Notes on Computerized Tomography

51:060 Bioimaging Fundamentals
Edwin L. Dove

Table of Contents

Notes on Computerized Tomography ................................................................. 1

1. Introduction ..................................................................................................... 3

2. Calculation of tomographic images ............................................................. 6

2.1 Basic concepts .............................................................................................. 6

2.2 Projection Details and the Radon Transform ............................................. 9

2.3 Sampling ...................................................................................................... 13

2.4 Image reconstruction .................................................................................. 13

2.4.1 Brute Force ............................................................................................ 13

2.4.2 Iterative Reconstruction ......................................................................... 14

2.4.3 Backprojection Method .......................................................................... 16

2.4.4 Fourier Reconstruction .......................................................................... 17

2.4.5 Filtered Backprojection (FBP) ............................................................... 18

2.4.5.1 Back Projection ................................................................................ 18

2.4.5.2 Filtered backprojection method ....................................................... 21

2.4.5.3 Discrete FBP ................................................................................... 22

2.4.5.4 Implementation ................................................................................ 24

2.4.6 Fan beam FBP ...................................................................................... 25

2.5 CT number .................................................................................................. 28

2.6 X-ray detectors ........................................................................................... 29
2.7 Imaging in three dimensions................................................................. 29
  2.7.1 Multiple 2D acquisitions................................................................. 29
  2.7.2 Spiral CT ...................................................................................... 30
  2.7.3 Cone-beam CT ............................................................................. 30
2.8 Image quality ..................................................................................... 31
  2.8.1 Image contrast ............................................................................. 31
  2.8.2 Spatial resolution ......................................................................... 31
  2.8.3 Noise .......................................................................................... 32
  2.8.4 Artifacts ....................................................................................... 33
    2.8.4.1 Under-sampling ....................................................................... 33
    2.8.4.2 Beam hardening ....................................................................... 33
    2.8.4.3 Partial volume effects .............................................................. 33
3. Appendix I – Fourier analysis................................................................. 35
  3.1 Introduction: .................................................................................... 35
  3.2 Response of LTIL system to eigenfunctions ....................................... 35
  3.3 Fourier Analysis of Periodic Signals .................................................. 36
  3.4 Calculating the Fourier Coefficients .................................................. 39
  3.5 Fourier Analysis of Long and Possibly Non-periodic Waveforms ........ 40
  3.6 Practical Procedure for Performing Fourier Analysis ....................... 41
  3.7 Frequency Analysis of Random Signals .......................................... 43
References............................................................................................... 47
1. Introduction

In this chapter we consider systems that provide important tomographic or three-dimensional (3D) capability. The tomogram is effectively an image of a slice taken through a 3D volume. Ideally, it is free of the effects of intervening structures, thus providing distinct improvement on the ability to visualize structures of interest. A CT image is the result of the superposition of all planes normal to the direction of propagation. Essentially, the system has infinite depth of focus.

Even though a CT image provides information not available in a traditional X-ray, the information is not provided at zero cost. Consider the following table in which the radiation dose of one CT slice is compared to the doses delivered in other common medical procedures.

Table 1.1 Radiation dose received by the patient in several X-ray procedures. (from Sung et al.)

<table>
<thead>
<tr>
<th>X-ray procedure</th>
<th>Dose (per exposure)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chest</td>
<td>20 mR</td>
</tr>
<tr>
<td>Brain</td>
<td>250 mR</td>
</tr>
<tr>
<td>Abdomen</td>
<td>550 mR</td>
</tr>
<tr>
<td>Dental</td>
<td>650 mR</td>
</tr>
<tr>
<td>Breast</td>
<td>54 mR</td>
</tr>
<tr>
<td>Xeromammography</td>
<td>200 mR</td>
</tr>
<tr>
<td>CT/slice</td>
<td>1000 mR</td>
</tr>
</tbody>
</table>

1 R = 0.01 Sv
(ICRP allows 50 mSv=5R per worker/5 year; 1 mSv=100mR/person/5year)

Are these doses dangerous? The answer to this question is not straightforward and simple, but rather involves, in a complex way, many factors or parameters. The biological effects of radiation have been studied for many years, and based on epidemiological studies, the answer to this question seems to depend on five parameters:

1. Threshold: There is no single threshold that serves as an upper limit of acceptable exposure. In general, since radiation effects are accumulative over a lifetime, the threshold should be as low as possible based on an analysis of the benefits of exposure versus the iatrogenic effects of radiation.
2. Exposure time: The severity of the biological effects depends on the duration of exposure. A given dose will produce fewer effects if divided than if it were given in a single exposure.
3. Exposure area: The larger the body areas exposed, the greater the chances of damage. Those areas of the body not directly near the part being studied are highly shielded.
4. Individual sensitivity: There is a wide variation in the radiosensitivity of various species. For this reason, the lethal dose is expressed in statistical terms. For humans, the lethal dose is 450 rads for whole body irradiation, that is, 50% of the population will die if exposed to 450 rads over a 30-day period.
5. Cell sensitivity: Within the same individual, the most sensitive cells to radiation damage are those cells that are rapidly dividing. Furthermore, specialized cells are more sensitive than non-specialized cells. Cells that are very sensitive to radiation effects include egg cells, sperm cells, white blood cells, immature red blood cells, and epithelial cells. Cells less sensitive include muscle and nerve cells.

If a dose of radiation of unusually large quantity (over 100 rads) is delivered to the body in a very short time, biological effects that occur within hours or days after irradiation are referred to as short-term effects. The signs and symptoms of these effects include nausea, vomiting, malaise, and ultimately fever, shock and possible death. These effects are collectively known as the acute radiation syndrome.

In diagnostic radiation, long-term and subtler effects caused by low dose radiation may become manifest years after the exposure. Of these effects, the carcinogenic and genetic effects have received the most attention. There is mounting evidence now to indicate that diagnostic X-rays can induce these effects. The National Council on Radiation Protection and Measurement established the maximal allowable dose per person per year is 5 R, or about 1/5 of the typical dose delivered in one CT image. In other words, a CT study of more than 5 slices typically exceeds the maximal allowable radiation dose in one year. The need for a large CT study must be medically compelling, especially in the young, and in those in childbearing years.

So what are the reasons for using CT rather than a traditional X-ray? There are primarily three reasons. When we look at a chest X-ray (Figure 1.1), certain anatomic features are immediately apparent. For example, the ribs show up as light structures because they attenuate the X-ray beam more strongly than the surrounding soft tissue, so the film (or detector) receives less exposure in the shadow of bones. The air-filled lungs show up as darker regions.
A short calculation illustrates the type of structure that one could expect to see with this sort of conventional radiograph. The linear attenuation coefficients for the energy spectrum of a typical diagnostic X-ray beam for air, bone, muscle and blood are

\[
\begin{align*}
\mu_{\text{air}} &= 0 \\
\mu_{\text{bone}} &= 0.48 \text{ cm}^{-1} \\
\mu_{\text{muscle}} &= 0.18 \text{ cm}^{-1} \\
\mu_{\text{blood}} &= 0.178 \text{ cm}^{-1}
\end{align*}
\]

Assuming that the tissue transmission follows Beer’s law that states:

\[
I(x) = I_0 \exp(-\mu x) \tag{1.1}
\]

then the following contrasts can be calculated for a slab of soft tissue with a 1-cm cavity in it:

**Table 1.2 Contrast in a transmission radiogram**
Modern X-ray film allows contrasts of the order of 2% to be easily seen, so a 1-cm diameter air-filled trachea can be visualized. However, blood in vessels and other soft tissue details (e.g., heart anatomy) cannot be seen on conventional radiographs. Thus, the first reason for using CT is to improve contrast to enable visualization of different types of soft tissue with X-ray energies.

The second reason for using CT is to enable 3D visualization (i.e., enable depth visualization). In a conventional radiogram, the 3D structure is collapsed onto a 2D film of detector.

The third reason is to allow visualization of a slice of the anatomy along any imaging plane. Conventional X-ray images produce a 2D image of the anatomy perpendicular to the X-ray beam. CT imaging allows visualization of a slice of anatomy at any oblique angle.

### 2. Calculation of tomographic images

#### 2.1 Basic concepts

Consider the case of an X-ray passing through a slice of tissue with four attenuation coefficients as shown in Figure 2.1. The output intensity $I$ is given by

$$I = I_0 \exp \left( - \left( \mu_1 + \mu_2 + \mu_3 + \mu_4 \right) x \right)$$  \hspace{1cm} (2.1)

where $x$ is the distance traveled by the X-ray beam. Taking the logarithm of Equation (2.1) and rearranging, we obtain the projection function

$$p(x) = \sum_{i=1}^{4} \mu_i$$  \hspace{1cm} (2.2)

The projection function represents the summation of the attenuation coefficients along a given X-ray path.

<table>
<thead>
<tr>
<th>Material in cavity</th>
<th>$I(x)/I_0$</th>
<th>Difference (%) with respect to muscle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air</td>
<td>1.0</td>
<td>+20</td>
</tr>
<tr>
<td>Blood</td>
<td>0.837</td>
<td>+0.2</td>
</tr>
<tr>
<td>Muscle</td>
<td>0.835</td>
<td>0</td>
</tr>
<tr>
<td>Bone</td>
<td>0.619</td>
<td>-26</td>
</tr>
</tbody>
</table>
Figure 2.1. A simplified X-ray beam passing through a segment of tissue with different linear attenuation coefficients (from Cho et al.).

In X-ray CT the contrast between one tissue type and another is associated with the different attenuation coefficients of the materials involved. The projection data taken at different views are the basis for tomographic image reconstruction. An example is shown in Figure 2.2. If the tissue block is rasterized (i.e., sampled into rows and columns), then for a particular row $i$ the projection function across all $m$ columns becomes:

$$p_i(x) = \mu_{i1} + \mu_{i2} + \mu_{i3} + \cdots + \mu_{im}$$

(2.3)

Figure 2.2. A 2D matrix of linear attenuation coefficients representing a segment of tissue (from Cho et al.).

The scan represented in Figure 2.2 is a scan taken at zero degrees with respect to the horizontal. In standard X-ray-based CT practice, a radiation source scans linearly along an object at a given view angle $\theta$, and after completion of the scan, the source and detector pair rotate a small angle $\Delta \theta$ and repeat the linear scan. This process is continued until the completion of 180° rotation. Thus for a parallel-beam data collection
scheme, a total of $180^\circ /N$ scans with an angular displacement of $\Delta \theta$ is collected for a scan angle of $180^\circ$. This is illustrated in Figure 2.3.

For the final image reconstruction in an image of $N \times N$, each beam is backprojected in the image matrix with some from of spatial or spatial-frequency filtering. The goal of CT imaging is to derive the unknown attenuation coefficients based only on the set of projection functions. The basic hardware for performing this procedure is illustrated in Figure 2.4.
2.2 Projection Details and the Radon Transform

Consider the parallel-beam geometry in Figure 2.5. The function $f(x, y)$ represents the linear attenuation coefficients.

Figure 2.4 Basic scanning system for computerized tomography (from Macovski).

Figure 2.5 Parallel beam geometry and coordinate system (from unpublished material).
The X-ray beam makes an angle $\theta$ with the $x$-axis. The un-attenuated intensity of the X-ray beams is $I_0$. A new coordinate system $(r,s)$ is defined by rotating $(x,y)$ counterclockwise over the angle $\theta$. The coordinate change obeys the following transformation formulae:

\[
\begin{bmatrix}
  x \\
  y
\end{bmatrix} = \begin{bmatrix}
  \cos \theta & -\sin \theta \\
  \sin \theta & \cos \theta
\end{bmatrix} \begin{bmatrix}
  r \\
  s
\end{bmatrix}
\]  

(2.4)

and

\[
\begin{bmatrix}
  r \\
  s
\end{bmatrix} = \begin{bmatrix}
  \cos \theta & \sin \theta \\
  -\sin \theta & \cos \theta
\end{bmatrix} \begin{bmatrix}
  x \\
  y
\end{bmatrix}
\]  

(2.5)

For a fixed angle $\theta$ the measured intensity profile as a function of $r$ is shown in Figure 2.6 and is given by

\[
I_\theta (r) = I_0 \exp \left[ -\sum_s f(x,y) \Delta s \right]
\]

\[
= I_0 \exp \left[ -\sum_s (f(r \cos \theta - s \sin \theta, r \sin \theta + s \cos \theta) \Delta s \right]
\]

(2.6)

\[\text{Figure 2.6 Intensity profile (from unpublished material).}\]

For each measured intensity, we can compute the corresponding attenuation line sum. Each intensity profile is then transformed into an attenuation profile as

\[
p_\theta (r) = -\ln \frac{I_\theta (r)}{I_0} = \sum_s (f(r \cos \theta - s \sin \theta, r \sin \theta + s \cos \theta) \Delta s
\]

(2.7)
The function $p_\theta(r)$ is called the projection of the function $f(x,y)$ along the angle $\theta$. A typical attenuation profile (or projection) is shown in Figure 2.7. Usually the projection profile is measured for $\theta$ ranging from 0 to $\pi$.

![Figure 2.7. Typical projection profile (from unpublished material).](image)

Stacking all the projections $p_\theta(r)$ together results in a 2D data set $p(r,\theta)$, which is called a sinogram. The transformation of a function $f(x,y)$ into the sinogram $p(r,\theta)$ is called the Radon transform, named after Johann Radon, who some contend is the founder of tomographic reconstruction. The Radon transform for a given $\theta$ is given by

$$p(r,\theta) = \Re \{ f(x,y) \} = \sum f(r \cos \theta - s \sin \theta, r \sin \theta + s \cos \theta) \Delta s$$

(2.8)

To recapitulate, the Radon Transform computes projections of an image along specified directions. A projection of a two-dimensional function $f(x,y)$ is a line integral in a certain direction. For example, the line integral of $f(x,y)$ in the vertical direction is the projection of $f(x,y)$ onto the $x$-axis; the line integral in the horizontal direction is the projection of $f(x,y)$ onto the $y$-axis. As an illustrative example, the function $f(x,y)$ in Figure 2.8 has a projection onto the $x'$ axis ($s$ in Equation (2.8)) found by integrating along the $y'$ direction ($r$ in Equation (2.8)). The projections are at an angle of $q$ with respect to the original orientation ($q$ is $\theta$ in Equation (2.8)).
Figure 2.8 Geometry of the Radon Transform (from MATLAB help files).

Another way of viewing this is to consider the parallel projections illustrated in Figure 2.9 in which parallel beams of X-rays passing through an object produce the projection function $R_q(x')$, which plays the role of $p_\theta(r)$ in our development.

Figure 2.9 Parallel beam projections through an object (from MATLAB help).
2.3 Sampling

In practice there is a limited number $M$ of projections or views and a limited number of detector samples. Under these limits, the sinogram $p(r,\theta)$ given by Equation (2.8) then becomes $p(n\Delta r, m\Delta \theta)$ with $M$ rows and $N$ columns. Given a particular beam aperture, it is possible to calculate the minimum number of detector samples required to prevent distortion due to aliasing. This minimum is known as the Nyquist criterion. This criterion states that if $\Delta s$ is the beam width, then the maximum sampling distance is $\Delta s/2$, or at least two samples per beam width are required.

The number of views is not so straightforward. A general rule is followed that states that the number of views should be $\pi/2$ times the number of detector samples. For example, if the field of view is 50 cm and the beam width is 1 cm, then 100 detector samples and 150 views are required.

2.4 Image reconstruction

Image reconstruction is the process of estimating an image $f(x,y)$ from a set of projections $p(r,\theta)$. Several algorithms exist to accomplish this task. These are:

- Brute force inversion technique
- Iterative reconstruction technique
- Backprojection technique
- Fourier-based technique
- Filtered backprojection technique

Of these, the last is the technique that is used in the majority of the modern medical image processing machines. We will, however, consider some of the other methods in some detail in order to develop an understanding of the process.

2.4.1 Brute Force

With this method, the projection set defines a system of simultaneous linear equations that in theory can be solved using algorithms from linear algebra. Consider the following 2x2 image with 6 projections:

$$
\begin{bmatrix}
w & x \\
y & z
\end{bmatrix}
= 
\begin{bmatrix}
12 \\
8
\end{bmatrix}
$$

$$
\begin{bmatrix}
13 & 11 & 9 & 7
\end{bmatrix}
$$
These can be organized into a system of equations given as:

\[
\begin{bmatrix}
1 & 1 & 0 & 0 \\
0 & 0 & 1 & 1 \\
1 & 0 & 1 & 0 \\
0 & 1 & 0 & 1 \\
1 & 0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
w \\
x \\
y \\
z
\end{bmatrix}
= 
\begin{bmatrix}
12 \\
8 \\
11 \\
9 \\
7 \\
13
\end{bmatrix}
\] (2.9)

This is an over-determined system of equations; there are 5 equations and 4 unknowns. As a consequence, some of the rows and columns are linearly dependent. The pseudo-inverse can be used to solve for the unknowns. This approach is not practical for real systems, which can have hundreds of simultaneous equations for a single slice.

### 2.4.2 Iterative Reconstruction

The iterative reconstruction (also known as algebraic reconstruction technique-ART) was the original reconstruction method used to solve the tomographic problem. The method consists of three steps:

1. Make an initial guess at the solution
2. Compute projections based on the guess
3. Refine the guess based on the weighted difference between the desired (real) projections and the actual (calculated) projections:

   \[ p^{i+1} = p^i + g(desired - actual) \] (2.10)

For example, if we have our favorite small image and its projections

\[
\begin{bmatrix}
w & x \\
y & z
\end{bmatrix}
\begin{bmatrix}
12 \\
8
\end{bmatrix}
\begin{bmatrix}
13 \\
11 \\
9 \\
7
\end{bmatrix}
\]

then:

1. Initial guess and projections

\[
\begin{bmatrix}
0 & 0 \\
0 & 0
\end{bmatrix}
\begin{bmatrix}
0 \\
0 \\
0 \\
0
\end{bmatrix}
\begin{bmatrix}
0 \\
0 \\
0 \\
0
\end{bmatrix}
\begin{bmatrix}
0 \\
0 \\
0 \\
0
\end{bmatrix}
\begin{bmatrix}
0 \\
0 \\
0 \\
0
\end{bmatrix}
\]
2. Refine horizontal ($\theta = 0^\circ$) estimates (assume $g = \frac{1}{2}$)

\[
\begin{align*}
&\begin{bmatrix} w \\ x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 + (12 - 0)/2 \\ 0 + (12 - 0)/2 \\ 0 + (8 - 0)/2 \\ 0 + (8 - 0)/2 \end{bmatrix} = \begin{bmatrix} 6 \\ 6 \\ 4 \\ 4 \end{bmatrix} \\
&\begin{bmatrix} 6 & 6 & 12 \\ 4 & 4 & 8 \end{bmatrix} \\
&10 \quad 10 \quad 10 \quad 10
\end{align*}
\]

3. Refine vertical ($\theta = 90^\circ$) estimates

\[
\begin{align*}
&\begin{bmatrix} w \\ x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 6 + (11 - 10)/2 \\ 6 + (9 - 10)/2 \\ 4 + (11 - 10)/2 \\ 4 + (9 - 10)/2 \end{bmatrix} = \begin{bmatrix} 6.5 \\ 5.5 \\ 4.5 \\ 3.5 \end{bmatrix} \\
&\begin{bmatrix} 6.5 & 5.5 & 12 \\ 4.5 & 3.5 & 8 \end{bmatrix} \\
&10 \quad 11 \quad 9 \quad 10
\end{align*}
\]

4. Refine diagonal ($\theta = \pm 45^\circ$) estimates

\[
\begin{align*}
&\begin{bmatrix} w \\ x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 6.5 + (7 - 10)/2 \\ 5.5 + (13 - 10)/2 \\ 4.5 + (13 - 10)/2 \\ 3.5 + (7 - 10)/2 \end{bmatrix} = \begin{bmatrix} 5 \\ 7 \\ 6 \\ 2 \end{bmatrix} \\
&\begin{bmatrix} 5 & 7 & 12 \\ 6 & 2 & 8 \end{bmatrix} \\
&13 \quad 11 \quad 9 \quad 7
\end{align*}
\]
We are now done; the image data and projections match.

The ART was the original reconstruction method used in medical imaging. The method works, but is slow and susceptible to noise. With more noise, the convergence is very slow, and is not guaranteed.

2.4.3 Backprojection Method

It is obvious that we desperately need a rigorous mathematical answer to the burning question that we all have: given the sinogram (family of projections) \( p(r, \theta) \), what is the original image \( f(x, y) \)? As \( p(r, \theta) \) is the Radon transform of \( f(x, y) \), this means that what we actually need is an expression of the inverse Radon transform

\[
f(x, y) = \mathcal{R}^{-1} \{ p(r, \theta) \}
\] (2.11)

The projection theorem, also called the central slice theorem or the central section theorem, provides an answer to this burning question. Let \( F(k_x, k_y) \) be the 2D Fourier transform of \( f(x, y) \)

\[
F(k_x, k_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \exp\left\{-2\pi j (k_x x + k_y y)\right\} dx dy
\] (2.12)

The variables \( k_x \) and \( k_y \) are orthogonal frequencies with units cycles (or radians) per distance. Let \( P_\theta(k) \) be the 1D Fourier transform of the projection \( p_\theta(r) \)

\[
P_\theta(k) = \int_{-\infty}^{\infty} p_\theta(r) \exp\left\{-2\pi j (kr)\right\} dr
\] (2.13)

The variable \( k \) is frequency (cycles or radians per distance). Allowing \( \theta \) to vary from 0 to \( \pi \) and stacking all \( P_\theta(k) \) together, we obtain a 2D dataset \( P(k, \theta) \). The projection theorem (or central slice theorem-CST) then states that

\[
P(k, \theta) = F(k_x, k_y)
\] (2.14)

with

\[
\begin{align*}
k_x &= k \cos \theta \\
k_y &= k \sin \theta \\
k &= \sqrt{k_x^2 + k_y^2}
\end{align*}
\] (2.15)

In words, the set of 1D Fourier transform in \( r \) of the Radon transform of a function is the 2D Fourier transform of that function. This means that if all the Fourier
transforms of the projections $P_\theta(k)$ of a function $f(x, y)$ are known, then it is possible to calculate the value of the function $f(x, y)$ for each point in the plane by calculating the inverse Fourier transform of $P(k, \theta)$. Proof of the CLT is straightforward (but a bit tedious) and in all CT textbooks.

### 2.4.4 Fourier Reconstruction

The computational problem with the Central Section Theorem is that a 2D inverse transform is required. In addition, CT requires various coordinate system shifts and interpolations that complicate the calculations even further. To make this point more clear, consider the following outline of the algorithm thus far. The reconstruction scheme based on the CST consists of the following steps:

1. Calculate the 1D Fourier transform of all projections $F \{ p_\theta(r) \} = P_\theta(k)$.
   Practically this is the FFT of the projection function in the $\theta$ direction, for all $\theta$.
2. Place all $P_\theta(k)$ on a polar grid to obtain $P(k, \theta)$, which after resampling in Cartesian space, is equal to $F(k_x, k_y)$

![Diagram](image)

Figure 2.10. Polar (a) and Cartesian (b) sampling grid.

3. Calculate the 2D inverse Fourier transform of $F(k_x, k_y)$ to yield $f(x, y)$.
   Practically this is the inverse FFT along each row (or column) for each column (or row). The result is the desired image.

In practice, re-sampling from polar to rectangular coordinates involves considerable interpolation, which makes the resultant estimate of $f(x, y)$ noisy. In other words, the spatial interpolation produces enough noise in the image to make this direct implementation of the CST unacceptable for clinical diagnostic purposes. There are two
ways around this problem. The first is to use smaller values for $\Delta \theta$ and $\Delta s$, which will decrease the noise introduced by the spatial interpolation scheme. Unfortunately, this will also increase the hardware requirements of the machines, it will cause many more calculations to be performed, both of which can lead to round-off and truncation errors. The second method is to use the filtered backprojection method.

### 2.4.5 Filtered Backprojection (FBP)

#### 2.4.5.1 Back Projection

For reference, Figure 2.5 is reproduced below:

![Figure 2.5 (repeated for reference)](image)

In back projection the measurements obtained at each projection are projected back along the same line (or equivalently same $\theta$). Thus the measured values are “smeared” across the unknown density line as if a line of wet ink is drawn across the line. This is illustrated in Figure 2.11 for the case of an object consisting of a single point.
Figure 2.11 Projections of a point at the origin back projected.

In this illustration, each projection is identical. Mathematically the back projection of a single measured projection along an unknown density is given by:

\[
b_b(x, y) = \int p_\theta(r) \delta(x \cos \theta + y \sin \theta - r) dr
\]

where \(b_b(x, y)\) is the back-projected density due to the projection \(p_\theta(r)\). Essentially we are stating that we know that the point of density is somewhere along the line so that a “crude” reconstruction will result if we assign the measured value along the entire line. The \(\delta(\cdot\cdot)\) function is the Dirac delta function that has the following properties:

\[
\delta(r) = \begin{cases} 1 & \text{if } r = 0 \\ 0 & \text{otherwise} \end{cases}
\]

Adding up these densities at all angles, we obtain the laminogram:

\[
f_\theta(x, y) = \int_0^\pi b_b(x, y) d\theta
\]

\[
= \int_0^\pi \int_0^\infty p_\theta(r) \delta(x \cos \theta + y \sin \theta - r) dr d\theta
\]

The laminogram represents a distorted picture of the anatomy through which the X-ray beam passes. We will now correct this distortion.

Each projection of a delta function at the origin is also a delta function. If these delta functions are back-projected, the resulting response (known as the impulse response) is simply:
The result is derived after knowing something about generalized functions; however, the lack of knowledge of generalized functions does not invalidate the result, which is correct by all accounts. Operations on generalized functions can be “esoteric.” For example, generalized functions predict that

\[ \delta[f(x)] = \sum \delta(x-x_n) \left| f'(x_n) \right| \]

where \( x_n \) are the roots of \( f(x) \).

Knowing that the response to an impulse is \( \frac{1}{r} \) and assuming linear system theory applies, then the following relationship holds:

\[ f_b(x,y) = f(x,y)*\frac{1}{r} \]  \hspace{1cm} (2.20)

The double asterisk (**) signifies convolution. Practically this means that the actual image \( f(x,y) \) is blurred by the \( \frac{1}{r} \) term. Thus, the image that one calculates with the backprojection method (the laminogram) is a blurred version of the actual image. The early reconstructions were based on Equation (2.20), with marginal results at best.

The Fourier transform of Equation (2.20) is

\[ F_b(k,\theta) = \frac{F(k,\theta)}{k} \]  \hspace{1cm} (2.21)

since the 2D Fourier transform of \( \frac{1}{r} \) is \( \frac{1}{k} \).

The \( \frac{1}{r} \) blurring must be removed. There have been two methods for accomplishing this task. The first is to calculate the 2D Fourier transform of Equation (2.20), weight the result with the 2D Fourier transform of the reciprocal of \( \frac{1}{r} \), and then perform the inverse 2D Fourier transform of the product. The computational requirements and potential for noise introduction (due to re-sampling, etc.) are enormous. The second method is to use the filtered backprojection algorithm.
2.4.5.2 Filtered backprojection method

It is desirable to be able to use the elegant simplicity of back projection and to undo the \(1/r\) blur without requiring 2D transforms. This is accomplished with our powerful friend, the Central Section Theorem.

We begin by restating the back-projection relationship for the laminogram \(f_b(x, y)\) as

\[
f_b(x, y) = \int_0^\pi \int_-\infty^\infty p(r, \theta) \delta(x \cos \theta + y \sin \theta - r) dr d\theta \tag{2.22}
\]

Restructuring this into a Fourier transform mode and applying the CST to substitute the inverse transform of \(P(k, \theta)\)

\[
f_b(x, y) = \int_0^\pi \int_-\infty^\infty \left[ \int_-\infty^\infty F(k, \theta) \exp(i2\pi kr) dk \right] \delta(x \cos \theta + y \sin \theta - r) dr d\theta \tag{2.23}
\]

Performing the integration over \(r\), we obtain

\[
f_b(x, y) = \int_0^\pi \int_-\infty^\infty F(k, \theta) \exp\left[j2\pi k(x \cos \theta + y \sin \theta)\right] dk d\theta \tag{2.24}
\]

This equation is actually quite important. To appreciate this statement, we can rewrite the 2D Fourier transform in polar form, yielding

\[
f(x, y) = \int_0^{2\pi} \int_0^\infty F(k, \theta) \exp\left[j2\pi k(x \cos \theta + y \sin \theta)\right] kdk d\theta \tag{2.25}
\]

This can be modified to conform to the form of Equation (2.24) yielding

\[
f(x, y) = \int_0^{2\pi} \int_0^\infty F(k, \theta) \exp\left[j2\pi k(x \cos \theta + y \sin \theta)\right] |k|dk d\theta \tag{2.26}
\]

It is necessary to use \(|k|\) instead of \(k\) because the integration includes negative values.

Comparing Equation (2.26) with Equation (2.24), we see that they differ only by the \(|k|\) weighting term, as expected. Substituting \(\mathcal{F}\{\rho_\theta(r)\}\) for \(F(k, \theta)\) in Equation (2.24), and multiplying and dividing by \(|k|\), we obtain
\[
 f_b(x, y) = \int_0^\infty \int_{-\infty}^{\infty} \mathcal{F}\{g_{\delta}(r)\} \exp\left[j 2\pi k (x \cos \theta + y \sin \theta)\right] |k| \, dk 
\]

This equation states that the transform of each projection has been weighted by $|k|$ along each radial line in Figure 2.5 (or Figure 2.11). This accounts for the blurred reconstruction of $f_b(x, y)$. The “fix” for this blurring is to pre-weight the transform of the projection prior to back projecting, that is

\[
 f(x, y) = \int_0^\infty \int_{-\infty}^{\infty} \mathcal{F}\{p_{\delta}(r)\} \exp\left[j 2\pi k (x \cos \theta + y \sin \theta)\right] |k| \, dk 
\]

This reconstruction is known as the filtered back-projection system, and is the current standard algorithm for calculating a tomographic image. This system involves only a one-dimensional Fourier transform that is fast (because the FFT algorithm is actually used).

The scheme is:

1. Each projection is individually transformed,
2. The transformed projection is weighted with $|k|$,
3. The result is inverse transformed, and
4. The result is back projected.

### 2.4.5.3 Discrete FBP

Of course digital computers require discrete representations of Equation (2.28). Going from continuous space to discrete space is usually straightforward: $\int$ become $\sum$, $d\theta$ becomes $\Delta \theta$, and $dk$ becomes $\Delta k$. Conversion would be straightforward if it were not for the pesky $|k|$ term. This term is the ramp filter, which is not stable (there are values of $k$ for which things are not well behaved). The instability makes this filter unusable in practice. To see this, consider the inverse Fourier transform of the filter:

\[
 F^{-1}\{|k|\} = \int_{-\infty}^{\infty} |k| \exp\left(j 2\pi kr\right) dk 
\]

\[
 = -\frac{1}{2\pi^2 r^2} 
\]

Note that the inverse Fourier transform of the filter has a singularity at $r=0$. Is this important? Unfortunately it is; the CT image “blows up” for each $\theta$, which is bad if the goal is to derive an image that faithfully reproduces the anatomy.
What to do? We know from our discussion on sampling that the measured data are limited to frequencies below the Nyquist frequency $k_N$ as shown in Figure 2.12(a).

![Figure 2.12: Decomposition of Ram-Lak filter (a) into the difference between a block (b) and a triangle (c).](image)

The resulting filter (graphed in Figure 2.12(a)) is known as the Ram-Lak filter, named after its inventors Ramachandran and Lakshminarayanan. The Ram-Lak filter can be written as the difference of a block and a triangle (Figure 2.12(b) and (c)). The inverse Fourier transform then gives the following:

$$q(r) = \frac{k_N \sin(2\pi k_N r)}{\pi r} - \frac{1 - \cos(2\pi k_N r)}{2\pi^2 r^2}$$  \hspace{1cm} (2.30)

The resulting filter function has no singularity at $r=0$. To make the filter discrete, substitute $r_n = n\Delta r$ and $k_n = 1/(2\Delta r)$. Thus Equation (2.30) becomes:

$$q(r_n) = \begin{cases} 
\frac{1}{4(\Delta r)^2} & \text{if } n=0 \\
0 & \text{if } n \text{ is even} \\
-\frac{1}{(\pi n(\Delta r))^2} & \text{if } n \text{ is odd}
\end{cases}$$  \hspace{1cm} (2.31)

Usually frequencies below $k_N$ are corrupted by aliasing and noise. Applying a smoothing filter (usually the Shepp-Logan filter, or other low-pass filters such as Hann, Hamming, or other) suppresses the highest spatial frequency and reduces noise. A generalized Hamming window is given by

$$H(k) = \begin{cases} 
\alpha + (1-\alpha) \cos \frac{\pi k}{k_N} & \text{for } |k| \leq k_N \\
0 & \text{for } |k| \geq k_N
\end{cases}$$  \hspace{1cm} (2.32)
For $\alpha = 0.5$, we have the Hann (or Hanning) window, and for $\alpha = 0.54$, we have the Hamming window. Figure 2.13 shows the results of applying the Hann smoothing filter to the Ram-Lak filter.

![Figure 2.13 Product of a smoothing filter and the Ram-Lak filter.]

After all of this development, the complete formula of discrete FBP can be written by combining a discrete convolution using the kernel in Equation (2.31) and a discrete back projection given in Equation (2.22) as

$$f(x_i, y_j) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} p(\theta_m, r_n) \cdot q \left( x_i \cos \theta_m + y_j \sin \theta_m - r_n \right) \Delta r \Delta \theta$$

(2.33)

### 2.4.5.4 Implementation

Actual implementation of the FBP algorithm might follow this outline (assuming parallel projections):

1. Design a smoothing filter (Shepp-Logan, Hann, Hamming, or other) in the Fourier (frequency) domain;
2. Calculate 1D FFT of the projections for each $\theta$;
3. Filter the projections by multiplying the filter with the Fourier transformed projections (remember: convolution in the space domain is multiplication in the frequency domain);
4. Determine image size; this involves either a desired size, or a calculation of the diagonal similar to $\text{diag} = 2 \cdot \text{ceil}(N/\sqrt{2}) + 1$, where $N$ is the specified number of rows, or if unspecified, $N = 2 \cdot \text{floor(size} (P, 1)/(2 \cdot \sqrt{2}))$ where $P$ is the projection matrix;
5. Back-project; this may involve interpolation if the size of the image is different from the size of the projection matrix. The interpolation scheme could be nearest neighbor, linear, spline, or other;
6. Scale the image elements (pixels) if necessary;
7. Add ad hoc image processing to prepare the image for ultimate clinical use. This ad hoc processing is usually vendor specific, and often defines the individual characteristics of the particular CT machine.
2.4.6 Fan beam FBP

The CT equipment has undergone various “generations” of development as illustrated in Figure 2.14.

The first generation was Hounsfield’s configuration and consisted of a pencil beam source and detector, and then linear translation across the patient. The second generation consisted of a small fan beam source, linear detector, and translation and rotation about the patient. The third generation consisted of a wider fan beam (about 30°) where the fan beam could cover the entire object under study. This generation of machines acquired data with rotation only; no translation was necessary. The fourth generation of machine consisted of a rotating fan beam, and a stationary detector array.

There is a “fifth generation” machine that consists of an electronically steered electron beam swept across a target ring of tungsten. This electron-beam CT (EBCT) shows high spatial and temporal resolution – there are no moving parts. This is illustrated in Figure 2.15.
From Imatron’s literature:

Single Slice mode scanning utilizes 100 millisecond sweeps along one of the four target rings. This sequence provides two distinct acquisition features: Continuous Volume Scanning (CVS) and Step Volume Scanning (SVS). In a CVS sequence data is acquired by moving the patient continuously through the x-ray beam path. In a SVS sequence timed scans are acquired while the patient moves incrementally through the scanner. In SVS diagnosticians can utilize real time ECG triggering to provide a volume acquisition of the heart at a prescribed phase of the cardiac cycle. The ability of the ECG to trigger the Imatron EBT scanner at high heart rates, without skipping a beat, provides extremely high temporal resolution and high patient compliance without exposure to unnecessary dose.

In Multi-Slice sequence scanning, the electron beam sweeps along one to four target rings in rapid succession covering up to 8 centimeters of the body without moving the patient through the scanner. This unique functionality, along with real time ECG cardiac triggering provides cine and flow capabilities. This very fast scanning speed and temporal resolution is ideal for evaluating ventricular function, joint motion, airway physiology and blood flow analysis. The exceptional scan speed of the Imatron EBT scanner provides the diagnostician with critical diagnostic information when imaging cardiac, pediatric, trauma and geriatric patients where speed of acquisition is essential.

The images from the EBCT machine are impressive. For example, the Left Anterior Descending (LAD) coronary artery looks like:
The reconstruction algorithms that we have developed are all based on the assumption that the data have been acquired with parallel-beam geometry. Data from modern machines are actually not ordered in parallel subsets, but rather in fans as shown in Figure 2.17.

The fan angle is the formed by the fan. How can the mathematical models developed in the previous section be used in this new geometry? Three are two approaches;
1. Record the fan beam data, and interpolate into parallel data. This process is known as *re-binning*. These data then require the same algorithm as developed above.

2. Adopt the FBP algorithm. This approach uses a modified convolution back-projection system of Gullberg and Denton (see refs.). Here the convolution kernel is slightly different, and the back projection involves a quadratic weighting factor rather than the uniform weighting of the parallel rays. This approach is widely used in existing scanners.

### 2.5 CT number

The reconstruction methods studied so far indicate how a 2D function can be reconstructed from its line integrals. These line integrals are the sum of the function in different directions. In the case of X-ray attenuation, however, we measure the exponent of the desired line integral as

$$I = I_0 e^{-\int \mu(z) dz} \quad (3.1)$$

Note that there is a nonlinear relationship between the measured projections and the desired line integral. This nonlinearity is removed if we use the log of the measured transmission where

$$\ln \frac{I_0}{I} = \int \mu(z) dz \quad (3.2)$$

Unfortunately, this is not going to work. Equation (3.1) was derived with the assumption that the source is mono-energetic, and the beam is infinitesimally small. Most machines compensate for this distortion on an *ad hoc* basis by using a nonlinear function of the log of the measurements.

All machines report derived linear attenuation coefficients using a *CT number*. The CT number of a pixel with attenuation coefficient $\mu_x$ is:

$$\text{CT Number} = K \frac{\mu_x - \mu_{\text{water}}}{\mu_{\text{water}}} \quad (3.3)$$

If the scale factor $K=1000$, then the units of the CT Number is Hounsfield Units (HU). The following table shows data for various materials.
Table 2.1 CT Number and $\mu$ for various materials at 60 keV.

<table>
<thead>
<tr>
<th>Material</th>
<th>CT Number (HU)</th>
<th>$\mu$ (cm$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bone</td>
<td>808</td>
<td>0.38</td>
</tr>
<tr>
<td>Water</td>
<td>0</td>
<td>0.21</td>
</tr>
<tr>
<td>Muscle</td>
<td>-48</td>
<td>0.20</td>
</tr>
<tr>
<td>Fat</td>
<td>-142</td>
<td>0.18</td>
</tr>
<tr>
<td>Air</td>
<td>-1000</td>
<td>0.0</td>
</tr>
</tbody>
</table>

2.6 X-ray detectors

There are three kinds of X-ray detectors used in modern CT equipment. These are scintillation detectors, gas ionization chambers, and solid-state (semiconductor) detectors.

Scintillation detectors, used in the earliest CT machines, consist of a scintillation crystal (NaI, CaF$_2$) in combination with a photo-multiplier tube. The advantages of this technique are high quantum efficiency, and fast time response. The disadvantage is the low packing density.

Gas ionization chambers consist of a pressurized gas chamber (Xe or Kr) with two electrodes. Some of the incident X-ray photons cause ionization of the gas (by the photo-electric absorption). The resulting free electrons and ions drift towards the anode and cathode, giving rise to measurable electric signals. Gas ionization chambers have a low quantum efficiency and slow response time. The advantage is high packing density.

The newest scanners typically use semiconductor detectors. A solid-state conductor consists of a scintillation crystal (CsI) and a photo-diode. The photo-diode is based on a pn-junction. When a scintillation photon passes through the depletion layer (boundary layer containing no mobile carriers), electron-hole pairs are produced, giving rise to a measurable electric current. These detectors have a very fast response time, high quantum efficiency, a good stability, and good reliability. On the other hand, they are susceptible to scattered radiation and thus require external collimation.

2.7 Imaging in three dimensions

2.7.1 Multiple 2D acquisitions

The most straightforward way to image an entire volume is to scan number of consecutive slices; this is called a sequential scan. This is illustrated in Figure 2.18(a) for a 360° acquisition.
\( \beta \) represents the angular position of the X-ray tube, and \( z \) its axial position relative to the patient. Data are acquired for discrete axial positions \( z_i \) and for angular tube positions \( \beta \) ranging from 0 to \( 2\pi \).

In order to acquire a complete 3D data set containing all information needed to calculate a reconstruction, and in order to not miss any lesions or other anatomical details, the axial sampling density must obey the Nyquist criterion. Let’s assume a rectangular slice thickness (or profile) of \( \Delta z \). This means that the real data is convolved with an axial block of width \( \Delta z \). We come to the conclusion that the maximum distance between slices is \( 2\Delta z \), or in words, we need to acquire at least two slices per slice thickness.

### 2.7.2 Spiral CT

A technique that is widely used is spiral (or helical) CT. The X-ray source rotates continuously around the patient, but simultaneously the patient is slowly translated through the gantry. The tube describes a helical orbit relative to the patient. Figure 2.18(b) illustrates this type of data collection scheme. In this figure, data are simultaneously collected in \( \beta \) and in \( z \). The table feed (TF in the figure) is the axial distance over which the table translates during the time needed by the source to complete a 360° orbit. The pitch ratio is the ratio between the table feed and the slice thickness.

In practice, a pitch ratio between 1.0 and 2.0 is used; for example, a slice thickness of 2 mm is combined with a table feed of 3 mm (pitch=1.5). A higher pitch, and thus faster scanning, has several advantages: lower dose, less tube load, and less motion artifacts.

### 2.7.3 Cone-beam CT

Some experimental machines have a 2D detector array in order to measure the entire volume on one single orbit of the X-ray source. Unfortunately, this technique has not lived up to the original expectations. This technique could use some smart biomedical engineer to continue its development.
2.8 Image quality

2.8.1 Image contrast

Image contrast depends on several factors, the major factors being:

1. Tube current (mAs) per slice (the more current supplied to the tube, the better the contrast, but also the more radiation delivered to the patient).
2. Spectrum of the X-ray tube. The effective linear attenuation coefficient of the tissue depends on the incident spectrum and therefore the contrast depends on the incident spectrum. The narrower the spectrum (more like mono-energetic photons) the better the contrast.
3. The dynamic range of the hardware limits the contrast.
4. Nonlinearities such as scatter, etc. limit the image contrast.

2.8.2 Spatial resolution

The spatial resolution of a CT image depends on several factors, the most important being:

1. Size of the focal spot, which is the area on which the electrons hit the anode.
2. Size of the detector elements; the size of the detector elements and the size of the focal spot determine the beam width.
3. Pre- and post-patient (axial) collimators, which determine the slice thickness.
4. Reconstruction kernel (or convolution filter), which can be tuned to enhance high frequencies (for determining edges), or low frequencies (for attenuate noise and other artifacts)
5. Better resolution can be obtained by:
   a. Increasing the number of detectors,
   b. Using a smaller pixel (or voxel) size, at the expense of more noise,
   c. Using a reconstruction kernel with less smoothing, at the expense of more noise.

A typical modern CT machine has sub-mm in-plane resolution, and up to 1 mm axial resolution.
Figure 2.19 Example of image artifacts: (a) test phantom, (b) second phantom, (c) noise, (d) detector under-sampling, (e) view under-sampling, (f) beam hardening, (g) scatter, (h) nonlinear partial volume effect, and (i) object motion. (unpublished results)

2.8.3 Noise

The main source of noise is quantum noise. The number of photons passing through the patient and striking the detectors is susceptible to quantum fluctuations, and can be described by the Poisson distribution. The variance of the number of photons is proportional to the actual number of photons; therefore an increase in the number of photons used to improve contrast will result in more noise. Increasing the energy of the X-ray beam can reduce noise.

Noise in the projections in CT exhibits a characteristic structure. A high noise level results in alternating dark and light thin streaks radiating from locations with high attenuation (bright areas) as shown in Figure 2.19(c).
2.8.4 Artifacts

2.8.4.1 Under-sampling

Taking too few samples results in two distinct phenomena (shown in Figure 2.19):

1. Aliasing due to the number of detectors being too small.
2. If the number of views is too small, then streaks appear.

2.8.4.2 Beam hardening

Low energy photons are preferentially absorbed. This causes the X-ray beam too “harden” as it passes through the body. As a consequence, hardening artifacts appear, such as “cupping” and streaks (Figure 2.19(f)).

2.8.4.3 Partial volume effects

Due to a finite beam aperture, every measurement represents an intensity that is averaged over the beam aperture. This average intensity is log-converted to give an effective beam attenuation. This value is always an underestimate of the actual attenuation. The larger the attenuation gradients perpendicular to the beam, the larger the underestimate. This results in streaks tangent to edges. This phenomenon is illustrated in Figure 2.19(h).

Another example is shown in Figure 2.20 in which the white arrows point to the streak artifacts due to the partial volume effect. The large white “blob” is highly absorbent contrast material in the patient’s stomach.
Figure 20. Image with artifactual streaks due to the partial volume effect.
3. Appendix I – Fourier analysis

3.1 Introduction:

The development of Fourier analysis has a long history involving many individuals. The Babylonians used "trigonometric series" to predict astronomical events. In 1748 L. Euler investigated the motion of a vibrating violin string and found that the vibrations along the length of the string were periodic, and the magnitude of the vibrations was sinusoidal. Euler performed the same calculations that we will perform in the next section. In 1753 Bernoulli criticized the use of trigonometric series to represent vibrations on strings. His criticism was based on his own belief that it was impossible to represent signals with corners (i.e., discontinuous slopes) using trigonometric series.

Jean Baptiste Joseph Fourier entered this controversy around 1800. Fourier was interested in diffusing the heat generated in cannons used by Napoleon Bonaparte. In 1807 Fourier submitted a manuscript for publication describing the Fourier Series and Fourier Transform. The manuscript went to four reviewers, including Laplace and Lagrange. Only Lagrange objected to the theory that signals can be represented as trigonometric series. The manuscript was never published. The theory was finally given the "light of day" in Fourier's book *The Analytical Theory of Heat* published in 1822. The ideas offered by Fourier have had an enormous impact in engineering, science, and mathematics. Circuits are design based on Fourier analysis. Blood pressure transducers are designed based on Fourier analysis. The list of engineering systems designed based on Fourier analysis is never-ending; the list grows daily.

3.2 Response of LTIL system to eigenfunctions

It is easy to show that if the input to a linear, time-invariant, lumped (LTIL) system is a sinusoid, then the output will also be a sinusoid of the same frequency, but scaled in magnitude and shifted in phase. If the input to a LTIL system is a complex exponential, the output is also a complex exponential with only a change in (possibly complex) amplitude. A signal for which the system output is a (possibly complex) constant times the input is referred to as an eigenfunction of the system, and the amplitude factor is referred to as the system's eigenvalue.

We know that a LTIL system is always characterized by the convolution of its impulse response $h(t)$ and the input $u(t)$. Let's let $u(t) = \exp(st)$. Thus we have:

$$y(t) = \int_{-\infty}^{\infty} h(\tau)e^{st-\tau} d\tau = e^{st} \int_{-\infty}^{\infty} h(\tau)e^{-st} d\tau$$

(4.1)

By the definition of the Laplace transform, this becomes:

$$y(t) = H(s)e^{st}$$

(4.2)
The function $H(s)$ is a complex constant the value of which depends on a complex number $s=\sigma+j\omega$. The value of $H(s)$ and the system impulse response are a Laplace Transform pair. Thus, the constant $H(s)$ for a specific value of $s$ is the eigenvalue associated with the eigenfunction $\exp(st)$.

We can do an example to illustrate this concept. Let

$$u(t) = c_1 e^{\sigma t} + c_2 e^{\sigma t} + c_3 e^{\sigma t}$$

Then by superposition we know that

$$y(t) = c_1 H(s_1) e^{\sigma t} + c_2 H(s_2) e^{\sigma t} + c_3 H(s_3) e^{\sigma t}$$

Extending this concept leads to a convenient expression for the response of a LTIL system. If the input is written as

$$u(t) = \sum_n c_n e^{s_n t}$$

then the output will be written as

$$y(t) = \sum_n c_n H(s_n) e^{s_n t}$$

Thus if the input can be represented as a series of complex exponentials, then the output can also be represented as a series of complex exponentials. Furthermore, each coefficient in this representation of the output is obtained as the product of the corresponding coefficient of the input $c_n$ and the system’s eigenvalue $H(s_n)$.

### 3.3 Fourier Analysis of Periodic Signals

The value of the variable $s$ in the above treatment is an arbitrary complex number. In Fourier analysis the value of $s$ is restricted to purely imaginary numbers, that is, $s=j\omega$. Thus in Fourier analysis only complex exponentials $\exp(j\omega t)$ are considered. We know however that complex exponentials are related to real functions. Euler proved that

$$e^{j\omega t} = \cos(\omega_0 t) + j \sin(\omega_0 t)$$

Thus, complex exponentials are periodic with period $T_0$ given by
\[ T_0 = \frac{2\pi}{\omega_0} = \frac{1}{f_0} \]  

(4.8)

Let’s look at an example of what this means.

Consider a periodic signal \( x(t) \) with fundamental frequency \( \omega_0 \) of \( 2\pi \). Let’s express the signal as

\[
x(t) = \sum_{k=-3}^{3} c_k e^{j2\pi k t} \tag{4.9}
\]

where \( c_0=1, \ c_1=c_1=1/4, \ c_2=c_2=1/2, \) and \( c_3=c_3=1/3 \). Substituting these values and collecting terms, we get

\[
x(t) = 1 + \frac{1}{24} (e^{j2\pi} + e^{-j2\pi}) + \frac{1}{2} (e^{j4\pi} + e^{-j4\pi}) + \frac{1}{3} (e^{j6\pi} + e^{-j6\pi}) \tag{4.10}
\]

Using Euler’s relation in Eq. 7, Eq. 10 can be written as

\[
x(t) = 1 + \frac{1}{2} \cos 2\pi t + \cos 4\pi t + \frac{1}{3} \cos 6\pi t \tag{4.11}
\]

Thus we see that the summation of a series of complex exponentials is equivalent to the summation of real periodic signals.

To generalize this result, Fourier showed that \( x(t) \) can be written as

\[
x(t) = \sum_{n=-\infty}^{\infty} c_n e^{-j\omega_0 n t} \tag{4.12}
\]

The constant \( c \) is, in general, complex. This can be rewritten as

\[
x(t) = c_0 + \sum_{n=1}^{\infty} [c_n e^{j\omega_0 n t} + c_{-n} e^{-j\omega_0 n t}] \tag{4.13}
\]

This can be further simplified by writing the coefficient \( c_n \) in rectangular form as \( c_n = A_n + jB_n \), where \( A_n \) and \( B_n \) are real. In this case, Eq. 13 can be rewritten as:
\[ x(t) = A_0 + 2 \left[ \sum_{n=1}^{\infty} A_n \cos n\omega_0 t - \sum_{n=1}^{\infty} B_n \sin n\omega_0 t \right] \]  

(4.14)

This is usually rewritten in the more familiar form

\[ x(t) = a_0 + \sum_{n=1}^{\infty} a_n \cos n\omega_0 t + \sum_{n=1}^{\infty} b_n \sin n\omega_0 t \]  

(4.15)

Equations 13 and 15 are equivalent if the following holds

\[ c_n = \frac{a_n}{2} - j \frac{b_n}{2} \]  

(4.16)

This is the Fourier Series in the trigonometric form. The fundamental frequency is \( \omega_0 \) rads/sec. The frequency \( n\omega_0 \) rads/sec is said to be the \( n \)-th harmonic frequency. The Fourier coefficients \( a_n \) and \( b_n \) are sometimes called the spectral coefficients of \( x(t) \). These coefficients are a measure of the portion of the signal \( x(t) \) that is at each harmonic of the fundamental component.

Dirichlet proved that under certain conditions, \( x(t) \) equals its Fourier Series representation, except at isolated values of \( t \) for which \( x(t) \) is discontinuous. At these values, the infinite series converges to the average value on either side of the discontinuity.

The Dirichlet conditions for the existence of the Fourier Series are:

1. Over any period, \( x(t) \) must be absolutely integrable, that is, the integral of \( x(t) \) over one period must be finite.
2. In any finite interval of time, \( x(t) \) must have a finite number of maxima and minima.
3. In any finite interval of time, there are only a finite number of discontinuities.

Parseval showed that for a continuous-time periodic signal \( x(t) \), then

\[ \frac{1}{T} \int_{-T/2}^{T/2} |x(t)|^2 \, dt = \sum_{k=-\infty}^{\infty} |a_k|^2 \]  

(4.17)

The term on the left-hand side of this equation is the average power (energy per unit of time) in one period of the signal \( x(t) \). Thus, Parseval's theorem proved that the total average power in a periodic signal equals the sum of the average powers in all of its harmonic components. Stated in another way, the average power of a periodic signal is
conserved across the Fourier Series. Thus, one can work in the time domain or the frequency domain, and power will not be lost.

### 3.4 Calculating the Fourier Coefficients

In order to take advantage of the power of Fourier analysis, the unknown coefficients $a_n$ and $b_n$ must be found. There are at least two methods that can be used to find the coefficients, but in practice, one uses the Fast Fourier Transform (more later). The original method of Fourier, however, was based on the orthogonality of trigonometric functions. Two functions are said to be orthogonal if their inner product is zero. For example,

\[
\int_{T} \cos n\omega t \sin m\omega t dt = 0 \quad (4.18)
\]

\[
\int_{T} \cos n\omega t \cos m\omega t dt = \begin{cases} 
0 & \text{if } m \neq n \\
\frac{T}{2} & \text{otherwise}
\end{cases} \quad (4.19)
\]

\[
\int_{T} \sin n\omega t \sin m\omega t dt = \begin{cases} 
0 & \text{if } m \neq n \\
\frac{T}{2} & \text{otherwise}
\end{cases} \quad (4.20)
\]

The Fourier Series is

\[
x(t) = a_0 + \sum_{n=1}^{\infty} a_n \cos n\omega_0 t + \sum_{n=1}^{\infty} b_n \sin n\omega_0 t \quad (4.21)
\]

If we multiply both sides of this equation by $x(t)$ and integrate over $T_0$ (one period) and take advantage of the fact that integrating a sine or cosine over one period is zero, then we get that

\[
a_0 = \frac{1}{T_0} \int_{0}^{T_0} x(t) dt \quad (4.22)
\]

Thus the first term in the Fourier Series is simply the average value of the waveform.

If we multiply both sides of the Fourier Series by $\cos(m\omega_0 t)$ and $\sin(m\omega_0 t)$, and take advantage of the orthogonality of sinusoids, we get:

\[
a_n = \frac{2}{T_0} \int_{0}^{T_0} x(t) \cos(n\omega_0 t) dt \quad (4.23)
\]
Thus we have an easy way to calculate the Fourier Series coefficients for any periodic waveform $x(t)$. Unfortunately there are two problems with this formulation. First most biomedical signals are not periodic. Second, performing this integration for each harmonic can be time consuming. This problem grows quadratically as $T_0$ increases.

### 3.5 Fourier Analysis of Long and Possibly Non-periodic Waveforms

Given a time function $f(t)$, one can predict the response of a linear system (filter) by first converting the time function to the frequency domain, and then altering the signal in the new domain. If the signal is periodic with period $T_0$ seconds, or if a computer is used to sample the function with a sampling window of $T_0$ seconds in width, then $f(t)$ can be written as a Fourier Series.

Euler's relations for the cosine and sine can be used to develop a practical approach for calculating the Fourier Series coefficients $c_n$. Remember that Euler said that

$$\cos x = \frac{\exp(jx) + \exp(-jx)}{2} \quad \text{and} \quad \sin x = \frac{\exp(jx) - \exp(-jx)}{2j}$$

Using the definition of $c_n$ in Eq. 16, the Fourier Series coefficients can be written as

$$c_n = \frac{1}{T_0} \int_0^{T_0} f(\tau) \exp(-j\omega_0 \tau) d\tau \quad (4.25)$$

The $c_n$ coefficients are found from solving this equation, or from the Fast Fourier Transform (FFT) algorithm.

The FFT algorithm was proposed by Cooley and Tukey in the early 1960's (see Brigham, E. O. “The Fast Fourier Transform,” Prentice-Hall, 1974). This algorithm calculates the complex coefficients $c_n$ at a blazingly fast speed. The algorithm is easily implemented in software using only a few multiplications and many shifts; this is the reason that the algorithm is fast. The algorithm is also easily implemented in hardware. One can buy an FFT on a chip for just a few dollars.
3.6 Practical Procedure for Performing Fourier Analysis

The procedure for handling large amounts of data, and for performing Fourier analysis of non-periodic signals, is as follows:

1. Sample the waveform for $T_0$ seconds at a sampling rate of $f_s$, samples per second. Thus the total number of samples is $N_p = T_0 f_s$. The value of $T_0$ is the sampling window. The fundamental frequency is that frequency that produces one complete cycle in this sampling window. Thus $T_0 = \frac{2\pi}{\omega_0} = \frac{1}{f_0}$.

2. Assemble the $N_p$ samples $data_0, data_1, \ldots$ into an array. The array will usually be complex with the imaginary parts all zero. The real parts will be the sampled data. Thus the data array will be

$$D = \begin{bmatrix} data_0 + j0 \\ data_1 + j0 \\ \vdots \\ data_{N_p-2} + j0 \\ data_{N_p-1} + j0 \end{bmatrix} \quad (4.26)$$

3. Calculate the values of $c_n$ using the FFT algorithm. By convention, one must scale the results (output) of the FFT by $\frac{1}{N_p}$, where $N_p$ is the number of points sampled in $T_0$ seconds. Thus the $c_n$ coefficients from the FFT must be multiplied by $\frac{1}{N_p}$, or equivalently, the points in the vector of data passed to the FFT algorithm must be scaled by $N_p$. In the conventional FFT algorithm, the number $N_p$ must be a power of two (2), such as 32, 64, 128, 256, 1024, etc. The results from the FFT will be arranged as $n=0$ to $N_p/2$, and in most implementations, these are followed in reverse order as $n=-(N_p/2)$ to -1. The coefficients from the FFT algorithm (assuming that scaling has already been applied) are thus arranged as a vector $C$ of complex numbers given by

$$C = \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_{N_p/2} \\ c_{N_p/2} \end{bmatrix} \quad (4.27)$$

Thus, only the first $N_p/2$ points in $C$ are useful. Once the $C$ coefficients (or equivalently the $a_n$ and $b_n$ coefficients) are obtained, one can reconstruct the time function from the Fourier Series equation (Eq. 20).
In sum, the FFT yields a scaled version of the $c_n$ coefficients. From these, one can filter the time signal $f(t)$ by altering the $c_n$ coefficients in the frequency domain. One can then use the Fourier Series representation to reconstruct the time waveform after filtering.

Often the values of $c_n$ are displayed as a graph with the magnitude of $c_n$ on the ordinate, and frequency on the abscissa. This graph is sometimes called the line spectrum.

There are many FFT algorithms available. All of the available FFT routines require that the number of points (samples) passed to the routine be a power of 2. If this is not the case, one can either pad the data sequence with zeros, or truncate the sequence to a power of two, or use the integration formula for calculating the $a$ and $b$ coefficients directly.

The IMSL library has an easy-to-use algorithm that is callable from FORTRAN, C, or C++. The MATLAB package has a discrete FT (DFT) algorithm that calculates the coefficients for sampled data. The corresponding frequencies of the DFT are summarized in the following table. The first two columns are digital frequencies, and the last two columns are analog frequencies. The two are related by $\omega = WT_k$ where $\omega$ is the analog frequency and $W$ is the digital frequency. If $x(t) = \sin(\omega t)$, then $x[k] = \sin(WT_k)$. Note that in the Table the variable $T_k$ is the time step (in seconds) between data samples, and the variable $N$ is the total number of samples. The frequency Hz-sec is obtained by multiplying the frequency in Hz by the time $T_k$, the inter-sample time, in seconds.

<table>
<thead>
<tr>
<th>Frequencies of DFT Components in Different Units</th>
<th>Hz-sec</th>
<th>Rad</th>
<th>Hz</th>
<th>rad/sec</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symbol</td>
<td>N</td>
<td>$\omega$</td>
<td>$f$</td>
<td>$W$</td>
</tr>
<tr>
<td>Frequency of $F_1$</td>
<td>$1/N$</td>
<td>$2\pi/N$</td>
<td>$1/NT$</td>
<td>$2\pi/NT$</td>
</tr>
<tr>
<td>Frequency of $F_m$</td>
<td>$m/N$</td>
<td>$2\pi m/N$</td>
<td>$m/NT$</td>
<td>$2\pi m/NT$</td>
</tr>
<tr>
<td>Frequency of $F_N/2$</td>
<td>0.5</td>
<td>$\pi$</td>
<td>$1/2T$</td>
<td>$\pi/T$</td>
</tr>
<tr>
<td>Sampling Frequency</td>
<td>1.0</td>
<td>$2\pi$</td>
<td>$1/T$</td>
<td>$2\pi/T$</td>
</tr>
</tbody>
</table>

The frequency of the final component, $F_{N/2}$, of the DFT of any real series is exactly one-half the sampling frequency in accordance with Shannon's Sampling Theorem. The period of the $F_m$ component is just the reciprocal of the frequency in all cases. Thus, the period of $F_m$ is $N/m$ samples or $NT/m$ seconds.
3.7 Frequency Analysis of Random Signals

The above analysis is not the same as finding the power spectrum. Spectral estimation is the determination of the spectral content of a random process based on a finite set of observations from the process. Formally, the power spectral density (PSD), which is denoted by $P_{xx}(f)$, of a wide-sense stationary random process $x_n$ is defined as (Kay, S.M. “Modern Spectral Estimation,” Prentice-Hall, Englewood Cliffs, 1988)

$$P_{xx}(f) = \mathcal{F}\{R_{xx}(t)\}$$  \hspace{1cm} (4.28)

$\mathcal{F}\{\}$ represents the Fourier transform. The function $R_{xx}(t)$ is the autocorrelation function given by

$$R_{xx}(t) = E\{f(\tau)^* f(\tau + t)\}$$  \hspace{1cm} (4.29)

where $E\{\}$ represents the expected value (expectation) operator. The superscript * represents the complex conjugate operator. The PSD represents the distribution of power with frequency of the random process.

For finite data lengths, the autocorrelation function cannot be found exactly - only estimates are available. For this case, the definition given for the autocorrelation function is changed to the following:

$$P_{xx}(f) = \frac{1}{M} \left| \mathcal{F}\{x(t)\} \right|^2$$  \hspace{1cm} (4.30)

This new definition says to take the magnitude squared of the Fourier transform of the data (calculated with the FFT) and then divide by the record length. This approach is the one used by MATLAB, which assumes that $M$, the number of samples, is large. There are many other methods for estimating the PSD. In contrast to the methods for estimating the line spectrum, no one method is universally accepted for estimating the PSD.

Often the values of $P_{xx}(f)$ are displayed on a graph with the magnitude (in dB) of the PSD on the ordinate and frequency (in Hz) on the abscissa. This is called the periodogram and is similar to a probability distribution function that describes how the power of a random signal is distributed over frequency. This is not the same as the line spectrum that describes how the time signal is decomposed into its sinusoidal components.

The area of the periodogram for positive frequencies gives the total power in the random signal. The units of power are squared units of the sampled data $x_n$. The units of power density are therefore power per hertz. It turns out that $R_{xx}(0)$ is the variance of the
random signal, which, if the signal is Gaussian, is also the total power. (Stearns, S.D., David, R.A. “Signal Processing Algorithms,” Prentice-Hall, Englewood Cliffs, 1988).

In sum, Fourier analysis is based on the idea that one can decompose a signal (either a function of time, or space) into the summation of sine and cosine waves. Fourier theory supports the statement that any continuous signal can be exactly written as the sum of a constant, an infinite number of weighted sine waves, and an infinite number of cosine waves. For example, if the signal $f(x)$ is periodic, then Fourier analysis states:

$$f(x) = a_0 + \sum_{n=1}^{\infty} a_n \cos(n\omega_0 x) + \sum_{n=1}^{\infty} b_n \sin(n\omega_0 x) \quad (4.31)$$

If $f(x)$ is not periodic, then one can write that the Fourier Transform of $f(x)$ is:

$$F(\omega) = \int_{-\infty}^{\infty} f(x) \exp(-j\omega x) \quad (4.32)$$

The notion of Fourier transform is graphically illustrated below in Figure A in which the complex signal is decomposed into its constituent sinusoids of different frequencies by applying either Equation (4.31) or Equation (4.32).

**Figure A – Decomposition of signal into summation of sine and cosine functions.**

If the signal is discrete, then the Fourier Transform becomes the Discrete Fourier Transform (DFT) that is given below (copied from MATLAB help files)

$$X(k) = \sum_{n=1}^{N} x(n)e^{-j2\pi(k-1)(n-1)/N} \quad 1 \leq k \leq N$$

$$x(n) = \frac{1}{N} \sum_{k=1}^{N} X(k)e^{j2\pi(k-1)(n-1)/N} \quad 1 \leq n \leq N$$
If the signal is real, as ours will most likely be, then the above equation becomes (again, from MATLAB):

\[ x(n) = \frac{1}{N} \sum_{k=1}^{N} a(k) \cos\left(\frac{2\pi(k-1)(n-1)}{N}\right) + b(k) \sin\left(\frac{2\pi(k-1)(n-1)}{N}\right) \]

where

\[ a(k) = \text{real}(X(k)), \quad b(k) = -\text{imag}(X(k)), \quad 1 \leq n \leq N \]

There is also a two-dimensional Fourier transform that is given by the following discrete equation (again copied from MATLAB):

\[ F(p, q) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m, n) e^{-j(2\pi/M)pm} e^{-j(2\pi/N)qn} \quad p = 0, 1, \ldots, M-1 \]
\[ q = 0, 1, \ldots, N-1 \]

\[ f(m, n) = \frac{1}{MN} \sum_{p=0}^{M-1} \sum_{q=0}^{N-1} F(p, q) e^{j(2\pi/M)pm} e^{j(2\pi/N)qn} \quad m = 0, 1, \ldots, M-1 \]
\[ n = 0, 1, \ldots, N-1 \]

As an example, consider the 2D figure below.

![Figure B – Test figure from MATLAB](image)
The result of applying the 2D DFT to the test figure is shown below.

![Image](image_url)

**Figure C – 2D DFT of object in Figure B.**

The actual calculation of the DFT is performed with an algorithm called the Fast Fourier Transform (FFT) invented by Tukey and Cooley. The MATLAB function `fft()` and `ifft()` calculate the forward FFT and inverse FFT, respectively. Try the following MATLAB code:

```matlab
load sunspot.dat
year = sunspot(:,1);
wolfer = sunspot(:,2);
plot(year,wolfer)
title('Sunspot Data')
Y = fft(wolfer);
N = length(Y);
Y(1) = [];
power = abs(Y(1:N/2)).^2;
nyquist = 1/2;
freq = (1:N/2)/(N/2)*nyquist;
figure, plot(freq,power), grid on
xlabel('cycles/year')
title('Periodogram')
period = 1./freq;
figure, plot(period,power), axis([0 40 0 2e7]), grid on
ylabel('Power')
xlabel('Period(Years/Cycle)')

[mp index] = max(power);
period(index)
```
References

1. Berger, S., W. Goldsmith, and E. Lewis (eds.) *Introduction to Bioengineering*, Oxford, New York, 1996. This reference is complete, but it is written at a beginning graduate level. The section on medical imaging is brief, but useful.

2. Brown, B., R. Smallwood, D. barber, P. Lawford, and D. Hose. *Medical Physics and Biomedical Engineering*, Institute of Physics Publishing, Bristol, 1999. This was to be our textbook for the semester. There are about 30 copies in the IMU bookstore, but we could not get more than 30 from the publisher. Too bad, so sad.

3. Macovski, A. *Medical Imaging Systems*, Prentice-Hall, Englewood Cliffs, 1883. This small text contains the mathematical treatment of imaging that we all need to know. Unfortunately, the book is out of print.

4. Shung, K. M. Smith, B. Tsui. *Principles of Medical Imaging*, Academic Press, San Diego, 1992. This is a wonderful brief text from which many of the images are drawn. Unfortunately, this text is out of print.