2

Fundamental Limits in Confocal Microscopy

James B. Pawley

INTRODUCTION

The previous chapter described how the confocal approach developed from conventional light microscopy and outlined the basic advantages gained by the use of confocal sampling techniques, primarily that the exclusion of light by the confocal pinhole makes it possible to record data from optical sections. This chapter will discuss the fundamental considerations that limit the performance of all confocal microscopes. Though at present no commercially available equipment approaches these limits, some simple tests will be described to help the user assess how well a given instrument performs. Additional information to help the user to operate the confocal microscope in an optimal manner can be found in Chapter 35, "A Tutorial on Confocal Microscopy," and Chapter 36, "Practical Confocal Microscopy." These also include methods for measuring resolution and other useful parameters.

What Limits?

The task of the confocal light microscope is to measure optical or fluorescent properties within a number of small, contiguous subvolumes of the specimen (Fig. 2.1; Pawley and Centonze, 1994). The fundamental limits on this process, therefore, are related to the quantitative accuracy with which these measurements can be made, a factor that depends on the number of photons that pass into n_1 and out of, n_2, the sub-volume; its size (δx, δy, δz); and its position (x, y, z). Additional limitations are imposed on the rate at which these measurements can be made by the effects of photodamage to the specimen, finite source brightness, and fluorescence saturation. Finally, limitations are imposed by the fact that the continuous specimen must be measured in terms of discrete volume elements called voxels [a voxel is the three-dimensional (3D) equivalent of a pixel, which is the smallest element of a two-dimensional (2D) digital image]. This chapter will try to define the factors that ultimately limit the accuracy with which these measurements can be made. As such, it will serve as an introduction to many of the chapters that follow in which the practical and theoretical aspects of these problems are discussed in greater detail. The discussion should be applicable to the consideration of any type of confocal microscope, though here, as elsewhere in this volume, microscopes in which scanning is accomplished by moving the light beam(s) rather than the specimen will be emphasized because they are more easily applied to living specimens. Most of the discussion will focus on the performance of the confocal laser scanning microscope (CLSM), but in some cases, differences between the mirror-scanning and disk-scanning instruments (including both tandem and single-sided disks) will dictate a separate consideration.

The data recorded from a confocal microscope will, in the simplest case, be a set of intensity values (usually representing the concentration of fluorophore) for every voxel throughout a 3D volume within the specimen. Though these data may often be displayed as an image, it should always be remembered that the single-beam confocal microscope is intrinsically a serial or sampling instrument, not a parallel-imaging instrument. While it is true that one may choose to sample a plane by sequentially scanning the illumination over a large number of overlapping sub-volumes that cover the focus plane and that by doing so, one may produce a 2D image, given sufficiently flexible equipment, one could also use the same total sampling (imaging) time to measure a single point at a great many different times, a smaller volume within the sample or, indeed, any other small collection of points within the specimen a great many times.

The distinction between sampling and imaging is, of course, not absolute; after all, most of us will view the final result as some sort of image. However, the distinction is still useful because it requires one to explicitly confront many problems that are not always so obvious when microscope images are viewed directly by eye or after photographic recording.

The sampling approach, which is covered in more detail in Chapter 4, allows an image to be built up from a number of individual measurements, each of which reflect properties within a specific region of the sample. If the measured properties are optical, the measurements involve counting photons, and this process itself implies limits on both the data rate and the statistical accuracy that are often ignored in normal widefield (WF) microscopy. These limitations are associated with factors such as counting statistics, source brightness, and specimen response, and are discussed next.

Although the points raised so far constitute a fairly complete list of the **physical limits** on the accuracy and completeness of the data stored in the image memory of a confocal microscope, they say nothing about factors affecting the response of either the dye molecules in the specimen or the response of these molecules to the excitation. A more complete understanding of these and other important variables can be found outlined in Pawley (2000) and in greater detail throughout the other chapters of this book.

Counting Statistics: The Importance of *n*

The accuracy of any particular measurement involving fundamental, quantum interactions (such as counting photons) is limited by Poisson statistics. Without going into the details, this means that if the same measurement is made repeatedly, and the average result of these measurements is n photons/measurement, the chance that any specific measurement is in the range between $\mathbf{n} + \sqrt{\mathbf{n}}$ and n

James B. Pawley • University of Wisconsin, Madison, Wisconsin 53706

 Handbook of Biological Confocal Microscopy, Third Edition, edited by James B. Pawley, Springer Science+Business Media, LLC, New York, 2006.

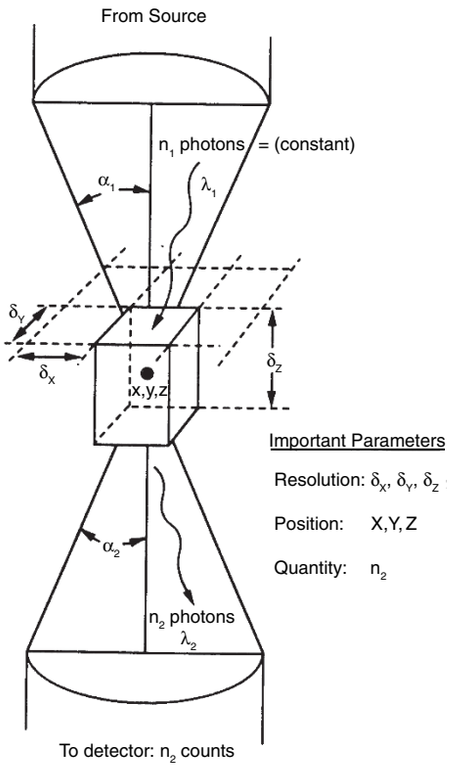

FIGURE 2.1. Parameters related to the fundamental limitations of confocal microscopy.

– $\sqrt{n}$ is only 63%. For example, if n is 100 photons, 63% of the measurements will be in the range from $100 - \sqrt{100}$ to $100 + \sqrt{100}$ or between 90 and 110. Such a measurement is said to have 10% statistics. It can be seen that increasing the precision to 1% requires that n be increased to $(100)^2$ or 10,000 photons. While similar considerations limit the performance of all types of image measurements, they are more explicit in their effect on confocal microscopy where photons are routinely counted individually. It might help to have a better idea about the numbers of photons involved. Figure 2.2 shows how a 1 mW laser beam composed of 10^{15} photon/s becomes a detected signal of only 10–100 photons/pixel.

The uncertainty associated with counting quantum-mechanical events is often spoken of in terms of it being the source of **intrinsic** or **statistical** noise, and this usage is the basis of the common belief that, while a single-scan image is "noisy," the "noise" can be "reduced" by summing or Kalman-averaging the data from many frames. This usage is accurate to the extent that because the summed image contains more data, it is better statistically determined and appears less noisy. However, it is important to keep intrinsic noise separate in one's mind from **extrinsic** noise such as that introduced by detector dark-current or electronic noise, or that produced by stray or out-of-focus light[1] because, unlike intrinsic noise, these factors are susceptible to being reduced by careful technique and technological improvements. Furthermore, in the case of fixed pattern noise, the effect may not be reduced by averaging many frames.

[1] A simple test for stray is to collect a frame with the room lights on but the laser turned off. If the average brightness of this frame or one with the laser on but no specimen in place, is higher than that with the laser and room lights off, you may have a stray light problem.

While it is misleading to think that the **only** way to reduce "noise" is to average more data, it is also true that in a well-designed CLSM, the major noise source is intrinsic noise. This fact highlights the importance of making sure that as many as possible of the available photons are recorded as part of the signal. Photon efficiency is discussed later.

Source Brightness

A fundamental law of optics states that the brightness of light in the image (measured in watts/cm^2/steradian) can never be greater than it was in the object (or source). In the case of laser-scanning microscopes, the intrinsic brightness of the laser source is so high that this law does not present a practical limitation on performance (though photodamage of the specimen may do so). However, it does pose a limitation on the performance of disk-scanning confocal instruments, many of which currently use mercury arc sources that lose 90% to 98% of their intensity in passing through the disk (Chapters 6, 10, *this volume*). In the latter case, the source brightness and the optical design of the illuminating optics are crucial to being able to detect enough photons to produce a statistically well-defined fluorescence image in a reasonable amount of time. At present, the total power of the narrow-band illumination needed to excite fluorescence that emerges from the best disk-scanning instruments is at least an order of magnitude less than that commonly used with the laser instruments. However, improved non-laser sources are constantly being developed (Chapter 6, *this volume*) and, in addition, the laser-powered disk-scanning microscopes are beginning to take over some of the fast-scanning market previously dominated by the arc-illuminated disk scanners (Chapter 10, *this volume*).

Specimen Response: Dye Saturation

In normal, widefield (WF) microscopy, it is safe to assume that photons interact with the specimen in a manner that is independent of the illumination intensity. However, this linear response is not characteristic of laser-based confocal microscopes operated at effective laser power levels of >1 mW. Conceivable departures include the possibility that absorption in the specimen may cause sufficient warming to produce damage or that the electric field strength may become sufficient to produce a nonlinear response (such as 2-photon excitation, Chapters 28 and 37, *this volume*). However, the most obvious complication is the phenomenon of singlet-state fluorescence saturation.

This phenomenon occurs when the flux of exciting illumination is so intense that, at any instant, a significant fraction of the fluorescent molecules in the illuminated area are already in the singlet-excited state. As excited molecules no longer absorb light at the usual wavelength (λ), this has the effect of lowering the effective dye concentration. This saturation threshold can easily be reached at flux levels around 10^6 W/m^2 such as are found in CLSMs (see Chapters 16, 17, and 39, *this volume*). The problem is more severe when using dye molecules with high absorption crossections, or long fluorescent decay times especially when the dye must be used at low concentration.

Effects of Saturation

Because fluorescence saturation is a property of the beam, it affects **all** dye molecules equally, so those in "bright" pixels are affected the same as those in "dim" pixels. As a result, the contrast of the final image seems unaffected: bright areas are bright, dark areas, dark. In other words, an image recorded with a beam intense

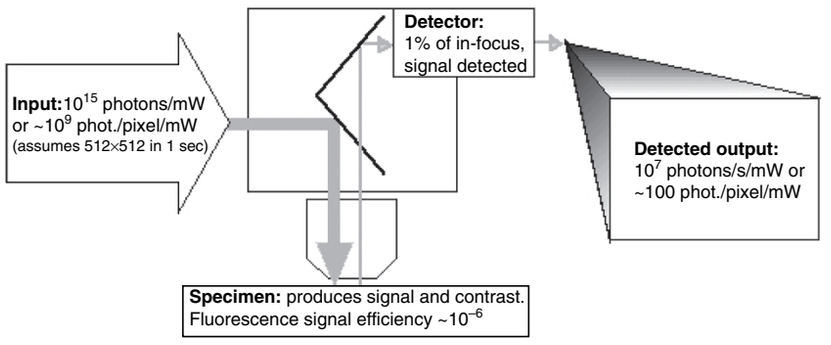

FIGURE 2.2. Schematic diagram showing photon numbers at various stages of fluorescence microscopy.

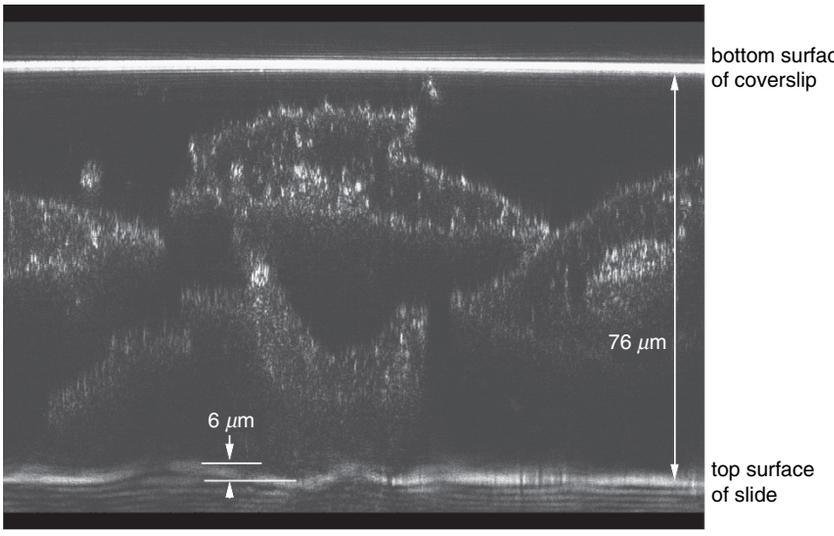

FIGURE 2.3. Confocal XZ, backscattered light image of cheek cells in a chamber made of a coverslip held away from the microscope slide by dots of dried nail polish. The image of the upper glass/water interface is straight and smooth. The lower one is not. This is an indication that the optical properties of the cheek cells have distorted the optical section. Bio-Rad MRC-600 confocal mounted on a Nikon upright microscope with a 60× NA 1.4 Plan-Apo oil-immersion objective.

enough to cause saturation does not look "clipped" in intensity in the same way that an overexposed photograph does. However, the danger of fluorescence saturation is that in its presence, the signal from **any particular** pixel becomes a function of variables other than dye concentration. These variables include the local fluorescent decay time constant, which can depend strongly on the molecular environment, and local defocus effects that may affect the peak intensity in the focused spot and hence the degree of saturation. It follows from this last point that, as saturation occurs mainly at the center of the focused spot, relatively more signal will be produced from **adjacent** planes when one operates near saturation, an effect that marginally reduces *z*-resolution.

Finally, although the mechanisms of photodegradation are as yet imperfectly understood (see Chapters 16 and 39, *this volume*), it now seems likely that the absorption of a second photon by a molecule still in the singlet-excited state may turn out to be a common bleaching mechanism. If high levels of light flux increase the **rate** at which fluorescent dye is bleached (per illuminating photon), operation near saturation may increase the bleach rate.

This fundamental limitation on the rate of fluorescent data acquisition can be side-stepped if the light illuminates more than one focused spot on the specimen. However, to preserve confocal conditions, a separate excitation and detection pinhole must be used for each spot, a condition that is only present in the disk-scanning and line-scanning instruments at this time.

The Effect of Refractile Cellular Structures

As is diagramed in countless textbooks, the pinhole of the confocal optical system excludes light emitted or scattered from features that are above or below the focus plane, creating the "optical section." However, all these diagrams assume that the specimen is optically uniform. While this assumption may be met for fixed specimens that are embedded in a mounting media having a high refractive index (RI), it is less accurate in describing the optics of living cells. The fact that living cells produce high-contrast images when viewed using phase-contrast or differential-interference contrast optics, is a clear indication that the RI of many cellular structures is substantially different from that of water. Figure 2.3 is an *xz*-image of a clump of cheek cells mounted between the coverslip (top) and the slide (bottom) made using **backscattered light** (BSL).[2] As it should be, the flat surface of the coverslip is imaged as a thin horizontal white line. However, because of the presence of the cheek cells, the equally flat surface of the glass slide is rendered as a fuzzy line that undulates up and down by as much as

[2] Backscattered light is a more general term for what is often referred to as "reflected light." It is recommended that the term "reflected" be reserved to describe light scattered by features sufficiently large and smooth so that "the angle of incidence equals the angle of reflection." Most cells produce much more backscattered light than reflected light.

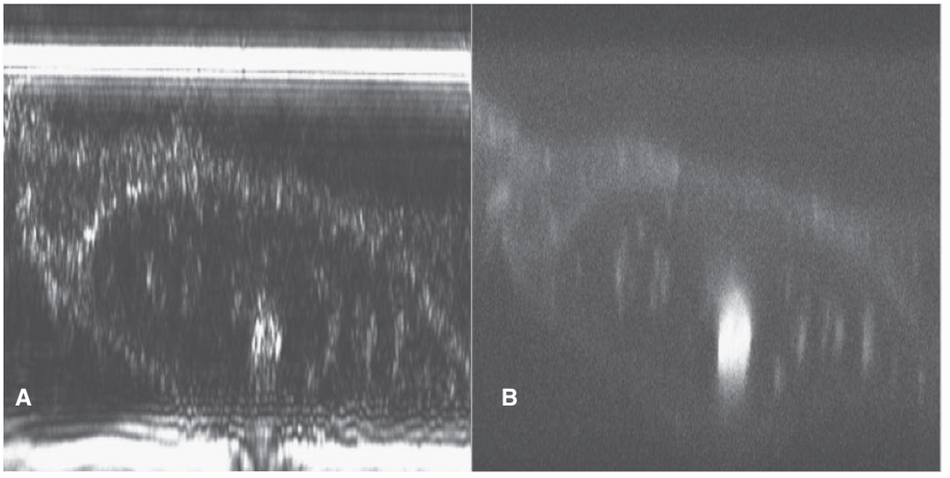

FIGURE 2.4. (A) Confocal XZ, backscattered light image of cheek cells in a chamber made of a coverslip held away from the microscope slide by dots of dried nail polish. (B) Acridine orange fluorescence image of the same specimen as in (A). It is easy to see how the presence of the nucleus (the large blob in the fluorescence image) has displaced the BSL image of the top surface of the slide. Optical system the same as in Figure 2.3. Field width, 50 μm; field height, ~60 μm.

6 μm. Figures 2.4 and 2.5 make clear the role that the nuclei in some cells can play in distorting the "plane of focus." In this case, the "surface of best focus" is clearly not a plane (Pawley, 2002).

The idea that the RI anomalies caused by common cellular structures, such as nuclei, are sufficiently severe to displace the location of any optical section on their far side by distances on the order of micrometers is deeply disturbing. As living cells necessarily have nuclei and other refractive organelles, it is hard to see how, or even if, we can compensate for the sequellae of their optical properties. Although one might imagine computer calculations that would model these optical effects and thereby improve matters, such calculations will be immensely complex and therefore take time: time seldom available when viewing living cells. In general, it seems that we must accustom ourselves to more modest performance when viewing living specimens in a 3D light microscope.

Given the distortion noted above, it seems inevitable that these RI anomalies must also deform the point-spread-function (PSF) of the microscope in nearby regions. This raises questions about the confidence one can have in images produced by widefield/image-deconvolution of living specimens. It is not that deconvolution will not produce a sharper image. The problem is, does this sharper image accurately represent the 3D structure of the specimen?

In some tissues, cellular structures are naturally less optically disruptive: cornea and lens come to mind. For other living tissues, it seems that one must be mindful that the visibility of peri-nuclear structures may depend not only on their presence or absence but also on whether these structures are located on the near or far side of the nucleus. When they are located on the far side, mounting the specimen between two coverslips so that it can be flipped over may be worthwhile (see Chapter 15, *this volume*).

When looking far below the surface of a tissue, it may be worth the effort to try to orient the specimen so that the foreground contains as few nuclei as possible.

A first approximation of how serious this problem is on any particular specimen can be obtained by viewing the specimen using phase-contrast or darkfield optics. Contrast in both of these imaging modes is proportional to changes in local RI.

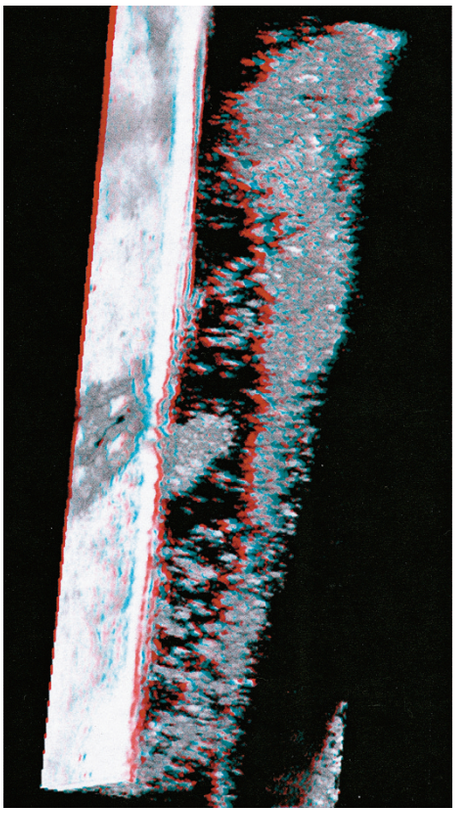

FIGURE 2.5. Stereo view from part of a z-stack of 128, 512 × 512 BSL images of a cheek cell specimen. The projection orientation was chosen to emphasize the spatial relationship between the nucleus (central blob) and the distorted area of the image of the water/slide interface originally located below it. The four-way symmetry of the "distortion" image is caused by the interaction of the nucleus and the polarized laser beam and modified by the fact that high-NA light is more likely to reflect from the slide interface. Bio-Rad Radiance confocal mounted on a Zeiss Axiophot upright microscope with a 40× NA 1.2 C-PlanApo water-immersion objective. To obtain stereo effect, view through red/blue or red/green glasses.

A Typical Problem

The many limits on CLSM performance do not act alone but can combine in complex and subtle ways. To highlight the interactions between counting statistics and more widely recognized limits such as spatial resolution, let us define a characteristic microscopical problem shown schematically in Figure 2.6. The specimen is a cell in which some of the protein sub-units making up the cytoskeletal fibers have been replaced with fluorescent analogs. These fibers are very small, about 10 times smaller than the *x,y*-resolution limit of light optics, and excitation of the fluorescent dye causes it to bleach, a process that is probably toxic to the cell. Because the object is to observe the formation and movement of the linear cytoskeletal elements within the living cell, one cannot take advantage of the higher spatial resolution of the electron microscope because this instrument cannot be used on living cells.

This example, though perhaps overly specific, is not uncharacteristic, and it has the advantage of highlighting interactions between many of the fundamental limitations. The important features of this example are as follows:

- The observer would benefit from high spatial resolution in all three dimensions.
- Accuracy in measuring intensity-of-stain/unit-length will be important, because doing so may permit determination of the number of sub-resolution, linear polymers bundled together to make up each visible fiber.
- A number of images must be recorded at different times in order to show change/motion.
- Each measurement will cause bleaching and cellular toxicity.

Clearly, these conditions contain an inherent contradiction: to obtain more quantitative temporal or spatial accuracy, more light must pass through the sample but this will produce more fading and cytotoxicity. As a result, the improved images may be of a dying cell rather than a living one, and, therefore, the **biological reliability** of the measurement may be inversely proportional to the **physical accuracy** set by counting statistics.

Other interactions between these parameters are also possible. High spatial resolution implies that the micro-volumes (pixels) excited by the beam and sampled by the detector must be very close together (actually the spacing between lines is less than one quarter the diameter of the Airy disk). The measurement at each pixel is really the detection of the small fraction (<1%) of the fluorescent excitations excited by the beam that actually produces signal in the detector while this pixel is illuminated.

Higher spatial resolution implies smaller pixels and, hence, the need to count more photons from any given volume of the specimen. The increases in detected signal implied by improved resolution are not insignificant: maintaining the statistical accuracy of the measurements when the resolution increases by a factor of 2 requires four times more signal in order to image a single plane and eight times more if a 3D volume is to be sampled. Fortunately, the only way to actually increase the resolution by 2× is to increase the numerical aperture (NA) by 2×, and doing so increases the fraction of the photons emitted from the specimen that are captured by the lens by a factor of $(\Delta NA)^2$ or 4×. As a result, some of the improved image quality can sometimes be retained by collecting a larger fraction of the emitted light rather than simply by exposing the specimen to increased illumination.[3]

This interaction emphasizes the importance of keeping the pixel size appropriate to the operating resolution. Those who usually use WF microscopes to record images on film may be less familiar with the idea that pixel size is an explicit experimental variable.[4] It is important to understand that the "zoom" magnification factor used to change the size of the area scanned on the specimen on most commercial CLSMs usually does so by changing the pixel size (referred to the specimen). Therefore, although the ability to vary the display magnification by a factor of about 10 : 1 may seem to be great convenience, **with a given λ and NA, only one zoom setting provides optimal Nyquist sampling** of the optical image data. All other zoom settings must necessarily either over- or under-sample the data and either produce more beam damage or record less resolution than they should. (See Chapter 4, *this volume*, for how to choose the optimal zoom setting.)

Given the interrelated constraints highlighted by these examples, the two features needed for a confocal microscope to approach its ultimate performance are:

- *Photon Efficiency:* The system must count as many as possible of the photons transmitted or emitted by the sample at the plane of focus.
- *Spatial and Temporal Resolution:* Generally one should focus the laser into a spot that is as small as possible though, on occasion, it may be advantageous to make this volume somewhat larger to avoid saturation or photodamage effects.

These two topics will now be discussed in more detail.

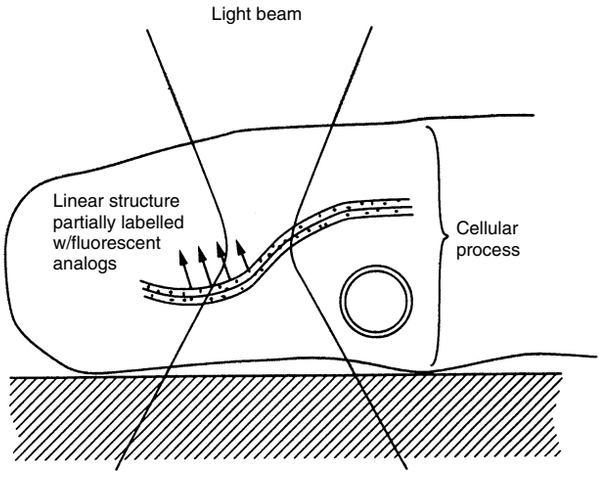

FIGURE 2.6. Diagram of a notional specimen: a living cell process containing a filamentous structure smaller than the resolution limit and sparsely stained with fluorescent analog molecules.

PRACTICAL PHOTON EFFICIENCY

Photon efficiency (γ) is a measure of the fraction of the signal photons generated by the action of the laser beam on the specimen that are actually represented in the final image data set stored in the computer. Although photons can be lost both between the light source and the specimen and between the specimen and the detector, those lost before reaching the sample can usually be "replaced"

[3] Although this reciprocity works in the *xy*-plane, it doesn't compensate in the *z*-direction because *z*-resolution is proportional to $1/(NA)^2$.

[4] These problems are strictly analogous if we relate pixel size to the size of the film grain. To record faint fluorescence, we choose a "fast" grainy film with a high ISO rating only to discover that it does not record details as small as those visible when using a film with a lower ISO rating.

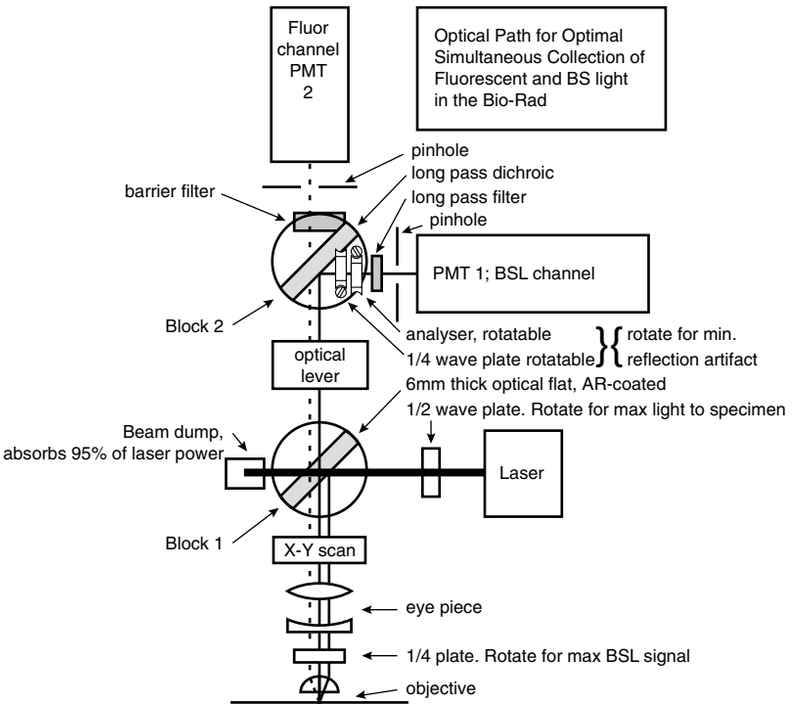

FIGURE 2.7. A system of two filter blocks for use with the Bio-Rad MRC-500–1024 which permits optimal removal of specular reflection artifacts and the simultaneous collection of BSL and fluorescence signals. Block I contains a beam-splitter consisting of a 6-mm-thick Pyrex optical flat, anti-reflection coated on its rear (upper) side. A 1/2-wave plate after the laser is used to rotate the laser polarization so that 0.8% to 5% of the laser light is reflected down the optic axis and through a 1/4-wave plate between the ocular and the objective. Almost all of the returning BSL and fluorescent light passes through the beam-splitter to Block II, where the two signals are separated with a dichroic mirror. The long λ signal passes to PMT 1 via an emission filter. Signal at the excitation wavelength passes through a second 1/4-wave plate and analyzer (both rotatable) and a laser line filter to PMT 2. The orientation of the 1/2-wave plate is adjusted to give maximum light from the objective, then the rotatable elements in Block II are adjusted to eliminate spurious reflections. Finally, the 1/4-wave plate in the microscope body is rotated to produce the maximum BSL signal from a living cell.

with relative ease by an increase in laser power.[5] However, photons lost after leaving the sample represent a more fundamental loss: they carry information obtained at the expense of radiation damage to the specimen.

Current instruments waste photons in a number of straightforward ways, some of which are discussed later in this section. Two somewhat less obvious cases of photon inefficiency that can be grouped under the heading "wasted light" are worthy of mention here.

Whenever the laser beam passes from a region having one RI to another region having a different RI, some light is scattered back towards the source. As a result, almost any biological specimen, particularly a living one, produces BSL. In most current instruments used in the fluorescent mode, the BSL is reflected back to the source by the dichroic beam-splitter and thereby wasted. As is also pointed out in Chapter 9, it is possible to design the confocal microscope in such a way that both the BSL and the fluorescent signal can be detected separately and in a totally non-interfering manner to produce a BSL image that is perfectly aligned with the fluorescent image (Pawley *et al.*, 1993a, Pawley, 2002). This image provides information on the optical inhomogeneities in the cell at no additional cost in terms of radiation damage to the specimen (Fig. 2.7).

The second example of wasted light is that which is elicited by the laser during line retrace in raster-scanning instruments. As photons excited from the specimen by this light are not detected, it represents a significant (30% for 1s scans, 10% for 4s scans) and unnecessary assault on the specimen. Fortunately, it can be eliminated by gating the light source during retrace using a Pockels Cell or the acousto-optical deflector (AOD), now available on most commercial instruments. Early video-rate scanners using resonant galvanometer scanning employed a system of blades to mask the sides of the raster so that light did not reach the specimen as the beam slowed down before turning around at either end of the sinusoidal motion.

Light generated inside the specimen can be lost through several mechanisms:

1. Absorption or scattering in either the objective lens or the medium that couples it to the specimen, or by the fixed and/or moving mirrors and transfer optics needed to "de-scan" the beam (see also Chapters 7 and 9, *this volume*).
2. Incorrect alignment of the optical system resulting in the improper placement or orientation of the pinhole or slit (see also Chapters 3, 8, 11, 22, 35, and 36, *this volume*).
3. Low quantum efficiency (QE) of the photon detector (see also Chapter 12, *this volume*).
4. Imprecise digitization of the output of the photon detector (see also Chapter 12, *this volume*).

While these subjects will be covered in much more detail in the chapters noted above, a few points will be made here as a background to descriptions of the practical tests of performance to be described later.

Losses in the Optical System

Objectives

To a first approximation, fluorescent light proceeds from the site of its generation in all directions equally. As a result, only a small fraction even strikes the objective lens: ~30% of the total for NA 1.4 and this fraction decreases with $1/(NA)^2$. The fraction of this light that emerges from the other side of the objective depends on its transmittance. Although measuring the absolute transmittance of an objective lens at full NA is far from easy (see Chapter 7, *this volume*, for details), useful measurements comparing the effective

[5] As noted above, such an increase is more difficult when using the arc sources commonly employed in the disk-scanning approach.

transmission of individual objectives can be made on epi-illuminated instruments with significantly less trouble. All that is needed is a photodiode light sensor (one **without** an air space between it and any clear glass "window") linked to a sensitive current meter (or better, a basic photometer) and a front-surfaced mirror. After adjusting the instrument to measure BSL and fitting objective lens A, align the instrument if necessary, then switch to a high zoom setting and measure the light emerging **from** the objective with the sensor (be sure to exclude stray light from the room and couple the lens to the sensor with immersion oil if it is an oil-immersion lens). Call this light reading I_a. Now place a front-surfaced mirror, slightly tilted, on the specimen stage and set up the microscope to image the reflecting surface using the BSL signal. Be careful to keep the illumination level very low so as not to damage the photomultiplier tube (PMT) in the microscope (and to realign the instrument if reducing the illumination requires the addition of ND filters in front of the laser, see Chapter 35, *this volume*). Focus on some dust on the mirror. Because of the slight tilt, the surface will only be near focus (bright) along a broad band. Use the computer controls to measure the peak brightness (B_a) of this band and that of some dark area (b_a) well away from the bright band (or better yet, with the laser obscured). This second, background reading is to compensate for non-specific reflections in the optical system, poor DC balance in the PMT head amplifier, and the Brightness or Black-level offset setting, etc. The reading should be slightly positive to ensure that "zero signal" is within the linear range of the digitizer.

Some microscopes produce severe specular reflection artifacts when used in the BSL mode: make measurements in a part of the image field unaffected by these reflections. Also be sure to adjust the PMT gain so that at a single setting, the signal remains well within the linear region of the digitizing system (about "three-quarters-full" or 192 counts/pixel in an 8-bit system) throughout the entire procedure.

Now change lenses, realign, and make a second set of measurements I_b, B_b, and b_b without changing the PMT gain.

To a reasonable approximation, the comparative transmission (T_{comp}) of the first lens as a fraction of the second will be the following:

$$T_{comp} = \frac{I_b(B_a - b_a)}{I_a(B_b - b_b)}100\% \quad \textbf{(1)}$$

(Transmission specifications for a number of common objectives can be found in tables in Chapter 7, *this volume*.)

Mirrors

On some scanners, a similar setup can be used to test the performance of the internal mirrors. Again, the light intensity leaving the objective, I_a, is measured. Then the beam is stopped (or made to scan a very small raster) at the part of the scan when it is focused on the mirror surface (brightest). After turning off the PMT(!), a second reading is made (P_a) with the same photodiode placed just in front of the pinhole, making sure that **all** of the light strikes the sensitive part of the sensor.

The P_a/I_a ratio will usually be depressingly small (~10%–20%), but the difference covers reflection losses at the various mirror and filter surfaces, including the beam-splitter (50%–60% loss on each pass) as well as those at the scan lens, the various transfer lenses and the objective. (A method for making a similar measurement that does not require placing the photodiode in front of the pinhole is given in the section "Measuring Photon Efficiency.")

Though indirect and somewhat cumbersome, such measurements can be useful for two reasons: (1) as a rough method of comparing the performance of different instruments (or different types of mirrors fitted to the same type of instrument) and (2) to monitor performance of a specific instrument over time. Performance degrades as dust and/or hydrocarbon or other atmospheric vapors condense onto the surfaces of mirrors and other optical components. In instruments having up to 11 reflective surfaces in the detector chain, a change in reflectance of even 1% can have marked effects on photon efficiency (see Chapter 7, *this volume*).

Pinhole

Is the Confocal Pinhole a "Good Thing"?

It is argued by some that the pinhole is the Achilles' heel of the confocal approach because it excludes photons that, while not originating from the plane of focus, do still carry relevant information regarding other planes because of the non-point nature of the focused optical probe. The process of "deconvolution" will "put the photons back where they came from" and all will be well (see Chapters 23, 24, and 25, *this volume*). Elaborate and authoritative mathematical analyses seem to support an argument that can be summarized, not too unfairly, as suggesting that, "as all measured photons are good, excluding photons must be bad." The problem is that the mathematics that describe so well the actions of light waves, generally has more trouble accounting for the behavior of photons, or more particularly of the errors in counting them caused by Poisson noise.

Leaving aside possible modifications to the confocal detector (noted in the next segment) that would allow photons from nearby planes-of-focus to be separately collected in a confocal instrument, the contention that light from out-of-focus planes contains information deserves some consideration. In the discussion that follows, all non-confocal (WF) images will be composed of light from both in-focus and out-of-focus planes while confocal images consist of light from **only** in-focus planes. The argument turns on the relative quantum efficiency (QE) of the detectors normally used with each technique [cooled charged-coupled device (CCD) vs. PMT] and on when and to what extent the out-of-focus light can yield information regarding a 3D structure **beyond t**hat contained in the in-focus light. This matter is covered in more detail in the appendix to Chapter 13, and also in Chapters 22, 23, and 24 and in Pawley (1994).

Let us take some examples. All will agree that an out-of-focus WF image of an isolated point object will be a larger and less intense image than an in-focus one (Fig. 2.8) and, furthermore, that a knowledge of the optical-transfer-function of the system, together with information from other adjacent planes as to whether the current plane of focus is above or below the plane containing the point object, will allow us to use computer image-processing techniques to gain some information from this out-of-focus data. Supposing then that only out-of-focus images were available, one would be justified in saying that, from an analysis of out-of-focus data, information had been gained as to the actual location of the object (i.e., in-focus information). But beyond that, would one be justified in saying that this would be a significant addition to the information present in an image of the plane containing the object (the in-focus image)? In other words, is the out-of-focus information "useful" only because in-focus information is lacking?

Another aspect of a measurement in which WF techniques will give a more accurate estimate is, for instance, a measurement of the total fluorescence from an entire specimen (say, all of the chromosomes in a nucleus). For a given dose to the specimen, the WF

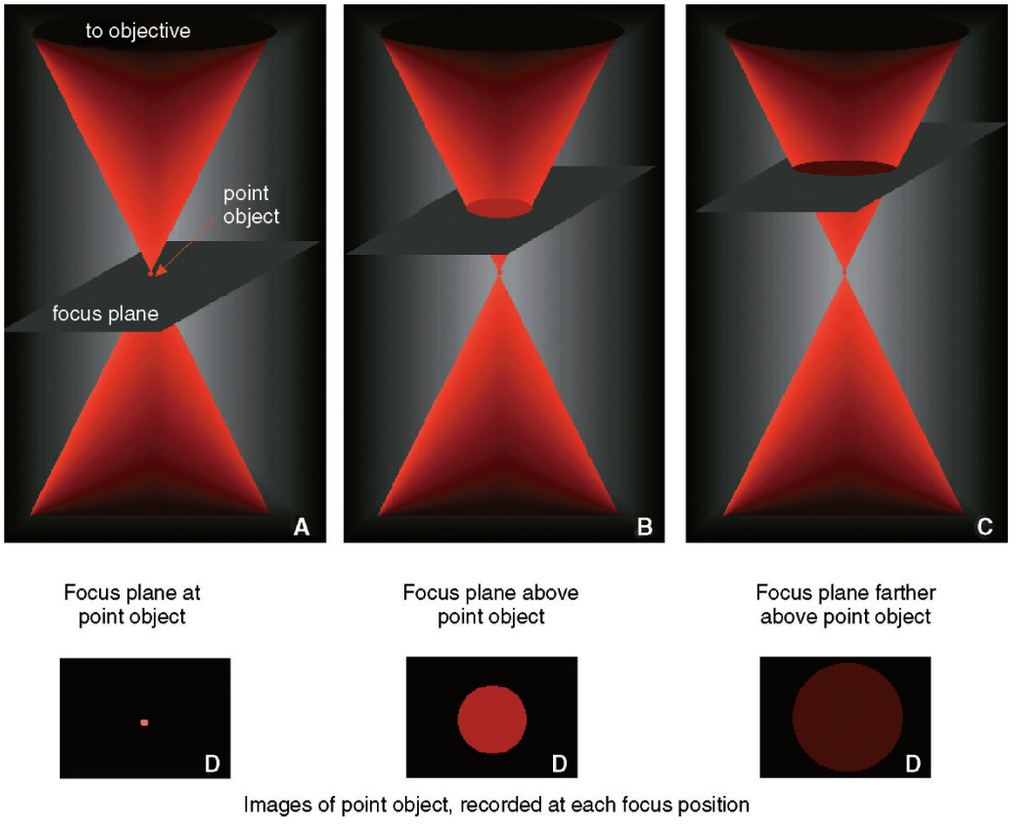

FIGURE 2.8. Image geometry. A schematic diagram showing how the cone of light leaving a point object at angles that will be accepted by the objective, looking at different focus planes.

technique can be more accurate simply because it will detect more total photons.

On the other hand, there must be some limit to the extent that light from sources remote from the in-focus plane can provide more information about that plane than it obscures (Chapter 10, *this volume*, Appendix). For instance, suppose the specimen consists of a small, 0.2-µm feature in the plane of focus and a larger and 10× brighter feature 1 µm above it. Let us also suppose that the point-spread function (PSF) of the optical system is such that, in the WF image focused on the plane containing the small object, 75% of the light reaching the pixels centered on this feature originates from the larger, brighter out-of-focus feature. As this extra signal will double the statistical noise associated with intensities measured in these pixels, it is hard to see how the WF measurement could be better than a confocal measurement that senses the in-focus feature but excludes most of the out-of-focus light before it reaches the detector.

As a more complex example, consider a BSL image of a tooth in which virtually every voxel scatters light. In such a thick "dense" specimen, the fraction of the WF signal originating from the in-focus plane is so small and so uniform that even detecting it using WF imaging and signal processing would be extremely difficult, and the suggestion that the out-of-focus signal actually contributes information **beyond** that available in the in-focus or confocal image seems difficult to support.

Somewhere in between these extremes might be the more realistic situations defined by the two, vertical "sight lines" labeled A and B, in Figure 2.9.

The question then becomes: Up to what level of 3D staining density (or sparsity) can **information** about the specimen (as distinct from **signal** defined as "output for the detector") be gained from out-of-focus light? The answer will not be a simple one, as it depends on a number of practical factors apart from the sparsity of the staining, such as (1) how well one knows the optical response function of the microscope and how it changes with z or, (2) the relative amounts of specific and non-specific staining and the exact effects of applying nonlinear image processing algorithms such as non-negativity or "maximum entropy" (see Chapters 23 and 24, particularly the Appendix, *this volume*).

Excellent results from the WF/image-processing approach are also shown in these chapters and elsewhere, but, because of the difficulty of producing fluorescent specimens that have an interesting 3D geometry but which do not bleach when they are looked at using first one method and then the other, no meaningful side-by-side comparison has yet been performed although there have been some computer simulations. More sophisticated comparisons are described in (Chapters 13 and 24, *this volume*), but it becomes clear that, to be both realistic and useful, any comparisons are likely to be complex as they must include differences in the QEs of the respective detectors (3–10× higher for WF) as well as the details of the optics, the geometry of the stained features in the specimen, and differences in the response of the dye to the much more intense light that must be used in any scanning technique.

The important parameters characterizing the three most popular methods of obtaining 3D images from fluorescent bio-

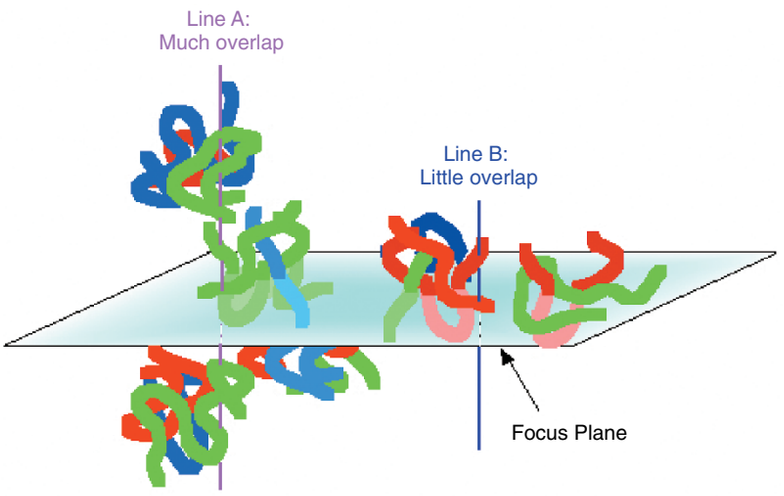

FIGURE 2.9. Schematic diagram showing areas of an imaginary fluorescent structure to illustrate how local variations in stain density change the levels of out-of-focus light detected.

logical specimens are summarized and compared in Chapter 35, Table 35.1 and 35.2.

Proper Use of the Confocal Pinhole

In a confocal microscope, the pinhole is present to prevent light originating from anywhere but the plane of focus from reaching the detector. It is mounted in an image plane, and, if it is misaligned or if its size is reduced beyond that corresponding to a diffraction-limited spot on the sample (i.e., a pinhole diameter of 0.7–1.5 Airy units), then it will severely reduce the number of photons reaching the detector while producing only a marginal improvement in *xy*- or *z*-resolution. Making the pinhole larger than the diameter of the diffraction-limited spot allows more photons to be detected but, as ~80% of those originating **from** the plane of focus were already being collected, **most of the additional signal comes from adjacent planes**,[6] reducing the *z*-resolution. Choice of the proper size and shape for the pinhole is a sensitive function of λ, and the objective lens NA and magnification. However, even an "optimum" 1 Airy aperture will exclude at least some photons (the ~20% present in the outer rings of the Airy disk), and this represents a fundamental cost of using the confocal pinhole.

Given these uncertainties, it seems unlikely that little, if any, additional, useful information will be gained from the out-of-focus light as long as:

- The "bright" voxels are stained at least 10× more intensely than background voxels and are each capable of producing at least **20 detectable photons/pixel** when present in the focus plane of a confocal instrument.
- The confocal detector pinhole is kept at 1 to 2 Airy units, the size needed to demonstrate a lateral resolution equal to that of a WF instrument with the same NA, λ, etc.
- Both instruments are designed and adjusted in accordance with the sampling and photon-efficiency criteria proposed in this book.
- The advantage of the confocal approach will increase as the total number of planes sampled (or imaged) is reduced.

[6] Assuming that fluorescent dye is located there.

Detection and Measurement Losses

The Detector

The detector characteristics of importance to 3D microscopy have been reviewed by several authors (Pawley, 1994; Sheppard *et al.*, 1992; Sandison *et al.*, 1994). The characteristics of most importance to photon efficiency are:

- *Effective quantum efficiency* (QE_{eff})*:* The proportion of the photons arriving at the detector that actually contribute to an output signal that is linearly proportional to the photon input. (QE is often a strong function of the λ of the detected photons, Fig. 2.10.)
- *Noise level:* This includes both additive noise, in the form of PMT or CCD dark current noise or electronic amplifier noise, and multiplicative noise in the form of random variations in size of the actual PMT output pulses derived from identical input photons.

The Photomultiplier Tube

Although the PMT is the most common detector used in the CLSM, it is not necessarily the ideal detector. While the raw QE of the most modern PMTs may be as high as 30%, this still means that 70% of the photons produce no signal. At 565 nm, only about 15% photons striking the outer surface of a selected, end-window PMT with an S-20 multi-alkali photocathode are absorbed in the thin photocathode (PC) layer evaporated onto the inner surface of the window (Fig. 2.10). The remaining 85% are either transmitted or reflected. As only those photons that are absorbed can produce photoelectrons (PE) for subsequent amplification in the multiplier section of the PMT, it is clear that any mechanism that reduces transmission or reflection losses may substantially improve PMT QE. This can be done by introducing the light into the end window of the PMT at an angle such that it is totally internally reflected by the PC (Gunter *et al.*, 1970). The light not absorbed in the PC then totally reflects off the outer surface of the window and strikes the PC again. This process continues until the light is either absorbed or it reaches the far side of the PC. Using such an optical enhancer, we have reported an **increase** in QE of 60% at 520 nm, 180% at 690 nm, and 220% at 820 nm (Pawley *et al.*, 1993).

Transmission losses can also be decreased by employing a new GaAs and GaAsP photocathode or using an opaque photocathode,

as is found in side-window PMTs (Fig. 2.10). These have markedly higher QE, especially in the red. On the other hand, it is important to remember that published QE curves tend to represent the very best performance measured and, especially on end-window tubes, there is lot of tube-to-tube variation. In addition, published QE curves only show the fraction of photons producing photoelectrons, not the fraction that actually reach the first dynode and start a cascade.

Though a significant achievement, improved PC performance in the red is usually accompanied by significantly higher dark-count rates. Although the dark current of a PMT is usually low (0.1 count/pixel), it is a strong function of the temperature (Fig. 2.11). In the heated confines of some commercial instruments, it may not always be small compared to the signal level of a weakly fluorescent sample, and this is especially true if excitation levels have been reduced to permit photon-counting without pulse-pileup. The fact that dark count rate is proportional to PC size explains the trend to PMTs with 10-mm PC diameters rather than the 25-mm tubes previously common. Dark current can also be reduced by cooling the PC.

Of more concern is the fact that PMT output pulses produced by single PEs vary in size by an order of magnitude because of statistical variations in the small number of particles present in the early stages of the electron multiplier [Fig. 2.12(A)]. The distribution in the height of these single PE pulses is shown in Figure 2.12(B) for several representative types of PMTs. Clearly, if one merely sums the output pulses, some electrons will count more than others. This produces multiplicative noise and the value sensed by integrating the pulses and recording the sum in the computer will have more uncertainty than had the photon pulses been counted directly.

As most of the scatter in the pulse-height distribution reflects Poisson noise applied to the small number of secondary electrons generated at the first dynode, it is important to keep this number as high as possible. Figure 2.13 shows that first-dynode gain can be as high as 20 on average but only if a relatively high voltage (600 V) is placed on the first stage of the voltage divider. Current commercial confocals are more likely to use less than half this voltage, leading to a likely gain of 10. Clearly, this will introduce an additional uncertainty in the process of trying to determine how many photons were detected in each pixel. Thirty percent multiplicative noise can only be overcome by collecting about 60% more signal.[7] In this way multiplicative noise reduces the effective QE well below that of the published graphs. Indeed, when one includes the "optimism" that often characterizes photocathode QE curves, the signal lost because some photoelectrons fail to propagate at the first dynode, and the effect of multiplicative noise, the **effective QE** of the entire PMT is usually less than 50% of what one would have predicted from the published curves alone.

There are some reasons for hope. Hybrid PMTs amplify photoelectrons by accelerating them directly onto a silicon photodiode [Fig. 2.14(A)]. By using an acceleration of tens of kilovolts, the electrons arrive at the diode with enough energy to make thousands of electron-hole (e-h) pairs. This extremely high "first-stage gain" substantially reduces multiplicative noise, as can be seen in Figure 2.14(B), which shows the pulse-height distribution from a hybrid PMT exposed to about 6 ± 2.5 photoelectrons. Because every photoelectron produces a nearly identical number of e-h pairs in the silicon, pixels in which, for instance, eight arrive, produce integrated output pulses that are clearly different in size from those made by either seven or nine photoelectrons. This pro-

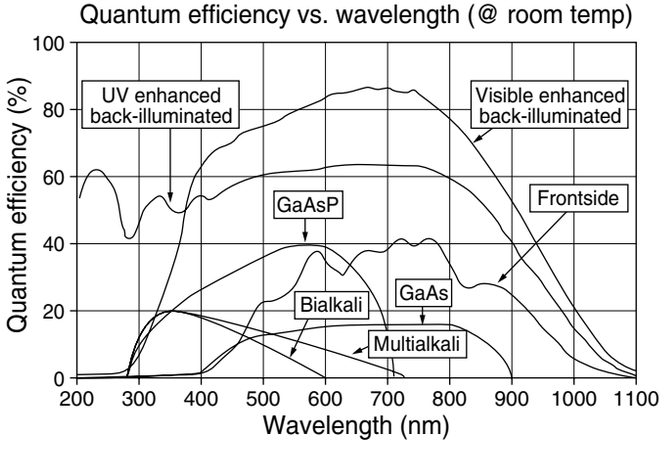

FIGURE 2.10. Variation of quantum efficiency with wavelength of some representative PMT photocathode materials and different types of CCDs. Unless they are thinned and rear-illuminated, the performance of the Si sensors in a CCD is substantially (50%) less in the green and blue because of absorption and scattering in the overlying polysilicon charge-transfer electrodes. Curves for the bialkyli, multialkyli, GaAs, and GaAsP photocathodes are "raw" figures of the photolelectrons leaving the photocathode, and do not take into account the ~30% of these that fail to multiply at the first dynode.

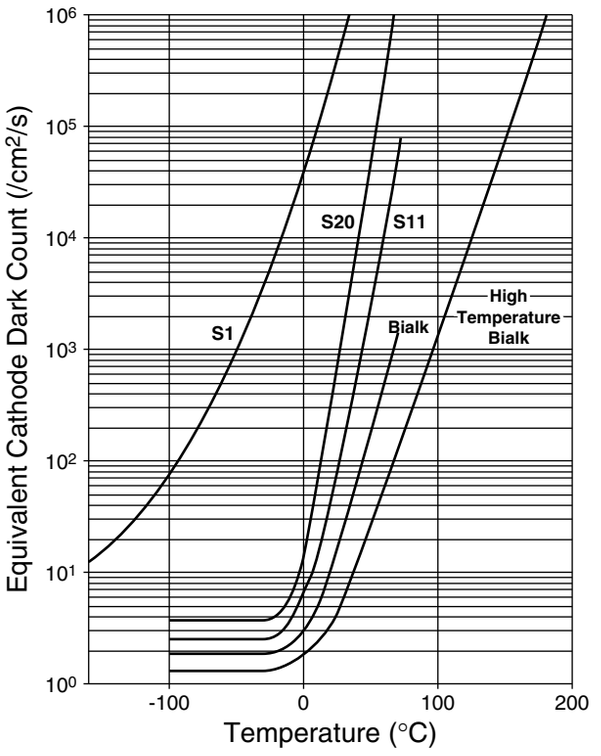

FIGURE 2.11. Variation of PMT dark current with photocathode temperature for a number of photocathode materials. The S1 curve represents performance of the only classical photocathode material with significant response in the near-IR. As the modern replacement of S1, GaAs, also has high dark current and is often used with a cooled photocathode. Please note that the vertical axis is logarithmic.

[7] Or by photon counting, an ability lost to the confocal world when Bio-Rad ceased making confocal microscopes.

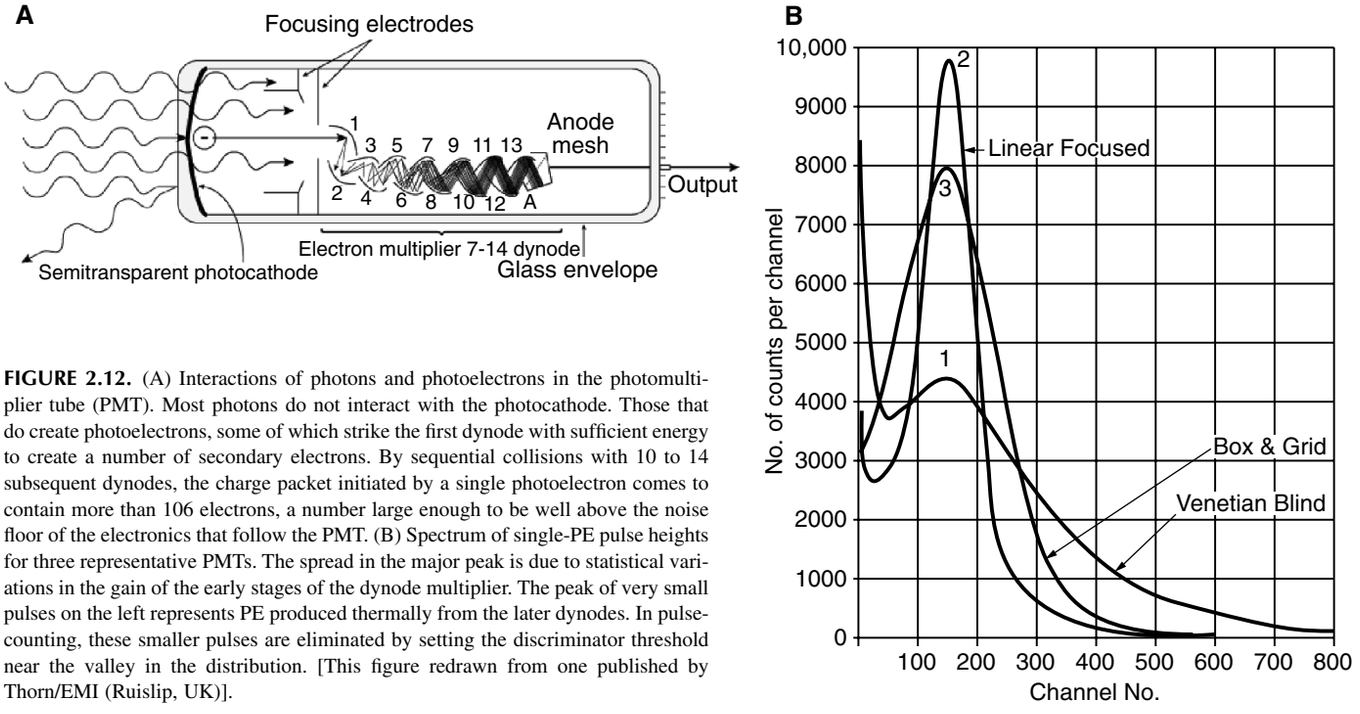

FIGURE 2.12. (A) Interactions of photons and photoelectrons in the photomultiplier tube (PMT). Most photons do not interact with the photocathode. Those that do create photoelectrons, some of which strike the first dynode with sufficient energy to create a number of secondary electrons. By sequential collisions with 10 to 14 subsequent dynodes, the charge packet initiated by a single photoelectron comes to contain more than 106 electrons, a number large enough to be well above the noise floor of the electronics that follow the PMT. (B) Spectrum of single-PE pulse heights for three representative PMTs. The spread in the major peak is due to statistical variations in the gain of the early stages of the dynode multiplier. The peak of very small pulses on the left represents PE produced thermally from the later dynodes. In pulse-counting, these smaller pulses are eliminated by setting the discriminator threshold near the valley in the distribution. [This figure redrawn from one published by Thorn/EMI (Ruislip, UK)].

vides the advantages of photon counting without the danger of saturation due to pulse pileup. The problem at present is high cost, relatively short lifetime and sensitivity to damage from overexposure.

Finally, as the PMT is a single-channel device, it can provide a measure of only those photons that pass through the pinhole mask. This makes the proper setting of this parameter very important. One cannot decide later that it should have been bigger or smaller.

The Cooled Charge-Coupled Device

The only practical alternative to the PMT is the solid-state detector of which the cooled charge-coupled device (CCD) is the optimal example. This detector has both different capabilities and different problems (see Chapter 12 and Appendix 3, *this volume*). The QE can be very high (70%–80%) and extends well into the infrared (Fig. 2.10). Furthermore, as each photon is recorded as an identical amount of current, there is no multiplicative noise. Unfortunately, to keep the noise level acceptably low (±5 photoelectron/measurement), this detector must be cooled to –40 to 80°C and read out at the relatively low rate of 100,000 pixels/s (vs. 400,000 pixels/s for a normal CLSM). This noise level is clearly too high if the **peak** signal level is only 10 photons/pixel, as it can be on many CLSMs.[8] It is less serious when the signal from the **darkest** pixel is >25 photoelectrons because then statistical variations in this number (i.e., $\sqrt{25} = 5$) are similar in size to the measurement noise. These features make the cooled CCD detector more suitable for slowly scanned images (10–100 s/frame) producing relatively high signal levels. (See discussion of this subject in Chapters 10, 12, and 23, *this volume*.)

In the disk-scanning and line-scanning confocal microscopes, the image data emerges as a real image rather than as a time sequence of intensity values from a single detector. Although this real image can be detected photographically or by eye, both of these sensors have fairly low QE. However, these confocal microscopes can approach the photon efficiency of the CLSM if they incorporate a cooled CCD sensor having detection performance similar to that described in the last paragraph. This combination is now implemented in several commercial instruments.

Of even more importance is the recent introduction of the electron multiplier CCD (EM-CCD). In the EM-CCD, a gain register is introduced between the end of the horizontal register and the read amplifier. The gain register consists of ~500 charge-coupled elements in which one of the three charge transfer voltages is much higher than normal. When exposed to this higher voltage, each electron in the charge pocket has about a 1% chance[9] of multiplying by collision with covalent electrons from the silicon lattice. This process produces two electrons and when repeated ~500 times, the average gain can be over 1000×, which is sufficient to lift the signal from even one signal electron well above the read noise of the output field effect transistor (FET) amplifier.

This produces an image detector with the QE of silicon and the noise characteristics of a PMT. There is a snag, of course: the random nature of the electron multiplication process produces

[8] Many users may find it hard to accept that the **brightest pixel** in their confocal images represents a signal of only 8–12 photons. Indeed, they may point to a histogram of such an image and note that there are some pixels present with every possible value from 0–255. Surely, if there are 256 possible values, the brightest must correspond to 255 photons. In fact, the relationship between photon signal level and digital value stored in the image memory is arbitrarily determined by the particular PMT gain and black-level setting chosen. The reason that a signal representing at most 8 photons is not displayed as a posterized image containing only 8 possible gray levels is that multiplicative **noise** in the PMT blends these 8 levels into each other so that the transitions are not apparent. Quite simply, some photoelectrons are counted as "more equal than others."

[9] The useful range is 0.5% to 1.5% and depends on how high the voltage on the transfer electrode is.

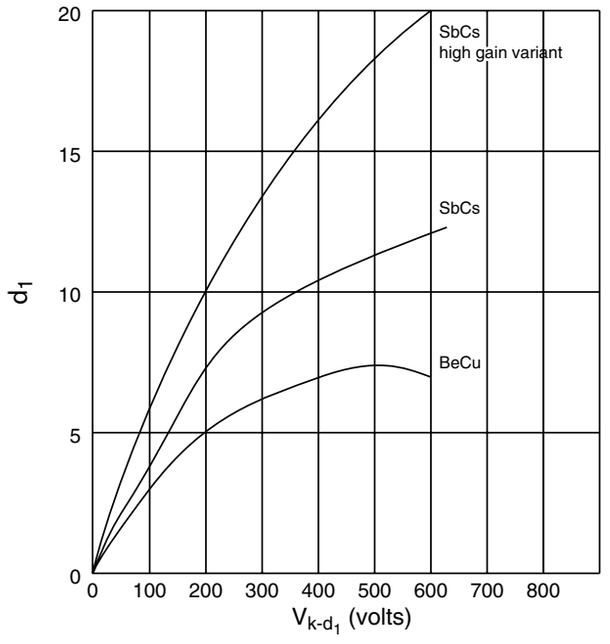

FIGURE 2.13. Variation of the gain from collisions that occur at the first dynode with cathode/first-dynode voltage.

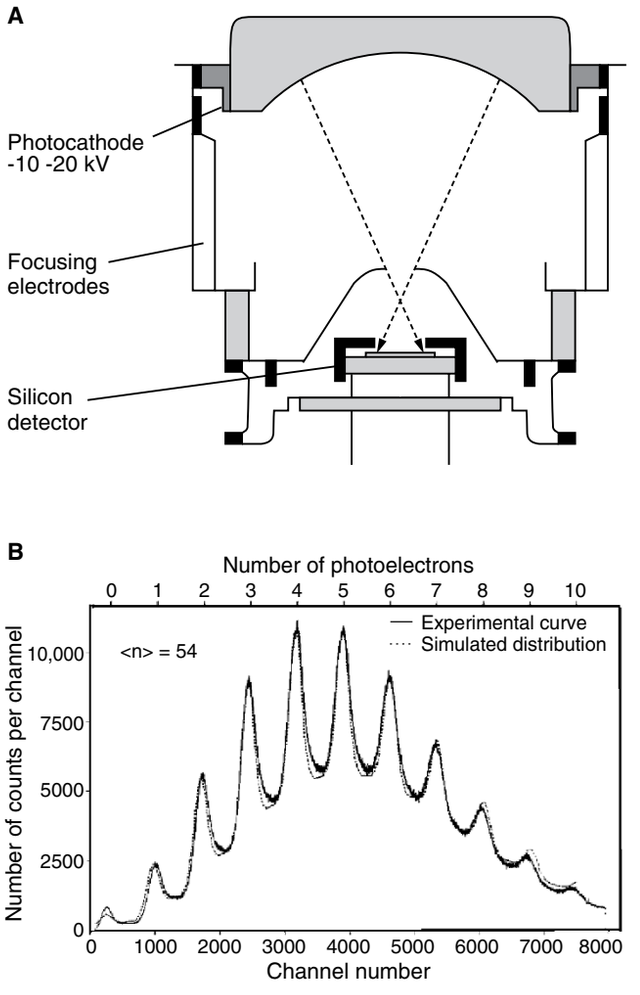

FIGURE 2.14. Hybrid PMTs. (A) Diagram of hybrid PMT layout. Photoelectrons leaving the photocathode are accelerated onto a silicon photodiode (or an array of such diodes) with sufficient energy to produce thousands of electron-hole pairs in the silicon. The very high gain of the first (and only) stage greatly reduces the effect of multiplicative noise. (B) A plot of single-pixel signal levels obtained from such a hybrid PMT under conditions where the photocathode was emitting on average, about 6 photoelectrons/pixel. The clearly defined peaks represent signal levels of from 1 to 11 photoelectrons/pixel.

even more multiplicative noise than is present in a good PMT. As a result, the output signal has about 1.4× more noise than can be attributed to Poisson noise alone. As this can only be reduced by counting 2× more photons, the easiest way to think of the EM-CCD is that it has no read noise, but the QE_{eff} is only 50% as great as it would have been in the absence of the multiplicative noise (i.e., a back-illuminated CCD detector having a QE of 80% would have an effective QE of only 40% if it were read out by an EM-CCD amplifier). This means that the EM-CCD is only optimal in the 0 to 20 PE/pixel signal range — just where the disk-scanning confocals operate[10] (see Chapters 10 and 43, *this volume*).

In addition to high QE, Si photon detectors have a variety of other practical advantages. As the sensitive element in such a detector is typically very small (5–30 µm on a side), selective use of only a few elements in a small, planar, 2D array could permit it to operate in the CLSM as a combination pinhole and detector. Figure 2.15 is a sketch of what such a detector might look like. After each pixel interval of the microscope, the charge pattern in the 5×5 sensor array at the top would be transferred to the read register and then the signal in all 25 pixels would be read out sequentially at about 35 MHz. These 25 values could then be "decoded" in a number of possible ways, the most straightforward of which would be to provide three separate signals corresponding to the summed signals from the brown, orange, and red areas of the sensor array. In this way, it would be possible to collect signal simultaneously at three different pinhole sizes.

With such a detector, pinhole alignment could be done electronically simply by searching for the detector element producing the most signal from a planar specimen and misalignment could be detected on the fly by comparing, for example, summed output from the 5 pixels on the left with the 5 on the right (Pawley, 1996).

Digitization

In the simplest CLSM system, the output of the PMT head amplifier is passed directly to the analog-to-digital converter (ADC), which samples the voltage for a few nanoseconds during the time of each pixel (t_p) and turns the sensed voltage into a digital number. As t_p is usually a few microseconds, it is important to ensure that the voltage present during the much shorter sampling time is a

[10] On first hearing, 20 c/pixel may sound like a very low level of signal. In fact, however, considerable experience shows that much fluorescence confocal microscopy is performed at much lower signal levels. The next section gives methods for those with instruments capable of fast-photon counting to calibrate their PMT gain controls. Those who follow this procedure may be surprised to find that when they are using "normal" PMT gain (8.00 and above on recent Bio-Rads), 256 stored in the memory corresponds to ~10 photons or less.

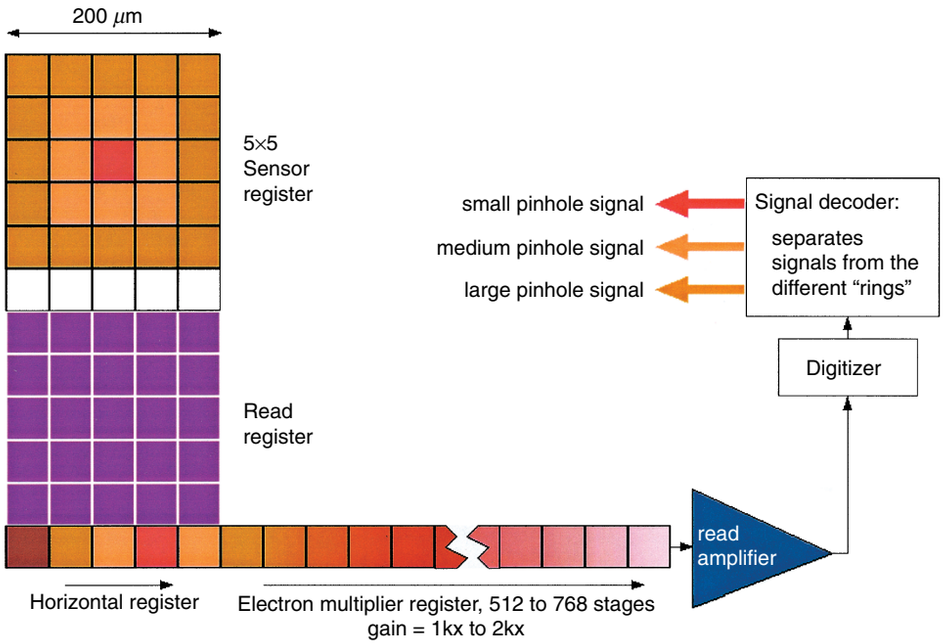

FIGURE 2.15. Schematic diagram showing a proposed solid-state detector that could be used to replace the PMT used in confocal microscopes. The detector uses an electron-multiplier register to reduce the read noise below 0.01 electron/pixel. By reading out all 25 pixels in the sensor array at each pixel of the confocal raster and then apportioning the signal to the three concentric areas of the sensor, this device would allow one to simultaneously collect signal at three different effective pixel sizes.

good measure of the average signal level during t_p. This is usually accomplished by limiting the bandwidth of the amplifier immediately preceding the ADC by giving it a time-constant of $t_p/4$. This approach is simple and effectively expands the sampling time from ~$t_p/1000$ to $t_p/4$ without excessively blending the signal from each pixel with that of its neighbors. However, this still means that the system is only "counting" about 25% of the time.

The situation can be improved if, as mentioned above, the system adjusts the electronic bandwidth to coincide with the optical bandwidth of the microscope and changes the pre-amp time constant to compensate for changes in optical resolution, scan speed (in μm/s as the focus plane) and t_p (see also Chapter 4, *this volume*).

Alternatively, it can also be improved if the CLSM uses a digitizer employing full integration. Such a system can be implemented in two ways. As first used by Sarastro/Molecular Dynamics and later by Bio-Rad, the output of the PMT is integrated by feeding a proportional current into a capacitor, then reading the capacitor voltage out to an ADC and finally resetting the capacitor voltage back to 0 (Fig. 2.16, lower left). The Bio-Rad MRC-600 and later instruments incorporate three circuits of this type in each digitizing channel: one accumulating, the second being read out, and the third being reset.

The second method of implementing full integration is to feed the output from a high-bandwidth ($t_p/20$) head amplifier into a high speed ADC running at, say, 10× the pixel rate and then utilizing fast digital circuitry to average the 10 successive digital readings needed to produce the single value actually stored for each pixel.

Compared to $t_p/4$ bandwidth limiting, either method of full integration effectively provides 4× more useful signal without any more light being detected.

Photon Counting

Obtaining a digital representation of optical data is ultimately a question of counting photons. Ideally, this means not only using a high-QE detector but also using a signal to which the contribution of each photon is equal. In the case of the solid-state sensors, the uniformity condition is automatically met by the sensing process (1 photon = 1 PE) but this condition can also be met by the PMT if it is operated in a pulse-counting mode.

In pulse-counting, the goal is not to measure the average level of the output current during the T_p but rather attempts to eliminate the effects of multiplicative noise by discriminating, and then counting, the individual output pulses resulting from the emission of individual PEs from the PC of the PMT. To reduce the effect of small noise pulses generated from the dynodes of the PMT, photon pulses are passed through a discriminator set to trigger on pulses larger than those at the bottom of the "valley" seen near the left edge of the pulse-height plot in Figure 2.12(B). Each time the PMT output goes above this preset threshold, one pulse is counted (Fig. 2.16).

Unfortunately, problems arise when large numbers of photons must be counted in a short time because the output of the PMT preamp does not immediately return to zero after each pulse. If a second pulse arrives before the first one is over, the second, or piled-up, pulse will be missed. Beyond the problem of pile-up is

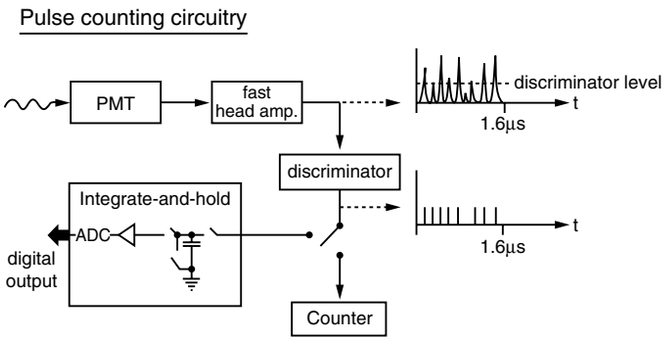

FIGURE 2.16. Two alternative approaches to counting single-photon pulses. In both, the signal from a fast head amplifier is passed to a discriminator. Uniform pulses from the discriminator can either be counted with digital circuitry or integrated in a capacitor and then read out through an analog-to-digital converter (ADC).

the difficulty of actually counting the output from the discriminator at very high asynchronous count rates.

Suppose a laser-scanning instrument scans a 512×512 raster in 1 s and 25% of this time is used for retrace. That leaves T_p = ~3 µs/pixel. If each photon pulse occupies 30 ns (t_p), the maximum number of pulses that one could possibly count in each pixel would be 100 but, because the photons arrive at random times, even at one tenth of this rate (i.e., 10 counts/pixel), 10% of the photons will still arrive when the circuit is already busy.

In general, the rate that pulses are recorded R_{rec} is given by

$$R_{rec} = T_p / t_p \left(1 - e^{-R_{input} t_p / T_p}\right) \quad \textbf{(2)}$$

The Bio-Rad MRC-1024 had a maximum count rate of 58 counts/pixel and its response was fairly linear up to about 15 counts/pixel. This instrument (and its successors) was unusual for allowing the user to switch to the photon-counting mode without any complicated adjustments. This ability made it fairly easy to occasionally switch modes, just to determine the number of photons/pixel that were actually being detected. On first hearing, 10 counts/pixel may sound like a very low level of signal. In fact, however, considerable experience with just this sort of mode switching shows that much fluorescence confocal microscopy is performed at much lower peak signal levels. Although one should be concerned about the signal saturation that occurs when one exceeds the linear range of the fast-photon counting circuitry, it is also important to remember that, to some degree, one can correct for piled-up losses with a simple look-up table.[11] Furthermore, such losses need not be reduced to zero but merely made small compared to the intrinsic statistical noise. Pile-up losses will become less important if more manufacturers switch to pulse-counting PMTs and use faster head amplifiers and counters.

With regard to the latter, digital counters are not strictly mandatory. All that must be done in order to remove PMT multiplicative noise is to clip all the single-PE pulses to a uniform size and feed them to a fully integrating ADC (Fig. 2.16). In fact, in some commercial instruments, much of the **beneficial effect of photon-counting** is incorporated into the **analog** digitization system by the simple expedient of arranging the electronic and PMT gain so that the single-PE pulses saturate a fast, high-gain amplifier installed between the PMT and the slower amplifier stage that leads to the ADC. Because this saturable amplifier is fast, each pulse is separately clipped to a uniform height, thereby meeting the criterion that each photon contribute equally to the final number recorded.

Where Have All the Photons Gone?

All present instruments embody design compromises that prevent them from obtaining the ultimate in photon efficiency throughout the four processes discussed above. Many systems employ more refractive optics than is absolutely necessary and most suffer transmission losses in the range of 30% to 65% (transmission figures as a function of λ for a number of modern objectives are given in Table 7.3, *this volume*). Although the metal mirrors that produced **losses of 85%** in early confocal instruments have now been replaced with broadband, dielectric-multilayer mirrors that reflect 98% of in-bandwidth light, the loss at each surface is multiplicative. The main culprit is often the triple dichroic beam splitter. Far from just reflecting at laser lines, this component often sends as much as 40% of the light between the laser lines back towards the laser. The use of an acousto-optical device as a beam-splitter (AOBS) by Leica has made a significant contribution to alleviating this problem (see Figs. 3.10 and 3.23, *this volume*).

Insufficient attention is often paid to the selection and adjustment of the PMT itself. While many recognize that any specific tube will operate best over only a narrow range of accelerating voltages and that those with bialkali photocathodes have lower noise and higher QE in the green, while those with S-20, multialkali, photocathodes are better in the red and near-infrared, it is usually forgotten that the performance of individual tubes often varies from the mean for their type by more than a factor of 2. In addition, the new side-window PMTs that are used in a number of systems have a response in the red that is almost 3× that of traditional end-window tubes. While this performance also implies a substantial increase in dark current, cooled versions are now available.

Additional degradation is imposed on the data by multiplicative noise and poor digitizing circuitry. Finally, signal is often lost because of poor alignment, and the improper choice of pinhole diameter may exclude as much as 90% of the useful signal from the detector in a vain attempt to extract an "imaginary" improvement in resolution (imaginary because the low signal levels prevent useful information from being obtained before the specimen is destroyed).

Taken together, all of these factors can add up to a factor of 100× or more in photon efficiency between state-of-the-art and sloppy operation on poorly chosen equipment. Every action that results in more efficient use of the photons generated within the specimen should be thought of as being directly responsible for making it possible to collect a proportionally larger number of images (or images with better statistics) before the specimen is damaged.

Measuring Photon Efficiency

The PMT output signal is the only source of readily available data with which to try to measure the photon efficiency. Unfortunately,

[11] If 10 were counted and this is known to cause 10% dead time, record 11.

a large number of parameters can have a major effect on this single measurement (Pawley, 2002). Those discussed above include:

- Laser power, which is a function of temperature, cavity gain, precision of stabilizing circuit, etc.
- The transmission of ND and other filters used at a particular λ.
- Proper alignment of the optics launching the light into the fiber.
- The NA and transmission of the objective lens and other optics.
- Reflectivity of the mirrors for the particular λ and polarization in use.
- The fraction of the laser beam that is actually accepted by the objective back entrance pupil (varies with NA/magnification).
- Pinhole diameter and alignment.
- PMT voltage and black-level setting.
- Staining density and type of dye.
- Focus level and RI of embedding media.

The number and diversity of these parameters make it difficult to obtain a measure of the fraction of the photons leaving the focused spot that are actually counted as part of the stored image data.

What is needed is a stable point-source of light of known intensity that can be mounted conveniently below the objective. One way to make such a source is by allowing a known amount of laser light to strike a stable phosphor. One first measures the light **emerging** from the objective (as noted above) and adjusts it to some standard amount. Specimens that maintain a constant level of fluorescent efficiency (i.e., ones that do not bleach or change with time) include such inorganic phosphors as single crystals of CaF_2–Eu or YAG–Ce and uranyl glass. Unfortunately, although these materials are very stable under intense laser illumination, they also have very high RI. Consequently, high NA objectives are unable to form an aberration-free focus within them, and because of this the signal that they generate at the PMT decreases rapidly as the focus plane moves into the material. However, they can be useful to those who normally use objectives of lower NA where spherical aberration effects are a less serious factor. Cubes of fluorescent plastic or uranyl glass also offer a convenient way of demonstrating correct alignment of the laser with the objective back-focal plane (BFP; see Chapter 36, *this volume*). An alternative fluorescence standard can be fabricated by dissolving dye in immersion oil or water (depending on the correction of the objective), but such specimens are not always as stable over time as one might wish.

A more direct approach to measuring photon efficiency involves using the microscope simply to image a small light source such as a light-emitting diode (LED) specimen, or the light formed by the microscope's normal transmission illumination system set up for Köhler illumination (Fig. 2.17). The arc or incandescent transmission source must be provided with a known and constant power supply. Once this has been set up, the only major variables remaining are the amount of metal deposited on the inside of the glass envelope surrounding the source, the bandpass effects of any filters that remain in the light path, the pinhole diameter, and the PMT voltage. In many instruments, it is relatively easy to remove all of the dichroic and bandpass filters and let the light from the image plane pass directly to the PMT. Under these conditions, one should get a standard reading with a given objective, pinhole size, and lamp power. Although the microscope is now a flying spot detector and only collects photons from one pixel at a time with the result that the effective signal intensity is about 250,000× less than if the PMT were measuring the entire field, it will often still be necessary to place ND filters between the source and the condenser lens to permit the PMT to be operated at a voltage similar to that normally used. Line-frequency variations in the filament heating current may be visible in the data as will any instability in the plasma of arc sources. For this reason, it may be more convenient to measure the PMT output with an analog DC voltmeter than with the ADC and image memory (see Chapter 36, *this volume*).

With such a setup, one can also calibrate the actual effective size of the pinhole. To do this, record the PMT signal level (I_s) as a function of the pinhole setting. I_s should vary as the square of the pinhole diameter, and a plot of $\sqrt{I_s}$ versus pinhole diameter should be a straight line passing through the origin. Departures from linearity may indicate that dust is obscuring part of the aperture or that the pinhole control does not give a very accurate estimate of pinhole diameter.

By introducing ND filters below the stage, it is possible to reduce the intensity of the light to the level at which photon-counting is appropriate. On those instruments that have this ability, one can easily make measurements to determine a rough ratio between the actual number of photons being detected (using photon-counting) and analog-intensity-values stored in the memory at the same PMT gain settings. This is done by recording the same brightfield "image" in both analog and photon-counting modes. Such information should be used to reach a rational deci-

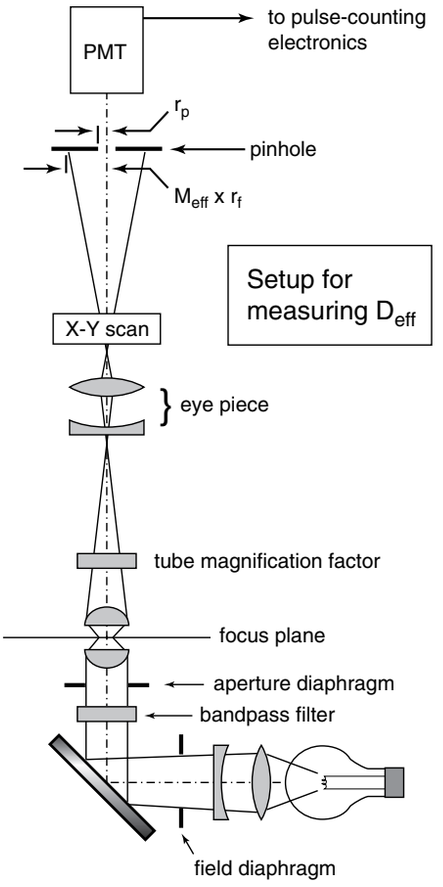

FIGURE 2.17. Optical setup for measuring the detection efficiency or the effective size of the pinhole using the internal transmitted illumination system of the microscope as a standard light source.

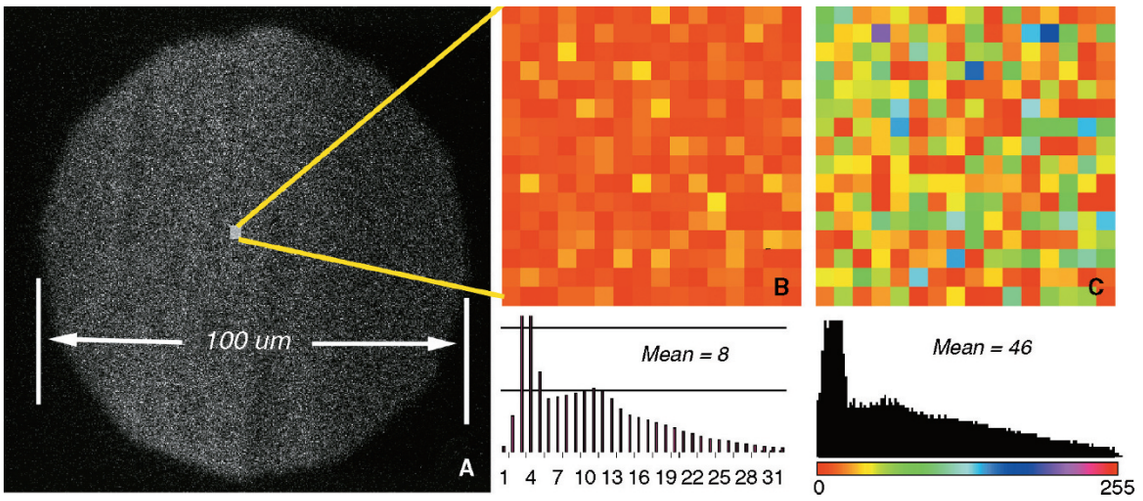

FIGURE 2.18. Effect of photon counting. The two images were made using the confocal detection system as a flying-spot scanner to record the light from the internal transmitted-light illumination system set up for Köhler illumination. Both images were recorded with a PMT setting of 8.00 but (A) was recorded in the analog mode while (B) was recorded using the fast-photon-counting mode (and the contrast was subsequently stretched as the brightest pixel had only 12 counts recorded in it).

sion about when one can use photon-counting without fear of pulse pile-up (i.e., that if the PMT gain is above XX, the signal level registering as 255 in the image memory is only "*x*" photons/pixel and so one can switch to photon-counting for lower multiplicative noise without any worry of pile-up losses). Figure 2.18 shows two such images with their accompanying histograms. The circle is formed by the field diaphragm that has been partially closed to provide a "zero-signal" reference. Different output look-up tables had to be used to display the two images because the photon-counting image had only 12 counts in the brightest pixel.

Finally, if a calibrated radiometer is available, it is possible to use this setup to measure the actual detection efficiency of the instrument, D_{eff}. From an image such as that in Figure 2.19(B), it is easy to measure the actual radius (r_f) of the field diaphragm in the focus plane in micrometers. If one can then measure the total light power (P) passing through this plane by placing the sensor of a photometer calibrated to measure milliwatts and one can specify the λ of these photons by placing a narrow-band filter in the illumination light path, one can then calculate the flux of photons passing through this plane in photons/μm^2, (F).

The energy of a single photon is $h\nu = hc/\lambda$, where c is the speed of light and h is Planck's constant, so the total number of photons having wavelength λ coming through the condenser and into the sensor of the photometer is $n = P\lambda/hc$ (photons/s). The fraction of those that should then reach the PMT is simply the ratio of the effective area of the pinhole (πr_o^2) to the effective area of the field diaphragm. The effective radius of the pinhole is its physical size (r_p) divided by the total magnification between it and the focus plane (M_{eff}). In the Bio-Rad, $r_p = 8$ mm with the pinhole all the way open and $M_{\text{eff}} = 53(M_{\text{tube}})(M_{\text{obj}})$. With a 10× lens and a $M_{\text{tube}} = 1.5$ (assuming 1.25× for the fluorescence attachment and 1.25× for the differential interference contrast (DIC) attachment. For modern, infinity-conjugate microscopes, $M_{\text{tube}} = 1$), $r_o = 5\,\mu$m. (Values of M_{eff} for a number of confocal instruments are tabulated in Appendix 2.)

If the radius of the field diaphragm is set to $r_f = 500\,\mu$m in the focus plane, then the fraction of the light from this plane entering the detector pinhole is $(r_o/r_f)^2 = (5/500)^2 = 0.01\%$, and the total number of photons striking the PMT, n_{PMT}, in sampling time t_s(s) will be

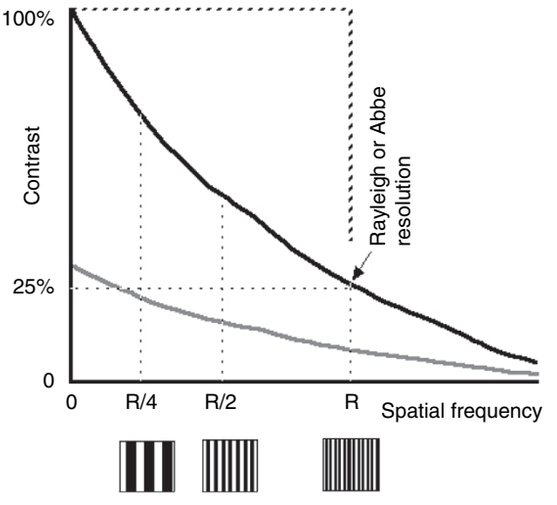

FIGURE 2.19. Contrast transfer function (CTF) of a microscope optical system. The black line shows how the contrast in the image of a high-contrast bar pattern varies with the spatial frequency of the pattern (spacings/μm). The gray line represents the response measured if the dark bars in the test target reflect (or transmit) 70% of the incident light rather than 0%. The dashed line represents the imaginary situation of a CTF in which all spatial frequencies up to the "resolution limit" really have equal contrast. The diagram makes clear why high frequency features of a test object are more likely to remain visible above any arbitrary noise level (5% of 26.5% shown) if they have 100% contrast rather than 30%.

$$n_{PMT} = \frac{D_{eff} N \, r_o^2}{r_f^2} t_s = \frac{D_{eff} P \lambda}{h \, c} t_s \frac{r_o^2}{r_f^2} = K D_{eff} P \qquad (3)$$

where D_{eff} is the transmission efficiency of the optical system and K is a constant that relates the power measured coming out of the condenser P, to D_{eff} and the number of photons counted in the PMT. For a Bio-Rad setup as noted above and with P measured in nanowatts and $\lambda = 546\,\text{nm}$, $K = 0.04$ photons/pixel/nW. The comparison between photons/pixel measured at the image plane with those detected by photon-counting is a worthwhile, though sobering, exercise.

Nothing increases the probability that normal operating procedures are optimal as much as practicing these techniques under test conditions because, in this case, one knows if one is getting the correct answer (see Chapters 35 and 36, *this volume*, for additional performance tests).

RESOLUTION: HOW MUCH IS ENOUGH?

A peculiar feature of any properly aligned confocal microscope is that it is always "in focus," in that the exciting and detecting light paths are always focused into the same diffraction-limited spot. Details of the size and shape of the volume sampled in this way are dependent almost entirely on the design of the optical system (especially the NA of the objective lens and the RI of the medium surrounding the specimen as formulated in the Abbe equation). In broad terms, the effective *xy*-resolution in biological confocal fluorescence microscopy is essentially the same as it is in widefield,[12] while the *z*-resolution is about 3× worse at NA 1.4 and 4× worse at NA 1.2. A more detailed analysis can be found in Chapters 1, 7, and 8. In addition, there is the important trade-off between spatial resolution, statistical accuracy, and radiation damage, as outlined above and related to the setting of the pinhole diameter.

Rather than repeat the straightforward aspects of theoretical resolution again here, this section will discuss three other aspects that are perhaps less obvious but which are both characteristic of confocal microscopy and that fundamentally limit the ability of the instrument to solve biological problems. These topics are:

- Circumstances under which it may be desirable to reduce resolution.
- The effect of image digitization on spatial and temporal resolution.
- Practical considerations that may degrade resolution or produce distortion.

Can Resolution Be Too High?

Normally, the microscopist makes every effort to obtain the highest possible spatial resolution, and, to the extent that the specimen is not degraded by being so observed, this is entirely proper. Although, in the case of biological samples, the condition of non-degradation is never entirely met, such samples can be rendered very robust by treatment with anti-bleaching agents and fade-resistant dyes (Chapters 16, 18, and 39, *this volume*). However, such fixed and chemically protected structures cannot be considered the general case. Recently, significant improvements in the photon efficiency of confocal instruments has led to them being used increasingly for viewing living specimens (Chapter 19, *this volume*). Indeed, experiments in which thousands of images have been recorded from particularly hardy cell lines are no longer unusual.

To determine when it may be advisable to intentionally reduce spatial resolution, it will be helpful to consider a second specimen type: a living cell that has been micro-injected with a fluorescent substance that changes its absorption spectrum in response to the local concentration of certain ionic species, a concentration expected to change on the scale of a few micrometers. The object of the experiment is to monitor changes in the concentration of this ion as a function of time and experimental conditions. The major complication is that the dye is cytotoxic, and this is especially true when it is excited by light. Furthermore, the fluorescence must be measured accurately at two different exciting wavelengths to determine the ionic concentration by ratioing the results. Because of diffusion, significant changes in ion concentration are not expected to occur on the scales smaller than 1 µm.

How does one optimize an instrument to perform this experiment? Clearly the cytotoxicity problem implies:

- Using a water-immersion lens of large NA to collect as much light as possible.
- Using the lowest dye concentration consistent with the statistical accuracy required.
- Using a magnification such that, when referred to the specimen, a pixel is on the order of $1\,\mu\text{m}^2$.

The amount of illuminating light that should be used must be determined on the basis of the number and accuracy of the measurements that are to be made. As it is important to measure changes in ionic concentration, it can be assumed that the time available for each measurement is a constant, probably set by the maximum scan speed of the galvanometers, although the intensity of the illumination used to make the images need not be. Both statistical accuracy and toxicity will increase with the intensity of the exciting beam. As "biological accuracy" cannot be sacrificed to obtain better statistical accuracy, conditions must be adjusted to maximize the number of measurements that can be made before the cell is damaged. Considering this problem from the point of view of a single-beam confocal will highlight the effect of fluorescence saturation.

The need for ratio imaging implies a fairly short scan-time to avoid changes in the specimen between the two component images. The need for statistical accuracy implies an intensely illuminated spot to produce many countable photons but, in single-beam confocal systems, saturation puts a limit on the maximum useful flux that can be used. Fortunately, the low spatial resolution required provides some flexibility. The flux of exciting illumination is highest at the neck of the cone of illumination formed by the objective lens, and the diameter of this cone will be smaller (and the flux higher) when the lens used to form it is operated at full NA. In the example described, however, the large NA was chosen to provide high collection efficiency, not high resolution. As the pixel size is to be 1 µm, there is no need to focus the entire beam into a 0.2-µm spot because this will produce a maximum flux density 25 times greater than that present in a 1-µm spot and incur a proportionally greater risk of dye saturation. (Note: To insure that the signal is collected from the entire area of the larger spot, a 5× larger pinhole diameter must be used.)

[12] Although other chapters in this volume note that the ultimate *xy*-resolution in confocal is 1.4x smaller than that in widefield, this assumes the use of a very small pinhole. Given the low signal levels that characterize stained biological specimens one virtually never uses such a small pinhole. The statement above assumes a pinhole diameter of 2 Airy units.

When using coherent laser illumination, the only way to make the spot larger is to under-fill the BFP of the objective lens, thereby reducing its effective NA. However, if one is using a more diffuse source such as an Hg arc, one can use a larger source "pinhole" in the intermediate image plane. Although under-filling the objective reduces the *z*-resolution, considerable depth discrimination remains, and the confocal microscope retains all of the other desirable features, such as quantitative measurement accuracy, and the ability to scan a plane within a larger specimen often makes it ideal for studies of this kind.

The lesson here, then, is that because of the twin considerations of dye diffusion and saturation, it is often important to be able to adjust the image-forming properties of the instrument to produce a spot size appropriate to the experiment at hand. The ability to intentionally under-fill a high-NA objective is poorly implemented on current commercial instruments.

Limitations Imposed by Spatial and Temporal Quantization

Although the image viewed in a disk-scanning or slit-scanning confocal microscope is, in principle, as continuous as that from a WF microscope, this distinction is lost when the image is finally sensed using a cooled CCD or a digital-video image sensor. The fact that all digital confocal images must be recorded and treated in terms of measurements made within discrete pixels can limit the effective resolution of the instrument in ways that may not be familiar to some who approach digital microscopy for the first time. In a sense, these limits are more practical than fundamental because, if the microscope operation follows the rules of Nyquist sampling theory as discussed below (and in greater detail in Chapter 4, *this volume*), these limits should present no obstacle to recording good images. However, because incautious use of the "zoom" magnification control found on all commercial CLSMs makes it relatively easy to operate these instruments outside the Nyquist conditions, a brief discussion of sampling theory is also included here. It is mentioned under the heading "Resolution" because it involves the ability to **record** "the separation of two closely spaced objects."

Spatial Frequencies and the Contrast Transfer Function

Sampling theory, like resolution itself, is often easier to think about accurately in the "**spatial frequency domain**" where one considers not the size of objects themselves but the inverse of their size: "How many objects are visible per millimeter?" In the spatial frequency domain, an image is seen as being composed of the **spacings** between features rather than the features themselves. The reason for using this seemingly obscure mental construct is that the ability of an optical system to transmit information depends entirely on the spatial frequency of this information. More specifically, all optical systems reduce the contrast of the high spatial frequencies (representing smaller spacings) more than they do lower spatial frequencies (representing larger features). This fact is made evident when one plots the contrast transfer function (CTF) of an optical system by measuring the contrast present in an image of a test object made up of regular arrays of black-and-white bars each having a specific spacing or frequency (Oldenbourg *et al.*, 1993; Chapter 1 and Figures 35.6–35.11 in Chapter 35, *this volume*). Such a CTF is shown by the black line in Figure 2.19 and the gray line below it represents the image contrast produced by a test target that is 100% white and 70% gray (i.e., one with only 30% contrast).

Contrast Transfer Function and Resolution

Although in common parlance the word "resolution" is often used as though it were an independent parameter, Figure 2.19 makes it clear that the number chosen to represent the "resolution" is really somewhat arbitrary. It refers to the highest spatial frequency at which the contrast is above some given value. For instance, the Rayleigh/Abbe criterion for bright, point-objects on a black background (i.e., stars on a clear night) really assumes that the minimum visible contrast is 25%. The problem with this simplistic attitude to resolution is that it can give one the idea that, no matter what their original contrast (i.e., the staining specificity) in the object, all spatial frequencies up to "the resolution" are equally visible as is implied by the upper dashed line in Figure 2.19. In fact, the most important message from Figure 2.19 is that the contrast-in-the-image is proportional to contrast-in-the-object as degraded by the CTF of the imaging system, and in particular, **the contrast of small features just within the "resolution limit" is much lower than that of larger features**.

Visibility, Resolution, and the Rose Criterion

The reciprocal relationship between contrast and resolution is important because usually what one is actually interested in is not resolution *per se* but **visibility**: the ability of an observer to recognize two closely spaced features as being separate. Visibility is a much more slippery concept because it depends not only on the calculable aspects of diffraction theory but on the higher functions that our visual system uses to determine how a particular pattern of light intensity should be interpreted. As a result, there are few standard methods for measuring visibility.

On the other hand, visibility has the advantage of requiring us to consider another important parameter of the image data: its signal-to-noise ratio (S/N). The Rose Criterion states that, to be visible, a dark feature that is a single pixel in size must have an intensity that differs from that of a white background by at least 5 times the noise level of the background[13] (Rose, 1948). Although the factor 5 is somewhat lower for lines and other geometrical features, the point is that visibility depends on more than geometrical optics and the CTF. Assuming that the contrast is measured in units that are proportional to the statistical precision with which each signal level is known, visibility also depends on both the absolute and the relative difference between the signal levels of the feature and the background. In the case of being able to resolve two point features, the Rose Criterion requires that the noise be low enough and the contrast be high enough that one can discriminate the lower intensity of the pixel between the two features as being different from that of the peaks. A number of authors have pointed out that, in fluorescence confocal microscopy, one seldom counts enough photons for the 25% contrast of the Abbe resolution to be visible with any degree of confidence.

As with the CTF curve, where resolution is defined in terms of contrast in the final image, visibility requires that, as smaller objects have lower contrast, more photons will have to be counted to reduce the Poisson noise of the measurement well below the low contrast of the features.

The reader can see that visibility brings together the contrast of the object (as embodied in its staining characteristics), the focus-

[13] The choice of a dark feature on a white ground is not arbitrary. A white signal has significant but measurable Poisson noise. A truly black background does not.

ing properties of the optical system as defined by the CTF and the statistical requirements of the Rose Criterion. However, there is still one other process that can limit whether or not a feature is visible in the final image and that is how the signal is digitized.

Digitization and the Nyquist Criterion

To convert any continuous, analog signal into a digital representation, it is sampled by measuring its intensity at regular intervals in time (or space). Suppose that one is recording the air temperature at a particular location over time. How often should one make such measurements? Clearly the answer depends on the rate of change of the "features" of the temperature record that are of interest. These might be the effects of breezes or thermals that may change over seconds or less, or they might be climatic changes that might only be noticeable over centuries or millennia. Nyquist's crucial insight was that there is a fixed relationship between the highest temporal (or spatial) frequency of interest in the data set and the minimum rate at which samples must be taken if they are to record all the possible significant variations in that signal accurately. Specifically, for non-periodic data, the sampling frequency should be at least 2.4× higher than the highest frequency in the data. Consequently, to preserve all of the information that could be recorded using a CTF such as that shown in Figure 2.19, it will be necessary to have the pixels smaller than 1/2.4 F_c where F_c is the cut-off frequency.[14] This means that the Airy figure image of a point object should be at least 4 to 5 pixels across the diameter of its first dark ring.

The practical problem that arises in CLSM is that, although changing the "zoom" magnification changes the area scanned on the specimen, it does not normally do so by varying the number of pixels, and consequently, pixel size is inversely proportional to the zoom factor. However, this does not change the fact that, for a given optical system, only one pixel size (and one zoom factor!) matches the Nyquist criterion. At zoom settings that provide smaller pixels, the data will be over-sampled with the result that it will be bleached more than necessary and only a smaller field of view can be scanned in a given period. It is becoming more common for manufacturers to display the pixel size as part of the setup display, but if this is not done, it is not difficult to calculate if one has a knowledge of the length of some feature in the image in both pixels and micrometers.

On the other hand, at lower zoom settings where the pixels are larger than those prescribed by Nyquist, not only may some smaller features be missed entirely but features not present in the object may "appear" in the data because of a phenomenon called aliasing.

One can understand aliasing by returning to the temperature analogy. Assume that the thermometer is capable of responding to changes in temperature on the order of minutes but that only annual changes are of interest. Nyquist would seem to say that one can find out the average yearly temperature by making on average 2.4 measurements/year. However, there could clearly be problems if one sample is made at 6 AM in December and the next at 11 AM in May! Large day-to-day and time-of-day variations will swamp out small changes in the average annual temperature. Nyquist can only work if the response time of the sensor is reduced until it is about 2.4× our sampling interval. This could be accomplished by burying the thermometer a few meters in the earth where the temperature changes much more slowly (Pollock and Chapman, 1993).

An analogous situation occurs in a CLSM operated at a low zoom setting. When images are recorded on a "continuous" medium, there is no possibility that small but bright objects will be entirely missed. However, this is not true of sampling systems. When pixels are large compared with the optical resolution (i.e., low zoom setting), it is possible that small features will be missed altogether (so-called "blind spots"). Blind spots only occur because Nyquist sampling conditions are not met and small features lying between scan lines are not sampled. However, on instruments that employ $T_p/4$ integration rather than full-integration in their digitizing circuitry, small features that the scanning beam passes during the dead time of the digitizer may be recorded with lower contrast or not at all.

Aliasing may cause features to appear larger, smaller, or in different locations from where they should be. Like aliasing, blind spots are only a problem when the pixel size is much larger than the spot size and where the object contains a lot of small features (see Chapters 4 and 35, *this volume*, for a more complete analysis).

There are two strategies for reducing these problems (both discussed in the previous sections): increasing the size of the focal spot and using a fully integrating digitization system. However, the latter approach only avoids blind spots in the horizontal direction.

The Nyquist sampling rule is often flaunted by practicing microscopists but the errors produced by this practice usually pass unnoticed for two reasons:

- Working images that are somewhat over-sampled tend to be more pleasant to view on the display screen. As the size of the individual pixels (referred to the sample) is considerably smaller than the Nyquist limit, the data looks "as it should" but the radiation dose needed to record it was higher than necessary.
- The pixel intensities making up the images involve so few quanta that statistical variations in the signal pose a greater limitation to defining the position of the edge of a feature than does the quantization of the image in space. This is less of a limitation if one uses more photons, which happens when one uses more pixels.

Nevertheless, as applications of confocal technology more closely approach the absolute limits (e.g., by more closely matching the pixel size to that of the data of interest and improving the photon efficiency so that the statistical noise is reduced) the costs of incorrect spatial (and temporal) quantizing will become more apparent.

Fortunately, a technique is available to help solve the noise problem. Because Poisson noise affects the measurement of intensity in each pixel independently, positive noise excursions can appear to be small bright features that are only one pixel wide. These can be eliminated by adhering to the second aspect of Nyquist sampling, the part related to **reconstructing** an analog image from the digital data. Nyquist mandates that the reconstructed signal be passed though an "amplifier" having the same bandwidth as that from which the original signal was obtained. In the case of microscopy, the input "bandwidth" in question is

[14] The original Nyquist theory related to 1-dimensional data. As 2D images may contain spatial frequencies at any angle to the raster axes, it might seem logical to increase the sampling frequency by 1.4 to account for features running at 45° to these axes. However, as is shown in Fig. 4.10, adjacent rows of pixels fill in this data. In addition, because the low S/N of most confocal data limits the effective resolution of final data more severely than does the Abbe criterion, the rule that pixels should be 2× smaller than the Abbe resolution limit works well.

defined by the CTF of the optics. As single-pixel noise includes spatial frequencies that are at least 4× higher than the Abbe limit, no information is lost by passing the recorded data through an "amplifier" having the same 3D bandwidth as the microscope. The simplest way of doing this is to deconvolve it.

This point bears repeating. Although deconvolution and confocal imaging are often seen as competing methods aimed at the same goal, namely producing images of 3D stain distributions, in fact, they are not exclusive, and there is much to be said for combining them. Not only does deconvolution suppress the "single-pixel" features created by Poisson noise, it also effectively averages the signal over 16 to 25 pixels in 2D and 64 to 125 voxels in 3D because this is the number of pixel/voxels that carry appreciable signal in the Nyquist-sampled image of a point object (Fig. 2.20). **In other words, it has the same effect of reducing the uncertainty in the estimate of the brightness in individual voxels as Kalman averaging for 64 to 125 frames.**[15] This point is so important that the present edition of this volume devotes an entire chapter to it: Chapter 25.

Figure 2.21 shows what can be accomplished. This example was kindly provided by Hans vander Voort of SVI, Delft, NE. The top row shows a through-focus series of an actual Nyquist-sampled PSF. The next two rows show two simulated through-focus series in which Poisson noise has been added as though the brightest pixel recorded only 25 photons, and the remainder proportionally less. Although the noise is different in each column, when they are deconvolved, the results (seen in the two bottom rows), are almost identical.

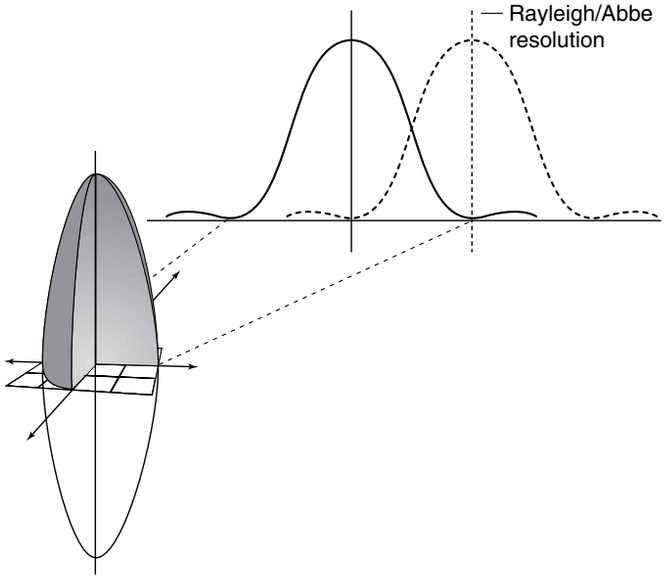

FIGURE 2.20. Relationship between the Rayleigh-criterion resolution, the point-spread function (PSF), and Nyquist sampling. If the PSF is sampled at the Nyquist level, each pixel will be less than 1/4 of the distance from one side of it to the other and ~16 pixels will have significant signal in them in each *xy*-plane.

Because the confocal PSF is so much more confined than that of WF data, it takes much less time to deconvolve confocal data. Indeed it is quite reasonable to approximate the process by a 3D Gaussian blur (see Fig. 19.2, *this volume*).

There are two other matters about sampling that should be considered:

- *Temporal aliasing effects*: The accuracy with which temporal changes can be imaged is limited by the frame-scan rate. Failure to sample sufficiently often can create artifacts such as the "back-rotating wagon-wheel effect" familiar from the movies. In the CLSM, temporal aliasing not only limits the ability to follow motion, it also has a resolution component in terms of the imprecision of the measurement of the location (or motion) of objects. Specifically, because the pixel time is short while the frame time is long, motion of the specimen produces distortion rather than the directional blurring that would occur if the whole field was recorded continuously over the same period of time.
- *Mismatch of probe and pixel shape*: There is a mismatch in shape between the circular shape of the moving Airy disk and the "hard-edged" square shape of a pixel on the liquid-crystal display (LCD), a difference made more pronounced by the innate tendency of the human visual system to emphasize edges.

In the case of a signal recorded by a cooled CCD detector on a disk-scanning confocal system, the CCD detector spatially quantizes the signal in a manner entirely determined by its sensor geometry and the total optical magnification. However, because the scanning mirrors in the CLSM follow ballistic rather than stepped trajectories, the output from the PMT of these instruments is, in principle, continuous in the horizontal direction. As a result, averaged over the digitizing time, the Airy disk is not round but is slightly more blurred in the horizontal direction.

Practical Considerations Relating Resolution to Distortion

To obtain the theoretical spatial resolution of a confocal microscope, it is, of course, necessary to have a diffraction-limited optical system, but this is not sufficient. Leaving the practical aspects of alignment and optical quality for Chapters 7 through 11, 22, and 35, this section discusses the limitations imposed by mechanical stability and the repeatability of the scanning system, a topic that highlights one of the most important differences between disk-scanning and laser-scanning instruments.

In disk-scanning instruments, the image is real and, therefore, cannot be distorted by the scanning system.[16] If a CCD sensor is used to detect it, geometrical distortion in the digitized image is extremely low, and because of the inherent mechanical stability of the sensor, any distortion that remains from the optics is stable and can be corrected by minor digital image processing. The mechanical problems of this instrument are, therefore, confined to the effects of vibration and the relative motion of the stage and lens.

In all confocal instruments, it is vital that relative motion between the objective and the specimen be kept to less than 10% of the resolution limit in *x*, *y*, and *z*. In tandem disk-scanning

[15] Although such filtering **will** reduce image contrast, the contrast can be increased arbitrarily by changing the display look-up table, a process that **does not** decrease the S/N.

[16] Although it can be distorted by any barrel or pincushion distortion present in the optics.

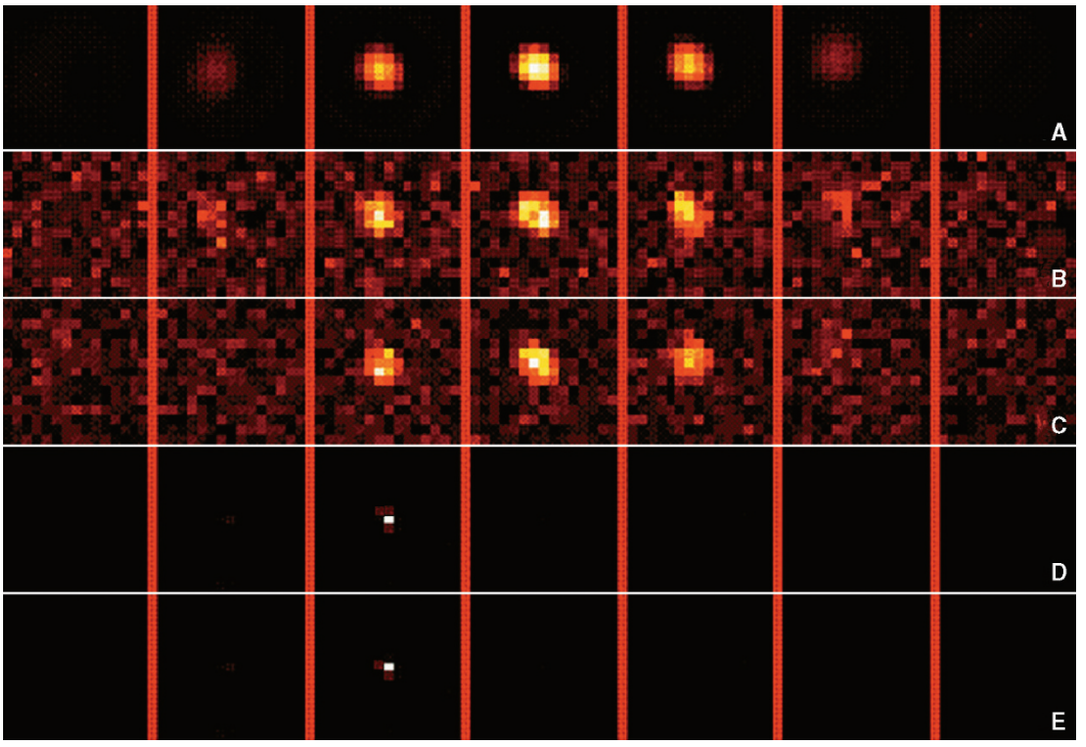

FIGURE 2.21. Deconvolution of simulated confocal data improves S/N and resolution. (A) shows a through-focus series of an actual Nyquist-sampled PSF. (B) and (C) show simulated images in which Poisson noise has been added as though the brightest pixel recorded only 25 photons. Although the noise is different in each column, when they are deconvolved, the results (D) and (E) are almost identical. (This example was kindly provided by Hans vander Voort of SVI, Delft, NE.)

instruments, both the rotating disk and the cooling system of the often-large illumination sources represent potential sources of vibration in addition to those normally present in the environment. All vibrations must be reduced below the tolerance levels or resolution will be degraded in a fairly straightforward manner. In this regard, it is not sufficient to mechanically isolate the microscope on some sort of isolation table because laser-cooling fans, spinning disks, scanning galvanometers, etc. can all **introduce** vibration into the isolated system.

Not so straightforward is the effect of vibration, and possibly stray electrical and magnetic fields acting either directly on the galvanometers or introducing spurious signals into the current amplifiers that control them. In these instruments, accurate imaging depends on the mirrors causing the beam to scan over the sample in a precise pattern duplicating the mathematically perfect raster represented by the data locations in the image memory. Failure to duplicate this raster has the result that data will be displayed in the wrong pixel, producing distortion. On a system with a 1000-line raster and a 10 : 1 zooming ratio, keeping beam placement to within 0.1 of a pixel requires accuracy of one part in 10^5. The electromechanical properties of the galvanometers (mass, spring constant, frequency response, overshoot, resonant frequency, bearing tolerance, rigidity, etc.) produce additional errors (Chapter 3, *this volume*). Image distortions produced by these errors are often masked by the paucity of test samples having an accurately defined geometry (Pawley *et al.*, 1993) and by the fact that, at high zoom, current instruments greatly over-sample the image so the smallest, visible structural features are many pixels wide.

This problem merits mention here because it is possible to measure the x, y, z **position of the centroid** of an object in a digital image to an accuracy that is **much smaller than the spatial resolution limit**. Indeed, WF light microscopy techniques have been used to measure motion on the order of 1 nm (Gelles *et al.*, 1988). However, due to the random imprecision in the systems used to position the mirrors, it is unlikely that measurements of similar reliability could be made on any present CLSM. In this context, then, the accuracy and precision with which mirror position can be controlled is a fundamental limitation on the ability of CLSM to determine position.

The presence of scan instability can be detected either by visually comparing sequential **single-scan** images of a diagonal knife-edge viewed at the highest possible magnification, contrast and signal level or, alternatively, by computing the apparent motion of the centroids of two fixed objects, each covering 100 to 200 pixels as they are recorded on a number of sequential scans. The variations measured by the second method should decrease rapidly with increasing illumination intensity because of improved statistical accuracy, although they may then begin to measure stage drift.

However, before all position errors are blamed on the galvanometers, it is important that living cells are far from the innocuous transparent objects often assumed in articles on microscope optics. As noted above the pronounced variations in RI that allow

TABLE 2.1. Fundamental Limits in Confocal Microscopy

Parameter	Theoretical	Practical
Resolution	Spatial: Y $\lambda_1, \lambda_2, \alpha_1, \alpha_2$	Alignment Off-axis aberrations
	X $\lambda_1, \lambda_2, \alpha_1, \alpha_2$	Alignment Off-axis aberrations Bandwidth Ballistic scan
	Z $\lambda_1, \lambda_2, \alpha_1, \alpha_2$ Pinhole diam.	Alignment Off-axis aberrations Δ tube length
	Temporal: Scan speed Signal decay times	Quantization of t
Position	Objective lens distortion Pixellation limitations Sampling time	Mirror accuracy Vibration
Quantitative measurement	Poisson statistics	Mirror and digitizing losses Detector QE and noise Bleaching/photodamage Saturation Source brightness

us to image cellular organelles in phase-contrast or differential-interference contrast microscopy can also profoundly distort the focus plane in the *z*-direction.

CONCLUSION

I have attempted to highlight some aspects of confocal instrumentation that must be addressed in order to attain performance limited only by fundamental considerations (Table 2.1). Although the constituent aspects of both photon efficiency and resolution have been addressed separately, I have tried to emphasize that the effects of both of these factors overlap and interact in a fairly complex manner. To summarize:

- Although resolution in the confocal microscope is primarily a function of the NA of the optical system and the λ of the light, it can also be further limited if the signal level represents so few quanta that the detected signal lacks the statistical precision to produce a "visible" feature or if the data are not correctly sampled because the pixels are too large.
- To improve the statistical precision of the signal, every effort should be made to count as many of the photons emerging from the specimen as possible:
  - — Reduce optical losses.
  - — Select the best type of PMT for the signal to be detected. New photodetectors may soon become available.
  - — Use photon-counting when appropriate.
  - — Check alignment often.
  - — Routinely check performance by making and analyzing images of test objects.
- To remove "single-pixel noise features" and provide the advantages of multi-voxel averaging, all Nyquist-sampled, 3D confocal results should be deconvolved (or at least 3D Gaussian filtered) before being viewed. Likewise, Nyquist-sampled 2D data should be filtered.
- The effects of image quantization should not be ignored. Only one pixel size (zoom setting?) is really optimal for each λ, NA, and lens magnification.
- Care should be taken to keep the pinhole diameter larger than that of the Airy disk at its half-power points (0.5 Airy unit). This setting will change with λ, NA, and lens magnification.
- Special circumstances may dictate breaking sampling and pinhole "rules," but they should still be recognized and acknowledged.
- It should be possible to operate the system both at its diffraction-limited resolution and at larger spot sizes.
- In laser-scanned instruments, imperfect scanning precision can introduce distortion.
- Fluorescence saturation places unexpected limits on the speed with which experiments can be performed with laser-scanning microscopes.

ACKNOWLEDGMENT

This chapter has benefited greatly from my conversations with Sam Wells (Cornell University) and Jon Art (University of Chicago). Thanks to Nick Doe (Bio-Rad, Hemel Hempstead, UK), who assisted with measurements of the digitization circuitry on the MRC-1000 and 1024. Figures 2.3, 2.4, and 2.5 were made while the author was a guest of the Key Centre for Microscopy and Microanalysis at the University of Sydney, in Sydney, Australia, supported by the Sabbatical Fund of the University of Wisconsin and by NSF grant # 9724515. I thank Guy Cox and other members of the Center for their patience, generosity, and assistance. The data for Figure 2.10 were obtained from Hamamatsu and data for Figure 2.14 were obtained Delft Photonics, Delft, NE. Figure 2.21 was kindly provided by Hans vander Voort of Scientific Volume Imaging, Delft, NE. Figure 2.18 was provided by P. Sims, University of Wisconsin, Zoology Department. Cheryle Hughes and Bill Feeney, artists in the Zoology Department, University of Wisconsin-Madison, made the other drawings. This work was supported by grant DIR-90–17534 from the US National Science Foundation (JP) and grant DRR-570 from the US National Institutes of Health to the Integrated Microscopy Laboratory, Madison, Wisconsin.

REFERENCES

Gelles, J., Schnapp, B.J., Steur, E., and Scheetz, M.P., 1988, Nanometer scale motion analysis of microtubule-based motor enzymes, *Proc. EMSA* 46:68–69.

Gunter, W.D. Jr., Grant, G.R., and Shaw, S., 1970, Optical devices to increase photocathode quantum efficiency. *Appl. Opt.* 9:251–257.

Oldenbourg, R., Terada, H., Tiberio, R., Inoué, S., 1993, Image sharpness and contrast transfer in coherent confocal microscopy. *J. Microscopy* 172:31–39.

Pawley, J.B., 1994, The sources of noise in three-dimensional microscopical data sets. In: *Three Dimensional Confocal Microscopy: Volume Investigation of Biological Specimens* (J. Stevens, ed.), Academic Press, New York, pp. 47–94.

Pawley, J.B., 2000, The 39 steps: A cautionary tale about "quantitative" 3D fluorescence microscopy, *BioTechniques*, 28:884.

Pawley, J.B., 2002, Limitations on optical sectioning in live-cell confocal microscopy, *Scanning*, 21:241–246.

Pawley, J.B., and Centonze, V., 1998, Practical laser-scanning confocal light microscopy: Obtaining optimal performance from your instrument. In: *Cell Biology: A Laboratory Handbook* (J.E. Celis, ed.) Academic Press, New York, pp. 149–169.

Pawley, J.B., Amos, W.B., Dixon, A., and Brelje, T.C., 1993a, Simultaneous, non-interfering, collection of optimal fluorescent and backscattered light signals on the MRC-500/600. *Proc. Microsc. Soc. Am.* 51:156–157.

Pawley, J., Blouke, M., and Janesick, J., 1996, The CCDiode: An optimal detector for laser confocal microscopes, *Proc. SPIE* 2655:125–129.

Pawley, J.B., Hasko, D., and Cleaver, J., 1993, A standard test and calibration specimen for confocal microscopy II. In: *Proceedings of the 1993 International Conference on Confocal Microscopy and 3-D Image Processing* (C.J.R. Sheppard, ed). Sydney, Australia, p. 35.

Pollock, H.N., and Chapman, D.S., 1993, Underground records of changing climate, *Sci. Am*. 68:44–50.

Rose, A., 1948, Television pickup tubes and the problem of noise. *Adv. Electron* 1:131.

Sandison, D.R., Piston, D.W., Williams, M., Webb, W.W., 1994, Quantitative comparison of background rejection, signal-to-noise ratio, and resolution in confocal and fullfield laser scanning microscopes. *Appl. Opt.* 1994: 3576–3588.

Sheppard, J.R., Gu, M., and Roy, M., 1992, Signal-to-noise ratio in confocal microscope systems. *J. Microsc.* 168:209–218.