The Diffraction Barrier in Optical Microscopy

The optical microscope has played a central role in helping to untangle the complex mysteries of biology ever since the seventeenth century when Dutch inventor Antoni van Leeuwenhoek and English scientist Robert Hooke first reported observations using single-lens and compound microscopes, respectively. Over the past three centuries, a vast number of technological developments and manufacturing breakthroughs have led to significantly advanced microscope designs featuring dramatically improved image quality with minimal aberration. However, despite the computer-aided optical design and automated grinding methodology utilized to fabricate modern lens components, glass-based microscopes are still hampered by an ultimate limit in optical resolution that is imposed by the diffraction of visible light wavefronts as they pass through the circular aperture at the rear focal plane of the objective. As a result, the highest achievable point-to-point resolution that can be obtained with an optical microscope is governed by a fundamental set of physical laws that cannot be easily overcome by rational alternations in objective lens or aperture design. These resolution limitations are often referred to as the diffraction barrier, which restricts the ability of optical instruments to distinguish between two objects separated by a lateral distance less than approximately half the wavelength of light used to image the specimen.
The process of diffraction involves the spreading of light waves when they interact with the intricate structures that compose a typical specimen. Due to the fact that most specimens observed in the microscope are composed of highly overlapping features that are best represented by multiple point sources of light, discussions of the microscope diffraction barrier center on describing the passage of wavefronts representing a single point source of light through the various optical elements and aperture diaphragms. As will be discussed below, the transmitted light or fluorescence emission wavefronts emanating from a point in the specimen plane of the microscope become diffracted at the edges of the objective aperture, effectively spreading the wavefronts to produce an image of the point source that is broadened into a diffraction pattern having a central disk of finite, but larger size than the original point. Therefore, due to diffraction of light, the image of a specimen never perfectly represents the real details present in the specimen because there is a lower limit below which the microscope optical system cannot resolve structural details.

In addition to the diffraction phenomenon that occurs with divergent light waves in optical instruments, the process of interference describes the recombination and summation of two or more superimposed wavefronts. Interference of light is perhaps the most ubiquitous phenomenon in optical microscopy and plays a central role in all aspects of image formation. In fluorescence or laser scanning confocal microscopy, the role of the objective is to focus the excitation light onto a focal point in order to ensure constructive interference of the focused wavefront at the specimen plane. In terms of this requirement, constructive interference (discussed below) ensures that the electric field vector of wavefronts incident from all available objective aperture angles resides in the same phase and therefore produces the smallest possible excitation spot.
Both interference and diffraction, which are actually manifestations of the same process, are responsible for creating a real image of the specimen at the intermediate image plane in a microscope. In brief, interference between two wavefronts occurs with addition to double the amplitude if the waves are perfectly in phase (constructive interference), but the waves cancel each other completely when out of phase by 180 degrees (termed destructive interference; however, most interference occurs somewhere in between). The photon energy inherent in a light wave is not itself doubled or annihilated when two waves interfere; rather this energy is channeled during diffraction and interference in directions that permit constructive interference. Therefore, interference and diffraction should be considered as phenomena involving the redistribution of light waves and photon energy.

A point object in a microscope, such as a fluorescent protein single molecule, generates an image at the intermediate plane that consists of a diffraction pattern created by the action of interference. When highly magnified, the diffraction pattern of the point object is observed to consist of a central spot (diffraction disk) surrounded by a series of diffraction rings (see Figure 1). In the nomenclature associated with diffraction theory, the bright central region is referred to as the zeroth-order diffraction spot while the rings are called the first, second, third, etc., order diffraction rings. When the microscope is properly focused, the intensity of light at the minima between the rings is zero. Combined, this point source diffraction pattern is referred to as an Airy disk (after Sir George B. Airy, a nineteenth century English astronomer). The size of the central spot in the Airy pattern is related to the wavelength of light and the aperture angle of the objective. For a microscope objective, the aperture angle is described by the numerical aperture (NA), which includes the term \( \sin(\theta) \), the half angle over which the objective can gather light from the specimen. In terms of resolution, the radius of the diffraction Airy disk in the lateral \((x,y)\) image plane is defined by the following formula:

\[
\text{Abbe Resolution}_{x,y} = \frac{\lambda}{2NA} \quad (1)
\]

where \( \lambda \) is the average wavelength of illumination in transmitted light or the excitation wavelength band in fluorescence. The objective numerical aperture (NA = \( n \cdot \sin(\theta) \)) is defined by the refractive index of the imaging medium (\( n \); usually air, water, glycerin, or oil) multiplied by the sine of the aperture angle (\( \sin(\theta) \)). As a result of this relationship, the size of the spot created by a point source decreases with decreasing wavelength and increasing numerical aperture, but always remains a disk of finite diameter. Thus, the image spot size produced by a 100x magnification objective having a numerical aperture of 0.90 in green light (550 nanometers) is approximately 30 micrometers, whereas the spot size produced by a 100x objective of numerical aperture 1.4 is approximately 20 nanometers, almost 50 percent smaller. The diffraction-limited resolution theory was advanced by German physicist Ernst Abbe in 1873 (see Equation (1)) and later refined by Lord Rayleigh in 1896 (Equation (3)) to quantitate the measure of separation necessary between two Airy patterns in order to distinguish them as separate entities.

\[
\text{Abbe Resolution}_z = \frac{2\lambda}{NA^2} \quad (2)
\]

According to Abbe's theory, images are composed from an array of diffraction-limited spots having varying intensity that overlap to produce the final result, as described above. Thus, the only mechanism for optimizing spatial resolution and image contrast is to minimize the size of the diffraction-limited spots by decreasing the imaging wavelength, increasing numerical aperture, or using an imaging medium having a larger refractive index. However, under ideal conditions with the most powerful objectives, lateral resolution is still limited to relatively modest levels approaching 200 to 250 nanometers (see Equation (1)) due to transmission characteristics of glass at wavelengths beneath 400 nanometers and the physical constraints on numerical aperture. In contrast, the axial dimension of the Airy disk forms an elliptical pattern that often referred to as the point-spread function (PSF). The elongated geometry of the point-spread function along the optical axis arises from the nature of the non-symmetrical wavefront that emerges from the microscope objective. Axial resolution in optical microscopy is even
worse than lateral resolution (as outlined in Equation (2)), on the order of 500 nanometers. When attempting to image highly convoluted features, such as cellular organelles, diffraction-limited resolution is manifested as poor axial sectioning capability and lowered contrast in the imaging plane. Furthermore, overall specimen contrast achieved in three-dimensional specimens is generally dominated by the relatively poor axial resolution that occurs due to out-of-focus light interference with the point-spread function.

Illustrated in Figure 1 is the effect of objective aperture angle on the size of a diffraction spot produced in a typical optical microscope. The point source and its conjugate \( P \) in the image plane where wavefronts converge and undergo constructive interference are illustrated for objectives having large (Figure 1(a)) and small (Figure 1(b)) numerical aperture. The point \( P_1 \) is moved laterally in the focal plane until destructive interference at a certain distance (dictated by the objective numerical aperture) defines the location of the first diffraction minimum and thus the radius of the diffraction spot. For the high resolution configuration in Figure 1(a), Points \( A \) and \( B \) in the wavefront produce a smaller spot size with 10 arbitrary units defining the imaged spot size. In contrast, for the lower resolution configuration presented in Figure 1(b), the reduced aperture angle increases the distance between \( A \) and \( B \) to 18 arbitrary units. In other words, light emitted by a fluorophore (the point source) is focused by the objective at the image plane where wavefronts traveling the same distance arrive at the image plane in phase and interfere constructively to produce a spot having high intensity. Destructive interference, leading to zero intensity, is generated by wavefronts that arrive one-half wavelength out of phase (see discussion above). Because the drop in intensity is gradual along the lateral axis of the spot, two point sources (or fluorescent molecules) closer together than the size of the spot will appear to be a single, larger spot and are unresolved.

As described above, the intensity distribution of an Airy disk in three dimensions is referred to as a point-spread function and completely describes the diffraction pattern of a point source of light (such as a single fluorophore) in the lateral \((x, y)\) and axial \((z)\) dimensions as modified by a diffraction-limited optical microscope. The size of the point spread function is determined by the wavelength of imaging light and the characteristics of the objective (numerical aperture) and the refractive index of the imaging medium. Resolution, in a practical sense, is often defined as the smallest separation distance between two point-like objects in which they can still be distinguished as individual emitters (and not amalgamated into a single spot). As a result, most resolution criteria (for example, the Rayleigh criterion, Sparrow limit, or the full width at half maximum; FWHM) are directly related to the properties and geometry of the point-spread function.

The Rayleigh Criterion for Lateral and Axial Resolution

![Figure 2](http://www.microscopyu.com/print/articles/superresolution/diffractionbarrier-print.html)

As described above, the intensity distribution of an Airy disk in three dimensions is referred to as a point-spread function and completely describes the diffraction pattern of a point source of light (such as a single fluorophore) in the lateral \((x, y)\) and axial \((z)\) dimensions as modified by a diffraction-limited optical microscope. The size of the point spread function is determined by the wavelength of imaging light and the characteristics of the objective (numerical aperture) and the refractive index of the imaging medium. Resolution, in a practical sense, is often defined as the smallest separation distance between two point-like objects in which they can still be distinguished as individual emitters (and not amalgamated into a single spot). As a result, most resolution criteria (for example, the Rayleigh criterion, Sparrow limit, or the full width at half maximum; FWHM) are directly related to the properties and geometry of the point-spread function.

\[
\text{Rayleigh Resolution}_{x,y} = 0.61\lambda/\text{NA} \ (3)
\]
According to the Rayleigh criterion, two point sources observed in the microscope are regarded as being resolved when the principal diffraction maximum (the central spot of the Airy disk; see Figure 2) from one of the point sources overlaps with the first minimum (dark region surrounding the central spot) of the Airy disk from the other point source. If the distance between the two Airy disks or point-spread functions is greater than this value, the two point sources are considered to be resolved (and can readily be distinguished). Otherwise, the Airy disks merge together and are considered not to be resolved. Stated in other terms, the Rayleigh criterion is satisfied when the distance between the images of two closely spaced point sources is approximately equal to the width of the point-spread function. In contrast, the Sparrow resolution limit is defined as the distance between two point sources where the images no longer have a dip in brightness between the central peaks, but rather exhibit constant brightness across the region between the peaks. The Sparrow resolution limit is closer to the Abbe value and approximately two-thirds (Equation (4)) of the Rayleigh resolution limit.

\[ \text{Sparrow Resolution}_{x,y} = 0.47\lambda/NA \] (4)

Presented in Figure 2 is a graphical representation of the Rayleigh criterion for both the lateral and axial dimensions of two closely positioned point sources. In Figure 2(a), the intensity of the point sources is represented by solid blue and dashed yellow curves. The total intensity generated by the combined point sources is represented by a red curve that is displaced along the ordinate for clarity. In order to distinguish between these point sources, the distance between the peaks should be sufficient to produce an intensity minimum that ranges between 20 and 30 percent of the peak intensity (Figure 2(a)). The same criterion applies to the axial dimension (Figure 2(b)). Note that the resolution (indicated in Figures 2(a) and 2(b) along the abscissa) is significantly lower along the \( z \) axis.

Although the Rayleigh criterion and similar measures are useful resolution gauges for observation of the specimen, there remain several shortcomings of such a definition for resolution. For example, in cases where the investigator is aware that two particles are merged to form a single point image, computer algorithms can be applied to discriminate between the particles to arbitrarily smaller distances. Determining the exact position of the two adjacent particles then becomes a question of experimental precision dictated by photon statistics rather than being described by the Rayleigh limit. Furthermore, resolution limits do not necessarily correspond to the level of detail that can be observed in images. While the Rayleigh limit is defined as the distance from the center of the first minimum of the point-spread function, this value can be rendered smaller by advanced optical systems or linear optics. Resolution criteria also do not rely on the fact that light is a diffracting wavefront that poses a finite limit to the level of detail that is actually contained within the waves.

The Abbe equation for resolution avoids the shortcomings of the Rayleigh criterion and Sparrow limit, but with a more indirect interpretation. The process of imaging a specimen in the microscope can be described by a convolution operation between the illumination and fluorescence emission (or transmitted light) point-spread functions. After being subjected to Fourier transformation (see Figure 3), objects observed in the microscope (whether they are periodic or not) can be uniquely described as a summation of numerous sinusoidal curves having different spatial frequencies. Note that the image of a specimen, present in all conjugate image planes, exists as the Fourier transform in the corresponding aperture planes where higher frequencies represent fine specimen detail and lower frequencies represent coarse details (Figure 3(a)). This point is illustrated with the waveform in objective rear aperture in Figure 3(b). The lower spatial frequencies reside near the center of the aperture, while the frequency progressively increases for regions approaching the edges of the aperture.
The concept of convolution in real space can be readily simplified by examining the equivalent operation in Fourier space. In the latter, the transformed object can be multiplied with the Fourier transform of the point-spread function to yield the Fourier transform of an ideal image lacking noise. After Fourier transformation, the point-spread function describes how efficiently each spatial frequency of the specimen is transferred to the final image. Thus, the Fourier-transformed point-spread function is referred to as the optical transfer function (OTF; see Figure 3(b)). The OTF defines the extent to which spatial frequencies containing information about the specimen are lost, retained, attenuated, or phase-shifted during the imaging process. Spatial frequency information that is lost during imaging cannot be recovered, so one of the primary goals for all forms of microscopy is to acquire the highest frequency range as possible for the specimen. The value of the OTF at each spatial frequency (measured in oscillations per meter) is a useful indicator to describe the contrast that a particular sinusoidal object feature achieves in the final image.

One of the important points to remember about the optical microscope is that the detection optical transfer function has a characteristic frequency that serves as a resolution "cut-off" border (the Abbe limiting frequency; see Figure 3(b)). Frequencies higher than the limiting value are not present in the image recorded by the microscope. The peak-to-peak distance for the highest spatial frequency able to pass through the objective (the value \( d \) for the green waveform in Figure 3(a)) is therefore commonly referred to as the Abbe limit, which is more formally defined as the smallest periodicity in a structure that can be detected in the final image. Due to the fact that a point source emits or transmits a wide range of spatial frequencies, the Abbe limit must also be present in the point-spread function spanning three dimensions.

Conclusions

A traditional widefield microscope generates an image of a point source by capturing the light in various locations in the objective and further processing the wavefronts as the pass through the optical train to finally interfere at the image plane. As a consequence of the reciprocity principle in optics, the Abbe limit in the lateral axis of the microscope corresponds to the maximum-to-maximum distance that can be obtained by interfering two waves at the most extreme angles captured by the objective. The Abbe resolution limit is attractive because it depends only on the maximal relative angle between different wavefronts leaving the specimen and captured by the objective. This limit therefore describes the smallest level of detail that can possibly be imaged, and that periodic structures have higher spatial frequency (shorter wavelengths) will not be transferred to the image.
Even in cases where an optical microscope is equipped with the highest available quality of lens elements, is perfectly aligned, and has the highest numerical aperture, the resolution remains limited to approximately half the wavelength of light in the best case scenario. In practice, the resolution typically achieved in routine imaging often does not reach the physical limit imposed by diffraction. This is due to the fact that optical inhomogeneities in the specimen can distort the phase of the excitation beam, leading to a focal volume that is significantly larger than the diffraction-limited idea. Additionally, resolution can also be compromised by the use of incompatible immersion oil, coverslips having a thickness outside the optimum range, and improperly adjusted correction collars.

Laser scanning confocal and multiphoton microscopy have been widely used to moderately enhance spatial resolution along both the lateral and axial axes, but the techniques remain limited in terms of achieving substantial improvement. The focused laser excitation coupled with pinhole-restricted detection in confocal microscopy can, in principle, improve the spatial resolution by a factor of 1.4, although this is only realized at a significant cost in signal-to-noise. Likewise, multiphoton fluorescence microscopy takes advantage of nonlinear absorption processes to reduce the effective size of the excitation point-spread function. Once again, however, the smaller and more refined point-spread function is counteracted by the necessity to use longer wavelength excitation light. As a result, rather than providing dramatic improvements to resolution, the primary advantage of confocal and multiphoton microscopy over traditional widefield techniques is the reduction of background signal originating from emission sources removed from the focal plane (out-of-focus light), which enables crisp optical sections to be obtained for three-dimensional volume-rendered imaging.

The resolution limits imposed by the physical laws that govern optical microscopy can be exceeded, however, by taking advantage of "loopholes" in the law that underscore the fact that the limitations are true only under certain assumptions. There exist three particularly important assumptions that are brought into play during assessment of resolution criteria, including the conventional geometry in which light is gathered by the objective, the uniformity of excitation light throughout the specimen, and the linear characteristics (absorption and emission) of fluorescence involving a single photon. In brief, resolution can be enhanced by gathering light over a larger set of angles around the specimen or using excitation light that varies with position. The diffraction barrier can also be broken using fluorescence processes that involve two or more photons in a non-linear manner.

Contributing Authors

Joel S. Silfies and Stanley A. Schwartz - Nikon Instruments, Inc., 1300 Walt Whitman Road, Melville, New York, 11747.

Michael W. Davidson - National High Magnetic Field Laboratory, 1800 East Paul Dirac Dr., The Florida State University, Tallahassee, Florida, 32310.