4

Image Formation and Interpretation

4.1. The Story So Far

Scanning electron microscopy is a technique in which images form the major avenue of information to the user. A great measure of the enormous popularity of the SEM arises from the ease with which useful images can be obtained. Modern instruments incorporate many computer-controlled, automatic features that permit even a new user to rapidly obtain images that contain fine detail and features that are readily visible, even at scanning rates up to that of television (TV) display. Although such automatic "computer-aided" microscopy provides a powerful tool capable of solving many problems, there will always remain a class of problems for which the general optimized solution may not be sufficient. For these problems, the careful microscopist must be aware of the consequences of the choices for the beam parameters (Chapter 2), the range of electron–specimen interactions that sample the specimen properties (Chapter 3), and finally, the measurement of those electron signals with appropriate detectors, to be discussed in this chapter. With such a systematic approach, an advanced practice of SEM can be achieved that will significantly expand the range of application to include many difficult imaging problems.

4.2. The Basic SEM Imaging Process

This chapter will consider the formation and interpretation of SEM images. One of the most surprising aspects of scanning electron microscopy is the apparent ease with which SEM images of three-dimensional objects can be interpreted by any observer, including young children, with no prior knowledge of the SEM instrument or the arcane subject of electron–solid interactions. This aspect of the SEM is often taken for granted, and yet it is one of the most important reasons for the great utility and wide acceptance of the instrument. SEM images are routinely presented in textbooks, popular scientific articles, and even advertisements with little or no mention of

J. I. Goldstein et al., *Scanning Electron Microscopy and X-ray Microanalysis*
© Springer Science+Business Media New York 2003

the type of microscopy employed in preparing the image, of the complex way in which the image was constructed, or with any guidance to aid the viewer's interpretation. It can safely be assumed that the reader will automatically perceive the true topographic nature of the specimen without any special knowledge concerning the origin of the image. For this to be true, the SEM imaging process must in some sense mimic the natural experience of human observers in visualizing the world around them. Such a situation is quite surprising in view of the unusual way in which the image is formed, which seems to differ greatly from normal human experience with images formed by light and viewed by the eye. In the SEM, high-energy electrons are focused into a fine beam, which is scanned across the surface of the specimen. Complex interactions of the beam electrons with the atoms of the specimen produce a wide variety of radiation products: backscattered electrons, secondary electrons, absorbed electrons, characteristic and continuum x-rays, etc. A sample of this radiation is collected by a detector, most commonly the Everhart–Thornley detector (positively biased scintillator–photomultiplier), and the collected signal is amplified and displayed on a cathode ray tube (CRT) or television screen scanned in synchronism with the scan on the specimen. Despite this complicated and unusual path to the formation of the image, the result is somehow straightforward to interpret, at least for the important class of objects that can be described as topographic (three-dimensional) in nature and are viewed at low to intermediate magnifications (up to $10{,}000\times$).

Although this "entry-level" approach to scanning electron microscopy will suffice for some problems, the experienced microscopist will frequently discover situations where it is necessary to make use of advanced concepts of image interpretation. Such situations are encountered, for example, with complex topography, especially when the specimen or feature has little familiarity; in the case of high-magnification images ($>10{,}000\times$), where questions of resolution limits and the delocalization of the imaging signals must be addressed; and with special types of specimens and contrast mechanisms, such as voltage contrast or magnetic contrast. The intellectual tools which the microscopist must bring to make a proper interpretation of an imaging problem include knowledge of electron optics, beam–specimen interactions, detector characteristics, signal quality/feature visibility relations, and signal/image processing. The first two topics, electron optics and beam–specimen interactions, were covered in detail in Chapters 2 and 3; the latter three topics will be the subject of this chapter.

Questions which constantly arise in microscopy studies include: What types of features can be seen in an image (i.e., what information about the specimen can be observed in an SEM image), and how small a region showing a particular characteristic can be detected (resolution)? These questions of image information and resolution are closely related, and in this chapter equations will be developed that permit the microscopist to make an estimate of what can be seen when specific operating conditions are chosen. SEM images are often perceived to be so spectacular that the impression arises that the technique is not subject to constraints. If a specimen contains a class of features by design and synthesis, then the SEM must be able to

image them. As a corollary, the absence of features from an image must guarantee their absence from that region of the specimen under study. In fact, the SEM, like all imaging devices, is constrained by fundamental statistical fluctuations in the signal, leading to limits of visibility. (The critical distinction between "visibility" and "resolution" will be made below.) For any choice of imaging conditions, there will always be a level of contrast below which features are not visible. Fortunately, these inevitable limitations on the performance of the SEM can be estimated numerically, and a strategy to improve visibility for a particular case can be developed. Even with a careful imaging strategy, however, it may not be possible to establish SEM visibility for some objects under any practical imaging conditions within the operating parameters of a particular SEM. Recognizing that such limitations exist and understanding how to optimize instrument parameters to achieve the best performance is critical to advanced use of the SEM.

Digital imaging has now been established as the dominant form of SEM image storage and manipulation. Indeed, for many users, SEM images may never appear in the form of photographs, but rather images will be recorded electronically, stored as a computer file, examined, modified, and measured on a computer display, and assembled into text reports or presentation graphics, and only in the final presentation to a client may a "hard copy" of the document be produced. Some reference will still be made to analog systems because SEMs often have a useful life of 20 years or more, so that readers of this book will likely use equipment spanning full-analog to full-digital control.

4.2.1. Scanning Action

The detailed description of the electron optics of the SEM, as given in Chapter 2, can be briefly summarized by considering the degree of control which the microscopist has in choosing the parameters that characterize the beam as focused to form a probe at the specimen plane: probe current i_B (pA to μA); probe diameter d, which ranges from a minimum of 1 nm in the highest resolution instruments (typically limited to 5 nm in conventional instruments) to a maximum of 1 μm; and probe divergence α (10^{-4}–10^{-2} rad). Of course, these three parameters cannot be selected independently because they are interrelated, as described by the brightness equation. The brightness equation gives the theoretical limit to the beam performance; lens defects (aberrations) result in beams which are either larger than expected for a given beam current and divergence or else lower in current for a specified diameter and divergence.

The SEM imaging process is illustrated schematically in Fig. 4.1. The electron beam, defined by the parameters d, i_B, and α, exits the electron column, enters the specimen chamber, and strikes the specimen at a single location on the optic axis of the column. The beam electrons interact elastically and inelastically with the specimen, forming the limiting interaction volume from which the various types of radiation emerge, including backscattered, secondary, and absorbed electrons, characteristic and bremsstrahlung x-rays, cathodoluminescence radiation (long-wavelength

photons in the ultraviolet, visible, and infrared ranges) in some materials (compounds, minerals, semiconductors, etc.), and finally, in certain semiconductor materials and structures, beam-induced currents. By recording the magnitude of these signals with suitable detectors, a measure of certain properties of the specimen, such as topography, composition, etc., can be made at the single location where the electron beam strikes. To study more than a single location and eventually construct an image, the beam must be moved from place to place by means of a scanning system, as illustrated in Fig. 4.1. This "scanning action" is usually accomplished by energizing electromagnetic coils arranged in sets consisting of two pairs, one pair each for deflection in the x and y directions. (The convention will be followed that an xyz coordinate system will be established with the z direction parallel to the optic axis of the microscope and the x and y axes defining a plane perpendicular to the beam.) A typical double-deflection scanning system, as shown in Fig. 4.1, has two sets of electromagnetic scan coils, located in the bore of the final (objective) lens. As controlled by the scan generator, the upper coils act to drive the beam off-axis, and the lower coils deflect the beam to cross the optic axis again, with this second crossing of the optic axis taking place in the final (beam-defining) aperture. This system has the advantage that by placing the scan coils within the lens, the region below is kept open and the specimen can be placed relatively close to the lens at a short working distance. A short working distance means the objective lens will be relatively more highly excited, which minimizes some of the lens aberrations, improving the beam performance. By locating the beam-defining aperture at the second crossover (pivot point), large scan angles necessary for low magnifications can be obtained without cutting off the field of view on the aperture (Oatley, 1972).

Scanning action is produced by altering the strength of the current in the scan coils as a function of time, so that the beam is moved through a sequence of positions on the specimen (e.g., locations 1, 2, 3, 4, etc., in Fig. 4.1) and the detector(s) samples the electron–specimen interaction at a defined sequence of points. In an analog scanning system, the beam is moved continuously, with a rapid scan along the x axis (the line scan)

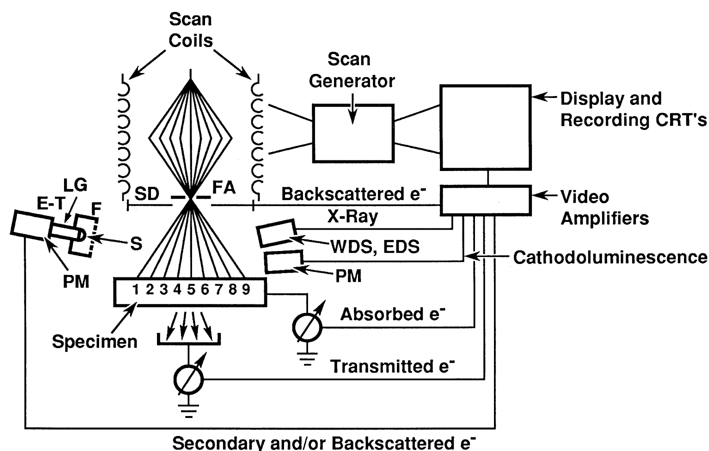

Figure 4.1. Schematic illustration of the scanning system of the SEM. Abbreviations: FA, final aperture; SD, solid-state backscattered electron detector; EDS, energy-dispersive x-ray spectrometer; WDS, wavelength-dispersive x-ray spectrometer; CRT, cathode ray tube; E–T, Everhart–Thornley secondary/backscattered electron detector, consisting of F, Faraday cage; S, scintillator; LG, light guide; and PM, photomultiplier. Successive beam positions are indicated by the numbered rays of a scanning sequence.

and a slow scan, typically at 1/500 of the line rate, at right angles along the y axis (the frame scan), so that a good approximation to an orthogonal scan is produced. In digital scanning systems, only discrete beam locations are allowed. The beam is addressed to a particular location (x, y) in a matrix, remains there for a fixed time (the dwell time), and is then rapidly moved to the next point. The beam current may actually be turned off during the movement between pixels by an electrostatic blanking system placed just below the gun. The time to shift the beam between points is negligible compared to the dwell time. For display, the image is constructed on a cathode ray tube (CRT), which is scanned in synchronism with the scan on the specimen, controlled by the same scan generator. The signal derived from one of the detectors is amplified and used to control the brightness of the CRT ("intensity modulation"), often with some form of signal processing applied to enhance the visibility of the features of interest, a topic to be covered in detail below. At the resolution at which a human observer usually examines or photographs the CRT image, there is no effective difference between images prepared with analog and high-density (e.g., 1024×1024, or higher) digital scans; both types of images appear continuous in nature. An added benefit of the digital scan is that the numerical address of the beam location is accurately and reproducibly known, and therefore the information on the electron interaction can be encoded in the form (x, y, I), where x and y give the address and I is the intensity. Such encoding provides the basis for digital image processing, to be discussed later in this chapter.

4.2.2. Image Construction (Mapping)

The information flow from the scanning electron microscope consists of the scan location in x–y space and a corresponding set of intensities from the set of detectors (backscattered electron, secondary electron, transmitted electron, specimen current, x-ray, cathodoluminescence, etc.) that monitor the beam–specimen interactions, all of which can be monitored in parallel. This information can be displayed to the observer in two principal ways: line scans and area scans.

4.2.2.1. Line Scans

In the line scan mode, the beam is scanned along a single vector on the specimen, e.g., in the x or the y direction. The same scan generator signal is used to drive the horizontal scan of the CRT. The resulting synchronous line scan on the specimen and the CRT produces a one-to-one correspondence between a series of points in the "specimen space" and on the CRT or "display space." In such a line scan displayed on a CRT, the horizontal position is related to distance along a particular line on the specimen. The effective magnification factor M between the specimen space and the CRT space is given by the ratio of the lengths of the scans:

$$M = L_{\mathrm{CRT}}/L_{\mathrm{specimen}}, \tag{4.1}$$

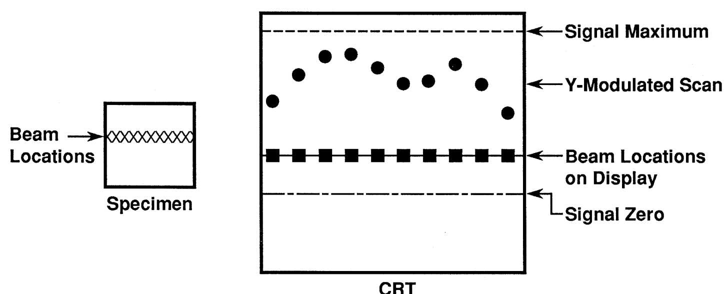

Figure 4.2. SEM line scan display: The beam is scanned along a locus of points on the specimen, and the beam on the CRT is scanned in a similar locus along the horizontal (x axis). A signal measured from one of the detectors (backscattered electrons, x-rays, etc.) is used to adjust the vertical position (y axis). For quantitative measurements, the trace corresponding to the locus of points should be recorded, along with traces corresponding to the signal zero and maximum values.

where L denotes the length of the scans. The numerical value of the magnification reported on the alphanumeric display typically refers to the final image format recorded on the SEM photographic system. The magnification value is thus appropriate to the size of the printed photographic image (often a 10×12.5-cm format), and it may underestimate the true magnification of the large viewing CRT by a large factor, often two or three times. On the other hand, the calibrated scale bar, usually expressed in micrometers or nanometers, refers to a specific distance in the scanned image. This value scales appropriately with any reproduction of the image with any device, including projection on a large screen, and is thus the meaningful dimensional quantity.

The intensity measured during the line scan by one of the detectors, for example, the signal from a backscattered electron detector or from an x-ray spectrometer, can be used to adjust the y deflection ("y modulation") of the CRT, which produces a trace such as that illustrated in Fig. 4.2. For example, if the signal is the characteristic x-ray intensity, the y deflection of the CRT will be proportional to the amount of a particular element present at each location. Although line scans are the simplest type of scanning action, they are almost never used alone and are typically superimposed on the more familiar two-dimensional SEM image. Line scans are extremely useful for diagnostic work, where the signal profile across a feature is needed. Line scans can be used to display small signal changes that can easily be detected as offsets in y modulation, but which would be difficult to discern in a conventional intensity-modulated area image. In recording a line scan, as shown in Fig. 4.2, it is important to record (1) the scan locus (i.e., the location in the area raster at which the line scan is taken), (2) the y-modulated scan of the signal, (3) the signal zero level, and (4) the signal maximum. This information is critical if any quantitative use is to be made of the line scan data.

4.2.2.2. Image (Area) Scanning

To form the SEM image with which we are most familiar, the beam is scanned on the specimen in an x–y pattern while the CRT is scanned

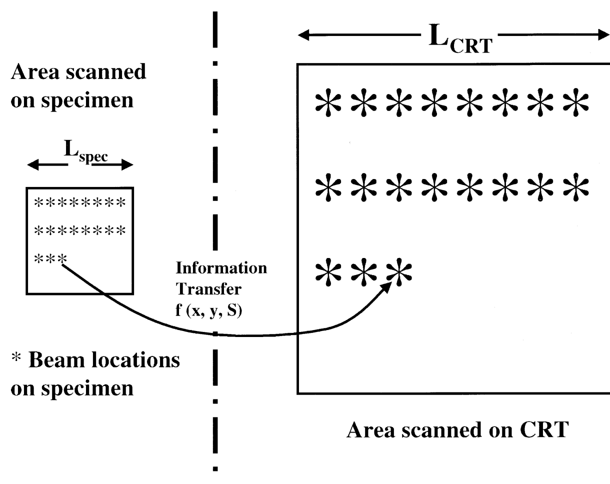

Figure 4.3. The principle of image display by area scanning. A correspondence is established between a set of locations on the specimen and a set on the CRT. Magnification = $L_{\mathrm{CRT}}/L_{\mathrm{specimen}}$.

in the same x–y pattern, as illustrated in Fig. 4.3. Again, a one-to-one correspondence is established between the set of beam locations on the specimen and the points on the CRT, and the linear magnification of the image is given by Eq. (4.1). To display the beam–specimen interaction information, the signal intensity S derived from a detector is used to adjust the brightness of the spot on the CRT, a process referred to as "intensity modulation," shown schematically in Fig. 4.4. Typically, an SEM contains one or more slow-scan CRT visual displays, a photographic CRT, and a television-rate display, or all of these displays may be replaced by a single computer monitor on a fully digital instrument. The television rate display operates with a rapid decay phosphor that can respond to scan speeds of

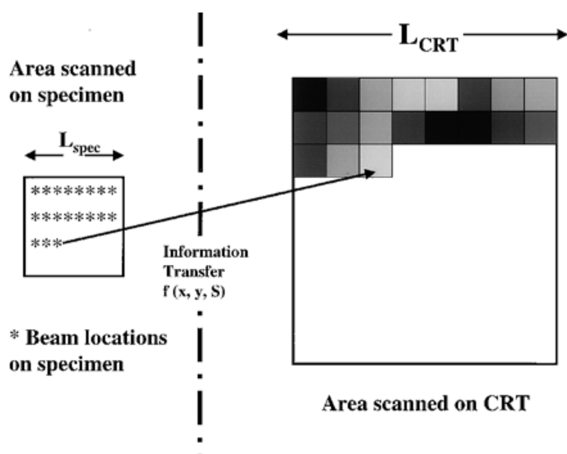

Figure 4.4. The principle of intensity modulation used to display the magnitude of the signal produced by electron–specimen interaction at the locations scanned in Fig. 4.3. Black represents low intensity; gray, intermediate intensity; white, high intensity.

1/30 s per frame, which the observer perceives as "flicker-free." The visual display CRT contains a phosphor with a relatively long persistence (usually 1 s or more) to permit the observer to view the image with frame scan speeds in the range 0.5–5 s. A high-resolution (small spot size) CRT with a short persistence phosphor is used for photographic recording, almost invariably with rapid-development film.

The creation of an SEM image consists in constructing an intensity map in the analog or digital domain. Unlike an optical microscope or transmission electron microscope, no true image exists in the SEM. In a true image, there are actual ray paths connecting points on the specimen to corresponding points in the image as displayed on a screen or detected by the eye or film. It is not possible to place a sheet of film anywhere in the SEM chamber and record an image, as it is in a light-optical microscope or a transmission electron microscope. In the SEM, image formation is produced by a mapping operation that collects information from the specimen space and passes the information to the display space. That such an abstract process of creating an image can produce a class of images of topographic objects which are readily understandable is a considerable surprise, fortunately a pleasant and extremely useful one.

The information contained in the image will convey the true shape of the object if the specimen and CRT scans are synchronous and are constructed to maintain the geometric relationship of any arbitrarily chosen set of points on the specimen and on the CRT. When this condition is satisfied, as shown in Fig. 4.5, a triangle on the specimen remains a triangle of the same shape on the CRT. The influence of projection distortion for three-dimensional objects and various scan defects on the shape of objects in SEM images will be described below. The nature of the intensity variations that produce the gray-level shading of various objects in the image will be discussed below in Section 4.4.

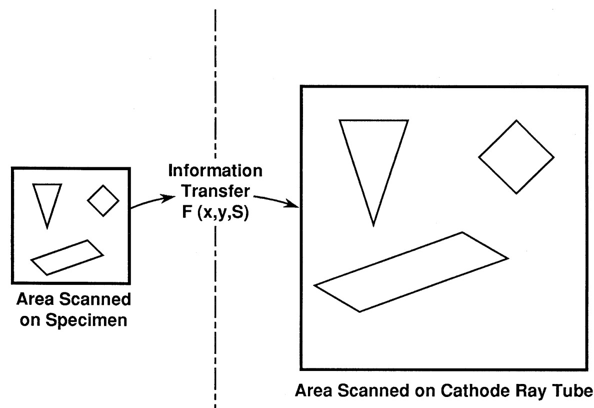

Figure 4.5. Shape correspondence between the scanned field on the specimen and the display on the CRT. In a perfect scan system, the shapes in the plane of the scan in the microscope are transferred without distortion to the CRT display.

4.2.2.3. Digital Imaging: Collection and Display

A digital image consists of a numerical array (x, y, S) stored in a computer memory, where each entry consists of three values, two for the position and one for the signal intensity. To create a digital image, the x–y pattern of scanned locations is made to correspond to a matrix array in a computer memory or in a framestore (a special dedicated computer memory board consisting of an array of memory registers adequate to record one or more complete images). Each scan position is generated as a digital address and converted with a digital-to-analog converter (DAC) into a voltage to drive the scan circuit. The number of discrete beam positions, the picture elements or "pixels," is specified as the digital resolution, for example, 512×512, 1024×1024, etc. The pattern does not have to be square, so that rectangular or other scan shapes can be generated, but it is important to maintain symmetric (i.e., square) pixels to avoid introducing a distortion into the digital image. When the beam has been addressed to a location on the specimen, the analog signal intensity is measured by the detector(s) (SE, BSE, x-rays, etc.), integrating for the length of the pixel dwell. To ensure proper matching of the measured signal to the processing chain, the analog signal amplifier is first adjusted with a separate oscilloscope or an equivalent y-modulated scan on the CRT to bring the range of the signal over the entire frame scan into the input acceptance range of the analog-to-digital converter (ADC). The voltage signal produced by the signal amplifier is digitized to a value S for the computer and stored as a discrete numerical value in the corresponding register (x, y, S). Typically, the intensity is digitized into a range of 8 bits, which gives $2^n = 2^8 = 256$ discrete levels.

With only 256 levels in the digital representation of the signal, it is only possible to recognize changes at the level of $1/256 = 0.4\%$ with subsequent digital functions, given that the input signal spans the entire range of the DAC. Generally, to avoid the possibility of saturating the response of the DAC, the input signal range is adjusted more conservatively so that the signal does not reach the 0 or 255 digital levels. Note that once the digitization has been performed, it is not possible to recover information not recorded in the original digitization step. For many applications, it is desirable to provide more discrete digital levels for subsequent processing, so the original signal can be digitized to 12 bits (4096 levels), 16 bits (65,536 levels), or even higher. This additional signal resolution is gained by digitizing for longer integration times or sampling repeatedly at each pixel and averaging. Generally, we do not wish to record all images at the maximum digital resolution available. The penalty for recording large digital images is the need to provide mass storage to save the images. A $1024 \times 1024 \times 256$-level image requires over 1 megabyte (1 MB) of mass storage, whereas a $2048 \times 2048 \times 256$-level image requires 4 MB. The choice of the digital x–y resolution should be made to slightly oversample the image in terms of the picture element (see Section 4.2.4), and the digital resolution of the intensity depends on the need for subsequent contrast manipulation.

The digital image is viewed by converting the numerical values stored in memory back into an analog signal (digital-to-analog conversion) for

display on a CRT. The density and speed of computer memory are such that large image arrays (512 × 512, or larger) can be stored digitally and read out in "real time," producing a "flicker-free" image. Such digitally stored images have many advantages. One special advantage previously mentioned is the possibility of minimizing the radiation dose necessary to view the specimen by viewing the digitally stored image until a fresh image field is needed. To continuously view an image with an analog SEM system, the beam must be constantly scanned on the specimen in order to refresh the display. The information in each scan is discarded unless recorded by film or videotape. Radiation damage or beam-induced contamination generally scales with the total dose, so that analog imaging with repeated scans exposes the specimen to possible degradation. Digital images are stored in memory, and can be repeatedly read out to the analog display for inspection of that particular frame. Thus, the beam can be scanned for a single frame on the specimen to accumulate data in a memory and then "blanked" or deflected off the specimen, generally into a Faraday cup for measurement of the current as a stability check, while the digital memory is repeatedly read to refresh a CRT display. As described in detail below, image processing functions can be applied to this stored image to enhance the visibility of objects of interest. Once the displayed image has been satisfactorily adjusted, the resulting image can be transmitted to the photographic recording CRT or other form of "photoquality" hard copy output device.

4.2.3. Magnification

Changing magnification in the SEM image is accomplished by adjusting the length of the scan on the specimen for a constant length of scan on the CRT, as described by Eq. (4.1), which gives the linear magnification of the image. There are several important points about the SEM magnification process:

1. *Numerical value of magnification:* Because the maximum CRT scan length is fixed to the full dimension L of the tube, for example, 10 cm, an increase in magnification is obtained by *reducing* the length l of the scan on the specimen. Table 4.1 gives the size of the area sampled on the specimen as a function of magnification.

Table 4.1. Area Sampled as a Function of Magnification

Magnification[a]	Area on sample
10×	$(1\ \text{cm})^2 = 100\ \text{mm}^2$
100×	$(1\ \text{mm})^2 = 1\ \text{mm}^2$
1,000×	$(100\ \mu\text{m})^2 = 0.01\ \text{mm}^2$
10,000×	$(10\ \mu\text{m})^2 = 10^{-4}\ \text{mm}^2$
100,000×	$(1\ \mu\text{m})^2 = 10^{-6}\ \text{mm}^2$
1,000,000×	$(100\ \text{nm})^2 = 10^{-8}\ \text{mm}^2$

[a] Assumes magnification is relative to a CRT screen that measures 10 cm ×10 cm.