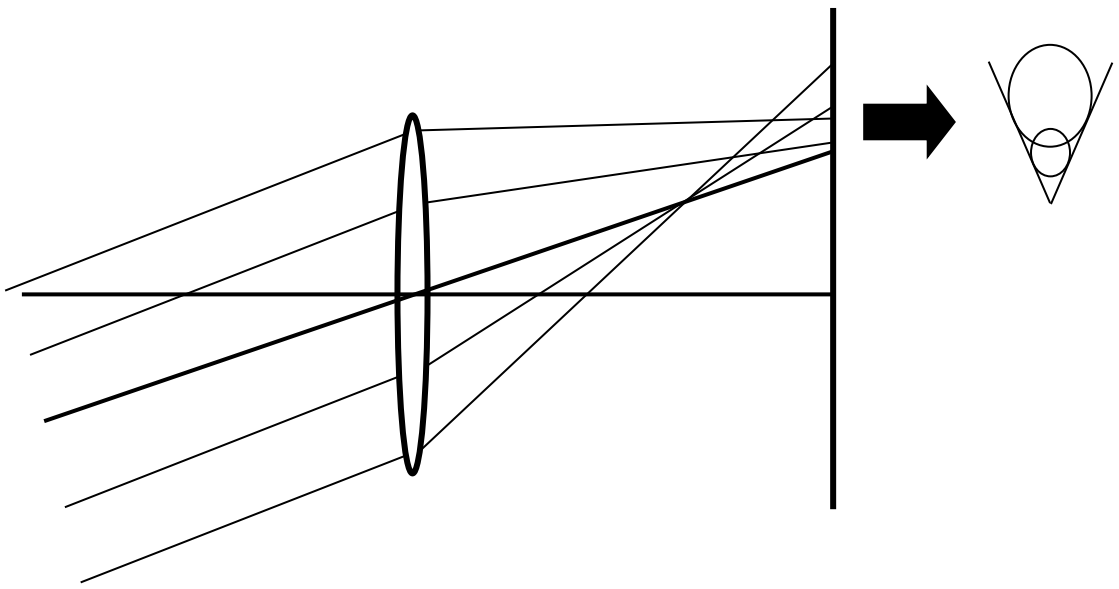
Curvature of field:

This aberration causes the best image sharpness to be obtained along a curved surface, rather than in a plane perpendicular to the optical axis.



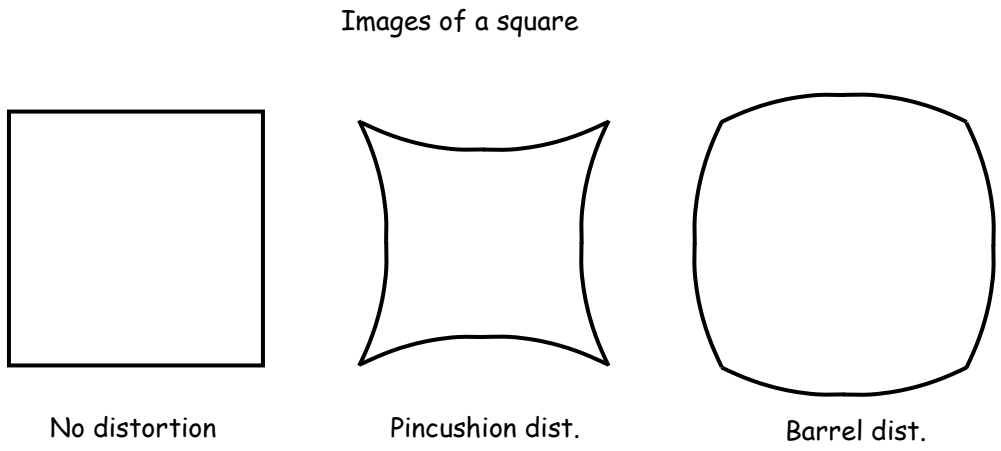
Coma:

This aberration causes the image of a point object to look like a comet (usually with the tail towards the periphery). The cause of the aberration is that, for oblique rays, different zones of a lens will give both different focus positions and different imaging scales.



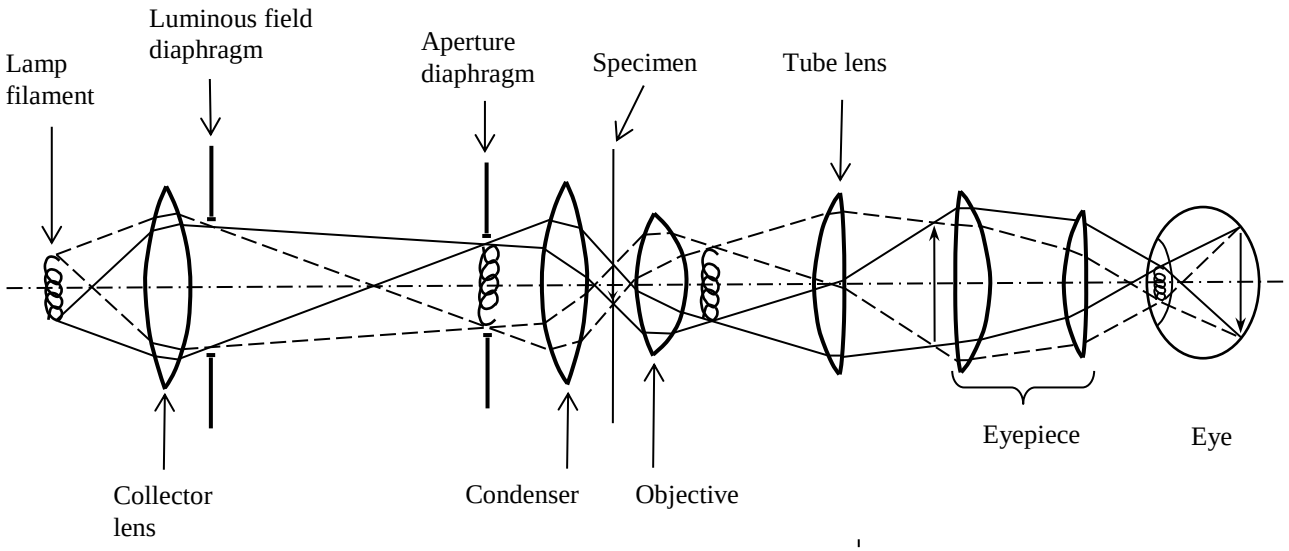
Distortion:

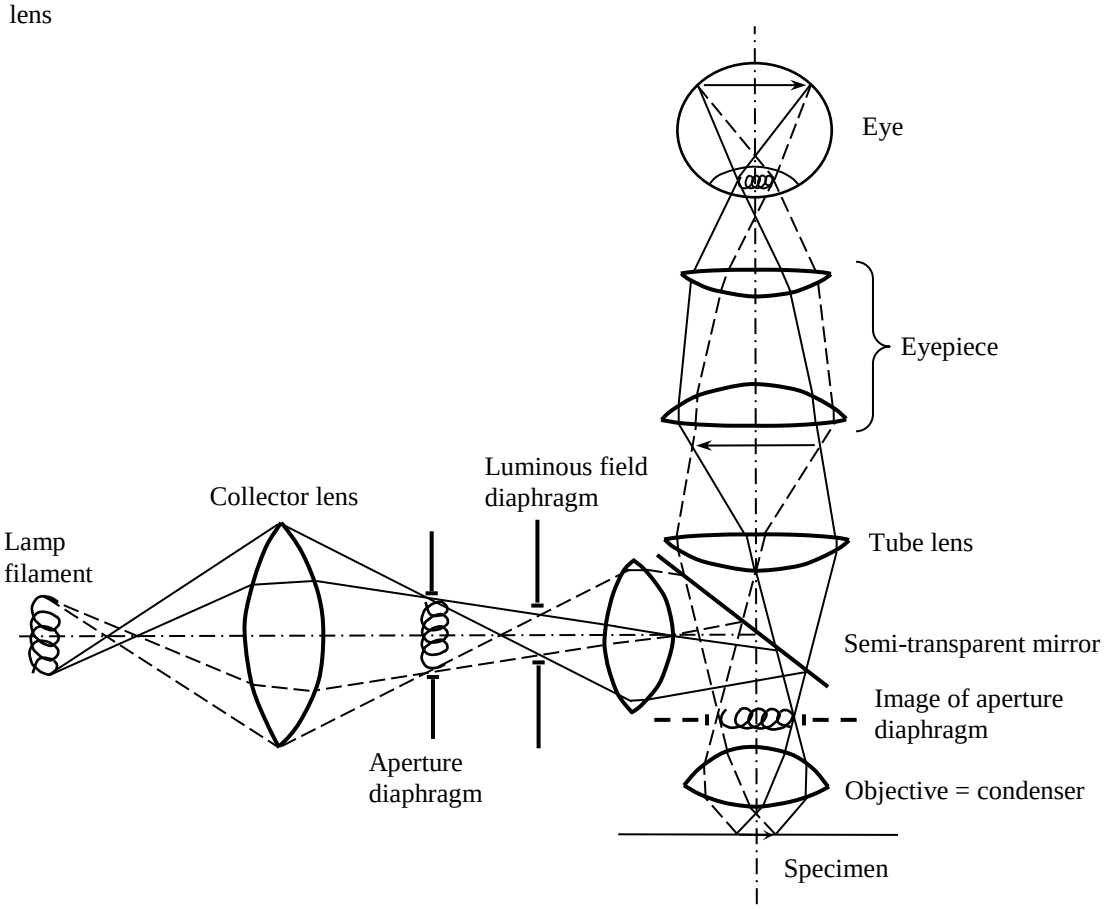
Distortion is an aberration that differs markedly from the rest, because it does not cause any blurring of the image but a geometric distortion. This is the effect of an imaging scale that varies with the distance to the optical axis. If the imaging scale increases as we move from the axis towards the periphery, we get so-called pincushion distortion. If instead the scale decreases towards the periphery, we get barrel distortion.



Appendix II (Illumination systems for infinite tube length microscopes)

For a microscope with infinite tube length figures 4 and 5 are changed into the figures shown below. To the user this difference means little, since all controls and functions of the Köhler illumination system will remain the same as for a finite tube length microscope. However, the optical components and their positions differ slightly for an epi-illumination system designed for infinite tube length. This may be of importance in cases where microscopes are modified or rebuilt by the user. For completeness the figures below are therefore included.





Appendix III (*psf* and MTF)

The diffraction pattern from a lens opening can be calculated using the Fresnel-Kirchoff formula (ref. [1]), which is essentially an extension of Huygens' principle, well known from school physics. The figure below illustrates the geometry involved. ξ and η are coordinates in the pupil plane, i.e. in the plane of the lens opening.

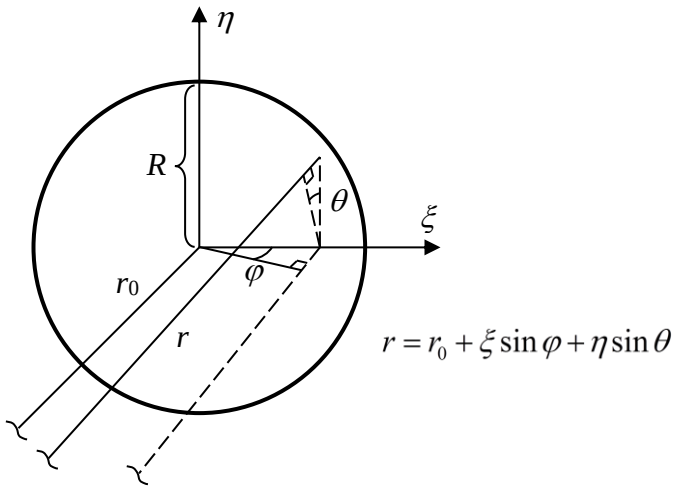


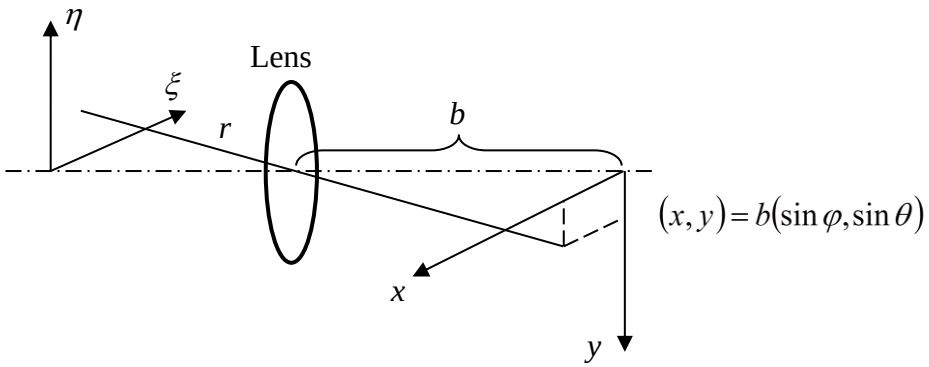
Illustration of the geometry involved in calculating the diffraction pattern from a lens opening of radius R.

Optical pathlengths from the lens opening to the image plane of the lens are given relative to the pathlength of a ray, r_0, emanating from the center of the lens (ξ = 0 and η = 0). The pathlength of another ray, parallel to r_0 and emanating from the pupil coordinates ξ and η, is, relative to r_0, ξsinφ+ ηsinθ. According to the Fresnel-Kirchoff formula the amplitude of the optical signal can then be expressed as (neglecting a constant factor)

$$h(\varphi,\theta) = \iint e^{\frac{2\pi i}{\lambda}(\xi\sin\varphi+\eta\sin\theta)} d\xi d\eta \qquad \text{(a1)}$$

where the area of integration covers the entire lens opening with radius R. All rays parallel to r_0, such as *r*, will be focused by the lens to the same point in the image plane, at a distance *b* from the lens. As illustrated in the upper figure on next page, the coordinates of this focus point can be expressed as x = bsinφ and y = bsinθ. Substituting $\sin\varphi = \frac{x}{b}$ and $\sin\theta = \frac{y}{b}$ in eq. a1 we get

$$h(x,y) = \iint e^{\frac{2\pi i}{\lambda}\left(\xi\frac{x}{b}+\eta\frac{y}{b}\right)} d\xi d\eta \qquad \text{(a2)}$$



Parallel rays forming angles φ and θ with the optical axis as illustrated in the figure on previous page will be focused to a point in the image plane of the lens.

Finally, by introducing the variables $\nu_x = \frac{\xi}{\lambda b}$ and $\nu_y = \frac{\eta}{\lambda b}$ we get

$$h(x,y) = \iint e^{2\pi i(\nu_x x + \nu_y y)} d\nu_x d\nu_y \qquad \text{(a3)}$$

where the integration is carried out over a circular area with a radius of $\frac{R}{\lambda b}$ in the ν_x, ν_y plane. What eq. a3 states is that h(x,y), the amplitude *psf*, is equal to the inverse Fourier transform of the lens aperture scaled by the factor $\frac{1}{\lambda b}$. This means that the FT of h(x,y), H (ν_x, ν_y), is equal to the lens aperture function scaled by $\frac{1}{\lambda b}$, i.e. a function that is constant within a radius of $\frac{R}{\lambda b}$ and zero outside, see figure below.

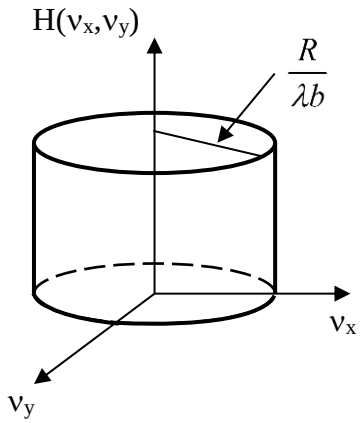


Illustration of the Fourier transform of h(x,y).

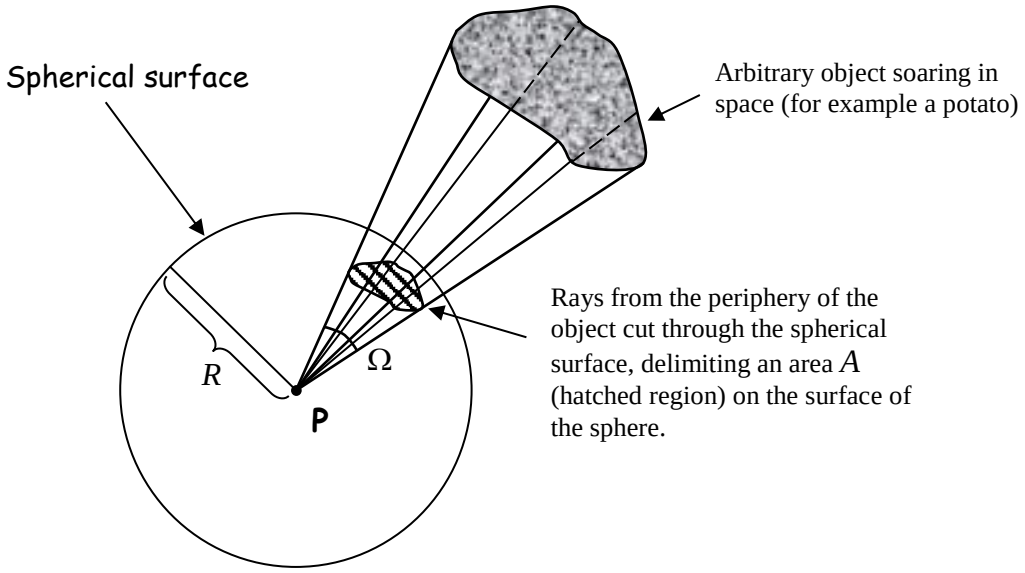
From Fig. 1 (sect. 1) we see that $\frac{R}{b} \approx \sin\beta$, so that the radius of H (ν_x, ν_y) can be written $\frac{\sin\beta}{\lambda}$. By using the Abbe sine condition (ref. [1]) we can write $\sin\beta = \frac{N.A.}{M}$, where N.A. is the numerical aperture of the objective, and M the magnification. Using this relationship, we can write the radius of H (ν_x, ν_y) as $\frac{N.A.}{\lambda M}$. It should be noted that so far the spatial frequencies refer to the *image* plane. The spatial frequencies in the *specimen* plane will be higher by a factor of *M*. Therefore, if we consider the frequencies in the specimen plane, the radius of H (ν_x, ν_y) will be $\frac{N.A.}{\lambda}$. This is the result we have used in eq. 5.

Appendix IV (Photometry)

The purpose of this little appendix is to give a summary of some fundamental and important quantities in radiometry and photometry. Radiometric quantities are used to describe radiating energy, power, power density etc. The basic units are those used in physics, such as joule and watt. Some basic radiometric quantities are given in the table below.

Quantity	Unit
Radiant flux	W
Radiance	W m^{-2} sr^{-1}
Irradiance	W m^{-2}

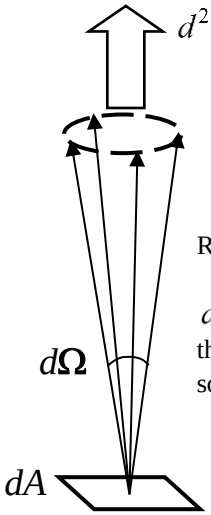
But first let's define what we mean by solid angle (measured in steradians, sr). This is defined in the figure below.



The solid angle, Ω, under which we view the object from point **P** is defined by the equation $\Omega = \frac{A}{R^2}$. The maximum solid angle possible is thus 4π, which means that the object completely surrounds us (when we stand on the surface of the earth, the sky subtends a solid angle of 2π). The unit for solid angle is called steradian, and is abbreviated sr.

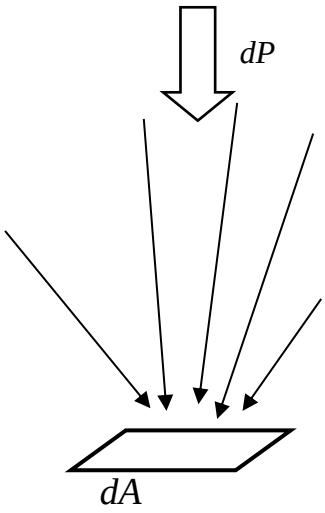
The quantity "Radiant flux" describes how much energy per unit time that is transferred via radiation. "Radiance" tells about emission, e.g. from a lamp filament. This quantity describes (as can be seen from the unit) how much power that is emitted per area and solid angle unit, see figure on next page♣. "Irradiance" tells about incoming radiation (e.g. onto a surface). It describes how much incident power we have per unit area.

♣ In reality there is a directional dependence that we omit in this course. The radiated power is usually highest in a direction parallel to the surface normal, and then becomes gradually smaller as the angle to the surface normal increases.



Radiance = $\frac{d^2P}{dA\,d\Omega}$, where

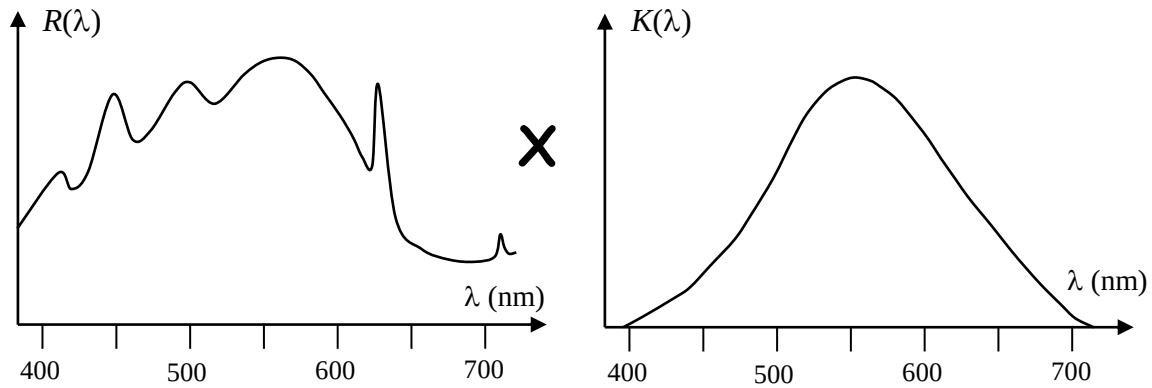
d^2P is the power emitted from the surface area dA within a solid angle of $d\Omega$.



Irradiance = $\frac{dP}{dA}$, where

dP = total irradiated power from all directions.

For each radiometric quantity there is a corresponding photometric quantity. The only difference is that photometric quantities are weighted with respect to the spectral sensitivity curve of the eye. Thus wavelengths around 550 nm will give the largest contribution, whereas shorter and longer wavelengths will contribute less (the further from 550 nm we go, the smaller the contribution will be). Wavelengths outside the visible range (approx. 400 – 700 nm) give no contribution at all. Radiant flux is thus replaced by luminous flux, radiance is replaced by luminance, and irradiance by illuminance. As an example, consider a spectral distribution for the radiant flux, $R(\lambda)$, according to the figure below. This function is multiplied by the spectral sensitivity function of the eye, $K(\lambda)$, and integrated over wavelength. The luminous flux, Φ, can then be expressed mathematically as $\Phi = \int_0^\infty R(\lambda)K(\lambda)d\lambda$.



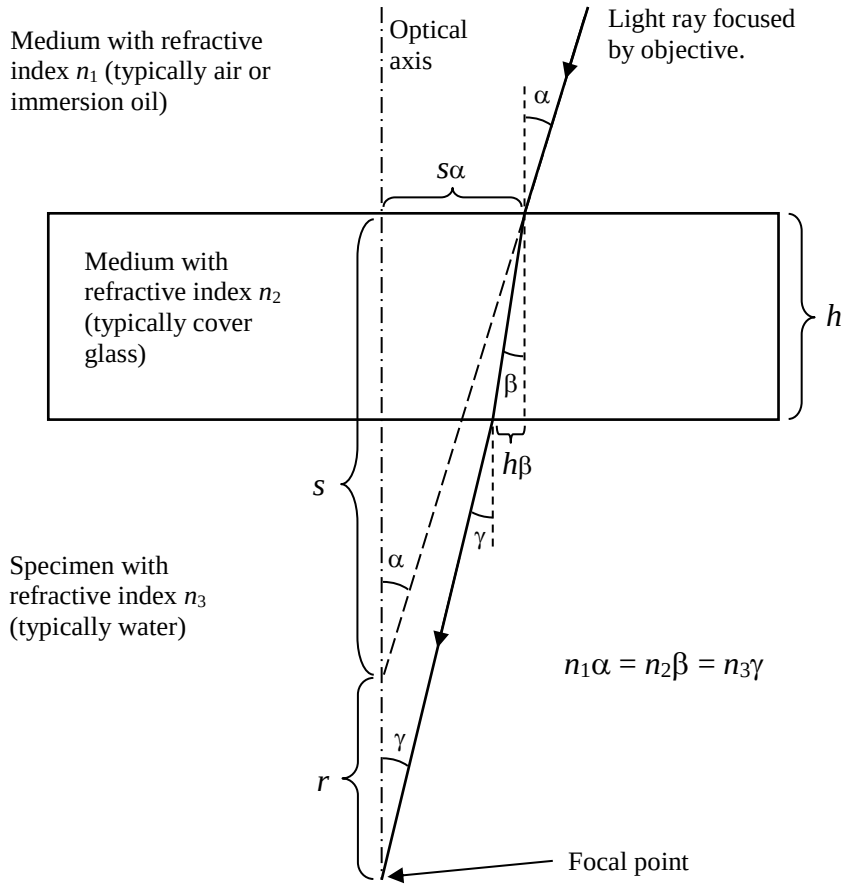
Spectral distribution of radiant flux **Eye sensitivity**

In the table below are listed the photometric quantities that correspond to the radiometric quantities given above.

Quantity	**Unit**
Luminous flux	lm (lumen)
Luminance	lm m^{-2} sr^{-1} (= nit)
Illuminance	lm m^{-2} (= lux)

Appendix V (Depth distortion)

When you gaze into a swimming pool, it doesn't look as deep as it really is. This well-known effect is caused by the difference in refractive indices for water and air. A similar effect often occurs when scanning a stack of images in a confocal microscope. If the refractive index of the specimen is different from that of the immersion medium (or air) in which the objective is working, a distortion of the depth scale occurs. Let's study this effect quantitatively for the paraxial approximation (i.e. small N.A.)♣. Below is a figure that shows how light is focused into a specimen. We can assume that the reflected or fluorescent light follows the same path but in the opposite direction.



Here are some useful equations we will use:

$$\frac{s\alpha - h\beta}{s - h + r} = \gamma \Rightarrow r = \frac{s\alpha - h\beta}{\gamma} + h - s \Rightarrow \frac{dr}{ds} = \frac{\alpha}{\gamma} - 1 = \frac{n_3}{n_1} - 1$$

♣ In practice, the results obtained are fairly good approximations also for high numerical apertures.

Now, let's assume that we move the specimen vertically, resulting in a small change, *ds*, in *s*. The focal point will then move a distance *ds* + *dr* further into the specimen volume. Using the equations on the previous page, we can write $ds + dr = ds + ds\left(\frac{n_3}{n_1} - 1\right) = ds\frac{n_3}{n_1}$. From this result we can deduce that the cover glass (with refractive index n_2) does not affect the depth scale. If $\frac{n_3}{n_1} > 1$, the movement of the focal point in the specimen will be > *ds*, resulting in a compression of the depth scale in a recorded stack of optical section images. This can be compensated for in the data processing. If $n_1 = n_3$ we get a correct depth scale. We can summarize these findings in the following equation

$$z' = \frac{n_1}{n_3} z$$

where *z*' is a vertical distance measured in the stack of recorded optical section images, and *z* is the true vertical distance in the specimen volume.
Example: Let's assume that we have recorded a stack of images where the microscope stage was refocused a distance of 1.0 μm between successive images. A dry microscope objective was used, and the specimen refractive index was 1.5. In the recorded stack of images the vertical extent of an object was measured to 100 pixels, which would correspond to 100 μm. Because of the scale distortion, however, the true distance would then be $100 \times \frac{n_3}{n_1} = 100 \times \frac{1.5}{1.0} = 150$ μm.

It is interesting to ask whether these results influence the recording of the *FWHM* depth response of a confocal microscope, for example when using the simple and popular mirror test (Fig. 25). Does it matter, for example, if we attach a cover glass onto the mirror with a drop of oil when testing a dry objective designed for use with cover glass♥? Will the result differ from what we would obtain if the oil layer were replaced with a thin layer of air? Without going into any details, we will just state that the results will not differ. The depth scale distortion we get during scanning is exactly compensated for by a distortion of the *psf* in the depth dimension.

Even if the depth distortion does not influence the *FWHM* measurement for an objective, it has implications concerning the depth resolution in a recorded volume. To illustrate this, we will again look at the example above where we assume a specimen volume whose refractive index is 1.5, and the use of a dry objective. The *FWHM* depth response has been measured to 2.0 μm. Because of refractive index mismatch, the recorded volume has to be expanded by a factor of 1.5 in the depth direction. This means, however, that the effective *FWHM* is also expanded by the same factor so that it becomes 3.0 μm. In this case it would have been better if the specimen volume had a lower refractive index. Better depth resolution could then have been obtained with the same objective.

♥ Attaching the cover glass with a drop of immersion oil is convenient when using an inverted microscope. Otherwise the cover glass will simply fall off.

Historical note

Confocal microscopy was described by Marvin Minsky in a patent from 1961. Development was then rather slow and sporadic over the next fifteen to twenty years, but in the last decades interest in this field has increased dramatically. The reason for this is that the rapid development in computer equipment means that the computing power necessary to utilize the data from a confocal microscope is now within reach of all scientific laboratories. Also, developments in laser technology have resulted in lasers that are better suited as light sources in confocal microscopes.

References

1. Born, M. & Wolf, E: Principles of Optics, Pergamon Press, New York, 1983.
2. Piller, H.: Microscope Photometry, Springer, Berlin, 1977.
3. Philip, J. & Carlsson, K.: 3D image deconvolution in light microscopy: theory and practice. KTH 2002. http://www.math.kth.se/~johanph/restore.pdf
4. Carlsson, K.: The influence of specimen refractive index, detector signal integration, and non-uniform scanning speed on the imaging properties in confocal microscopy. J. Microsc., **163**, 167-178 (1991).
5. Carlsson, K. & Mossberg, K.: Reduction of cross-talk between fluorescent labels in scanning laser microscopy. J. Microsc., **167**, 23-37 (1992).
6. Carlsson, K., Åslund, N., Mossberg, K. & Philip, J.: Using intensity-modulated scanning beams in combination with lock-in detection for recording multiple-labeled fluorescent specimens in confocal laser microscopy. Proc. SPIE **2184**, 21-29 (1994).
7. Hell, S.W. & Wichmann J.: Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy, Opt. Lett., **19**, 780-782 (1994).
8. Gustafsson, M.G.L.: Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy, J. Microsc., **198**, 82-87 (2000).
9. Carlsson, K. Compendium "Imaging Physics," KTH, 2009.
10. Castleman, K.R.: Digital Image Processing, Prentice-Hall, 1979.

Index