12

Practical Limits to Resolution in Fluorescence Light Microscopy

ERNST H. K. STELZER

INTRODUCTION

In a perfect optical system, the NA and the wavelength determine the resolution. In a real fluorescence microscope, the number of photons collected from a specimen determines the contrast and, thus, the resolution. The number of picture elements per resolvable unit area, the number of photons, and the aberrations affect the contrast (Castleman 1993). The theory of confocal fluorescence microscopy claims that the resolution in all three dimensions is improved over that of wide-field fluorescence microscopes. This implies that laser scanners should be used instead of video cameras and that the improved lateral resolution can be used to visualize otherwise unidentifiable features. Such claims are backed up by theoretical comparisons of confocal and wide-field microscopy. In transmission and reflection contrasts, the resolution in terms of the cut-off frequency is improved by a factor of two (Sheppard and Choudhury 1977), whereas the full width at half-maximum of a point object is improved by a factor $1/\sqrt{2} \approx 1/1.4 \approx 0.7$ (Brakenhoff et al. 1979). On the other hand, every user of a confocal fluorescence microscope knows that faint features in an object which provides a low signal are not easily distinguished. In fact, it is common experience that human visual perception becomes much worse in the dark.

This chapter summarizes 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 an image of two closely spaced small objects is affected by noise. The chapter outlines how appropriate numbers can be estimated. It does not attempt to establish a strict mathematical framework. The emphasis is placed on understanding how resolution, contrast, dynamic range, and S/N ratio are related and what affects them in fluorescence microscopy (Stelzer 1998).

AIRY DISKS DESCRIBE IMAGES OF POINT-LIKE LIGHT SOURCES

In microscopy the spatial distribution of a point light source in the image plane is equivalent to a system response. It is referred to as the amplitude point-spread function (amplitude PSF) and is used to describe the properties of the optical components (Born and Wolf 1980). Although the distribution of the amplitudes cannot be seen directly, the intensity PSF (i.e., an image) can be directly visualized by placing a piece of paper into the optical path or by indirect observation through a camera (Figure 12.1). The intensity PSF extends in all three dimensions. Due to the cylindrical symmetry of microscope objective 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 (Hopkins 1943).

The lateral Airy pattern tells us nothing about the component of the PSF along the optical axis. Just for reasons of simplicity, we purposely ignore the axial component and restrict all analyses to influences on the lateral resolution.

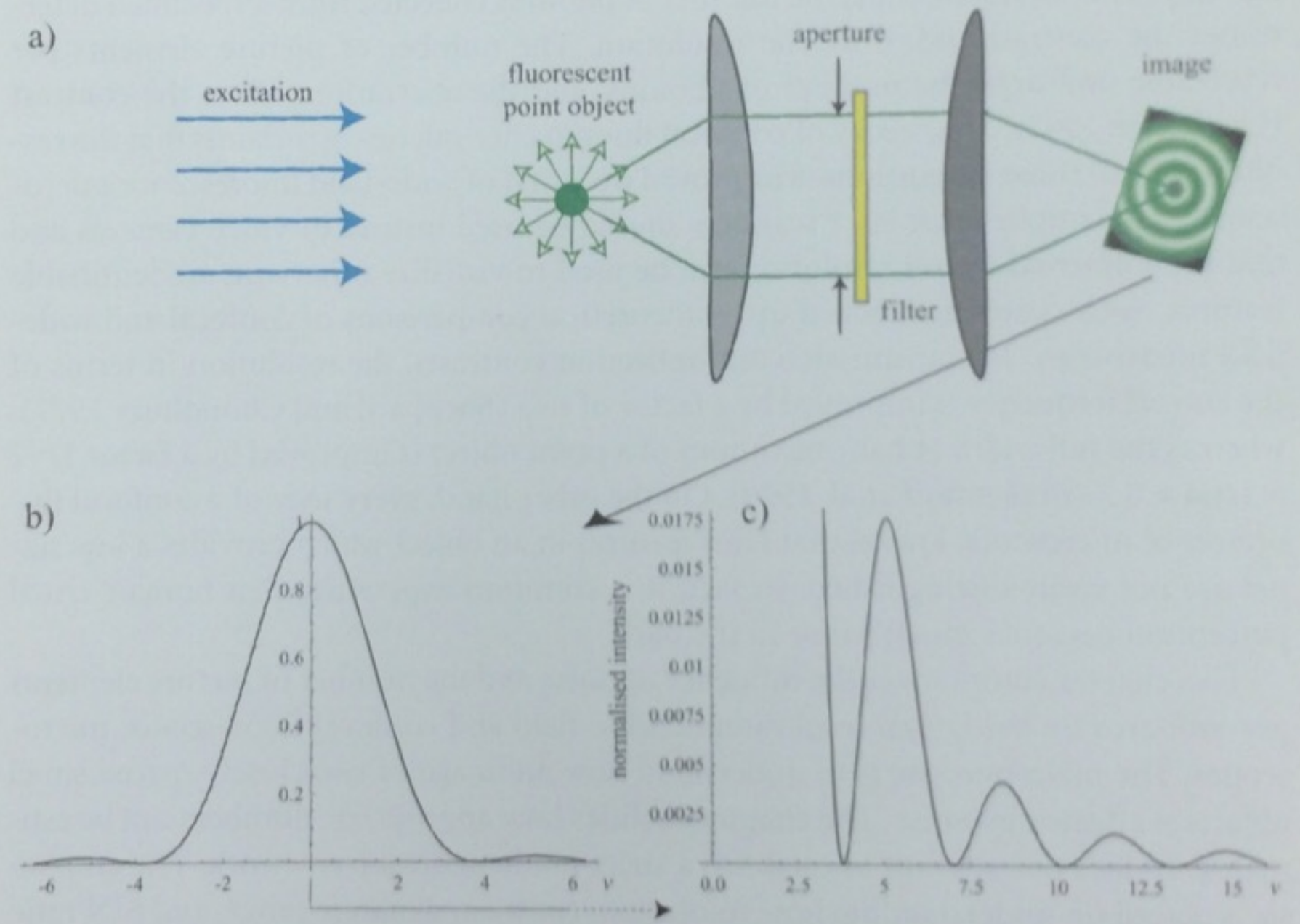

Figure 12.1. Airy pattern and Airy disk. (*a*) In a microscope, a beam of light excites the fluorophores of a point-like object. The fluorescence emission occurs with an even probability in all directions. A small fraction of the emitted light is collected by a set of lenses, filtered, and focused into an image plane where it appears as an Airy disk. (*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 NA. (*c*) The Airy pattern has a main maximum and higher order maxima with positions in between where the intensity becomes zero.

CONTRAST AND RESOLUTION ARE RELATED TERMS

In fluorescence microscopy the imaging process is described using intensity PSFs. The resultant image is a sum of Airy disks (Figure 12.2). Let us assume we form an image of two point objects. If the two objects are very far apart in object space, the images are very far apart, too, and 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 how well the two objects are separable, i.e., how well they are resolved, we introduce the term contrast. It is defined as the difference between the lowest intensity found between the images of the two point objects and the highest intensity in the image. Contrast can be calculated and plotted as a function of the center–center distance s between two objects (Figure 12.3). Because the height of a single Airy disk is one, and the minimum between two Airy disks is at best zero, the highest achievable contrast is one. It is only achieved when the distance between the two objects is large.

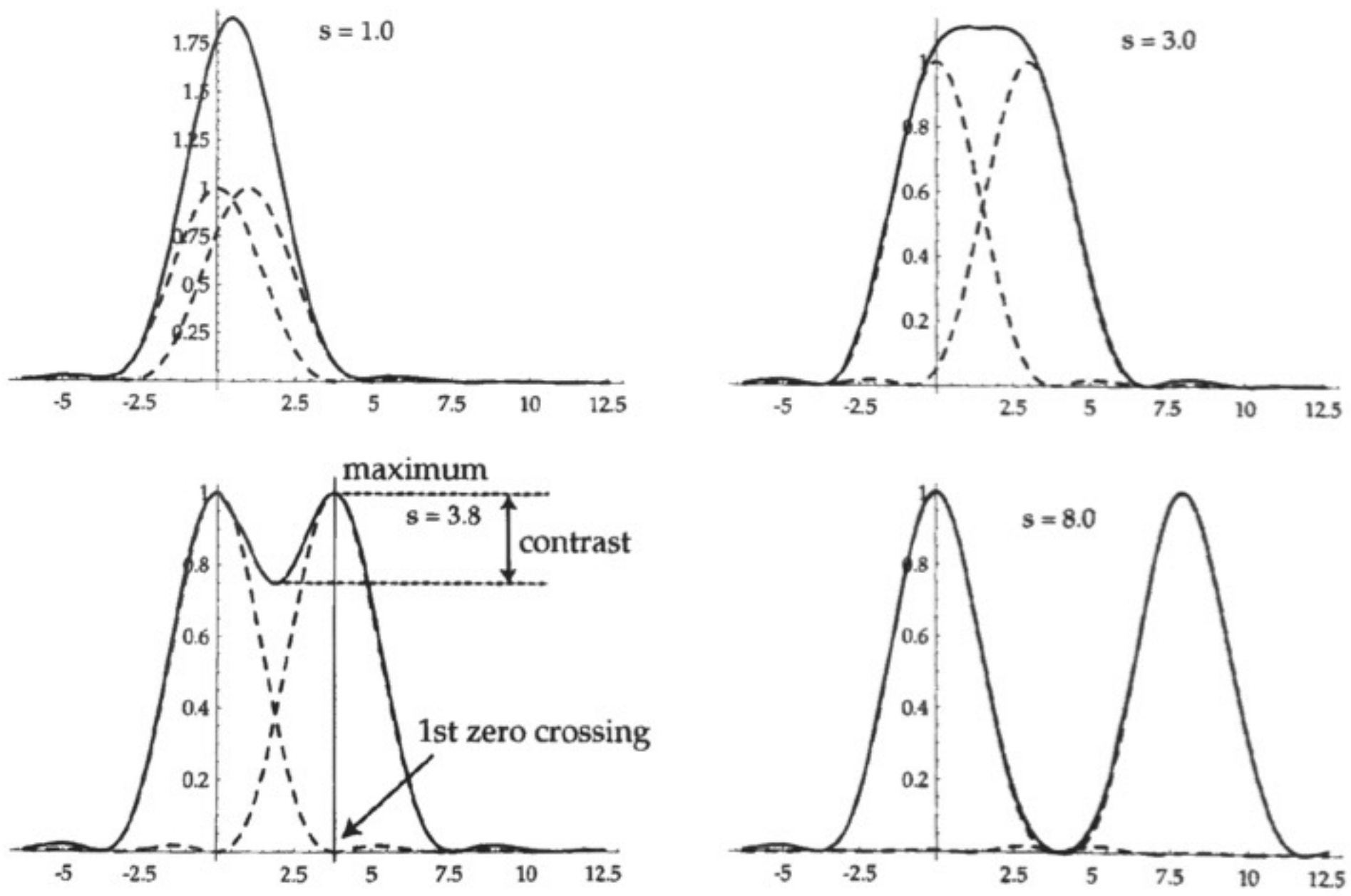

Figure 12.2. Definition of contrast versus distance functions. The depth of the gap between two Airy patterns depends on their center–center distance or separation s. If the two Airy patterns are very close to each other, the gap does not exist ($s = 1$ and $s = 3$). As the distance increases, the depth of the gap increases and hence the contrast improves. Please note that the scale is different in the upper two graphs. In the case $s = 3.8$ (the Rayleigh criterion), the maximum of the right Airy pattern overlaps with the first zero crossing of the left Airy pattern.

If two objects come closer to each other, the contrast in the image decreases. At a certain distance the contrast becomes zero. The two maxima are no longer distinguishable and hence a contrast ceases to exist.

The distance s at which the contrast becomes zero is referred to as the contrast cut-off distance and defines the smallest distance between two objects that can be resolved in the image. The relationship between contrast and distance is called the contrast-versus-distance function, which is similar in concept to contrast transfer function.

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

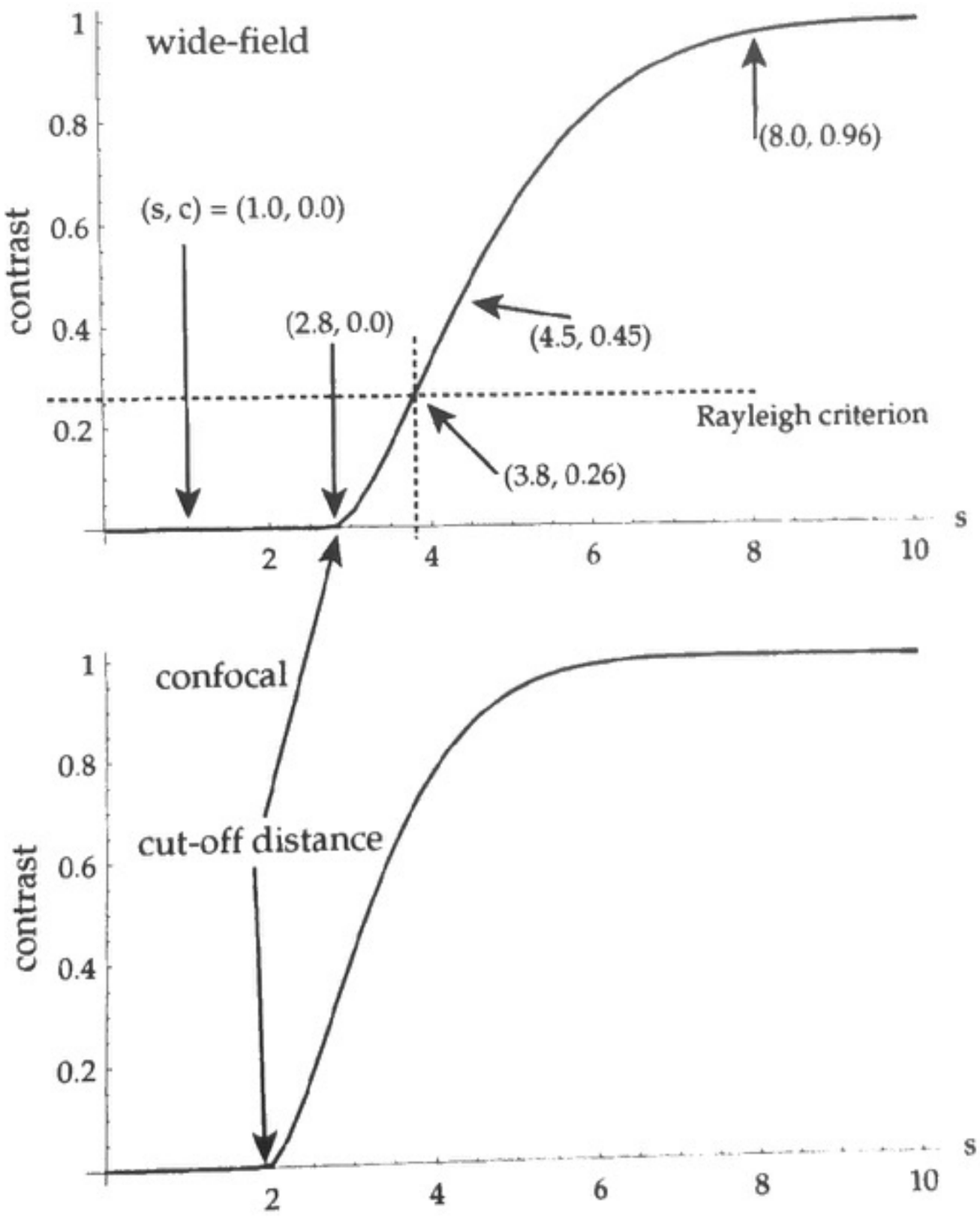

Figure 12.3. The contrast c can be plotted as a function of the center–center distance s of two Airy disks. Only beyond a certain distance does the contrast start to exist. The upper figure refers to wide-field fluorescence microscopy and the lower figure refers to confocal fluorescence microscopy. The cutoff distance and the slope depend on the width of the Airy disk. The cutoff distance in a wide-field microscope is larger than in a confocal microscope. The traditional way is to plot contrast over the inverse of the distance. This measure describes how many objects are present per unit length and is called spatial frequency. The smaller the width of the Airy disk, the higher is the cutoff frequency. The dashed lines indicate the Rayleigh criterion.

- The cutoff distance is the distance beyond which two objects cannot be distinguished. The Sparrow criterion defines the resolution of an optical system through its cutoff distance.

- The Rayleigh criterion uses the distance at which the contrast is 26.4%. At this distance the maximum of one Airy pattern coincides with the first minimum of the other Airy pattern.

Any contrast between 0% and 99.99% can be used to define a resolution. No definition is fundamentally better. The important message of this discussion is that it does not make any sense to discuss contrast and resolution as if they were independent.

SAMPLING 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 sample the image (Figure 12.4). Each pixel summarizes the response of the optical system in a certain area. When looking at the Airy pattern, note that the zero crossings occur at points. However, because the pixels have a finite extent, one never looks at points, but always at areas. The areas integrate the intensity. Therefore, it is impossible to detect zero intensity. Since this effect increases the minimum value as well as it decreases the maximum value, the consequence is that the sampling process reduces the contrast. The cutoff distance is increased and consequently the resolution (independent of the definition) is decreased. Thus, whenever one uses a camera in an optical system, one will never be able to achieve the theoretical maximal resolution.

The effect of sampling can be estimated by relating the extent of a pixel to the diameter of the Airy disk. The question is, How many pixels are required to cover the area

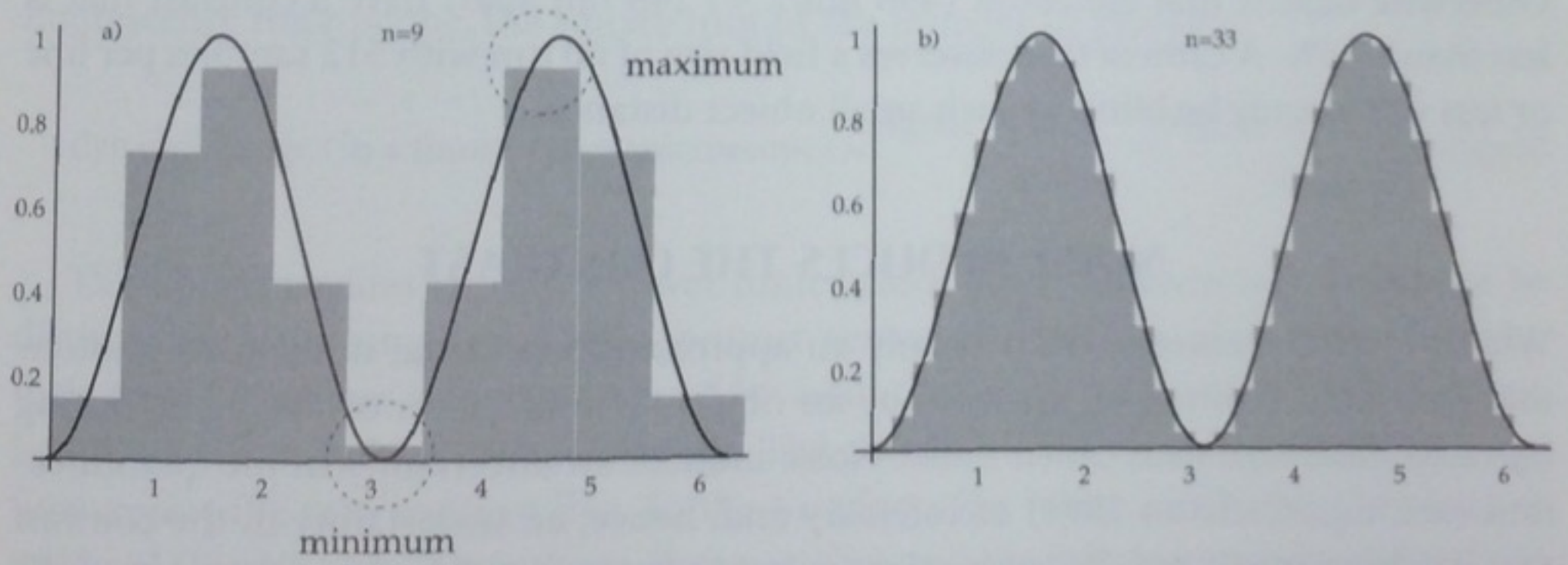

Figure 12.4. Influence of sampling on contrast. The finite number of pixels in a camera or a scanner breaks the continuous function down into a series of intensity values. Each of these values represents the intensity integrated over a small area. This procedure causes maxima to become smaller and minima to become larger. The main effects are a decrease of the contrast and an ambiguity concerning the positions of minima and maxima.

Table 12.1. *The contrast ranges in percent achievable for various numbers of picture elements per Airy pattern, per Airy disk, and per Airy object consisting of two Airy disks at the Rayleigh distance (s=3.8)*

Pixels per Airy pattern	Disk	Object	Contrast (%)
64	4096	5888	26.4–26.2
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

below two adjacent Airy disks to achieve a certain contrast? The more nonoverlapping pixels are used, the smaller becomes the effect due to sampling (see Table 12.1). Analysis 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%. 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 water-immersion objective lens with a NA of 1.2 at a wavelength of 488 nm and assuming the typical field to consist of 512 samples per axis, the pixel size, i.e., the pixel–pixel distance has to be on the order of 60 nm.

$$1.22 \cdot \frac{488\ \text{nm}}{1.2} \Big/ 8 = 496\ \text{nm} \Big/ 8 = 62.0\ \text{nm}$$

The horizontal field size should be at most 32 µm.

$$512\ \frac{\text{samples}}{\text{line}} \cdot \frac{62.0\ \text{nm}}{\text{sample}} = 31.8\ \frac{\text{micron}}{\text{line}}$$

Otherwise objects that are about (496 nm/2 =) 248 nm apart have a contrast that is less than 17.6%. A camera that observes a field size of 60 µm with 512 samples per line or less will already be blind to such small object distances.

NOISE REDUCES THE CONTRAST

Whatever value is measured, it is only an approximation of 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 the quantification (see, e.g., Carlsson 1991) of intensity and, hence, an uncertainty in the contrast that has been measured. Because an error has to be accepted in the estimation of the intensity, it has to be accepted that the contrast is in general underestimated. However, every reduction of the contrast causes an increase of the cutoff distance and, hence, a decrease of the resolution.

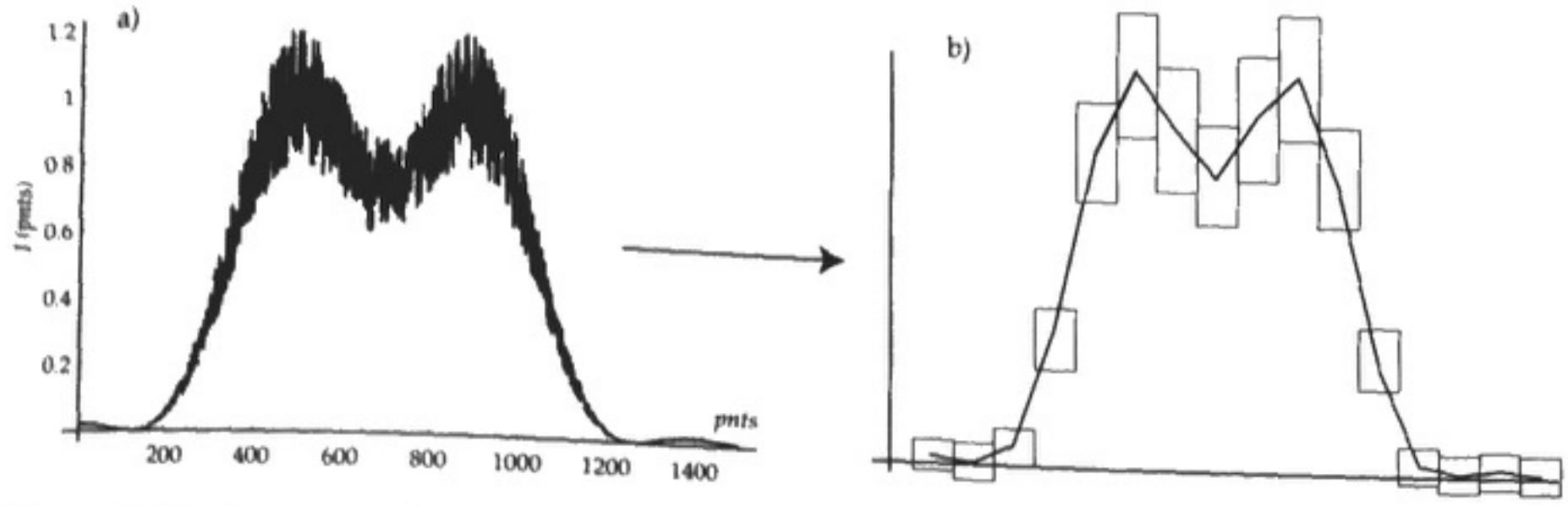

Figure 12.5. Influence of noise on contrast. (*a*) About 20% noise has been added to the image of two point-like objects at the Rayleigh distance. (*b*) Twelve points are used to represent the sum (i.e., eight points per Airy pattern). The heights of the boxes indicate the uncertainties associated with each intensity value in the Airy pattern.

It is obvious that the effects of noise and sampling have to be combined (Figure 12.5) and will always increase the cutoff distance. The effects of sampling can be estimated, but it is not as easy to estimate the noise in a single pixel. In transmission or reflection contrasts, the number of photons will be very high and the most likely sources of noise will be due to the electronic equipment and variations in the illumination intensity. In fluorescence microscopy, the number of fluorescence photons is very small and Poisson noise will be dominating.

DYNAMIC RANGE INFLUENCES THE CUTOFF FREQUENCY

Once the noise relative to the signal has been estimated, one can calculate the number of distinguishable gray 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.

$$\text{dynamic range (in a fluorescence microscope)} = {}^{\text{Signal}}\Big/_{\text{Noise}} = {}^{\text{Signal}}\Big/_{\sqrt{\text{Signal}}} = \sqrt{\text{Signal}}$$

The noise provides us with a lower limit below which different signals cannot be distinguished. Coming back to the contrast-versus-distance functions defined earlier, their cutoff distances were determined by assuming an infinitely high S/N ratio and, hence, an infinitely high dynamic range. The cutoff is the distance at which the function crosses the zero contrast line. In the simplest case, noise is introduced by raising this level (Figure 12.6). Only those distances that provide at least a certain contrast can be distinguished. In estimating the resolution in a noise-limited situation, this level must be used as a reference. Obviously, the cutoff distance is directly affected. It should be noted that the digital number that describes the intensity of a pixel is only propor-

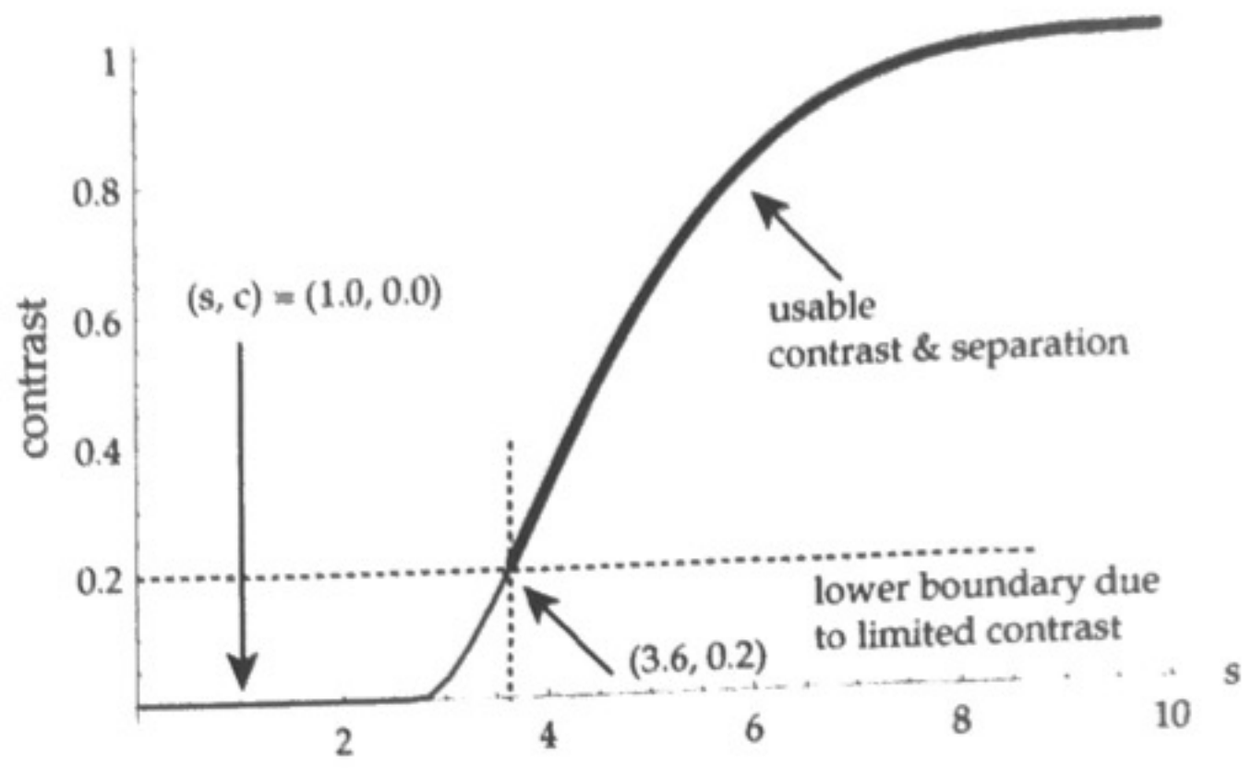

Figure 12.6. Influence of noise on cutoff distance. Noise is introduced by imposing a lower limit to the contrast. The simplest case is a constant distance-independent noise level. A noise level of 20% is assumed. This reduces the usable portion of the contrast versus distance function. Objects whose separation is smaller than the cutoff due to the lower boundary cannot be distinguished.

tional to the number of photons but not identical to it. The microscope manufacturer should be able to provide a function that calculates it.

IMPROVING THE RESOLUTION INCREASES NOISE

Improving the resolution means that the observed area or volume is decreased (Stelzer and Haar 1999). If object and object preparation procedures remain the same, this means that the number of observable fluorophores and, in consequence, the number of collectable photons, is decreased. Because 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 PSF. If, for example, the observable volume is decreased by a factor of two, the signal is reduced by at least a factor of two and the observation period has to be increased by a factor of four simply to maintain the S/N ratio.

REFERENCES

Born M. and Wolf E. 1980. Elements of the theory of diffraction. In *Principles of optics*, pp. 370–458. Pergamon Press, Oxford, United Kingdom.

Brakenhoff G.J., Blom P., and 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.

Castleman K.R. 1993. Resolution and sampling requirements for digital image processing,