CHAPTER 11

NUCLEAR MEDICINE IMAGING DEVICES

M.A. LODGE, E.C. FREY
Russell H. Morgan Department of Radiology and Radiological Sciences,
Johns Hopkins University,
Baltimore, Maryland, United States of America

11.1. INTRODUCTION

Imaging forms an important part of nuclear medicine and a number of different imaging devices have been developed. This chapter describes the principles and technological characteristics of the main imaging devices used in nuclear medicine. The two major categories are gamma camera systems and positron emission tomography (PET) systems. The former are used to image γ rays emitted by any nuclide, while the latter exploit the directional correlation between annihilation photons emitted by positron decay. The first section of this chapter discusses the principal components of gamma cameras and how they are used to form 2-D planar images as well as 3-D tomographic images (single photon emission computed tomography (SPECT)). The second section describes related instrumentation that has been optimized for PET data acquisition. A major advance in nuclear medicine was achieved with the introduction of multi-modality imaging systems including SPECT/computed tomography (CT) and PET/CT. In these systems, the CT images can be used to provide an anatomical context for the functional nuclear medicine images and allow for attenuation compensation. The third section in this chapter provides a discussion of the principles of these devices.

11.2. GAMMA CAMERA SYSTEMS

11.2.1. Basic principles

The gamma camera, or Anger camera [11.1], is the traditional workhorse of nuclear medicine imaging and its components are illustrated in Fig. 11.1. Gamma camera systems are comprised of four basic elements: the collimator, which defines the lines of response (LORs); the radiation detector, which counts incident γ photons; the computer system, which uses data from the detector to create 2-D

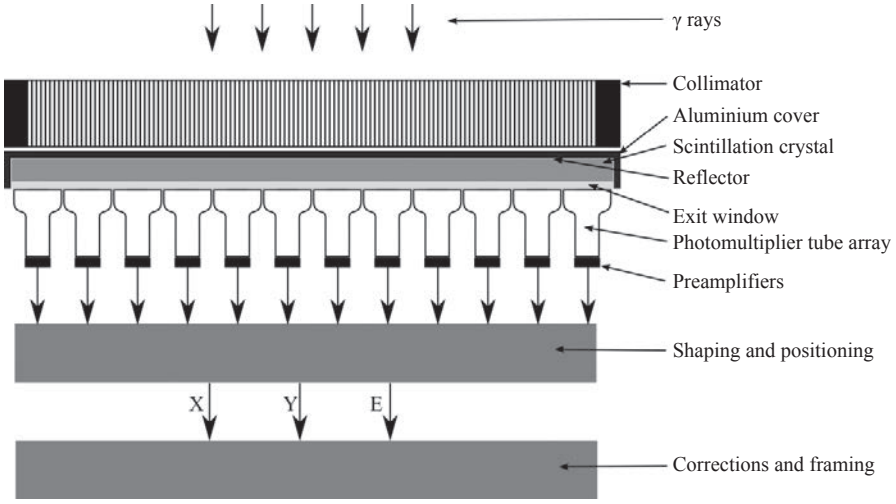

FIG. 11.1. Schematic diagram showing the major components of a gamma camera.

histogram images of the number of counted photons; and the gantry system, which supports and moves the gamma camera and patient. The overall function of the system is to provide a projection image of the radioactivity distribution in the patient by forming an image of γ rays exiting the body. Forming an image means establishing a relationship between points on the image plane and positions in the object. This is sometimes referred to as an LOR: ideally, each position in the image provides information about the activity on a unique line through the object. In gamma cameras, single photons are imaged, in contrast to PET where pairs of photons are imaged in coincidence. Thus, in order to define an LOR, a lens is required, just as in optical imaging. However, the energies of γ rays are so high

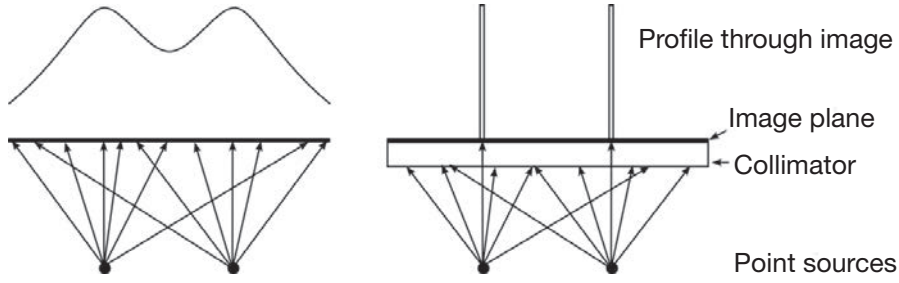

FIG. 11.2. The drawing on the left demonstrates the image of two point sources that would result without the collimator. It provides very little information about the origin of the photons and, thus, no information about the activity distribution in the patient. The drawing on the right illustrates the role of the collimator and how it defines lines of response through the patient. Points on the image plane are uniquely identified with a line in space.

that bending them is, for all practical purposes, impossible. Instead, collimators are used to act as a mechanical lens. The function of a collimator is, thus, to define LORs through the object. Figure 11.2 illustrates the function of and need for the collimator, and, thus, the basic principle of single photon imaging.

11.2.2. The Anger camera

11.2.2.1. Collimators

As mentioned above, the collimator functions as a mechanical lens: it defines LORs. The collimator accomplishes this by preventing photons emitted along directions that do not lie along the LOR from reaching the detector. Thus, collimators consist of a set of holes in a dense material with a high atomic number, typically lead. The holes are parallel to the LORs. Ideally, each point in the object would contribute to only one LOR. This requires the use of collimator holes that are very long and narrow. However, such holes would allow very few photons to pass through the collimator and be detected. Conversely, increasing the diameter or decreasing the length of the holes results in a much larger range of incident angles passing through the collimator. As illustrated in Fig. 11.3, this results in degraded resolution. As can be seen from this figure, each hole has a cone of response and the diameter of the cone of response is proportional to the distance from the face of the collimator.

As discussed above, changing the diameter of collimator holes changes the resolution and also the number of photons passing through the collimator. The noise in nuclear medicine images results from statistical variations in the number of photons counted in a given counting interval due to the random nature of radiation decay and interactions with the patient and camera. The noise is described by Poisson statistics, and the coefficient of variation (per cent noise) is inversely proportional to the square root of the number of counts. Thus, increasing the number of counts results in less noisy images. As a result, there is an inverse relationship between noise and spatial resolution for collimators: improving the resolution results in increased image noise and vice versa.

Another important characteristic of collimators is the opacity of collimator septa to incident γ rays. In an ideal collimator, the septa would block all incident radiation. However, in real collimators, some fraction of the radiation passes through or scatters in the septa and is detected. These phenomena are referred to as septal penetration and scatter. The amount of septal penetration and scatter depends on the energy of the incident photon, the thickness and composition of the septa, and the aspect ratio of the collimator holes. Since gamma cameras are used to image radionuclides with energies over a wide range, collimators are typically available that are appropriate for several energy ranges: low energy

collimators are designed for isotopes emitting photons with energies lower than approximately 160 keV; medium energy collimators are designed for energies up to approximately 250 keV; and high energy collimators are designed for higher energies. It should be noted that in selecting the appropriate collimator for an isotope, it is important to consider not only the photon energies included in the image, but also higher energy photons that may not be included in the image. For example, in ^{123}I there are a number of low abundance high energy photons that can penetrate through or scatter in septa and corrupt the images. As a result, medium energy collimators are sometimes used for ^{123}I imaging, despite the main γ photopeak at 159 keV.

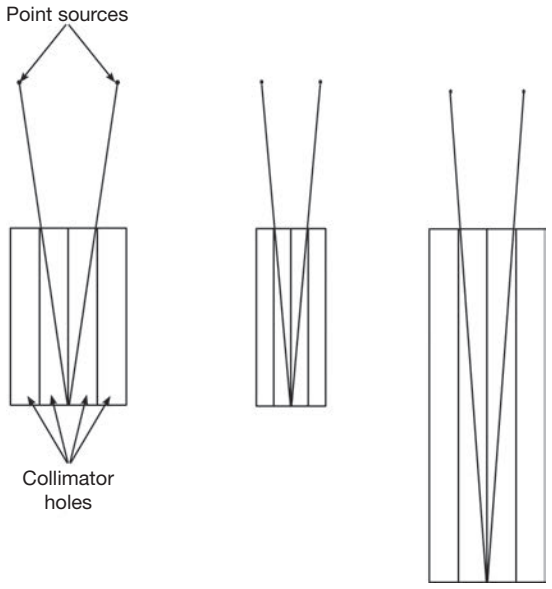

FIG. 11.3. Illustration of the concept of spatial resolution and how collimator hole length and diameter affect spatial resolution. The lines from the point source through the collimator indicate the furthest apart that two sources could be and still have photons detected at the same point on the image plane (assumed to be the back of the collimator). Thus, sources closer together than this would not be fully resolved (though they might be partially resolved). From this, we see that the resolution decreases as a function of distance. It should also be noted that the resolution improves proportionally with a reduction in the width of the collimator holes and improves (though not proportionally) with the hole length.

For multi-hole collimators, hole shape is an additional important factor in collimator design. The three most common hole shapes are shown in Fig. 11.4. Round holes have the advantage that the resolution is uniform in all directions in planes parallel to the face of the collimator. However, as discussed below, the

sensitivity is relatively low because there is less open area for a given resolution and septal thickness. Hexagonal hole collimators are the most common design for gamma cameras using continuous crystals. They have the advantage of relatively direction independent response functions and higher sensitivity than a round hole collimator with the same resolution and septal thickness. Square hole collimators are especially appropriate for detectors that have pixelated crystals. Having squares holes that match the spacing and inter-gap distance of these detectors results in good sensitivity with these detectors. However, the resolution varies significantly depending on the direction, being worse by a factor of a $\sqrt{2}$ along the diagonal direction compared to parallel to the rows of holes.

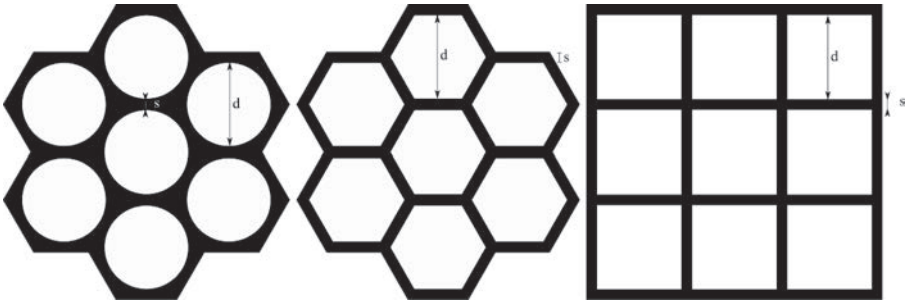

FIG. 11.4. Examples of the three major hole shapes used in multi-hole collimators. They are, from left to right: round, hexagonal and square holes. In all cases, black indicates septa and white open space. The diameter (often called flat-to-flat for square and hexagonal holes) is d and the septal thickness is s.

Multi-hole collimators typically have many thousands of holes. The uniformity of the image critically depends on the holes having uniform sizes and spacing. As a result, high quality fabrication is essential. The septa must be made of a high density, high *Z* material in order to stop the incident γ rays. Lead is the material of choice for most multi-hole collimators due to its relatively low cost, low melting temperature and malleability. Lead multi-hole collimators can be divided into two classes based on fabrication techniques: cast and foil collimators.

Fabrication of cast collimators involves the use of a mould that is filled with molten lead to form the collimator. The mould typically consists of two plates with holes at opposing positions that match each of the holes in the collimator. A set of pins is placed in the holes. Lead is then poured between the plates and the entire assembly is then carefully cooled. The plates and pins are removed leaving behind the collimator. This technology is especially well suited to making high and medium energy collimators as well as special purpose collimator geometries.

Foil collimators are created from thin lead foils. The foils are stamped and then glued together to build up the collimator layer by layer. Figure 11.5 shows a schematic of how two layers are stamped and glued to form the holes. It should be noted that in the stamping the septa that are glued must be thinner than the other walls in order to retain uniform septal thickness and, thus, maximize sensitivity. It is clear that precise stamping, alignment and gluing is essential to form a high quality collimator. The septa in foil collimators can be made thinner than in cast collimators. As a result, foil fabrication techniques are especially appropriate for low energy collimators. Understanding the fabrication technology can help in diagnosing problems with the collimator. For example, Fig. 11.6 shows an image with non-uniformities appearing as vertical stripes in the image of a sheet source. This was a foil collimator and the non-uniformities apparently originated from fabrication problems, resulting in some layers having different sensitivities compared to other layers.

FIG. 11.5. Illustration of fabrication of foil collimator by gluing two stamped lead foils. It should be noted that the foils must be stamped so that the portions of the septa that are glued are half the thickness of the rest of the septa. Furthermore, careful alignment is essential to preserve the hole shapes.

FIG. 11.6. Uniformity image of a defective foil collimator. The vertical stripes in the image result from non-uniform sensitivity of the collimator due to problems in the manufacturing process. The peppery texture is due to quantum noise and is visible because the grey level was expanded to demonstrate the non-uniformity artefacts.

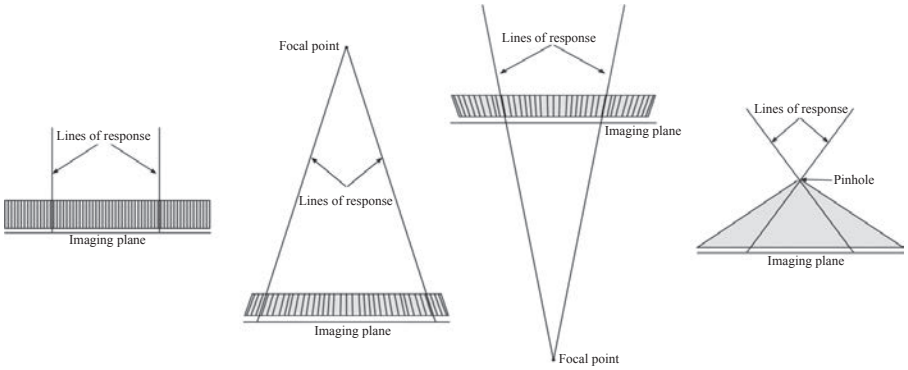

FIG. 11.7. Illustration of the four common collimator geometries: (left to right) parallel, converging, diverging and pinhole.

A final important characteristic of collimators is the hole geometry. There are four common geometries in nuclear medicine: parallel, converging, diverging and pinhole, as illustrated in Fig. 11.7. Parallel holes are the most commonly used collimators. The LORs are parallel, and there is, thus, a one to one relationship between the size of the object and image. In converging hole collimators, the LORs converge to a point (focal point) or line (focal line) in front of the collimator, and there is, thus, magnification of the image. These two collimator types are referred to as cone-beam and fan-beam collimators, respectively. These are useful for imaging small objects, such as the heart or brain, on a large camera as they provide an improved trade-off between spatial resolution and noise. In diverging hole collimators, the LORs converge to a point or line behind the collimator. This results in minification of the image. Diverging hole collimators are useful for imaging large objects on a small camera. However, they result in a poor resolution versus noise trade-off, and are, thus, infrequently used. Pinhole collimators use a single hole to define the LORs. In terms of geometry, they are similar to cone-beam collimators but with the focal point between the image plane and the object being imaged. As a result of this, the image is inverted compared to the object. In addition, the object can be either minified or magnified depending on whether the distance from the image plane to the focal point is less than or greater than the distance from the pinhole to the object plane. Pinhole collimators provide an improved resolution noise trade-off when objects are close to the pinhole. They are, thus, useful for imaging small objects such as the thyroid or small animals. Another advantage of pinholes is that there is only a single hole, referred to as an aperture, which determines the amount of collimator penetration and scatter. As a result, it is possible to fabricate the aperture from high density and high atomic number materials (e.g. tungsten, gold or depleted uranium), which can reduce

collimator penetration and scatter. This makes these collimators appropriate for imaging radionuclides emitting high energy γ rays such as ^{131}I. In addition, pinhole collimators with changeable apertures can have different diameter pinholes. This allows selection of resolution/sensitivity parameters relatively easily.

The collimator properties can be most completely described by the collimator point source response function (PSRF), the noise-free image of a point source in air with unit activity using an ideal radiation detector, as a function of position in the object space. The shape of the collimator PSRF completely describes the resolution properties, and when normalized to unit total volume is referred to as the collimator point spread function (PSF).

Figure 11.8 shows some sample collimator PSFs for an ^{131}I point source imaged with a high energy general purpose collimator and a medium energy general purpose collimator. These PSFs are averaged over the position of the source with respect to the collimator and, thus, do not show the hole pattern. There are several things to note from this figure. First, using a properly designed collimator reduces septal scatter and penetration to very low levels, while they become significant for a collimator not designed to handle the high energies of ^{131}I. Second, the response function becomes wider as a function of distance, demonstrating the loss of resolution as a function of increasing source to collimator distance. Finally, there is evidence of the shape of the holes, which, in this case, were hexagonal. The shape can be barely discerned in the shape of the central portion of the response, which is due to photons passing through the collimator holes. The geometry of the collimator is more evident in the septal

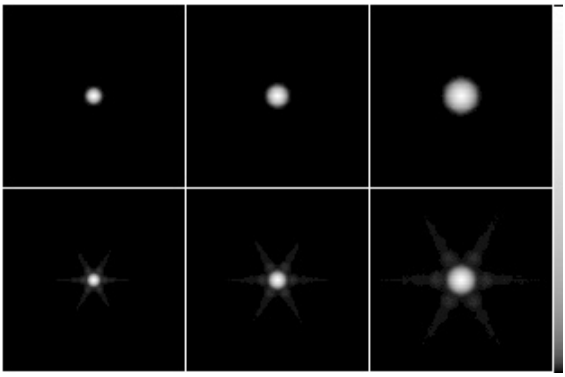

FIG. 11.8. Sample images of the point spread function for an ^{131}I point source at (left to right) 5, 10 and 20 cm from the face of a high energy general purpose collimator (top row) and a medium energy general purpose collimator (bottom row). The images are displayed on a logarithmic grey scale to emphasize the long tails of the point spread function resulting from septal penetration and scatter. The brightness of the image has been increased to emphasize the septal penetration and scatter artefacts.

penetration and scatter artefacts. In fact, the septal penetration is highest along angular directions where the path through the septa is thinnest, giving rise to the spoke-like artefacts in the directions perpendicular to the walls of the hexagonal holes.

Another useful way to describe and understand the resolution properties of the collimator is in terms of its frequency response. This can be described by the collimator modulation transfer function, which is the magnitude of the Fourier transform of the collimator PSF. Figure 11.9 shows some sample profiles through the collimator modulation transfer function. It should be noted that the collimator response does not pass high frequencies very well and, for some frequencies, the response is zero. This attenuation of high frequencies results in a loss of fine detail (i.e. spatial resolution) in the images. Finally, the cut-off frequency decreases with distance from the collimator and different collimator designs have different frequency responses.

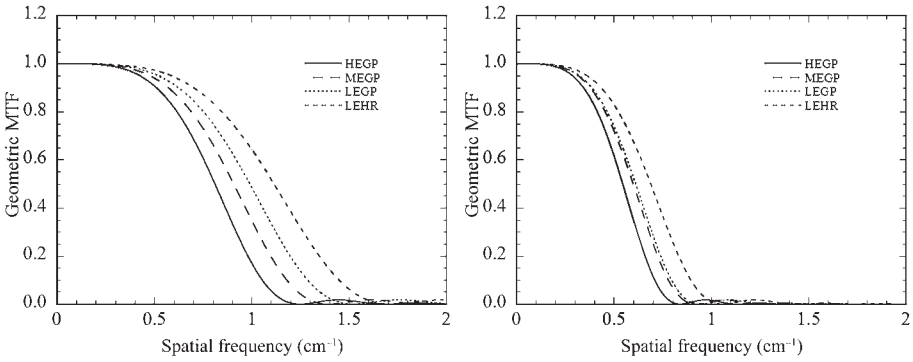

FIG. 11.9. Sample profile through the geometric modulation transfer functions (MTFs) for low, medium and high energy (HE, ME and LE, respectively) general purpose (GP) and high resolution (HR) collimators for a source 5 cm (left) and 20 cm (right) from the face of the collimator.

It is often desirable to summarize the collimator resolution in terms of one or two numbers rather than the entire response function. This is often done in terms of the width of the collimator PSRF at a certain fraction of its maximum value. For example, Fig. 11.10 shows a sample profile through a collimator PSF and the position of the full width at half maximum (FWHM) and the full width at tenth maximum (FWTM). To good approximation, the FWHM of a collimator is proportional to the distance from the face of the collimator. This holds for all distances except those very close to the collimator face, as illustrated in Fig. 11.11. The FWTM is useful for assessing the amount of septal penetration and scatter that are present. For a Gaussian response function, which is a good

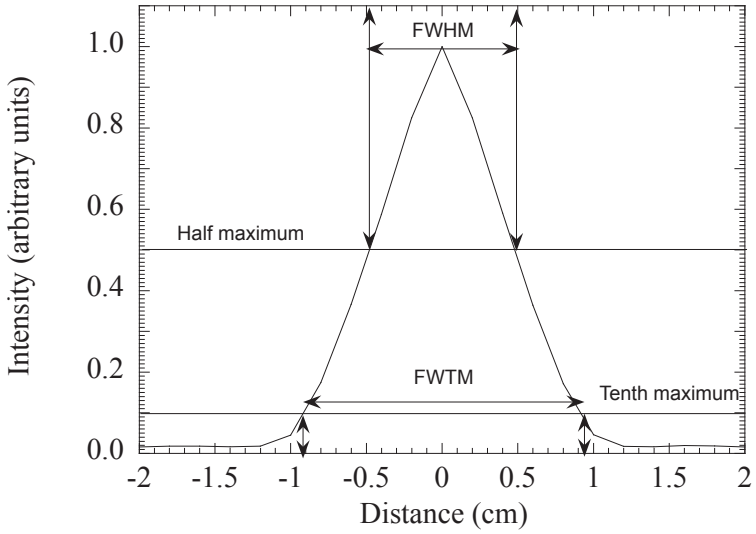

FIG. 11.10. Plot of the total collimator–detector point spread function for a medium energy general purpose collimator imaging ^{131}I, indicating the positions of the full width at half maximum (FWHM) and the full width at tenth maximum (FWTM).

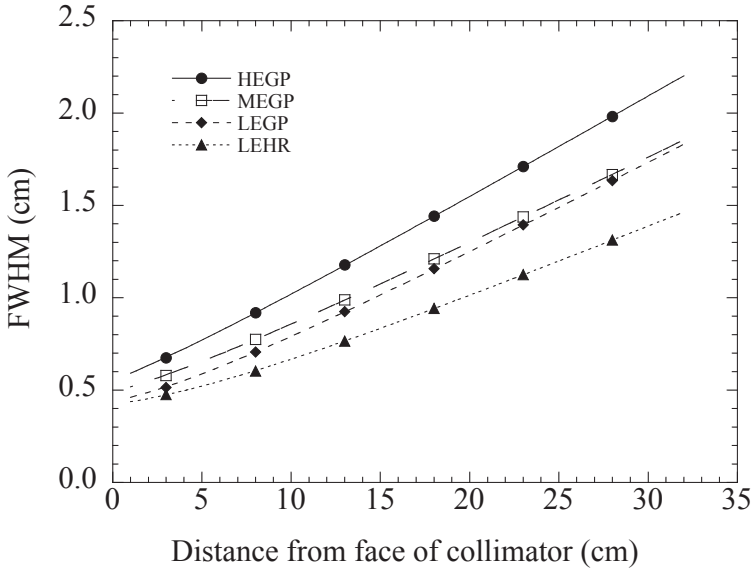

FIG. 11.11. Plot of the full width at half maximum (FWHM) of the geometric collimator–detector point source response function including a Gaussian intrinsic response with an FWHM of 4 mm as a function of distance from the face of the collimator for the same set of collimators described in Fig. 11.9.

approximation for the combination of a Gaussian intrinsic detector response with the geometric collimator response, the relationship between the FWHM and FWTM is given by FWTM/FWHM ≈ 1.86. Thus, if the FWTM is substantially larger than this factor times the FWHM, the response has been affected by factors other than the geometric response, such as septal penetration and scatter. It should be noted that the effects of septal penetration and scatter on the FWTM are less visible in a PSRF than they are in a line source response function.

The FWHM of the collimator resolution can be estimated from geometric arguments. Figure 11.12 shows a schematic that can be used to derive the resolution for a point source a distance Z from the face of the collimator. The collimator hole has a length L and a width d. The image plane (often taken to be the mean path of the primary photons in the crystal plus any physical gap) is a distance B behind the back face of the collimator. The photon passing through the collimator holes with the most oblique angle of incidence will have an incident angle defined by $\tan \theta = d/L$. Thus, the extreme limits of the response function will be defined by this limit. If it is assumed that the geometric response function is triangular in shape, then the FWHM in Fig. 11.12 will be half of this distance. Using similar triangles, it can be shown that the FWHM is given by:

$$\mathrm{FWHM} = \frac{d}{L}(Z + L + B) \tag{11.1}$$

Thus, it can be seen that, as described above, the FWHM is linearly related to the distance from the face of the collimator and is proportional to that distance when the distance is large compared to $L + B$.

The resolution of the collimator–detector system depends on both the resolution of the collimator and the intrinsic resolution of the gamma camera. For continuous-crystal gamma cameras, the intrinsic resolution can be modelled with a Gaussian function. In this case, the total response function for the collimator–detector is the convolution of the intrinsic resolution and the collimator response. If the collimator response is approximated by a Gaussian function, the FWHM is given by the Pythagorean sum of the intrinsic and collimator FWHMs:

$$\mathrm{FWHM}_{\mathrm{total}} = \sqrt{\mathrm{FWHM}^2_{\mathrm{collimator}} + \mathrm{FWHM}^2_{\mathrm{intrinsic}}} \tag{11.2}$$

Figure 11.11 shows a plot of the total geometric FWHM resolution for several collimators including the effects of a 4 mm Gaussian intrinsic resolution. The curves through the points represent a fit with a function of the above Pythagorean sum. It should be noted that except for distances close to the collimator, the resolution is linear with distance, indicating that the total resolution is dominated by the collimator resolution.

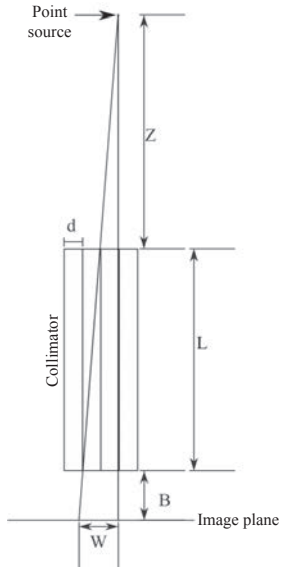

FIG. 11.12. Diagram illustrating the collimator geometry used to derive the expression for the full width at half maximum.

The integral of the collimator PSRF gives the sensitivity of the collimator. This is, in principle, a dimensionless quantity which gives the fraction of emitted photons that pass through the collimator, and is of the order of 10^{-3}–10^{-4} for typical nuclear medicine collimators. It is often useful to express the sensitivity in terms of counts per unit activity per unit time, for example counts per second per megabecquerel. For a parallel-hole collimator, the sensitivity is a function of two terms: the solid angle of the hole, which is a function of $(d/L)^2$, and the fraction of the active area of the collimator that is open area (hole) as compared to septa. The second term can be described in terms of the unit cell: the smallest geometric region that can be used to form the entire collimator by a set of simple translations. The sensitivity S of a parallel-hole collimator is given by:

$$S = \frac{a_{\text{open}}}{4\pi L^2}\frac{a_{\text{open}}}{a_{\text{total}}} \tag{11.3}$$

where

a_{open} is the open area of a collimator unit cell, i.e. the area of the hole itself;

and a_{total} is the total area of the cell including the part of the collimator septa lying in the unit cell.

From the above it can be seen that, for a parallel-hole collimator, the sensitivity is independent of the distance to the collimator face. This is because there is a balance between the decrease in sensitivity from a single hole and the increase in the number of holes through which photons can pass as a function of increasing distance from the collimator face. It should also be noted that a_{open} is proportional to d^2, which is proportional to the FWHM resolution. The term a_{open}/a_{total} varies slowly as a function of d if $d \gg s$. Thus, the sensitivity is proportional to the square of the resolution:

$$S = k \times \mathrm{FWHM}^2 \tag{11.4}$$

where the constant k depends only weakly on the FWHM. Since noise is directly related to the number of counts, there is a fundamental trade-off between resolution and noise. From the above, it is also evident that maximizing the ratio of a_{open}/a_{total} is important in terms of reducing noise for a given resolution.

11.2.2.2. Scintillation crystals

The scintillation crystal in the gamma camera converts γ ray photons incident on the crystal into a number of visible light photons. The characteristics and principle of scintillation radiation sensors are described in more detail in Chapter 6. Ideally, the crystal would be dense and of a high Z material in order to stop all incoming γ rays with photoelectric events. It should have high light output to provide low quantum noise for energy and position estimation. The decay time of the light output needs to be fast enough to avoid a pile-up of pulses at the count rates experienced in nuclear medicine imaging procedures. The wavelength spectrum of the scintillation photons should be matched to the sensitivity of the photodetectors used to convert the scintillation signal into an electrical signal. In addition to these technical properties that directly affect image quality, there are a number of desirable material properties that influence the cost of the device. These include the cost of the raw material, the ease of growing large single crystals and the sensitivity to environmental factors such as humidity. Owing to its unique combination of desirable properties, the crystals used in gamma cameras based on photomultiplier tubes (PMTs) are typically made of NaI(Tl). Gamma cameras based on solid state photodetectors require a different light spectrum and typically CsI(Tl) is used in these devices. The scintillation properties of these materials are discussed in detail in Chapter 6.

As will be described below, the interaction position of the γ ray with the detector is estimated based on the distribution of the scintillation light to an array of PMTs. It is important that the distribution of light be as independent as possible of the depth of interaction in the crystal and depends in a predictable

way on the lateral position. Further, absorption of scintillation photons by defects in the crystal is highly undesirable, as it will adversely affect the accuracy and precision of energy and position estimation. Thus, in order to make the lateral light distribution predictable, as even as possible and to minimize absorption, Anger cameras employ a large single crystal equal to the size of the field of view (FOV) of the camera. This can be as large as 60 cm × 40 cm in modern cameras. As NaI(Tl) is hygroscopic, it is important to hermetically seal the crystal in an enclosure. The back of the crystal must be optically coupled to the photodetector, so that the back part of the crystal enclosure consists of an optical glass exit window optically coupled to the crystal. The exit window lies between the crystal and the photodetector array. It serves several functions. First, it serves to hermetically seal the crystal. Second, the exit window allows scintillation photons to pass from the crystal into the photodetector array, and, thus, must be transparent in the emission range of the scintillator. To reduce internal reflections, it is desirable that the index of refraction be matched as closely as possible to that of the scintillator (n = 1.85 for NaI(Tl)) and the photodetector ($n \approx 1.5$ for borosilicate glasses used in the entrance windows of PMTs). The remainder of the enclosure should be light-tight to block out ambient light. The front face should be thin and of a low Z material — typically Al is used — to reduce the probability of incident γ ray absorption. Finally, to help collect incident light photons and improve energy resolution, the inside of the enclosure is a reflective layer. The use of specular versus diffusive reflectors affects the nature of the light response and has an impact on the variation in the light response on the photodetectors as a function of interaction depth, and, thus, impacts the precision of the position estimation.

One important parameter of the scintillation crystal related to camera performance is its thickness. The thickness is a trade-off between two characteristics: intrinsic resolution and sensitivity. The intrinsic resolution depends on the crystal thickness via the variation in the light distribution as a function of depth of interaction. Since the depth of interaction can vary over a wider range in a thicker crystal, there will be a larger variation in the light distribution and, thus, a larger uncertainty in the estimated lateral position of the interaction. In other words, thicker crystals generally have poorer intrinsic resolution. The functional relationship between the thickness and intrinsic resolution is complicated and depends on the details of the surface treatment of the scintillator, the photodetector array and the position estimation algorithm. For GE Millenium VG cameras, the FWHM intrinsic resolution for 140 keV photons using 0.953, 1.587 and 2.54 cm thick crystals is 3.5, 3.9 and 5.2 mm, respectively.

The intrinsic sensitivity of the crystal is related to crystal thickness by:

$$S_i = 1 - e^{-\mu t} \tag{11.5}$$

where

S_i is the intrinsic sensitivity;
μ is the linear attenuation coefficient of the crystal;

and t is the crystal thickness.

Since the linear attenuation coefficient decreases with energy, the intrinsic sensitivity also decreases with energy. Figure 11.13 shows a plot of the intrinsic sensitivity as a function of energy for several crystal thicknesses. For 140 keV, the sensitivity is ~92% for a 0.953 cm (3/8 in) thick crystal. This is the most common crystal thickness in commercial systems, though cameras with 5/8 and 1 in are available. These crystal thicknesses provide substantially improved sensitivity for radionuclides emitting medium and high energy photons such as ^{111}In and ^{131}I at the cost of a relatively minor reduction in intrinsic spatial resolution.

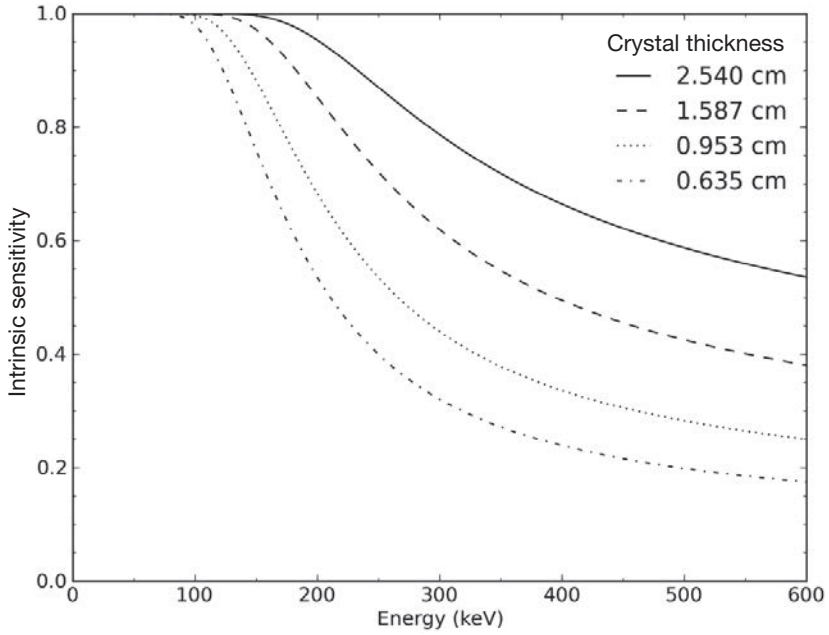

FIG. 11.13. Plot of the intrinsic sensitivity of a NaI scintillation crystal as a function of energy for several crystal thicknesses.

A final important property of the scintillation crystal is the light output. This is a characteristic of the scintillator material, and is the number of scintillation photons per unit energy deposited in the crystal by a γ photon. Thus, the total light is proportional to the energy deposited in the crystal, and can, therefore, be used to estimate the energy of the γ ray. The number of scintillation photons produced for a given event is a Poisson random variable. Thus, the larger the number of scintillation photons the smaller the coefficient of variation (standard deviation divided by the mean) of the mean number of photons, and, hence, the estimated photon energy. Thus, scintillators with high light output will provide higher energy resolution. In addition, as will be seen below, the light distribution over the photodetector array is used to estimate the interaction position. Since the light collected by each element in the array is also a Poisson random variable that is proportional to the light output, a larger light output will result in higher precision in the estimated position, and, thus, improved intrinsic spatial resolution. One reason that NaI(Tl) is used in gamma cameras is its high light output.

11.2.2.3. Photodetector array

The next element in the radiation detector is the photodetector array. This array measures the distribution of scintillation photons incident on the array and converts it into a set of pulses whose charge is proportional to the number of scintillation photons incident on each corresponding element in the array. As described below, the output of this array is used to compute the interaction position of the γ ray in the scintillator. In clinical gamma cameras, the photodetector array is comprised of a set of 30–90 PMTs arranged in a hexagonal close packed arrangement, as illustrated in Fig. 11.14. More details on the operation and characteristics of PMTs are provided in Chapter 6. In brief, PMTs have the advantage that they are very well understood, have a moderate cost, are relatively sensitive to low levels of scintillation light and have a very high gain. In some commercial designs, PMTs have been replaced by semiconductor detectors such as photodiodes. Generally, these devices are somewhat less sensitive and have a lower gain than PMTs, resulting in more noise in the charge signal and, thus, less precision in the energy and position estimated from the charge signal.

Since the position and energy are estimated from the set of charge signals from the elements in the photodetector array, it is highly desirable that the proportionality constants relating light intensity to charge be the same for all of the photodetectors. This can be ensured by choosing matching devices and by carefully controlling and matching the electronic gain. For PMTs, the gain is controlled by the bias voltage applied to the tubes. Since gain is also a function of temperature, the temperature of the photodetectors must be carefully controlled. The gains of PMTs are very sensitive to magnetic fields, even those as small

as the Earth's magnetic field. Thus, the PMTs must be magnetically shielded using mu-metal. Finally, since the gains of tubes can drift over time, periodic recalibration is necessary.

One of the major advantages of the gamma camera is that the number of PMTs is much smaller than the number of pixels in images from the gamma camera. In other words, in contrast to semiconductor detectors where a separate set of electronics is required for each pixel, the gamma camera achieves a great reduction in cost and complexity by estimating the interaction position of the γ ray based on the output of the array of PMTs.

To understand the position estimation process, Fig. 11.15 is considered. This figure shows a cross-section through two PMTs, and the crystal and exit window. The number of photons collected by a PMT directly (i.e. without reflection) will be proportional to the solid angle subtended by it at the interaction point. As can be seen in the figure, the interaction position is offset to the right, and there is a smaller solid angle subtended by PMT 1 than by PMT 2. Thus, the signal from PMT 1 will be smaller than for PMT 2. If the interaction position moves to the left, so that it lies along the line separating the two PMTs, there will be an equal amount of light collected by each PMT. The relationship between the light collected by the two PMTs and the lateral interaction position can be used to estimate the interaction position, as will be described in more detail below. In addition, the total scintillation light collected by all of the PMTs is proportional to the energy deposited by the γ ray in the crystal. Thus, the total charge can be used to estimate the energy of the photon.

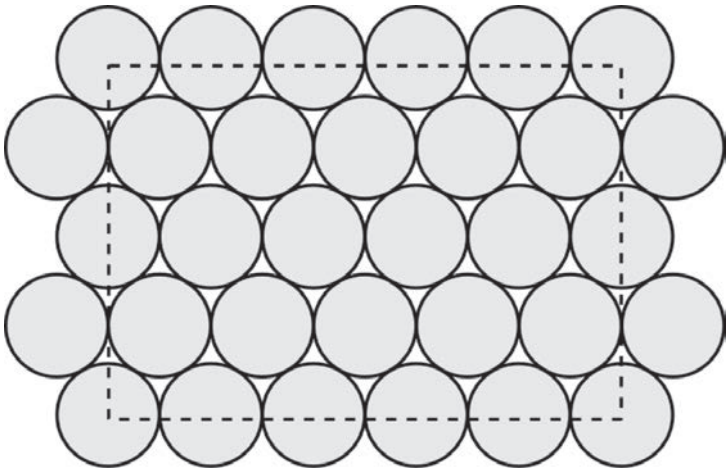

FIG. 11.14. Cross-section of a gamma camera at the back face of the entrance window showing the hexagonal close packed array of photomultiplier tubes. The dotted line indicates the approximate region where useful images can be obtained.

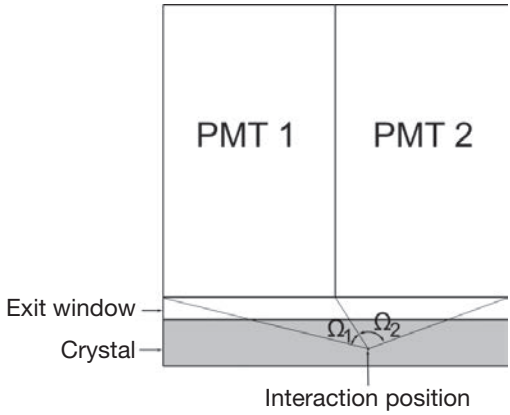

FIG. 11.15. Cross-section through two photomultiplier tubes (PMTs), the exit window and crystal in a gamma camera. The interaction position of a γ ray photon is indicated. The solid angles subtended by PMT 1 and 2 are Ω_1 and Ω_2, respectively.

11.2.2.4. Amplifiers and pulse shaping

The charge pulse from each PMT is very small and, thus, subject to noise. In addition, the scintillation photons are emitted randomly over a finite time (given by the scintillator's decay constant), making the charge pulse rather noisy. To make subsequent analysis of the pulse easier and more resistant to electrical noise, the pulse is amplified and shaped prior to processing to estimate the interaction position and photon energy. The components of this stage are a preamplifier and shaping amplifier.

The preamplifier integrates the charge pulse from the PMT to form a voltage pulse with height proportional to the input charge. The design of the preamplifier should be such that the voltage height is as independent as possible of the details of the charge pulse, such as decay and rise times. Preamplifiers are typically mounted directly to the PMT outputs in order to avoid corruption of the tiny charge pulses by electrical noise and interference.

Ideally, output pulses would have a very flat top to allow easy digitization of the pulse height and be very narrow to allow high pulse rates without pulse pile-up. However, the output pulse from the preamplifier typically has a relatively long decay time and is not very suitable for digitization and handling high pulse rates. As a result, the output of the preamplifier is fed into a shaping amplifier. Typically, shaping amplifiers use a combination of integration and differentiation stages to produce near Gaussian pulses. It should be noted that more recent commercial gamma cameras have used digital pulse processing methods to

perform this function. This involves digitizing the output waveform from the preamplifier. This has a number of advantages including providing the ability to change the trade-off between energy resolution and count rate, depending on the requirements of the particular imaging procedure. In addition, this method also provides digital estimates of the pulse heights that can be used in digital position and energy estimation algorithms.

11.2.2.5. Position and energy estimation

The goal of the radiation detector is to provide an estimate of the energy and interaction position of each γ ray incident on the detector. The output of the photodetector array and amplifier system is a set of voltage signals for each photon. The sum of these voltages is proportional to the gamma camera energy and the position is a function of the set of voltage values. The position and energy estimation circuits estimate the γ ray energy and position from the set of voltage values from the photodetector array.

One way of doing this is to use a resistive network to divide the signals from the array elements among a set of four signals often referred to as X_+, X_-, Y_+ and Y_-, as illustrated in Fig. 11.16. The resistor values for each PMT are chosen so that the charge is divided in proportion to its position with respect to the centre of the array. For example, for a PMT in the centre horizontally, the resistances for X_+ and X_- would be equal. Similarly, for a PMT in the centre vertically, the resistances for Y_+ and Y_- would be equal.

Using the scheme described above, the energy E can be computed using:

$$E = X_+ + X_- + Y_+ + Y_- \tag{11.6}$$

However, one limitation of this method is that the total amount of light collected is dependent on position. For example, if the interaction is directly under a PMT, a larger fraction of the total light will be collected, resulting in a larger value of E than if the interaction is in the gap between PMTs. This means that the estimate of the energy will vary spatially. As discussed below, this has an impact on camera uniformity. As a result, the energy must be corrected based on the interaction position.

Under the assumption that the light collected by a PMT is proportional to the distance from its centre, and with the correct resistor values, the interaction position, defined by x and y can be computed using:

$$x = \frac{X_+ - X_-}{E} \quad \text{and} \quad y = \frac{Y_+ - Y_-}{E} \tag{11.7}$$

In early gamma cameras, the computations above were performed using analogue circuits. However, in recent cameras, the pulse heights (or the entire pulse) are digitized and the computations performed digitally.

Using the resistive summing and simple estimation approaches above results in a number of problems. First, light collected by phototubes is not linearly related to the distance from the interaction point. For example, the amount of light changes relatively little for PMTs at a large lateral distance from the interaction position. As a result, thresholding circuits are often added to exclude the signal from PMTs with small outputs (and, thus, far from the interaction point) from the position calculation. In addition, the distribution of light between two tubes changes more quickly when the interaction position lies between two tubes than it does when the interaction position is directly over a tube. This results in spatial non-linearities where images of line sources are bent towards the centre of PMTs. A final difficulty is that it is not possible to reliably estimate the position of photons interacting near the edge of the camera. In this case, almost all of the light will be collected by the nearest PMT and there will be little change in the relative amount of light as the interaction position moves.

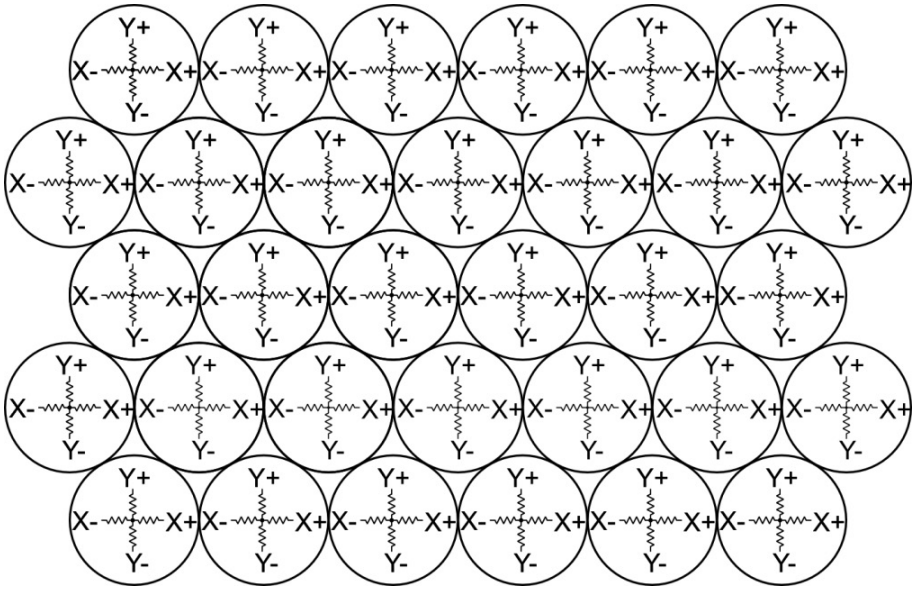

FIG. 11.16. Illustration of a resistive network used to implement position estimation. The output from each photomultiplier tube/preamplifier is divided by a resistive network with four outputs, X_+, X_-, Y_+ and Y_-. The corresponding signals from all of the photomultiplier tubes are connected to provide the summed signal.

As a result of the above difficulties, modern cameras use sophisticated correction and position estimation methods. The correction methods will be discussed in more detail below. Advanced position estimation methods involve digitizing the outputs of all of the PMTs and using them in the position estimation. In this case, position and energy estimation, and the various corrections are done by a digital signal processor or microprocessor, allowing a great deal more sophistication in the choice of algorithms. In some systems, these are done using statistical estimation techniques such as maximum-likelihood estimation.

As alluded to above, the values of the interaction position (x, y) and energy E are computed using equations similar to those shown above. The input variables X_+, X_-, Y_+ and Y_- are related to the charge signal from the PMT. These signals are proportional to the number of photoelectrons emitted from the photocathode. The emission of photoelectrons is the end result of a series of random processes that includes the number of scintillation photons produced, the number of these collected by the PMT, the number of photoelectrons produced for a given photon and the number of photoelectrons emitted which are focused onto the first dynode. The net result is that there is statistical variation in the values of the input variables for a given interaction position and energy. Thus, the values of position and energy are not exact, and are only estimates of the true quantities. Thus, there will be imprecision in the energy and position estimates resulting in finite intrinsic energy and spatial resolutions (see Chapter 8 for more details).

To a good approximation, both the energy and intrinsic spatial resolution can be characterized by a Gaussian function. The energy resolution results from variations in the total number of photoelectrons incident on the first dynode. The random variations can be approximated by a Poisson distribution and the variance in the energy resolution is, thus, approximately equal to the number of these photoelectrons. The approximate energy resolution of a gamma camera can, thus, be estimated as follows. A NaI(Tl) crystal produces, on average, 38 photons/keV. The quantum efficiency (fraction of incident scintillation photons that produce photoelectrons) of a PMT for the 415 nm emission of NaI(Tl) is approximately 12%. Thus, for a 140 keV photon, the number of photoelectrons collected is 140 × 38 × 0.12 = 638 electrons. The FWHM is equal to approximately 2.35 times the standard deviation, so the fractional energy resolution is equal to $2.35 \times \sqrt{638} / 638 = 9.3\%$. This is approximately equal to the energy resolution for a state of the art scintillation camera. It should also be noted from the above that the energy resolution is proportional to $E^{-0.5}$ and spatial variations in the collection efficiency will produce spatial variations in the energy resolution. Estimating the intrinsic spatial resolution is more difficult than estimating the energy resolution because of the more complicated estimation equation. However, typical intrinsic spatial resolutions are in the range of 3–5 mm, depending on the number of PMTs used and details of the estimation procedure.

11.2.2.6. Corrections

As mentioned above, the energy and position estimation are non-ideal, resulting in errors in energy and position estimates. These errors give rise to non-uniform sensitivity in the camera. Thus, to obtain clinically acceptable images, energy, spatial and uniformity corrections are needed. The need for these corrections is illustrated in Fig. 11.17. An image is shown resulting from a uniform distribution of γ rays on the camera with the collimator removed, often referred to as an intrinsic flood image. The substantial non-uniformity, the presence of edge packing artefacts near the edge of the FOV, and the visibility of the tube pattern should be noted.

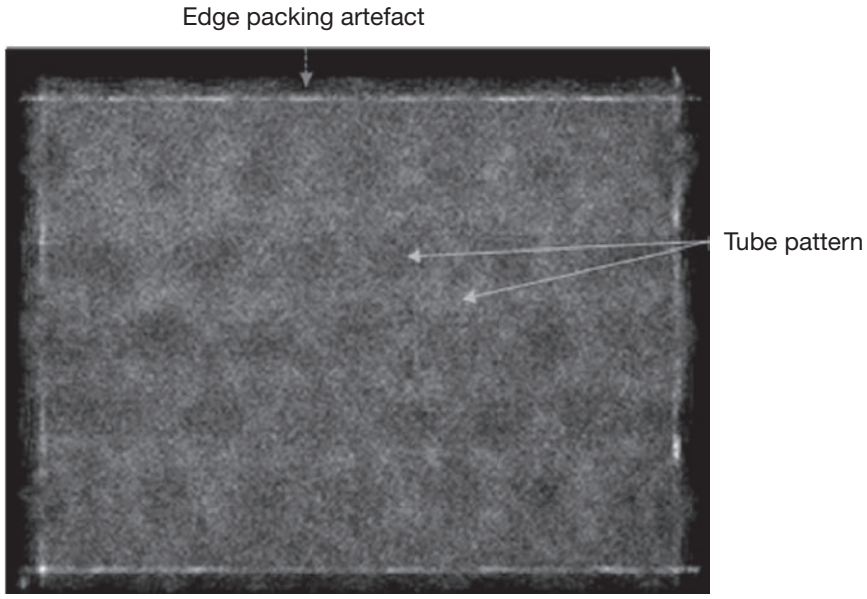

FIG. 11.17. Intrinsic flood image of gamma camera without energy, spatial or sensitivity corrections.

Energy corrections are needed because the estimated energy depends on spatial position. This behaviour can be understood in terms of variations of the fraction of the scintillation light collected as a function of interaction position. Since the energy is proportional to the light collected, differences in the fraction of light collected will result in a proportional change in the estimated energy. Figure 11.18 shows an example energy spectrum where the fraction of collected light was 2% lower or 2% higher than the average. This results in energy spectra

shifted to lower or higher energy, respectively. Only photons falling within the acquisition energy window, in this case having a full width of 20% of the photopeak energy, centred at 140 keV, are counted. There are about 1.7% fewer photons counted in the two sample energy spectra than for the case of the average energy spectrum. Thus, the sensitivity is about 2% lower at these points. A typical energy correction algorithm measures the energy spectrum as a function of position in the image using a source or sources with known energies. A linear or higher order correction is then made to the estimated energy.

Spatial corrections are needed because of biases in estimated interaction positions. These biases result, as described above, from discrepancies in the formulas used to estimate the position and the actual behaviour. As mentioned, these usually result in lines being bent towards space between PMTs. Typically, separate corrections are made for the axial (y) and transaxial (x) directions. These corrections involve imaging a mask with a grid of holes or lines in combination with a flood source. This results in a series of bright spots or lines in the image that correspond to the known positions of the holes or lines in the mask. A function, typically a polynomial, is fit to the set of true points as a function of the set of measured points. This function can then be used to correct a measured position.

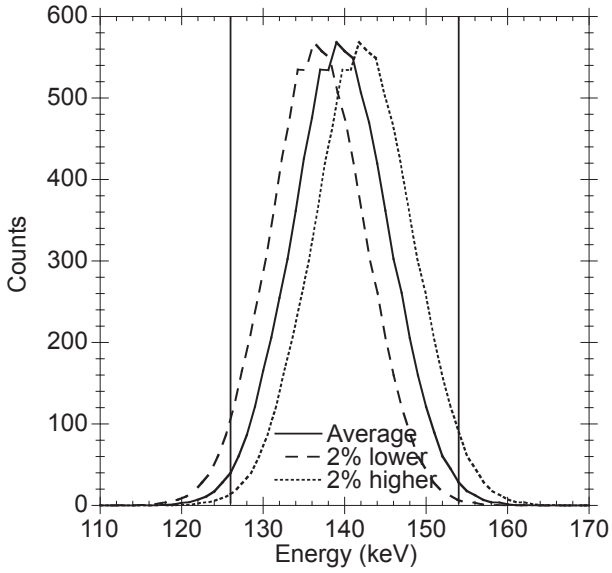

FIG. 11.18. Sample energy spectrum for 140 keV photons for the cases of average, 2% lower than average and 2% higher than average light collection efficiency. The variation in light collection efficiency results in a shift of the energy spectrum, which results in non-uniform sensitivity.

The final type of correction applied is a uniformity or sensitivity correction. The goal of this correction is to make images of a flood source as uniform as possible (see Fig. 11.19). There are two types of uniformity corrections: intrinsic, which corrects only for non-uniform sensitivity of the detector system (i.e. excluding the collimator), and extrinsic, which corrects for both detector and collimator non-uniformities. Uniformity corrections are made using a high-count flood image. Uniformity correction is implemented by, in essence, multiplying each pixel in acquired images by a factor equal to the average counts in the active portion of the flood image divided by the counts in the corresponding pixel in the flood image.

The number of counts in the flood image is critical in determining the ultimate uniformity of the image. This is especially important in SPECT where local non-uniformities can result in ring artefacts. To achieve this, the counts in the flood image should be such that the relative standard deviation (coefficient of variation) of the pixel counts resulting from Poisson counting statistics is less than the desired uniformity. For example, if uniformity correction to better than 1% is desired, the average number of counts per pixel in the uniformity flood should be greater than $1/0.01^2 = 10$ kcounts per pixel. The total number of counts in the flood, thus, depends on the number of active pixels in the image. Since non-uniformities are generally relatively low frequency, this restriction can be relaxed to some degree by the use of low pass filtering applied to the flood image.

Intrinsic flood images are usually acquired using a point (or syringe) source containing a small quantity of the isotope of interest. If the source is placed at a distance of more than five times the largest linear dimension of the camera

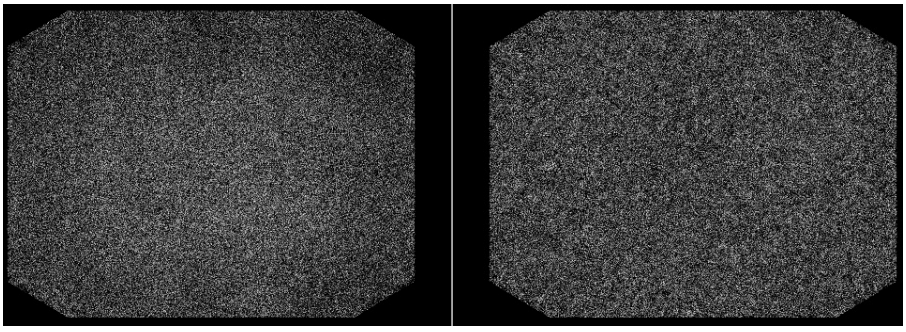

FIG. 11.19. Intrinsic flood images for a gamma camera having a poor (left) and good (right) set of corrections applied. It should be noted that the images are windowed so that the brightness represents a relatively small range of count values in order to amplify the differences. The peppery texture is due to quantum noise and is to be expected. The quantum noise is exaggerated because of the windowing used, and can be reduced by acquiring very high-count flood images.

FOV from the camera face, then the irradiation of the camera can be considered uniform. Since the uniformity of the camera will, in general, vary depending on the energy of the isotope and energy window used, this correction should ideally be made for each isotope and energy window used. The count rate for the acquisition should be within acceptable limits to avoid high count rate effects.

Extrinsic flood images are made using a flood or sheet source. Fillable flood sources have the advantage that they can be used for any isotope. However, great care must be made in filling the phantom to remove bubbles, mix the activity and maintain a constant source thickness. In addition, images must be obtained for each collimator used with a given isotope. As a result, ^{57}Co sheet sources are often used to obtain extrinsic flood images. These have the advantage of convenience but are, strictly speaking, valid for only a single isotope.

One way to take advantage of both approaches is to perform uniformity corrections using a combination of intrinsic flood images for each isotope used and an extrinsic flood image obtained for the collimator in question. This approach assumes that collimator uniformity is independent of energy and can, thus, be measured with, for example, a ^{57}Co sheet source. Not all equipment vendors support this approach. Some vendors assume that the energy and linearity corrections produce uniformity that is energy independent, and they, thus, recommend only the use of an extrinsic flood image for uniformity correction. Another approach is to first confirm the uniformity of all collimators via a sheet source flood image with intrinsic correction for ^{57}Co. Then, an extrinsic flood image for each isotope used is acquired and used in uniformity correction for that isotope, assuming that the collimator is sufficiently uniform. The best approach depends on the characteristics of the individual camera.

11.2.2.7. Image framing

The final step in generating gamma camera images is image framing. Image framing refers to building spatial histograms of the counts as a function of position and possibly other variables. This involves several steps, and is typically done either by microprocessors in the camera or in an acquisition computer. In this step, position is mapped to the elements in a 2-D matrix of pixels. The relationship between pixel index and physical position is linearly related to the ratio of the maximum dimension of the FOV of the camera to the number of pixels, the zoom factor and an image offset. The zoom factor allows enlarging the image so that an object of interest fills the image. This can be useful, for example, when imaging small objects. It results in a pixel size in the image that is a factor of 1/zoom factor as large as in the unzoomed (zoom factor equals one) image. It should be noted that even though the pixel size is decreased, the resolution of the image will not necessarily be improved as long as the original

pixel size is smaller than the intrinsic resolution. For example, if the native pixel size is 3.2 and the intrinsic resolution is 4 mm, a zoom factor of two will result in a pixel size of 1.6 mm, but the intrinsic resolution will still be 4 mm. An image offset can be used to shift the image, so that an object of interest is in the centre of the acquired image.

In addition to adding counts to the appropriate pixel spatially, the framing algorithm performs a number of other important functions. The first is to reject photons that lie outside of the energy window of interest. This is done to reject scattered photons. Gamma cameras typically offer the ability to simultaneously frame images corresponding to more than one energy window. This can be useful for isotopes having multiple photopeaks, for acquiring data in energy windows used by scatter compensation algorithms or for acquiring simultaneous images of two or more radionuclides. Framing software typically enables the summation of photons from multiple, discontiguous energy windows into one image as well as simultaneously framing multiple images from different energy windows into different images. There is often a limited number of energy windows that can be framed into a single image, and a limit on the number of images that can be framed at one time. These limits may depend on the image size, especially if the framing is done by a microprocessor in the camera that has limited memory.

A second important function provided by the framing system is the ability to obtain a sequence of dynamic images. This means that photons are recorded into a set of images depending on the time after the start of acquisition. For example, images could be acquired in a set of images with a frame duration of 10 s. Thus, for the first 10 s, photons are recorded into the first image; for the second 10 s, they are recorded into a second image; and so on. Thus, just as multiple images are obtained in the case of a multi-energy window acquisition, multiple images are obtained corresponding to a sequential set of time intervals. This is illustrated in Fig. 11.20, where seven dynamic frames are acquired for a time interval T. Dynamic framing is used for monitoring processes such as kidney function, gastric emptying, etc. The time frames are often not equal in duration as there may be more rapid uptake at early times and a later washout phase in which the change in activity with time is slower. The number or acquisition rate of dynamic frames is often limited due to constraints in framing memory and this limit can depend on the image size.

Gated acquisition is similar to dynamic acquisition in that photons are recorded into a set of images depending on the time they are detected. However, in gated acquisition, the time is relative to a physiological trigger, such as an electrocardiogram (ECG) signal that provides a signal at the beginning of each cardiac cycle. This is appropriate for processes that are periodic. The photons are counted into a set of frames, each of which corresponds to a subinterval of the time between the two triggers. For example, the bottom two illustrations

in Fig. 11.20 should be considered. In both cases, the interval between gates is divided into four subintervals (in cardiac imaging, 8 or 16 frames are typically used, but 4 are illustrated in this example for simplicity). The photons arriving in each of the subintervals are counted in the corresponding frame. Thus, for cardiac imaging, the activity distribution during the first quarter of the cardiac cycle is imaged in frame 1, the second quarter in frame 2, etc. This is useful for assessing wall motion and thickening. Just as in dynamic and multiple window acquisitions, there may be limits on the size and number of frames due to limits in framing memory on the acquisition computer.

Since the gate is derived from a physiological signal, and physiological signals are not necessarily exactly periodic, the framing algorithm must handle the case of variable time intervals between gate signals. In cardiac imaging, this corresponds to variations in the heart rate. Figure 11.20 also illustrates two ways of dealing with this. In the first method, the frames correspond to fixed time intervals. What is typically done is to measure the average heart rate and to divide this by the number of frames per cycle. The photons arriving during each of these time intervals are then binned into the image corresponding to each time interval. Difficulty arises when there are variations in the length of the cardiac cycle. Figure 11.20 illustrates the case when there are beats that are 5% longer and 5% shorter than average. For fixed time interval framing, the actual interval for the fourth frame will be lengthened or shortened when the length of the cardiac cycle changes. This will result in motion blurring of the gated images. The alternative is to change the length of the subintervals for each beat based on the time between gates for that beat. This eliminates the problem with motion blurring described above. However, this also requires buffering of the events for the period between gates before they can be framed. One additional advantage of this approach is that bad beats (ones longer or shorter than the average beat by more than a certain fraction of the average beat length) can be discarded.

The final acquisition mode is list-mode acquisition. In list-mode acquisition, the energies and positions of incoming photons are simply saved to a file in the order in which they appear. Additional information is recorded in the form of events in the list-mode stream. These events include things such as physiological triggers, gantry or table motion, start and stop of acquisition, and timing marks which are injected at regular intervals. The advantage of list-mode data is that they contain all of the information obtained during the acquisition. As a result, it can be retrospectively reframed using different pixel sizes, energy windows, frame intervals, etc. However, the downside is that list-mode files are very large (typically eight or more bytes of data are stored for each photon). List-mode is often not made available for routine clinical use, but can be very useful for research use.

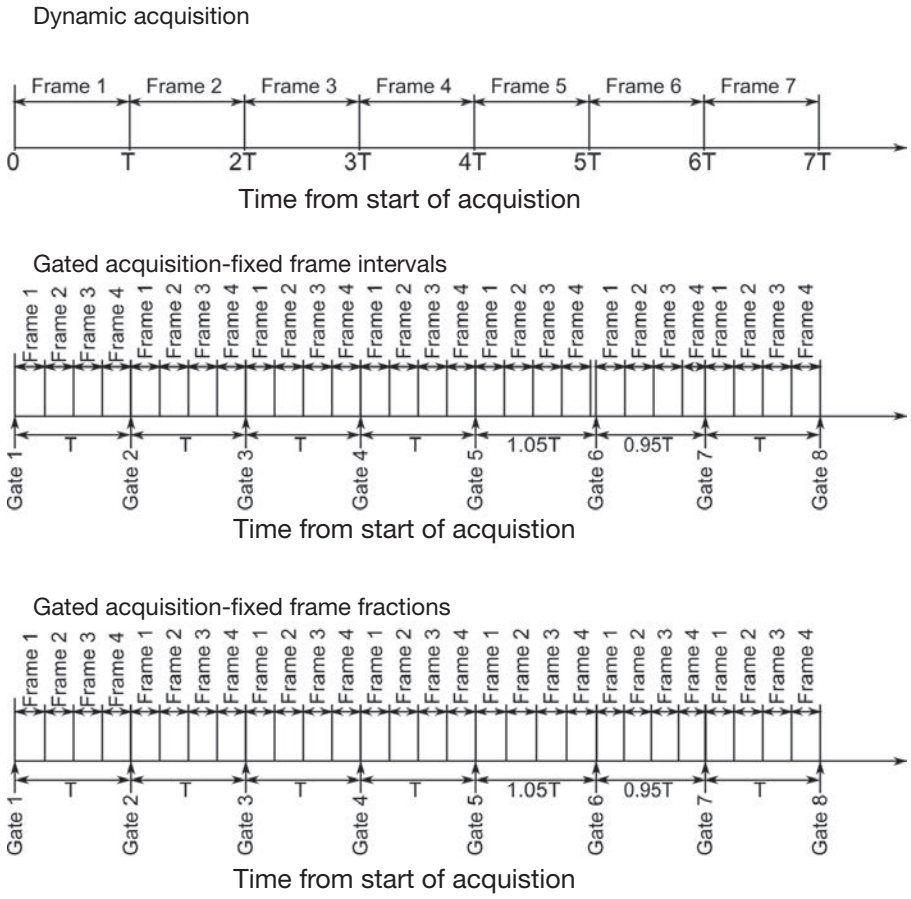

FIG. 11.20. Comparison of dynamic and gated acquisition modes. In all cases, time is along the horizontal axis. In dynamic acquisition, photons are framed into different images based on the time after acquisition. In this case, the interval between dynamic frames is T. For gated acquisition, photons are framed into a set of images based on the time in relation to the previous gate signal. The gate signal is derived from a physiological trigger such as the R wave in an electrocardiogram signal. For the two dynamic gating examples, the time interval between gate signals 1–5 and 7–8 are the same, but the interval between gates 5–6 and 6–7 are 5% longer and 5% shorter than the average interval. Two methods of dividing the time interval between gates, fixed intervals and fixed fractions are illustrated. In the fixed interval method, the photons are framed into frames with fixed widths. In this case, the extra time (5% of T) between gates 5–6 is not counted, and the interval into which photons are counted in frame 4 in the time interval between gates 6–7 is shortened by 20% (5% of T).

11.2.2.8. Camera housing

The camera housing contains the radiation detector and provides a mount for the collimators. The housing performs a number of important functions. First, it must provide radiation shielding so that photons can only enter the camera and be detected via the collimator. This shielding is made of lead and, since the crystal and PMT array is large, the housing must also be large and is rather heavy. In addition, the PMTs are very sensitive to magnetic fields. The housing, thus, includes mu-metal shields around the PMTs to screen magnetic fields. Without these, there can be variations in the sensitivity and uniformity due to changes in the camera position relative to ambient magnetic fields, including the magnetic field of the Earth. The PMTs and detection electronics are sensitive to variations in temperature and generate a non-negligible amount of heat. As a result, the housing must include a temperature control system, typically in the form of fans to circulate air and provide ventilation.

A final important function of the camera housing is to provide mounting for the collimators. High energy parallel and pinhole cameras can weigh more than 100 kg, so the mounting system must be sufficiently strong to securely support this much weight. The back face of the collimator (excluding pinhole collimators) must be in close proximity to the crystal in order to provide the highest possible resolution. Since collimators are often changed several times per day, the mounting system must provide for easy collimator removal and change. Most modern cameras include automatic or semi-automatic collimator exchange systems. Thus, the collimator retaining system often includes an automated locking system. Finally, modern cameras include touch sensors on the face of the collimators and often include proximity sensors. The touch sensors activate when the collimator face touches the patient or bed and disable camera motion. This is done for patient safety and in order to avoid injuring the patient or damaging the camera. A proximity sensor is sometimes also included. This typically consists of an array of light emitting diodes with an opposing array of photodiodes. They are mounted so that the light beam from the light emitting diodes is parallel to the face of the camera. Proximity to the patient can be detected when one of the light beams is interrupted by part of the patient. The proximity sensor can be used for automated sensing of the camera orbit that positions the cameras close to the patient at each camera position. Electrical connections between the housing and the collimator provide communication with the touch and proximity sensors as well as providing information about which collimator type is mounted.

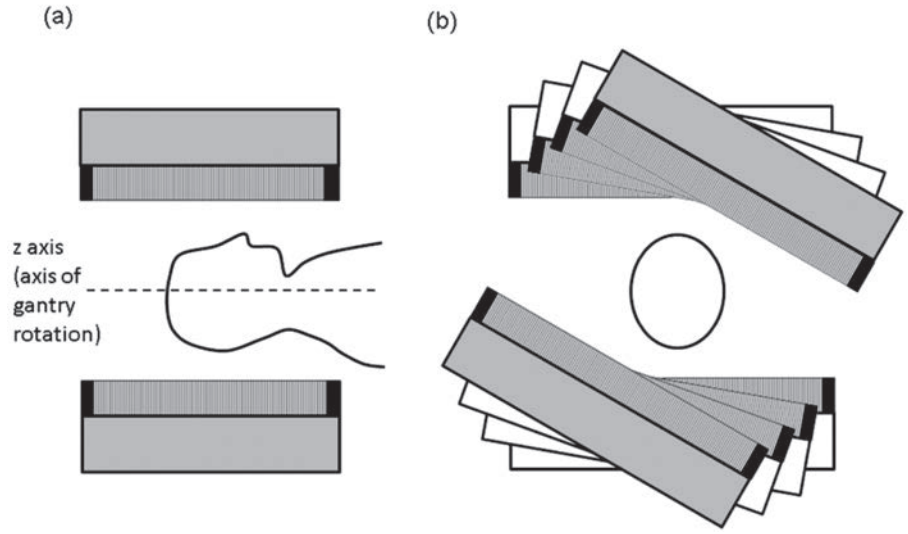

FIG. 11.21. (a) A cross-section of a dual head gamma camera capable of acquiring two views simultaneously. It should be noted in this example that the heads are oriented 180° apart, although a 90° degree configuration is also possible. SPECT data acquisition requires rotation of the gamma camera heads about the long axis of the patient as indicated. (b) A transverse slice with the position of four different camera orientations superimposed to illustrate the acquisition of multiple angular views.

11.2.3. SPECT

11.2.3.1. Gamma camera SPECT systems

In addition to the software requirements for image reconstruction, SPECT is associated with hardware requirements that are beyond those needed for planar imaging. Although various SPECT configurations have been developed, the most common implementation involves the use of a conventional gamma camera in conjunction with a gantry that allows rotation of the entire detector head about the patient [11.2]. The gantry rotation is about the long axis of the patient (see Fig. 11.21) and is typically performed in discrete steps (step and shoot), although continuous motion may also be supported. During rotation of the gantry, the patient bed typically does not move, so SPECT data acquisition is more similar to conventional CT than to spiral CT, in this respect. In this way, planar views of in vivo radioactivity distribution can be acquired at different angular orientations and these data can be used to form the projections that are required for image reconstruction by computed tomography.

In principle, rotation of the gamma camera about 180° allows for the acquisition of sufficient projections for tomographic reconstruction. However, in practice, opposing views acquired 180° apart differ due to various factors (photon attenuation, depth dependent collimator response) and SPECT data are commonly acquired over 360°. The theory of computed tomography determines the number of angular samples that are required, but for many SPECT studies, around 128 views may be acceptable. The time needed to acquire these multiple projections with adequate statistical quality is a practical problem for clinical SPECT where patient motion places a limit on the time available for data acquisition. In an effort to address this issue, a common approach in modern SPECT designs is to increase the number of detector heads, so that multiple views can be acquired simultaneously. Dual detector head systems currently predominate, although triple detector head gamma cameras also exist. Increasing the number of detector heads increases the effective sensitivity of the system for SPECT, at the expense of increasing cost. Dual head gamma cameras are often considered the preferred configuration for systems intended not just for SPECT but also for general purpose applications, including whole body studies where simultaneous acquisition of anterior and posterior planar images is required.

In addition to the rotational motion required for SPECT, flexibility is also required in the relative positioning of the detector heads. For general purpose SPECT with a dual head system, the two heads are typically oriented in an opposing fashion (sometimes referred to as H-mode) and 360° sampling is achieved by rotation of the gantry through 180°. In contrast, cardiac SPECT is often performed with the detectors oriented at 90° to each other (sometimes referred to as L-mode). In this mode, the gantry rotates through 90° and the two detectors acquire projections about 180° from the right anterior oblique position to the left posterior oblique position. Despite acquiring only 180° of data, this mode has advantages for cardiac applications as it can minimize the distance between the heart and the detectors, thus reducing attenuation and depth dependent losses in spatial resolution. Other approaches to minimizing the distance between the detectors and the patient during SPECT data acquisition involve further control of the rotational motion of the detector heads. For detectors rotating about a circular orbit, this involves adjusting the radius of rotation for individual studies, so as to minimize the source to collimator distance. Other options include detectors that rotate, not in a circular orbit, but in an elliptical orbit, or, alternatively, a variable rotational motion that contours to the outline of the body.

The flexibility of the motions that are available in modern SPECT systems makes it particularly important to ensure that the detectors are correctly aligned. This means that the specified angle of rotation is accurately achieved at all angles. The detector heads also need to be perfectly oriented parallel to the z axis of the system, such that each angular view is imaging the same volume. Furthermore,

it is important that the centre of each angular projection is consistent with the centre of mechanical rotation. Errors due to these factors can potentially lead to a loss of spatial resolution and the introduction of image distortion or ring artefacts. In order to identify and correct these issues, an experimental centre of rotation procedure is employed. A small point source is placed in the FOV at an off-centre location. SPECT data acquisition is performed and deviations from the expected sinusoidal pattern are measured in the resulting sinograms.

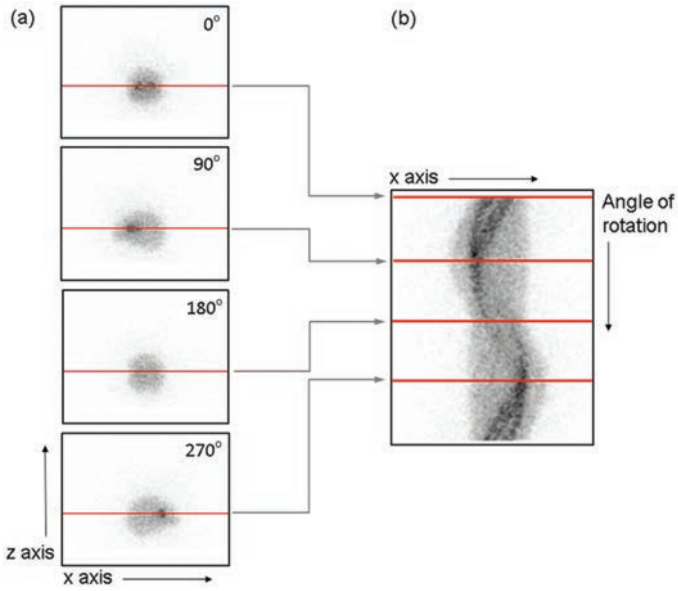

FIG. 11.22. (a) A series of planar views acquired at different angular orientations. A sample of four views has been extracted from a total of 128 views acquired about 360°. It should be noted that the z axis represents the axial position and is the axis of gantry rotation. (b) A sinogram corresponding to a particular axial location. The red lines in (b) indicate 1-D projections that have been extracted from the corresponding planar views shown in (a).

Although SPECT can be performed with a variety of collimator geometries, such as cone-beam or pinhole, much of the present discussion has assumed parallel-hole collimation. In this case, each planar view consists of multiple 1-D projections, each measured at different axial positions (Fig. 11.22). Each projection is defined by the holes of the collimator and approximates a series of parallel line integrals of the activity distribution in the FOV. In practice, these line integrals are substantially corrupted by the effects of photon attenuation, scatter and depth dependent collimator response. Each of these factors requires software correction and these corrections are described in the following paragraphs.

11.2.3.2. Attenuation correction

Standard tomographic reconstruction algorithms, such as those based on filtered back projection, assume that measured projections are line integrals through the object. However, in SPECT, the interaction of photons via photoelectric absorption and Compton scatter within the patient results in attenuated projections. The attenuated projections $P_\theta(t)$ can be described for the 2-D case by the equation:

$$P_\theta(t) = \int_0^\infty a(l\boldsymbol{n}_\theta + t\boldsymbol{m}_\theta) \mathrm{e}^{-\int_0^l \mu(l'\boldsymbol{n}_\theta + t\boldsymbol{m}_\theta)\,\mathrm{d}l'} \,\mathrm{d}l \tag{11.8}$$

where the geometry is illustrated in Fig 11.23. In this equation, the unit vectors $\boldsymbol{n}_\theta$ and $\boldsymbol{m}_\theta$ are as described in the legend of Fig. 11.23, t is the transaxial distance in the projection from the projected position of the origin and l is the distance along the projection line from the face of the detector. $a(\boldsymbol{x})$ is the activity distribution and gives the activity at point $\boldsymbol{x}$. It should be noted that the integral in the exponent represents the integral through the attenuation distribution $\mu(\boldsymbol{x})$ from the point $\boldsymbol{x} = l\boldsymbol{n}_\theta + t\boldsymbol{m}_\theta$. Thus, the exponential represents the attenuation of photons emitted at $\boldsymbol{x}$ as they travel back towards the detector.

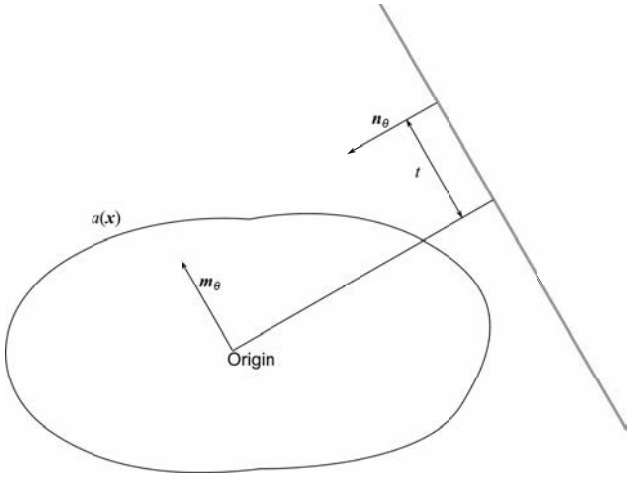

FIG. 11.23. Projection geometry used to describe the attenuated projection in Eq. (11.8). In this figure, the projection is at an angle θ. A parallel-hole collimator is assumed, and the unit vector $\boldsymbol{n}_\theta$ is perpendicular to the collimator and parallel to the projection rays. The unit vector $\boldsymbol{m}_\theta$ is parallel to the collimator face and perpendicular to $\boldsymbol{n}_\theta$. The variable t is the distance along the detector from the projected position of the origin.

As can be seen from the above equation, unlike PET, the attenuation is not constant for a projection ray, but instead varies along the ray. Using reconstruction methods that do not model this effect produces both artefacts and a loss of quantitative accuracy in the resulting images. The artefacts can include streak artefacts, resulting from highly attenuating objects such as bones, catheters or medical devices; shadows, due to higher attenuation between an object in some views than in others (e.g. breast or diaphragm artefacts in cardiac SPECT); and a generally reduced image intensity in the centre of the image.

The first requirement to compensate for attenuation is knowledge of the attenuation distribution in the patient. This is done by either assuming uniform attenuation inside the object and extracting information about the body outline from the emission data or using a direct transmission measurement. Assuming a uniform attenuation distribution in the patient is only valid in regions such as the head. Even in the head, bone and regions containing air, such as the sinuses, result in imperfect estimates of the attenuation distribution and lead to imperfect attenuation compensation. Myocardial perfusion imaging is an important application for SPECT and, since attenuation can produce artefacts that obscure actual perfusion defects, a number of commercial devices have been developed to allow measurement of the attenuation distribution in the body. All of these devices use transmission CT techniques to reconstruct the attenuation distribution inside the body. The devices that have been developed can be divided into two general classes: devices using radionuclide sources and devices based on X ray tube sources. In both cases, a source of X rays or γ radiation is aimed at the body and a detector on the opposite side of the body measures the transmitted intensity. The intensity $I_\theta(t)$ passing through the body for a source with incident intensity I_0, projection position t and projection view θ is given by:

$$I_\theta(t) = I_0(t)\mathrm{e}^{-\int_0^\infty \mu(l\mathbf{n}_\theta + t\mathbf{m}_\theta)\,\mathrm{d}l} \tag{11.9}$$

where the symbols and geometry are as in Fig. 11.23.

Acquiring sets of these transmission data for various angles allows reconstruction of the attenuation distribution. Tomographic reconstruction methods can be applied directly by noting that the negative of the log of the fraction of transmitted photons is a line integral through the attenuation distribution.

A number of transmission devices based on radionuclide sources have been developed and marketed (Fig. 11.24). All of these devices use the gamma camera to detect the transmission photons. The simplest of these designs is a sheet source of radioactivity. To avoid contaminating the projection data, either the

transmission scan is acquired separately from the emission data or simultaneously using a radionuclide with a lower photopeak energy. Typically, ^{153}Gd is used as it has an energy lower than that of ^{99m}Tc and the transmission photons, thus, do not interfere with collection of emission data. To reduce patient dose and scatter in the transmission data, the source should be collimated. The disadvantages of sheet source designs are that they are expensive and high activities make them dangerous to handle. Using lower activity sources results in contamination of the transmission data.

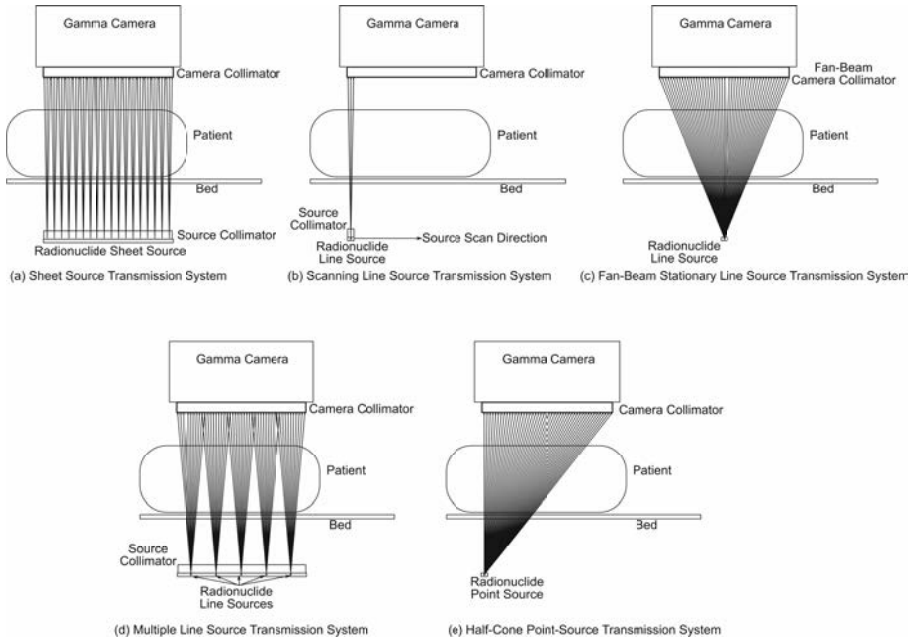

FIG. 11.24. Illustration of a number of proposed transmission scanning devices. It should be noted that in all cases the drawing represents a sagittal view of the system. However, for (d), the multiple line source system, the line sources are normally parallel to the long axis of the patient and, thus, perpendicular to the direction shown. Similarly, for (c), the fan-beam system, the fan beam is in the transaxial direction, the opposite of that pictured above. Furthermore, for the sheet source system, the source is continuous and the fan of rays is only shown to illustrate the limited range of directions of transmission source photons that pass through the source collimator.

As a result of these difficulties, line source transmission systems have been developed. The line sources are either scanned opposite the patient and parallel to the collimator or positioned at the focal spot of a fan-beam collimator. In the latter case, the projection geometry is different from the parallel-beam geometry

described above, but the same basic principles apply. The fan-beam geometry has the advantage of having much higher sensitivity and, thus, of requiring lower activity sources. However, the fan-beam geometry results in magnification and, thus, the FOV is smaller than the size of the detector. It should be noted that in the fan-beam geometry the fan lies in the transaxial plane and, thus, the truncation is in the transaxial direction. For the scanning line source systems, an electronic window is used so that transmission data are acquired only in the region directly under the line source. To overcome the low sensitivity of parallel-beam geometry, high activities (1.85–18.5 GBq) are used. This reduces the contamination of the transmission data by emission activity. The scanning line source adds additional mechanical complexity to the system and, as a result, several other designs have been proposed.

In multiple line source systems, a small number of line sources are used. The line sources are placed close enough together so that the object is covered due to the finite acceptance angle of the camera collimator, in effect acting like a non-uniform sheet source. One advantage of this geometry is that weaker sources can be used for the sources near the edge of the camera, since these correspond to thinner regions of the body. This can help reduce the effects of high count rates. The final geometry proposed is the half-cone geometry shown above. In this design, a high energy ^{133}Ba point source is used. Many of the high energy photons penetrate through the collimator, creating a half-cone-beam geometry. This allows a parallel-hole collimator to be used both for transmission and emission imaging. The use of a point source simplifies shielding of the source.

In general, radionuclide transmission sources have a number of disadvantages. These include the fact that the source decays and must be replaced. There are also limits on transmission count rates imposed by the gamma camera, resulting in relatively noisy transmission images. In addition, if the count rate due to the emission activity within the patient is high, the transmission images can be degraded, resulting in inaccurate attenuation maps. Finally, with the sheet source, scanning line source and multiple line source geometries, the resolution of the transmission scan is limited by the combination of source and camera collimators. In general, these provide lower resolution transmission scans. However, one advantage of radionuclide based transmission systems is the potential to perform simultaneous imaging, thus eliminating the need for an additional transmission scan. Another advantage, especially when acquired simultaneously, is that registration of the emission and transmission images is guaranteed. Finally, the use of radionuclide sources with a small number of high energy photopeaks makes converting the transmission images into an attenuation map at the energy of the emission source easier than for X ray CT based systems, which use X ray tubes having continuous X ray energy spectra.

The second major kind of transmission scanning apparatus uses X ray sources. There are two major kinds: slow rotation and hybrid SPECT/CT systems. In the slow rotation devices, a low power X ray tube is attached to the gantry opposite a conventional X ray detector. The gantry is rotated around the patient, resulting in acquisition of X ray transmission data. In one commercially available slow-rotation device, the X ray detector has 1–4 rows of detectors with an effective axial length of 0.5–1.0 cm at the centre of rotation and can, thus, acquire transmission data for 1–4 slices per revolution. The patient is then translated on the bed to acquire data from additional slices. The resolution of these scanners can be as good as 1 mm transaxially but is 0.5–1 cm axially. Typical revolution times for this type of scanner are 20–30 s/rotation. Another commercial implementation uses a flat panel detector and cone-beam geometry. This provides very high resolution, though there is the potential for cone-beam artefacts in the reconstructed images. In hybrid SPECT/CT systems, a SPECT gantry and an independent diagnostic CT system are integrated. The CT images are acquired very rapidly using conventional CT technology either just before or just after acquisition of the SPECT data. The CT images acquired using these systems, as well as the acquisition protocols and options, are similar to those on diagnostic CT scanners.

X ray tube based methods for obtaining transmission images have a number of advantages and disadvantages. The major advantages are acquisition speed, the quality of the attenuation maps (high resolution and low noise) and the convenience of not needing to replace radionuclide sources. Among the disadvantages are that the image is not acquired simultaneously and is often acquired with the bed in a different position than used for the SPECT scan. There is, thus, the potential for mis-registration of the SPECT images and attenuation maps, resulting in degraded attenuation compensation. A second disadvantage is that the effects of motion, especially respiratory motion, during the attenuation scan are different to those during the emission scan. For example, in slow-rotation devices, streak artefacts in the diaphragm region caused by respiratory motion during transmission data acquisition are common. For hybrid devices, the CT scans are acquired very rapidly, so may freeze the patient at one particular phase of the respiratory cycle. Despite these potential disadvantages, the advantages of X ray tube based transmission scanning (including hybrid SPECT/CT systems) have meant that they have largely replaced devices based on radionuclide sources.

Using the above systems for attenuation compensation requires transforming the transmission CT image into an emission attenuation map. For radionuclide sources, this basically means translating the attenuation coefficients at the energy of the transmission source to that of the emission source. For emission and transmission source energies above 100 keV, this can be done by scaling the transmission CT image by the ratio of the attenuation coefficient in

water at the emission energy divided by that at the transmission source energy. However, more accurate results can be obtained using the bilinear scaling method described below for use with X ray CT images.

For hybrid SPECT/CT systems where a conventional X ray CT image is transformed into an attenuation map, there are a number of additional considerations. First, transmission data are obtained using a polychromatic source. There can, thus, be substantial beam hardening. Fortunately, beam hardening and other corrections routinely applied in X ray CT scanners to produce images in Hounsfield units (HU) eliminate many of these concerns. In this case, the CT image in Hounsfield units can be transformed to the attenuation map via piecewise linear scaling, where, in effect, pixels with values less than 0 HU are treated as water with densities ranging from 0 to 1, pixel values between 0 and 1000 HU are treated as a mixture of bone and water, and pixel values greater than 1000 HU are treated as dense bone. Thus, for a pixel having a value h in Hounsfield units, the attenuation map value μ is given by:

$$\mu(h) = \begin{cases} \dfrac{1000+h}{1000}\mu_{\text{water}} & \text{for } h \leq 0 \\ \mu_{\text{water}} + \dfrac{h}{h_{\text{bone}}}(\mu_{\text{bone}} - \mu_{\text{water}}) & \text{for } 0 < h < h_{\text{bone}} \\ \dfrac{h}{h_{\text{bone}}}\mu_{\text{bone}} & \text{for } h > h_{\text{bone}} \end{cases} \tag{11.10}$$

where μ_{water} and μ_{bone} are the attenuation coefficients of water and bone, respectively, for the energy of the photopeak of the imaging radionuclide.

Once the attenuation map is obtained, attenuation correction can be implemented using analytical, approximate or statistical image reconstruction algorithms. Generally, analytical methods are not used due to their poor noise properties. Approximate methods include the Chang algorithm. This method is often used in regions of the body where the attenuation coefficient is assumed to be uniform and the actual attenuation distribution is not measured, but instead is approximated from the boundary of the object, which is usually assumed to be an ellipse. In the Chang method, an image reconstructed using filtered back projection (i.e. without attenuation compensation) is approximately compensated for attenuation. This approximate compensation is obtained for each voxel by dividing the uncorrected image signal by the average of the attenuation factors that correspond to each projection view. For a uniform attenuator with an assumed elliptical boundary, these attenuation factors can be calculated analytically. However, the Chang method is approximate and has poor noise properties.

Thus, for the best attenuation compensation, statistical iterative reconstruction methods should be used. Statistical iterative reconstruction methods are discussed in more detail in Chapter 13. These methods can be used to compensate for attenuation by incorporating a model of the attenuation process in the imaging matrix. Accurate attenuation compensation can be obtained with fast statistical iterative reconstruction methods, such as the ordered-subsets expectation-maximization (OSEM) algorithm, if an accurate attenuation map is available.

11.2.3.3. Scatter correction

In gamma camera imaging, a significant fraction of the detected photons are scattered in the body. This is due to the finite energy resolution of the gamma camera, which results in imperfect energy based scatter rejection. The scatter to primary ratio (SPR) depends on the radionuclide, energy window, energy resolution, source depth and the size of the object. For example, for ^{99m}Tc, the SPR is in the range of 0.2 for brain imaging and 0.6 for cardiac imaging. On the other hand, for ^{201}Tl cardiac imaging, the SPR can be greater than 1. Scatter results in loss of contrast, especially for cold objects in a warm background, and loss of quantitative accuracy.

Scatter correction requires estimating the scatter component of the projection data combined with a compensation method. Most frequently, the scatter component is estimated using data acquired in auxiliary energy windows. Perhaps the most common and flexible such method is the triple energy window (TEW) method. This method uses two scatter energy windows, one above and one below the photopeak window, as illustrated in Fig. 11.25. The scatter is estimated from the counts in the scatter windows using a trapezoidal approximation where counts in the scatter windows divided by their window widths are treated as the sides, and the scatter in the photopeak window is the area of the trapezoid. The estimated scatter counts in the photopeak window estimated using TEW s_{TEW} are given by:

$$s_{\text{TEW}} = \left[\frac{c_{\text{lower}}}{w_{\text{lower}}} + \frac{c_{\text{upper}}}{w_{\text{upper}}}\right]\frac{w_{\text{peak}}}{2} \tag{11.11}$$

where

c_{lower} and c_{upper} are the counts in the lower and upper scatter windows, respectively;

and w_{peak}, w_{lower} and w_{upper} are the widths of the photopeak, lower scatter and upper scatter windows, respectively.

It should be noted that in Fig. 11.25 the scatter windows are adjacent to the photopeak energy window, but this is not necessary, nor, in all cases, desirable. In fact, it is desirable to position the windows as close as possible to the photopeak window, while having only a small fraction of the photopeak photons detected in the scatter windows. The width of the scatter windows is a compromise between obtaining as accurate a scatter estimate as possible, which favours a narrow energy window, but having an estimate that is low in noise, which favours a wide energy window.

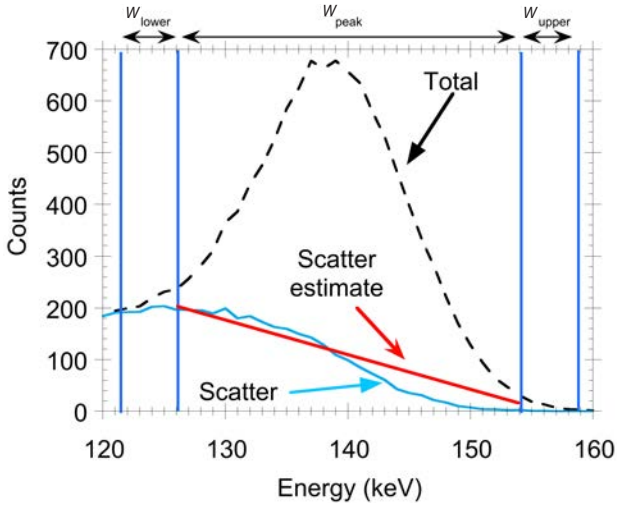

FIG. 11.25. Illustration of the use of a trapezoidal approximation to estimate the scatter in the photopeak energy window in the triple energy window method scatter compensation for ^{99m}Tc. It should be noted that in this example the windows are not necessarily optimally placed. In particular, the scatter windows are positioned such that there is a non-zero contribution from unscattered photons in the scatter energy windows. This is especially evident for the upper energy window. For the case of ^{99m}Tc, the counts in the upper window are often assumed to be zero.

Another method to estimate the scatter component in the projection data is via the use of scatter modelling techniques. These techniques use an estimate of the activity distribution and a mathematical algorithm to compute the scatter that would have been detected from the activity distribution. The mathematical techniques used range from accurate approximations to full Monte Carlo simulations.

As mentioned above, scatter correction is accomplished by combining scatter estimation and compensation methods. Methods of compensating for scatter include subtracting the scatter estimate from the projection data and, for

SPECT, including the projection data in the iterative reconstruction process. Owing to noise and errors in the scatter estimate, subtraction of a scatter estimate from measured projection data can lead to some negative pixel values. This can be a problem when the data are subsequently used with statistical iterative reconstruction algorithms. For scatter estimates obtained from energy windows, this effect can be reduced by low pass filtering the scatter estimate. Nevertheless, truncation of negatives is often used, though this can increase bias in the scatter compensation.

For SPECT, a better way to accomplish scatter compensation is to add the scatter estimate to the computed projection during the iterative reconstruction process. For example, algorithms such as OSEM and maximum-likelihood expectation-maximization (MLEM) involve back projecting the ratio of the measured and computed projections. In this case, the scatter estimate can simply be added to the computed projections. Another approach is to include scatter modelling in the projection matrix. Either of these approaches is superior to pre-subtraction of the scatter estimate.

11.2.3.4. Collimator response compensation

Images obtained with a gamma camera are degraded by the spatially varying collimator–detector response (CDR). For parallel-hole collimators, the CDR depends approximately linearly on the distance from the collimator face. The CDR has geometric, septal penetration and septal scatter components. These correspond, respectively, to photons passing through the collimator holes, photons passing through the septa without interacting, and photons scattering in the septa and resulting in a detected photon. The latter two effects tend to reduce image contrast, can produce star artefacts and introduce distance dependent sensitivity variations. Septal penetration and scatter can be reduced with the use of a properly designed collimator. However, for isotopes emitting photons with energies greater than ~300 keV, including isotopes having low abundance high energy photons not used for imaging, some level of these effects is almost unavoidable.

Since SPECT images contain information about the distance from the source to the collimator, it is possible to provide improved compensation for the CDR as compared to planar imaging. This can be accomplished using both analytical and iterative methods. However, analytical methods involve approximations and usually have suboptimal noise properties and are, generally, not commercially available. As a result, methods based on statistical iterative reconstruction have been developed and are commercially available. These methods model the CDR in the projection and back projection process. In SPECT, the matrix elements are not explicitly calculated and stored, but are implicitly calculated during

the reconstruction using a projector and back projector algorithm. Thus, CDR compensation is accomplished by modelling the CDR in the projector and back projector. This is often implemented by rotating the image estimate, so that it is parallel to the collimator face at each projection view. In this orientation, the CDR is constant in planes parallel to the collimator face and can, thus, be modelled by convolution of the CDR for the corresponding distance. In order to do this, the distance from the plane to the face of the detector is needed. This is somewhat complicated by the use of non-circular orbits and, in this case, manufacturers need to store the orbit information (distance from the collimator face to the centre of rotation for each projection view) with the projection image. In addition, a way of estimating the CDR is needed. Analytical formulas exist for calculating the geometric component of the CDR. Alternatively, a Gaussian function fit to a set of measured point response functions measured in air can be used. Compensation for the full CDR, including septal penetration and scatter, requires a CDR that includes these effects. Analytical formulas do not exist, so either numerically calculated (e.g. using Monte Carlo simulations of the collimator–detector system) or measured CDRs are used. Various optimization and speed-up techniques have been implemented to reduce the time required for CDR modelling.

It should be noted that CDR compensation does not fully recover the loss of resolution of the collimator: the resolution remains limited and spatially varying and partial volume effects are still significant for small objects. In addition, CDR compensation results in correlated noise that can give a ‘blobby’ texture to the images (though the images do indeed seem qualitatively ‘less noisy’), and results in ringing artefacts at sharp edges. Despite these limitations, CDR compensation has generally been shown to improve image quality for both detection and quantitative tasks and has been used as a way to allow reduced acquisition time.

11.3. PET SYSTEMS

11.3.1. Principle of annihilation coincidence detection

Radioactive decay via positron emission is at the heart of the PET image formation process [11.3]. Positrons are emitted from the nucleus during the radioactive decay of certain unstable, proton-rich isotopes. These isotopes achieve stability by a decay process that converts a proton to a neutron and is associated with the creation of a positron. A positron is the antimatter conjugate of an electron and has the same mass as an electron but positive charge. As with β decay, positrons are emitted from the nucleus with different energies. These energies have a continuous spectrum and a specific maximum value

that is characteristic of the parent isotope. Once emitted from the nucleus, the positron propagates through the surrounding material and undergoes scattering interactions, changing its direction and losing kinetic energy (Fig. 11.26). Within a short distance, the positron comes to rest and combines with an electron from the surrounding matter. This distance is dependent on the energy of the positron, which is itself a function of the parent isotope and is typically on the order of a millimetre. The combination of a positron and an electron results in the annihilation of both particles and the creation of two photons, each with an energy of 511 keV, equivalent to the rest masses of the two original particles. Conservation of momentum, which is close to zero immediately before annihilation, ensures both photons are emitted almost exactly 180° apart. These characteristic photon emissions (known as annihilation radiation) — always 511 keV, always emitted simultaneously and almost exactly 180° apart — form the basis of PET and result in distinct advantages over single photon imaging in terms of defining the LOR.

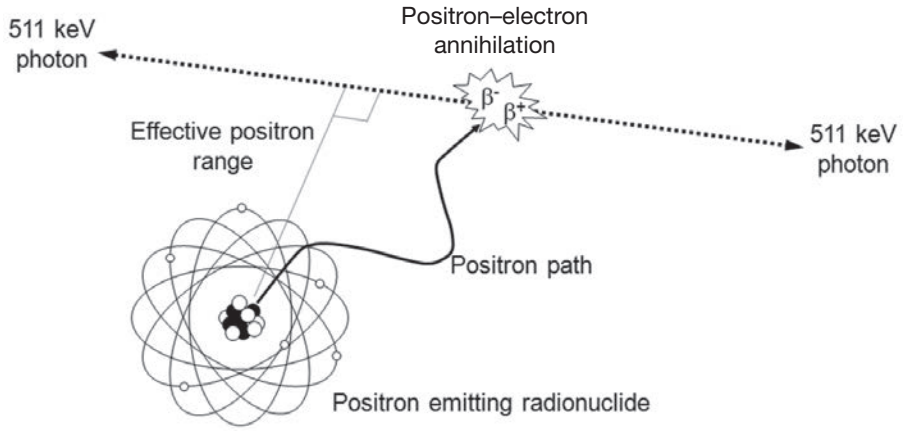

FIG. 11.26. Positrons emitted from a radioactive nucleus propagate through the surrounding material before eventually coming to rest a short distance from their site of emission. At this point, the positron annihilates with an electron, creating two 511 keV photons that are emitted approximately 180° apart. The perpendicular distance from the line defined by the two photons to the site of positron emission places a limit on the spatial resolution that can be achieved with PET systems.

The advantages of PET over SPECT, in terms of improved spatial resolution, statistical quality and quantitative accuracy, can be attributed to the fact that PET does not require a collimator and, therefore, eliminates the weakest link in the SPECT image formation process. Instead of physical collimation,

PET systems employ a form of detection that can be thought of as electronic collimation. If a positron source is surrounded by suitable detectors, both back to back photons from an individual positron decay can potentially be detected (Fig. 11.27). As both photons are emitted simultaneously, they will be detected at approximately the same time, allowing temporal acceptance criteria to be used to associate pairs of corresponding detection events. This mode of detection is referred to as coincidence detection and allows corresponding photon pairs to be distinguished from other unrelated, potentially numerous, photon detection events. As both photons that arise from positron decay are emitted almost exactly 180° apart, coincidence detection can be used to help localize the source of the photon emissions. In general, a line drawn between corresponding detectors can be assumed to intersect the point of photon emission, although information is usually not available about exactly where along that line the emission occurred. However, if a system of detectors is arranged at different positions around the source, multiple coincidence events can be recorded at different angular orientations. Over the course of an extended scanning period, a large number of coincidence events will be recorded and angular projections of the activity distribution can be estimated. These projections may then be used to reconstruct 3-D images using the methods of computed tomography.

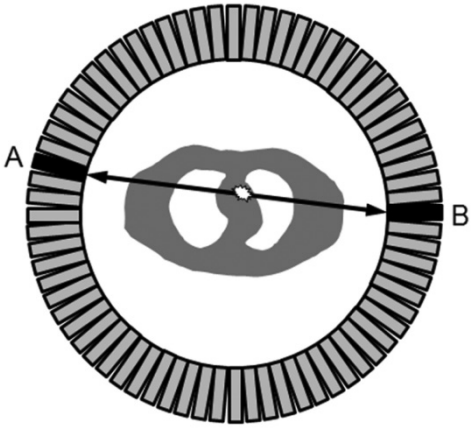

FIG. 11.27. The back to back photons that result from positron–electron annihilation can potentially be measured by detectors placed around the source. Coincidence detection involves the association of detection events occurring at two opposing detectors (A and B) based on the arrival times of the two photons. A line of response joining the two detectors is assumed to intersect the unknown location of the annihilation event. Coincidence detection obviates the need for a collimator and is sometimes referred to as electronic collimation.

11.3.2. Design considerations for PET systems

11.3.2.1. Spatial resolution

High spatial resolution is clearly an important design objective for PET imaging systems. As such, the trend in modern scanner systems has been to decrease the width of individual detectors and to increase the total number of detector elements surrounding the patient. The increased concentration of detector elements decreases the sampling interval and generally improves spatial resolution. Although modern designs involve detectors that are only a few millimetres wide, the need for high sensitivity means that they are often a few centimetres long. Problems can occur when photons are incident on one detector but penetrate through to an adjacent detector (Fig. 11.28(a)). The location within the detector (depth of interaction) is typically not measured and the detection event is assigned to a location at the face of the detector. This problem frequently occurs when detectors are arranged in a ring (or similar) configuration and gives rise to a loss of resolution at more peripheral locations. This resolution loss generally occurs in the radial direction as opposed to the tangential direction due to the angle of incidence of the photons on the detectors.

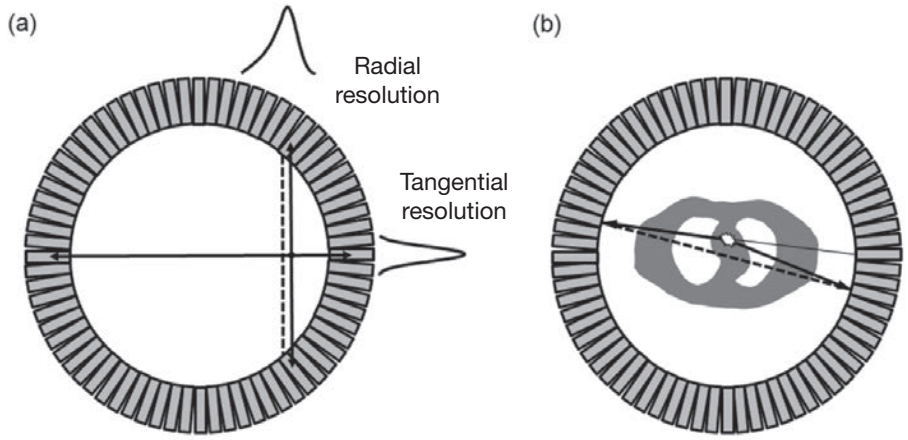

FIG. 11.28. (a) Photon penetration between adjacent detectors in a ring based system leads to mis-positioning of events. This primarily affects the radial component of spatial resolution which degrades with distance from the centre of the field of view. (b) Residual momentum of the positron and electron immediately before annihilation causes the two 511 keV photons to deviate slightly from the expected 180° angle. As a result, a line joining detection events does not intersect the exact point of annihilation. The extent of this non-collinearity is greatly exaggerated in the figure, but it does contribute to a loss of spatial resolution, especially for large diameter PET systems.

Another factor that influences spatial resolution is the distance between opposing detectors. This distance is relevant because of a small uncertainty in the relative angle of the 511 keV annihilation radiation (Fig. 11.28(b)). Although the basic assumption of coincidence detection is that annihilation radiation is emitted 180° apart, this is not strictly true. Positrons frequently annihilate before they have lost all momentum, and this residual momentum translates to a small deviation of about ±0.25° from the expected back to back emissions. This effect is referred to as non-collinearity and tends to degrade spatial resolution as detector separation increases. For PET systems with opposing detectors separated by only a few centimetres, such as those optimized for specific organs such as the brain or breast, this is not a major issue. However, for whole body systems, in which opposing detectors are typically separated by about 80 cm, the effect of non-collinear photons contributes a blurring with an FWHM of approximately 2 mm.

The distance travelled by a positron between its point of emission and annihilation is an additional factor that degrades the spatial resolution that can be achieved by PET systems. As previously discussed, this distance, or positron range, is dependent on the energy of the positron and also the type of material through which the positron is passing; a greater range is expected in tissue such as lung compared to soft tissue. It should be noted that software corrections have been implemented that model the effect of positron range and can potentially reduce the loss of resolution in reconstructed images.

11.3.2.2. Sensitivity

The best possible spatial resolution that can be obtained by a PET system is not always achieved in clinical practice due to statistical noise in the measured data (Fig. 11.29). In order to suppress this noise, clinical protocols generally employ low-pass filters or other similar image reconstruction methods, but the consequence is invariably a loss of spatial resolution. Improving the statistical

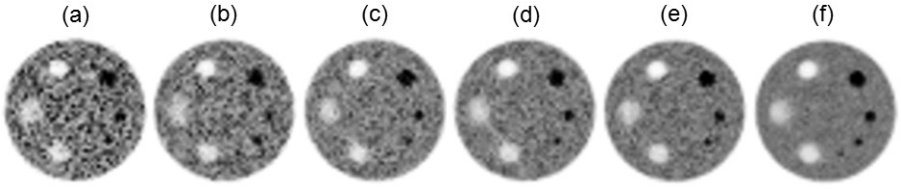

FIG. 11.29. Images of the same phantom, each showing different statistical quality. The images shown in (a), (b), (c), (d), (e) and (f) were acquired for 1, 2, 3, 4, 5 and 20 min, respectively. Increasing the acquisition time increases the total number of true coincidence events and reduces statistical variability in the image.

quality of the measured coincidence data not only reduces image noise but also allows the opportunity to reduce image smoothing and improve spatial resolution. The need to optimize this trade-off between statistical noise and spatial resolution influences both image reconstruction development and scanner design.

Noise in PET images is influenced by a number of factors, including the sensitivity of the detector system, the amount of radioactive tracer administered to the patient and the amount of time the patient can remain motionless for an imaging procedure. Limitations on the latter two factors mean that high sensitivity is an important objective for scanner design. Sensitivity is determined by the geometry of the detector arrangement and the absorption efficiency of the detectors themselves. Reducing the distance between opposing detectors increases the solid angle of acceptance and increases sensitivity. However, the requirement to accommodate all regions of the body imposes a minimum ring diameter for whole body systems. For such systems, extending the axial field of view (AFOV) of the detector system provides a mechanism for increasing sensitivity. Cost constraints have prevented the construction of PET systems that cover the entire length of the body. However, extended axial coverage can be achieved using systems with much smaller AFOVs by scanning sections of the body in a sequential fashion. For such systems, extending the AFOV of the scanner not only increases sensitivity but also reduces the number of bed translations required for whole body coverage. In addition to the geometry of the detector system, sensitivity is also determined by the absorption efficiency of the detectors. A high absorption efficiency for 511 keV photons is desirable in order to make best use of those photons that are incident upon the detectors. Absorption efficiency or stopping power of the detector material is, therefore, an important consideration for PET system design.

11.3.2.3. Quantitative accuracy

One of the strengths of PET is its capability to quantify physiological processes in vivo. A prerequisite for this kind of quantitative analysis is that the images accurately reflect the local activity concentration in the body. In order to ensure this kind of quantitative accuracy, it is important to minimize effects that corrupt the data and to correct residual corruption as necessary. Quantitative error can arise from many sources but is primarily due to random coincidence events, photon scatter within the body, photon attenuation within the body and detector dead time. Figure 11.30 illustrates some of these situations.

Figure 11.30(b) illustrates a type of unwanted coincidence event that can occur when photons from unrelated annihilation events are detected at approximately the same time. Two photons detected within a short time interval (coincidence timing window) will be associated with each other under the

assumption that they originated from the same positron–electron annihilation. However, when there is a large amount of radioactivity within the FOV, it is very possible that photons from unrelated annihilation events will be detected within this time interval, leading to a spurious random coincidence. Figure 11.30(c) illustrates the case where a scattered photon is deflected from its original direction but still reaches the detectors. In this case, a coincidence event can potentially be recorded between the scattered and unscattered photons, leading to an inaccurate coincidence event whose LOR does not pass through the true location of the original annihilation. If not adequately corrected, such scattered coincidences contribute a spurious background to the image that is dependent upon the size and composition of the patient's body and is non-uniform across the FOV. In Fig. 11.30(d), one of the two annihilation photons has been attenuated within the body and only one photon is detected. No coincidence event is possible and the

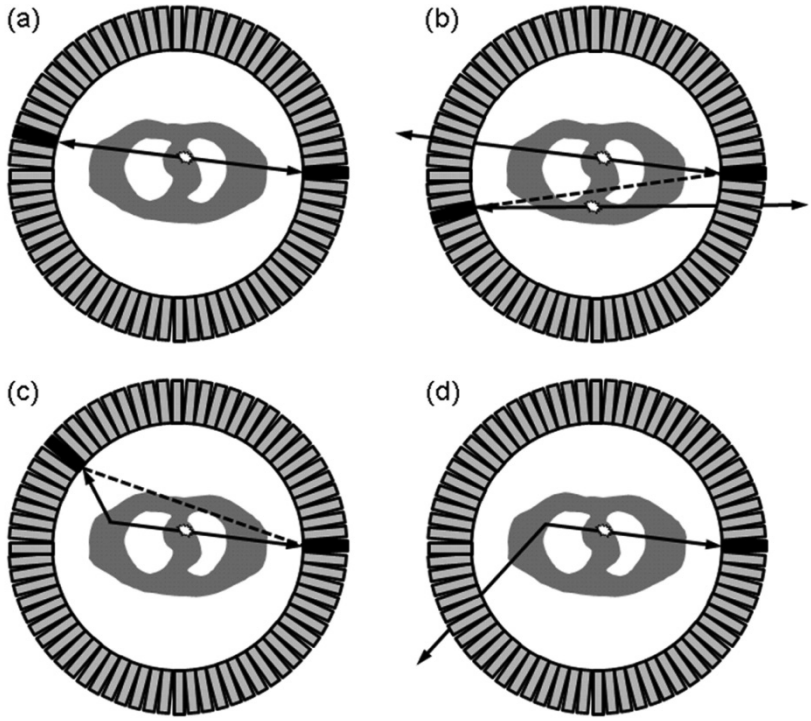

FIG. 11.30. (a) A true coincidence event can occur when both photons escape the body without interacting. (b) A random coincidence event occurs when two photons from unrelated annihilation events are detected at approximately the same time. (c) A scattered coincidence event can occur when either photon is scattered within the body but is still detected. (d) No coincidence event is recorded when one or both photons are attenuated, typically due to scatter out of the field.

scanner underestimates the signal that would be measured in the absence of attenuation. The degree of underestimation depends on the size and distribution of the patient's body and results in quantitative error and gross image artefacts that are non-uniform across the FOV. Attenuation can occur when one of the two photons undergoes photoelectric absorption within the body. However, a more likely occurrence is that one of the two photons is scattered within the body and is deflected out of the FOV, leaving only one photon to be detected. Another problem that can occur when large amounts of radioactivity are present is that true coincidence events can potentially be lost due to the limited count rate capability of the detection system. Each component of the system requires a finite amount of time to process each event and, if another photon is detected during this time, the second event will be lost. This dead time effect becomes significant at high count rates and leads to an underestimation of the local activity concentration in the images.

Software processing prior to (or during) image reconstruction can mitigate the above effects, but the accuracy of these corrections may not be reliable if the contamination overwhelms the signal from true coincidence events. PET systems are, therefore, designed to minimize the contribution of the various degrading factors described above. In terms of scanner design, very little can be done to reduce attenuation as photons that are absorbed within the body do not reach the detectors. However, scattered photons can potentially be rejected by the detection system if their energy falls outside a predetermined acceptance range. Annihilation radiation that is scattered within the body will emerge with energies less than 511 keV. The exact energy of the scattered photon will depend on the number of Compton scattering interactions that have occurred and the angle through which the photon was scattered. Energy discrimination can, therefore, be used to reject photons that have energies less than 511 keV and are assumed to have been scattered. This approach relies on the detection system having high energy resolution and, in practice, energy discrimination reduces but does not eliminate scattered coincidence events. The limited energy resolution of current PET systems means that, in order to avoid rejecting too many true (unscattered) coincidence events, the energy acceptance window is usually set to accept quite a broad range of energies around 511 keV. High energy resolution is, nevertheless, an important design objective, particularly for those scanner systems that detect a large proportion of scattered photons.

Decreasing the coincidence timing window decreases the number of random events as the shorter time interval reduces the likelihood of a coincidence occurring by chance between two unrelated photons. There is, however, only limited scope for reducing the duration of the coincidence time window as it is restricted by the timing resolution of the detector system and the fact that off-centre annihilations result in real time differences between detection events.

Optimization of the coincidence timing window for a particular scanner represents a compromise between wanting to reduce the number of random coincidence events without significantly reducing the number of true coincidences. Detector systems that are able to measure photon detection times with low variability (high timing resolution) are, therefore, desirable from the perspective of randoms reduction. Timing resolution also contributes to detector dead time as a shorter coincidence timing window reduces the likelihood of more than two photon detection events occurring. Other contributions to dead time include the time required by the detector to measure an individual photon event and the time spent processing coincidence events.

11.3.2.4. Other considerations

Spatial resolution, sensitivity and quantitative accuracy are the main physics issues influencing PET system design, but there are various other factors that need to be considered. One obvious issue is the overall cost of the system. This significantly influences design decisions as the detectors represent a significant fraction of the overall production cost. The choice of detector material, the thickness of the detectors, the diameter of the detector ring and the axial extent of the detectors all contribute to the total cost of the system. Another important design issue that affects the overall cost is integration of the PET system with a second modality in a combined scanner. In most cases, this means integration of the PET subsystem with a CT subsystem, although combined PET/magnetic resonance (MR) scanners exist and present more significant technical challenges. Combined PET/CT scanners have effectively replaced stand-alone PET for clinical applications and the optimal CT configuration included in a combined system is limited mainly by cost concerns. In practice, this means the number of slices to include in the multi-detector CT system. High performance multi-detector CT is important for systems intended for cardiac PET/CT applications, whereas lower slice capability may be adequate for oncological PET/CT. Other relevant design issues are related to the computer workstation used for acquiring and processing data. These include fast image reconstruction times, convenient integrated control of both PET and CT subsystems, and seamless integration with other institutional information technology systems such as clinical information systems and image archive systems.

11.3.3. Detector systems

11.3.3.1. Radiation detectors

Although different radiation detector designs have been used in PET, almost all current systems adopt an approach based on scintillation detectors. Scintillation detectors are inorganic crystals that emit scintillation light in the visible range when high energy photons are incident upon them (Fig. 11.31). This light is converted to an electrical signal by means of a photodetector, usually a PMT, coupled to the crystal material. Various scintillator materials have been used in PET, including thallium doped sodium iodide (NaI(Tl)), bismuth germanate (BGO) and cerium doped lutetium oxyorthosilicate (LSO). Table 11.1 shows some of the properties of the crystal materials that are relevant for PET applications. The properties of an ideal crystal for PET would include a high stopping power for 511 keV photons (high linear attenuation coefficient); short scintillation light decay time to reduce dead time and allow short coincidence time windows to reduce random coincidences; and high light output. High light output enables good energy resolution, which gives rise to improved scatter

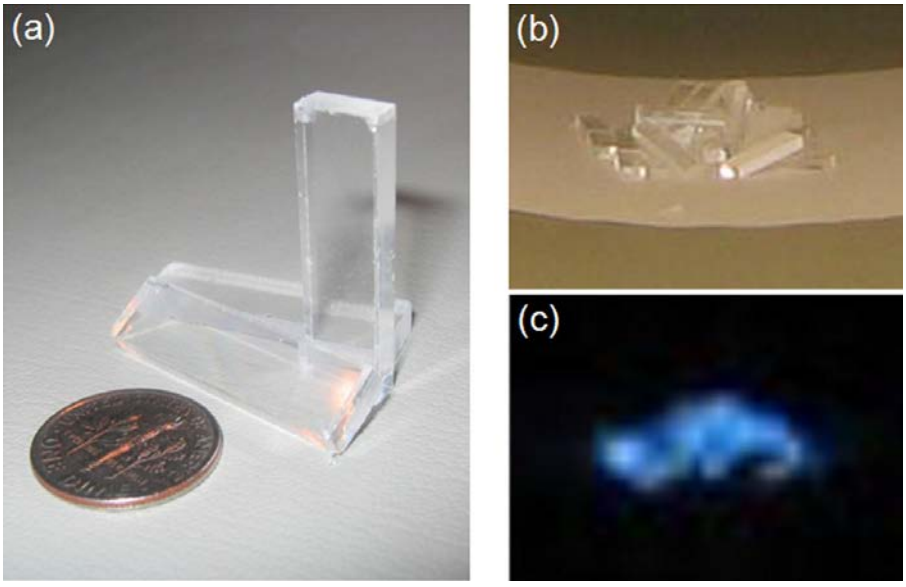

FIG. 11.31. Example of bismuth germanate crystals used for PET (a). Bismuth germanate samples photographed under room lighting (b) and in the presence of X ray irradiation and dimmed room lighting (c). The scintillation light seen in (c) is due to the interaction of radiation with the crystals, which causes electrons to become excited. When they return to their ground state, energy is emitted, partly in the form of visible light.

rejection. It also affords cost savings in the construction of a complete scanner system as the number of photodetectors required to resolve a given number of crystal elements can potentially be reduced.

TABLE 11.1. PROPERTIES OF SOME OF THE SCINTILLATORS USED IN PET. LINEAR ATTENUATION COEFFICIENTS AND ENERGY RESOLUTION ARE QUOTED FOR 511 keV

Property	NaI	BGO	LSO
Linear attenuation coefficient (cm^{-1})	0.34	0.95	0.87
Scintillation decay constant (ns)	230	300	40
Relative light output	100%	15%	75%
Energy resolution (%)	6.6	10.2	10.0

Note: BGO: bismuth germanate; LSO: lutetium oxyorthosilicate.

Although NaI(Tl) is ideal for lower energy single photon imaging, its relatively low linear attenuation coefficient for 511 keV photons makes it less attractive for PET applications. Sensitivity could potentially be increased by increasing the thickness of the crystals, which are typically 1–3 cm thick. However, the scope for substantially increasing crystal thickness is limited as it results in a loss of spatial resolution. This is because thicker crystals are prone to more significant depth of interaction problems as the apparent width of the detector increases for sources located off-centre. Thin crystals composed of a material with a high stopping power for 511 keV photons are, thus, desirable to ensure best possible sensitivity while maintaining spatial resolution. For this reason, BGO and, more recently, LSO have replaced NaI(Tl) as the scintillator of choice for PET.

BGO has the advantage of a high stopping power for 511 keV photons and has become an important scintillator for PET applications. However, it is not ideal in many respects as it has relatively poor energy resolution and a long crystal decay time. The poor energy resolution translates into a limited ability to reject scatter via energy discrimination. In addition, the long decay time translates into a greater dead time and increased number of random coincidences at high count rates. As such, BGO is well suited for scanner designs that minimize scatter and count rate via physical collimation, such as those with interplane septa (see Section 11.3.4.2). Attempts to increase sensitivity by removing the interplane septa typically result in a high scatter, high count rate environment for which BGO is not ideal.

Although LSO has a lower linear attenuation coefficient than BGO, its shorter crystal decay time and slightly improved energy resolution convey

significant advantages as a PET scintillator. LSO has become the scintillator of choice for scanner designs that operate without interplane septa because its short decay time makes it well suited for high count rate applications. The fast decay time of LSO also enables a time of flight (TOF) data acquisition mode that will be discussed in Section 11.3.4.4. LSO has proved to be a successful crystal for PET detector applications despite the fact that the material contains around 2.6% ^{176}Lu, which is itself radioactive. A component of the emissions from ^{176}Lu is detected within the energy acceptance window and the dominant effect is to contribute random coincidences. In practice, the increased randoms rate is not a major problem for clinical studies, and the naturally occurring radiation has even been used for quality assurance. Lutetium-176 has a half-life of 3.8×10^{10} a and, thus, provides a long lived source of radiation that can be used to check consistency of detector response without the need for external sources. It should be noted that, for commercial reasons, some PET systems employ cerium doped lutetium yttrium oxyorthosilicate (LYSO(Ce)) which has substantially similar properties to LSO.

11.3.3.2. Detector arrangements

As discussed above, the interaction of 511 keV annihilation radiation with the scintillation crystals gives rise to optical light that can be detected by a suitable photodetector. A photodetector is a device that produces an electrical signal when stimulated by light of the sort emitted by a scintillation detector. For most PET applications, PMTs have been the preferred photodetector because their high gain results in an electrical output with a good signal to noise ratio. In addition, PMT output is proportional to the intensity of the incident light and, thus, proportional to the energy deposited in the crystal. This provides a mechanism for selective acceptance of only those detection events with energies within a specific range and can be used to reject scattered photons. In addition, PMTs provide high amplification with little degradation in the timing information that is essential for electronic collimation. Although PMTs are by far the most widely used photodetector for PET applications, they are somewhat bulky and highly sensitive to magnetic fields. For these reasons, they are often not used in combined PET/MR systems where space is limited and operation in high magnetic fields is a requirement. In these and some other applications, semiconductor based photodiodes are an alternative to PMTs. It should be noted that in these applications, the semiconductor device is used in conjunction with a scintillation detector and is used to detect scintillation light, not the annihilation photons. Avalanche photodiodes have much lower gains than PMTs but can be very small and have been shown to be effective in high magnetic field environments. Their low gain requires very low noise electronics and they are also sensitive to small temperature variations.

Space and cost constraints mean that individual scintillation crystals are not usually coupled directly to individual photodetectors in a one to one fashion. Instead, the most common arrangement is a block detector in which a group of crystal elements share a smaller number of PMTs (Fig. 11.32). The design of each block varies between manufacturers and scanner models but usually involves a matrix of crystal elements, a light guide and four PMTs. An example configuration might be an 8 × 8 array of closely packed 4.4 mm × 4.0 mm × 30 mm crystal elements, where the longest dimension is in the radial direction to maximize detection efficiency. The light guide allows light to be shared between four circular PMTs and the relative light distribution depends on the location of the crystal in which the photon interacted. The (x, y) position of the detection event is calculated from the outputs of the four PMTs using a weighted centroid algorithm, similar to the Anger logic of a gamma camera. Although individual crystals can be identified in this way, the response is not linear throughout the block due to differences in the locations of the different crystal elements relative to the PMTs. Experimentally determined look-up tables are used to relate the measured (x, y) position to a corresponding detector element, effectively performing a form of linearity correction. In this way, only four PMTs are needed to localize signals from a much greater number of crystal elements. The number of crystal elements divided by the number of PMTs in a PET system has been referred to as the encoding ratio. A high encoding ratio implies lower production costs and is, therefore, desirable.

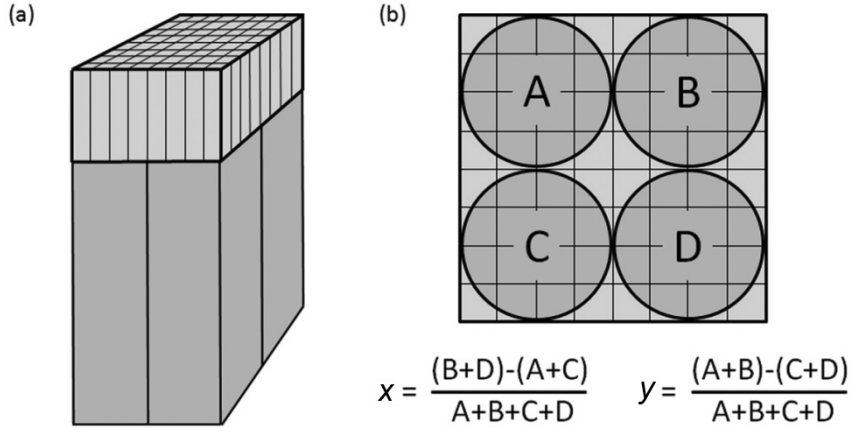

FIG. 11.32. (a) A PET detector block consisting of scintillator material coupled to an array of photomultiplier tubes. The scintillator is cut into an array of individual crystal elements. Four photomultiplier tubes are used to read out the signal from the 8 × 8 array of crystal elements. (b) The x and y position of each photon is determined from the signal measured by each of the four photomultiplier tubes labelled A–D, using the equations shown.

One of the advantages of the design described above is that each block operates independently of its surrounding blocks. This leads to good count rate performance as light is not transferred between blocks and the PMTs of one block are unaffected by detection events in an adjacent block. An alternative arrangement, referred to as quadrant sharing, increases the encoding ratio by locating the PMTs at the corners of adjacent blocks. This arrangement differs from the conventional block design in that each PMT can now be exposed to light from up to four different blocks. This can result in better spatial resolution and a higher encoding ratio but is also susceptible to greater dead time problems at high count rates.

Another alternative to the block design adopts an approach similar to that used in conventional gamma cameras. These Anger-logic designs involve detector modules that have a much larger surface area compared to conventional block detectors, e.g. 92 mm × 176 mm. Each module is comprised of many small crystal elements which are coupled, via a light guide, to an array of multiple PMTs. Light is spread over a larger area than in the block design and positional information is obtained using Anger-logic in the same way as a gamma camera. The PMTs used in this design are typically larger than those used in block detectors, increasing the encoding ratio. The larger area detector modules encourage more uniform light collection compared to block designs, which leads to more uniform energy resolution. However, a disadvantage of this design is that the broad light spread among many PMTs can lead to dead time problems at high count rates.

11.3.3.3. Scanner configurations

The detectors described above form the building blocks used to construct complete scanner systems. Various scanner configurations have been developed, although the dominant design consists of a ring of detectors that completely surrounds the patient (or research subject) in one plane (Fig. 11.33(a)). As with other scanner systems, this plane is referred to as the transverse or transaxial plane and the direction perpendicular to this plane is referred to as the axial or z direction. Several rings of detectors are arranged in a cylindrical geometry, allowing multiple transverse slices to be simultaneously acquired. As coincidence detection requires two opposing detectors, a full ring system of this sort allows coincidence data to be acquired at all angles around 180°. Although complete angular coverage is achieved in the transverse plane, there is much more limited coverage in the axial direction. Cost constraints and, to some extent, limited patient tolerance of extended tunnels mean that the detector rings usually extend for only a few centimetres in the axial direction. Human whole body systems typically have an AFOV of around 15–20 cm, although the trend in scanner

design has been to increase the AFOV, thus increasing both sensitivity and the number of transverse slices that can be simultaneously acquired.

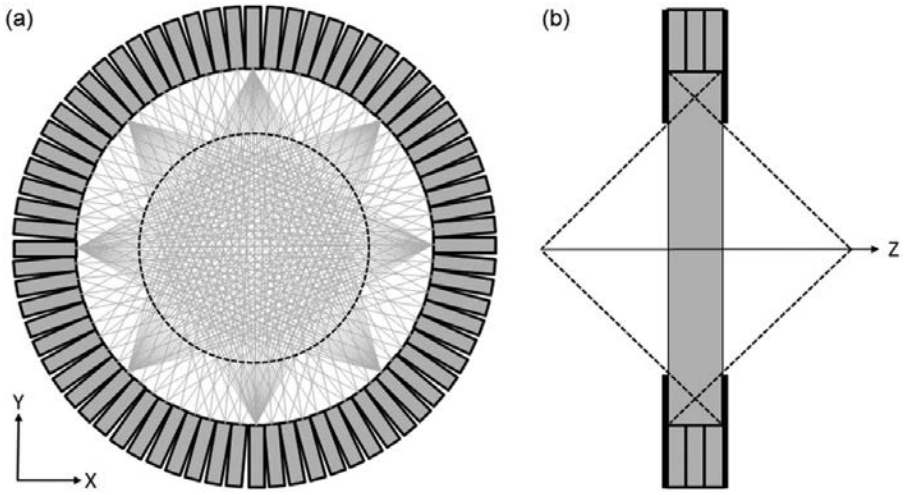

FIG. 11.33. (a) Full ring PET system shown in the transverse plane, indicating how each detector can form coincidence events with a specific number of detectors on the opposite side of the ring. For clarity, this fan-like arrangement of lines of response is shown for only eight detectors. The dashed line indicates how the imaging field of view is necessarily smaller than the detector ring diameter. (b) PET system shown in side elevation, indicating the limited detector coverage in the z direction. The shaded area indicates the coincidence field of view. The dashed lines indicate the singles field of view. End shields reduce, but do not eliminate, detection of single photons from outside of the coincidence field of view when operating in 3-D mode.

The diameter of the detector ring varies considerably between designs and this dimension reflects the intended research or clinical application. Small animal systems may have ring diameters of around 15 cm, brain oriented human systems around 47 cm and whole body human systems around 90 cm. Systems with ring diameters that can accommodate the whole body are clearly more flexible in terms of the range of studies that can be performed. However, smaller ring diameters have advantages in terms of increased sensitivity, owing to a greater solid angle of acceptance, and potentially better spatial resolution, owing to reduced photon non-collinearity effects. It should be noted that the spatial resolution advantage is complicated by greater depth of interaction problems as the detector ring diameter decreases and shorter crystals or depth of interaction measurement capability may be required. Furthermore, the effective imaging FOV is always smaller than the detector ring diameter because the acquisition of coincidence events between all possible detector pairs (such as those between nearby detectors in the ring)

is not supported. In addition, PET systems have annular shields at the two ends of the detector ring that reduce the size of the patient port. These end shields are intended to decrease the contribution of single photons from outside the coincidence FOV (Fig. 11.33(b)). The coincidence FOV refers to the volume that the detector system surrounds, within which coincidence detection is possible. Single photons originating from outside the coincidence FOV cannot give rise to true coincidence events but may be recorded as randoms and can also contribute to detector dead time. Reducing the size of these end shields allows the patient port size to be increased but also leads to greater single photon contamination.

Unlike rotating camera SPECT systems, where different projections are acquired in a sequential fashion, full ring PET systems simultaneously acquire all projections required for tomographic image formation. This has an obvious advantage in terms of sensitivity, and it also enables short acquisition times, which can be important for dynamic studies. Full ring systems are, however, associated with high production costs and, for this reason, some early PET designs employed a partial ring approach. In these designs, two large area detectors were mounted on opposite sides of the patient and complete angular sampling was achieved by rotating the detectors around the z axis. Gaps in the detector ring led to reduced sensitivity and the partial ring design is now usually reserved for prototype systems. Another related approach to PET system design was to use dual head gamma cameras modified to operate in coincidence mode. The use of modified gamma cameras allowed for lower cost systems capable of both PET and SPECT. However, the poor performance of NaI based PET means that this approach has now been discontinued.

Current clinical systems have an AFOV that is adequate to cover most individual organs but in order to achieve coverage of the whole body, patient translation is required. Given the clinical importance of whole body oncology studies, the mechanism for translating the patient through the scanner has become an important component of modern PET systems. The patient bed or patient handling system has to be made of a low attenuation material but must still be able to support potentially very heavy patients. It must be capable of a long travel range, so as to allow whole body studies in a single pass without the need for patient repositioning. Precise motion control is also critical, particularly for PET/CT systems where accurate alignment of the two separately acquired modalities is essential. Advanced patient handling systems have been specifically developed for PET/CT to ensure that any deflection of the bed is identical for both the CT and PET acquisitions, thus ensuring accurate alignment irrespective of patient weight. Although most patient beds have a curved shape for improved patient comfort and better mechanical support, many manufacturers can also provide a flat pallet that is more compatible with radiation treatment positioning.

11.3.4. Data acquisition

11.3.4.1. Coincidence processing

The basis of coincidence detection is that pairs of related 511 keV annihilation photons can be associated together by the detector system based upon their times of measurement. Two photons detected within a short time interval are assumed to have arisen from the same positron–electron annihilation and a coincidence event is recorded. The time interval determining when events are considered to be coincident is denoted 2τ and is a system parameter that is not usually adjustable by the user. In order to minimize random coincidences, this interval should be kept as short as possible and for typical BGO based systems, 2τ may be around 12 ns. Shorter time windows are made possible by detector materials such as LSO that have faster scintillation decay times and, thus, better time resolution. Further reductions in the coincidence window are limited by differences in the arrival times of the two photons. For an electron–positron annihilation taking place at the edge of the transverse FOV, one photon would have to travel only a short distance, whereas the other photon might have to travel almost the diameter of the detector ring. Assuming a 90 cm ring diameter and a speed of light of 3×10^8 m/s, it can be seen that a maximum time difference of around 3 ns can be expected.

When a photon is incident upon a PET detector, an electrical pulse is generated (Fig. 11.34). A constant fraction discriminator then produces a digital logic pulse when the detector signal reaches a fixed fraction of the peak pulse height. This digital logic pulse is defined to have a duration τ and is fed into a coincidence module that determines whether it forms a coincidence event with any of the signals from other detectors in the ring. A coincidence event is indicated if there is an overlap in time between separate logic pulses from two different detectors. In other words, a coincidence would be identified if two detection events were recorded within a time interval no greater than τ. A photon detected at time t can, thus, form a coincidence with another photon detected within the interval $t + \tau$. It can alternatively form a coincidence with an earlier photon detected within $t - \tau$. For this reason, 2τ is often referred to as the coincidence time window. Consistent timing of signals from every detector in the system is clearly essential to ensure that true coincidences are effectively captured within the coincidence time window. Differences in performance of the various detector components and cable lengths can introduce variable time offsets. For this reason, time alignment corrections are performed to characterize and compensate for timing differences between different detectors in the ring.

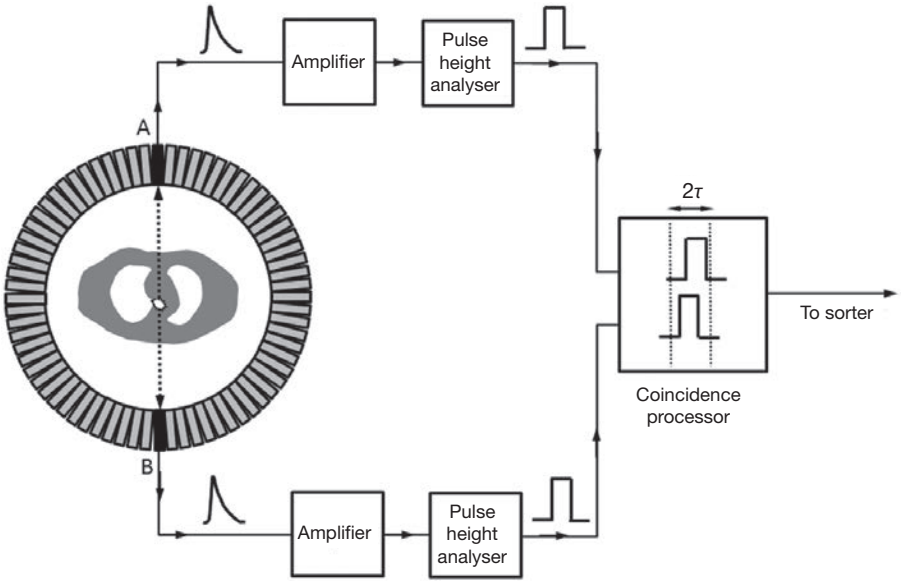

FIG. 11.34. Coincidence circuits allow two photon detection events to be associated with each other based upon their arrival times. Photons detected at A and B produce signals that are amplified and analysed to determine whether they meet the energy acceptance criteria. Those signals that fall within the energy acceptance window produce a logic pulse (width τ) that is passed to the coincidence processor. A coincidence event is indicated if both logic pulses fall within a specified interval (2τ).

Coincidence detection assumes that only two photons were detected, but with multiple disintegrations occurring concurrently, it is possible for three or more photons to be recorded by separate detectors within the coincidence time window. When this occurs, it is unclear which pair of detectors corresponds to a legitimate coincidence and multiple events of this sort are often discarded. This circumstance is most likely to occur when there is a large amount of activity in or around the FOV, and it contributes to count loss at high count rates. Another possible scenario is that only one photon is detected within the coincidence time window and no coincidence event will be recorded. These single photon detection events are a result of a number of reasons: the angle of photon emission was such that only one of the two annihilation photons was incident upon the detectors; one of the two annihilation photons was scattered out of the FOV; one of the two annihilation photons was absorbed within the body; and photons originating from outside the coincidence FOV. Although these single photons cannot form true coincidences, they are a major source of randoms.

The mechanism described above records what are known as prompt coincidence events which consist of true, random and scattered coincidences.

The relative proportion of each component depends on factors such as the count rate, the size of the attenuation distribution (patient size) and the acquisition geometry (2-D or 3-D). Only the trues component contributes useful information and the randoms and scattered coincidences need to be minimized. The randoms component is maintained as low as possible by setting the coincidence time window to the shortest duration consistent with the time resolution of the system. The scatter component is maintained as low as possible by energy discrimination. In addition to providing positional and timing information, the pulse produced by the detectors can be integrated over time to provide a measure of the energy deposited in the detector. In a block detector with four PMTs, the sum of the signals from each PMT is proportional to the total amount of scintillation light produced and, thus, the total energy deposited in the detector material. Under the assumption that the photon was completely absorbed in the detector, this signal provides a measure of the photon's energy and can be used to reject lower energy photons that have undergone Compton scattering within the patient. In practice, the energy resolution of most PET detector systems is such that the energy acceptance window must be set quite wide to avoid rejecting too many unscattered 511 keV photons. For BGO based systems, an energy acceptance range of 350–650 keV is typical. As small-angle scatter can result in only a small loss of energy, many of these scattered photons will be accepted within the energy window, despite the fact that they do not contribute useful information. Energy discrimination, therefore, reduces, but does not eliminate, scattered coincidence events and additional compensation is required.

11.3.4.2. Data acquisition geometries

Scanners consisting of multiple detector rings provide extended axial coverage and are advantageous for rapid acquisition of volumetric data. However, the presence of multiple detector rings raises issues concerning the optimum combinations of detectors that should be used to measure coincidence events. In a system with only one ring of detectors, the acquisition geometry is simple as each detector measures coincidence events with other detectors on the opposite side of the same ring. When additional detector rings are added to the system, it is possible to allow coincidence events to be recorded between detectors in different rings. This means including photons that were emitted in directions that are oblique to the transverse plane. Given that photons are emitted in all directions, increasing the maximum ring difference increases the angle of obliqueness that is accepted and, therefore, increases system sensitivity. The data acquisition geometry refers to the arrangement of detector pairs that are permitted to form coincidence events and, in practice, involves the presence or absence of interplane septa (Fig. 11.35). Data acquisition with septa in place is referred

to as 2-D mode; data acquisition without any interplane septa is referred to as 3-D mode. The 2-D/3-D designation refers to the acquisition geometry rather than the resulting images as both modes produce similar volumetric images.

In 2-D acquisition mode, an array of septa is inserted between the detector rings. These septa are annular and are typically made of tungsten. The purpose of the septa is to physically absorb photons incident at large oblique angles relative to the transverse plane, allowing only those photons incident approximately orthogonal to the z axis of the scanner. These septa differ significantly from gamma camera parallel-hole collimators as in the PET case, no collimation is provided within the transverse planes. By physically rejecting almost all oblique photons from reaching the detectors, the count rate is substantially reduced, resulting in a low rate of random coincidences and low detector dead time. In addition, 2-D acquisition is associated with a low rate of scattered coincidence events since only photons emitted in and scattering within a transverse plane can pass through the septa. However, if a Compton interaction occurs, the likelihood is that the scattered photon will emerge at an oblique angle. Photons that undergo a Compton interaction are scattered through an angle that is distributed over 4π, so although the photon may be scattered in-plane, out-of-plane scatter is more likely. Photons scattered through oblique angles will be absorbed by the septa, effectively reducing the fraction of scattered coincidences that are measured.

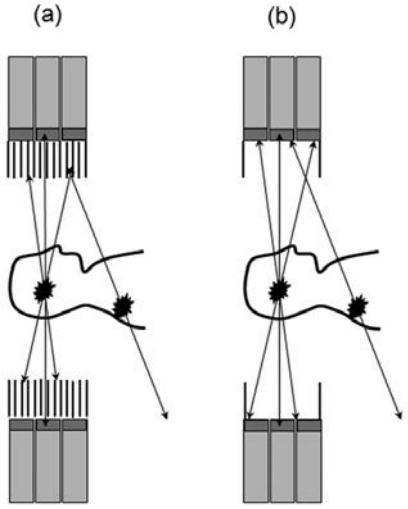

FIG. 11.35. (a) 2-D and (b) 3-D acquisition geometries. In 2-D mode, a series of annular septa are inserted in front of the detectors so as to absorb photons incident at oblique angles. In 3-D mode, these septa are removed, allowing oblique photons to reach the detectors. 3-D mode is associated with high sensitivity but also increased scatter and randoms fractions, the latter partly due to single photons from outside the coincidence field of view.

In 3-D acquisition mode, the septa are entirely removed from the FOV and there is no longer any physical collimation restricting the photons that are incident upon the detectors. Coincidence events can be recorded between detectors in different rings and potentially between all possible ring combinations. Photons emitted at oblique angles with respect to the transverse plane are no longer prevented from reaching the detectors, and system sensitivity is substantially increased compared to 2-D acquisition. Sensitivity gains by a factor of around five are typical, although the exact value depends on the scanner configuration and the source distribution. In 2-D mode, sensitivity varies slightly between adjacent slices but does not change greatly over the AFOV. In 3-D mode, the sensitivity variation in the axial direction is much greater and has a triangular profile with a peak at the central slice. The triangular axial sensitivity profile can be understood by considering a point source centrally located in the first slice at one of the extreme ends of the scanner. True coincidence events can only be recorded between detectors in the first ring. As the source is moved towards the central slice, coincidence events can be recorded between an increasing number of detector ring combinations, leading to an increase in sensitivity. As a consequence of the substantial sensitivity increase, 3-D acquisition is associated with higher detector count rates, leading to more randoms and greater dead time than corresponding acquisitions in 2-D mode. Furthermore, 3-D mode cannot take advantage of the scatter rejection afforded by interplane septa and, as a result, records a greatly increased proportion of scattered coincidence events.

The advantage of 3-D acquisition is its large increase in sensitivity compared to 2-D acquisition. This would be expected to result in images with improved statistical quality or, alternatively, comparable image quality with shorter scan times or reduced administered activity. In practice, evaluating the relative advantage of 3-D acquisition is complex as it is associated with substantial increases in both the randoms and scatter components. Both of these unwanted effects can be corrected using software techniques, but these corrections can themselves be noisy and potentially inaccurate. Furthermore, the relative contribution of randoms and scattered photons is patient specific. The randoms and scatter fractions are defined as the randoms or scatter count rate divided by the trues rate, and both increase with increasing patient size. Both randoms and scatter fractions are substantially higher in 3-D compared to 2-D mode. In 3-D mode, scatter fractions over 50% are common, whereas 15% is more typical for 2-D mode. Randoms fractions are more variable as they depend on the study, but randoms often exceed trues in 3-D mode.

A figure of merit that is sometimes useful when considering the performance of scanner systems is the noise equivalent count rate (NECR). The NECR is equivalent to the coincidence count rate that would have the same noise properties as the measured trues rate after correcting for randoms and scatter.

NECR is commonly used to characterize 3-D performance and, since the relative proportion of the different kinds of coincidence events is strongly dependent on object size, standardized phantoms have been developed. It is computed using:

$$\mathrm{NECR} = \frac{T^2}{T + S + 2fR} \tag{11.12}$$

where

T, S and R are the true, scatter and random coincidence count rates, respectively;

and f is the fraction of the sinogram width that intersects the phantom.

For a given phantom, the NECR is a function of the activity in the FOV and is usually determined over a wide activity range as a radioactive phantom decays (Fig. 11.36). The reason for this count rate dependence is twofold: the randoms rate increases as the square of the single photon count rate (which is approximately proportional to the activity in the FOV) and the sensitivity of the scanner for trues decreases with increasing count rates as detector dead time becomes more significant.

An important factor when considering the relative performance of 2-D and 3-D acquisition modes is the characteristics of the detector material. In 2-D mode, the septa substantially reduce dead time, randoms and scatter, making the poor timing and energy resolution of BGO less of a limitation. BGO based systems are, thus, well suited to 2-D acquisition mode. However, for BGO, the sensitivity advantage of 3-D acquisition mode is substantially offset by the high randoms and scatter fractions that are encountered. For systems based on detectors such as LSO, the improved timing resolution can be used to reduce the coincidence time window and, thus, reduce the randoms fraction. The improved energy resolution also allows the lower level energy discriminator to be raised, resulting in a lower scatter fraction. LSO or similar fast detector materials are, thus, well suited to 3-D acquisition mode. The introduction of these detectors, along with improved reconstruction algorithms for 3-D data, means that 3-D acquisition mode now dominates. Many scanner systems no longer support 2-D mode as this allows the septa to be completely removed from the design, reducing cost and potentially increasing the patient port diameter.

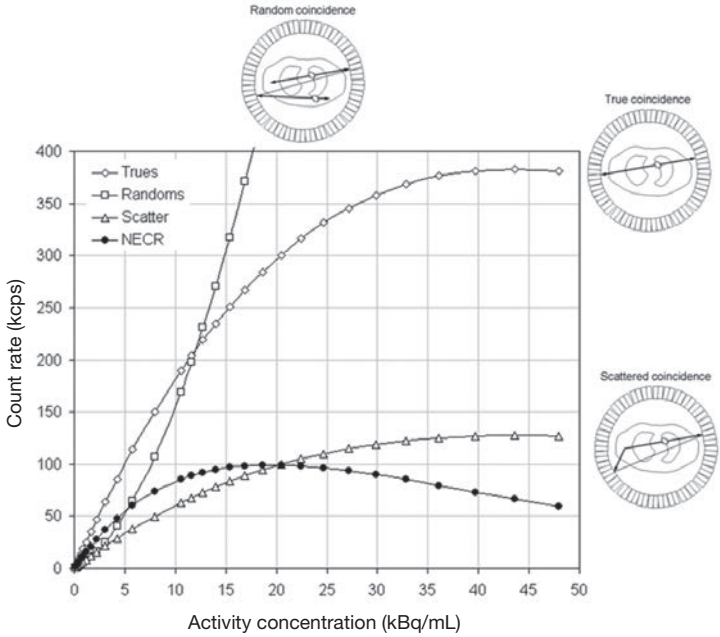

FIG. 11.36. The relative proportion of true, random and scattered coincidence events as a function of activity in the field of view. At low activities, the true coincidence count rate increases linearly with activity. However, at higher activities, detector dead time becomes increasingly significant. The trues rate increases less rapidly with increasing activity and can even decrease at very high activities. The randoms count rate increases with increasing activity as a greater number of photons are detected. The scatter count rate is assumed to be proportional to the trues rate. Scanner count rate performance can be characterized using the noise equivalent count rate (NECR), which is a function of the true, random and scatter coincidence count rates.

11.3.4.3. Data organization

The data recorded during a conventional PET acquisition are the total number of coincidence events measured between the various detector pairs. These data are typically binned into 2-D matrixes known as sinograms (Fig. 11.37). If a 2-D acquisition geometry is considered, each row of the sinogram represents a projection of the radionuclide distribution at a particular angle around the patient. These projections consist of coincidence events recorded between pairs of detectors, where each detector pair forms LORs that are approximately parallel to each other. The number of counts in each element of the projection is proportional to a line integral of the radionuclide distribution within the limitations imposed by the various physical effects such as scatter and attenuation. The sinogram is

indexed along the y axis by angle and the x axis by distance. For a full ring system, angular sampling is usually evenly spaced over 180° but the sampling along each row is slightly non-linear. The separation of adjacent elements in the projection decreases towards the edges of the FOV owing to the ring geometry. Correction for this effect, known as arc correction, is required and is usually implemented during image reconstruction. Adjacent elements within a particular sinogram row would be expected to be associated with two parallel LORs joining detector pairs that are next to each other in the ring. In practice, improved sampling is achieved by also considering LORs that are offset by one detector. Despite the fact that these LORs are not exactly parallel to the others, these data are inserted into the sinogram rows as if they came from virtual detectors positioned in the gaps between the real detectors.

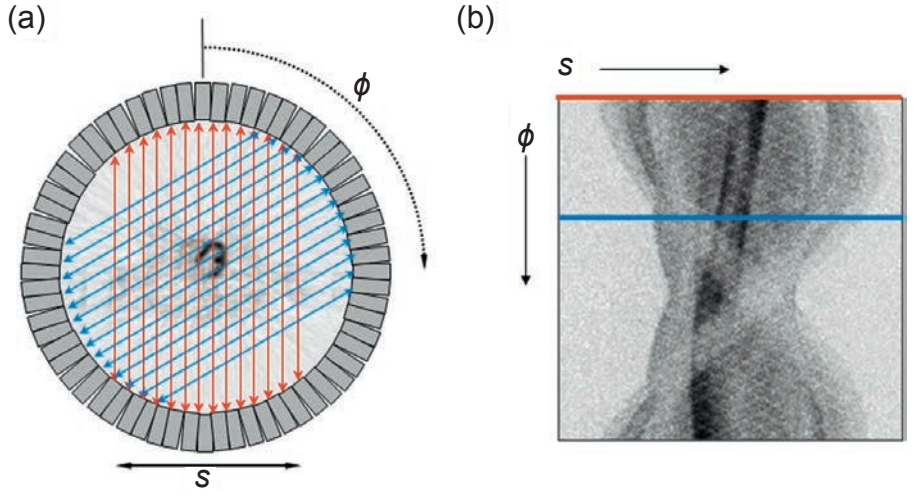

FIG. 11.37. Full ring PET scanners simultaneously measure multiple projections at different angles ϕ with respect to the patient. An example showing the orientation of two parallel projections is shown in (a). Projection data of this sort are typically stored in sinograms; an example is shown in (b). In a sinogram, each row represents a projection at a different angle ϕ. Each projection is made up of discrete elements that are indexed by s and contain the number of coincidence counts recorded along individual lines of response. The two example projections shown in (a) are also highlighted in sinogram (b).

The prior discussion of 2-D acquisition mode only considered coincidence events between detectors in a single ring, referred to as a direct plane. In practice, the interplane septa do not completely eliminate the possibility of coincidence events being detected between different nearby rings (Fig. 11.38). Inclusion of these slightly oblique events is advantageous as it increases sensitivity.

Coincidence events between detectors in immediately adjacent rings are combined into a sinogram that is considered to have been measured in a plane located between the two detector rings. This plane is referred to as a cross plane and is considered to be parallel to the direct planes, despite the fact that the contributing LORs are slightly oblique to these planes. As well as increasing sensitivity, inclusion of these cross planes increases axial sampling by producing $2N - 1$ slices from an N ring scanner. Sensitivity is further increased for cross planes by extending the ring difference between which coincidences are allowed from one to three or higher odd numbers. This principle is also applied to direct planes, resulting in coincidence events not just within the same ring but also between detectors with ring differences of two or higher even numbers. It should be noted that the data obtained from these different ring combinations are added together, so individual sinograms actually consist of LORs that were measured at slightly different oblique angles. The total number of ring combinations contributing to a direct plane plus those contributing to a cross plane is sometimes referred to as span. Within the limitations imposed by the septa, span can be increased, resulting in increased sensitivity and degraded spatial resolution in the axial direction.

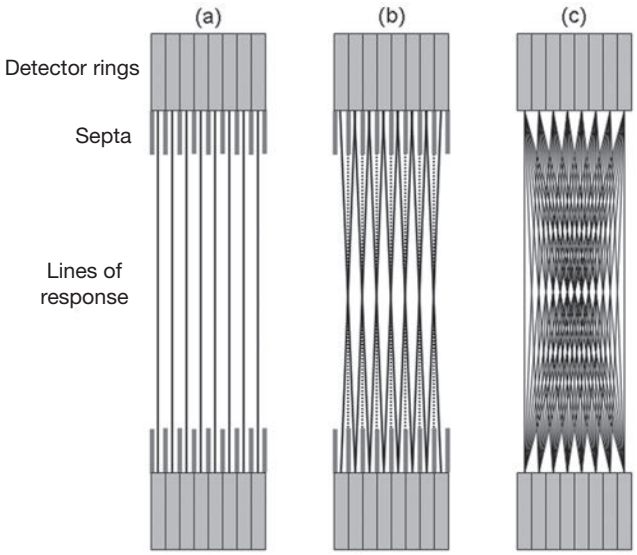

FIG. 11.38. Side elevation of an eight ring PET scanner in 2-D ((a) and (b)) and 3-D (c) acquisition modes. (a) Lines of response joining opposing detectors in the same ring forming direct planes. (b) Lines of response between detectors in adjacent rings. These lines of response are averaged to form cross planes (dotted line) that are assumed to be located at the mid-point between adjacent detectors. Both direct and cross planes are simultaneously acquired during 2-D acquisition. (c) 3-D acquisition in which each ring is permitted to form coincidence events with all other rings.

When the septa are removed, as is the case in 3-D acquisition mode, there is no longer any physical restriction on the detector rings that can be used to measure coincidence events. An N ring scanner could have a maximum ring difference of $N - 1$, resulting in up to N^2 possible sinograms. In 2-D mode, such a system would have a total of $2N - 1$ sinograms, so it can be seen that the total volume of data is substantially higher in 3-D mode. In order to reduce this volume for ease of manipulation, the maximum ring difference can be reduced. This has the effect of introducing a plateau on the axial sensitivity profile, converting it from a triangular to a trapezoidal form. Additionally, several possible ring combinations can be combined in a similar fashion to that indicated for 2-D acquisition mode. It should be noted that 3-D acquisition mode results in data that are redundant in the sense that only a subset of the sinograms (those in the transverse planes) are required for tomographic image reconstruction. The purpose of acquiring the additional oblique data is to increase sensitivity and reduce statistical noise in the resulting images.

In addition to the sinogram representation described above, some scanners also support list-mode acquisition. In this mode, coincidence events are not arranged into sinograms in real time but are recorded as a series of individual events. This stream of coincidence events is interspersed with time signals and potentially other signals from ECG or respiratory gating devices. These data can be used as the input to list-mode image reconstruction algorithms but may also be sorted into sinograms prior to image reconstruction. The advantage of acquiring in list-mode is that the sorting of the data into sinograms can be performed retrospectively. This provides a degree of flexibility that is very helpful when data are acquired in conjunction with physiological gating devices or when sequential images over time are of interest. For example, separate ECG gated and dynamic time series images can be obtained from the same cardiac list-mode acquisition. Furthermore, certain parameters can be retrospectively adjusted and do not have to match the parameters chosen at the time of acquisition.

11.3.4.4. Time of flight

Detectors operating in coincidence mode provide spatial information related to individual positron–electron annihilations but this information is not sufficient to determine the exact location of each event. A line joining the two detectors can be assumed to intersect the site of the annihilation but the exact position along this line cannot be determined. For this reason, PET systems measure signals from multiple events and the resulting projections are used to reconstruct images using computed tomography. However, it has long been appreciated that the difference in the detection times of the two annihilation photons provides a mechanism for precisely localizing the site of individual positron–electron

annihilations (Fig. 11.39). Given that photons travel at the speed of light, essentially irrespective of the composition of the material through which they pass, the difference in the arrival times of the two photons can potentially be used to localize their original point of emission. This is clearly attractive because it means that each coincidence measurement provides significantly more information, promising substantial improvements in image statistical quality.

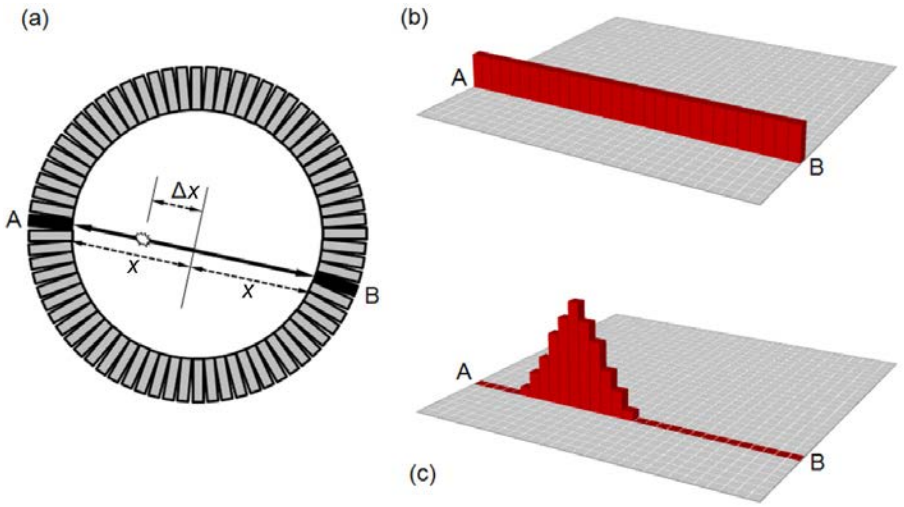

FIG. 11.39. (a) A coincidence event detected along a line of response between detectors A and B. The average time difference between the two detectors is given by $(x + \Delta x)/c - (x - \Delta x)/c = 2\Delta x/c$, *where c is the speed of light. (b) With conventional PET, no information is available about the location of the annihilation event along the line of response. During reconstruction, the event is assigned with equal weight to all pixels between A and B. (c) With time of flight PET, the time difference between the signals recorded at detectors A and B is used to estimate the position of the annihilation event along the line of response. During reconstruction, events are weighted according to the detector time difference and a function that reflects the limited time resolution of the system.*

Incorporating information derived from differences in the photon arrival times has been referred to as TOF mode and a number of PET systems have been developed that exploit this approach. A prerequisite for TOF PET systems is high timing resolution (Δt) as this determines the spatial uncertainty (Δx) with which the annihilation event can be localized. The two parameters are related by $\Delta x = c\Delta t/2$, where c is the speed of light (3×10^8 m/s). According to this equation, a timing resolution of 66 ps is required to achieve 1 cm depth resolution. Early TOF PET systems used detector materials that, although they had good timing resolution, suffered from a poor stopping power for 511 keV

photons. The resulting low sensitivity of these devices could not be offset by the improved signal to noise ratio provided by the TOF information and interest in the method declined. Interest was subsequently rekindled with the introduction of LSO based systems, which have been able to combine timing resolutions of around 600 ps with high sensitivity. A timing resolution of 600 ps translates to a spatial uncertainty of 9 cm which, although clearly worse than the spatial resolution that can be achieved with conventional PET, does represent useful additional information.

In addition to the high performance required for conventional PET, TOF PET requires scanners optimized for high timing resolution. The additional TOF information has data management considerations because an extra dimension has been added to the dataset. TOF data may be acquired in list-mode and fed directly to a list-mode reconstruction algorithm that is optimized for TOF. Alternatively, the data may be reorganized into sinograms where the sinograms have an additional dimension reflecting a discrete number of time bins. Each coincidence event is assigned to a particular sinogram depending on the difference in the arrival times of the two photons. TOF sinograms also require dedicated reconstruction algorithms that incorporate the TOF information into the image reconstruction. An interesting feature of TOF PET is that the signal to noise ratio gain provided by the TOF information is greater for larger diameter distributions of radioactivity. This is related to the fact that the spatial uncertainty Δx becomes relatively less significant as the diameter increases. This has potential benefits for body imaging, particularly in large patients where high attenuation and scatter mean that image quality is usually poorest.

11.3.5. Data corrections

11.3.5.1. Normalization

Normalization refers to a software correction that is applied to the measured projection data in order to compensate for variations in the sensitivity of different LORs. Without such a correction, images display systematic variations in uniformity and pronounced artefacts that include spike and ring artefacts at the centre of the FOV (Fig. 11.40). It is somewhat analogous to the uniformity correction applied to gamma camera images. Sources of sensitivity variations include:

— Detector efficiency variations: The detection efficiency of a particular LOR depends on the efficiencies of the individual detectors involved. Individual detectors can have variable efficiency due to differences in PMT

gain, differences in the performance of individual crystal elements and the position of the detector element within the larger detector array.

— Geometric effects: Geometric issues also influence the sensitivity of different LORs. Individual detector elements contribute to multiple LORs and measure photons that are incident over a range of angles. Photons incident normal to the face of the detector will have a shorter thickness of detector material in their path compared to those incident at more oblique angles. This results in a greater probability of detection for photons incident at more oblique angles. However, another geometric effect that exists in full ring block based systems occurs towards the edges of each projection where LORs are formed by detectors at larger oblique angles. Although oblique angles lead to a greater thickness of crystal, they also decrease the solid angle that the detectors present, reducing the sensitivity for those particular LORs.

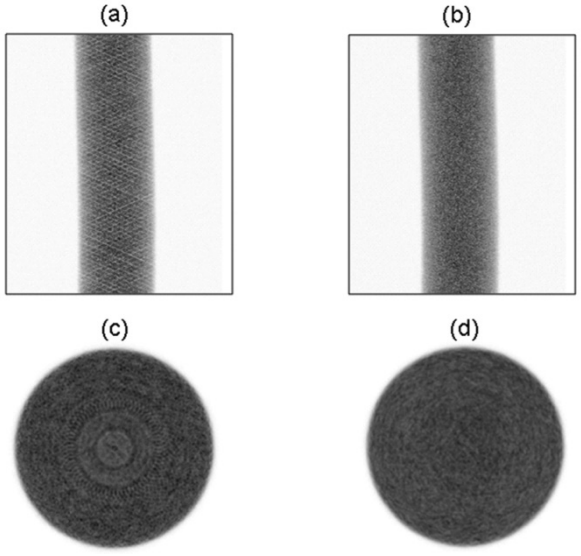

FIG. 11.40. Sinograms corresponding to a centrally located uniform cylinder before normalization (a) and after normalization (b). Transverse images are shown for reconstructions without normalization (c) and with normalization (d). The artefacts and non-uniformity seen in the image without normalization should be noted.

The acquisition geometry, either 2-D or 3-D mode, determines the way different detector pairs are combined and, thus, also influences sensitivity variations between different LORs. In addition, the presence of the interplane septa has a shadowing effect that reduces sensitivity. For this reason, separate

2-D and 3-D normalizations are required for systems capable of acquiring data in both modes. Normalization files are experimentally determined correction factors that are applied as multiplicative terms for each LOR. They are periodically updated to reflect the current state of the detector system and are applied to all subsequently acquired data. In order not to degrade the statistical quality of the patient data, normalization coefficients need to be measured with low statistical noise, and a variety of methods have been developed to achieve this.

11.3.5.2. Randoms correction

Randoms make up a potentially large component of all measured coincidence events and, if left uncorrected, will contribute to a loss of image contrast and quantitative accuracy. Random coincidences are generally smoothly distributed across the FOV but the magnitude of the randoms component depends on the count rate encountered during data acquisition. 3-D acquisition mode or studies involving large amounts of activity in or near the FOV are usually associated with high randoms fractions. Randoms correction is essential for all quantitative studies and is routinely implemented on almost all scanner systems.

One widely adopted correction method involves estimating the number of randoms contributing to the prompts (trues + scatter + randoms) using an additional coincidence circuit. This secondary coincidence circuit is acquired simultaneously with the prompt measurement but is only sensitive to random events. Preferential selection of randoms is achieved by delaying the logic pulse from one of the detectors such that it cannot form a true coincidence event between corresponding annihilation photons (Fig. 11.41). Although the time delay prevents measurement of true and scattered photons, it does not stop coincidence events being recorded by chance between unrelated 511 keV photons. As this secondary, or delayed, coincidence circuit is identical to the prompt circuit in all other respects, the number of counts in the delayed channel provides an estimate of the randoms in the prompt channel. In many implementations, the delayed data are automatically subtracted from the corresponding prompt data, providing an on-line randoms correction. Alternatively, the prompt and delayed data can be stored as separate sinograms for retrospective off-line subtraction, often as part of a statistical image reconstruction algorithm.

The delayed channel does not identify and remove individual random coincidences from the prompt measurement but, instead, estimates the average number of randoms that might be expected. As the delayed channel records counts in individual LORs over a limited scan duration, the randoms estimate is often noisy, leading to increased statistical uncertainty in the corrected data after randoms subtraction.

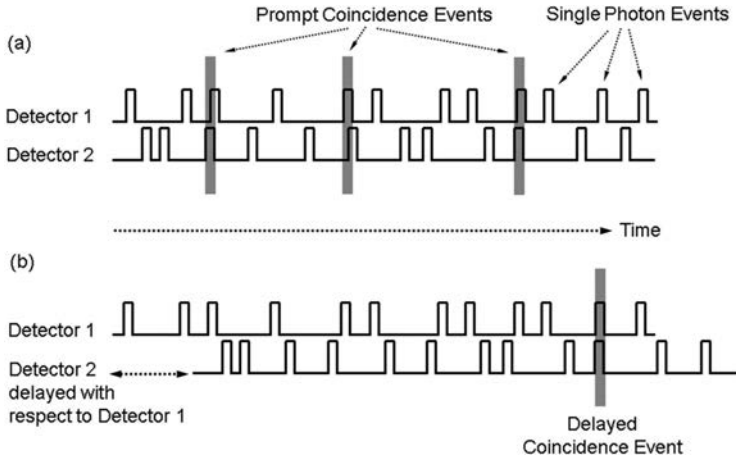

FIG. 11.41. Diagram illustrating the concept of how a delayed coincidence circuit can be used to estimate the number of random events in the prompt circuit. (a) Detection events from two opposing detectors, indicating three coincidence events in the prompt circuit. (b) Data from detector 2 delayed with respect to detector 1 and indicating one coincidence event in this delayed circuit. The temporal delay prevents true coincidence events from being recorded in the delayed circuit, but random coincidence events still occur with the same frequency as in the prompt circuit. If data are acquired with sufficient statistical quality, the total number of delayed coincidence events provides an estimate of the total number of randoms in the prompt circuit.

An alternative method of randoms correction is to estimate the randoms for each LOR using the singles rates at the corresponding detectors. If N_1 and N_2 are the singles rates at two opposing detectors, the rate of random coincidences between these detectors is given by $2\tau N_1 N_2$ where 2τ is the coincidence timing window. This approach has the advantage that singles count rates are substantially higher than coincidence count rates and, therefore, lead to randoms estimates with better statistical quality than the delayed channel method.

11.3.5.3. Attenuation correction

Despite their high energy, only a small fraction of the emitted 511 keV photon pairs escape the body without undergoing some form of interaction. Compton interaction is the most likely mechanism and, depending on the energy and direction of the scattered photon, may result in the detection of a scattered coincidence event. However, it is more likely that the scattered photon will not result in a coincidence event for a variety of reasons. The scattered photon may emerge from the Compton interaction along a path that is not incident upon the detectors. Alternatively, the scattered photon may undergo further Compton

interactions, resulting in a lower energy and a greater likelihood of photoelectric absorption. Even if the scattered photon does reach the detectors, it may have lost so much energy that it does not meet the energy acceptance criteria of the scanner and will be rejected. Thus, as well as creating scattered coincidence events, Compton interactions lead to a much greater loss of true coincidence events. This underestimation of true counts is referred to as attenuation.

One of the advantages of PET over SPECT is the ease with which attenuation correction can be performed. This feature of coincidence detection can be understood by considering a point source located at depth x within a uniformly attenuating object with attenuation coefficient μ. Figure 11.42(a) shows an LOR passing through the point source and intersecting a thickness D of attenuating material. The probability of photon 1 escaping the object without undergoing any interactions p_1 is given by:

$$p_1 = \frac{I(x)}{I(0)} = \mathrm{e}^{-\mu x} \tag{11.13}$$

where

$I(x)$ is the beam intensity after passing through attenuating material of thickness x;

and $I(0)$ is the intensity in the absence of attenuation.

The probability that the corresponding photon 2 will also escape the object is given by p_2:

$$p_2 = \frac{I(D-x)}{I(0)} = \mathrm{e}^{-\mu(D-x)} \tag{11.14}$$

The probability of both photons escaping the body such that a coincidence event can occur is given by the product of p_1 and p_2:

$$p_1 \times p_2 = \mathrm{e}^{-\mu x} \times \mathrm{e}^{-\mu(D-x)} = \mathrm{e}^{-\mu D} \tag{11.15}$$

It can be seen that the probability of a coincidence event occurring decreases as the thickness of attenuating material increases. However, this probability is not dependent on the location of the source along a particular LOR. This result differs from the SPECT case where the attenuation experienced by a single photon source along a particular LOR is strongly dependent on the distance between the source location and the edge of the attenuating medium.

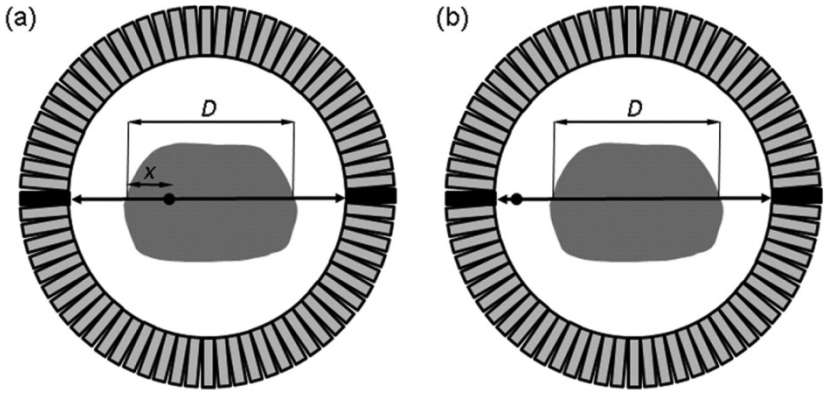

FIG. 11.42. When considering both back to back photons along a particular line of response, the attenuation experienced by a point source within the body (a) is independent of its location along the line and is given by $e^{-\mu D}$ for a uniformly attenuating object. In (b), the positron emitting transmission source is outside the patient but experiences the same attenuation as the internal source when considering coincidence events along the same line of response.

From Eq. (11.15), it can be seen that photon attenuation reduces the number of true coincidence events by a factor of $e^{-\mu D}$. In practice, the body is not composed of uniformly attenuating material, and the attenuation factor AF for a particular LOR is given by:

$$\mathrm{AF} = \mathrm{e}^{-\int \mu(x)\,\mathrm{d}x} \tag{11.16}$$

where $\mu(x)$ refers to the spatially-variant distribution of linear attenuation coefficients within the body and the integral is over the LOR joining opposing detectors.

As this attenuation factor is dependent only on the material along the LOR, it is identical to the attenuation experienced by an external source of 511 keV photons along the same line (Fig. 11.42(b)). Consider an external source of 511 keV photons that gives rise to a total of N_0 coincidence events along a particular LOR in the absence of any attenuation. When a patient is interposed, the number of coincidence events falls to N_x where:

$$N_x = N_0 \mathrm{e}^{-\int \mu(x)\,\mathrm{d}x} \tag{11.17}$$

From this equation, it can be seen that the attenuation factor can be obtained by dividing N_x by N_0. The attenuation correction factor is simply the reciprocal of the attenuation factor and is given by N_0/N_x. N_0 can be obtained from what

is referred to as a blank scan, as it is acquired with nothing in the FOV. N_x can be obtained from what is referred to as a transmission scan, as it measures the coincidence events corresponding to the photons that pass through the patient while in the scanning position.

A variety of transmission systems have been developed for measuring patient specific attenuation correction factors. Early transmission systems used ^{68}Ge (271 d half-life, decays to the positron emitter ^{68}Ga) ring sources, although poor scatter rejection meant that the ring configuration was superseded by rotating rod sources. These sources, also consisting of ^{68}Ge, were oriented parallel to the z axis of the scanner and could be inserted close to the detectors at the periphery of the PET FOV during a transmission acquisition. Once in position, they would rotate around the patient in a similar fashion to the motion of an X ray tube in CT. Coincidence events were recorded in 2-D acquisition mode for all LORs as the source rotated continuously around the patient. The patient transmission data were used in conjunction with a similarly acquired blank scan to determine an attenuation correction sinogram that was then applied to the patient emission data, under the assumption that the patient did not move between scans. The rotating rod configuration had the advantage that it enabled transmission data to be acquired in conjunction with a spatial 'window' that tracked the current position of the rotating source. Coincidence events that were not collinear with the current position of the source, such as scattered coincidences, could be rejected, improving the quality of the transmission data. Rod windowing also helped reduce contamination of the transmission measurement by coincidence events not originating from the rotating sources but from a radiopharmaceutical within the patient. The ability to acquire transmission data in the presence of a positron emitting tracer within the body was of great practical significance as, without this capability, lengthy protocols were required involving transmission acquisition prior to tracer administration.

Rod windowing also provided the potential for simultaneous acquisition of emission and transmission data, although cross-contamination meant that separate emission and transmission acquisitions were usually preferred. The disadvantage of acquiring emission and transmission data in a sequential fashion was that scan times for both modes were necessarily lengthy in order to obtain data with sufficient statistical quality. Increasing the rod source activity was not an effective way of reducing transmission scan times as the source was located close to the detectors and dead time at the near-side detectors quickly became a limiting factor. As an alternative to ^{68}Ge, single photon emitters such as ^{137}Cs (30 a half-life, 662 keV) were used as a transmission source. Single photon transmission sources had the advantage that they could be shielded from the near-side detectors and could, thus, use much greater amounts of activity, leading to data with improved statistical quality and, in practice, shorter scan times.

The fact that the single photon emissions were at 662 keV as opposed to 511 keV and that the transmission data had a large scatter component was problematic but could be effectively suppressed using software segmentation algorithms.

With the introduction of PET/CT, the need for radionuclide transmission systems was eliminated as, with careful manipulation, the CT images can be used not just for anatomic localization but also for attenuation correction. CT based attenuation correction has a number of advantages, including the fact that the resulting attenuation correction factors have very low noise due to the high statistical quality of CT; rapid data acquisition, especially with high performance multi-detector CT systems; insensitivity to radioactivity within the body; and no requirement for periodic replacement of sources as is the case with ^{68}Ge based transmission systems. CT based attenuation correction is significantly different from radionuclide based transmission methods because the data are acquired on a separate, albeit well integrated, scanner system using X ray photons with energies that are very different from the 511 keV photons used in PET. Unlike monoenergetic PET photons, the photons used in CT consist of a spectrum of energies with a maximum value that is dependent on the peak X ray tube voltage (kVp). CT Hounsfield units reflect tissue linear attenuation coefficients that are higher than those applicable to PET photons as they are measured at an effective CT energy (~70 keV), which is substantially lower than 511 keV. An important step in the process of using CT images for attenuation correction is to scale the CT images to linear attenuation coefficients that are applicable to 511 keV photons. A number of slightly different approaches have been employed but usually involve multi-linear scaling of the CT Hounsfield units using functions specific for the X ray tube kVp setting (Fig. 11.43). The methods used are very similar to those described in Section 11.2.3.2 (Eq. (11.10)) for SPECT. After scaling, the CT images are filtered, so as to have a spatial resolution that is similar to that of the PET data and attenuation factors are calculated by integration in a manner indicated by Eq. (11.16). The integration, or forward projection, is performed over all directions measured by the PET system and, thus, provides attenuation correction factors for all LORs.

CT based attenuation correction has proved to be very effective although a number of potential problems require consideration. Patient motion, commonly motion of the arms or head, can cause the CT and PET images to be misregistered, leading to incorrect attenuation correction factors, which in turn cause image artefacts and quantitative error. Respiratory motion can also lead to similar problems as the CT and PET data are acquired over very different time intervals. CT data acquisition is extremely short and usually captures a particular phase in the respiratory cycle. In contrast, PET data are acquired over multiple breathing cycles and the resulting images represent an average position that will be somewhat blurred in regions where respiratory motion is significant.

In the area around the lung boundary where there is a sharp discontinuity in the body's attenuation properties, respiratory motion can lead to localized misregistration of the CT and PET images, and pronounced attenuation correction artefacts. Another consideration for CT based attenuation correction arises when the CT FOV is truncated such that parts of the body, usually the arms, are not captured on the CT image or are only partially included. This leads to under-correction for attenuation and corresponding artefacts in the PET images.

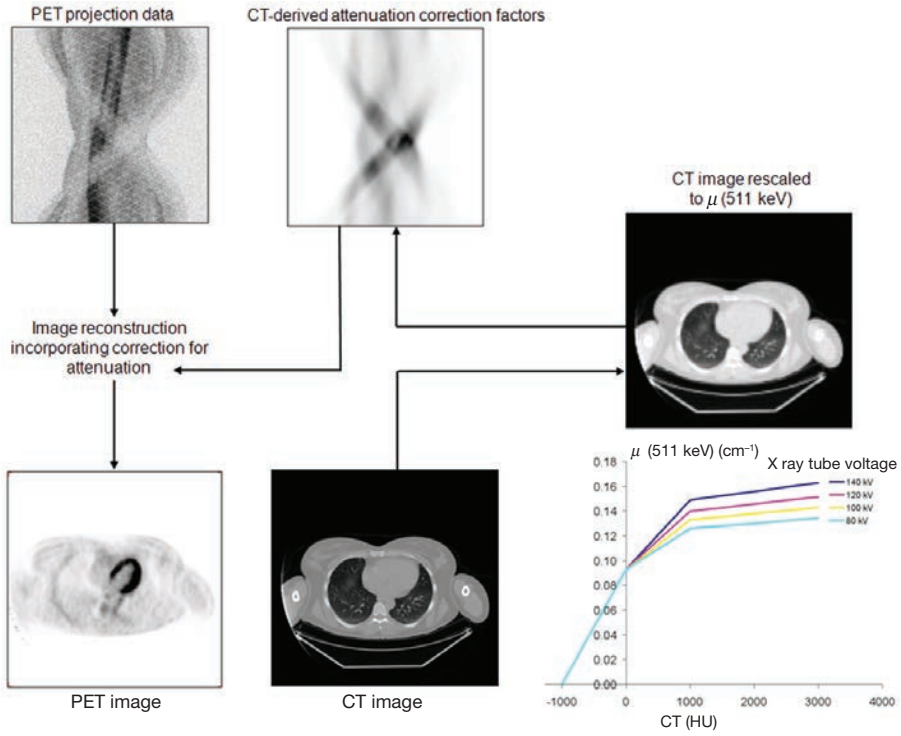

FIG. 11.43. For PET attenuation correction, CT images have to be rescaled from Hounsfield units (HU) to linear attenuation coefficients (μ) appropriate for 511 keV. Multi-linear scaling functions have been used. The rescaled CT images are then forward projected to produce attenuation correction factors that are applied to the PET emission data, prior to or during image reconstruction.

11.3.5.4. Scatter correction

Scatter correction is required because the limited energy resolution of PET systems means that scattered photons can only be partially rejected by

energy discrimination. Uncorrected scatter forms a background in reconstructed images that reduces lesion contrast and degrades quantitative accuracy. This scatter background is a complex function of both the emission and attenuation distributions and is non-uniform across the FOV. In 2-D mode, physical collimation ensures that the scatter contribution is relatively low compared to 3-D mode and approximate corrections, based on scatter deconvolution, have been widely used. The form of the scatter distribution function can be measured experimentally using line sources at different positions in a water phantom. Analytical expressions derived from this scatter function can be determined and convolved with the projection data from individual patient studies to estimate the scatter distribution. This method assumes a uniform scattering medium and has limited accuracy in areas such as the thorax. It also cannot account for scatter between adjacent planes, which is significant for 3-D acquisition mode.

An alternative algorithm that has been applied to 3-D brain studies involves a tail fitting approach. In brain studies, the LORs that do not pass through the head are comprised of scatter that can be modelled by fitting a Gaussian function to the tails of each projection. This function can be interpolated to the projections that do pass through the head and used as an estimate of the scatter contribution along these LORs. This method provides a first order correction for scatter and has limited accuracy in areas of non-uniform attenuation or any area where the tails of the projections cannot be accurately measured.

More accurate scatter correction can be achieved in 3-D using a model based approach. This method makes use of the physics of Compton scattering to model the distribution of coincidence events for which one of the two photons experienced a single scattering interaction. The assumption that the scatter component in the measured data is dominated by these single scatter events has been shown to be reasonable. An initial estimate of the radionuclide distribution is first obtained from a reconstruction that does not include any scatter correction. This image is then used, in conjunction with an attenuation map derived from CT or transmission data, to estimate the scatter contribution to each LOR. Different implementations have been developed, but each makes use of the Klein–Nishina formula to determine the probability of a photon scattering through a certain angle and being detected by a certain detector. Determining these probabilities for all possible scattering positions and LORs is computationally demanding. However, good scatter estimates can be obtained by interpolating data estimated using a coarse grid of scattering locations and a subset of LORs. These data can then be interpolated for all LORs, resulting in an estimate of the scatter distribution that has been found to be highly accurate over a range of anatomical locations.

11.3.5.5. Dead time correction

Detector count rates vary between patients and can also vary within studies performed on the same patient. In order to achieve quantitatively accurate images, the true count rate should ideally increase linearly with increasing activity in the FOV. Although this is usually the case at low count rates, the PET scanner's response becomes increasingly non-linear at high count rates. Individual detector modules within a scanner require a finite period of time to process each detected photon. If a second photon is incident upon a detector while an earlier photon is still being processed, the secondary photon may be lost. The likelihood of this occurring increases at high count rates and results in an effective loss of sensitivity. This kind of count loss is referred to as dead time.

Detector dead time losses occur mostly in the detector front-end electronics. The signal produced by an individual photon is integrated for a fixed period of time in order to determine the position and energy of the incoming event. If a second photon is recorded during this integration time, the two signals will combine (pulse pile-up) in such a way as to become indistinguishable from each other. The resulting combined signal will be rejected if it exceeds the upper level energy discriminator. Alternatively, if it is recorded within the energy acceptance window, it may be treated as a single event with a position somewhere between the locations of the two individual photons. In this case, pulse pile-up contributes to a loss of spatial resolution as well as a loss of true counts. Other sources of dead time arise during coincidence event processing. When more than two events occur within the coincidence time window, it is impossible to determine the correct coincidence pair. In this circumstance, all detection events may be discarded, contributing to dead time losses, or alternatively, all possible coincidence events can be included, increasing the randoms component.

Dead time correction compensates for this loss of sensitivity. Corrections are usually based upon experimental measurements of the scanner's response to a decaying source of activity. After randoms correction, residual non-linearity in the scanner's response can be attributed to dead time. An analytical model of the scanner's count rate response can be determined from these experimental data and used for dead time correction of subsequent patient data. A global correction factor can be applied for a particular acquisition, assuming that dead time effects are similar for all detectors in the ring. Alternatively, different corrections can be applied to each detector block or group of blocks. Corrections can be determined based upon an estimate of the fraction of the acquisition period that each detector was busy processing events and unable to process other photons. Alternatively, the single photon rate at a particular detector can be used as input for a model of the scanner's dead time performance to estimate the magnitude of the dead time effect. Dead time correction only compensates for count losses and does

not compensate for the event mis-positioning that can occur as a result of pulse pile-up.

11.3.5.6. Image calibration

The above corrections substantially eliminate the image artefacts and quantitative errors caused by the various physical effects that degrade PET data. As a result, the reconstructed images reflect the activity distribution within the FOV, within the limitations imposed by the system's limited spatial resolution. Furthermore, these reconstructed images can be used to quantify the in vivo activity concentration in a particular organ or tissue. Although this capability is not always fully exploited, the potential to accurately quantify images in terms of absolute activity concentration facilitates a range of potential applications.

After image reconstruction, including the application of the various physical corrections, PET images have arbitrary units, typically counts per voxel per second. Quantitative data can be extracted from the relevant parts of the image using region of interest techniques but cannot be readily compared with other related data such as measurements made with a radioactivity calibrator ('dose' calibrator). In order to convert the PET images into units of absolute activity concentration such as becquerels per millilitre, a calibration factor is required. This calibration factor is experimentally determined, usually using a uniform cylinder phantom. The cylinder is filled with a known volume of water, to which a known amount of radioactivity is added. After ensuring the radioactivity is uniformly distributed within the phantom, a fully corrected PET image is acquired. The calibration factor CF can be determined using:

$$\mathrm{CF} = \frac{\mathcal{A}}{V} \times \frac{p}{C} \tag{11.18}$$

where

$\mathcal{A}/V$ is the known activity concentration (Bq/mL) within the phantom;
C is the mean voxel data (counts · voxel^{-1} · s^{-1}) from a large region well within the cylinder part of the image;

and p is the positron fraction of the radionuclide used in the calibration experiment (typically ^{18}F, positron fraction 0.97).

The positron fraction is a property of the radionuclide and is the fraction of all disintegrations that give rise to the emission of a positron.

The above calibration assumes that the true activity within the phantom is accurately known. This can usually be achieved to an acceptable level of tolerance using an activity calibrator that has been calibrated for the isotope of interest using a long lived standard source that is traceable to a national metrology institute. In principle, a single calibration factor can be applied to subsequent studies performed with different isotopes as long as the positron fraction is known. Calibrated PET images can, thus, be determined by multiplying the raw image data by the calibration factor and dividing by the positron fraction for the particular isotope of interest.

11.4. SPECT/CT AND PET/CT SYSTEMS

11.4.1. CT uses in emission tomography

SPECT and PET typically provide very little anatomical information, making it difficult to precisely localize regions of abnormal tracer accumulation, particularly in oncology studies where disease can be widely disseminated. Indeed, it is often the case that the more specific the radiopharmaceutical, the less anatomical information is available to aid orientation. Relating radionuclide uptake to high resolution anatomic imaging (CT or MRI (magnetic resonance imaging)) greatly aids localization and characterization of disease but ideally requires the two images to be spatially registered. Retrospective software registration of images acquired separately on different scanner systems has proved to be effective in certain applications, notably for brain studies where rigid body assumptions are realistic. However, for most other applications, the rigid body assumption breaks down and the registration problem becomes much more difficult. Combined scanner systems, such as SPECT/CT and PET/CT, provide an alternative solution. The advantage of this hardware approach is that images from the two modalities are inherently registered with no need for further manipulation. Of course, this assumption can become unreliable if the patient moves during data acquisition, but, in general, combined scanner systems provide an accurate and convenient method for achieving image registration.

In addition to the substantial clinical benefit of registered anatomical and functional images, the coupling of CT with SPECT and PET systems provides an additional technical benefit. Although radionuclide sources have been used for attenuation correction, the availability of co-registered CT is particularly advantageous for this purpose. In the case of SPECT, the main advantages of CT based attenuation correction are greater accuracy and reliability compared to radionuclide sources, while in PET, the main advantage is an effective reduction in the overall duration of the scanning procedure owing to the speed with which

CT images can be acquired. In addition, there are a number of other subsidiary benefits to the introduction of CT to SPECT and PET systems. Radionuclide transmission sources and their associated motion mechanisms are somewhat cumbersome, particularly in the case of SPECT, and the addition of the CT allows this component to be removed from the design. Elimination of the transmission source from PET systems enabled the patient port size to be enlarged, allowing larger patients to be accommodated. Other benefits include use of the CT subsystem's localizing projection image to aid patient positioning, particularly for single bed-position PET scans. In addition, the potential of acquiring both a radionuclide study and diagnostic quality CT in the same scanning session has advantages in terms of convenience. In terms of quantitative image analysis, the availability of registered CT along with a radionuclide study can sometimes be useful for region of interest definition, particularly in research applications. Although giving rise to a number of significant benefits, it should be noted that the addition of CT to SPECT and PET instrumentation has led to an appreciable increase in the radiation dose received by patients undergoing radionuclide imaging procedures.

11.4.2. SPECT/CT

In many respects, SPECT might be expected to benefit more than PET from the addition of registered CT. SPECT has lower spatial resolution than PET. Many SPECT tracers are quite specific and often do not offer the kind of anatomical orientation that is provided by normal organ uptake with PET tracers such as fluorodeoxyglucose (FDG). Radionuclide transmission scanning is more awkward in SPECT compared to PET because gamma cameras need to be capable of multiple flexible modes of acquisition. Despite these considerations, the adoption of combined SPECT/CT instrumentation (Fig. 11.44) has been slower than that of PET/CT. Cost considerations no doubt contribute and there remains uncertainty about the level of CT performance that is required for a SPECT/CT system.

FIG. 11.44. Clinical SPECT/CT systems.

Early SPECT/CT designs, including successful commercial offerings, coupled SPECT with low performance CT. The SPECT subsystem was placed in front of the CT and data were acquired sequentially with the patient being translated between gantries by a common patient bed. Sequential as opposed to simultaneous data acquisition was required due to cross-talk between the CT and SPECT subsystems and was quickly established as the standard mode of operation for both SPECT/CT and PET/CT. Low power X ray tubes and slow rotation speeds meant that the CT component was by no means optimized for diagnostic quality imaging. The aim was to provide attenuation correction and a much needed anatomical context for the SPECT, while maintaining a relatively low cost.

The desire for improved CT image quality has led to the introduction of SPECT/CT systems that incorporate a high performance CT component with capabilities comparable to dedicated CT scanners. With this development, SPECT/CT now benefits from substantially improved CT image quality, faster data acquisition and a broader range of CT protocols. A number of different multi-detector CT slice configurations are available, as well as alternative designs including those based upon flat panel detectors and improvements of the original low cost, non-diagnostic CT. This broad range of CT capability may reflect a diversity of opinion about the role of combined SPECT/CT in clinical practice.

11.4.3. PET/CT

Concurrent technical developments in both PET and CT were exploited in the development of the combined PET/CT system [11.4]. In PET, new detector materials and approaches to image reconstruction increased spatial resolution and improved image statistical quality. In CT, the introduction of spiral scanning and multi-detector technology enabled extremely fast data acquisition over an extended FOV. These developments meant that the combined PET/CT system was well positioned to take advantage of the growing evidence that PET imaging with FDG could play a valuable role in oncology (Fig. 11.45). The addition of spatially registered anatomical information from the CT component of the combined scanner provided the impetus for widespread acceptance of PET in oncology and has driven the rapid growth of PET/CT instrumentation. PET/CT has now been rapidly accepted by the medical community, so much so that stand-alone PET systems are no longer being developed by the major commercial vendors.

Early PET/CT systems were not closely integrated and consisted of separately developed PET and CT components that operated independently of each other. Given that PET/CT acquisitions occur sequentially, as opposed to simultaneously, this approach was reasonable but it meant that multiple computer

systems were required and the user interface could be awkward. In addition, the availability of the CT for attenuation correction meant that the PET transmission scanning system was somewhat redundant. Subsequent designs removed the PET transmission sources and moved the two subsystems towards greater integration. In some cases, this meant a more compact system with a continuous patient tunnel. In other cases, the PET and CT gantries were separated by a gap which allowed greater access to the patient during scanning. Removing the transmission scanning system also provided scope for increasing the size of the patient port, so as to accommodate larger patients. This was further achieved by removing the septa from the PET subsystem and decreasing the size of the end-shields used to reject out of FOV radiation. Although the PET and CT detectors remain separate subsystems, many software functions of a modern PET/CT system run on a common platform, including a common patient database containing both PET and CT data.

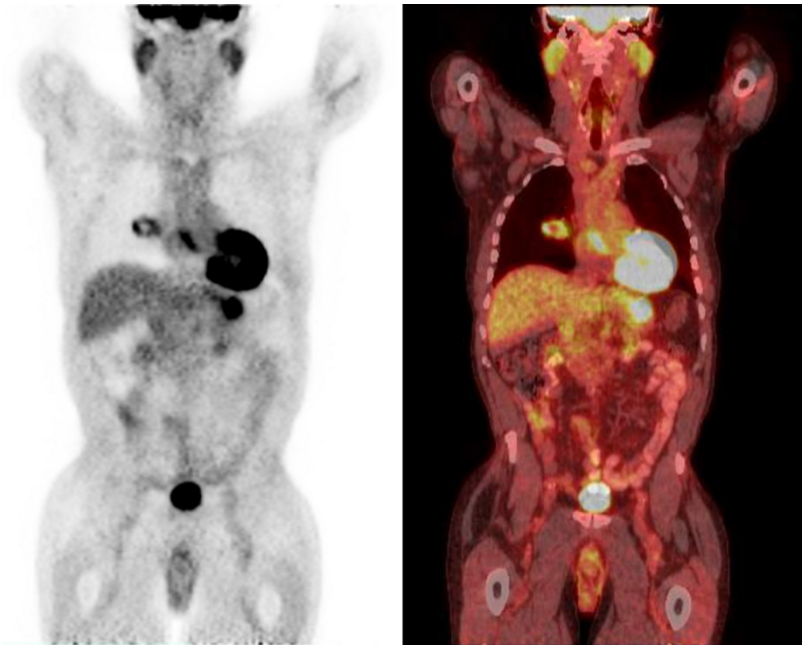

FIG. 11.45. Coronal images from a whole body fluorodeoxyglucose PET/CT study. (a) The PET data are shown in inverse grey scale. (b) The same PET data are shown in a colour scale, superimposed on the CT in grey scale.

The CT component of combined PET/CT systems continues to evolve as greater multi-detector capability becomes available. The initial PET/CT design

involved single slice CT, whereas later systems have incorporated 4, 16, 64 or greater slice CT. Advanced CT detector capability allows for extended volume coverage in a single breath-hold or contrast phase and also facilitates a range of rapid diagnostic CT protocols, particularly those intended for cardiology. The level of CT capability required by a combined PET/CT system depends on the extent to which the system will be used for diagnostic quality CT. For many PET/CT oncology studies, state of the art CT is not necessary and low dose protocols are favoured. In these cases, the X ray tube current is reduced and intravenous contrast would typically not be administered. In addition, for whole body studies, data would not be acquired under breath-hold conditions in order to improve spatial alignment of the CT with the PET data that are acquired over multiple respiratory phases.

As far as CT based attenuation correction is concerned, the main advantage for PET is not so much improved accuracy, as ^{68}Ge transmission sources are perfectly adequate when obtained with sufficient statistical quality. The main advantage of CT based attenuation correction is the speed with which the data can be acquired. This is particularly important for whole body PET studies where transmission data are needed at each bed position, requiring extended scan durations to achieve adequate statistical quality. With multi-detector CT scanners, low noise images can be acquired over the whole body in only a few seconds. Replacing radionuclide transmission scans with CT had the effect of substantially reducing the overall duration of scanning procedures, particularly those requiring extended axial coverage. In turn, shorter scans increase patient comfort and reduce the likelihood of motion problems. With the shorter scan times, ‘arms up’ acquisition is better tolerated by patients, leading to reduced attenuation and improved image statistical quality. In addition to these methodological considerations, shorter scan durations allow for more patient studies in a given time period and, thus, more efficient utilization of the equipment.

Despite the relatively rapid scanning afforded by modern PET/CT systems, the PET component still typically requires many minutes of data acquisition and management of patient motion continues to be a problem. Motion can potentially cause misalignment of the PET and CT images, degradation of image spatial resolution and the introduction of characteristic artefacts. Although the CT images are acquired independently of the PET, the PET images involve CT based attenuation correction and are, thus, particularly susceptible to patient motion between the two scans. The types of motion usually encountered include gross movement, often of the arms or head, and respiratory motion. The former is hard to correct retrospectively, although external monitoring devices such as camera based systems can potentially provide solutions for head motion. External monitoring devices can also help reduce problems due to respiratory motion. Respiratory gated PET can reduce motion blurring by reconstructing images

from data acquired only during particular phases of the respiratory cycle. This can be further refined by including a motion model into the PET reconstruction and incorporating CT data also acquired under respiratory control.

REFERENCES

[11.1] ANGER, H.O., Scintillation camera, Rev. Sci. Instrum. **29** (1958) 27–33.

[11.2] CHERRY, S.R., SORENSON, J.A., PHELPS, M.E., Physics in Nuclear Medicine, Saunders, Philadelphia, PA (2003).

[11.3] BAILEY, D.L., TOWNSEND, D.W., VALK, P.E., MAISEY, M.N., Positron Emission Tomography: Basic Sciences, Springer, London (2005).

[11.4] BEYER, T., et al., A combined PET/CT scanner for clinical oncology, J. Nucl. Med. **41** (2000) 1369–1379.