Journal of Microscopy, Vol. 189, Pt 1, January 1998, pp. 15–24.
Received 7 April 1997; accepted 15 August 1997

Contrast, resolution, pixelation, dynamic range and signal-to-noise ratio: fundamental limits to resolution in fluorescence light microscopy

E. H. K. STELZER
Light Microscopy Group, Cell Biophysics & Cell Biology Programmes, European Molecular Biology Laboratory (EMBL), D–69117 Heidelberg, Germany

Key words. Airy disc, confocal, conventional, distances, reflection, transfer function, wide-field.

Summary

In a perfect optical system numerical aperture and wavelength determine resolution. In a real optical system, however, the number of photons collected from a specimen determines the contrast and this limits the resolution. Contrast is affected by the number of picture elements per unit area, the number of photons and the aberrations present in every optical system. The concept of contrast vs. distance functions is used to compare the resolution achievable in confocal and wide-field fluorescence microscopes and the effect of a further reduction of the observable volume. *In conclusio*: (a) real optical systems will never be able to achieve the theoretical resolution, (b) wide-field fluorescence microscopy will often provide a better resolution than confocal fluorescence microscopy, (c) decreasing the observed volume does not necessarily increase the resolution and (d) using multiple fluorophores can improve the accuracy with which distances are measured. Some numbers for typical situations are provided.

Introduction

Advocates of confocal fluorescence microscopy claim that the three-dimensional resolution in their instruments is better than that of wide-field fluorescence microscopes. They therefore propose that laser scanners are used instead of video cameras and that the improved lateral resolution can be used to visualize otherwise unidentifiable features. These comments are backed up by theoretical comparisons of confocal and wide-field microscopy which prove that in transmission and reflection contrasts the resolution in terms of (a) the cut-off frequency is improved by a factor of 2 (Sheppard & Choudhury, 1977) and (b) the full width at half-maximum of a point object is improved by a factor of $1/\sqrt{2} \approx 1/1{\cdot}4 \approx 0{\cdot}7$ (Brakenhoff *et al.*, 1979). On the other hand, every user of a confocal fluorescence microscope knows that features in objects which provide a low signal are not easily distinguished and that the amount of signal from a fluorophore-labelled object can be improved by opening the pinhole in front of the detector. (In fact, it is an everyday experience that features are less well distinguished in the dark.) Hence, in practice, the full resolution of a confocal fluorescence microscope is often not used. The instrument instead operates in a mode that should not be much better than that achieved with a good video camera.

In this paper we are interested in the influence of noise and the number of picture elements per unit area on the lateral resolution of wide-field and confocal fluorescence microscopes. The main purpose is to understand how the image of two closely spaced small objects is affected and how this can be evaluated. The paper outlines in physical terms how appropriate numbers could be calculated but it does not attempt to establish a strict mathematical framework. The emphasis is placed on understanding how the (often-confused) terms resolution, contrast, dynamic range and signal-to-noise-ratio are related and what affects them in fluorescence microscopy.

The Airy disc describes the image of a point-like light source

In a microscope the field of a point-like light source in the image plane is equivalent to the system response. It is referred to through the amplitude point spread function and used to describe the properties of the optical components (Born & Wolf, 1980, pp. 370–458). While the distribution of the amplitudes cannot be seen directly, the intensity point spread function (i.e. an image) can be visualized, e.g. by placing a piece of paper into the optical path or by observing it directly using a camera orientated normal to the optical axis (Fig. 1). The intensity point spread function extends in

© 1998 The Royal Microscopical Society

all three dimensions. Owing to the cylindrical symmetry of lenses the two lateral components can be regarded as equal. The rotational symmetric Airy pattern describes the intensity distribution in the plane of focus as a function of the distance from the optical axis. The lateral distance is normalized by the numerical aperture and the wavelength and therefore has no unit (Hopkins, 1943). Important features of the Airy pattern are (Born & Wolf, 1980, pp. 370–458):

1 it is normalized to one on the optical axis;
2 the zero crossings occur at $v = 3{\cdot}8, 7{\cdot}0, 10{\cdot}2, 13{\cdot}3, \ldots$;
3 the heights of the first four higher order maxima are 1·7%, 0·4%, 0·2% and 0·08%;
4 about 84%, 91% and 94% of the total photon flux are inside the disc enclosed by the first through third zero crossings, respectively.

The Airy pattern tells us nothing about the component of the point-spread-function along the optical axis. In this paper we purposely ignore the axial component and restrict all analyses to influences on the lateral resolution.

Contrast and resolution are related terms

Let us assume we form an image of two point objects. Let us also assume the image formation process is incoherent, as is the case in fluorescence microscopy. Incoherence means that the imaging process is described using intensity point spread functions and interference effects are not present in the image. The resultant image then contains the sum of two Airy discs (Fig. 2). If the two objects are very far apart in object space the images are very far apart also and may be easily separable. If the two objects are very close their individual images overlap and the combined image may appear as a single image resulting from a brighter and/or larger single object. In general, the two images will overlap to a certain extent and will consist of two peaks with a gap between. The deeper the gap the easier it is to distinguish, i.e. to resolve the two objects in the image.

To quantify resolution we introduce the term contrast, which we define as the difference between the lowest intensity, found between the images of the two objects and the maximum intensity. We can calculate and plot contrast as a function of the centre-to-centre distance s between two objects (Fig. 3). Since the maximum height of a single Airy disc is 1, the highest achievable contrast is also 1 and achieved only when the spacing between the two objects is large. As the two objects come closer the contrast in the image decreases and becomes zero. The two maxima are then no longer visible and hence a contrast ceases to exist. The distance at which the contrast becomes zero is referred to as the contrast cut-off distance and defines the smallest distance between two objects which can be resolved in the image. The relationship between contrast and distance is called the contrast-vs.-distance function or contrast transfer function.

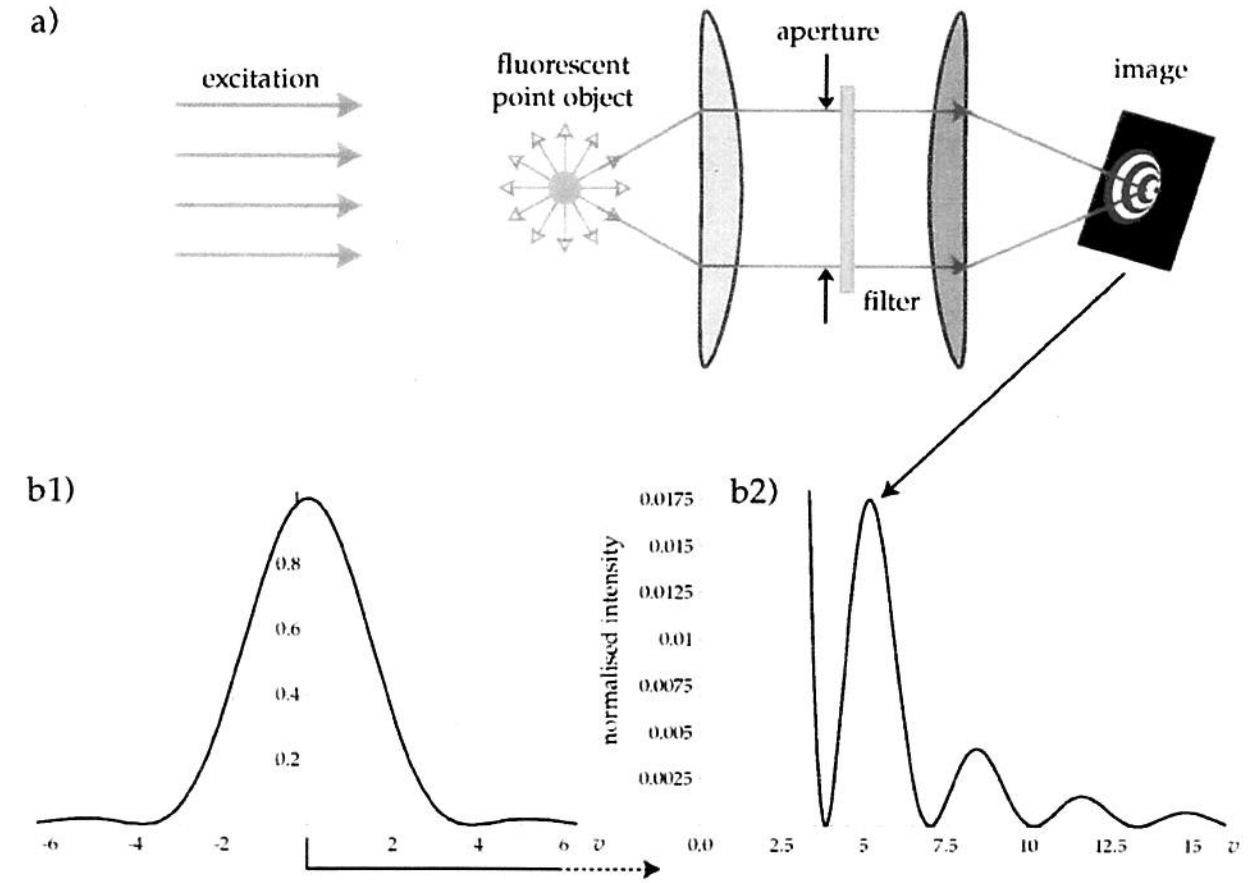
Now what is resolution? Resolution is the distance of two objects at which a certain contrast is achieved in their image. The ambiguity associated with the term resolution stems from the fact that several definitions make perfect sense. Some examples:

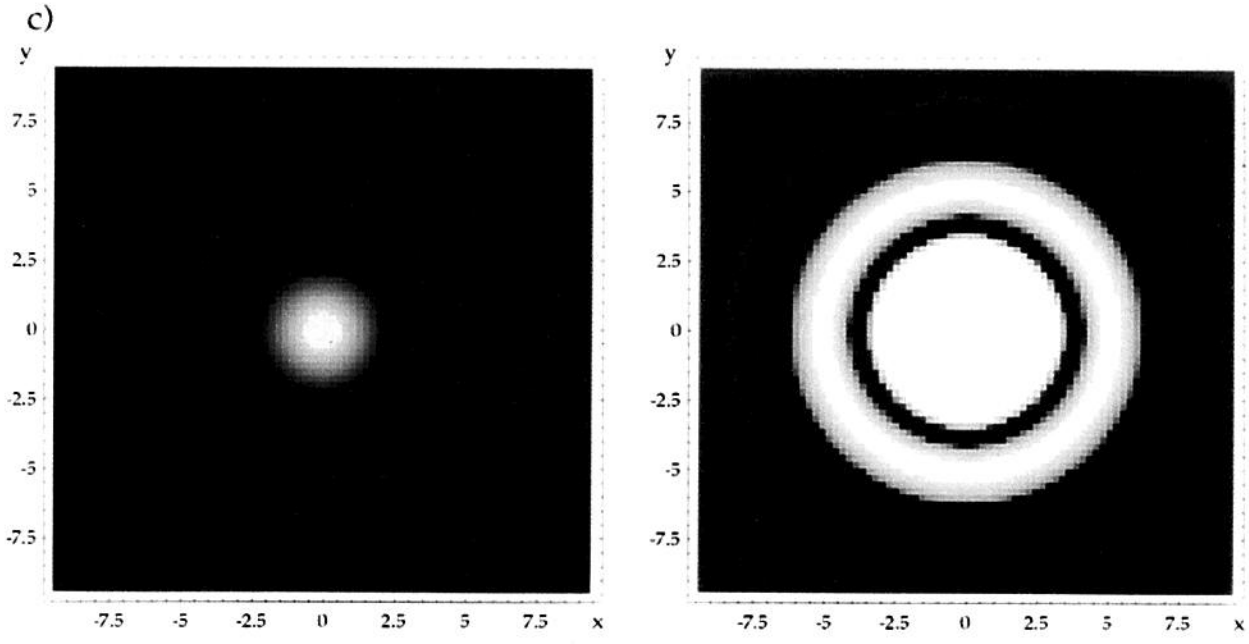
- The cut-off distance is defined as the distance below which two objects cannot be distinguished by means of contrast. The Sparrow criterion defines the resolution of an optical system as being equal to the cut-off distance.
- The Rayleigh criterion proposes to use the distance at which the contrast in an image is 26·4% as the resolution. Under ideal circumstances at this distance the maximum of one Airy pattern is above the first minimum of the other Airy pattern.
- The area below the contrast-vs.-frequency curve can be used to determine the resolving power and is especially useful since aberrations tend to cause bends, which are in general missed by the cut-off distance and may not affect certain contrasts.

In essence, any contrast between 0% and 99·9% can be used to define a resolution and one should be careful not to believe one definition to be fundamentally better than any other. It should also be clear that it does not make any sense to discuss contrast and resolution as if they were independent terms.

Although resolution is not defined by the extent of a single Airy disc, it is related to its full width at half-maximum (50%-diameter). An optical system with an Airy disk of a smaller 50%-diameter will provide the same contrast at a shorter distance. The Airy disc of,

Fig. 1. Airy pattern and Airy disc. (a) In a microscope a beam of light excites the fluorophores of a point-like object. The fluorescence emission occurs with an even probability in all directions. A small fraction of the emitted light is collected by a set of lenses, filtered and focused into an image plane where it appears as an Airy disc. (b) The intensity distribution of a point-like light source in the focal plane of a perfect optical system is described by the Airy pattern. The variable v is the distance from the optical axis normalized by the wavelength and the numerical aperture. The Airy pattern has a main maximum and higher order maxima with positions in between where the intensity becomes zero. (c) The Airy disc describes only the lateral extent of the intensity point spread function in terms of the intensity as a function of the distance v from the optical axis. It requires a cylindrical symmetric optical system. The image on the left is normalized to the main maximum in the centre of the Airy disc while the image on the right is normalized to the second maximum.


a)
excitation
fluorescent
point object
aperture
filter
image
b1)
b2)
normalised intensity
0.0175
0.015
0.0125
0.01
0.0075
0.005
0.0025
0.8
0.6
0.4
0.2
-6
-4
-2
2
4
6
v
0.0
2.5
5
7.5
10
12.5
15
v


c)
y
7.5
5
2.5
0
-2.5
-5
-7.5
-7.5
-5
-2.5
0
2.5
5
7.5
x

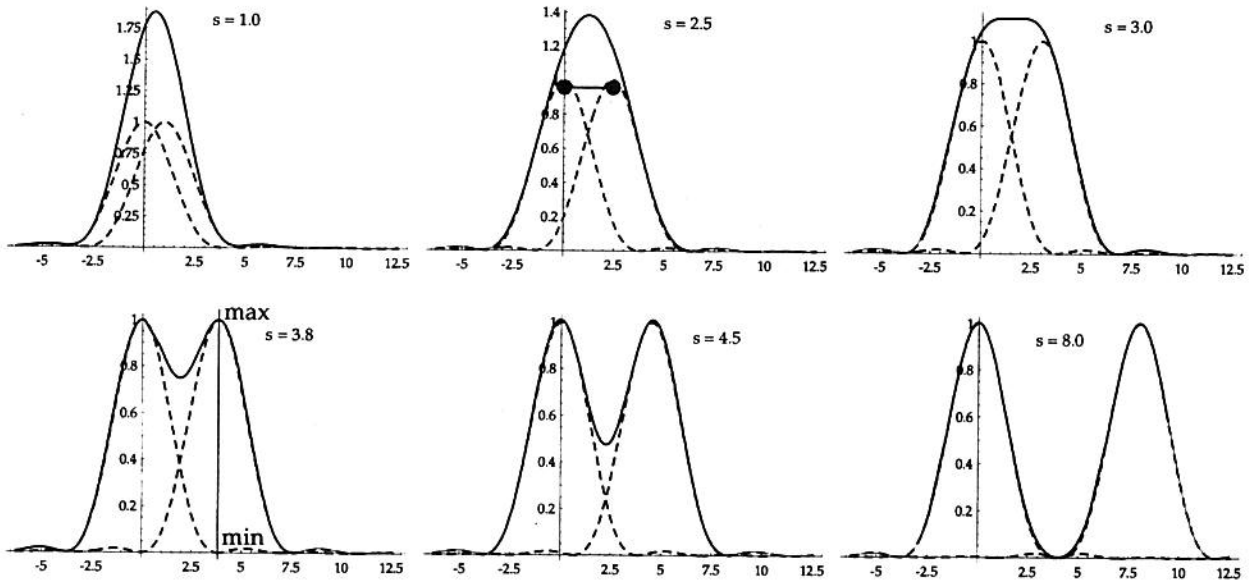


Fig. 2. Definition of contrast transfer functions. The depth of the gap between two Airy discs depends on their centre–centre distance s. If the two Airy discs are very close the gap does not exist. As the distance increases the depth of the gap increases and hence the contrast improves. Please note that the scale is different in the upper three graphs. In the case $s = 3{\cdot}8$ (the Rayleigh criterion) the maximum of the right Airy pattern overlaps with the first zero crossing of the left Airy pattern.

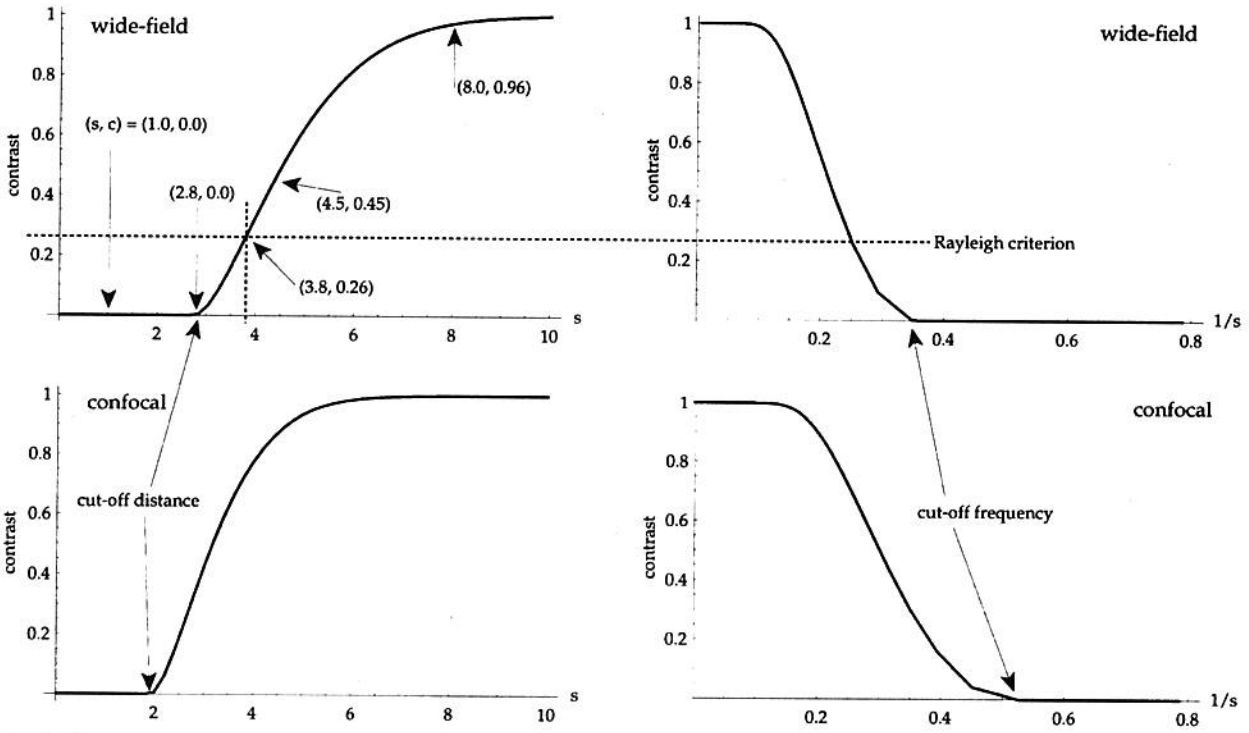


Fig. 3. The contrast c can be plotted as a function of the centre–centre distance s of two Airy discs. Only after a certain distance does the contrast start to exist. The two figures on the left show contrast vs. distance plots while the curves on the right show contrast vs. frequency plots. The upper two figures refer to wide-field fluorescence microscopy while the lower two refer to confocal fluorescence microscopy. The cut-off distance and the slope depend on the width of the Airy disc. The cut-off distance in a wide-field microscope is larger than in a confocal microscope. The more traditional way is to plot contrast over the inverse of the distance. This measure (which describes how many objects are present per unit length) is also referred to as the spatial frequency. The smaller the width of the Airy disc the higher is the cut-off frequency. The dashed lines indicate the well-known Rayleigh criterion.

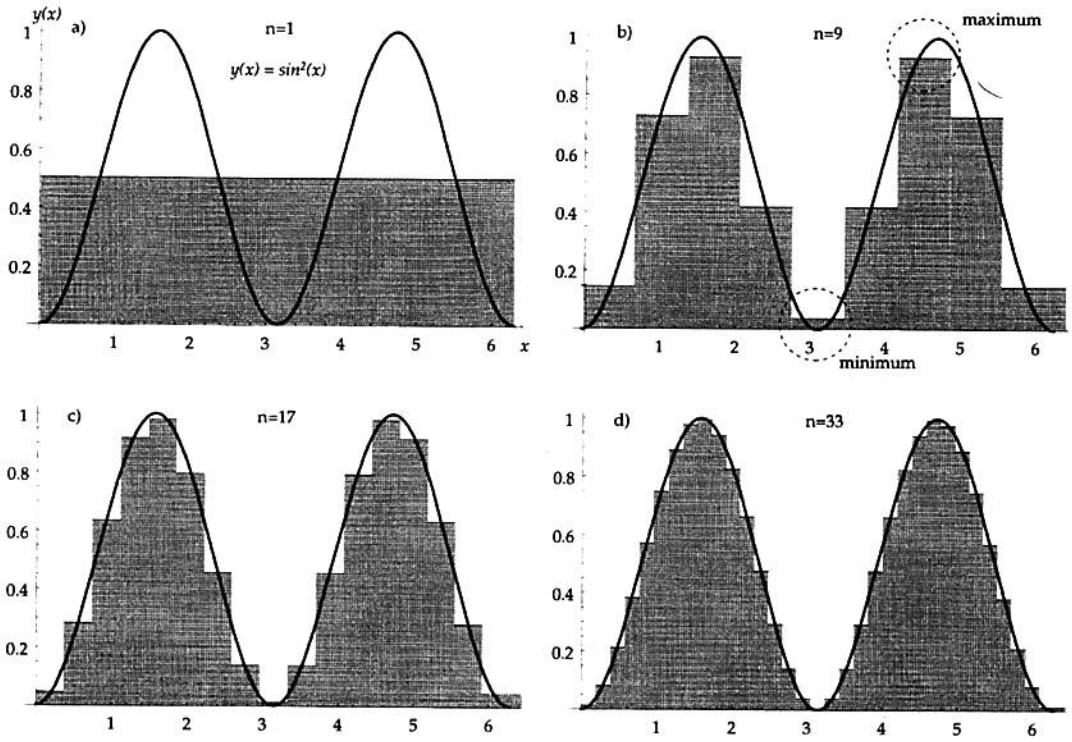


Fig. 4. Influence of pixelation on contrast. The finite number of pixels in a camera or a scanner breaks the ideal, continuous function down into a series of intensity values. Each of these values represents the intensity integrated over a small area. This procedure causes maxima to become smaller and minima to become larger. The main effects are a decrease of the contrast and an ambiguity concerning the positions of minima and maxima.

for example, a confocal fluorescence microscope (the Airy disc of a confocal fluorescence microscope is the square of the Airy disc of a wide-field fluorescence microscope) has the same zero crossings but a smaller 50%-diameter and therefore the cut-off distance is shorter.

Table 1. The contrast ranges in per cent achievable for various numbers of picture elements per Airy pattern, per Airy disc and per Airy object consisting of two Airy discs at the Rayleigh distance ($s = 3{\cdot}8$).

Pixels per Airy:			
pattern	disc	object	Contrast (%)
32	1024	1536	26·2–25·8
16	256	384	25·8–24·1
8	64	96	24·0–17·6
4	16	24	17·0–0·0
2	4	6	0·0
1	1	2	0·0

Pixelation reduces the contrast

Until now we assumed that the Airy pattern was smooth and described by an infinite number of samples. In a real optical system, however, we use a camera or a laser scanner to generate a finite number of picture elements (pixels), i.e. we pixelate the image (Fig. 4). Each pixel then summarizes the response of the optical system in a certain area. Referring again to the Airy pattern please note that the zero crossings occur at points. Since we never look at points, but always at areas, we will never be able to detect a value of zero intensity. The consequence for our contrast is that pixelation reduces the contrast since it increases the minimum and decreases the maximum. The cut-off distance is increased and consequently the resolution (independent of the definition) is decreased. So, whenever we use a camera in an optical system we will never be able to achieve the theoretical resolution.

The effect of pixelation can be estimated by relating the extent of a pixel to the diameter of the Airy disc. The question is: how many pixels are required to cover the area below two adjacent Airy discs to achieve a certain contrast? The more nonoverlapping pixels are used the smaller becomes the effect due to pixelation.

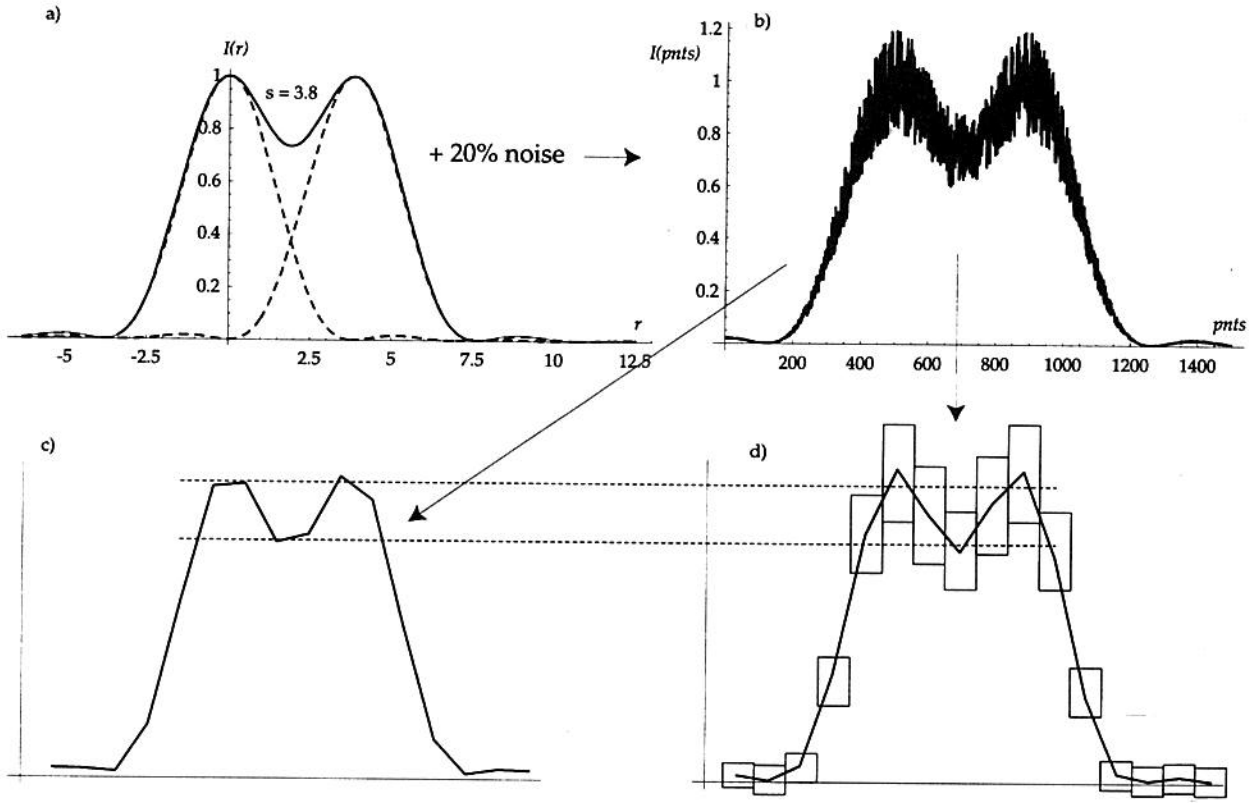


Fig. 5. Influence of noise on contrast. (a) Ideal image representation of two point-like objects at the Rayleigh distance. (b) The same conditions but now a 20% noise level has been included. (c) and (d) Twelve points are used to represent the sum (i.e. eight points per Airy pattern). Both presentations are valid. The boxes indicate the uncertainties associated with each intensity value in the Airy pattern.

A detailed analysis (Table 1) shows that four samples per Airy pattern along the main axis of two objects (which under ideal conditions provide a contrast of 26·4%) result in contrasts between 0% and 17%. Therefore, at least eight samples per Airy pattern are required to guarantee a contrast of at least 17·6% in an Airy object under all circumstances. Using a lens with a numerical aperture of 1·3 at a wavelength of 500 nm and assuming the typical field to consist of 512 samples per axis

$$1{\cdot}22\frac{500\,\text{nm}}{1{\cdot}3}/8 = 465\,\text{nm}/8 = 58{\cdot}2\,\text{nm}$$

$$512\frac{\text{samples}}{\text{lines}}\cdot\frac{58{\cdot}2\,\text{nm}}{\text{samples}} = 29{\cdot}8\frac{\mu\text{m}}{\text{line}}$$

we conclude that the pixel size (equivalent to the pixel–pixel distance) must be of the order of 60 nm while the field size should be at most 30 μm. Otherwise objects that are about 240 nm apart have a contrast that is less than 17%. Keeping the number of samples per axis fixed a field size of 60 μm will already be blind to objects of such an extent.

Noise reduces the contrast

In a real situation we always have the problem of noise. Whatever value we measure it is only an approximation for the number of photons that have been emitted or scattered by an object. The variation of the signal during repeated observations is called noise. Noise induces an uncertainty in our quantification (e.g. Carlsson, 1991) of the intensity and hence an uncertainty in the contrast that has been measured. Since we have to accept an error in our estimation of the intensity we have to accept that the contrast is in general underestimated. However, every reduction of the contrast causes an increase of the cut-off distance and hence a decrease of the resolution.

It is probably needless to say that the effects of noise and pixelation have to be combined (Fig. 5) and will always increase the cut-off distance. While the effects of pixelation can be estimated quite easily, it is not so easy to estimate the noise in a pixel. Sources of noise are, for example, the finite number of photons and electronic noise. In transmission or reflection contrasts the number of photons will be very high

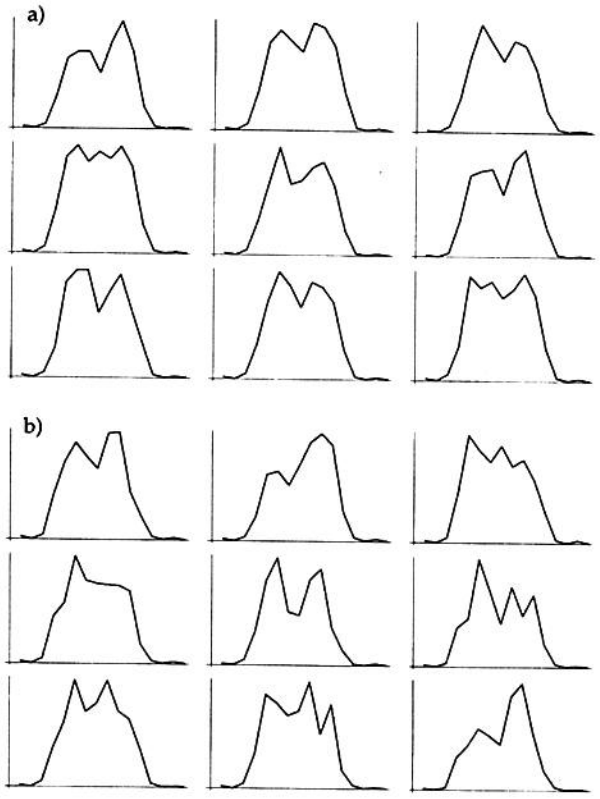


Fig. 6. Influence of noise on contrast of two point-like objects located at the Rayleigh distance. (a) A noise level of 20% is included; (b) a noise level of 40% is included in the calculations. The images vary considerably in both galleries, indicating that the images could stem from one large as well as from three small objects.

and the most likely sources of noise will be electronic equipment and variations in the illumination intensity. In fluorescence microscopy the number of fluorescence photons is very small and Poisson noise will dominate (Fig. 6).

Dynamic range influences the cut-off frequency

Once we have estimated the amount of noise relative to the amount of signal we can estimate the number of distinguishable grey levels, i.e. the dynamic range in a pixel or in an image. The dynamic range is essentially the ratio of the signal over the noise and in the case of a fluorescence microscope the square root of the number of photons. (The digital number that describes the intensity of a pixel is only proportional to the number of photons but not identical to it.)

$$\text{Dynamic range (in a fluorescence microscope)} = \text{signal/noise} = \text{signal}/\sqrt{\text{signal}} = \sqrt{\text{signal}}.$$

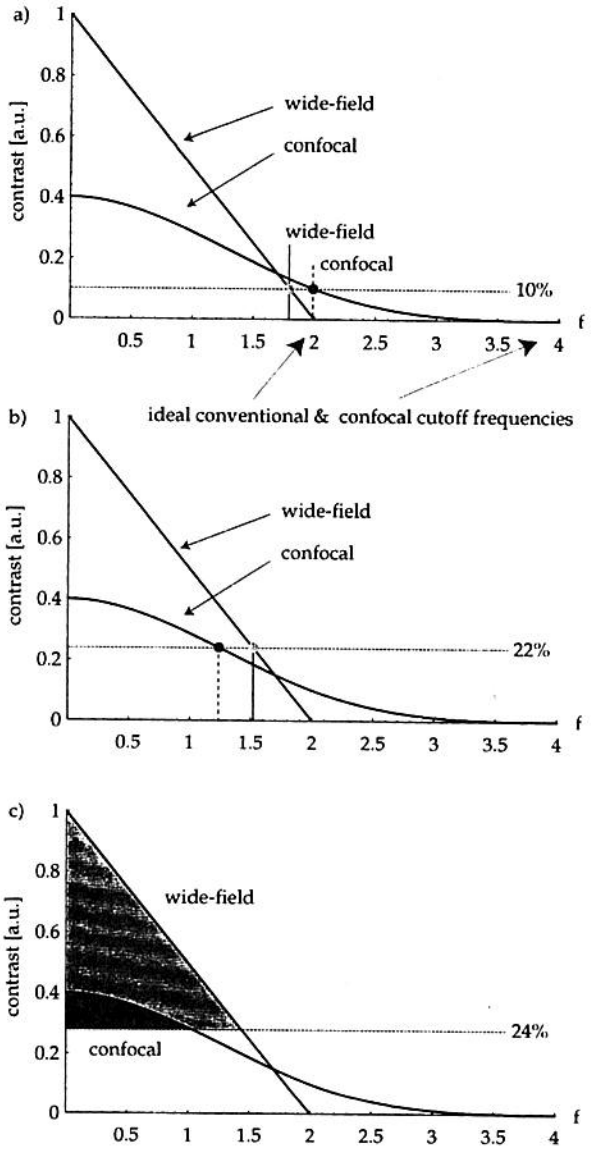


Fig. 7. Influence of noise on cut-off frequency. Indicated are the optical transfer functions (OTF) for wide-field and confocal fluorescence microscopy. Since the signal is lower in confocal microscopy the level of the confocal OTF is lower. Noise is introduced by imposing a lower limit to the contrast. The simplest case is a constant frequency-independent noise level. (a) A noise level of 10% is assumed. The cut-off frequency of the confocal microscope is not much higher than that of the wide-field microscope. (b) At a level of 22% the cut-off frequency of the wide-field microscope is higher. (c) A different criterion is the area below the curve. Appropriately weighted it is proportional to the energy or the number of photons collectable from a diffraction-limited object.

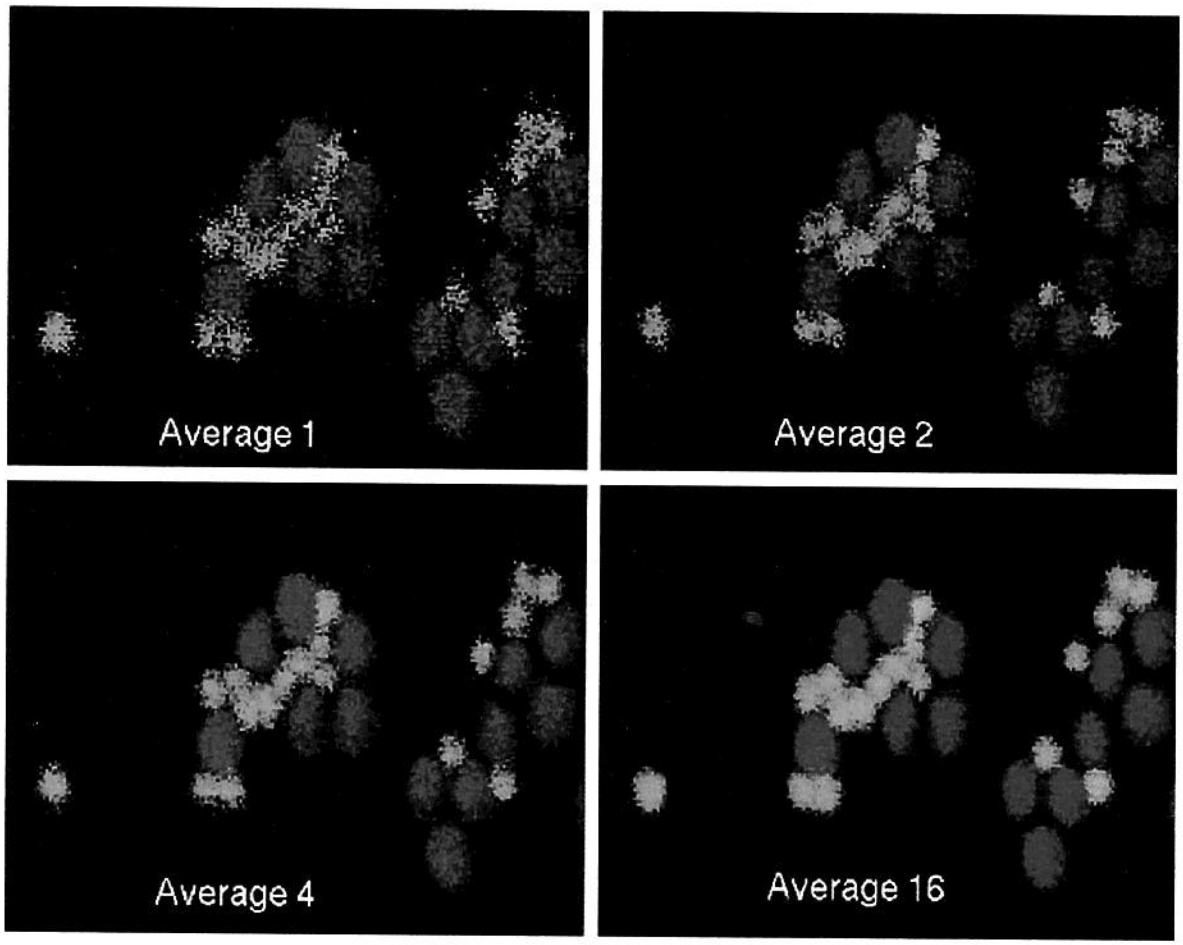


Fig. 8. Four images of fluorescent latex beads. The green latex beads have a diameter of 0·5 μm; the red latex beads have a diameter of 1 μm. The images were recorded with averages of 1, 2, 4 and 16 frames. The dwell time was of the order of 80 ns. The dots in the upper left image indicate single events, i.e. each dot corresponds to one photon. The objects had been focused to the centre of the smaller beads. The larger beads therefore appear smaller than they actually are. While the red beads can be distinguished easily in the nonaveraged image, eight averages are required for beads with a diameter of 0·5 μm and 16 averages are already too many.

The noise provides us with a lower limit below which we cannot distinguish between different signals. Coming back to the contrast-vs.-distance functions we defined earlier, their cut-off distances were determined by assuming an infinitely high signal-to-noise ratio and hence an infinitely high dynamic range. The cut-off is the distance at which the function crosses the zero contrast line. In the simplest case noise is introduced by raising this level (Fig. 7). Only those distances that provide at least a certain contrast can be distinguished. If we wish to estimate the resolution in a noise-limited situation we must use this level as a reference. Obviously the cut-off distance and cut-off frequency are affected.

Figure 8 illustrates this important aspect. In the image in the upper left the beads consist of clusters of individual dots, each of which is created by a single photon. Small beads that are far apart and large beads can be easily distinguished but small beads that are very close form a large cluster. As we increase the number of averages and hence improve the signal-to-noise ratio more of the beads become visible. However, more than eight averages are not required since in the lower left and in the lower right frame all objects are discernible. An average of eight is therefore optimal for this specimen. While a nonaveraged image is sufficient to distinguish the 1-μm beads, an average of less than $4^2 = 16$ is required to distinguish the 0·5-μm beads. This is understandable because they are a factor of 2 smaller in diameter and a factor of 4 smaller in area.

Decreasing the diameter of the point-spread-function increases the noise

Improving the resolution means that the observed area or volume is decreased. If object and object preparations stay the same, this means that the number of observable fluorophores and in consequence the number of collectable photons is also decreased. Since the dynamic range ultimately depends on the number of photons that can be collected from a given number of fluorescent molecules any improvement that decreases the observable region has to increase the noise. This imposes a limit to any technique that intends to improve resolution by decreasing the size of a point-spread function. If, for example, the observable volume is decreased by a factor of 2, the signal is reduced by at least a factor of 2 and the observation period has to be increased by a factor of 4 simply to maintain the signal-to-noise ratio.

Acknowledgments

S. Lindek, N. Salmon and E.-L. Florin are thanked for critical reviews of the manuscript. Thanks to James B. Pawley for discussing the signal-noise ratio in electron microscopy. N. Salmon recorded the images in Fig. 8.

References

Born, M. & Wolf, E. (1980) *Principles of Optics*. Pergamon Press, Oxford.

Brakenhoff, G.J., Blom, P. & Barends, P. (1979) Confocal scanning light microscopy with high aperture immersion lenses. *J. Microsc.* **117**, 219–232.

Carlsson, K. (1991) The influence of specimen refractive index, detector signal, integration and non-uniform scan speed on the imaging properties in confocal microscopy. *J. Microsc.* **163**, 167–178.

Florin, E.L., Hörber, J.K.H. & Stelzer, E.H.K. (1996) High-resolution axial and lateral position sensing using two-photon excitation of fluorophores by a continuous-wave Nd:YAG laser. *Appl. Phys. Lett.* **69**, 446–448.

Goodman, J.W. (1968) *Introduction to Fourier Optics*. McGraw-Hill, San Francisco.

Hopkins, H.H. (1943) The Airy disk formula for systems of high relative aperture. *Proc. Phys. Soc.* **55**, 116–120.

Saxton, M. & Jacobson, K. (1997) Single-particle tracking: Application to membranes. *Ann. Rev. Biophys. Biomolec. Struct.* **26**, 373–399.

Sheppard, C.J.R. & Choudhury, A. (1977) Image formation in the scanning microscope. *Opt. Acta*, **24**, 1051–1073.

Appendices

What an object would look like

In this paper we investigate images of point objects. These objects are so small that their features cannot be resolved. If we use the terminology that was developed earlier we define point objects as objects whose diameters are much smaller than the cut-off distance of an optical system. In fluorescence microscopy we look at objects in which, for example, the diameter of the area over which the fluorescent molecules are scattered or, for example, the maximum distance between two fluorescent molecules is very often smaller than the cut-off distance. Each of the molecules creates an Airy disc in the image but they are so close to each other that their sum is indistinguishable from a single Airy disc.

Dynamic range and number of grey levels

Because of the properties of fluorescent dyes a confocal fluorescence microscope will typically record less than 30 photons pixel^{-1} μs^{-1}. In most cases it will record at most 10 photons pixel^{-1} μs^{-1}. The dynamic range will thus vary between $\sqrt{10} \approx 3$ and $\sqrt{30} \approx 5$. The six distinguishable grey value ranges are 0, 1, 2–4, 5–9, 10–16 and 17–25. (Or in a different notation: 0, 1±0, 3±1, 7±2, 13±3, 21±4.) This is also the reason why it is sufficient to store the square root of the count rate. That number will contain all the available information.

Distance measurements are not strictly resolution dependent

In this paper we have so far looked at two objects that emitted light in the same band of wavelengths. The wavelength bands of the emitted photons were identical and therefore a single image showed the sum of two objects. If, however, the point-like objects emit at discernible wavelength bands, two images can be recorded independently. Each image will show a single object and the two images will not affect each other. The exact location of each object can be calculated using the intensity-weighted 'centre of mass equivalents' and the distance of any two objects can be determined with a noise-limited precision. Again we are limited by the number of photons, but as pointed out earlier fewer photons are required to identify the location of a single object in an image than to distinguish two objects in a single image.

The same argument also applies to single-particle tracking in video sequences (e.g. Saxton & Jacobson, 1997). The issue is to determine distances from intensity-weighted 'centre of mass equivalents' in independently recorded images. Such distances can be below 20 nm.

Another example is the photonic force microscope (Florin *et al.*, 1996) which uses the position of a single bead to determine a three-dimensional structure. The position of the bead inside the focal volume can be determined with a precision that is most likely below 10 nm. The distance of

topological structures in an object can hence be determined with a resolution around 15 nm.

A final example is the determination of the surface topology of integrated circuits using, for example, confocal reflection microscopes. The height differences of planes that are sufficiently far apart can be determined with an unlimited precision. Here the surface roughness, the precision with which the position of the object can be measured along the optical axis and the reflectivity of the surface probably limit the resolution.

Transfer functions

The concept of transfer functions (Goodman, 1968, p. 395) may seem complicated but it has the advantage of being able to cope with aberrations, with noise, and with a number of nonlinear effects at the same time. In this paper we used the lateral distance between Airy disc centres to define the abscissa. But transfer functions can also be defined based on frequency of rectangles or triangles. The most useful concept is to base contrast transfer on the frequencies of sine functions since all objects can be approximately described by finite harmonic series and therefore all transfer functions can be generated from sine transfer functions. Since the sine function has a value of zero in a single point the maximum contrast is only achieved at infinitely high distances, i.e. close to the dc component.

The shape of the object we wish to analyse has to be taken into account. This is often forgotten when regular patterns are used to determine the resolution of an optical system. It is probably a common error to confuse the transfer function of line pairs which are rectangular (or square wave) functions with those of sine waves.

Aberrations such as the spherical aberration cause continuous bends in the transfer function. It is then lower than the unaberrated transfer function but at no distance (or frequency) larger.

The existence of transfer functions is not obvious. A number of conditions have to be met by an optical system:
1 The optical system must be linear, i.e. no light must be absorbed or scattered and no object must be found in the shadow of another object.
2 The optical system must be invariant, i.e. features in the image of an object must be independent of the position of the object.

Since these conditions are never perfectly fulfilled the concept of transfer functions is only approximately correct.