Tomography

An X-ray or a nuclear medicine scan (each sometimes referred to as planar scans) suffers from the loss of information inherent in making a two-dimensional representation of a three-dimensional object. All that you can learn from either is the combined effect of attenuation (and radioisotope distribution for NM) all along a path through the patient. Finding the particular arrangement of attenuating material and/or radioactive material along that line is impossible with a single picture. This problem is familiar to anyone who has had to wear an eye patch – if you cover one eye and hold your head still, you lose your depth perception. The effect is also used to generate 3D movies: two cameras are placed side by side (so that their lenses are about as widely separated as your eyes) to record a scene from slightly different angles, and the resulting films are each directed to a different eye using either colored filters or polarizing filters. Basically, a planar X-ray of a patient with a bullet resting on his chest would look the same as an X-ray of a patient with a bullet in his chest.

The need for a more accurate determination of the three-dimensional structure of the body led to the invention of tomography, or imaging by slices. In practice, the patient is surrounded by a ring of detectors or (more cheaply) a gamma camera is moved around the patient. By viewing the body from many different angles, we can reconstruct the data to get an understanding of the full picture. When this is done with X-rays, it’s called a CT, or computed tomography scan (formerly computed axial tomography, or CAT scan). If it is instead done with a gamma camera (or multiple cameras), it is known as a Single Photon Emission Computed Tomography (SPECT) scan. Positron decays are used in the Positron Emission Tomography (PET) scanner. We’ll look at emission tomography shortly, but for now, let’s concentrate on CT.

What are we really measuring when we take an X-ray? It’s just the radiation received at a point. When we’re using regular film, realize that we have nothing but the density of silver grains at a point (a single number). Ideally, that represents (only) the number of X-ray photons that traveled in a straight line from the X-ray source through the patient and onto that spot of the film. We know that there will also be some photons that don’t follow that path but get scattered to land on the tiny region of film we’re examining, but we have no way of separating those from the “good” photons.

If we look at a two-dimensional slice of this process, it gets even simpler. Imagine that we’re taking a chest X-ray of a person (the person’s torso will be shaped like a cylinder to make this easier). In the drawing below, we’re looking at a horizontal plane of the process (parallel to the floor the patient is standing on). In that plane, the cylindrical chest is just a circle and the rectangular X-ray film is just a line.
The edges of the film should be well exposed since there was no tissue between them and the source. As we move in from the edges towards the center, there should be less exposure since more of the patient was in the way to block the beam. If we were to plot the degree of exposure along the X-ray film, it might look something like the situation below:

In a more complicated case, let’s assume we have something that is very efficient at blocking X-rays stuck in the patient (like a piece of lead)
We could tell from this change in the X-ray that there is something blocking radiation somewhere along the dotted red line through the patient. The only way we could find out where exactly (which would be important if you’re operating!) is to take more X-rays from different angles, like below:

In the unrealistically simple case we have above (symmetric patient with uniform attenuation everywhere except at the location of the bullet), two views would let us find the bullet. For a more realistic human, where there are many regions of differing attenuation (fat, muscle, bone, air-filled lungs, etc.) that aren’t circularly symmetric, it takes more angles to get a complete picture.

How are these 2-D X-rays (1-D lines in our example above) used to reconstruct the true 3-D (2-D in our example) picture of the patient’s body? We can use something known as the **Fourier slice theorem**. Before doing that, though, we need to investigate the **Fourier transform** itself.

**Fourier Analysis**

There are two equivalent ways to describe any waveform. We can talk about it as the amplitude of something (speaker position, electric field, location of molecules in rocks during an earthquake, etc.) as time changes (the **time domain**) or we can think of it as a collection of sine or cosine waves of different amplitudes and different frequencies (the **frequency domain**). If we look for the connection between these two ideas, the Fourier transform takes us from the time domain to the frequency domain, and the **inverse Fourier transform** goes in the other direction.

If we were to plot the position of one of the tines in a recently-struck tuning fork as a function of time, we would get a sine wave. Because of the uncertainty principle, we could only be completely certain of the frequency of this wave if we had an infinite number of cycles to count. To understand this, imagine that you’re taking your pulse while exercising. If you count it for 10 seconds, you know that you need to multiply your
answer by six to get your true pulse. You can also quickly see that you will never get an answer like 63 beats per minute (since that's not a multiple of six). If you wanted to narrow this uncertainty in the pulse down to something smaller, you could count for thirty seconds and double your answer. Of course, you're limited to even numbers for the pulse in this case. You could extend it to one minute and get to the nearest whole number, but what if your heart doesn't happen to beat exactly x times in a minute? You need to count for two minutes and then divide by two, etc. You can see that reducing the uncertainty in your pulse frequency to zero would mean counting for an infinite time. You can see the two plots below (time domain on the left, frequency domain on the right).

If a signal is narrow in one domain, it must be wide in the other domain, just as a particle with a small uncertainty in position has to have a large uncertainty in momentum. In a similar way, if you want to produce an instantaneous displacement of a speaker cone (or your eardrum), a huge range of frequencies will really be present. The graphs below are constructed from plotting a simple function from $-\pi$ to $\pi$. It can be written as

$$f(x, j) = \sum_{i=1}^{j} \cos(i \cdot x)$$

For example, $f(x, 1)$ is just $\cos(x)$, $f(x, 2) = \cos(x) + \cos(2x)$, $f(x, 3) = \cos(x) + \cos(2x) + \cos(3x)$, etc. In the frequency domain, this function will just look like $j$ spikes. What do we get in the time domain? Look below.
As you can see, adding more frequencies (making the signal broad in the frequency domain) bunches it up in the time domain. To get a perfect spike, which is infinitely tall and infinitesimally wide (called a delta function), you would add an infinity of these frequency terms, so the frequency domain plot would look like a single horizontal line at some nonzero height.

Being able to represent any waveform either way means that if you could have millions (really, an infinity) of people, each with a tuning fork of a different frequency, and you could control how hard and at what instant each person hits his or her fork, you could exactly reproduce the Gettysburg address in Lincoln’s voice, or any other sound you can imagine.

Other properties of the Fourier transform that can affect the imaging process include aliasing, where a frequency higher than the sampling frequency $f_s$ is incorrectly identified as being a different frequency (lower than $f_s$), as shown below. Imagine that the green vertical lines represent the instants in time when samples are taken. The red line is the signal we’re trying to sample, but notice that everywhere the red line intersects a green line, it also intersects the blue line, which is a lower frequency signal. There are an infinite number of sine waves that have the same value as the red line at each sampling interval (green line). We will always assume that the lowest frequency wave that fits the data is the signal actually being detected because it’s the most reasonable assumption. It is up to the people designing/using the equipment to know its limitations and to not exceed them.
The highest frequency we can reliably measure when we sample at a frequency \( f_s \) is just twice that frequency. This is known as the Nyquist condition, and it’s the reason CDs are sampled at a rate of 44.1 kHz, even though the upper limit of human hearing is about 20 kHz. We can’t accurately detect 20 kHz sound waves if we don’t sample at 40 kHz or better. Aliasing can be seen when someone on TV wears clothes with very thin and closely-spaced stripes – they tend to look like they’re moving or made of different (solid) colors because the spacing is smaller than the resolution of the TV signal.

**Projections and the Radon Transform**

When we have taken multiple images of the patient from different angles, the simplest reconstruction possible will involve the backprojection of those images. Imagine our detector is a long, narrow line rather than a plane. We will get images by placing the detector on one side of the patient and the X-ray tube (if we’re not doing SPECT) on the other side. The line-shaped detector will be oriented perpendicular to the line connecting it to the patient, as seen below.
We can describe the projection data completely in terms of intensity at a given position \( s \) along the line which is rotated an angle \( \theta \) from our arbitrary starting point. We do this for each slice as well, but we'll ignore that for now. Each \( \theta \) (or projection angle) will produce a simple plot of intensity vs. \( s \). Backprojection is the process of taking all the plot for each \( \theta \) and "smearing" it back in the direction it came.

For example, in the projection below, we have a high intensity spike on the left and a gradually decreasing intensity to the right. Assume that this projection came from \( \theta = 0^\circ \). We would then backproject it as shown.

We have no way of knowing where the source of high intensity on the left side of the projection originates. It could be near the detector, it could be on the other side, or it could be evenly distributed along the line perpendicular to the detector. We can’t know, so we just say that there is high intensity, on average, somewhere along the left side of the patient when seen from the angle \( \theta \). We are measuring one number (intensity) at each position \( s \), which is necessarily a combined effect of everything along the line perpendicular to our film at position \( s \). The information about how the intensity is distributed along that perpendicular line is unknown and unknowable (with a single projection).

We can then look at our next projection below, which we'll say is from 90°.
It has a high intensity at the right edge, meaning we backproject it and get the figure on the right. When combined with the earlier projection, we get this:

Using the two projections, we have narrowed the source of the high intensity to the lower left part of the image. We can continue this for more projections and better locate the source. (The terminology I’m using is a little sloppy – the basic ideas are the same for both CT and SPECT, but **high intensity** in SPECT would mean lots of escaping gamma rays and little attenuation, while a high-intensity region on an X-ray or CT scan would mean a region where X-rays had no problem penetrating, so we’d be finding something like an air pocket). The image on the left shows several projections taken with a high-intensity source inside the patient while the image on the right is the image reconstructed from backprojections.
There are a couple of things to notice here. First, more projections will give us a better image. The downside, though, is that we end up with an artificial darkening in the center of the image, since every backprojection will pass through it. We’ll see how to correct for this later.

One more thing to be aware of is that the frequency domain here refers to oscillations in position rather than time. In other words, realize that high-frequency sound is sound that changes significantly over very short time intervals. High frequencies here represent things that change significantly over a very few pixels, or picture elements. If we were looking at a picture that had only low frequency components, that means that each pixel would not look dramatically different from its neighbors. A picture with large high-frequency components can make nearby pixels very different. For example, the pictures below show some of the simplest two-dimensional intensity plots, just as our earlier single-frequency sine wave plot was one of the simplest one-dimensional intensity plots.
These pictures show the result of a sine wave executing a single period in the x direction, a single period in the y direction, and one period in each direction. This can be compared to the grayscale image we might expect in a medical scan below, where we show the result of a single period in the x direction and 4 periods in the y direction.

The reason we gather these projections and then perform a Fourier transform on them is the Fourier slice theorem. It says that the FT of the projection data along the line at a particular projection angle $\theta$ is equal to what we’d get if we looked at the FT of the full 2-D image along a the line going through the center at an angle $\theta$.
This is the way we can reconstruct the distribution of intensity (or activity, or attenuation, or whatever we're mapping). In the above image (left), we see that we really only have data wherever a blue line is present. In between the lines, we can only make an educated guess. To get a perfect reconstruction, we would need to have blue lines at all angles so that we know the full details of the data. Remember that the only way we can get more lines is to take more projections in the first place. This tells us that, in general, more projections will yield a better reconstruction.

What’s the problem with doing multiple projections? Well, if the imaging time is going to be limited by patient comfort and financial constraints, it means that we still have the same hour or less to get the full scan done. Taking four times as many projections means each one will have to happen in one-fourth the time (ignoring the time taken to reposition the camera). A shorter scan at each projection means fewer photons, and the noise will go up significantly (even more than the simple $N^{1/2}$ projection because of the peculiarities of the backprojection process).

A trade-off has to be made between having a larger number of noisy projections or a smaller number of better projections. One solution to this problem is to use two (or more) gamma cameras; we can obviously expect $N$ cameras will give us $N$ times the efficiency of a single camera. Of course, we can also expect that these multi-head machines will cost about $N$ times as much!
Filter Choice

If you’re trying to decide how to reconstruct and process a typical image (128 x 128, for example), one of the things you have to deal with is noise. It might occur to you that you could smooth the noise by having some kind of weighting factor to tone down a pixel which is much brighter than its neighbors or to boost a pixel which is much dimmer than its neighbors. This is the principle behind filters, and this could be thought of as a low-pass filter. Low frequencies (which will show significant changes only over a few pixels or more) will survive this kind of manipulation but high frequencies (dramatic changes from pixel to pixel) will not. In some respects, this is a good way to get rid of noise, since it will certainly vary from pixel to pixel due to its random nature. Of course, what would a tiny cancerous lesion look like in your image? A bright pixel surrounded by dimmer pixels. Your filter will tend to smear this into the background and make it much harder (if not impossible) to detect. This is the inherent trade-off in filtering.

One simple kind of filter that can be applied to the data is the ramp filter. It’s called a ramp because the amplitude vs. frequency graph of the filter is just a straight line with positive slope, or a ramp. This boosts high frequencies more than low frequencies, which is required to compensate for the backprojection’s tendency to pile up activity in a 1/r way around hot spots. Because the camera itself has a finite resolution, the ramp can’t be allowed to increase indefinitely. This is why a significantly better filter, known as the Butterworth filter, starts off as a ramp but drops to zero relatively quickly for frequencies above the point where noise dominates the signal and amplification would only worsen the problem. The Butterworth filter has two parameters, known as the cutoff and the order. These determine the precise shape of the filter, but a few examples are shown below.
The black line is a simple ramp filter which is chopped at 0.5 cycles per pixel (the Nyquist limit) while the other filters have cutoffs at 0.1, 0.2, and 0.3 cycles per pixel.

**Other Reconstruction Methods**

Filtered backprojection is simple and fast, but does it give the most useful images? It is exact for the "simple" problem of measuring attenuation only (meaning it will work fine or CT scans), but not for SPECT. In the past few years, iterative reconstruction techniques have become increasingly popular in SPECT. The idea here is that we start with a model of the imaging process known as a projector. This is something that describes the physics of the process from generation of radiation to the passage through the patient, collimator, and ultimately into the camera. The projector is applied to an initial estimate of activity distribution that may be uniform (i.e., assume the activity is the same everywhere within the patient’s body) to generate projection data. These are compared with the real projection data acquired in the clinic, and differences are noted. The estimate of activity is updated with this new information (more activity here, less over here, etc.) and new projections are made which are again compared with the actual data. After repeating this process through a few cycles (iterations), we will typically end up with a very good reconstruction of the true situation. The drawback to this method is that it is much more computationally intensive than ordinary filtered backprojection. Of course, that is a diminishing problem as computers get faster and faster.

One of the most popular forms of this iterative reconstruction is known as OS-EM, or Ordered Subsets-Expectation Maximization. This is just a way to group the projection data to speed up the reconstruction process and minimize deviations between the computer-generated projections and the real ones. One of the reasons for the superior performance of iterative reconstruction is the ability to incorporate the physics of the process into the projector; for example, we can supply the algorithm with an attenuation map so that it knows a small amount of activity in a region of low attenuation can give brighter projections than a larger amount of activity in a region of high attenuation. Also, we can include estimations of things like scatter in the patient as well as in the collimator itself, and penetration of the collimator by higher-energy photons.

Of course, we have to get the attenuation map first if we are going to use it in the reconstruction process. To do this, there are two popular methods. Originally, a long line source would be rotated around the patient with the camera on the opposite side. This would be kind of like an X-ray CT scan made with gammas instead of X-rays. Knowing the strength and position of the source would then give us a way to calculate the attenuation coefficient for the path through the body.

A method that has become more popular recently involves hybrid machines; these may be SPECT and CT or PET and CT. The CT scanner gives an attenuation map while the PET or SPECT machine provides emission data. Together, they will give an accurate map of activity in the body.
Other problems that can be at least partially addressed in an iterative reconstruction include scatter correction, both in the patient and in the collimator itself. This is a difficult process which depends on the attenuation map of the patient and on the particular collimator and radioisotope used.

The SPECT Process

SPECT imaging presents different challenges than those found in X-ray CT. In addition to the distinction between transmission/emission (or attenuation/activity) discussed earlier, the acquisition protocols themselves are different. Since each X-ray in CT travels through the entire patient (or is stopped), complete data can be obtained by rotating 180° around the patient. In SPECT, since the emission event can happen anywhere along a line through the patient, we need a full 360° of data for the best reconstruction. While true in general, cardiac SPECT is usually done with only 180° of data since the heart will only be poorly imaged from the back. The imperfections in the reconstruction that arise from this limited data set affect image quality less than the mostly-scattered photons that would be picked up in the posterior scan.

As the camera is moved around the patient, we have another choice to make. We can move the camera slowly but continuously or we can move it in a series of discrete steps, taking pictures only when the camera is stopped. The latter method is called step and shoot, and makes grouping of projections easy. In the continuous acquisition case, since we won’t have an infinity of projections, these have to be binned by angle. This means there will be a certain amount of blurring since the camera is taking data while moving (in other words, some of the data in a bin might be taken from when the camera was at 270° while other data comes from 271°, 272°, etc.). In general, the step and shoot method is more useful.

There is also a choice to be made about how to divide the camera face into pixels. More pixels will yield a higher-resolution image in general, although the collimator and camera itself provide a limit that pixel choice cannot fix. The downside to increasing the number of pixels is that each pixel will record fewer detected photons, and the noise (because of the reconstruction process) will actually get worse than would be predicted by simple $N^{1/2}$ statistics.

Common matrix choices are 64 x 64 and 128 x 128 (with 256 x 256 becoming more common). Compare this to the number of pixels in the cheapest digital cameras (usually at least 640 x 480). Can you think of the reason why a machine costing hundreds of thousands of dollars would have poorer resolution than a $20 camera?

How many different projection angles should be used? What size projection matrix is “right”? The technique of filtered backprojection fixes the smearing caused by running each projection through the center of the image, but the correction moves radially outward as the number of projections is increased. It would seem that we should use many projections to enlarge the area of acceptable reconstruction as much as possible. Of course, if we are still limited by patient comfort and/or economics to a scan of approximately thirty minutes, more projections necessarily means less time spent at
each position, which means fewer photons and more noise. In general, the number of projections should match the linear dimension of the projection matrix (i.e., 128 rotations for a 128 x 128 pixel image). The pixel size itself is chosen to be around 1/2 to 1/3 the resolution of the camera.

**Collimators**

The choice of collimator for a SPECT study is very important. There are multiple designs available, but the most common is the **parallel-hole collimator** (or **PHC**). This is just a honeycomb design where the hexagonal tubes are all parallel to one another and perpendicular to the face of the collimator. There are two basic ways to make these collimators, known as the **cast** method or the **foil** method. In the cast method, tens of thousands of tiny hexagonal pins (kind of like small Allen wrenches) are all lined up in a mold and molten lead is poured around them. When it cools and solidifies, the pins are carefully tapped out and a collimator is left behind. Because the lead walls are thin, soft, and hazardous, the collimator is enclosed. Dropping the (heavy!) collimator will almost certainly cause some of the walls to collapse, ruining it.

The foil method uses adhesive to fasten corrugated foils together, as shown below.

Because of the potential for nonuniformities with the foil method, the cast method has largely replaced it.

The collimator's face will have the same dimensions as the camera it is built to fit – a common size is a 40 cm by 50 cm rectangle. The thickness of the collimator (length of the hexagonal tubes) depends on the intended purpose. A high-resolution collimator will generally be thicker due to the longer holes than a lower-resolution collimator. Thickness ranges from a few cm to several cm, but decreasing efficiency and increasing weight tend to limit the thickness on the high end. As discussed earlier, the thickness of the septa will be determined by the energy of the gammas emitted by the isotope used. The tradeoff is again between efficiency (thinner septa are better) and excessive penetration (thicker septa are better). For the relatively low-energy isotope $^{99m}$Tc, septa may be only 0.15 mm (150 $\mu$m) thick, while collimators used at higher energies (such as when using a SPECT machine to do PET) may have septa a few mm thick.

The septal thickness is not the end of the story when looking at efficiency or resolution. Since parallel-hole collimators reject photons based on their angular spread, the linear resolution of such a collimator will degrade as the source is moved away. For this reason, great pains are taken to minimize the distance between the collimator and the patient. Because of this and the fact that people don't generally have a circular cross-section, most SPECT machines have the option to move the camera in a **non-circular**
orbit (NCO). This keeps the camera at a constant (and small) distance from the patient in an effort to improve resolution. While resolution can also be improved by the use of longer, narrower holes in the collimator, that will degrade efficiency. As a rule of thumb, we can say that efficiency is roughly proportional to the square of resolution (remember that larger efficiency is better, but smaller resolution is better).

The parallel-hole collimator’s resolution can be found from the formula

\[ R = \frac{a + b + c}{a} \times d \]

where the length of the collimator holes is \(a\), the distance between the source and collimator is \(b\), the distance between the collimator and the point where the gamma interacts in the crystal is \(c\), and the hole diameter is \(d\). This clearly shows the way resolution climbs (gets poorer) with increasing depth in the patient.

In addition to differences in hole dimensions, collimators can be made in different shapes as well. A parallel-hole collimator will produce an image on the gamma camera that is the same size as the organ being imaged. Depending on the particular application, we might need to fit a larger image on the camera or we might not want to “waste” so much area on the camera face when imaging something very small, like the thyroid.

Other collimator choices include the fan beam, cone beam, pinhole, and rotating slant-hole. In the fan beam design, the holes remain parallel along one axis but converge along another axis, kind of like a stack of triangular rakes on top of one another. This means that all of the holes point to a line, just as all of the holes below point to the green line.

The fan beam can be either a diverging or a converging collimator, depending on which way it’s facing. If the holes are closer together on the camera face than towards the patient, the image will be smaller than the object and it is a diverging configuration.

If the holes are closer together on the patient side, we have magnification of the image and a converging collimator. The benefit of the diverging collimator is to give a wider field of view than we would otherwise have, while the converging collimator gives a
larger image and is more efficient than a parallel-hole collimator (if the radioactivity is being emitted isotropically, a hole pointed at the patient is more likely to catch a photon than a hole pointed at the wall behind the patient).

We can go further and focus the holes in two dimensions, giving a cone beam collimator. Instead of all of the holes pointing towards a single axis, they are now all directed towards a single point, as shown below.

For even smaller organs, a pinhole collimator may be used. This is the gamma-ray version of an ordinary pinhole camera, where rays going through a hole travel as shown below.

The image is magnified but the resolution drops off as you move away from the region immediately behind the hole. In addition to thyroid imaging, pinhole SPECT is useful in the rapidly growing field of small animal imaging. Rather than gathering large numbers of mice for medical studies and having to dissect a statistically-significant number of them periodically to track their progress, a smaller group can be used and kept alive if we can image them using SPECT, PET, CT, etc. rather than a microscope.

The final kind of collimator we'll look at (briefly) is experimental and is known as a rotating slant-hole (RSH) collimator. This collimator has a circular face which is broken into segments, like slices of a pie. Typically, there will be four segments and the holes within a segment will be parallel to each other, but the individual segments are not angled the same way. If you imagine light streaming out of the camera through a collimator, the standard parallel-hole collimator would look just like a regular searchlight. The RSH collimator would look more like four spotlights all aimed at the same thing, as shown below. Where the PHC has to move all around the patient (dozens of positions throughout 360°, only a few of which are shown in the image on the left below), the RSH collimator can take a complete set of data needed for reconstruction in just three camera positions.
The RSH collimator manages this by rotating the face of the collimator multiple times at each of the three camera positions. Although the first guess at the improvement in efficiency might be a factor of four due to the four different segments, it’s really modified by the cosine of the slant angle, so it ends up being around three times as efficient as a PHC design. Of course, there are always tradeoffs, and in this case the tradeoff comes in a significantly smaller imaged volume. The region seen by every segment in every camera rotation at every collimator position is relatively small – typically a spherical (roughly) shape with a diameter of about 12 cm. This may make the RSH collimator suitable for breast imaging or possibly for cardiac imaging in an emergency-room situation.

**SPECT Cardiac Imaging**

The most important application of SPECT imaging today is **myocardial perfusion imaging**. This generally takes the form of a stress test combined with the injection of either $^{99m}$Tc or $^{201}$TI which will be distributed throughout the heart due to exercise or a chemical heart stimulant for patients who cannot use the treadmill. The procedures vary from clinic to clinic, but the basic idea is to compare the resting distribution of an isotope to the post-exercise distribution. Blood starved regions of the heart are darker in the SPECT images since tissue in the healthy, high-flow region gets brighter during exercise. If $^{99m}$Tc redistributes over time (the “rest” image), defects are reversible by bypass or cleaning. If not, the dark tissue is essentially dead cardiac muscle. There needs to be some delay between the two imaging sessions so that cross-contamination of the images is minimized. The delay could be hours or it could be a full day. Some clinics are using dual-isotope SPECT scans where $^{99m}$Tc and $^{201}$TI are injected at different points in the procedure (one rest, one stress) and imaged without a lengthy delay due to the relatively small overlap of their energies.

The results of a SPECT scan of the heart are shown below:
If a SPECT study of the heart could be completed in a second or less (as some ultrafast CT scans can be), motion of the heart would not be much of a problem. As it is, the movement of the heart back and forth throughout the scan will blur the image, making diagnostic information harder to determine. For an idea of what we’re facing, imagine that you would like to film a complete cycle of a pendulum in a darkened room. If you take a quick snapshot, there will be very little motion blurring but you won’t leave the shutter open long enough to collect enough photons to adequately expose the film. On the other hand, if you lock the shutter open long enough to get a bright image, the pendulum’s position will be smeared throughout its entire range.

One solution to this is to take many fast, dim pictures and then combine them, but combine them after sorting them by position. If we break the pendulum’s swing of (for example) 30° into 15 distinct angular positions, we can then put each brief, dim exposure into one of the 15 piles. Eventually, we’ll combine the 30 separate “families” of pictures into 15 pictures, one every 2°, as shown below.
This can be done in cardiac imaging by using ECG (electrocardiogram) data to identify different phases of the heartbeat, and is called **gating**. Typically, either 8 or 16 “bins” (positions in the cycle) are used. Recent developments have also explored the idea of **respiratory gating**, which makes an effort to repeat this process for different phases of the breathing cycle. Although the lungs aren’t the organs of interest in cardiac imaging, you can see for yourself that your entire torso moves when you inhale or exhale, and the heart moves during this process as well.

### Other Applications of SPECT

In addition to myocardial perfusion imaging, SPECT can be used for brain imaging (\(^{99m}\)Tc HMPAO used to detect Alzheimer’s disease), thyroid imaging, breast cancer, prostate cancer, etc. In addition to \(^{99m}\)Tc and \(^{201}\)TI, other commonly-used isotopes include \(^{123}\)I (a nice 159 keV gamma, but other high-energy gammas are also emitted less often. This isotope may be coupled with a regular low-energy collimator, but the presence of the high-energy component suggests a collimator with thicker septa may be a better choice), \(^{67}\)Ga, \(^{133}\)Xe, and \(^{111}\)In. Many others are used with less frequency.

### Computational Simulation

As in other areas of physics, computational simulation has emerged as a third structural element in the subject, along with theoretical and experimental techniques. These techniques are generally referred to as **Monte Carlo** methods. The name is derived from a famous casino in Monaco. The basic idea is that statistical information can be obtained by applying a relatively simple set of rules (things like Compton scattering, etc., in our case) to processes which have a pseudorandom component to them (which tiny area in the body will emit a gamma ray next, and which way will it be aimed?). We call the numbers that determine these outcomes **pseudorandom** because everything a traditional computer does is deterministic. For example, if someone read the last digit of each of the numbers in the phone book to you, you would be unable to distinguish that sequence from something truly random (like the roll of a ten-sided die). It wouldn’t be truly random since anyone who knew the algorithm (had the phone book) could predict what numbers will come next with 100% accuracy. In many cases, pseudorandom processes are more desirable for simulation than truly random.
processes would be, since it removes one more variable and enables you to repeat a simulation exactly (something that would be impossible with truly random numbers).

All phases of the imaging process can be modeled in a computer and insights into the true nature of what actually happens can be gained. Simulation is not a substitute for experimentation, but rather an alternative means of exploring the physics involved. The benefits include 1) the ability to exactly duplicate a simulation, changing only the variables of your choice 2) additional information which could not otherwise be obtained, either through human subject experiments or those done with physical phantoms, such as the complete history of a photon from emission through scatter to detection or absorption 3) no need to purchase a SPECT scanner or wait until after the nuclear medicine clinic is closed to gather data 4) simulation of the imaging process does not require patient/physician cooperation, IRB (institutional review board) approval, etc. 5) the ability to remove or quantify background radiation and 6) knowledge of the true activity distribution.

Drawbacks include 1) simulation is not experimentation. If there is a conflict between the two, the experiment is correct. 2) It is frequently slower than an actual experiment done with realistic activity. This of course depends on the number and power of the computers at your disposal. 3) The programs are usually written with one thing in mind – some have simple models for the things they are not most interested in, such as (typically) scatter 4) (related to the first disadvantage) The simulation is only as correct as the algorithms making it up. In SPECT, this is not necessarily a major problem since the physics of Compton scattering and the photoelectric effect are well understood.

The simulation needs to have a means for defining or importing phantoms. These are the representations of the activity and attenuation of a real human body. Bitmaps are most commonly used for this, and they may be generated by a different program. There must be an engine that determines where the next decay will occur, propagates the photon through the body, the collimator, and the camera, and handles things like patient scatter, collimator penetration and scatter, the production of characteristic X-rays from the lead in the collimator, and possible absorption along the way. Also, the photon may not be completely stopped by the crystal and may only deposit some of its energy there (partial deposition). Conversely, two photons may hit at the same time (or during the time the light is being collected) and the camera will record that as the arrival of a single, larger photon at a position between the two actual points of interaction.

There are several programs available for simulation of the imaging process, including SimSET, SIMIND, MCNP, GEANT, EGS4, and GATE, among others. These have different strengths and weaknesses. Some are for a specific modality (e.g. SIMIND, which is designed for SPECT) and some have been modified from other purposes for imaging simulation (GEANT, originally used for high-energy physics, and MCNP, originally used at Los Alamos for nuclear weapons design, among others).

The images below (taken from the NCAT phantom produced by the Division of Medical Imaging Physics at Johns Hopkins - [http://dmip.rad.jhmi.edu/](http://dmip.rad.jhmi.edu/) are activity phantoms commonly used in simulations of the $^{99m}$Tc SPECT imaging process. The views
(negative images) are from the front, top, and side of the body. As you can see, the liver and kidneys are very dark, corresponding to high activity, while the lungs are barely visible.

Projection data (negative) from a parallel-hole collimator simulation of a mammography image is shown on the left below, along with a few projections from a rotating slant-hole SPECT simulation.

The reconstruction of the SPECT data gives us a 3-D image. One slice of that image is shown below (the arrow indicates the presence of a lesion).

All of this could be done by experiment, but the next part is only possible using simulations. The photons are tracked by origin and history to get the pictures below. The three columns are three different camera positions and the rows represent 1) the source of all detected photons 2) the point of last interaction for all detected photons 3) the point of last interaction for all scattered photons. This tells us that the middle
camera position (row 1, column 2) picks up a great deal of activity that came from the heart. This is undesirable for breast imaging. The other two camera positions in that column show much greater contributions from the breast and much smaller from the heart.

The second row shows that the breast is the last point of interaction for most photons. A filtered backprojection algorithm can only determine the last point of interaction, since it’s apparently where the photon in question came from. We don’t want a lot of interaction in the breast after the photons have already been emitted, since that will be evenly distributed (scatter) and will reduce the overall contrast or signal to noise ratio.

Finally, the bottom row shows the last interaction location for scattered photons. This is different from row 2, which includes primary (unscattered) photons. Row 2 is therefore a mix of good (primary) breast photons and bad (scattered) photons from all over – the breast, the heart, etc.
One of the results of this research was the discovery that covering the patient with a lead apron (except for the breast) would substantially reduce the scatter reaching the camera. This was relatively easy to discover through simulation, but would not have been as easy to determine via experiment.

**PET – General Principles**

A more recent advance in nuclear medicine is **Positron Emission Tomography**, or **PET**. PET is similar to SPECT in that they are both based on emission rather than transmission and they both image function rather than structure. In PET imaging, the patient is given a radioisotope which undergoes $\beta^+$ decay, producing positrons. The positrons will travel short distances (typically a few mm) before slowing enough to meet up with an electron and annihilate, producing two gamma rays of 511 keV each. When the positron and electron are at rest (forming **positronium**), the two gamma rays will be emitted in exactly opposite directions (180° apart). In the center of momentum frame where the positrons are not moving, the angle is exactly 180°. If the positron still has some forward momentum, the angle between the two gammas will be slightly reduced in the lab frame. This has an effect on resolution which depends on the diameter of the ring of detectors that will ultimately intercept the gammas. For a ring diameter of about one meter, this means a minimum resolution of around 2 mm.

One of the large advantages of PET over SPECT is the possibility of **electronic collimation**. Rather than trying to locate decay events by using lead tunnels to restrict the field of view of the detectors, specialized circuitry looks at decays received by two detectors within a **coincidence window**, which is usually about 10 ns. If two opposing detectors in the ring of many detectors each register a gamma interaction within that coincidence window, it is assumed that they are the two photons produced from a positron-electron annihilation somewhere along the line joining them. Choice of the coincidence window is another tradeoff. A window that is too wide will allow unrelated photons to be considered to be related while a window that is too narrow may reject genuine coincidences. Even with this restriction, the number of photons captured is far better than the 1 in 10,000 of SPECT, meaning the signal to noise ratio is correspondingly larger.

The coincidence count rate comes from three sources: true coincidences, random coincidences, and scatter coincidences. The true coincidence rate increases linearly with the amount of activity in the patient and is what the machine is designed to detect. The random coincidence rate occurs because unrelated gammas will sometimes hit two detectors within the coincidence window, making them appear to be related when they are in fact from two separate decays.

The rate $R$ of random coincidences between two detectors with non-coincidence count rates of $\sigma$ will be $R = \tau \sigma^2$ where $\tau$ is the width of the coincidence window. This means that the rate of randoms will increase with the **square** of patient activity instead of linearly with patient activity, as true coincidences do. The effects of this are partially alleviated by not connecting every detector to every other detector in a coincidence circuit; two detectors which are adjacent to one another would not be wired in
coincidence since there is no part of the patient between the two detectors. See the diagram below.

If one or both of the gammas is scattered, the line connecting them (known as the line of response, or LOR) will not contain the point where the actual annihilation occurred, as shown below.

Two gammas emitted from green star – lower one is scattered away from green detector towards orange detector. LOR is incorrectly calculated as black line.

We can try to look at the energies of the two gammas and reject scattered photons just as we do in SPECT, but there are a couple of problems here. First, some of the best detector materials (in other respects) for PET have energy resolution which is even poorer than the NaI(Tl) crystals commonly used in SPECT. This means photons in a wide energy range must be accepted. A photon that would be 511 keV when unscattered can be deflected by about 12 degrees and still have an energy of 500 keV. If the ring diameter is a little over a meter, a 12 degree change in direction could displace the LOR by several centimeters! We can’t set the lower energy limit too high, though, because these high-energy gammas may not lose all of their energy in the crystal. If they only deposit 400 keV in the crystal, we’ll miss them. The location of this energy window is yet another choice to be made.

**Detectors**

The detector system in a PET machine is similar in many ways to that in a regular gamma camera. There is a scintillator which turns gamma photons into visible photons and a collection of photomultiplier tubes recording those visible photons. One major difference is in the number of crystals and PMTs – a PET scanner may have more than a thousand PMTs, each watching 16 detectors.
There are several possible choices for the scintillating crystal material in a PET scanner. We could use NaI(Tl), just as in the gamma camera. It has superior energy resolution, but its low density and low effective value of $Z$ (number of protons) means that while it is quite efficient at stopping 140 keV photons, it is not particularly good at stopping the 511 keV gammas produced in PET imaging. This is even more of a problem than you might think at first because of the necessity of stopping **both** gammas for this process to work. The attenuation coefficient of NaI(Tl) for 511 keV photons is $0.34/\text{cm}^{-1}$. If a certain thickness of NaI(Tl) has a 75% chance of stopping a gamma, what is the chance that both gammas will be caught?

The stopping power of the crystal is important because using a thicker crystal means there will be **depth of interaction** effects. If the annihilation event does not occur at the geometric center of the ring (as will of course be the case for the vast majority of events), there are different possibilities for the detectors that will record the arrival of the gammas, as shown below (greatly exaggerated).

Two of the four different detectors could record the two gammas (red), with the resulting LOR being estimated to be any of the four dashed lines which join the **centers** of the scintillator blocks. Depending on the actual point of origin along the red line, the reconstructed LOR could be far from the actual LOR.

Other choices include BGO (Bismuth Germanate), LSO (Lutetium Oxyorthosilicate) and YSO (Yttrium Oxyorthosilicate), and BaF$_2$ (Barium Fluoride), among others. BGO has a much larger attenuation coefficient ($0.92 \text{ cm}^{-1}$), meaning a thinner crystal of BGO will stop gammas as effectively as a thicker crystal of NaI(Tl), reducing the depth of interaction problem. Unfortunately, the light output of BGO is about 1/6 that of NaI(Tl), meaning its energy resolution will be significantly poorer.

The **decay time** of a scintillator is another important measure of a detector’s usability. This is the time required for the light level in a scintillator to return to a specific fraction of its maximum intensity. Longer decay times mean lower peak count rates. BGO has a slightly longer decay time than NaI(Tl) (300 ns vs. 230 ns), but BaF$_2$ has a remarkably short decay time of only 0.8 ns. For this reason, it was considered to be a good
candidate for the production of a time-of-flight (TOF) PET machine. The idea behind this technology is to measure the arrival times of the two gammas with high accuracy, since the difference in arrival time would be proportional to the difference in the paths taken by the two photons. A positron annihilation at the center of the ring of detectors would give simultaneous arrivals, but an annihilation closer to the surface of the patient’s body would place one gamma significantly nearer to the detector than the other. Ideally, TOF PET would remove the need for any reconstruction algorithm, since the actual point of emission would be determined instead of just the line of response.

The timing level required for this kind of TOF PET is not yet available, since the gammas will travel approximately one foot (about 0.3 m) in one nanosecond. Precise identification of the annihilation point along the LOR would require timing in the picosecond range. Additionally, BaF$_2$ has a light output even smaller than BGO, at about 1/20 of NaI(Tl).

LSO has the advantage of a low decay time along with a high effective Z and high density, with an attenuation coefficient of 0.87 cm$^{-1}$, almost as large as that of BGO. It also has a light output about twice as great as BGO, giving superior energy resolution (though still not as good as NaI(Tl)). Although LSO is itself radioactive, that is not a problem for use in a PET system. Why?

One disadvantage of PET vs. SPECT is the ultimate resolution limit. The positron emitted in a nuclear decay can travel several millimeters before finding an electron to annihilate with. The two gammas produced will be tracked back to the point where the electron and positron were destroyed, **not** the point where the positron was produced, which is what we really want since the radioisotope was carried to a particular spot by design. The maximum range of the positron is dependent on the particular isotope involved. SPECT does not have this limitation since the gamma imaged is emitted at the location of the radioisotope. As a practical matter, this is not particularly important since efficiency considerations and the resolution of the typical gamma camera & collimator system (around 1 cm) combine to give resolution poorer than the average PET scan (about 5 mm).

**PET Isotopes**

The most commonly used isotope in PET imaging is $^{18}$F, which takes the place of a hydrogen atom in $^{18}$FDG, or 2-deoxy-2-$[^{18}$F$]$ fluoro-D-glucose. This sugar is used in studies of the brain, myocardium, bones, and other organs. $^{18}$F decays with a half-life of 109 minutes and is therefore transportable over reasonable distances. This fact allows a single cyclotron to provide $^{18}$F for multiple PET imaging centers.

Other useful isotopes include 1) $^{15}$O (in the form of water, O$_2$, and CO$_2$ for use in blood flow studies of the brain & heart) with a very short half-life of 2.03 minutes, meaning it will be useful only if the cyclotron producing it is on site 2) $^{13}$N (as ammonia, again studying blood flow in the heart) with a $T_{1/2}$ of 10 minutes 3) $^{11}$C (in many different forms for studies of brain & heart function) with a $T_{1/2}$ of 20 minutes, and 4) $^{82}$Rb (as RbCl, which acts like potassium, used measuring myocardial perfusion) with a half-life of only 76 seconds. Fortunately, $^{82}$Rb is produced via a generator and its very short
half-life allows for repeated imaging sessions of the same patient in a short visit. Unfortunately, the high energy of its positron (3.15 MeV maximum) means that the positron’s range is long and the resolution is poor.

**PET Systems**

PET systems contain multiple rings of detectors with intervening lead or tungsten septa that can be extended to reduce the chance that gammas can be scattered between imaging planes. This is known as **2-D** imaging. If the septa are retracted (**3-D** imaging), coincidences between planes are allowed, effectively doubling the axial resolution of the system since now a coincidence between planes 2 and 3 indicates the annihilation event was located in “plane” 2.5. The 3-D imaging mode will dramatically increase the number of both scatter and random coincidences, however, requiring a correction algorithm.

It is possible to do PET imaging with a dual-camera SPECT machine by orienting the cameras so that they are 180° away from each other. At this point, the two 511 keV photons can be detected by electronic coincidence systems or by regular physical collimation. The downsides include the fact that the two gamma cameras will cover a significantly smaller region than the typical large ring detectors in dedicated PET machines, as well as (in the electronic collimation case) the necessity for cameras with very high count rates, since the removal of the collimator will dramatically increase the number of detected photons. Additionally, the lead collimators used in the physical collimation scheme need **very** thick septa (2-3 mm or so) meaning the already-low efficiency of physical collimation is dropped even lower. The benefit is the ability to actually do some PET imaging with a machine that almost every nuclear medicine clinic already has.

**Attenuation & Reconstruction in PET**

Attenuation correction is a little simpler in PET than in SPECT because of the reliance on two photons. If the photons exit the body on a line which has a length $L$ inside the body, one of the photons will travel an unknown distance $x$ and the other will travel $L-x$. The photon traveling the longer distance has, in the uniform attenuation case, a smaller chance of making it out of the body. However, since detection of one depends (through coincidence circuitry) on the detection of the other, the combined probability is just

$$e^{-\mu x} e^{-\mu (L-x)} = e^{-\mu L}$$

Recall that this is of the same form as the attenuation in a regular CT scan. The SPECT case is much more difficult by virtue of the fact that it only requires single photons, so the (again unknown) length traveled in the body $x$ is not cancelled by any other term as it is in the equation above.

Because the attenuation is not uniform throughout the body, a map still has to be made. This can be done using the PET machine itself by moving a positron-emitting rod
source around the ring between the ring and the patient. This would correspond to the equation above where \( x = 0 \) since one photon will not enter the body at all before reaching the detector. A growing trend is the use of hybrid PET/CT systems where the CT part of the system is used to make an attenuation map which is later used in the reconstruction process of the PET image. The hybrid machine essentially removes the difficulties otherwise inherent in the process of registration, where images from two different modalities must be geometrically aligned.

The reconstruction process can be either simple filtered backprojection or the physics of the scatter and attenuation processes can be modeled and incorporated into an iterative reconstruction algorithm, similar to that found in SPECT. The reconstruction can be either 2-D or 3-D, but the amount of processing and scatter correction necessary for 3-D imaging goes up significantly.

The results of a PET scan using 22 mCi of \(^{18}\text{FDG}\) are shown below (animated on the web page from which it was taken). A transmission scan was taken for attenuation correction.

![PET Scan Images]

Bibliography

http://www.crump.ucla.edu/software/lpp/lpphome.html
http://depts.washington.edu/nucmed/IRL/pet_intro/
http://www.bnl.gov/pet/FDG.htm
http://www.med.harvard.edu/JPNM/TF03_04/Dec9/Write.html
http://www.scionixusa.com/pages/navbar/scin_crystals.html