

Tomographic Reconstruction in Nuclear Medicine

A basic problem in conventional radionuclide imaging is that the images obtained are two-dimensional (2-D) projections of three-dimensional (3-D) source distributions. Images of structures at one depth in the patient thus are obscured by superimposed images of overlying and underlying structures. One solution is to obtain projection images from different angles around the body (e.g., posterior, anterior, lateral, and oblique views). The person interpreting the images then must sort out the structures from the different views mentally to decide the true 3-D nature of the distribution. This approach is only partially successful; it is difficult to apply to complex distributions with many overlapping structures. Also, deep-lying organs may have overlying structures from all projection angles.

An alternative approach is *tomographic imaging*. Tomographic images are 2-D representations of structures lying within a selected plane in a 3-D object. Modern *computed tomography* (CT) techniques, including positron emission tomography (PET), single photon emission tomography (SPECT), and x-ray CT, use detector systems placed or rotated around the object so that many different angular views (also known as *projections*) of the object are obtained. Mathematical algorithms then are used to *reconstruct* images of selected planes within the object from these projection data. Reconstruction of images from multiple projections of the detected emissions from radionuclides within the body is known as *emission computed tomography* (ECT). Reconstruction of images from transmitted emissions from an external source (e.g., an x-ray tube) is known as *transmission computed tomography* (TCT or, usually, just

CT, See Chapter 19, Section B). The mathematical basis is the same for ECT and TCT, although there are obviously differences in details of implementation.

ECT produces images in which the activity from overlying (or adjacent) cross-sectional planes is eliminated from the image. This results in a significant improvement in contrast-to-noise ratio (CNR), as already has been illustrated in Figure 15-11. Another advantage of SPECT and PET over planar nuclear medicine imaging is that they are capable of providing more accurate quantitation of activity at specific locations within the body. This is put to advantage in tracer kinetic studies (Chapter 21).

The mathematics underlying reconstruction tomography was first published by Johann Radon in 1917, but it was not until the 1950s and 1960s that work in radioastronomy and chemistry resulted in practical applications. The development of x-ray CT in the early 1970s initiated application of these principles for image reconstruction in medical imaging. An interesting historical perspective on the origins and development of tomographic image reconstruction techniques is presented in reference 1.

Instrumentation for SPECT imaging is discussed in Chapter 17 and instrumentation for PET is discussed in Chapter 18. Although the instruments differ, the same mathematics can be used to reconstruct SPECT or PET images. In this chapter, we focus on the basic principles of reconstructing tomographic images from multiple projections. A detailed mathematical treatment of image reconstruction is beyond the scope of this text. The reader is referred to references 2 to 4 for more detailed accounts.

A. GENERAL CONCEPTS, NOTATION, AND TERMINOLOGY

We assume initially that data are collected with a standard gamma camera fitted with a conventional parallel-hole collimator. (Applications involving other types of collimators are discussed in Section E.) To simplify the analysis, several assumptions are made. We consider only a narrow cross-section across the detector. The collimated detector is assumed to accept radiation only from a thin slice directly perpendicular to the face of the detector. This reduces the analysis to that of a 1-D detector, as shown in Figure 16-1. Each collimator hole is assumed to accept radiation only from a narrow cylinder defined by the geometric extension of the hole in front of the collimator. This cylinder defines the *line of response* for the collimator hole. For further simplification, we ignore the effects of attenuation and scatter and assume that the counts recorded for each collimator hole are proportional to the total radioactivity contained

within its line of response. The measured quantity (in this case, counts recorded or radioactive content) sometimes is referred to as the *line integral* for the line of response. A full set of line integrals recorded across the detector is called a *projection*, or a *projection profile*, as illustrated in Figure 16-1.

Obviously, the assumptions noted earlier are not totally valid. Some of the effects of the inaccuracies of these assumptions are discussed in Chapter 17, Section B and in Chapter 18, Section D.

A typical SPECT camera is mounted on a gantry so that the detector can record projections from many angles around the body. PET systems generally use stationary arrays of detector elements arranged in a ring or hexagonal pattern around the body. In either case, the detectors acquire a set of projections at equally spaced angular intervals. In reconstruction tomography, mathematical algorithms are used to relate the projection data to the 2-D distribution of activity within the projected slice. A schematic illustration of the data acquisition process is shown in

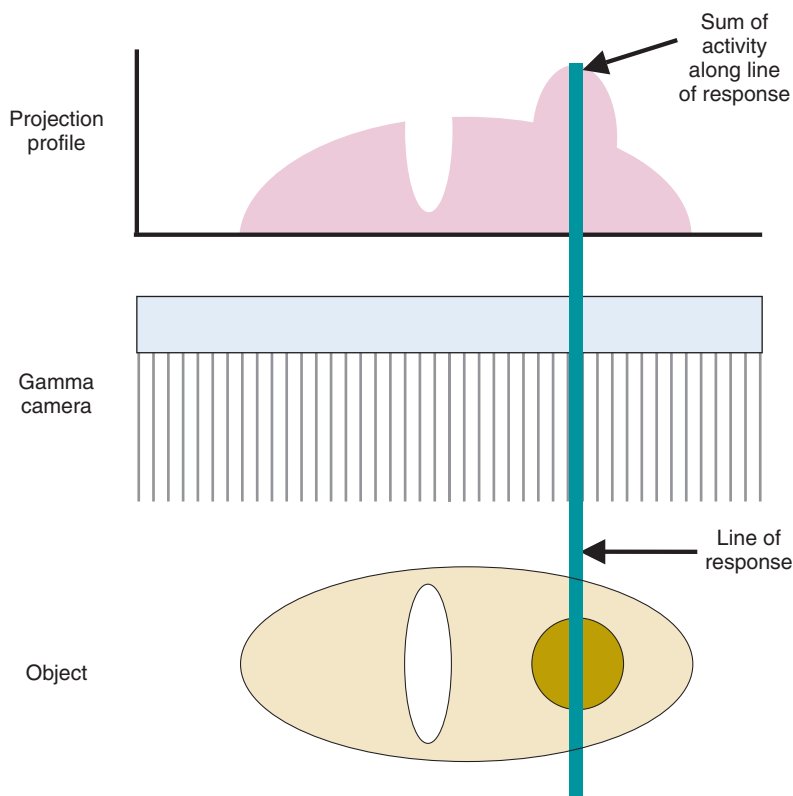


FIGURE 16-1 Cross-section of the response characteristics of an idealized gamma camera. Each collimator hole views the radioactivity within a cylinder perpendicular to the face of the gamma camera, called its *line of response*. Under idealized conditions (such as no attenuation or scatter) the signal recorded by the detector at that point reflects the sum of activity within the line of response. For a row of holes across the detector, the gamma camera generates a projection profile as shown. The projection profiles provide the data from which the image is reconstructed.

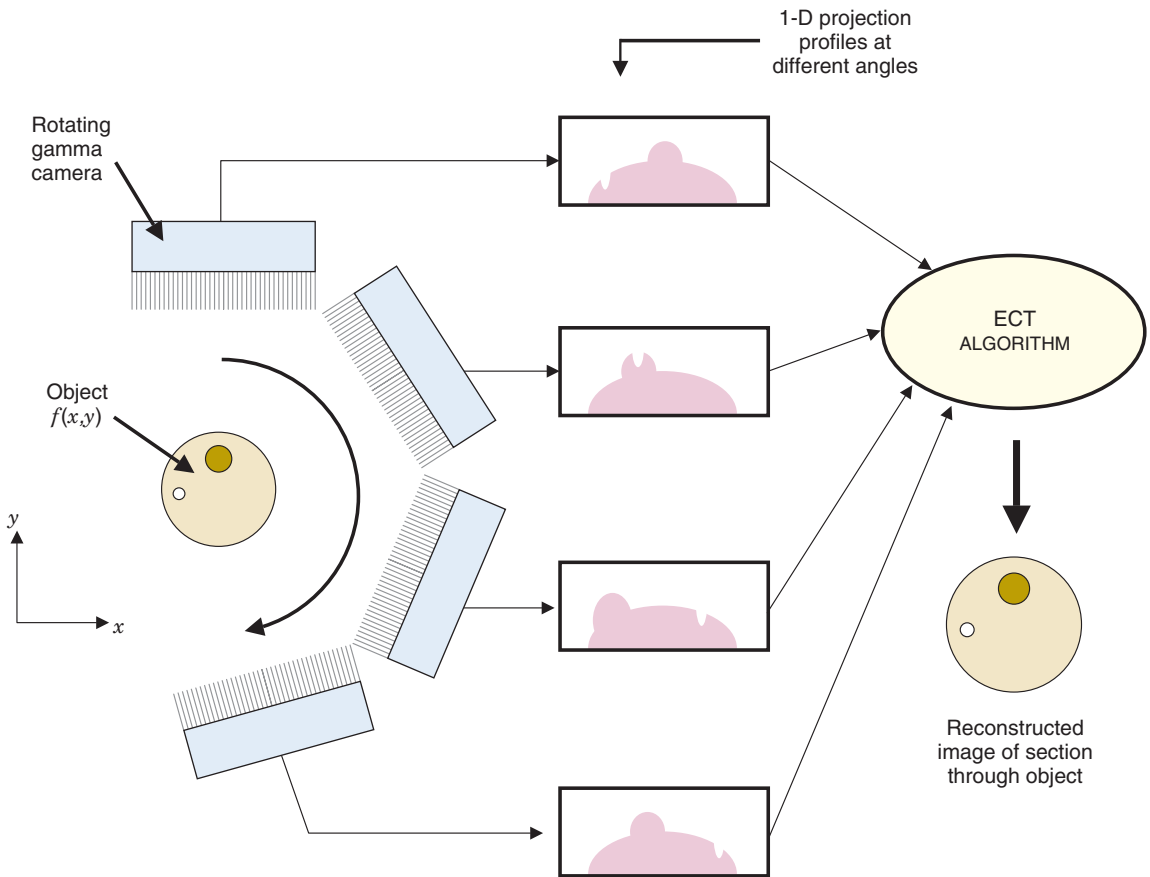


FIGURE 16-2 Rotating the gamma camera around the object provides a set of one-dimensional projection profiles for a two-dimensional object, which are used to calculate the two-dimensional distribution of radioactivity in the object. ECT, emission computed tomography.

Figure 16-2. Note that the data collected correspond to a slice through the object perpendicular to the bed and that this is called the *transverse* or *transaxial* direction. The direction along the axis of the bed, which defines the location of the slice, is known as the *axial* direction.

We assume that N projections are recorded at equally spaced angles between 0 and 180 degrees. Under the idealized conditions assumed here, the projection profile recorded at a rotation angle of $(180 + \phi)$ degrees would be the same (apart from a left-right reversal) as the profile recorded at ϕ degrees. Thus the data recorded between 180 and 360 degrees would be redundant; however, for practical reasons (e.g., attenuation), SPECT data often are acquired for a full 360-degree rotation. This is discussed further in Chapter 17.

For purposes of analysis, it is convenient to introduce a new coordinate system that is stationary with respect to the gamma camera detector. This is denoted as the (r,s)

coordinate system and is illustrated in Figure 16-3. If the camera is rotated by an angle ϕ with respect to the (x,y) coordinate system of the scanned object, the equations for transformation from (x,y) to (r,s) coordinates can be derived from the principle of similar triangles and are given by

$$r = x \cos \phi + y \sin \phi \quad (16-1)$$

and

$$s = y \cos \phi - x \sin \phi \quad (16-2)$$

These equations can be used to determine how radioactivity at a location (x,y) in the object contributes to the signal recorded at location r in the projection acquired at rotation angle ϕ .

One commonly used way to display a full set of projection data is in the form of a 2-D matrix $p(r,\phi)$. A representation of this matrix, generically known as a *sinogram*, is shown for

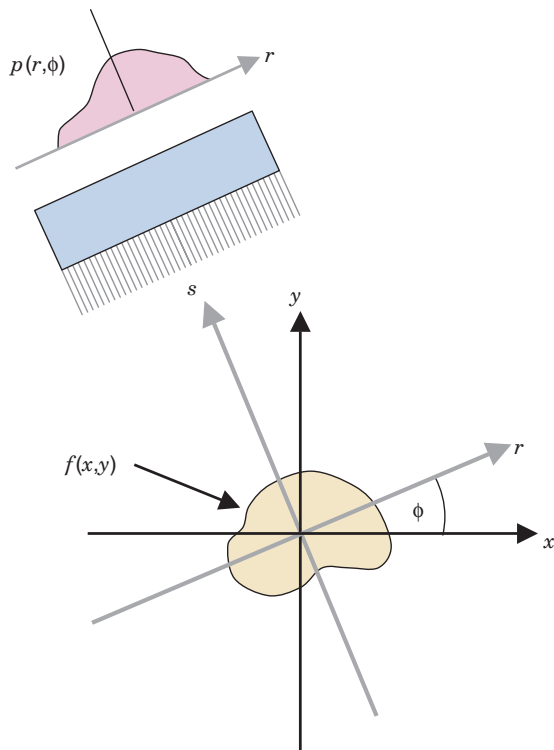


FIGURE 16-3 The (r, s) coordinate system is rotated by projection angle ϕ with respect to the (x, y) coordinate system of the object and is fixed with respect to the gamma camera.

a simple point-source object in Figure 16-4. Each row across the matrix represents an intensity display across a single projection. The successive rows from top to bottom represent successive projection angles. The name *sinogram* arises from the fact that the path of a point object located at a specific (x, y) location in the object traces out a sinusoidal path down the matrix. (This also can be deduced from Equations 16-1 and 16-2.) The sinogram provides a convenient way to represent the full set of data acquired during a scan and can be useful for determining the causes of artifacts in SPECT or PET images.

B. BACKPROJECTION AND FOURIER-BASED TECHNIQUES

1. Simple Backprojection

The general goal of reconstruction tomography is to generate a 2-D cross-sectional image of activity from a slice within the object, $f(x, y)$, using the sinogram, or set of projection profiles, obtained for that slice. In practice, a set of projection profiles, $p(r, \phi_i)$, is acquired at discrete angles, ϕ_i , and each profile is sampled at discrete intervals along r . The image is reconstructed on a 2-D matrix of discrete

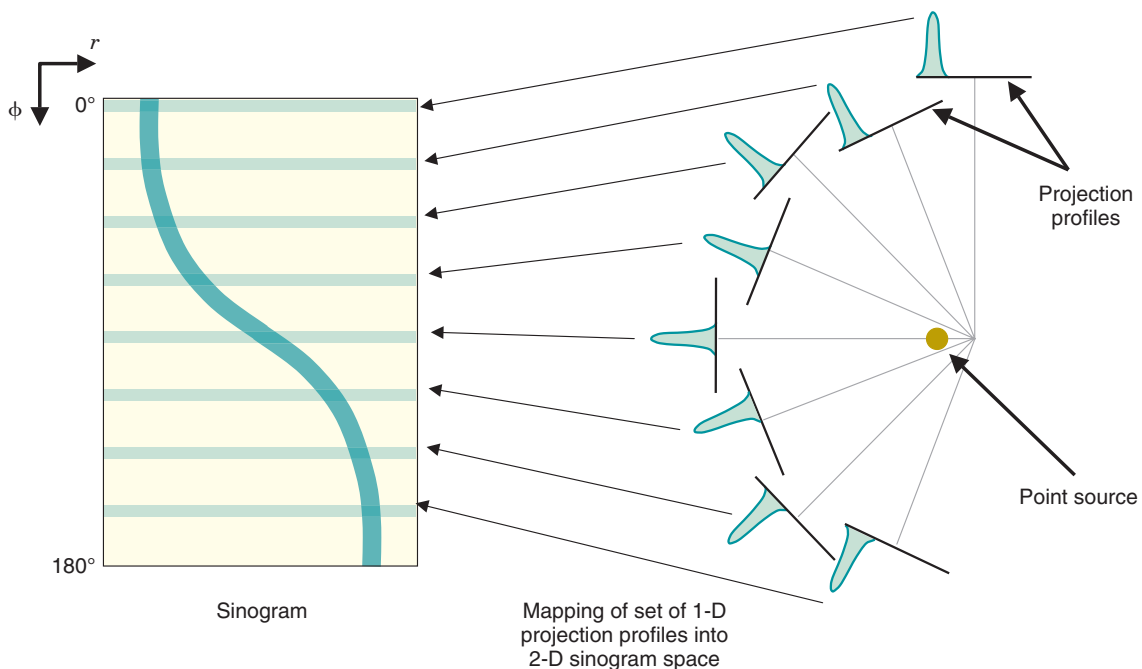


FIGURE 16-4 Two-dimensional (2-D) intensity display of a set of projection profiles, known as a *sinogram*. Each row in the display corresponds to an individual projection profile, sequentially displayed from top to bottom. A point source of radioactivity traces out a sinusoidal path in the sinogram.

pixels in the (x,y) coordinate system. For mathematical convenience, the image matrix size usually is a power of 2 (e.g., 64×64 or 128×128 pixels). Pixel dimensions Δx and Δy can be defined somewhat arbitrarily, but usually they are related to the number of profiles recorded and the width of the sampling interval along r .

The most basic approach for reconstructing an image from the profiles is by *simple backprojection*. The concepts will be illustrated for a point source object. Figure 16-5A shows projection profiles acquired from different angles around the source. An approximation for the source distribution within the plane is obtained by projecting (or distributing) the data from each element in a profile back across the entire image grid (Fig. 16-5B). The counts recorded in a particular projection profile element are divided uniformly amongst the pixels that fall within its projection path.* This operation is called *backprojection*. When the backprojections for all profiles are added

together, an approximation of the distribution of radioactivity within the scanned slice is obtained. Mathematically, the backprojection of N profiles is described by

$$f'(x,y) = \frac{1}{N} \sum_{i=1}^N p(x \cos \phi_i + y \sin \phi_i, \phi_i) \quad (16-3)$$

where ϕ_i denotes the i^{th} projection angle and $f'(x,y)$ denotes an approximation to the true radioactivity distribution, $f(x,y)$.

As illustrated in Figure 16-5B, the image built up by simple backprojection resembles the true source distribution. However, there is an obvious artifact in that counts inevitably are projected outside the true location of the object, resulting in a blurring of its image. The quality of the image can be improved by increasing the number of projection angles and the number of samples along the profile. This suppresses the “spokelike” appearance of the image but, even with an infinite number of views, the final image still is blurred. No matter how finely the data are sampled, simple backprojection always results in some apparent activity outside the true location for the point source. Figure 16-6 shows an image reconstructed by simple backprojection for a somewhat more complex object and more clearly illustrates the blurring effect.

*In practice, counts are assigned to a pixel in proportion to the fraction of the pixel area contained within the line of response for the projection element. However, owing to the complexity of the notation, this part of the algorithm is not included here.

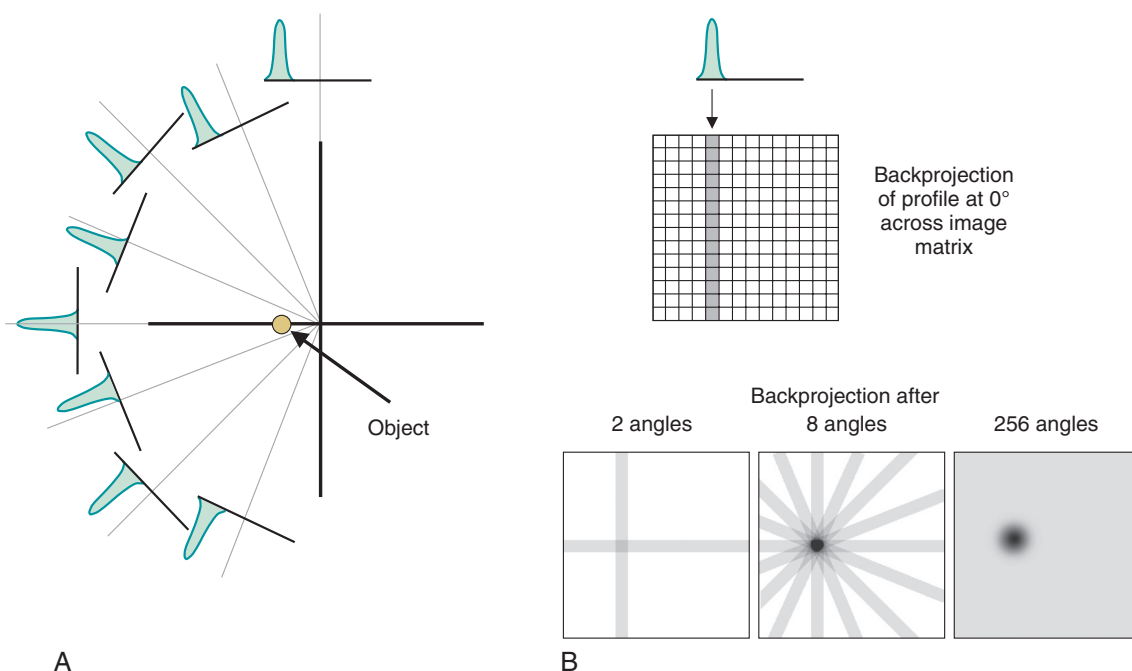


FIGURE 16-5 Illustration of the steps in simple backprojection. A, Projection profiles for a point source of radioactivity for different projection angles. B, Backprojection of one intensity profile across the image at the angle corresponding to the profile. This is repeated for all projection profiles to build up the backprojected image.

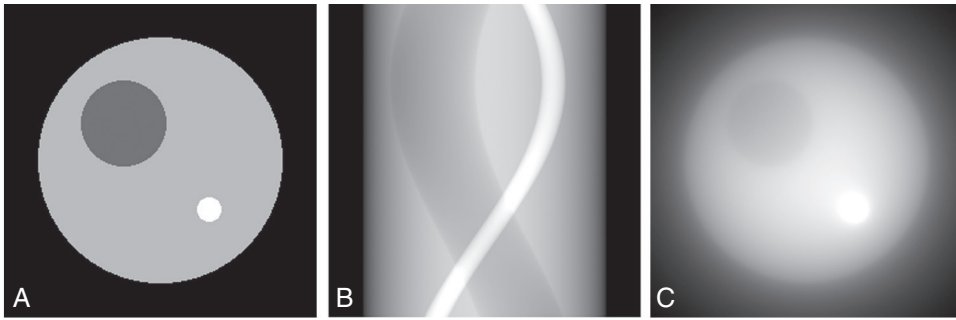


FIGURE 16-6 A, Computer-simulation phantom used for testing reconstruction algorithms. B, Sinogram of simulated data for a scan of the phantom. C, Image of simulation phantom for simple backprojection of data from 256 projection angles. $1/r$ blurring is apparent in the object, and edge details are lost. (Computer simulations performed by Dr. Andrew Goertzen, University of Manitoba, Canada)

Mathematically, the relationship between the true image and the image reconstructed by simple backprojection is described by

$$f'(x, y) = f(x, y) * (1/r) \quad (16-4)$$

where the symbol $*$ represents the process of convolution described in Appendix G. A profile taken through the reconstructed image for a point source that is reconstructed from finely sampled data decreases in proportion to $(1/r)$, in which r is the distance from the center of the point-source location. Because of this behavior, the effect is known as $1/r$ blurring. Simple backprojection is potentially useful only for very simple situations involving isolated objects of very high contrast relative to surrounding tissues, such as a tumor with avid uptake of a radiopharmaceutical that in turn has very low uptake in normal tissues. For more complicated objects, more sophisticated reconstruction techniques are required.

2. Direct Fourier Transform Reconstruction

One approach that avoids $1/r$ blurring is Fourier transform (FT) reconstruction, sometimes called *direct Fourier transform reconstruction* or *direct FT*. Although direct FT is not really a backprojection technique, it is presented here as background for introducing the filtered backprojection (FBP) technique in the next section.

Basic concepts of FTs are discussed in Appendix F. Briefly, in the context of nuclear medicine imaging, the FT is an alternative method for representing spatially varying data. For example, instead of representing a

1-D image profile as a spatially varying function, $f(x)$, the profile is represented as a summation of sine and cosine functions of different spatial frequencies, k . The amplitudes for different spatial frequencies are represented in the FT of $f(x)$, which is denoted by $F(k)$. The operation of computing the FT is symbolized by

$$F(k) = \mathcal{F}[f(x)] \quad (16-5)$$

The function $f(x)$ is a representation of the image profile in *image space* (or “object space”), whereas $F(k)$ represents the profile in “spatial frequency space,” also called *k-space*. FTs can be extended to 2-D functions, $f(x, y)$, such as a 2-D image. In this case, the FT also is 2-D and represents spatial frequencies along the x - and y -axes, $F(k_x, k_y)$, in which k_x and k_y represent orthogonal axes in 2-D k -space. Symbolically, the 2-D FT is represented as

$$F(k_x, k_y) = \mathcal{F}[f(x, y)] \quad (16-6)$$

Mathematically, a function and its FT are equivalent in the sense that either one can be derived from the other. The operation of converting the FT of a function back into the original function is called an *inverse FT* and is denoted by

$$\mathcal{F}^{-1}[F(k_x, k_y)] = f(x, y) \quad (16-7)$$

FTs can be calculated quickly and conveniently on personal computers, and many image and signal-processing software packages contain FT routines. The reader is referred to Appendix F for additional information about FTs.

The concept of k-space will be familiar to readers who have studied magnetic resonance imaging (MRI), because this is the coordinate system in which MRI data are acquired. To reconstruct an image from its 2-D FT, the full 2-D set of k-space data must be available (Equation 16-7). In MRI, data are acquired point-by-point for different (k_x, k_y) locations in a process known as “scanning in k-space.” There is no immediately obvious way to directly acquire k-space data in nuclear medicine imaging. Instead, nuclear medicine CT relies on the *projection slice theorem*, or *Fourier slice theorem*. In words, this theorem says that the FT of the projection of a 2-D object along a projection angle ϕ [in other words, the FT of a profile, $p(r, \phi)$], is equal to the value of the FT of the object measured through the origin and along the same angle, ϕ , in k-space (note, the *value* of the FT, not the *projection* of the FT). Figure 16-7 illustrates this concept. Mathematically, the general expression for the projection slice theorem is

$$\mathcal{F}[p(r, \phi)] = F(k_r, \phi) \quad (16-8)$$

where $F(k_r, \phi)$ denotes the value of the FT measured at a radial distance k_r along a line at angle ϕ in k-space.

The projection slice theorem provides a means for obtaining 2-D k-space data for an object from a series of 1-D measurements in object space. Figure 16-7 and Equation 16-8 provide the basis for reconstructing an object from its projection profiles as follows:

1. Acquire projection profiles in object space at N projection angles, ϕ_i , $i = 1, 2, \dots, N$ as previously described.
2. Compute the 1-D FT of each profile.
3. Insert the values of these FTs at the appropriate coordinate locations in k-space. Note that values are inserted in polar coordinates along radial lines through the origin in k-space. For a specific value of k_r in the FT of the projection acquired at a rotational angle ϕ , the data are inserted at rectangular coordinates given by

$$\begin{aligned} k'_x &= k_r \cos \phi \\ k'_y &= k_r \sin \phi \end{aligned} \quad (16-9)$$

where primed notation is used to indicate that the coordinate locations do not correspond exactly to points on a rectangular grid. The inserted values are closely spaced near the origin and more widely spaced farther away from the origin. This “over-representation” of data near the origin in k-space is one explanation for the $1/r$

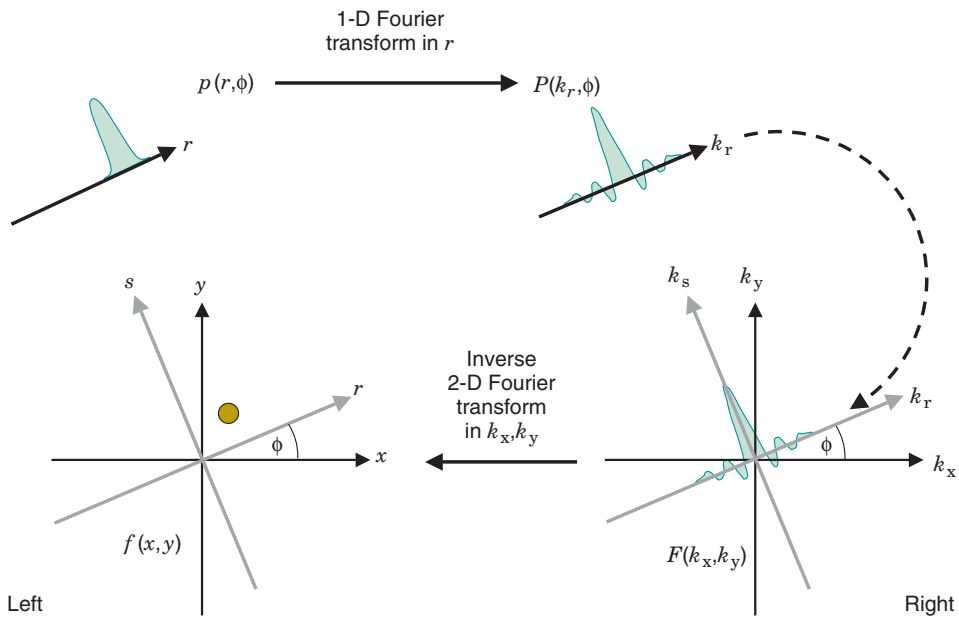


FIGURE 16-7 Concepts of the projection slice theorem. (Left) $p(r, \phi)$ is a one-dimensional (1-D) profile of the 2-D object $f(x, y)$ at projection angle ϕ . The theorem states that the 1-D Fourier transform of this projection profile (right) is equal to the values of the 2-D Fourier transform of the object, $F(k_x, k_y)$, along a line through the origin of k-space at the same angle ϕ .

blurring that occurs in simple backprojection, as was discussed in Section B.1.

4. Using the values inserted in polar coordinates, interpolate values for k_x and k_y on a rectangular grid in k-space.
5. Use the interpolated values in k-space and a standard 2-D (inverse) FT (Equation 16-7) to compute the image of the object.

With noise-free data, perfect projection profiles (i.e., line integrals that represent precisely the sum of activity along a line measured through the object) and perfect interpolation, the direct FT reconstruction technique is capable of producing an exact representation of the object. Additional criteria regarding the required numbers of projection profiles and sampled points across each profile are discussed in Section C.

A major drawback of direct Fourier reconstruction is that the interpolation from polar to rectangular coordinates in k-space is computationally intensive. As well, it can lead to artifacts in the image, if not done carefully. A more elegant (and practical) approach, called *filtered backprojection (FBP)*, is described in the next section.

3. Filtered Backprojection

Like the direct FT algorithm, FBP employs the projection slice theorem but uses the theorem in combination with backprojection in a manner that eliminates $1/r$ blurring. The steps are as follows:

1. Acquire projection profiles at N projection angles (same as direct FT).
2. Compute the 1-D FT of each profile (same as direct FT). In accordance with the projection slice theorem (see Fig. 16-7 and Equation 16-8), this provides values of the FT for a line across k-space.
3. Apply a “ramp filter” to each k-space profile. Mathematically, this involves

multiplying each projection FT by $|k_r|$, the absolute value of the radial k-space coordinate at each point in the FT. Thus the value of the FT is increased (amplified) linearly in proportion to its distance from the origin of k-space. Figure 16-8 illustrates the profile of a ramp filter, with filter amplitude denoted by $H(k_r)$. Applying the ramp filter produces a modified FT for each projection, given by

$$P'(k_r, \phi) = |k_r| P(k_r, \phi) \quad (16-10)$$

where $P(k_r, \phi)$ is the unfiltered FT.

4. Compute the inverse FT of each filtered FT profile to obtain a modified (filtered) projection profile. This is given by

$$\begin{aligned} p'(r, \phi) &= \mathcal{F}^{-1}[P'(k_r, \phi)] \\ &= \mathcal{F}^{-1}[|k_r| P(k_r, \phi)] \end{aligned} \quad (16-11)$$

5. Perform conventional backprojection using the filtered profiles. Mathematically, the result is

$$f(x, y) = \frac{1}{N} \sum_{i=1}^N p'(x \cos \phi_i + y \sin \phi_i, \phi_i) \quad (16-12)$$

Step 5 is essentially the same as simple backprojection, but with filtered profiles. However, unlike Equation 16-3, in which $f'(x, y)$ is only an approximation of the true distribution, FBP, when applied with perfectly measured noise-free data, yields the exact value of the true distribution, $f(x, y)$. Figure 16-9 schematically illustrates the process of FBP for a pointlike object.

The only difference between simple and filtered backprojection is that in the latter method, the profiles are modified by a *reconstruction filter* applied in k-space before they are backprojected across the image. The effect

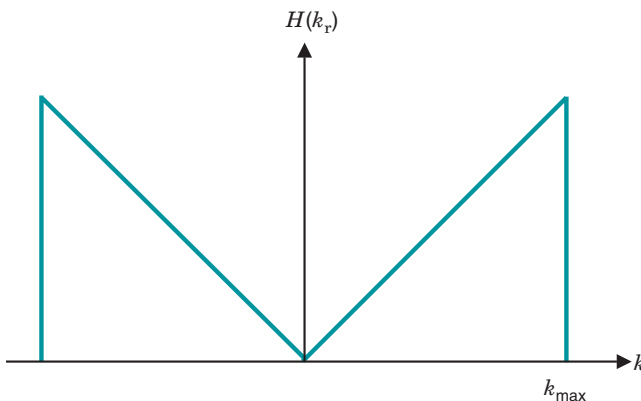


FIGURE 16-8 Ramp filter in the spatial-frequency (k-space) domain. The filter selectively amplifies high-frequency components relative to low-frequency components. The filter removes the $1/r$ blurring present in simple backprojection and sharpens image detail, but it also amplifies high-frequency noise components in the image.

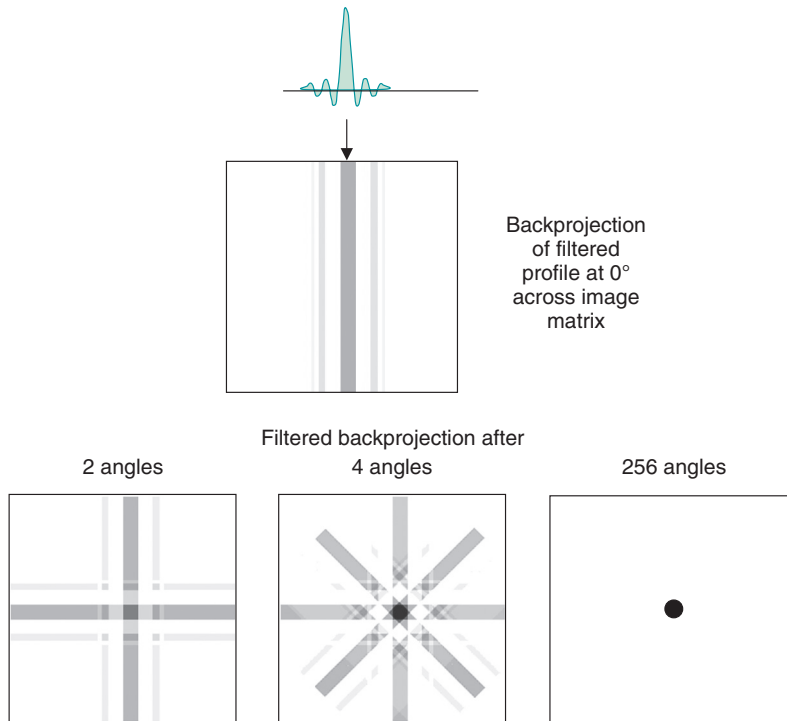


FIGURE 16-9 Illustration of the steps in filtered backprojection. The one-dimensional Fourier transforms of projection profiles recorded at different projection angles are multiplied by the ramp filter. After taking the inverse Fourier transform of the filtered transforms, the filtered profiles are backprojected across the image, as in simple backprojection.

of the ramp filter is to enhance high spatial frequencies (large k_r) and to suppress low spatial frequencies (small k_r). The result of the filtering is to eliminate $1/r$ blurring.* One way to visualize the effect is to note that, unlike unfiltered profiles (see Fig. 16-5), the filtered profiles have both positive and negative values (see Fig. 16-9). The negative portions of the filtered profiles near the central peak “subtract out” some of the projected intensity next to the peak that otherwise would create $1/r$ blurring.

Amplification of high spatial frequencies in FBP also leads to amplification of high-frequency noise. Because there usually is little signal in the very highest frequencies of a nuclear medicine image, whereas statistical noise is “white noise” with no preferred

frequency, this also leads to degradation of signal-to-noise ratio (SNR). For this reason, images reconstructed by FBP appear noisier than images reconstructed by simple backprojection. (This is a general result of any image filtering process that enhances high frequencies to “sharpen” images.) In addition, filters that enhance high frequencies sometimes have edge-sharpening effects that lead to “ringing” at sharp edges. This is an unwanted byproduct of the positive-negative oscillations introduced by the filter, illustrated in the filtered profile at the top of Figure 16-9.

To minimize these effects on SNR and artifacts at sharp edges, the ramp filter usually is modified so as to have a rounded shape to somewhat suppress the enhancement of high spatial frequencies. Figure 16-10 illustrates a ramp filter and two other commonly used reconstruction filters. Also shown are the equations describing these filters. A variety of reconstruction filters have been developed, each with its own theoretical rationale. Filters also play a role in the suppression of artifacts caused by aliasing in FT-based reconstruction techniques, as discussed in Section C. Additional discussions of these filters can be found in reference 4.

*More precisely, $1/r$ blurring is the convolution of the true image with a blurring function, $b(r) = 1/r$ (Equation 16-4). As discussed in Appendix G, convolution in image space is equivalent to multiplying by the FT of the blurring function in k-space, which for $b(r)$ is $B(k_r) = 1/|k_r|$. Thus multiplying by $|k_r|$ in k-space is equivalent to *deconvolving* the blurring function in image space, thereby eliminating the blurring effect.

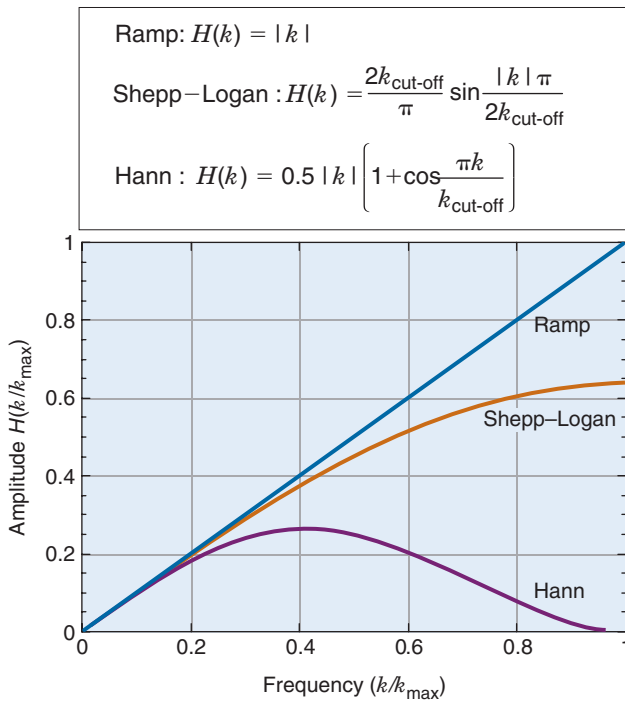


FIGURE 16-10 Ramp filter and two other reconstruction filters that are designed to prevent artifacts and noise amplification caused by the sharp cut-off of the ramp filter at the maximum frequency k_{max} . Note that all of the filters shown have the same response at lower frequencies and that cut-off frequencies are set so that $k_{\text{cut-off}} = k_{\text{max}}$. However, the Hann and Shepp-Logan filters roll off gradually at higher frequencies, thereby minimizing artifacts and noise amplification.

Because of its speed and relative ease of implementation, FBP became a widely used reconstruction method in nuclear medicine. A single 2-D image slice can be reconstructed in a fraction of a second on a standard computer. Under idealized conditions (noise-free data acquisition, completely sampled data, and so forth), FBP produces an accurate representation of the distribution of radioactivity within the slice. However, FBP is not without its limitations. First, it is susceptible to major artifacts if data from the object are measured incompletely (possibly because of collimator defects, portions of the object outside the FOV of the camera for some projections, etc.). Second, in datasets that have poor counting statistics or random “noise” spikes (perhaps caused by instrument malfunction), FBP produces annoying “streak artifacts.” These artifacts can be suppressed by employing a k-space filter with a strong roll-off at high spatial frequencies, but this results in loss of image resolution as well.

Finally, the FBP algorithm cannot readily be modified to take into account various physical aspects of the imaging system and data acquisition, such as limited spatial resolution of the detector, scattered radiation, and the fact that the sensitive volume of the detector collimator holes actually is a cone rather than a cylinder, as assumed for the reconstruction process. These factors require

additional preprocessing or postprocessing data manipulations that work with varying degrees of success. They are discussed further in Chapter 17, Section B.

By contrast, another set of reconstruction methods, known as *iterative reconstruction techniques*, can build these steps directly into the reconstruction algorithm and are less prone to the artifacts described in the preceding paragraph. These techniques are described in Section D.

4. Multislice Imaging

The analysis presented earlier for backprojection and Fourier-based reconstruction techniques applies to single-slice images. In practice, as described in Chapters 17 and 18, both SPECT and PET imaging are performed with detectors that acquire data simultaneously for multiple sections through the body. Projection data originating from each individual section through the object can be reconstructed as described. Individual image slices then are “stacked” to form a 3-D dataset, which in turn can be “resliced” using computer techniques to obtain images of planes other than those that are directly imaged. Thus 2-D multislice imaging of contiguous slices can be used to generate 3-D volumetric images. In many SPECT systems, the distance between image slices and slice thickness can be adjusted to achieve different axial

resolutions, much the same as the sampling interval, Δr , across the image profile can be adjusted to vary the in-plane resolution. In PET systems, the distance between image slices and the slice thickness often is fixed by the axial dimensions of the segmented scintillator crystals typically used in the detectors (See Chapter 18, Sections B and C).

C. IMAGE QUALITY IN FOURIER TRANSFORM AND FILTERED BACKPROJECTION TECHNIQUES

In this section, we discuss some general issues involving image quality in reconstruction tomography based on the direct FT and FBP techniques. These issues affect all reconstruction tomography based on these techniques, including both x-ray CT and ECT. Additional aspects that are specifically relevant to SPECT and PET image quality are discussed in Chapters 17 and 18, respectively. The issues discussed here do not pertain directly to iterative reconstruction techniques, which are discussed separately in Section D.

1. Effects of Sampling on Image Quality

Projection data are not continuous functions but discrete point-by-point samples of projection profiles. The distance between the sample points is the *linear sampling distance*. In addition, projection profiles are acquired only at a finite number of *angular sampling intervals* around the object. The choice of linear and angular sampling intervals and the cut-off frequency of the reconstruction filter (see Fig. 16-10), in conjunction with the spatial resolution of the detector system, determine the spatial resolution of the reconstructed image. The effects of the imaging system depend on the type of detector, collimator, and so forth and are discussed in Chapters 17 and 18. Here we discuss briefly those aspects that are related to the reconstruction process, which are applicable to all types of imaging devices.

The sampling theorem⁵ states that to recover spatial frequencies in a signal up to a maximum frequency k_{\max} requires a linear sampling distance given by

$$\Delta r \leq 1/(2k_{\max}) \quad (16-13)$$

This means that the highest spatial frequency to be recovered from the data must be sampled

at least twice per cycle. Coarser sampling does not allow higher spatial frequencies to be recovered and leads to image artifacts known as *aliasing*. Mathematical aspects of aliasing are discussed in detail in Appendix F, Section C.

Thus the linear sampling distance sets a limit on spatial resolution for the imaging system. This limit (k_{\max} in Equation 16-13) also is known as the *Nyquist frequency*, k_{Nyquist} (see also Equation F-9). The highest spatial frequency that is present in an image profile depends on the spatial resolution of the collimator-detector system. Higher resolution implies higher frequency content. As a rule of thumb, the sampling requirement for an imaging detector is

$$\Delta r \leq \text{FWHM}/3 \quad (16-14)$$

where FWHM is the full width at half maximum of its point-spread function (see Chapter 15, Section B.2).

Figure 16-11 shows images of a computer-simulation phantom that were reconstructed with progressively coarser sampling of the image profiles. Undersampling not only results in image blurring but also creates image artifacts resulting from the effects of aliasing.

The Nyquist frequency is the highest spatial frequency represented in k-space and thus defines an upper frequency limit for the reconstruction filter. However, a lower-frequency filter cut-off, $k_{\text{cut-off}} < k_{\text{Nyquist}}$ can be used in the reconstruction. This improves SNR by suppressing the high-frequency end of the spatial frequency spectrum, where a large fraction of the signal is statistical noise. Lowering the cut-off frequency also degrades spatial resolution, because the higher frequencies also contain the fine details of the image. Thus the choice of the reconstruction filter and its cut-off frequency involve a tradeoff between spatial resolution and SNR in the image. This is illustrated in Figure 16-12, which shows images of a computer-simulation phantom reconstructed with a Shepp-Logan filter with different cut-off frequencies.

The angular sampling interval (angle between projections) should provide sampling around the periphery at approximately the same intervals as the linear sampling distance. Thus if projections are acquired around a FOV of diameter D , the minimum number of angular views, N_{views} , should be approximately the length of the 180-degree arc over

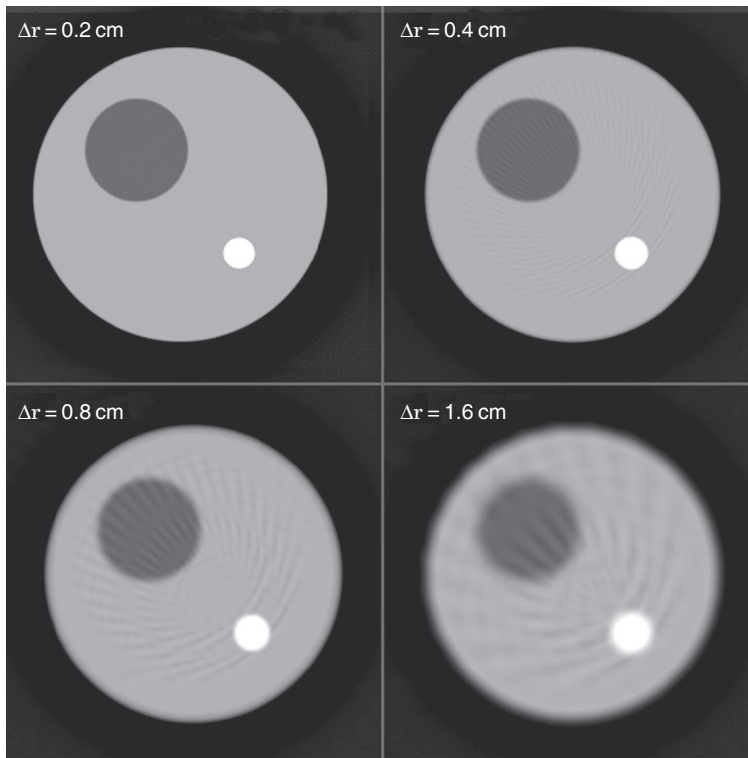


FIGURE 16-11 Images of a computer-simulation phantom reconstructed with progressively coarser sampling of the image profiles. Linear under-sampling results both in loss of resolution and image artifacts. (Computer simulations performed by Dr. Andrew Goertzen, University of Manitoba, Canada.)

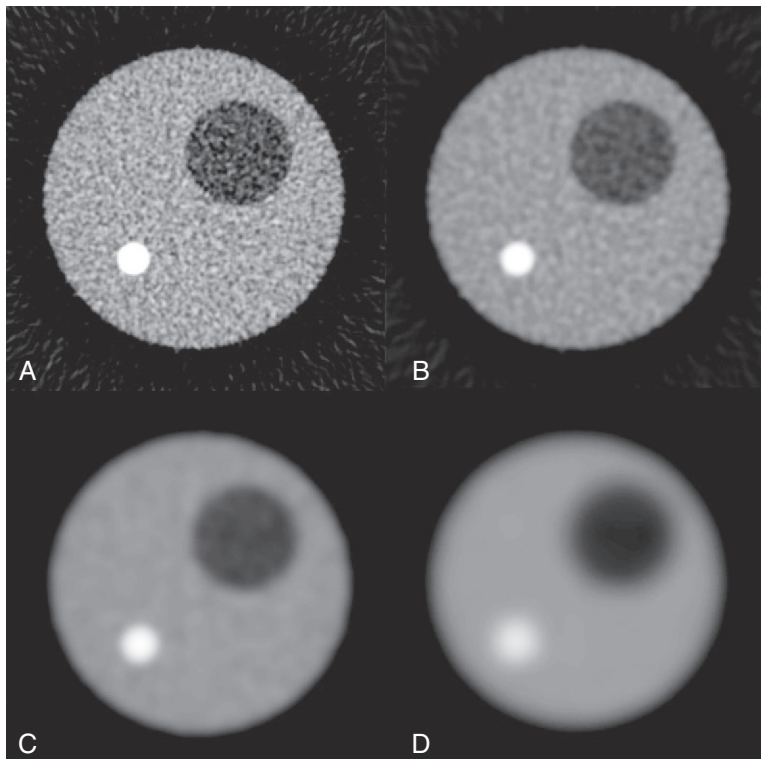


FIGURE 16-12 Filtered backprojection reconstructions of the computer-simulation phantom shown in Figure 16-6, using a Shepp-Logan filter with different cut-off frequencies. A, $k_{\text{cut-off}} = k_{\text{max}}$; B, $k_{\text{cut-off}} = 0.8 k_{\text{max}}$; C, $k_{\text{cut-off}} = 0.6 k_{\text{max}}$; and D, $k_{\text{cut-off}} = 0.2 k_{\text{max}}$. Note the tradeoff between image detail and signal-to-noise ratio. (Computer simulations performed by Dr. Andrew Goertzen, University of Manitoba, Canada.)

which projections are acquired ($\pi D/2$) divided by the linear sampling distance, Δr :

$$N_{\text{views}} \geq \pi D/2\Delta r \quad (16-15)$$

Figure 16-13 illustrates the effect of angular sampling interval on images of a computer-simulation phantom. Spokelike artifacts are evident around high-intensity objects when the number of angular samples is inadequate.

EXAMPLE 16-1

Suppose you are working with an ECT system that has spatial resolution $\text{FWHM} \approx 1$ cm and $\text{FOV} = 30$ cm. Estimate the sampling interval, Δr , and the number of angular views, N_{views} , that would support the available spatial resolution of the system.

Answer

From Equation 16-14, the sampling interval should be

$$\Delta r \leq 1 \text{ cm} / 3 \approx 0.33 \text{ cm}$$

For $\text{FOV} = 30$ cm, this amounts to

$$N_{\text{samp}} \geq 30/(1/3) \approx 90 \text{ samples per profile}$$

According to Equation 16-15, the number of views should be such that

$$N_{\text{views}} \geq (\pi \times 30)/[2 \times (1/3)] \approx 140 \text{ views}$$

Thus 140 views over a 180-degree degree arc, with linear sampling at approximately 0.33-cm intervals, would fully support the available system resolution.

The closest power-of-two image reconstruction and display matrix that would meet the sampling requirements in Example 16-1 is 128×128 . One possibility would be to interpolate the sampled profiles from 90 samples to 128 samples. A more practical option, however, is to acquire 128 samples over 30 cm, which would somewhat exceed the linear sampling requirement. If this were done, Equation 16-15 would suggest that additional views would be needed to support the smaller value of Δr ;

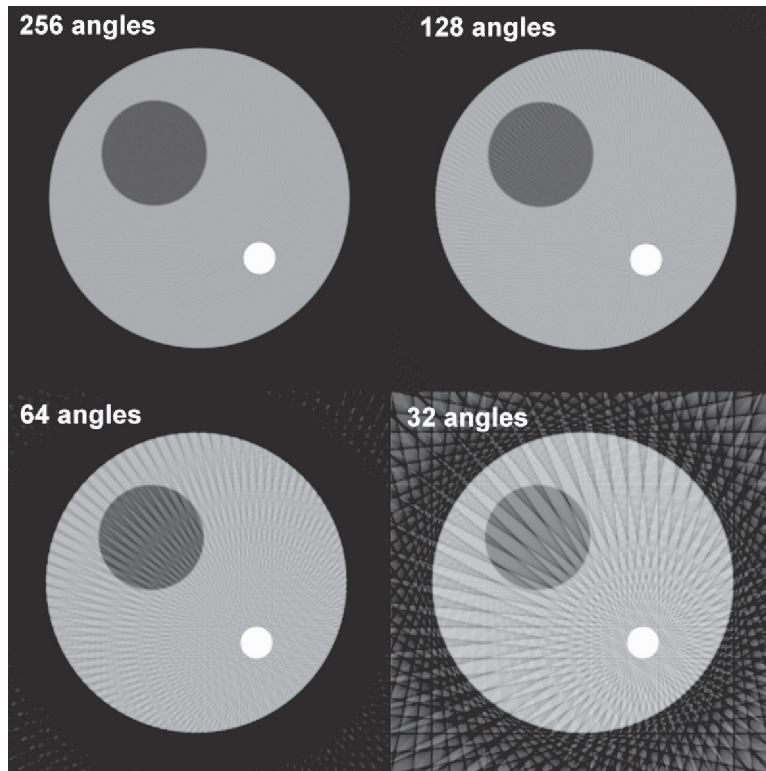


FIGURE 16-13 Effect of the number of angular samples recorded on the reconstructed image of a computer-simulation phantom. Spokelike streak artifacts are evident when an inadequate number of projections are used. (Computer simulations performed by Dr. Andrew Goertzen, University of Manitoba, Canada.)

however, 140 views still would provide the number of angular views needed to support the *system resolution* and this number would not have to be increased. On the other hand, going in the opposite direction, that is, acquiring only 64 samples and fewer angular samples for reconstruction on a 64×64 matrix would lead to a loss of image detail and introduce the possibility of image artifacts, as illustrated in Figures 16-11 and 16-13.

2. Sampling Coverage and Consistency Requirements

In addition to meeting the requirements described in the preceding section regarding linear and angular sampling intervals, the data acquired must provide full coverage of the object. Thus it is necessary that data be acquired over a full 180-degree arc. If an arc less than 180 degrees is used, geometric distortions are produced. Figure 16-14 demonstrates that an inadequate angular-sampling range causes data to flare out past the true objects and produces geometric distortions perpendicular to the direction of the absent projections. This is a problem for a number of systems developed in nuclear medicine that are classified as “limited-angle tomography” (e.g., rotating slant-hole tomography

and some positron emission mammography systems).

A second requirement for coverage is that the entire object (or at least the parts containing radioactivity) must be included in all projections. If some parts of the object are not included in all projections, the data will be *inconsistent* between different projections. There are a number of ways in which this can happen. For example, the FOV of the detector may be insufficient to provide full coverage from all directions. Figure 16-15 illustrates the effect of incomplete coverage of the object during some parts of the scan.

Two other possible sources of inconsistency between projections are patient movement and missing or distorted values in individual profiles caused by instrumentation failures, such as an unstable element in a detector array. Figure 16-16 illustrates some effects of these types of inconsistencies.

3. Noise Propagation, Signal-to-Noise Ratio, and Contrast-to-Noise Ratio

Noise propagation, SNR, and CNR differ in ECT from their behavior in conventional planar imaging. In conventional planar imaging, the SNR for an individual pixel is essentially equal to $\sqrt{N_{\text{pixel}}}$, in which N_{pixel} is

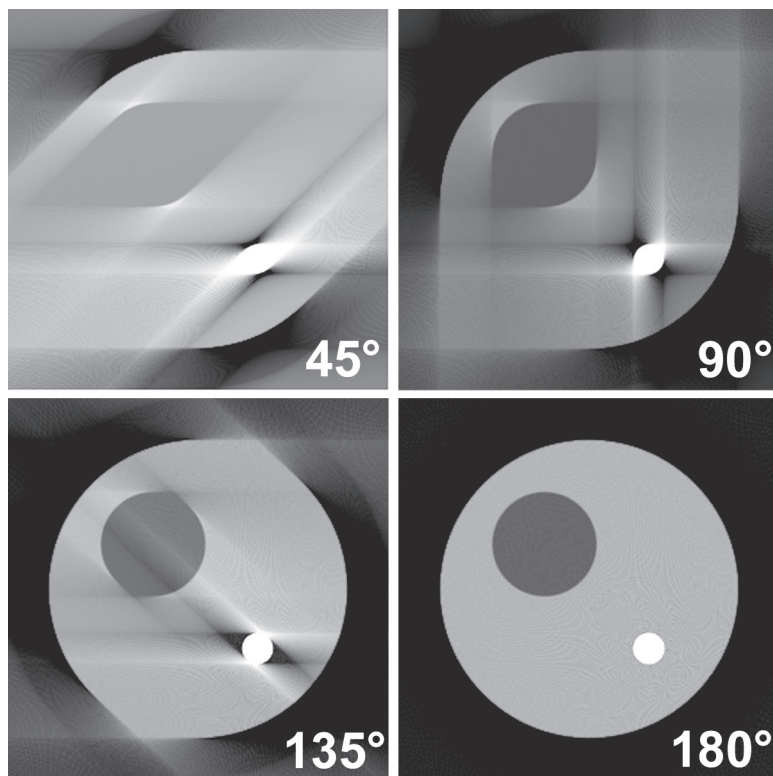


FIGURE 16-14 Effects of angular sampling range on images of a computer-simulation phantom. Images obtained by sampling over 45 degrees, 90 degrees, 135 degrees, and 180 degrees. Sampling over an interval of less than 180 degrees distorts the shape of the objects and creates artifacts. (Computer simulations performed by Dr. Andrew Goertzen, University of Manitoba, Canada.)

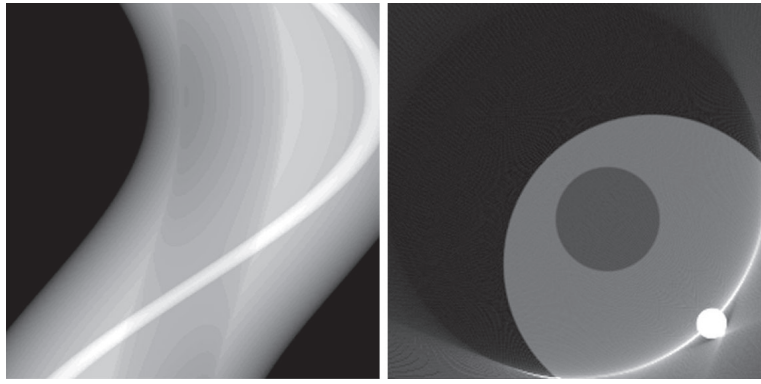


FIGURE 16-15 Effects of having some profiles that do not cover the entire object. *Left*, Sinogram of computer-simulation phantom. *Right*, Reconstructed image. (Computer simulations performed by Dr. Andrew Goertzen, University of Manitoba, Canada.)

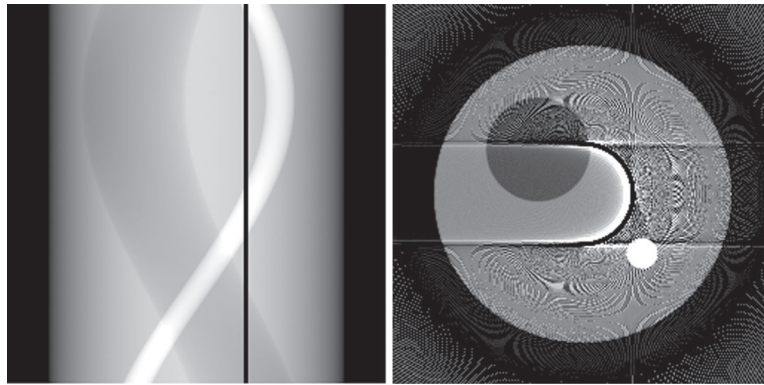


FIGURE 16-16 Effects of missing projection elements on reconstructed image. *Left*, Sinogram of computer-simulation phantom. *Right*, Reconstructed image. This simulation would apply to a SPECT image reconstructed from profiles acquired over a 180-degree sampling range with a single-headed camera, with one region of the detector “dead.” (Computer simulations performed by Dr. Andrew Goertzen, University of Manitoba, Canada.)

the number of counts recorded for that pixel. In ECT, the computation of noise and SNR is much more complicated because the intensity level for each pixel is derived by computations involving different views and many other pixels in the image. In addition, a variety of mathematical manipulations, such as filtering operations, are performed along the way. As a result, although SNR still depends on the square root of the total number of counts recorded during the imaging procedure, the relationship between those counts and the SNR of individual pixels is more complicated.

Suppose that an ECT image is acquired of a cylindrical object of diameter D containing a uniform concentration of radioactivity. Suppose further that projection data are acquired with a linear sampling interval Δr across all projection profiles, that a total of N_{image} counts are recorded during the imaging procedure, and that the image is reconstructed by FBP with a ramp filter on a

square matrix of size $D \times D = D^2$ with pixel size $\Delta r \times \Delta r = \Delta r^2$. It can be shown that the SNR for an individual pixel in the resulting image of the object is given by⁶

$$\text{SNR}_{\text{pixel}} \approx \sqrt{\frac{12N_{\text{image}}}{\pi^2(D/\Delta r)^3}} \quad (16-16)$$

Equation 16-16 indicates that SNR decreases when pixel size, Δr , is made smaller, that is, as spatial resolution is improved.* The

*Note that Equation 16-16 specifically assumes that pixel width is the same as the sampling interval, Δr . Often in nuclear medicine, interpolation techniques are used to generate images with pixels that are smaller than the sampling interval. Equation 16-16 is valid in these situations provided that the sampling interval rather than pixel size is used in the equation. Some texts describe Δr as the “resolution element” to avoid confusing it with pixel size.

dependence is relatively strong, as illustrated by the following example.

EXAMPLE 16-2

Suppose that an image of a 20-cm diameter cylinder containing a uniform distribution of activity is generated by FBP reconstruction. The image is reconstructed on a square matrix, 20×20 cm in size with 1×1 cm pixel size. A total of 1 million counts are acquired for the image. Calculate the number of counts required to generate an image with the same SNR per pixel if both the sampling interval and pixel size are reduced to 0.5 cm.

Answer

For the image with 1 cm resolution, from Equation 16-16

$$\text{SNR}_{\text{pixel}} \approx \sqrt{\frac{12 \times 10^6}{\pi^2 \times (20/1)^3}} \\ \approx 12.33$$

To maintain the same value of $\text{SNR}_{\text{pixel}}$ with 0.5-cm pixels, the required number of counts in the image, $N_{0.5 \text{ cm}}$, must be such that

$$\sqrt{\frac{12 \times N_{0.5 \text{ cm}}}{\pi^2 \times (20/0.5)^3}} \approx 12.33$$

Solving this equation yields the requirement of $N_{0.5 \text{ cm}} = 8$ million counts.

Example 16-2 indicates that if Δr is decreased by a factor of 2 (to $\Delta r/2$), the total number of counts required to keep the SNR per pixel constant increases by a factor of 8, that is, as the inverse cube of the size of the pixel.

The total number of pixels in the reconstructed image is $n_{\text{pixels}} = (D/\Delta r)^2$. Equation 16-16 can be rewritten as

$$\text{SNR}_{\text{pixel}} \approx \sqrt{12/\pi^2} \times \frac{\sqrt{\langle N_{\text{pixel}} \rangle}}{\sqrt[4]{n_{\text{pixels}}}} \quad (16-17)$$

where $\langle N_{\text{pixel}} \rangle$ is the average number of counts recorded per reconstructed pixel in the object. This can be simplified even further by noting that $\sqrt{12/\pi^2} = 1.103 \approx 1$, for more approximate work.

Equation 16-17 indicates that SNR per pixel improves in proportion to the square root of the average number of counts recorded per pixel. This part of the equation is

consistent with conventional counting statistics. However, as compared with conventional planar imaging (or photon counting), there is an additional factor equal to the fourth root of the total number of pixels (or resolution elements) in the denominator of the equation. This places stronger requirements on counting statistics for reconstruction tomography as compared with planar imaging, to achieve a specified level of $\text{SNR}_{\text{pixel}}$, as illustrated in the following example.

EXAMPLE 16-3

Consider two images of a cylindrical cross-section 20 cm in diameter with a uniform concentration of radioactivity, one a planar image and the other generated by FBP reconstruction with a ramp filter. Each image is 20×20 cm in size with 1×1 cm pixel size, and a total of 1 million counts are acquired for each image. What is the percent noise level, relative to signal level, in each image?

Answer

For both images,

$$\begin{aligned} \text{percent noise level} \\ &= (\text{noise/signal}) \times 100\% \\ &= 100\% / \text{SNR}_{\text{pixel}} \end{aligned}$$

For the planar image, the average number of counts per pixel within the area occupied by the 20-cm diameter object is

$$\begin{aligned} n_{\text{pixel}} &= [10^6 / (\pi \times 10^2)] \text{ counts/cm}^2 \times 1 \text{ cm}^2/\text{pixel} \\ &\approx 3180 \text{ counts/pixel} \end{aligned}$$

This yields an SNR given by

$$\begin{aligned} \text{SNR}_{\text{pixel}} &= \frac{n_{\text{pixel}}}{\sqrt{n_{\text{pixel}}}} = \sqrt{n_{\text{pixel}}} \\ &= \sqrt{3180} = 56 \end{aligned}$$

from which the percent noise level is $100\%/56 \approx 1.8\%$.

For the image reconstructed by FBP, the result is as given in Example 16-2:

$$\text{SNR}_{\text{pixel}} \approx 12.33$$

from which the percent noise level is $100\%/12.33 \approx 8.1\%$.

The “noise enhancement” factor for reconstruction tomography illustrated in Example 16-3 is the result of noise propagation from

pixels at many locations in the imaged object into the pixel of interest in the backprojection process, as well as the ramp filtering operation.

Example 16-3 applies to the SNR of a single pixel in images of a uniform object. The result would seem to imply a statistical disadvantage for the detection of low-contrast objects by ECT. However, for purposes of applying the Rose criterion for detectability of lesions and other objects (see Chapter 15, Section D.2) this must be converted to CNR for the object of interest. Using the definitions given in Chapter 15, it can be shown that the CNR for a lesion that occupies n_ℓ pixels in an ECT image is

$$\text{CNR}_\ell \approx |C_\ell| \times \sqrt{n_\ell} \times \text{SNR}_{\text{pixel}} \quad (16-18)$$

where the absolute value indicates that CNR always is a positive quantity. Although the noise characteristics in ECT differ somewhat from those of planar imaging (particularly regarding possible artifacts), the same general rules for detectability apply for ECT and planar images, that is, $\text{CNR}_\ell \geq 4$.

EXAMPLE 16-4

Consider the situation described for radiopharmaceutical B in Example 15-6. In that example, the radiopharmaceutical produced “cold” lesions with uptake that was 10% of the surrounding normal tissue and a CNR of only 0.9 for a 1-cm diameter lesion. Using the same parameters, estimate the CNR that would be achieved using the same radiopharmaceutical, spatial resolution, and total imaging time with ECT. Assume that the normal tissue fills a volume of $(10 \times 10 \times 10 \text{ cm})$ and, as in Example 15-6, ignore the effects of attenuation and source-to-detector distance.

Answer

The planar image described for Example 15-6 could be obtained by facing the detector toward any face of the cubic volume of tissue and acquiring counts for a 1-min imaging time. For purposes of computing the SNR of an ECT image, many projection views would be required (e.g., 60 1-sec views), but the total number of counts recorded in 1 min of imaging time, in the absence of attenuation and distance effects, would be the same as for planar imaging. The total number of pixels in the ECT image is $n_{\text{pixels}} = 10 \times 10 = 100$, and the average number of counts per pixel in the

image is $\langle N_{\text{pixel}} \rangle \approx 10$ counts per minute (cpm) $\times 1 \text{ min} = 10$. Thus from Equation 16-17, the SNR per pixel for the ECT image would be

$$\text{SNR}_{\text{pixel}} \approx \frac{\sqrt{10}}{\sqrt[4]{100}} \approx \frac{\sqrt{10}}{\sqrt{10}} \approx 1$$

Using the definition given in Equation 15-6, the contrast of the lesion in the ECT image is

$$C_\ell = (1 - 10)/10 = -0.9$$

Substituting these values into Equation 16-18, one obtains

$$\text{CNR}_\ell \approx |-0.9| \times \sqrt{1} \times 1 \approx 0.9$$

which is the same result as obtained for planar imaging. In neither case would the lesion be detectable using the Rose criterion.

Example 16-4 shows that, for the same level of object contrast and total number of counts in the image, and in the absence of attenuation and distance effects, *there is no intrinsic difference in CNR between ECT and planar imaging*. This result is obtained, in spite of the apparent statistical disadvantage of ECT illustrated in Example 16-3, because of the increased contrast of the low-contrast lesion in an ECT image as compared with a projection image. On the other hand, the sophisticated data manipulations of ECT do not improve the detectability of the lesion. This is not too surprising, because it should not be possible to improve CNR when noise is generated by counting statistics by applying mathematical manipulations (e.g., reconstruction tomography and contrast enhancement) of otherwise comparable data.

Thus it is inaccurate to conclude that ECT improves detectability of lesions or other objects by improving CNR. Rather, *the primary advantage of ECT for detecting low-contrast lesions derives from its ability to remove confusing overlying structures that may interfere with detectability of those lesions*, such as ribs overlying a lesion in the lungs. Not only does an object become more detectable when overlying clutter is removed by ECT, but its shape and borders become more clear.

An additional advantage of ECT is the ability to determine more accurately the concentration of radioactivity in a particular volume of tissue. For example, in Example 16-4, the same planar image would be obtained if the lesion were twice as thick

along the viewing direction, but with half the uptake suppression and thus twice the concentration as originally specified in Example 15-6. However, such a difference would be readily evident on the ECT image (assuming the CNR requirements for detectability were met). Some appreciation for all of these advantages can be gained by inspection of Figure 15-11.

D. ITERATIVE RECONSTRUCTION ALGORITHMS

A viable and increasingly used alternative to FBP is a class of methods known as *iterative reconstruction*. These methods are computationally more intensive than FBP and for this reason have been more slowly adopted in the clinical setting. However, as computer speeds continue to improve, and with a combination of computer acceleration techniques (e.g., parallel processors), and intelligent coding (e.g., exploiting symmetries

and precomputing factors), reconstruction times have become practical and iterative methods are finding their way into more general use.

1. General Concepts of Iterative Reconstruction

The general concepts of iterative reconstruction are outlined in Figure 16-17. In essence, the algorithm approaches the true image, $f(x,y)$, by means of successive approximations, or estimates, denoted by $f^*(x,y)$. Often the initial estimate is very simple, such as a blank or uniform image. The next step is to compute the projections that would have been measured for the estimated image, using a process called *forward projection*. This process is exactly the inverse of backprojection. It is performed by summing up the intensities along the potential ray paths for all projections through the estimated image. The set of projections (or sinogram) generated from the estimated image then is compared with

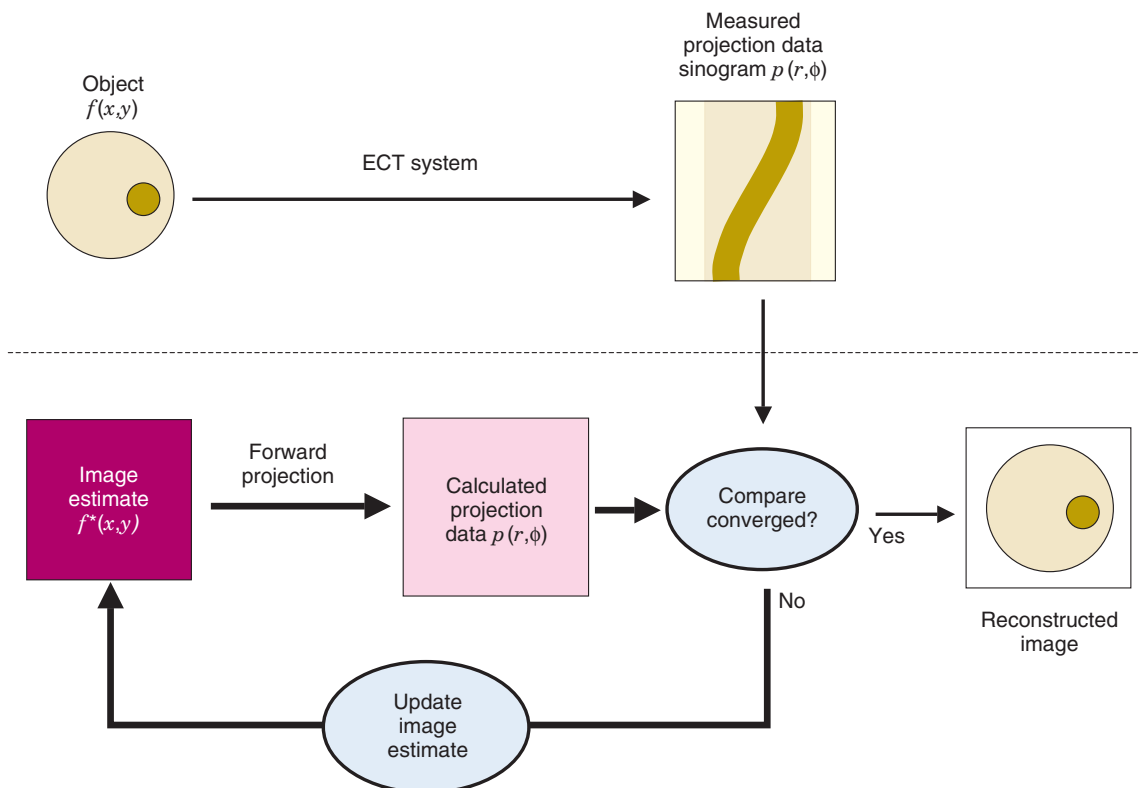


FIGURE 16-17 Schematic illustration of the steps in iterative reconstruction. An initial image estimate is made and projections that would have been recorded from the initial estimate are then calculated by forward projection. The calculated forward projection profiles for the estimated image are compared to the profiles actually recorded from the object and the difference is used to modify the estimated image to provide a closer match. The process is repeated until the difference between the calculated profiles for successively estimated images and the actually observed profiles reaches some acceptably small level.

the actually recorded projections (or sinogram). Most likely, they will not agree, because it is unlikely that the initial estimate of $f^*(x,y)$ closely resembles the true image. However, the difference between the estimated and actual projections can be used to adjust the estimated image to achieve closer agreement.

The update-and-compare process is repeated until the difference between the forward-projected profiles for the estimated image and the actually recorded profiles falls below some specified level. With proper design of the image updating procedure, the estimated image progressively converges toward the true image. Figure 16-18 shows the progress of the estimated image during iterative reconstruction with an increasing number of iterations.

The two basic components of iterative reconstruction algorithms are (1) the method for comparing the estimated and actual profiles and (2) the method by which the image is updated on the basis of this comparison. In generic terms, the first component is performed by the *cost function*, which measures the difference between the profiles generated by forward projections through the estimated

image and the profiles actually recorded from the scanned object. The second component is performed by the *search* or *update function*, which uses the output of the cost function to update the estimated image. A general goal of algorithm development is to devise versions of these functions that produce convergence of the estimated image toward the true image as rapidly and accurately as possible. One area of algorithmic differences is the method for dealing with statistical noise. For example, some algorithms give more weight to portions of projections (or sinograms) that contain the highest number of counts, and thus the lowest percentage levels of statistical noise (see Chapter 9, Section B.1). Another approach is to incorporate some sort of “prior information,” such as the expected shape or smoothness of the image. Some algorithms also “force” the reconstructed image to be non-negative. A concise history and review of iterative reconstruction methods are presented in reference 7.

Two factors make iterative reconstruction computationally more intensive than FBP. First, most iterative algorithms require several iterations to converge to an acceptable

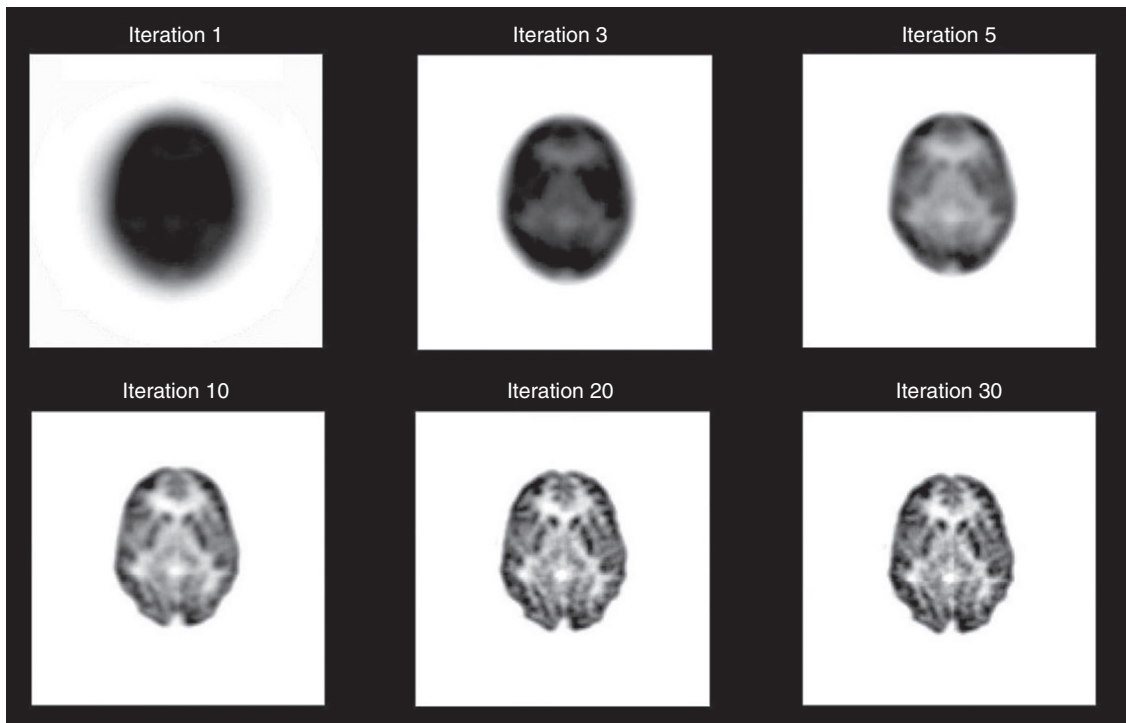


FIGURE 16-18 Brain images generated for different numbers of iterations by an iterative reconstruction algorithm. Image resolution progressively improves as the number of iterations increases. In practice, the iterations are performed until an acceptable level of detail is achieved or until further iterations produce negligible improvement. (Courtesy Dr. Richard Leahy, University of Southern California, Los Angeles, CA.)

image, and each of these iterations is essentially equivalent to a separate backprojection procedure. Backprojection is the most time-consuming part of the FBP algorithm but only needs to be done once for FBP. Forward projection is similarly time-consuming in iterative reconstruction algorithms.

Second, iterative algorithms often incorporate factors that account for the specific characteristics of the imaging device, such as collimator and object scatter, system geometry, and finite detector resolution. Simple forward projection along a single ray path no longer is used to calculate the projection profiles for the estimated image. Instead, all image pixels are considered to have a finite probability of contributing data to virtually all ray paths. In practice, very distant pixels might not be considered. Nevertheless, this adds to the computing time, because the reconstruction must include effects not only from pixels directly along a ray path but from pixels outside that ray path as well.

A number of methods have been developed to speed up these advanced algorithms. One of the most popular is called *ordered subsets*. In this method only a small number (or subset) of projection angles are used in the initial iterations. As the image is refined, a larger number of projection angles are included. This speeds up the algorithm, because the time per iteration is directly proportional to the number of projection profiles that must be computed. The ordered-subsets approach can be used to speed up both simple forward projection-based iterative algorithms as well as the advanced algorithms that use complex modeling of the imaging system.

Although they are more challenging to implement compared with FBP, iterative algorithms have the potential for providing quantitatively more accurate reconstructions. An example of one algorithm is presented in the following section.

2. Expectation-Maximization Reconstruction

The *expectation-maximization* (EM) algorithm incorporates statistical considerations to compute the “most likely,” or *maximum-likelihood* (ML), source distribution that would have created the observed projection data, including the effects of counting statistics. Specifically, it assigns greater weight to high-count elements of a profile and less weight to low-count regions. (By comparison, backprojection algorithms assign a uniform statistical weighting to all elements of a

profile.) Because of the statistical weighting factor, the algorithm often is referred to as the *ML-EM method*. A detailed discussion of this algorithm and its theoretical underpinnings are beyond the scope of this text but can be found in references 8 and 9. Here we present only a description of how it is implemented.

In the EM algorithm, the reconstruction process is formulated as follows

$$p_j = \sum_i M_{i,j} f_i \quad (16-19)$$

where f_i is the intensity (or activity) in the i^{th} pixel in the image, p_j is the measured intensity in the j^{th} projection element, and $M_{i,j}$ is the probability that radiation emitted from the i^{th} pixel will be detected in the j^{th} projection element. Note that, unlike previous uses of i and j to represent different (x,y) locations in a 2-D image (or a set of projections), the indices here each apply to the full set of the subscripted quantities. Thus, if the image is reconstructed on a grid of 128×128 pixels, the subscript i runs from 1 to 16,384 (128×128). If the imaging system records projections at 128 different angles around the object, and each projection has 256 elements, the index j runs from 1 to 32,768 (128×256). In essence, all of the image pixels and projection elements are “strung together” to form a single list for each set. The matrix M is very large, even for a single-slice image ($16,384 \times 32,768$ in the previous example). It can be extended to three dimensions as well, in which case it becomes even larger.

The matrix approach described above provides a potentially much more accurate model for relating projection profiles to the underlying source distribution than simple forward projection. The matrix could be determined by calculations, simulations, or a combination of both. For example, one could position a point source at all locations within the imaged slice (or volume) and record the counts in all elements of all possible projection profiles. However, this would be very time consuming. Symmetry considerations could somewhat shorten the project. In practice, many of the geometric effects can be calculated from simple models (e.g., collimator response—see Fig. 14–15) and others, such as collimator scatter, can be simulated or derived from theoretical models.

Once the matrix M has been determined and projection profiles have been recorded, the operating equation for computing the

estimated intensity value f of pixel i in the $(k + 1)^{\text{st}}$ iteration of the EM algorithm is as follows:

$$f_i^{k+1} = \frac{f_i^k}{\sum_j M_{i,j}} \times \sum_j M_{i,j} \frac{p_j}{\left(\sum_l M_{l,j} f_l^k \right)} \quad (16-20)$$

where k refers to the immediately preceding k^{th} iteration. The term in parentheses in the denominator on the right hand side of Eq. 16-20 represents a summation over all image pixels. This term must be evaluated first before the summation over the j projection elements can be computed. Therefore it is given a different pixel index, l , instead of i , to avoid confusion.

The number of iterations can be fixed, or the iteration process can be terminated when some measure of the difference between images from one iteration to the next (e.g., the sum of the squares of differences for all pixels in the reconstructed image) falls below some predetermined value. In theory, with perfectly measured noise-free data and an exact matrix M , the algorithm eventually would converge to the point where the estimated projection data, $\sum_l M_{l,j} f_l^k$ exactly equals the measured projection data, p_j , for each profile. At that point

$$f_i^{k+1} = f_i^k \quad (16-21)$$

that is, there is no further change in the estimated image and the estimated activity image is identical to the true activity distribution. In practice, this never happens, owing to inaccuracies or simplifications in M and statistical noise. Therefore some practical limit must be set for an acceptable difference that will be used to terminate the reconstruction process.

The computational issues relating to iterative reconstruction techniques already have been mentioned. Equation 16-20 (which represents only a single-slice version of the algorithm) illustrates this point. Nevertheless, the ML-EL algorithm can produce high-quality images with good quantitative accuracy and is now a selectable option on many PET and SPECT cameras.

The sampling and noise-propagation rules summarized in Section C do not apply to iterative reconstruction. Although insufficient sampling also has consequences for iterative algorithms, the aliasing and streaking artifacts associated with FBP are not seen. More typically, undersampling in the linear sampling distance or in the number

of angular views results in a more or less uniform loss of spatial resolution across the reconstructed image. If only partial angular coverage of the object is obtained (e.g., from 0 to 120 degrees instead of over the full 180 degrees), the resolution is likely to be degraded along the direction of the missing data. Because iterative algorithms are non-linear in nature, the exact effects of undersampling are object and algorithm dependent.

E. RECONSTRUCTION OF FAN-BEAM, CONE-BEAM AND PINHOLE SPECT DATA, AND 3-D PET DATA

The discussion thus far has focused on reconstructing projection data in which the acquired rays for a given projection angle are parallel and the projection data arises from parallel sections through the body. This is the situation when a parallel-hole collimator is used. Tomographic reconstruction also can be performed using data acquired with fan-beam, cone-beam, or pinhole collimators. The rationale for using these collimators is that they can provide higher spatial resolution (converging-hole or pinhole collimators) or greater coverage (diverging-hole collimators—see Chapter 14, Section D). However, these collimators introduce an added degree of complexity for reconstruction tomography of SPECT data, because they do not provide simple parallel-ray line integral projections such as were illustrated in Figure 16-2. Similar issues arise in PET scanning. In addition to acquiring projection data for transverse sections through the body, PET scanners, as discussed in Chapter 18, Section C, also are capable of acquiring additional projection data at oblique angles with respect to these transverse slices. Accurately incorporating this additional projection data requires 3-D reconstruction algorithms.

1. Reconstruction of Fan-Beam Data

We should first distinguish between fan-beam versus cone-beam collimators. Figure 16-19 schematically illustrates the difference. Consider first the *fan-beam collimator* shown at the top of the figure. In this collimator, each row of holes across the collimator has its own focal point. Sequential rows of collimator holes are stacked and evenly spaced, parallel to each other, along the z-axis of the object. Apart from overlapping coverage resulting from the finite diameters of the collimator

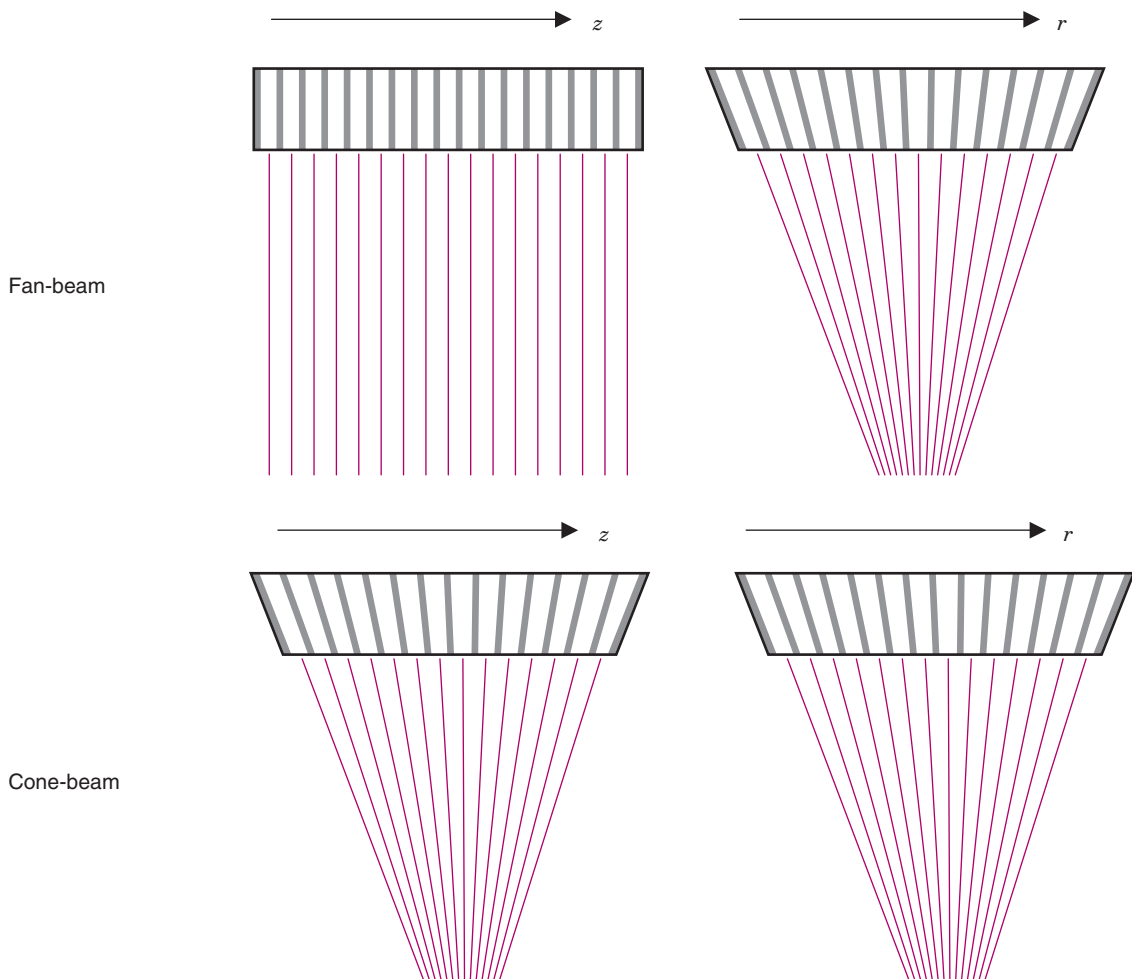


FIGURE 16-19 Schematic illustrations of fan-beam and cone-beam collimators. Cross-sections are shown for perpendicular viewing angles.

holes, each row of holes provides its own independent and nonoverlapping projection profile.

Data from a fan-beam collimator cannot be inserted directly into algorithms used for reconstructing data acquired with a parallel-beam collimator. However, the data can be rearranged so that these algorithms can be used. One approach is to re-sort the fan-beam data into parallel-beam data. Figure 16-20 illustrates how this is done for a few elements of adjacent projection profiles. Once the data have been re-sorted, any of the algorithms discussed in the preceding sections for parallel-beam collimators can be used. Alternative, and more elegant, approaches reformulate the FBP algorithm itself to handle fan-beam data. These are discussed in references 2 and 7.

A fan-beam collimator provides complete 3-D coverage of a volume of tissue in a single

rotation around the object. However, whereas complete coverage can be obtained with a 180-degree rotation using a parallel-hole collimator, the required rotation for a converging-beam collimator is $(180 + \theta)$ degrees, in which θ is half the fan angle for the collimator (see Fig. 14-20). Conversely, for a diverging collimator, the required angle of rotation is $(180 - \theta)$ degrees.

2. Reconstruction of Cone-Beam and Pinhole Data

In a *cone-beam collimator* (Fig. 16-19, bottom), all of the holes are directed toward (or away from) a common focal point. Each row of holes across the center of the collimator provides a projection profile, but the profiles all intersect at the center. (This also applies to the pinhole collimator.) It is not possible to re-sort the data acquired from a single rotation of a

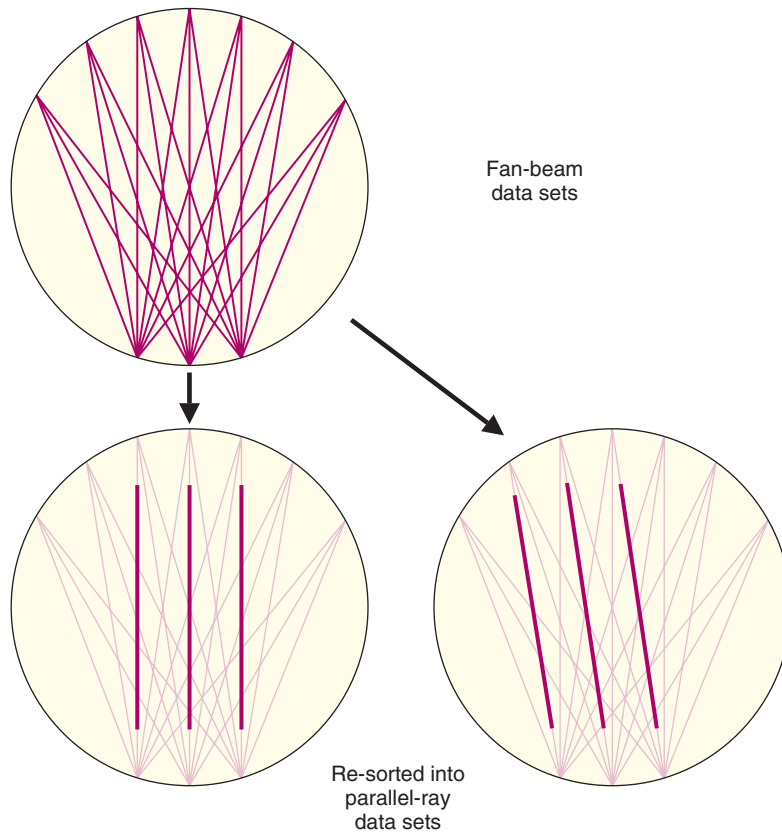


FIGURE 16-20 Procedure for creating parallel-beam projections from a set of fan-beam projections.

cone-beam collimator around the object into a full set of parallel-ray projections. Only one set (corresponding to the projections acquired from a single slice across the center of collimator, oriented perpendicular to the axis of rotation) can be re-sorted in this way. Therefore to obtain complete projection coverage of a volume of tissue to allow accurate reconstruction of multiple slices, a more complex rotation is required.

One approach is to perform a helical scan around the object, translating the collimator along the z-axis as it rotates about that axis. This provides a dataset that can be re-sorted into a complete set of parallel projections for multiple slices through the object. An alternative approach is to use approximations and interpolations to convert the cone-beam data into fan-beam data. The most popular of these methods is called the *Feldkamp algorithm*, described in reference 10. These methods work best when the cone angle is small.

Finally, iterative algorithms, conceptually similar to those described in Section C, have been developed for direct reconstruction of

3-D cone-beam data; however, computing time increases dramatically as compared with already time-consuming single-slice iterative algorithms. The matrix M (Equation 16-19) becomes very large for a full 3-D algorithm and, even with accelerated approaches and specialized computer hardware, full 3-D image reconstructions are typically at least an order of magnitude slower than multislice 2-D image reconstructions.

3. 3-D PET Reconstruction

PET scanners typically consist of multiple detector rings (see Chapter 18, Section B). Projection data acquired within a given detector ring can be reconstructed into a transverse image with the methods described previously in Sections B and D. However, PET scanners also can acquire projection data at oblique angles between detector rings (see Chapter 18, Section C and Fig. 18-24). To incorporate these additional projection angles requires some form of 3-D reconstruction algorithm. 3-D algorithms have been developed based on both FBP and iterative

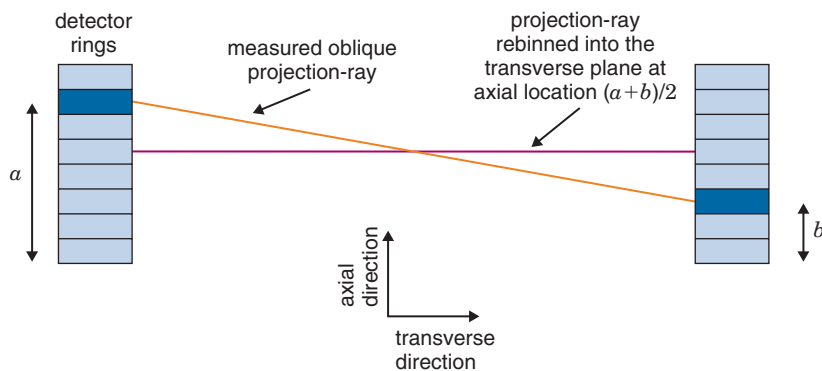


FIGURE 16-21 Illustration of single-slice re-binning in which an oblique projection-ray between the detector pair a and b is “re-assigned” to the projection data for the non-oblique slice corresponding to a transverse detector pair at axial location $(a + b)/2$.

reconstruction methods. A detailed description of these algorithms is beyond the scope of this text. However, some general concepts will be presented.

One common approach is to “re-bin” the 3-D dataset, such that each oblique projection ray is placed within the projection data for a particular nonoblique 2-D transverse slice. In effect, the 3-D dataset is collapsed back into a multislice 2-D dataset. The most simple method to accomplish this is to assign each ray to its average axial location.¹¹ Thus an oblique projection-ray between a detector at location a and a detector at location b would be positioned as if it were a projection from a directly opposed pair located halfway between them, i.e., at location $(a + b)/2$ (Fig. 16-21). Processing of all the projection rays in this manner results in a series of sinograms of parallel-ray projections, each corresponding to different axial locations through the object. Each sinogram can then be reconstructed using the 2-D FBP or iterative algorithms described previously. This method is known as single-slice rebinning.

Figure 16-21 illustrates that for events originating close to the center of the field of view (FOV) of the scanner, only small errors in positioning are made by this approximation. However, for events originating close to the edge of the scanner, and for projections at a large angle with respect to the transverse plane, significant mispositioning errors are made in the axial direction using this method. To overcome this, more accurate methods that include more sophisticated re-binning algorithms and an axial filtering step¹² or that use Fourier re-binning techniques¹³ have been developed.

An alternative approach is to formulate the iterative reconstruction equations (Eqs. 16-19 to 16-21), for fully 3-D reconstruction, thus implicitly accounting for the exact orientation of each line of response. Because of the additional dimensionality of the projection data, and the fact that the images are being reconstructed into a 3-D volume rather than a 2-D slice, the matrix M in Equation 16-19 becomes very large. The number of elements can be in the range of 10^{13} to 10^{15} with modern PET scanners that have large numbers of detector rings. Furthermore, the backprojection and forward-projection steps must now be performed in 3-D, tracing each ray through a volume rather than across a 2-D slice.

Thus the computational challenges are formidable, although great progress has now been made in reducing the matrix size using sparse storage techniques and symmetry arguments, and multiprocessor hardware and efficient coding have produced fast methods for 3-D backprojection and forward projection. Fully 3-D iterative algorithms are now available on some systems, especially small-animal imaging systems in which the small FOV typically leads to more manageable projection dataset sizes.

REFERENCES

1. Webb S: *From the Watching of Shadows: The Origins of Radiological Tomography*, Bristol, England, 1990, Adam Hilger.
2. Kak AC, Slaney M: *Principles of Computerized Tomographic Imaging*, Philadelphia, 2001, SIAM.
3. Herman GT: *Fundamentals of Computerized Tomography: Image Reconstruction from Projections*, ed 2, London, 2009, Springer-Verlag.
4. Natterer F: *The Mathematics of Computerized Tomography*, New York, 1986, Wiley.

5. Oppenheim AV, Wilsky AS: *Signals and Systems*, Englewood Cliffs, NJ, 1983, Prentice-Hall, pp 513-555.
6. Hoffman EJ, Phelps ME: Positron emission tomography: Principles and quantitation. In Phelps ME, Mazziotta JC, Schelbert HR, editors: *Positron Emission Tomography and Autoradiography: Principles and Applications for the Brain and Heart*, New York, 1986, Raven Press, pp 237-286.
7. Leahy RM, Clackdoyle R: Computed tomography. In Bovik A, editor: *Handbook of Image and Video Processing*, Burlington, MA, 2005, Elsevier Academic Press, pp 1155-1174.
8. Shepp LA, Vardi Y: Maximum likelihood reconstruction for emission tomography. *IEEE Trans Med Imag* 1:113-122, 1982.
9. Lange K, Carson R: EM reconstruction algorithms for emission and transmission tomography. *J Comput Assist Tomogr* 8:306-316, 1984.
10. Feldkamp LA, Davis, LC, Dress JW: Practical cone-beam algorithm. *J Opt Soc Am* 1:612-619, 1984.
11. Daube-Witherspoon ME, Muehllehner G: Treatment of axial data in three-dimensional PET. *J Nucl Med* 28:1717-1724, 1987.
12. Lewitt RM, Muehllehner G, Karp JS: Three-dimensional image reconstruction for PET by multi-slice rebinning and axial image filtering. *Phys Med Biol* 39:321-339, 1994.
13. Defrise M, Kinahan PE, Townsend DW, Michel C, Sibomana M, Newport DF: Exact and approximate rebinning algorithms for 3D PET data. *IEEE Trans Med Imag* 16:145-148, 1997.