A Clinical System for the measurement of Regional Metabolic Rates in the Brain

Claude Nahmias

A thesis submitted to the University of Surrey in fulfilment of the degree of Doctor of Philosophy

June 1985
Abstract

The study of the chemical events that regulate the function of the human brain is particularly difficult. The introduction by Hounsfield, in 1973, of a tomographic technique based on the attenuation of X-rays by tissues has proved invaluable in the study of the morphology of the brain. An extension of this technique, employing the concepts of computerised tomography in combination with the use of specific molecules labelled with positron emitters, is now making the direct regional measurement of metabolic rates during life possible. Although some positron tomography systems are available commercially, they do not necessarily fulfil the specific needs of all researchers.

Faced with the problem of quantitating the regional distribution of the essential neurotransmitter, dopamine, in the human brain a positron tomography system, which forms the basis of this work, was designed and built based on a series of experiments aimed at optimizing spatial resolution and detection efficiency. The performance of the tomograph has been evaluated through a series of phantom studies; and the system has been used to measure the local cerebral metabolic rate of glucose and the local distribution of dopamine in the healthy and diseased brain. It is felt that the ability of this tomograph to resolve metabolic structures in the brain as small as $10^3 \, \text{mm}^3$ will only be surpassed at the cost of unduly increasing the radiation dose to the subject.

The results of positron tomographic studies performed using different positron labelled molecules and those obtained using X-ray computerized tomographic techniques and magnetic resonance techniques in the same subject have been compared. The results have been found to be complementary, each technique providing a clue to the proper understanding of the functioning of the brain.
Acknowledgements

Without the help of Nicholas M. Spyrou and E. Stephen Garnett this thesis could not have been started, and without their guidance and support it would not have been completed.

Special thanks are due to E.S. Garnett, department of Nuclear Medicine, McMaster University Medical Centre, and to the department of Physics at the University of Surrey for permission to undertake this collaborative venture.

Finally, I would like to thank E. Stephen Garnett, D. Bev Kenyon and Gunter Firnau for the many discussions, advice and assistance that I have received from them throughout the last fifteen years.
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>iii</td>
</tr>
<tr>
<td>Table of contents</td>
<td>iv</td>
</tr>
<tr>
<td>List of figures</td>
<td>viii</td>
</tr>
<tr>
<td>List of tables</td>
<td>xiii</td>
</tr>
<tr>
<td>Chapter 1</td>
<td>1</td>
</tr>
<tr>
<td>Chapter 2</td>
<td>29</td>
</tr>
<tr>
<td>Chapter 3</td>
<td>60</td>
</tr>
<tr>
<td>Chapter 4</td>
<td>94</td>
</tr>
<tr>
<td>Chapter 5</td>
<td>141</td>
</tr>
<tr>
<td>Chapter 6</td>
<td>186</td>
</tr>
<tr>
<td>Bibliography</td>
<td>245</td>
</tr>
</tbody>
</table>
Table of Contents

Chapter 1: Organization of the brain.

1.1. Introduction 1

1.2. Anatomical organization of the brain 2
   1.2.1. Subdivisions of the brain 3
      1.2.1.1. The cerebral hemispheres 3
      1.2.1.2. The brain stem 3
      1.2.1.3. The cerebellum 4
   1.2.2. The cerebral hemispheres 6
      1.2.2.1. The frontal lobe 9
      1.2.2.2. The temporal lobe 9
      1.2.2.3. The parietal lobe 11
      1.2.2.4. The occipital lobe 11
      1.2.2.5. The insula 13
      1.2.2.6. The limbic lobe 14
      1.2.2.7. The thalamus 15
      1.2.2.8. The basal ganglia 15

1.3. Functional organization of the brain 17
   1.3.1. The cerebral hemispheres 17
      1.3.1.1. The frontal lobes 20
      1.3.1.2. The temporal lobes 21
      1.3.1.3. The parietal lobes 22
      1.3.1.4. The occipital lobes 23
      1.3.1.5. The insula 23
      1.3.1.6. The limbic system 24
      1.3.1.7. The thalami 25
      1.3.1.8. The basal ganglia 25

1.3.2. Asymmetry of the cerebral hemispheres 25

1.4. Discussion 27

Chapter 2: Positron emission and energy losses.

2.1. The positron 29

2.2. Positron emission 30
   2.2.1. Positron decay 30
   2.2.2. Electron capture 32
   2.2.3. Competition between electron capture and positron decay 34
2.3. Positron energy loss

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.3.1. Elastic scattering with nuclei</td>
<td>36</td>
</tr>
<tr>
<td>2.3.2. Elastic scattering with atomic electrons</td>
<td>44</td>
</tr>
<tr>
<td>2.3.3. Inelastic collisions with atomic electrons</td>
<td>44</td>
</tr>
<tr>
<td>2.3.4. Radiative collisions with atomic nuclei</td>
<td>47</td>
</tr>
</tbody>
</table>

2.4. Positron range

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.4.1. Path length and range of positrons</td>
<td>49</td>
</tr>
<tr>
<td>2.4.2. Range-energy relation</td>
<td>50</td>
</tr>
</tbody>
</table>

2.5. The annihilation process

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5.1. One photon annihilation</td>
<td>52</td>
</tr>
<tr>
<td>2.5.2. Two photons annihilation</td>
<td>52</td>
</tr>
<tr>
<td>2.5.3. Formation of positronium atoms</td>
<td>54</td>
</tr>
<tr>
<td>2.5.4. Decay of positronium atoms</td>
<td>56</td>
</tr>
</tbody>
</table>

2.6. Discussion

58

Chapter 3: Detectors for positron tomography

3.1. Introduction

60

3.2. General properties of radiation detectors

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2.1. Energy resolution</td>
<td>61</td>
</tr>
<tr>
<td>3.2.2. Detection efficiency</td>
<td>64</td>
</tr>
<tr>
<td>3.2.3. Photopeak fraction</td>
<td>65</td>
</tr>
<tr>
<td>3.2.4. Time resolution</td>
<td>65</td>
</tr>
<tr>
<td>3.2.5. Spatial resolution</td>
<td>68</td>
</tr>
</tbody>
</table>

3.3. Proportional counters

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.3.1. General principles</td>
<td>70</td>
</tr>
<tr>
<td>3.3.2. Multiwire proportional counters</td>
<td>71</td>
</tr>
<tr>
<td>3.3.3. Discussion</td>
<td>74</td>
</tr>
</tbody>
</table>

3.4. Solid state detectors

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.4.1. General principles</td>
<td>76</td>
</tr>
<tr>
<td>3.4.2. Germanium</td>
<td>78</td>
</tr>
<tr>
<td>3.4.3. Cadmium telluride</td>
<td>79</td>
</tr>
<tr>
<td>3.4.4. Mercuric iodide</td>
<td>79</td>
</tr>
<tr>
<td>3.4.5. Discussion</td>
<td>79</td>
</tr>
</tbody>
</table>

3.5. Scintillation detectors

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.5.1. General principles</td>
<td>80</td>
</tr>
<tr>
<td>3.5.2. Sodium iodide</td>
<td>88</td>
</tr>
<tr>
<td>3.5.3. Caesium fluoride</td>
<td>88</td>
</tr>
<tr>
<td>3.5.4. Barium fluoride</td>
<td>89</td>
</tr>
<tr>
<td>3.5.5. Bismuth germanate</td>
<td>90</td>
</tr>
</tbody>
</table>

3.6. Conclusions

90
### 5.4. The electronics
- 5.4.1. Pulse height discrimination
- 5.4.2. Timing
- 5.4.3. Coincidence logic

### 5.5. The data handling
- 5.5.1. Reconstruction
- 5.5.2. Display and manipulation

### 5.6. Performance of the tomograph
- 5.6.1. Detection efficiency
- 5.6.2. Linearity
- 5.6.3. Scattered fraction
- 5.6.4. Spatial resolution
- 5.6.5. Axial resolution

### 5.7. Summary

<table>
<thead>
<tr>
<th>Chapter 6: The study of regional cerebral metabolism during life.</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1. Introduction</td>
</tr>
<tr>
<td>6.2. The tracer technique</td>
</tr>
<tr>
<td>- 6.2.1. Assumptions in tracer techniques</td>
</tr>
<tr>
<td>- 6.2.2. Advantages of substrate analogues in tracer techniques</td>
</tr>
<tr>
<td>6.3. The measurement of energy requirements in the brain</td>
</tr>
<tr>
<td>- 6.3.1. Introduction</td>
</tr>
<tr>
<td>- 6.3.2. The deoxyglucose method</td>
</tr>
<tr>
<td>- 6.3.2.1. Rationale of the method</td>
</tr>
<tr>
<td>- 6.3.2.2. Compartmental analysis</td>
</tr>
<tr>
<td>- 6.3.2.3. Discussion</td>
</tr>
<tr>
<td>6.4. The measurement of intracerebral neurotransmitter</td>
</tr>
<tr>
<td>metabolism</td>
</tr>
<tr>
<td>- 6.4.1. Introduction</td>
</tr>
<tr>
<td>- 6.4.2. The fluorodopa method</td>
</tr>
<tr>
<td>- 6.4.2.1. Rationale of the method</td>
</tr>
<tr>
<td>- 6.4.2.2. Compartmental analysis</td>
</tr>
<tr>
<td>- 6.4.2.3. Discussion</td>
</tr>
<tr>
<td>6.5. Correlation between CT, NMR and PT findings in the</td>
</tr>
<tr>
<td>brain</td>
</tr>
<tr>
<td>- 6.5.1. Introduction</td>
</tr>
<tr>
<td>- 6.5.2. Examination of nigro-striatal pathway</td>
</tr>
<tr>
<td>- 6.5.3. Examination of the hypoxic brain</td>
</tr>
<tr>
<td>- 6.5.4. Discussion</td>
</tr>
<tr>
<td>6.6. Conclusions</td>
</tr>
</tbody>
</table>
List of Figures

Chapter 1: Organization of the brain.

1.1 Photograph and drawing of midsagittal section of the brain identifying the main anatomical structures.  
1.2 Horizontal and coronal sections illustrating the location of the longitudinal fissure and the corpus callosum.  
1.3 Lateral view of the skull and the brain showing the relation between the bones of the skull and the lobes of the brain.  
1.4 The frontal lobe.  
1.5 The parietal lobe.  
1.6 The temporal lobe.  
1.7 The occipital lobe.  
1.8 The insula.  
1.9 The limbic lobe.  
1.10 The thalamus and the basal ganglia.  
1.11 Main functional organization of the brain.  

Chapter 2: Positron emission and energy losses.

2.1 Energetics of positronium formation  

Chapter 3: Detectors for positron tomography.

3.1 Energy resolution.  
3.2 Comparison of energy spectra obtained using a Ge (Li) detector and a Na I (Tl) detector.  
3.3 Photopeak fraction.
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.4</td>
<td>Time resolution.</td>
<td>66</td>
</tr>
<tr>
<td>3.5</td>
<td>Spatial resolution.</td>
<td>69</td>
</tr>
<tr>
<td>3.6</td>
<td>Multiwire proportional counter.</td>
<td>72</td>
</tr>
<tr>
<td>3.7</td>
<td>Multiwire proportional counter with lead converters.</td>
<td>72</td>
</tr>
<tr>
<td>3.8</td>
<td>Energy band representation of conductors, insulators and semi-conductors.</td>
<td>77</td>
</tr>
<tr>
<td></td>
<td>Chapter 4: Principles of image reconstruction from projections.</td>
<td></td>
</tr>
<tr>
<td>4.1</td>
<td>Transverse section through an object.</td>
<td>96</td>
</tr>
<tr>
<td>4.2</td>
<td>System of coordinates used to describe the density function ( f(x,y) ).</td>
<td>96</td>
</tr>
<tr>
<td>4.3</td>
<td>Relation between the fixed and rotated coordinate systems.</td>
<td>99</td>
</tr>
<tr>
<td>4.4</td>
<td>Polar representation of the projection data.</td>
<td>99</td>
</tr>
<tr>
<td>4.5</td>
<td>Simple and filtered backprojection of a point source.</td>
<td>101</td>
</tr>
<tr>
<td>4.6</td>
<td>Iterative least squares technique.</td>
<td>105</td>
</tr>
<tr>
<td>4.7</td>
<td>Algebraic reconstruction technique.</td>
<td>106</td>
</tr>
<tr>
<td>4.8</td>
<td>Relation between the orthogonal and polar coordinate systems.</td>
<td>109</td>
</tr>
<tr>
<td>4.9</td>
<td>Illustration of the projection-slice theorem.</td>
<td>109</td>
</tr>
<tr>
<td>4.10</td>
<td>Illustration of the frequency windows used in filtered backprojection.</td>
<td>113</td>
</tr>
<tr>
<td>4.11</td>
<td>Parallel ray geometry.</td>
<td>120</td>
</tr>
<tr>
<td>4.12</td>
<td>Divergent ray geometry.</td>
<td>122</td>
</tr>
<tr>
<td>4.13</td>
<td>Scattered true coincidences.</td>
<td>126</td>
</tr>
<tr>
<td>4.14</td>
<td>Wobble motion.</td>
<td>137</td>
</tr>
<tr>
<td>4.15</td>
<td>Dichotomic motion.</td>
<td>139</td>
</tr>
</tbody>
</table>
# Chapter 5: Design and construction of the tomograph

<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>Energy spectrum obtained using a bismuth germanate crystal.</td>
<td>146</td>
</tr>
<tr>
<td>5.2</td>
<td>Variation in photopeak efficiency as the crystal face is reduced while maintaining the length constant.</td>
<td>148</td>
</tr>
<tr>
<td>5.3</td>
<td>Variation in the photopeak efficiency as the crystal length is shortened.</td>
<td>150</td>
</tr>
<tr>
<td>5.4</td>
<td>Geometry used to measure the line response of two banks of detectors operated in coincidence.</td>
<td>154</td>
</tr>
<tr>
<td>5.5</td>
<td>Line response of the two banks of detectors operated in coincidence.</td>
<td>154</td>
</tr>
<tr>
<td>5.6</td>
<td>Isoresponse curves illustrating the effects of reducing the depth of shielding.</td>
<td>156</td>
</tr>
<tr>
<td>5.7</td>
<td>Photograph of the gantry.</td>
<td>158</td>
</tr>
<tr>
<td>5.8</td>
<td>Photograph of the bismuth germanate crystals fitted with their photomultiplier tubes.</td>
<td>158</td>
</tr>
<tr>
<td>5.9</td>
<td>Photograph of a portion of the detector assembly.</td>
<td>159</td>
</tr>
<tr>
<td>5.10</td>
<td>Photograph of the tomograph and the patient's couch.</td>
<td>159</td>
</tr>
<tr>
<td>5.11</td>
<td>Time spectrum obtained using two channels of the tomograph.</td>
<td>163</td>
</tr>
<tr>
<td>5.12</td>
<td>Block diagram of the coincidence logic.</td>
<td>164</td>
</tr>
<tr>
<td>5.13</td>
<td>The electronics: Detecting a 511 keV photon.</td>
<td>166</td>
</tr>
<tr>
<td>5.14</td>
<td>The electronics: Closing the input gates.</td>
<td>167</td>
</tr>
<tr>
<td>5.15</td>
<td>The electronics: Rejecting a 'singles' event.</td>
<td>168</td>
</tr>
<tr>
<td>5.16</td>
<td>The electronics: Rejecting a coincidence event involving more than two sectors.</td>
<td>169</td>
</tr>
<tr>
<td>5.17</td>
<td>The electronics: Processing a valid coincidence.</td>
<td>169'</td>
</tr>
<tr>
<td>5.18</td>
<td>Illustration of a black and white and two different colour displays.</td>
<td>173</td>
</tr>
</tbody>
</table>
5.19 Response of the tomograph to decreasing amounts of radioactivity in the field of view. 179
5.20 True and scattered coincidences count rates. 181
5.21 Transaxial spatial resolution of the tomograph. 182
5.22 Axial resolution of the tomograph. 184

Chapter 6: The study of regional cerebral metabolism during life.
6.1 The glycolytic pathway. 192
6.2 Comparison of the pathways for glucose and 2-deoxyglucose. 194
6.3 Compartmental model for the 2-deoxyglucose method. 196
6.4 Mathematical analysis of the 2-deoxyglucose model. 197
6.5 The operational equation for the calculation of regional cerebral metabolic rates of glucose. 202
6.6 Tomographic slices of normal brain from a single individual obtained using $[^{18}\text{F}]$ 2-fluoro 2-deoxyglucose. 206
6.7 Diagram of some of the main connections of the basal ganglia. 215
6.8 Pathways of dopamine biosynthesis and metabolism. 218
6.9 Compartmental model for the fluorodopa method. 220
6.10 Mathematical analysis of the fluorodopa model. 223
6.11 Rate of change of $^{18}\text{F}$ activity in striatal and occipital regions. 227
6.12 Tomographic slices of normal brain from a single individual obtained using $[^{18}\text{F}]$ fluorodopa. 230
6.13 CT, NMT and PT scans of a normal subject. 233
6.14 CT and PT scans of a patient suffering from Huntington's disease. 236
6.15 CT and PT scans of a patient suffering from Parkinson's disease. 236
6.16 CT and PT scans of a patient suffering from secondary dystonia. 236
6.17 CT and PT scans of a six day old infant. 238
6.18 NMR and PT scans of a patient having suffered and infarct. 238
**List of Tables**

5.1 Comparison of some of the physical properties of sodium iodide, caesium fluoride and bismuth germanate. 144

5.2 Effects of reducing the area of the exposed face of a bismuth germanate crystal while maintaining its length constant. 148

5.3 Effects of reducing the length of a bismuth germanate crystal while maintaining the area of the exposed face constant. 150

5.4 Effects of varying the crystal height while maintaining the opening at 0.8 cm X 2 cm. 152.
Chapter 1:
Organization of the brain

Chapter 2:
Positron emission and energy losses

Chapter 3:
Detectors for positron tomography

Chapter 4:
Principles of image reconstruction from projections

Chapter 5:
Design and construction of the tomograph

Chapter 6:
The study of regional metabolism during life
CHAPTER 1

ORGANIZATION OF THE BRAIN.

1.1. Introduction

1.2. Anatomical organization of the brain
   1.2.1. Subdivisions of the brain
   1.2.1.1. The cerebral hemispheres
   1.2.1.2. The brain stem
   1.2.1.3. The cerebellum
   1.2.2. The cerebral hemispheres
   1.2.2.1. The frontal lobe
   1.2.2.2. The temporal lobe
   1.2.2.3. The parietal lobe
   1.2.2.4. The occipital lobe
   1.2.2.5. The insula
   1.2.2.6. The limbic lobe
   1.2.2.7. The thalamus
   1.2.2.8. The basal ganglia

1.3. Functional organization of the brain
   1.3.1. The cerebral hemispheres
   1.3.1.1. The frontal lobes
   1.3.1.2. The temporal lobes
   1.3.1.3. The parietal lobes
   1.3.1.4. The occipital lobes
   1.3.1.5. The insula
   1.3.1.6. The limbic system
   1.3.1.7. The thalamus
   1.3.1.8. The basal ganglia

1.3.2. Asymmetry of the cerebral hemispheres

1.4. Discussion
1.1. Introduction

The brain is an assembly of cells that continually receives information, elaborates and perceives it, and makes decisions about it. All information we have concerning the world around us is conveyed centrally to the brain by an elaborate sensory system. Receptors of many kinds act as transducers which translate physical and chemical stimuli in our environment into nerve impulses which travel to the brain where an appropriate response is elicited. The brain is also concerned with all kinds of motor activity and with the regulation of visceral, endocrine and somatic functions. Attention, consciousness and emotional experiences are all central neuronal functions. The ability of the brain to discriminate between stimuli of the same and different types forms one of the bases for learning. The receptive and expressive use of symbols and signs that underlie communication is also one of the main concerns of the human brain. Higher functions such as memory, imagination, thought and creative ability, though poorly understood, must also be related to complex neuronal activity.

Although biochemists have developed a very sophisticated understanding of the chemical events occurring in brain tissue, their work has been hampered by their inability to make direct measurements in vivo. Simplified test tube experiments may not always accurately represent the situation.
in an intact organism and more complex experiments in animals cannot necessarily be extrapolated to man. Furthermore, a class of studies such as studies of mood can only be performed in an intact human being. Some progress has been made towards the measurement of global cerebral blood flow and metabolism, but regional differences are obscured, if not completely lost, by such an approach. Direct measurements of metabolic rates are being made possible by the combined use of specific molecules labelled with positron emitters and computerized tomographic techniques. This method of investigation, positron tomography, has the potential to provide unique information on the biochemistry and physiology of the human central nervous system in health and disease.

The purpose of this work was to design, to build and to characterize a clinical instrument with which to study regional metabolic rates in the brain. Some understanding of the anatomical and functional organization of the human brain is essential to the design of such an instrument.

1.2. **Anatomical organization of the brain**

The human brain is a relatively small structure weighing about 1400 g and constituting about 2 % of the total body weight. It consists of three basic subdivisions: the cerebral hemispheres, the brain stem and the cerebellum.
(fig. 1.1). Three-quarters of the brain cells are found in its outer layer, the cortex. The cerebral cortex does not have a uniform histological structure throughout all parts of the hemispheres. Brodmann and others have described many different cortical areas having individual architectonic characteristics, and the existence and general boundaries of some of these regions are now well established. Experimental and pathologic research indicates that there are certain specific differences in function associated with certain brain parts.

1.2.1. Subdivisions of the brain

1.2.1.1. The cerebral hemispheres

The paired cerebral hemispheres consist of a highly convoluted grey cortex, an underlying white matter of considerable magnitude and a collection of deeply located neuronal masses, known as the basal ganglia. The grey matter represents nerve cell collections while the white matter represents axons which connect the various regions of the brain with each other and with the spinal cord.

The cerebral hemispheres are concerned with sensory and motor functions as well as complex behavioural processes.

1.2.1.2. The brain stem

The brain stem is an extension upwards of the spinal cord. Below, the medulla becomes continuous with the
spinal cord. Above, the pons is continuous with the mid-brain which connects it to the cerebrum. The mid-brain is the smallest and least differentiated part of the brain stem. The substantia nigra, the largest single nuclear mass in the mid-brain, has connections with parts of the basal ganglia and the thalamus.

The brain stem is concerned primarily with the control of important bodily functions such as respiration, cardiovascular function and gastrointestinal function. It contains all the nuclei of the cranial nerves except those of smell and sight. It also carries up the sensory impulses from the spinal cord to the higher centres, and carries down corticospinal tracts from higher levels to the motor cells of the cord.

1.2.1.3. The cerebellum

The cerebellum lies behind the pons and the medulla and extends laterally under the tentorium to fill the greater part of the posterior fossa.

The cerebellum is concerned primarily with motor coordination and the control of muscle tone and equilibrium. It is not concerned with proprioceptive perception and no information transmitted to it enters the conscious sphere, that is, the cerebrum.
Figure 1.1: Photograph and drawing of midsagittal section of the brain identifying the main anatomical structures.
1.2.2. The cerebral hemispheres

The cerebral hemispheres are partially separated from each other by the longitudinal fissure (fig. 1.2). In frontal and occipital regions, the separation of the hemispheres is complete, but in the central region the two hemispheres are united by a thick band of white matter, the corpus callosum which is the chief functional link between them (fig. 1.2). The massive white matter of the cerebral hemispheres forms the medullary core of the cortical convolutions and contains basically three types of fibres: projection fibres that convey impulses either from the cortex to distant loci, or from distant loci to the cortex, association fibres that interconnect various cortical regions of the same hemisphere and commissural fibres that interconnect corresponding cortical regions of the two hemispheres.

Each cerebral hemisphere is subdivided into lobes by various sulci. The major lobes of the brain are named for the bones of the skull which overlie them (fig.1.3). Although the boundaries of the various lobes as seen in the gross specimen are somewhat arbitrary, multiple cortical areas in each lobe are histologically distinctive. The grey cellular mantle of the cerebral cortex in man is highly convoluted. The crest of a single convolution is referred to as a gyrus; sulci separate the various gyri,
Figure 1.2: Horizontal and coronal sections illustrating the location of the longitudinal fissure and the corpus callosum.
Figure 1.3: Lateral view of the skull and the brain showing the relation between the bones of the skull and the lobes of the brain.
producing patterns with more or less constant features. On the basis of the more constant sulci and gyri, the cerebrum is divided into six lobes: frontal, temporal, parietal, occipital, insula and limbic (fig. 1.3).  
1.2.2.1. The frontal lobe

The frontal lobes are the most recently philogenetically developed parts of the brain. In man, the frontal lobe is the largest of all the lobes of the brain and comprises about one third of the hemispheric surface. The frontal lobe extends from the central sulcus to the frontal pole; its inferior lateral boundary is the lateral sulcus (fig. 1.4). The convexity of the frontal lobe has four principal convolutions: a precentral gyrus that parallels the central sulcus and three horizontally oriented convolutions, the superior, middle and inferior frontal gyri. The precentral gyrus and the anterior bank of the central sulcus comprise the primary motor area, where the parts of the body are represented in a distorted, but topographical, manner. Regions of the frontal lobe anterior to the primary motor area are referred to as premotor and prefrontal areas.

1.2.2.2. The parietal lobe

The boundaries of the parietal lobe are not precise, except for its anterior border on the lateral convexity formed by the central sulcus, and its posterior border on the medial aspect of the hemisphere formed by the
Figure 1.4: The frontal lobe

Figure 1.5: The parietal lobe
parieto-occipital sulcus (fig. 1.5). The posterior bank of
the central sulcus and the postcentral gyrus constitute
the primary somesthetic area, the cortical region where
impulses concerned with tactile and kinesthetic senses from
superficial and deep receptors converge and are somatotopically
represented. The majority of cortical neurons in the
postcentral gyrus are concerned with fixed receptive fields
on the contralateral side of the body that are place-specific,
modality-specific and related to discriminative aspects
of sensation.

1.2.2.3. The temporal lobe

The temporal lobe lies below the lateral sulcus.
It displays on its lateral surface three obliquely oriented
convolutions, the superior, middle and inferior temporal
gyri (fig. 1.6). It is concerned with the auditory perception
as well as memory and integration.

1.2.2.4. The occipital lobe

The small occipital lobe rests on the tentorium.
Its anterior boundary is the parieto-occipital sulcus on
the medial aspect of the hemisphere (fig. 1.7). Its lateral
surface is poorly delimited from the parietal lobe. On
the medial aspect of the hemisphere the occipital lobe is
divided by the calcarine sulcus into the cuneus and the
lingual gyrus. The calcarine sulcus joins the parieto-
occipital sulcus posteriorly into a Y-shaped formation.
Figure 1.6: The temporal lobe

Figure 1.7: The occipital lobe
The cortex on both banks of the calcarine sulcus represents the primary visual cortex.

1.2.2.5. The insula

The insula is a cortical area that lies in the depth of the lateral sulcus, buried by overgrowth of the adjoining cortical areas, the opercula (fig. 1.8).

Figure 1.8: The insula (coronal and horizontal sections).
1.2.2.6. The limbic lobe

The limbic lobe consists of large cortical convolutions on the medial aspect of the hemispheres which surround the anterior part of the brain stem and the interhemispheric commissure. It includes the cingulate and parahippocampal gyri. It also includes associated subcortical nuclei: the hypothalamus, the amygdala and the hippocampus (fig. 1.9).

The limbic lobe plays an important role in the regulation of internal organs and the expression of emotional behaviour.

Figure 1.9: The limbic lobe
1.2.2.7. The thalamus

The thalamus is an oblique mass of grey matter on either side of the midline and forms the walls of the third ventricle. It lies between the fibres of the internal capsule, a broad band of projection fibres, and is flanked by the body and tail of the caudate nucleus. It is composed of numerous separate nuclear units of differing and largely unknown functions.

The thalamus is an important relay centre on the sensory pathway to the cortex.

1.2.2.8. The basal ganglia

The basal ganglia are large subcortical masses which are generally divided into the corpus striatum, and the amygdaloid nuclear complex. The amygdala is functionally related to the hypothalamus and is regarded as an integral part of the limbic system. The striatum is composed of the caudate nucleus and the putamen. The caudate nucleus is a long, arched mass of grey matter which is closely related throughout its length to the lateral ventricle where it forms the inferior boundary. The tail of the caudate nucleus turns downward behind the thalamus and ends in the amygdala. The putamen is the largest and most lateral part of the basal ganglia and is separated from the head of the caudate nucleus by the internal capsule. The globus pallidus lies medial to the putamen and is divided
into an internal and external section (fig. 1.10).

The basal ganglia are mostly concerned with voluntary motor activities.

Figure 1.10: The thalamus and the basal ganglia (lateral representation and horizontal section).
1.3. Functional organization of the brain

1.3.1. The cerebral hemispheres

In a general way, the central sulcus divides the brain into a posterior, receptive, portion and an anterior portion closely related to efferent and motor function (fig. 1.11).

The posterior part contains all the primary receptive areas which receive specific sensory impulses from the lower centres of the brain and hence, indirectly, from the peripheral sensory receptors. Impulses entering these primary areas produce sensations of a sharply defined character. However, these sensations have not yet attained the perceptual level necessary for the recognition of an object. This requires the integration of primary stimuli into more complicated sensory patterns. Each primary sensory cortex is adjoined by a parasensory association area. All of the parasensory areas receive cortical inputs, either directly or indirectly, from only one primary sensory cortex. These areas serve for the combination and elaboration of the primary impulses into more complex unisensory perceptions. The parasensory areas project, in turn, to polymodal zones which, by receiving two or more different types of sensory information, have functions which transcend the specific modalities of the primary sensory and parasensory areas. In addition to projecting to the polymodal
Figure 1.11: Main functional organization of the brain.
areas, each second order parasensory area also projects to one or more paralimbic association area, therefore establishing a link with the limbic system. This is one way whereby highly processed sensory information is invested with emotional and motivational significance.

Similarly, in the frontal lobe and anterior to the primary motor cortex is another large region of association cortex, the frontal association cortex. It is subdivided into premotor and prefrontal areas. The premotor cortex has reciprocal connections with the motor cortex. It also receives inputs from first order parasensory areas and, in this way, uses the sensory information it receives to influence the activity of the motor cortex. The prefrontal association cortex receives cortical inputs from premotor, rather than primary motor cortex, and projects to paralimbic areas, specifically the cingulate gyrus and the anterior temporal lobe. There is also a network of reciprocal connections between frontal and parasensory association areas and between frontal and paralimbic areas. This pattern of connectivity suggest that this region allows the organism to take into consideration both motivational as well as complex sensory data in preparation for eventual motor activity.

Thus as a sensory input reaches the cerebral cortex it first passes through successive stages of intramodality
elaboration allowing progressively more complex discrimination of the features of a particular stimulus. Then, by a series of further connections, this sensory information, now in a highly complex form, is conveyed to polymodal zones for cross-modal interchange of information, to paralimbic and limbic areas for investment with emotional tone and storage in memory, and to the frontal association areas where both sensory and limbic data are integrated in preparation for the organism to respond to sensory stimuli by an appropriate action.

1.3.1.1. The frontal lobes

The frontal lobes can be subdivided into motor cortex, premotor cortex (motor organization) and prefrontal cortex (higher integration). The precentral gyrus constitutes the motor area of the cortex and contains the higher centres controlling the individual movements, primarily the terminal muscles of the extremities, on the opposite side of the body. The areas of the left inferior frontal gyrus close to the lateral sulcus control language output. This area, Broca'a area, lies close to the portion of the motor cortex that controls the muscles involved in the execution of speech. Injury to this area interferes with normal speech production, even though comprehension of language is normal. The premotor area organizes, from ideas, specific movements which are purposive and transmits
them to the motor area to be analysed and routed to the specific muscles. The prefrontal regions have rich connections with the reticular activating system of the brain stem and the thalamus, and with all other cortical zones. Through the first set of connections, the prefrontal areas are intimately concerned with the state of alertness of the organism; while the rich connections with the posterior receptor areas and motor cortex allow the prefrontal areas to organize and execute the most complex of man's goal directed or purposive activity.

1.3.1.2. The temporal lobes

The organization of the temporal lobes is very complex. It is related to the sense systems of olfaction and audition whose primary projection areas and perceptual elaboration lie within its boundaries. Although the temporal lobes are neither concerned with the primary reception of visual information nor with its elaboration into meaningful wholes, they are concerned however with the integration of visual experience with all forms of sensory information coming from the receptors of the other special senses and from the receptors of the bodily senses. An area close to the primary auditory cortex in the dominant hemisphere, Wernicke's area, controls language input. Injury to this area causes loss of comprehension of language even though speech production is retained.
The temporal lobes play an important role in memory in both its specific and general aspects. They contain systems which help to preserve the record of conscious experience. Finally, they have such an intimate connection with the structures of the limbic system, which itself has far reaching connections, that their functional boundaries as well as morphological boundaries are ill-defined. It is through this system that the temporal lobes help to provide part of the anatomical substrate for the integration of the emotional and motivational aspect of the organism with information coming from all those sensory systems situated behind the central sulcus and through its connections with the frontal lobes, with those systems for plans of action which are formulated in these regions.

1.3.1.3. The parietal lobes

The parietal lobes are strategically situated between the frontal, occipital and temporal lobes and are closely related in function to each of these regions of the brain. The post central sulcus delimits the postcentral gyrus which is concerned with somatic sensations. The primary reception area for the numerous forms of somatic sensations has its principal locus in the post central gyrus. The secondary or association cortex posterior to this is thought to deal with the elaboration of the discrete elements into meaningful wholes.
1.4.1.4. The occipital lobes

The occipital lobes contain the primary visual cortex and a secondary sensory area believed to be concerned with the elaboration and synthesis of visual information. A third zone adjacent to the secondary sensory area possesses abundant connections with other regions of the hemispheres so that it appears to be chiefly involved in the integration of visual information with information gathered by the auditory and other sensory systems and it unites visual information with the brain system subserving speech and other executive functions. It is also concerned, along with areas in the temporal lobes, with visual memory.

1.3.1.5. The insula

The exact connections and functions of the insula are yet unknown.

1.3.1.6. The limbic system

The limbic system occupies a central position in the neuronal mechanism that governs behaviour and emotion, and visceral functions. The hypothalamus is primarily concerned with the regulation of internal organs and is related to endocrine and somatic activities. The amygdala and hippocampus are associated with the expression of emotional behaviour; the first has an excitatory function whereas the latter has an inhibitory function.
1.3.1.7. The thalami

The thalamus is an important relay centre on the sensory pathway to the cortex. It also acts as a centre for integrating impulses from many sources (somatic, sensory, visual, visceral, etc.) before passing them on to the cerebral cortex. The thalamic nuclei can be classified either as specific relay nuclei or as association nuclei. The specific relay nuclei project to and receive fibres from well defined cortical areas considered to be related to specific functions. The association nuclei do not receive direct fibres from ascending systems, but project to association areas of the cortex. With the exception of olfaction, impulses involved in all forms of sensation reach localized areas on the cerebral cortex via thalamocortical projection systems. The localized regions that receive the principal projections of the specific sensory relay nuclei of the thalamus are the focal regions in the cerebral cortex where specific sensory modalities are most extensively and critically represented. They are the primary sensory areas and are concerned specially with integration of sensory experience. Evidence suggest that near each primary receptive area there are cortical zones which may receive sensory inputs directly, or indirectly, from the thalamus. These cortical zones, adjacent to primary sensory areas, but outside of the principal projection
area of the specific sensory relay nuclei of the thalamus are referred to as the parasensory association areas.

1.3.1.8. **The basal ganglia**

The connections of the striatum are intricate and are only partially known. Fibres from the frontal, temporal and parietal cortical areas and from the thalami end in the putamen and caudate nucleus. Connections between the striatum and the substantia nigra appear truly reciprocal, and topographically interrelate specific regions of both nuclei. Unlike the striatum, the globus pallidus does not appear to receive direct afferents from the cerebral cortex or from the thalamic nuclei. This complex arrangement of connections linking the cortex, the striatum and the brain stem influences various voluntary motor activities.

1.3.2. **Asymmetry of the cerebral hemispheres**

Although the two cerebral hemispheres appear as duplicates of each other, there are many functions which are not represented equally at a cortical level. This appears true even though impulses from receptors on each side of the body seem to project nearly equally, although largely contralaterally, to symmetrical areas, and certain information received in the cortex of one hemisphere can be transferred to the other via interhemispheric commissures. In certain higher functions, believed to be cortical in nature, one hemisphere appears to be the leading one, and,
in this sense is referred to as the dominant hemisphere. The most remarkable feature of cerebral dominance in man is the fact that in the adult the capacity for speech is overwhelmingly controlled by the left hemisphere even in left-handers. Cerebral dominance is most complete, therefore, in relation to the complex and highly evolved aspects of language. With respect to most of the other higher functions, cerebral dominance appears to be one of degree. The dominant hemisphere deals with language, computation and logical thinking whereas the nondominant hemisphere deals with music, spatial perception and non-verbal tasks.

At a biochemical level, asymmetries between the cerebral hemispheres have also been found. A moderate degree of asymmetry in the distribution of the neurotransmitter dopamine has been demonstrated in the striatum, at least in the rat, and it has been speculated that this asymmetry may be optimal for maximal overall learning ability. Unilateral sensory and visual stimuli have been shown to evoke reciprocal changes in the rate of dopamine formation not only between the caudate nucleus and the substantia nigra on the same side, but also between the right and left sides.
1.4. Discussion

The study of the human brain, and therefore the understanding of its function in health and disease, has been hampered by an inherent inability to make regional measurements of blood flow and energy requirements and to map the distribution of essential biochemical molecules such as neurotransmitters and drugs that affect the central nervous system. The last ten years has seen the advent of three new technologies, each of which can probe a different aspect of the intact brain. Computerized axial tomography defines the intracerebral anatomy on the basis of differences in mass attenuation coefficients. It differentiates some tumours and the consequences of cerebrovascular accidents from normal brain tissue. However, it has a limited ability to distinguish grey matter from white and it gives no information at all about cerebral function. Nuclear magnetic resonance also defines intracerebral anatomy, but in greater detail than CT. It has the ability to discriminate well between grey and white matter and to distinguish very small intracerebral lesions such as plaques of multiple sclerosis. Unlike CT, NMR devices have the potential to measure some parameters of brain metabolism. These relate, so far, to measurements of the phosphorous containing molecules associated with energy metabolism and may be useful in
stroke research. The third modality, positron tomography, presents, at least theoretically, no limit as to the molecules of biological interest that can be studied. It has made it possible to monitor, for the first time, biochemical reactions as they occur in small regions of the living brain, and to do so quantitatively, repetitively and non-destructively without altering the function of the brain itself. Its principle relies on the coincidence detection of two 511 keV photons which are predominantly emitted as a result of the annihilation of a positron with an electron. This measurement defines the line along which a nuclear disintegration must have occurred, and hence, along which the radionuclide of interest must have been located. In its simplest form, such a detection system consists of a pair of detectors operating in coincidence and scanning the object at different angles. The only radioactive isotopes that can be used with this technique are those that decay by emitting positrons. Therefore, the emission of positrons and their interaction with matter will be discussed next.
CHAPTER 2

POSITRON EMISSION AND ENERGY LOSSES.

2.1. The positron

2.2. Positron emission
   2.2.1. Positron decay
   2.2.2. Electron capture
   2.2.3. Competition between electron capture and positron decay

2.3. Positron energy loss
   2.3.1. Elastic scattering with nuclei
   2.3.2. Elastic scattering with atomic electrons
   2.3.3. Inelastic collisions with atomic electrons
   2.3.4. Radiative collisions with atomic nuclei

2.4. Positron range
   2.4.1. Path length and range of positrons
   2.4.2. Range-energy relation

2.5. The annihilation process
   2.5.1. One photon annihilation
   2.5.2. Two photons annihilation
   2.5.3. Formation of positronium atoms
   2.5.4. Decay of positronium atoms

2.6. Discussion
2.1. The positron

The positron is the antiparticle of the electron. Its existence was postulated on theoretical grounds in 1928. In that year, P.A.M. Dirac announced his relativistic theory of quantum mechanics. The solutions for $E$ in the equation expressing the relativistic properties of mass and energy and the relation between them:

$$E^2 = m^2c^4 = c^2p^2 + m_0^2c^4$$

lead to the prediction that the allowed values of total relativistic energy $E$ for a free electron are:

$$E = \pm \sqrt{c^2p^2 + (m_0c)^2}$$

where $m_0$ is the electron rest mass. The negative energy levels force the postulate that all these are normally occupied at all points in space, thus removing the objectionable concept of negative total relativistic energy. Then, the exclusion principle would prevent a free electron from dropping into any of these levels, but would allow an electron to be excited into one of the unoccupied positive energy levels. The minimum energy required for this process is:

$$h\nu = 2m_0c^2$$

and the process results in the production of an electron in a positive energy level, leaving a hole in a negative energy level. This hole has the properties of a particle with the
same mass as the electron and a charge of the same magnitude, but opposite in sign: the positron. This process of pair production of an electron and a hole, or of an electron and a positron, was first observed by C.D. Anderson during the course of an experiment on cosmic rays in 1932.

However, positrons are not created only in the process of pair production. They are also emitted during the decay of some unstable nuclei. A nucleus with too few protons tends to become stable by increasing its nuclear charge, while a nucleus with too many protons tends to become stable by decreasing its nuclear charge. In this latter case, the nuclear charge can be decreased either by emission of a positron or by capture of an orbital electron.

2.2. **Positron emission**

2.2.1. **Positron decay**

In positron decay, one proton in the parent nucleus changes into a neutron in the daughter nucleus. Simultaneously, a positron and a neutrino are expelled from the nucleus. The nuclear conversion is equivalent to:

$$\frac{1}{1}p \rightarrow \frac{0}{0}n + \frac{0}{1}e$$

and the type equation is:

$$\frac{A}{Z} \rightarrow \frac{A-1}{Z-1} + \beta^+ + \nu + Q$$
The Q value is given by:

\[ \frac{Q}{c^2} = M_n \left( \frac{A_X}{Z^X} \right) - M_n \left( \frac{A_Y}{Z-1^Y} \right) - m \]

\[ = M_a \left( \frac{A_X}{Z^X} \right) - Zm - M_a \left( \frac{A_Y}{Z-1^Y} \right) + (Z-1)m - m \]

\[ = M_a \left( \frac{A_X}{Z^X} \right) - M_a \left( \frac{A_Y}{Z-1^Y} \right) - 2m \]

where \( M_n \) and \( M_a \) refer to the nuclear mass and the atomic mass respectively, and \( m \) refers to the electronic mass. Since positron emission decreases the atomic number by one unit, the orbitals of the daughter nuclide must lose an electron as soon as the parent nucleus ejects the positron. Thus the atomic mass of the daughter nuclide must be at least two electron masses lighter than that of the parent nuclide.

For positron emission to be energetically possible, the atomic mass of the parent nuclide must therefore be greater than the atomic mass of its isobar with nuclear charge one unit smaller by at least two electron masses:

\[ M_a \left( \frac{A_X}{Z^X} \right) > M_a \left( \frac{A_Y}{Z-1^Y} \right) + 2m \]

The energy made available by the transition is equal to the energy equivalent to the change in mass between the parent and daughter nuclides. Part of this energy is seen as the rest energy of the positron and the electron \((2 m_0 c^2)\) ejected, that of the neutrino (zero), and as the recoil energy of the residual atom (negligible in \(\beta\) decay). The
remaining available energy is carried off as kinetic energy by the positron and the neutrino.

e.g.:

\[
\frac{11}{6} C + \frac{11}{5} B + \beta^+ + \nu + Q
\]

\[
\Delta E = M_a\left(\frac{11}{6} C\right) - M_a\left(\frac{11}{5} B\right)
\]

\[
= 10 \, 256.856 - 10 \, 254.876
\]

\[
= 1.98 \text{ MeV}
\]

The maximum kinetic energy of the positron will be:

\[
1.98 - 1.02 = 0.96 \text{ MeV}
\]

\[
\frac{11}{6} C \quad 20.3 \text{ m}
\]

\[
\frac{11}{5} B
\]

\[
\beta^+ \quad 99\%
\]

\[
\text{EC} \quad 0.19\%
\]

\[
Q_{\beta^+} = 0.96 \text{ MeV}
\]

\[
Q_{\text{EC}} = 1.98 \text{ MeV}
\]

2.2.2. Electron capture

Some proton rich nuclei must decay by a $p \rightarrow n$ conversion, but have daughter products whose mass is greater than the acceptable mass difference for positron emission. These nuclei can only decay by the capture of one of the orbital electrons. The orbital electrons, in the course of their motion, often approach close to the nucleus and, according to wave mechanics, may even penetrate it. The K electrons are those who have the greatest probability density of being in the nucleus and of being captured. The
negative charge thus acquired by the nucleus reduces its atomic number by one unit and yields the same daughter nuclide that would have been produced by positron emission, had this been energetically possible. The nuclear conversion is equivalent to:

\[ \frac{1}{1^p} + \frac{0}{-1^e} \rightarrow \frac{1}{0^n} \]

and the type equation is:

\[ A^X_{Z} \rightarrow A^Y_{Z-1} + \nu + Q \]

The Q value is given by:

\[ \frac{Q}{c^2} = M_n \left( \frac{A^X}{Z^X} \right) - M_n \left( \frac{A^Y}{Z-1} \right) \]

\[ = M_a \left( \frac{A^X}{Z^X} \right) - M_a \left( \frac{A^Y}{Z-1} \right) \]

The condition that orbital electron capture be energetically possible is that the mass of the parent nuclide be greater than that of its isobar with nuclear charge one unit smaller:

\[ M_a \left( \frac{A^X}{Z^X} \right) > M_a \left( \frac{A^Y}{Z-1} \right) \]

The energy made available by the transition is carried off as kinetic energy by the neutrinos which are therefore emitted in one or more monoenergetic groups corresponding to the few excited states of the daughter nuclide. This is in contrast with the continuous distribution of neutrino energy found in positron decay. The vacancy in the K shell is usually filled by an electron from an outer shell with
the resultant emission of x-rays characteristic of the daughter nuclide.

e.g.:

\[ ^{51}_{24}\text{Cr} \rightarrow ^{51}_{23}\text{V} + \nu + Q \]

\[ \Delta E = M_a\left(^{51}_{24}\text{Cr}\right) - M_a\left(^{51}_{23}\text{V}\right) \]

\[ = 47,453.931 - 47,453.179 \]

\[ = 0.752 \text{ MeV} \]

\[ ^{51}_{24}\text{Cr} \quad 27.8 \text{ d} \]

\[ ^{51}_{23}\text{V} \]

\[ 0.320 \text{ MeV} \]

2.2.3. Competition between electron capture and positron decay

Positron decay is possible energetically only if \( M_a\left(A_X\right) > M_a\left(A_Y\right) + 2m \). Electron capture transition has no specific energy requirement except for \( M_a\left(A_X\right) > M_a\left(A_Y\right) \) so this mode of decay is always in competition with positron emission. Nuclides that decay by positron emission will also decay by electron capture.

2.3. Positron energy loss

In electron capture, following the capture of an orbital electron by a nucleus, there will be an orbital vacancy, usually in the K shell. This vacancy will promptly
fill from outer orbits and the transitions will produce the characteristic x-ray spectrum of the daughter element.

In positron emission, as the ejected positron passes through matter, it gives up its kinetic energy to excitations and ionizations. When slowed down enough, the positron will eventually pair with one of the electrons within its range and the two particles will annihilate each other, disposing of their rest energy by emission of two or three gamma rays.

The mechanism by which a positron looses its kinetic energy, or is deflected from its original path, involves four principal types of interactions:

(i) elastic scattering with a nucleus:
In elastic nuclear scattering, the incident positron is deflected but does not radiate, nor does it excite the nucleus. The incident positron looses only the kinetic energy required for conservation of momentum between the two particles.

(ii) elastic scattering with atomic electrons:
In elastic electronic scattering, an incident positron is deflected in the field of the atomic electrons. Energy and momentum are conserved, and the energy transfer is generally less than the lowest excitation potential of the orbital electrons. Such collisions are significant for the case of very low energy (<100 eV) incident positrons.

(iii) inelastic collisions with a nucleus:
In a close, non capture, encounter with a nucleus, the
incident positron experiences a deflection. In some, but not all, such deflections a quantum of radiation is emitted (bremsstrahlung), and a corresponding amount of kinetic energy is lost by the incident positron.

(iv) inelastic collisions with atomic electrons: Inelastic collisions with bound electrons are usually the predominant mechanism by which a positron loses kinetic energy in an absorber. As a result of each such collision, one or more atomic electrons experience a transition to an excited state (excitation) or to an unbound state (ionization).

In an absorbing material, a moving positron is slowed down by the combined action of all four of these elastic and inelastic processes and finally annihilates with an electron. Which type of interaction, if any, will occur when a swift positron passes a particular atom is described only by the laws of chance. From collision theory, the probability of any particular type of collision, of any particular energy loss, and of any particular change in the direction of motion of the incident positron are obtained.

2.3.1. Elastic scattering of positrons by nuclei

A collision of a positron with an atomic nucleus involves Coulomb interactions in which the positron is deflected from its original path. If this interaction is elastic, the energy of the scattered positron is essentially equal to that prior to the collision. This can be deducted from
the fact that the total kinetic energy for the colliding system has not changed, and since the atom is at least several thousand times heavier than the positron, the recoil energy of the nucleus will be negligible.

In the laboratory (L-) coordinates system, the target particle is assumed to be at rest before collision, and is approached by the incident positron.

\[
\begin{align*}
&m, v_0 \\
\rightarrow & v_c \\
O & \rightarrow M
\end{align*}
\]

The centre of mass of these particles moves with a velocity \( v_c \), in a direction parallel to \( v_0 \) such that the linear momentum of the system is:

\[
m v_0 = (M + m) v_c
\]

or

\[
v_c = v_0 \frac{m}{M + m}
\]

In the centre of mass (C-) coordinates system, the centre of mass of the target and incident particles is considered to be at rest. This is equivalent to translating the L-coordinates system at a uniform speed \((-v_c)\).

\[
\begin{align*}
&m, v_0 -v_c \\
\rightarrow & O \\
M & \rightarrow -v_c
\end{align*}
\]

Before collision, the incident particle moves with speed

\[
v_0 - v_c = v_0 \left(1 - \frac{m}{M + m}\right) = v_0 \frac{M}{M + m}
\]
and the target particle moves with speed \((-v_c\)).

The total momentum of the system is zero because the centre of mass is stationary:

\[
m (v_0 - v_c) - M (v_c) = m v_0 \frac{M}{M + m} - M v_0 \frac{m}{M + m}
\]

\[= 0\]

After collision, since the total linear momentum must be conserved, its value must also be zero. The incident particle will move at an angle \(\phi\) with respect to its initial direction; the target particle must move at an angle \((\pi - \phi)\) with respect to the direction of the incident particle. The net results of the collision are that the line joining the two particles is rotated through an angle \(\phi\), and the two particles recede from the centre of mass after the collision.

From conservation of kinetic energy:

\[
\frac{1}{2} m v_0^2 \left(\frac{M}{M + m}\right)^2 + \frac{1}{2} M v_0^2 \left(\frac{m}{M + m}\right)^2 = \frac{1}{2} m v_m^2 + \frac{1}{2} M v_M^2
\]

and from conservation of momentum:

\[m v_m + M v_M = 0\]
the velocities of the two particles after the collision are obtained:

\[ v_m = v_0 \frac{M}{M + m} \quad ; \quad v_M = v_0 \frac{m}{M + m} \]

The total effect in the C-coordinates system is, therefore, to change the direction of the velocities, but not their magnitudes.

In the L-coordinates system, in which the target nucleus was originally at rest, the magnitude of the velocities are changed, and the directions are not opposite. The incident particle, which is scattered through an angle \( \theta \), has now a velocity \( v_m \) which is the vector sum of the velocity of the incident particle in the C-coordinates system and the velocity of the centre of mass, so that:

\[
v_m^2 = v_0^2 \left(\frac{M}{M + m}\right)^2 + v_0^2 \left(\frac{m}{M + m}\right)^2 + 2 \ v_0 \left(\frac{M}{M + m}\right) \left(\frac{m}{M + m}\right) \cos \phi \]
\[
= v_0^2 \left[ 1 - 2 \ (1 - \cos \phi) \ \frac{Mm}{(M + m)^2} \right]
\]
Similarly, the velocity imparted to the target particle is:

\[ v_M^2 = v_c^2 + v_c^2 - 2v_c^2 \cos \phi \]

\[ = 2\left(v_0 \frac{m}{M + m}\right)^2 \left(1 - \cos \phi\right) \]

The relation between the scattering angle \( \theta \) in the L-coordinates system and the scattering angle \( \phi \) in the C-coordinates system is given by:

\[ \cot \theta = \frac{\cos \phi + \frac{m}{M}}{\sin \phi} \]

All these results are independent of the type of interaction between the particles, and the velocities are only the initial and final velocities, i.e., the velocities when the separation of the particles is sufficiently large so that their mutual potential energy is negligible.

If no forces were to exist between the two particles, they would eventually pass each other at a distance \( x \) called the impact parameter. However, as the particles approach each other, the forces which are to be effective in the collision gradually begin to be felt. Since the interaction between a positron and a nucleus is to be due to Coulomb repulsion, the force between the two particles will increase as \( 1/R^2 \), where \( R \) is the distance separating the two particles. During the effective time of impact, the velocities of both particles will change in magnitude and in direction.
The change in magnitude will correspond to the changing potential energy between the two particles, the potential energy of their interaction having been taken as zero when their separation \( R \) was very large.

The initial and final momentum vectors have the same magnitude. The speed of the positron is the same far from the region of collision, either before or after the collision:

\[
P_f = p_i = m v_0
\]

The rate of change of momentum of the positron during the interaction is equal to the force applied to the particle:

\[
\frac{dp}{dt} = F
\]

\[
\int_{p_i}^{p_f} dp = \int_0^\infty F \, dt
\]

\[
P_f - p_i = \Delta p = \int_0^\infty F \, dt
\]

\[
= 2 \, m \, v_0 \, \sin (\theta/2)
\]

The vector \( \int F \, dt \) is the integral of the force vector over the whole time taken for the positron to trace its orbit. It may be regarded as the sum of infinitely many vectors, each of infinitesimal length. They must add up to a vector parallel to the vector \( \Delta p \). All other components add up to zero. Thus, if \( \alpha \) is the angle between the vector \( F \) and \( \Delta p \) at each point of the path, then,
\[ 2 m v_0 \sin(\theta/2) = \Delta p = \int_{0}^{\infty} F \cos \alpha \, dt \]

\[ = \int_{-(\pi-\theta)/2}^{+(\pi-\theta)/2} F \cos \alpha \frac{dt}{d\alpha} \, d\alpha \]

But \( \frac{d\alpha}{dt} \) is the angular velocity \( \omega \). The angular momentum is constant since there is no torque on the positron around the nucleus. The initial value of the angular momentum is \( m v_0 x \). Hence:

\[ m v_0 x = m R^2 \omega \]

and:

\[ \frac{dt}{d\alpha} = \frac{1}{\omega} = \frac{R^2}{v_0 x} \]

Thus:

\[ 2 m v_0^2 x \sin(\theta/2) = \int R^2 F \cos \alpha \, d\alpha \]

Coulomb's law states that the force between the nucleus with charge \( Z e \) and the positron with charge \( e \) is:

\[ F = \frac{Ze^2}{R^2} \]

so:

\[ R^2 F = Ze^2 \]

and:

\[ 2 m v_0^2 x \sin(\theta/2) = \int \frac{Ze^2}{R^2} \cos \alpha \, d\alpha \]

\[ = \frac{Ze^2}{R^2} \cos(\theta/2) \]

or:

\[ \frac{m v_0^2 x}{Ze^2} = \cos \left( \frac{\theta}{2} \right) \quad ; \quad \frac{Ze^2}{m v_0^2 x} = \tan \left( \frac{\theta}{2} \right) \]
where $\theta$ is the angle of scattering of a positron by a nucleus of atomic number $Z$ when it approaches the nucleus with impact parameter $x$.

It is useful to define the collision diameter as the closest distance of approach in a head-on collision between repelling particles. At this distance, $b$, all the kinetic energy of the positron has been converted to electrostatic potential energy.

$$\frac{Z e^2}{b} = \frac{1}{2} m v_0^2$$

or:

$$b = \frac{2 Z e^2}{m v_0^2}$$

and:

$$\tan(\theta/2) = \frac{b}{2x}$$

or:

$$x = \frac{b}{2} \cot(\theta/2)$$

The expression for the differential cross section for an impact parameter between $x$ and $x + dx$ is simply the area of a ring of radius $x$ and width $dx$:

$$d\sigma = 2\pi x \, dx$$

$$= 2\pi \frac{b}{2} \cot(\theta/2) \frac{b}{4} \csc^2(\theta/2) \, d\theta$$

$$= \frac{\pi b^2}{4} \cot(\theta/2) \csc^2(\theta/2) \, d\theta$$

An impact parameter of $x$ or less will produce a deflection through an angle $\theta$ or greater. Therefore, the cross section for scattering through an angle $\theta$ or greater is:
\[ \sigma(\theta) = 2\pi \int_0^{x} x \, dx = \int_0^{\theta} d\sigma \]
\[ = \frac{\pi}{4} b^2 \cot^2(\theta/2) \]

which becomes, for a positron of charge e, velocity \( v_0 = \beta c \) and mass \( m = m_0 / \sqrt{1 - \beta^2} \):
\[ \sigma(\theta) = \pi Z^2 \left( \frac{e^2}{m_0 c^2} \right)^2 \left( 1 - \frac{\beta^2}{\beta^4} \right) \cot^2(\theta/2) \text{ cm}^2/\text{atom} \]

2.3.2. **Elastic scattering with atomic electrons**

The permissible scattering angles for an elastic collision between a positron and an atomic electron are always small because the two particles are of similar masses and because the energy imparted to the struck electron must not exceed its binding energy so as not to produce ionization or excitation.

2.3.3. **Inelastic collisions with atomic electrons**

In the L-coordinates system, a collision confers on the struck particle \( M \) a velocity \( v_M \) in the direction \( \theta \) if it was initially at rest and unbound. Thus the struck particle acquires a kinetic energy and this energy is lost by the incident positron. The energy loss is:
\[ Q = \frac{1}{2} M v_M^2 \]
\[ = \frac{1}{2} M 2 \left( v_0 \frac{m}{M + m} \right)^2 \sin^2(\theta/2) \]
and its differential is:
\[ dQ = 2 v_0^2 \frac{Mm^2}{(M + m)^2} \sin(\phi/2) \cos(\phi/2) \, d\phi \]

so that:
\[ d\sigma = 2\pi \times dx \]
\[ = \pi \frac{b^2}{4} \cot(\phi/2) \csc^2(\phi/2) \, d\phi \]
\[ = \pi \frac{b^2}{4} \frac{1}{\sin^2(\phi/2) \sin(\phi/2)} \cos(\phi/2) \, d\phi \]
\[ = \pi \frac{b^2}{4} v_0^2 \frac{Mm^2}{(M + m)^2} \frac{dQ}{Q^2} \]

This equation shows that the cross section for energy transfer will be most important for those struck particles having a small mass \( M \). Energy transfers to nuclei are therefore unimportant compared with those to atomic electrons. Hence most collisions of positrons with nuclei will be elastic collisions whereas most collisions with atomic electrons will be inelastic collisions.

The classical cross section per atomic electron for an energy transfer between \( Q \) and \( Q + dQ \) is therefore:
\[ d\sigma = \frac{2\pi e^4}{m v_0^2} \frac{dQ}{Q^2} \, \text{cm}^2/\text{electron} \]

When this incident particle traverses a distance \( ds \) in an absorber which contains \( N \) atoms/cm\(^3\), each with \( Z \) electrons/atom, the expectation value for the energy loss per unit path length \( dT/ds \) is:
\[
\frac{dT}{ds} = N Z \left[ \int_{Q_{\min}}^{Q_{\max}} Q \, dq \right] \text{ ergs/cm}
\]

\[= \frac{2\pi e^4}{m v_0^2} N Z \ln \left( \frac{Q_{\max}}{Q_{\min}} \right) \]

The minimum energy loss is the mean of all ionization and excitation potentials of the absorbing atom, I. The maximum energy loss is \(m v_0^2\), so that:

\[
\frac{dT}{ds} = \frac{2\pi e^4}{m v_0^2} N Z \ln \left( \frac{m v_0^2}{I} \right)
\]

With relativistic correction terms, the quantum mechanical formula per atom is:

\[
\frac{dT}{ds} = \frac{2\pi e^4}{m v_0^2} N Z \ln \left[ \frac{m v_0^2}{I^2 (1 - \beta^2)} \right] - \frac{\beta^2}{2}
\]

where \(T\) is the energy transfer and \(\beta = v_0/c\).

For nonrelativistic incident positrons (\(I << T << m_0 c^2\)) the equation becomes:

\[
\frac{dT}{ds} = \frac{2\pi e^4}{T} N Z \ln \left( \frac{T\sqrt{2}}{I} \right) \text{ ergs/cm}
\]

The cross section per atom (N atoms per unit volume) for a fractional energy loss \(dT/T\) due to an ionizing collision is a measure of the probability that such an energy loss will occur per length \((ds)\) of absorber:
2.3.4. Radiative collisions with atomic nuclei

Whenever an incident charged particle is deflected from its path or has its velocity changed, it should emit electromagnetic radiations whose amplitude is proportional to the acceleration. The acceleration produced by a nucleus of charge Ze on a positron of charge e and mass m is proportional to Ze²/m. Thus the intensity, which is proportional to the square of the amplitude, will vary as Z²e⁴/m².

In an individual deflection by a nucleus, the incident particle can radiate any amount of energy from zero up to its total kinetic energy T. The differential cross section for the emission of a photon in the energy range between hν and hν + d(hν) by incident positrons of kinetic energy T and total energy T + m₀c² can be written:

\[ d\sigma = \sigma_0 B Z^2 \frac{T + M_0c^2}{T} \frac{d(h\nu)}{h\nu} \text{ cm}^2/\text{nucleus} \]

\[ \sigma_0 = \frac{1}{137} \left( \frac{e^2}{m_0c^2} \right)^2 \]

and B is a very slowly varying function of Z and T, of the
order of magnitude of 10; \(1/137\) is the fine structure constant \(e^2/\hbar c\).

Integration gives the total energy loss per unit path length due to Bremsstrahlung:

\[
\frac{dT}{ds} = N \int \frac{h\nu}{d\sigma} \text{ erg/cm} = N \sigma_0 Z^2 (T + m_0 c^2) \int_0^1 B \frac{d\left(\frac{h\nu}{T}\right)}{B}
\]

The total bremsstrahlung cross section is defined as the fraction of the total energy \((T + m_0 c^2)\) of the positron which is radiated as the positron traverses an absorber of unit thickness that contains 1 atom/cm^2. Then:

\[
\sigma = \frac{\frac{dT}{ds}}{\frac{1}{N ds}} = \frac{1}{\sigma_0 Z^2} \int_0^1 B \frac{d\left(\frac{h\nu}{T}\right)}{B}
\]

\[
= \sigma_0 Z^2 \frac{1}{\overline{B}}
\]

where \(\overline{B}\) is the average value of \(B\) over the domain \(h\nu = 0\) to \(h\nu_{\text{max}} = T\). For the nonrelativistic case, \(T \ll m_0 c^2\), an approximate expression arises from the detailed quantum mechanical theories:

\[
\sigma = \frac{16}{3} \sigma_0 Z^2
\]

\[
= \frac{16}{3} \left(\frac{e^2}{m_0 c^2}\right)^2 \frac{Z^2}{137} \text{ cm}^2/\text{atom}
\]
2.4. Positron range

2.4.1. Path length and range of positrons

As the positron passes through matter, it loses its energy in ionizing and radiative collisions. In each of these it may suffer significant deflections. In addition, there is a large number of deflections due to elastic scattering. The net result is that the positron's path as it passes through an absorber is very tortuous and its range, the thickness of an absorber which the particle can just penetrate, is very much shorter than its path length.

The statistical distribution of energy losses, or straggling, for each positron will result in a broad distribution of path lengths. Some positrons will have large losses and short path lengths, others will suffer smaller losses per collision and have longer path lengths. However, it is theoretically possible to calculate the path length of a positron using the relations that have been developed. Thus the average path length of a positron of kinetic energy $T$ is:

\[
\bar{S} = \int ds = \int_{T_1}^T \frac{dT}{(dT/ds)}
\]

where:

\[
\frac{dT}{ds} = \left(\frac{dT}{ds}\right)_{\text{ion}} + \left(\frac{dT}{ds}\right)_{\text{rad}}
\]

Ionization losses vary roughly as $1/\beta^2$ so that they are largest for slow particles. On the other hand, radiative losses increase with increasing energy. In the energy range
of biologically useful positrons ($E_{\text{max}} < 2 \text{ MeV}$), most of the losses will be ionization losses.

2.4.2. **Range-energy relation**

The previously derived mean value for $(dT/ds)_\text{ion}$ and $(dT/ds)_\text{rad}$ can only serve as guides in the estimation of theoretical ranges for positrons. In most practical situations, completely empirical range-energy relations are used.

Ionization losses per unit distance along the path of the primary particle are proportional to the number of atomic electrons per cm$^3$ of the absorber, $N Z$. If $\rho \text{ g/cm}^3$ is the density of the absorber, $A$ its atomic weight, $N_0$ Avogadro's number and $N$ the number of atoms per cm$^3$, then:

$$N Z = \left(\frac{\rho N_0}{A}\right) Z = \rho N_0 \left(\frac{Z}{A}\right)$$

The ratio $Z/A$ is nearly constant for all elements and equal to $0.45 \pm 0.05$, except for hydrogen for which it is approximately 1. Therefore, $N Z / \rho$ is approximately constant for all elements. Consequently, if distances along the path of the positron are measured in units of $\rho ds = d\omega \text{ g/cm}^2$, the ionization losses $dT/d\omega$ (ergs/g cm$^{-2}$) become approximately independent of the absorbing material.

If the absorbing material is composed of a mixture of elements, the term $Z/A$ could be replaced by:

$$\left(\frac{Z}{A}\right) f_a + \left(\frac{Z}{A}\right) f_b + \ldots + \left(\frac{Z}{A}\right) f_n$$

where $f_a$, $f_b$, ..., $f_n$ represent the amounts of elements $a$, $b$, ..., $n$
as a proportion of the molecular weight of the material. However, since the factor $Z/A$ is nearly constant, the ionization losses will remain approximately independent of the absorbing material, and will depend only on its density. Ranges can therefore be converted directly from one absorber to another, since energy losses depend primarily on the number of electrons encountered and not upon the atomic number or chemical composition of the absorber. If a maximum range $r_0$ is in g/cm$^2$, then $r_0/\rho$ will be the maximum range in cm, where $\rho$ is the density in g/cm$^3$.

2.5. The annihilation process

The annihilation of positrons may occur either when the positron is in motion or when it has already been stopped. Because of the high ionization cross section compared to the annihilation cross section, the fast positrons lose their energy rapidly by ionization and by excitation; annihilation becomes significant only at a positron energy of a few eV. The fate of such low energy positrons prior to annihilation depends on the nature of the atomic or molecular systems in which they are present. The positron can capture an electron in a rearrangement collision and form the bound state positronium, a hydrogen equivalent positron-electron two body system. Other bound state formations such as positron-atom and positron-molecule
are also feasible, but are not predominant. The positrons that do not form bound states annihilate from scattering centres.

2.5.1. **One photon annihilation**

When slow positrons interact with electrons which are tightly bound to a nucleus capable of absorbing the recoil momentum, a single photon can be emitted. This is the only process in which energy and momentum conservation does not exclude one photon annihilation. This process yields essentially monoenergetic photons and occurs with a small, but finite, probability. When, however, annihilation takes place in the absence of a buffer capable of taking up the recoil momentum, one photon annihilation is rigorously forbidden by kinematics considerations, and two or more photons are simultaneously emitted at an angle to one another.

2.5.2. **Two photon annihilation**

In the absence of internal fields, conservation of energy and momentum require that the lowest order process yields two photons emitted (in the centre of mass system) with equal energy \( e, m_0 c^2 = 0.511 \text{ MeV} \), in opposite directions.

The cross section for two photon annihilation of a slow positron with a free low energy electron is given by:

\[
\sigma_{2\gamma} = \frac{\pi r_0^2 c}{v}
\]

where \( v \) is the relative positron-electron velocity (\( v \ll c \)) and
\[ r_0 = \frac{e^2}{mc^2} \] 

is the classical electron radius. Consequently the annihilation probability of a slow positron is given by:

\[ \lambda_{2\gamma} = \sigma_{2\gamma} n v = \pi r_0^2 c n \]

and is independent of the positron velocity and simply proportional to the electron density \( n \).

In the derivation of this formula, the positron-electron Coulomb interaction was not included. The changes required in the formula to take into account the full many body system (one positron-many electrons) encountered in real situations, such as in positron annihilation in a gas or in condensed matter, have been the subject of many theoretical investigations and yield the formula:

\[ \sigma_{2\gamma} = \pi r_0^2 c n_{\text{eff}} \]

where \( n_{\text{eff}} \) is the effective density of electrons as seen by the positron. In general, then, the rate of annihilation in collisions between free positrons and free electrons should be proportional to the density of available electrons and, hence, in a given substance, to the density. However, this proportionality of annihilation rate and density was observed not to be satisfied and lead to the demonstration that the annihilation of slow positrons sometimes involves the formation of a positron-electron bound state, i.e., a positronium.
2.5.3. Formation of positronium atoms

It was observed by Deutsh in 1951 that positrons in freon (CCl₂F₂) exhibit a decay rate indicating an electron density independent of gas pressure. It was also shown that, in this case, a large fraction of the annihilation spectrum presented the continuous energy distribution characteristic of three photon decay. It was suggested that positrons which have been slowed down sufficiently by passage through matter can form an intermediate metastable system with electrons in which both conjugate particles revolve about their mutual centre of gravity in a hydrogen-like atom: the positronium atom. The positronium formation probability depends on the nature of the environment; in water, about 36% of all positrons form positronium atoms. The gross structure of positronium states is similar to that of the hydrogen atom. However all energy levels are reduced by a factor of two due to the reduced mass ratio. The ionization potential is 6.8 eV and the first excited state is 5.1 eV.

When the kinetic energy $T$ of the positron is less than the lowest energy required to excite a level in the moderating medium, $V_c$, but larger than the minimum energy necessary for radiationless electron capture to form positronium ($V_i - 6.8$ eV), $V_i$ being the ionization energy, positronium formation is a predominant effect (fig. 2.1). This kinetic energy is called the Ore gap. Below 6.8 eV,
ionization potential, or binding energy of positronium

→ inelastic collisions →

0 6.8 eV $V_i$ $V_c$ energy

| ionization | minimum positron |
| energy of  | kinetic energy   |
| the gas   | required for     |
|           | inelastic collisions |

If $V_c > V_i - 6.8$ eV

then $V_c > E > V_i - 6.8$ eV

where $E = \text{range of energies leading to positronium formation.}$

Figure 2.1: Energetics of positronium formation.
positrons cannot form positronium atoms. Positroniums are formed through this mechanism in low density gases. In liquids and solids, positronium formation does not have to occur via the Ore mechanism; instead, positronium can be formed by the capture of an electron in the ionization trail created by the slowing down of the positron.

The ground states of the positronium atom are the singlet $^1s$ state or parapositronium, in which the orbital angular momentum $l$ and the total spin are zero; and the triplet $^3s$ state or orthopositronium, in which $l \neq 0$ and $s = 1$. Three-quarters of the positronium atoms are formed in the triplet state ($m = 1$, $l = 0$, $s = 1$, $m_s = -1$ or 0 or +1), the remaining one-quarter are formed in the singlet state ($m = 1$, $l = 0$, $s = 0$, $m_s = 0$).

2.5.4. Decay of positronium atoms

The charge parity of the positronium system is:

$$C = P_i P_1 P_s$$

where $P_i$ is the intrinsic parity which is negative for a particle-antiparticle pair, $P_1 = (-1)^l$ is the spatial parity and $P_s = (-1)^{s+1}$ is the spin parity. Thus for positronium, $C = (-1)^{l+s}$ is +1 for the singlet state and -1 for the triplet state.

The charge parity of the photon is negative and for a system of $n$ photons is $C = (-1)^n$.

The general selection rule for the annihilation of
the positronium atom is therefore given by:

\[ (-1)^{l+s} = (-1)^n \]

It follows that the parapositronium atom decays into an even number of photons and the orthopositronium atom into an odd number greater than one.

The state of positronium shows no first order (linear) Zeeman effect because the expectation value of the total magnetic moment of positronium vanishes in the triplet as well as in the singlet spin state. However, there are second order effects due to the mixing of singlet and triplet components in a magnetic field. In this case, due to the admixture of the two states, the triplet state can now decay by two photon annihilation, the applied magnetic field absorbing the unbalanced angular momentum. This process which causes atoms in the triplet state to be destroyed at a faster rate than the natural decay rate is called quenching. There are other quenching mechanisms, apart from that due to external magnetic fields: triplet-singlet conversion in elastic collisions, pick off annihilation in collisions with gas molecules which always contain electrons with the appropriate spin orientation for rapid two photon decay and formation of positronium compounds.
2.6. Discussion

Emission tomography attempts to map the distribution of a radioactive tracer in a cross section of the body. In single photon emission tomography, any radioactive nuclide emitting gamma rays can be used. In positron tomography, only those that decay by positron emission can be used.

Proton rich nuclides will decay either by electron capture or by positron emission or both. The positron ejected in such transformation will carry a maximum kinetic energy that is equal to the mass difference between the parent and daughter nuclides, minus two electron rest masses. In an absorber, the positron will lose most of its kinetic energy in inelastic collisions with atomic electrons. It will also be deflected from its path by elastic collisions with nuclei. The maximum path length and the range will vary from positron to positron depending on the amount of energy lost in excitation or ionization events and on the number of scattering events. The range of a positron in an absorber depends on the number of electrons encountered. However, if the range is expressed in g/cm$^2$, then it is independent of the atomic number and the chemical composition of the absorber. When the positron is slowed down sufficiently, it will eventually pair with a free electron and the two particles will annihilate each other, disposing of their rest energy by emission of two 0.511 MeV photons in most cases.
The finite range of positrons in tissue will limit the spatial resolution of positron imaging systems. This range depends on the energy of the emitted positron and is typically less than 2 mm for radioisotopes of biological interest. Another limit to spatial resolution is the deviation from strict colinearity between the emitted photons. Because of thermal motion, the conjugate positron and electron have a small, but finite, angular momentum that will cause an angular distribution of the order of 0.5° about a mean angle of 180° between the two annihilation photons. However, the most important restriction on spatial resolution is imposed by the relatively high energy (in imaging terms) of the annihilation photons. This determines to a large extent the choice of detectors and their size which will be discussed in the following chapter.
### Chapter 3

**DETECTORS FOR POSITRON TOMOGRAPHY**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1. Introduction</td>
<td>60</td>
</tr>
<tr>
<td>3.2. General properties of radiation detectors</td>
<td>61</td>
</tr>
<tr>
<td>3.2.1. Energy resolution</td>
<td>61</td>
</tr>
<tr>
<td>3.2.2. Detection efficiency</td>
<td>64</td>
</tr>
<tr>
<td>3.2.3. Photopeak fraction</td>
<td>65</td>
</tr>
<tr>
<td>3.2.4. Time resolution</td>
<td>65</td>
</tr>
<tr>
<td>3.2.5. Spatial resolution</td>
<td>68</td>
</tr>
<tr>
<td>3.3. Proportional counters</td>
<td>70</td>
</tr>
<tr>
<td>3.3.1. General principles</td>
<td>70</td>
</tr>
<tr>
<td>3.3.2. Multiwire proportional counters</td>
<td>71</td>
</tr>
<tr>
<td>3.3.3. Discussion</td>
<td>74</td>
</tr>
<tr>
<td>3.4. Solid state detectors</td>
<td>76</td>
</tr>
<tr>
<td>3.4.1. General principles</td>
<td>76</td>
</tr>
<tr>
<td>3.4.2. Germanium</td>
<td>78</td>
</tr>
<tr>
<td>3.4.3. Cadmium telluride</td>
<td>79</td>
</tr>
<tr>
<td>3.4.4. Mercuric iodide</td>
<td>79</td>
</tr>
<tr>
<td>3.4.5. Discussion</td>
<td>79</td>
</tr>
<tr>
<td>3.5. Scintillation detectors</td>
<td>80</td>
</tr>
<tr>
<td>3.5.1. General principles</td>
<td>80</td>
</tr>
<tr>
<td>3.5.2. Sodium iodide</td>
<td>88</td>
</tr>
<tr>
<td>3.5.3. Caesium fluoride</td>
<td>88</td>
</tr>
<tr>
<td>3.5.4. Barium fluoride</td>
<td>89</td>
</tr>
<tr>
<td>3.5.5. Bismuth germanate</td>
<td>90</td>
</tr>
<tr>
<td>3.6. Conclusions</td>
<td>90</td>
</tr>
</tbody>
</table>
3.1. **Introduction**

The study of radioactivity involves the detection and measurement of radiation. The quantitative measurement requires knowledge of the number of photons arriving at a detector per unit time, and their energy. In addition, in positron tomography, the precise time at which the photon strikes the detector is needed.

The gamma ray detectors are a crucial component of any positron imaging system. It is their detection efficiency, time resolution, spatial resolution and their configuration which determine the quality of the eventual image. The various detectors that have been used in positron tomography can be classified into three groups: proportional counters, solid state detectors and scintillation detectors. Each group has advantages and disadvantages that have to be considered before the choice for a system is made. However, it might be useful to first outline some general properties which apply to all types of detectors, including definition of terms such as detection efficiency and spatial resolution.
3.2. General properties of radiation detectors

When photons pass through the detector's material, they impart energy, through photoelectric and Compton interactions, to the charged particles of the absorber which then causes ionization. This ionization is the basis of nearly all the instruments used for the detection of photons and the measurement of their energies. The net result of the radiation interaction is therefore the appearance of a given amount of electric charge within the detector's active volume [this is strictly true for the first two types of detectors: proportional counters and solid state detectors; for scintillation detectors, the charge is formed, indirectly, in a photomultiplier tube]. This charge must then be collected to form an electrical signal that can be processed further. This electrical signal will contain information about the amount of energy deposited in the detector and about the instant in time at which the photon struck the detector.

3.2.1. Energy resolution

Each individual pulse amplitude carries information regarding the charge generated by the particular photon that interacted in the detector. If a large number of such pulses are observed, their amplitude will not be the same. Variations may be due either to differences in the radiation energy or to fluctuations in the inherent response of the detector.
to monoenergetic radiation. The pulse amplitude distribution is a fundamental property of the detector output and is generally characterized by the term energy resolution.

The energy resolution of a detector is conventionally defined as the full width at half maximum (FWHM) divided by the energy of the incident gamma ray (fig. 3.1). Detectors with good energy resolution (small $\text{FWHM} / E_0$) will be able to distinguish between two radiations whose energies lie near each other (fig. 3.2).

There are a number of potential sources of fluctuations in the response of a given detector which result in degradation in energy resolution. These include any drift of the operating characteristics of the detector during the course of the measurement, sources of random noise within the detector and statistical noise arising from the discrete nature of the measured signal itself. The statistical noise represents an irreducible minimum amount of fluctuation that will always be present in the detector signal no matter how perfect the remainder of the system is made. This statistical noise arises from the fact that the charge generated within the detector by a quantum of radiation is not a continuous variable, but instead represents a discrete number of charge carriers and is therefore subject to random fluctuations from event to event even though exactly the same amount of energy is
Figure 3.1: Energy spectrum. Energy resolution is defined as the full width at half maximum divided by the energy of the incident gamma ray.

Figure 3.2: Energy spectrum of $^{57}$Co obtained using a Ge(Li) detector (solid line) and a NaI (Tl) detector (dotted line). Note that because of its poorer energy resolution, NaI (Tl) does not resolve the 136 keV peak.
deposited in the detector.

3.2.2. Detection efficiency

In principle, all radiation detectors will give rise to an output pulse for each quantum of radiation which interacts within its active volume. Gamma rays can travel relatively large distances before interacting and can also traverse the detector without interacting and therefore without being detected. It is therefore necessary to have a precise figure for the detection efficiency, or ability to detect gamma rays, in order to relate the number of pulses counted to the number of photons incident on the detector.

Counting efficiencies are usually subdivided into two classes: absolute and intrinsic efficiencies. Absolute efficiency is defined as the ratio of the number of pulses recorded to the number of radiation quanta emitted by the source. It depends not only on detector properties, but also on the details of the counting geometry. A more useful term for comparison of various detectors is the intrinsic efficiency which is defined as the ratio of the number of pulses recorded to the number of radiation quanta incident on the detector. The two efficiencies are simply related for isotropic sources by the solid angle of the detector seen from the actual source position.

The intrinsic efficiency of a detector usually
depends primarily on the detector material, the radiation energy and the physical thickness of the detector in the direction of the incident radiation.

3.2.3. Photopeak fraction

Counting efficiencies are also characterized by the nature of the event recorded. If all pulses from the detector are accepted, then it is appropriate to use total efficiency. In this case, all interactions, no matter how low in energy, are assumed to be counted. The peak efficiency, however, assumes that only those interactions which deposit the full energy of the incident radiation are counted (fig. 3.3). The photopeak fraction is defined as the ratio of the number of events in the photopeak to the total number of events recorded. In applications where only the photopeak events are considered, this ratio must be high.

3.2.4. Time resolution

In positron tomography, the precise arrival time of a quantum of radiation in the detector is of particular interest. The accuracy with which timing can be performed depends both on the properties of the specific detector and the type of electronics used to process the signal. The best timing performance is obtained for the fastest detectors, that is, those in which the signal charge is collected most rapidly. For detectors with equal charge collection times, those that generate the greatest number
Figure 3.3: The photopeak fraction is defined as the ratio of the number of events in the photopeak to the total number of events in the spectrum.

Figure 3.4: Time spectrum (one branch of the time to amplitude converter has been delayed 500 nsec). The time resolution is defined as the full width at half maximum of the time peak.
of information carriers per pulse will demonstrate superior timing properties.

The fundamental measure in time resolution is the time spectrum (fig. 3.4). It bears a close relation to the pulse height spectrum and the fundamental measure in time resolution is the FWHM of the time distribution. When two independent detectors are irradiated by a common radioisotopic source which is assumed to emit at least two detectable quanta in true coincidence, that is, both radiations arise from the same nuclear event within the source and that there is no appreciable time delay between the emission of the two quanta, some fraction of all true coincidence events will give rise to radiations which are detected simultaneously in both detectors. In addition to the coincident events, each detector will produce typically a much larger number of pulses which correspond to the detection of one quantum for which there may not be a corresponding coincident emission, or for which the coincident radiation escapes detection in the opposite detector. For these events there can be, therefore, no true coincidence. However, because of their random distribution in time, some sequences will occur by chance in which a quantum will strike one detector within a very short time of an unrelated quantum striking the other detector. These events are called
chance coincidences and their intensity depends on the rate at which pulses are generated in either detector (fig. 3.4).

3.2.5. Spatial resolution

Spatial resolution, the ability to resolve detail in the image, is not an intrinsic property of a radiation detector, but is a crucial parameter in positron tomography. In an ideal imaging system, each point in the object would be represented as a point in the image. However, no ideal system exists and the image of a point is always smeared out in all directions (fig. 3.5). The amount of smearing is measured by the FWHM of the distribution and is referred to as the point spread function. The extent to which an imaging system will spread the image of a point will govern the minimum separation between two points that is necessary to image them as distinct objects, and therefore governs spatial resolution (fig. 3.5). Although spatial resolution does not depend directly on the radiation detector used, it does depend on detector configuration and should be considered in a discussion of detectors for positron tomography.
Figure 3.5: Spatial resolution. The spatial resolution is defined as the full width at half maximum of the response of the tomograph to a point source of radioactivity.
3.3. Proportional counters

3.3.1. General principles

The proportional counter is a type of gas filled detector that relies on the phenomenon of gas multiplication to amplify the charge represented by the original ion pair created within the gas. Gas multiplication is a consequence of increasing the electric field within the gas to a sufficiently high value. Free electrons in the gas are accelerated by the applied field and will collide with the gas molecules. If the kinetic energy of the electrons is greater than the ionization energy of the neutral gas molecules, it is possible for additional ion pairs to be created in the collisions. The electrons liberated by this secondary ionization process will also be accelerated by the electric field. During its subsequent drift, they undergo collisions with other neutral gas molecules and thus create additional ionization. The gas multiplication process therefore takes the form of a cascade in which free electrons created in a collision can potentially create more free electrons by the same process. The avalanche terminates when all free electrons have been collected at the anode.

Proportional counters are used for the detection of soft x rays or gamma rays whose energy is low enough to interact with reasonable efficiency in the counter gas. At higher photon energies, the direct interaction probability
of the photons in the gas drops rapidly with increasing energy. In this case, an interacting medium, with an appropriate stopping power can be intercalated between the source of photons and the proportional counter. Photons interacting in this medium will create secondary free electrons which will reach the counter chamber and create an avalanche. In this case, because of the variable loss of electron energy before it reaches the gas, the proportional counter cannot be used to measure the energy of the incident photons.

3.3.2. Multiwire proportional counters

In some situations, it is advantageous to provide more than one anode wire within a proportional counter. Detectors with large surface areas can be economically constructed by placing a grid of anode wires between two large flat plates which serve as cathodes on either side of the counter (fig. 3.6). Ions formed in the gas filled volume drift toward the nearest anode wire where they undergo multiplication. Because the signal appears only on one anode wire, the event is automatically localized in the dimension perpendicular to the anode wires.

In positron tomography, lead converters are placed in front of multiwire proportional counters detectors to convert the annihilation photons to electrons and so use the multiwire proportional counters to determine position.
Figure 3.6: A multiwire proportional counter.

Figure 3.7: A multiwire proportional counter with lead converter. Photons are stopped by the lead and produce fast electrons that can escape to an adjacent hole. The electric field multiplies and extracts the resulting electrons that are detected by the multiwire proportional counter.
The converter consists of a sandwich of lead and fibreglass sheets glued together, in which a large number of holes are drilled close together (fig. 3.7). Photons interact in the lead bars and produce fast electrons which can escape to an adjacent hole. The free electrons, resulting from gas ionization in a hole, may be multiplied and extracted by a strong electric field and detected by a standard wire chamber.

The choice of gases with which to fill the chambers and which governs the gas multiplication is important in the design of multiwire proportional chambers. Penning mixtures are preferred. Photoionization in these gas mixtures does not provide the electron multiplication; it results rather from the Penning effect. In this effect, electron collisions excite gas atoms to metastable states which cannot decay by photon emission; instead, the excited atoms ionize the secondary gas by collision, provided that the metastable energy exceeds the ionization potential of the secondary gas. Therefore a new mode of chamber operation is possible: electron amplification in the holes of the converters with further amplification in the wire chamber itself. Neon-carbon dioxide mixtures were introduced in 1982 and present several advantages over the argon-hydrocarbon and neon-hydrocarbon mixtures used previously. Time resolution, previously very poor for multiwire proportional chambers,
has been improved fourfold owing to the higher electron drift velocity in neon. The use of carbon dioxide, which also has a high drift velocity at high fields, instead of hydrocarbons results in a non polymerizing gas mixture for a long chamber life and a safer operation.

3.3.3. Discussion

In positron imaging using multiwire proportional chambers, the object to be imaged is placed between the two chambers. By backprojecting all the coincident gamma ray pairs detected by the chamber, a blurred three dimensional image of the object is obtained. The overall spatial resolution of the camera is determined by several factors: the chamber resolution and the parallax errors due to the converter thickness. Other factors to consider are energy resolution and detection efficiency.

Although there is no correspondance between photon energy and chamber pulse height, the camera does have an important intrinsic sensitivity to photon energy. As the photon energy falls, the interaction probability rises as the photoelectric effect dominates. This would increase the detection efficiency, but it is more than offset by the fact that the lower energy photoelectrons have lower probabilities of escaping from the lead. Down to about 200 keV, detection efficiency falls slowly, after which it drops rapidly. Thus the camera is insensitive to a significant
amount of the scattered radiation that exists in a clinical imaging situation.

Increasing the detection efficiency requires a thicker converter to stop more photons, whereas improving the time resolution demands a thinner converter, as the time resolution is dominated by the drift time of the free electrons through the converter. The drift time within a converter may be reduced by optimizing the gas mixture in the chamber and the drift voltage applied to the converter. A time resolution of 100 ns is achieved using a mixture of 70% methane and 30% neopentane; a time resolution of 20 ns can be achieved using a mixture of 93% neon and 7% carbon dioxide.

The camera's limitations on spatial resolution are the chamber wire pitch, the precision of the read out, the converter's hole geometry and the counter's parallax error. These errors can be kept to within a millimeter by careful construction of the chamber, yielding an overall spatial resolution of 2 to 3 mm.

Because of the limited solid angle acceptance of the camera, the spatial resolution in the direction perpendicular to the chambers is degraded, and so the image is treated as a series of thick slices parallel to the chamber. Another disadvantage of multiwire proportional chambers is the poor utilization of the available photons.
To improve spatial resolution and to increase the detection of the available photons, more chambers can be placed around the object to be imaged.

3.4. **Solid state detectors**

3.4.1. **General principles**

Semiconductors are a special class of solid state materials characterized by a regular, systematic bonding of the atoms in a crystal lattice structure. The discrete, allowed energy levels of the individual atoms are smeared into allowed energy bands because of the close proximity of all the atoms. In semiconductors and in insulators, the chemical bonding is such that the highest energy band (the valence band) is completely filled in the absence of thermal excitation. The only basic difference between insulators and semiconductors is in the magnitude of the energy gap to the next higher allowed energy band (the conduction band). In semiconductors, this difference is small enough so that some electrons can be thermally excited into the conduction band and contribute to conductivity (fig. 3.8).

When a gamma ray enters a detector, any of the three primary interactions may take place between it and the electrons of the material: a photoelectric interaction, a Compton interaction or pair production. Whereas a photoelectric event produces an amount of ionization corresponding directly to the incident gamma ray energy,
Figure 3.8: Distribution of electrons in energy bands. In a conductor, the valence band is full and the conduction band is partly filled. In an insulator and a semiconductor, the valence band is full and the conduction band is empty. In a semiconductor, the width of the forbidden gap is of the order of \( kT \), thus allowing electrons to gain enough thermal energy to jump to the conduction band and contribute to the flow of electricity. In an insulator, this gap is much greater than \( kT \) and cannot be overcome by electrons.
Compton events produce a variable amount of ionization. Only if the degraded, less energetic, secondary gamma ray resulting from Compton interactions is fully absorbed can it contribute useful information about the distribution of gamma ray energies.

The gamma photons, in interactions with the material of the semiconductor, will release electrons and create holes. These electrons and holes will travel at high velocities initiating showers of free electrons and holes, the final members of which are traveling at thermal velocities in the lattice. The free charges produced in the detector will either recombine or migrate under the influence of the applied electric field until they are collected at an electrode. The impulse of current flowing in the external circuit while the free charges are in transit measures the energy of the photon producing the signal.

3.4.2. Germanium

The energy gap, determining intrinsic conductivity, of germanium is 0.67 eV at room temperature. Thus germanium detectors must be cooled, usually to 150°K or lower, to reduce the number of thermally generated carriers to an acceptable level. A charge pair is produced on average for each 2.9 eV of energy from the incident photon. The energy resolution is therefore very good and is not limited by statistical variations. This allows a significant amount of
the scattered radiation to be eliminated. However, the efficiency of germanium is generally poor at energies above 100 keV.

3.4.3. Cadmium telluride

Cadmium telluride combines a relatively high Z value with a large enough band gap energy (1.47 eV) to allow room temperature operation. The probability of photoelectric absorption per unit path length is roughly a factor of four to five higher in cadmium telluride than in germanium.

Because of the rather poor collection efficiency for holes, energy resolution is generally not comparable with that obtained in germanium.

A persistent problem in the use of cadmium telluride is the phenomenon of polarization which, under certain conditions of operation, lead to a time dependent decrease in the counting rate and charge collection efficiency.

3.4.4. Mercuric iodide

Mercuric iodide combines a high Z constituent with a wide band gap energy (2.10 eV). This allows room temperature operation without excessive thermally generated noise. It exhibits an inherent low mobility-lifetime product for holes so that complete charge collection is very difficult.

3.4.5. Discussion

The main drawback of semiconductor detectors in positron tomography is their relatively low stopping power and hence detection efficiency. An additional drawback of
germanium detectors is the need to operate them at liquid nitrogen temperatures. The main advantage would be in improved energy resolution and hence better Compton scattering rejection.

Cadmium telluride and mercuric iodide detectors will probably see significant applications in areas where small room temperature detectors of x rays can be used, but material fabrication problems will restrict the size of these detectors for some time. This will rule out their use in imaging applications.

Semiconductor detectors can provide positron cameras of extended area and with a high density of imaging elements resulting in spatial resolutions of 3 to 4 mm. An added advantage of these cameras, specially with hyperpure germanium, is the almost total rejection of scattered radiation and very fast timing resolution.

3.5. Scintillation detectors

3.5.1. General principles

Scintillation detectors rely on the fact that some substances will absorb energy, such as that carried by a photon, and will reemit it as light. This light is picked up by a photomultiplier tube that converts the light into an electrical signal. Scintillation detectors are divided into two classes: organic scintillators, such as anthracene
and certain plastics, and inorganic scintillators, such as certain alkali halides. Each have properties that make them more suitable for certain applications. Organic scintillators are typically very fast, but yield little light and have a low effective Z, making them attractive for timing experiments in which energy resolution or detection efficiency is not of prime concern. Inorganic scintillators tend to have high effective Z, good light output but are relatively slow, making them suitable for applications in which energy resolution and detection efficiency are important, such as in positron tomography.

The scintillation mechanism in inorganic materials depends on the energy states determined by the crystal lattice in the material. Electrons in the crystal lattice have available the valence band that represent those electrons that are essentially bound at lattice sites and the conduction band that represent electrons that have sufficient energy to be free to migrate throughout the crystal. There exists an intermediate band of energies, the forbidden band, in which electrons can never be found in pure crystals. Absorption of energy can result in the elevation of an electron from its normal position in the valence band across the gap into the conduction band, leaving a hole in the normally filled valence band. In the pure crystal, the return of the electron to the valence
band with the emission of a photon is an inefficient process. Furthermore, typical gap widths are such that the resulting photon would be of too high an energy to lie in the visible range. In order to enhance the probability of visible photon emission during the deactivation process, small amounts of an impurity are commonly added to inorganic scintillators. These impurities create special sites in the lattice at which the normal energy band structure is modified from that of the pure crystal. As a result, there will be energy states created within the forbidden gap through which the electron can de-excite back to the valence band. Because the energy is less than that of the full forbidden gap, this transition can now give rise to a visible photon and therefore serve as the basis of the scintillation process. In sodium iodide, for example, small amounts of thallium are added as impurity. In certain other inorganic scintillators, no addition of impurity is necessary to produce luminescence. In bismuth germanate, for example, the observed luminescence is entirely due to the properties of the Bi$^{3+}$ ion in the particular environment present in this crystal.

As the gamma ray passes through the detector, it may undergo an interaction that will transfer all or part of the photon energy to an electron in the absorbing material. These electrons have a maximum energy equal to
the energy of the incident gamma photon and will slow down and lose their energy in interactions with other electrons in the lattice. These interactions will result in the formation of a large number of electron-hole pairs created by the elevation of electrons from the valence band to the conduction band. The positive holes will quickly drift to the location of an activator site and ionize it, because the ionization energy of the impurity will be less than that of a typical lattice site. Meanwhile, the electron is free to migrate through the crystal and will do so until it encounters such an ionized activator. At this point, the electron can drop into the impurity site, creating a neutral impurity configuration which has its own set of excited energy states. If the activator state is formed in an excited configuration with an allowed transition to the ground state, its de-excitation will occur very quickly and with high probability for the emission of a corresponding photon. If the activator is properly chosen, this transition can be in the visible energy range. Because the migration time for the electron is much shorter than the half life for such excited states, all the excited impurity configurations are formed essentially at once and will subsequently de-excite with the half life characteristics of the excited states. It is the decay time of these states therefore that determines the time characteristics of the emitted scintillation.
light. There are processes that compete with the one just described. For example, the electron upon arriving at the impurity site can create an excited configuration which transition to the ground state is forbidden. Such states require an additional increment of energy to raise them to a higher lying state from which de-excitation to the ground state is possible. One source of this extra energy is thermal excitation and the resulting slow component of light is called phosphorescence. It can often be a significant source of background light, or after glow, in scintillators. A third possibility exists when an electron is captured at an active site. Certain radiationless transitions are possible between some excited states formed by electron capture and the ground state, in which case no visible photon results. Such processes are called quenching and represent loss mechanisms in the conversion of the full photon energy to scintillation light.

One important consequence of luminescence through activator sites is the fact that the crystal can be transparent to the scintillation light. In the pure crystal, roughly the same energy would be required to excite an electron-hole pair as that liberated when that pair recombines. As a result, the emission and absorption spectra will overlap and there will be substantial self absorption. However, the emission from an activated crystal occurs at
an activator site where the energy transition is less than that represented by the creation of the electron-hole pair. As a result, the emission spectrum is shifted to longer wavelengths and will not be influenced by the optical absorption band of the crystal.

In any scintillation detector, the collection of the largest possible fraction of the light emitted as a result of the interaction of the photon with the scintillator material is desirable. Two effects arise in practice which lead to less than perfect light collection: optical self absorption with the scintillator and losses at the scintillator surfaces. Because the scintillator light is emitted in all directions, only a limited fraction can travel directly to the surface at which the photomultiplier tube is mounted. The remainder, if it is to be collected, must be reflected one or more times at the other scintillator's surfaces. In order to recapture the light that does not escape from the surface at which the photomultiplier is mounted, the scintillator is normally surrounded by a reflector at all surfaces except that viewed by the photomultiplier tube. Reflectors can be either specular or diffuse. A polished metallic surface will act as a specular reflector for which the angle of reflection equals the angle of incidence. Better results are usually obtained, however, with a diffuse reflector, such as magnesium
oxide or aluminium oxide. In this case, the angle of reflection is approximately independent of the angle of incidence. Although total internal reflection is desirable at reflecting surfaces, it must be minimized at the surface from which the scintillator is viewed by the photomultiplier tube to prevent internal trapping of the light. Ideally, the scintillator should be optically coupled to the photocathode of the photomultiplier tube through a transparent medium of the same index of refraction as the scintillator. Then, no internal reflection would occur and all the light incident on the surface would travel to the photocathode.

The photomultiplier converts light that typically consists of no more than a few thousand photons into a usable current pulse without adding a large amount of random noise to the signal. The two major elements of a photomultiplier tube are the photocathode and the dynode chain. The photocathode serves to convert as many of the incident light photons as possible into low energy electrons. If the light consists of a pulse from a scintillation crystal, the photoelectrons produced will also be a pulse of similar time duration. Because only a few hundred photoelectrons may be involved in a typical pulse, this charge is too small at this point to serve as a detectable electrical signal. The dynode chain provides an efficient amplifier to greatly increase the number of electrons. After
amplification through the multiplier structure, a typical scintillation pulse will give rise to $10^7$ to $10^{10}$ electrons, easily sufficient to serve as the charge signal for the original scintillation event. This charge is conventionally collected at the anode, or output stage, of the multiplier structure and can be transformed into a voltage pulse in a preamplifier. Most photomultiplier tubes perform this charge amplification in a very linear manner, producing an output pulse that remains proportional to the number of original photoelectrons, and hence to the energy deposited in the crystal, over a wide range of amplitudes. Much of the timing information of the original light pulse is also retained.

The photoemission at the photocathode occurs in three sequential stages. The absorption of the incident light photon and the transfer of energy to an electron within the photoluminescent material; the migration of that electron to the surface; the escape of the electron from the surface of the photocathode. The energy that can be transferred from the photon to an electron is given by the quantum energy of the photon $h\nu$. Some of that energy will be lost through electron-electron collisions in the migration process. Finally, there must be sufficient energy left for the electron to overcome the inherent potential barrier which always exist at any interface between material and
Vacuum: the work function of the material. Normal conduction electrons within the photocathode material will always have some thermal kinetic energy and can occasionally have an energy that exceeds the work function. If that electron is close enough to the surface, it may escape and give rise to a spontaneous, thermally induced, signal.

The multiplier portion of a photomultiplier tube is based on the phenomenon of secondary electron emission. Electrons from the photocathode are accelerated and caused to strike the surface of a dynode. If the dynode material is properly chosen, the energy deposited by the incident electron can result in the re-emission of more than one electron from the same surface. In this case, electrons within the dynode material are excited by the passage of energetic electrons originally incident on the surface rather than by an optical photon. The emission of secondary electrons is a statistical process, and therefore the number of electrons re-emitted at a given dynode will fluctuate from event to event about a mean value. This statistical spread will result in broadening of the peak.

3.5.2. Sodium iodide

Thallium activated sodium iodide has been, and still remains, the scintillator of choice in many applications, such as Nuclear Medicine. It can be grown into large ingots that can be cut into a variety of sizes and shapes. It is
hygroscopic and will deteriorate due to water absorption if exposed to the atmosphere for any length of time. Crystals must therefore be canned in an air tight container.

Sodium iodide thallium activated has an excellent light yield, which is the highest of any known scintillator material. The dominant decay time of the scintillation pulse is 230 nsec, uncomfortably long for some fast timing or some high count rate applications. Its specific gravity is 3.67 and its index of refraction at the wavelength of maximum emission (410 nm) is 1.85.

3.5.3. Caesium fluoride

The main advantage of caesium fluoride is its unusually fast decay time (5 nsec) making it the candidate of choice for fast timing or high count rate applications. However, caesium fluoride is very hygroscopic and special care must be taken in the encapsulation process. It also exhibits a low light output and its specific gravity is 4.11, slightly better than that of sodium iodide. Its index of refraction at the wavelength of maximum emission (390 nm) is 1.48.

3.5.4. Barium fluoride

The decay spectrum of barium fluoride is characterized by two components; one with a decay time of 0.6 nsec, the other with a decay time of 620 nsec. The very fast component of the decay spectrum makes it an attractive
candidate for fast timing applications.

Barium fluoride is non hygroscopic. Its specific gravity is 4.88. Its light yield is 5% that of sodium iodide. Its peak emission is at 225 nm (20%) and 310 nm (80%).

3.5.5. Bismuth germanate

The main attraction of bismuth germanate is the high Z value of the bismuth atom (83) that leads to a high photoelectric cross section for gamma rays. It is non hygroscopic and its mechanical and chemical properties make its use and handling easy. Its light output is poor. Its specific gravity is 7.13 and its index of refraction at the wavelength of maximum emission (480 nm) is 2.15. This causes a substantial amount of light to be trapped in the crystal lattice. Its decay constant is 300 nsec, comparable to that of sodium iodide thallium activated.

Bismuth germanate is a relatively new scintillator and improvements in the quality of the crystal have resulted in a steady improvement in resolving time (from 20 nsec to 3 nsec) and light output (energy resolution at 662 keV from 15% to 9%).

3.6. Conclusions

From the preceding exposition, it should be clear that there is no one ideal detector for positron tomography. Each type of detector has properties that make it more suitable for one type of application and less suitable for others. If a spatial resolution close to the
theoretical limit imposed by the finite positron range is required, multiwire proportional counters are the most appropriate solution. However, the price paid for this spatial resolution is the time it takes to make an examination because of the low counting efficiency of multiwire proportional counters. For example, Jeavons et al (1980) have produced remarkable pictures of the skeletal system of a rabbit, but the time required to gather sufficient information was of the order of half an hour. Furthermore, because of the complexity in fabricating multiwire proportional counters, the cameras are small in area and can only be used to image objects of the size of a rabbit, such as the thyroid gland in a human.

If rejection of scattered radiation is of paramount importance, then a camera based around solid state detectors is the solution because of the energy resolution of such detectors. However, this is done at the expense of detection efficiency because large solid state crystals are difficult to grow. Even though a camera could be constructed of strips of solid state detectors in order to cover a large area, the detectors cannot be obtained with sufficient thickness, causing a loss in stopping power. The University of California, San Francisco, have been developing such a camera for some time. To date, no experimental results, except for some phantom studies, are available.
In most medical applications, the number of photons available to form an image is limited. The half lives of the positron emitting isotopes of biological interest are short, ranging from 2 min for $^{15}$O to 110 min for $^{18}$F. If any chemical synthesis is required to label specific molecules, such as fluoro deoxyglucose or fluoro dihydroxyphenylalanine, the time taken for the synthesis and the radiochemical yield will limit the dose that can be obtained. Furthermore, restrictions are imposed as to the amount of radioactivity that can be safely given to a human being. Thus detection efficiency has to be maximized in order to obtain a statistically significant image. However, other factors, such as spatial resolution and scatter rejection, have to be taken into consideration as they, too, affect the quality and the information content of the final image.

The detectors that seem to be most appropriate for clinical positron tomography are therefore the inorganic scintillators. Sodium iodide was the first detector used in clinical applications, despite its limitations with respect to stopping power and time resolution. As new scintillators were developed by the chemical companies, their suitability was assessed. Of the newer scintillators, bismuth germanate (introduced in 1977) presents many advantages over the others. It is the scintillator with the highest stopping power available thus allowing for improved detection
efficiency. Its time resolution is poorer than that of other inorganic scintillators, but improvements in its performance have been made possible with continued research into its growth. Furthermore, its mechanical and chemical properties are ideally suited for optimizing the geometry of individual detectors in a multicrystal camera.

Caesium fluoride was introduced in 1980 and its main advantage is its very short decay time. It was hoped that this would allow for time of flight tomography where the difference in times of arrival of the two 511 keV photons can be translated into the location along the path joining the two detectors at which the positron annihilated. However, the best systems so far constructed only allow the determination of the location of the annihilation event to within 60 mm, clearly too crude for imaging an organ such as the human brain. Caesium fluoride has found applications in very high count rate studies in which random coincidence rejection is crucial. It has also found applications in time of flight assisted tomography where the added spatial information provided is used to reduce the uncertainty in the reconstruction of the image.
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1.</td>
<td>Introduction</td>
<td>94</td>
</tr>
<tr>
<td>4.2.</td>
<td>Reconstruction methods</td>
<td>95</td>
</tr>
<tr>
<td>4.2.1.</td>
<td>Definitions</td>
<td>95</td>
</tr>
<tr>
<td>4.2.2.</td>
<td>Simple backprojection</td>
<td>100</td>
</tr>
<tr>
<td>4.2.3.</td>
<td>Iterative reconstructions</td>
<td>102</td>
</tr>
<tr>
<td>4.2.3.1.</td>
<td>General principles</td>
<td>102</td>
</tr>
<tr>
<td>4.2.3.2.</td>
<td>ILST</td>
<td>104</td>
</tr>
<tr>
<td>4.2.3.3.</td>
<td>ART</td>
<td>104</td>
</tr>
<tr>
<td>4.2.3.4.</td>
<td>SIRT</td>
<td>107</td>
</tr>
<tr>
<td>4.2.4.</td>
<td>Analytical reconstructions</td>
<td>107</td>
</tr>
<tr>
<td>4.2.5.</td>
<td>Event by event reconstructions</td>
<td>116</td>
</tr>
<tr>
<td>4.2.5.1.</td>
<td>General principles</td>
<td>116</td>
</tr>
<tr>
<td>4.2.5.2.</td>
<td>Event by event filtered backprojection</td>
<td>117</td>
</tr>
<tr>
<td>4.2.5.3.</td>
<td>ARTIST</td>
<td>117</td>
</tr>
<tr>
<td>4.3.</td>
<td>Image reconstruction in positron tomography</td>
<td>117</td>
</tr>
<tr>
<td>4.3.1.</td>
<td>Introduction</td>
<td>117</td>
</tr>
<tr>
<td>4.3.2.</td>
<td>Parallel geometry</td>
<td>118</td>
</tr>
<tr>
<td>4.3.3.</td>
<td>Divergent geometry</td>
<td>121</td>
</tr>
<tr>
<td>4.4.</td>
<td>Corrections in positron tomography</td>
<td>125</td>
</tr>
<tr>
<td>4.4.1.</td>
<td>Introduction</td>
<td>125</td>
</tr>
<tr>
<td>4.4.2.</td>
<td>Efficiency corrections</td>
<td>127</td>
</tr>
<tr>
<td>4.4.3.</td>
<td>Attenuation corrections</td>
<td>128</td>
</tr>
<tr>
<td>4.4.4.</td>
<td>Randoms corrections</td>
<td>131</td>
</tr>
<tr>
<td>4.4.5.</td>
<td>Scattered corrections</td>
<td>132</td>
</tr>
<tr>
<td>4.5.</td>
<td>Discussion</td>
<td>134</td>
</tr>
<tr>
<td>4.5.1.</td>
<td>Linear sampling</td>
<td>134</td>
</tr>
<tr>
<td>4.5.2.</td>
<td>Angular sampling</td>
<td>135</td>
</tr>
<tr>
<td>4.5.3.</td>
<td>Wobble motion</td>
<td>136</td>
</tr>
<tr>
<td>4.5.4.</td>
<td>Dichotomic motion</td>
<td>138</td>
</tr>
<tr>
<td>4.5.5.</td>
<td>Continuous rotation</td>
<td>140</td>
</tr>
<tr>
<td>4.5.6.</td>
<td>Conclusion</td>
<td>140</td>
</tr>
</tbody>
</table>
4.1. Introduction

Image reconstruction from projections is a technique whereby measurements of a parameter are made around an object in an effort to find out the regional distribution of that parameter in a cross section of the object. For example, in X-ray computerized tomography, the function of interest is the linear attenuation coefficient. Measurements of the transmission of an X-ray beam are obtained; from these, an image is formed that represents the distribution of attenuation coefficients. Since every tissue is characterized by a different attenuation coefficient, the image also represents tissue distribution in the slice. In emission tomography, the function of interest is the activity of the radionuclide. Measurements of the number of photons being emitted, after the injection or the inhalation of a radioactive substance, are obtained and from these an image is reconstructed that represents the distribution of radioactivity. Since the radioactivity was associated with a particular molecule, the image also represents the distribution of this molecule in the slice. In nuclear magnetic resonance tomography, the function of interest is the distribution of the nuclear spin density. Measurements of spin density are obtained by aligning the nuclear spins in a large magnetic field, then exciting them by applying a small electric field and by watching them lose that excitation. Only nuclei with an odd number of
protons will exhibit this behaviour and by choosing the magnetic and electric fields appropriately, different nuclei can be selectively excited. From these measurements, an image is obtained that represents the distribution of spin densities. Since these represent a particular nucleus, or a nucleus in a particular chemical environment, the image also represents the distribution of these nuclei.

There are many approaches to the problem of recovering the unknown distribution of the parameter being studied. In general, these can be classified into iterative methods and analytical methods. In addition, corrections have to be applied to the data collected in order to account for problems associated with the gathering of data. These corrections will be discussed as they apply to positron tomography after an exposition of the reconstruction methods.

4.2. Reconstruction methods

4.2.1. Definitions

A set of data consists, in the simplest case, of a collection of \( n \) values, each representing the sum of the values of the parameter being examined that fall along a particular path through the transverse section of the object being examined (fig. 4.1). These paths are all coplanar and traverse the region at \( n_s \) regularly spaced
Figure 4.1: Transverse section through the object at angle $\theta$. The separation between each ray path is $n_s$.

Figure 4.2: System of coordinates used to describe the density function $f(x,y)$. 
intervals and at $n \theta$ regularly spaced angles, such that

$$n = n_s \cdot n \theta$$

More formally, an $(x,y)$ coordinate system is used to describe points in the plane under examination (fig. 4.2). The contribution of each point toward the detected signal is denoted by the density function $f(x,y)$. This density function may be the value of the linear attenuation coefficient, the concentration of radioactivity, etc, at location $(x,y)$.

A linear projection, $p$, consists of a set of ordered numbers, each representing the sum of the values of the parameter along a ray path (line integral) normal to the direction of projection. The orientation of the projection is expressed in terms of a rotated system of cartesian coordinates $(r,s)$; the new system is rotated by an angle $\theta$ with respect to the fixed frame of reference $(x,y)$. In the simplest case, line integrals are obtained along lines that are equidistant from each other and that are parallel to the rotated ordinate (fig. 4.3). The equation of any such line is given by:

$$R = \frac{r}{\cos(\phi-\theta)}$$

or

$$r = R \cos(\phi-\theta)$$

$$= R \cos \phi \cos \theta + R \sin \phi \sin \theta$$
Noting that
\[ x = R \cos \phi \quad ; \quad y = R \sin \phi \]
the coordinate pairs \((x,y)\) that specify which \(f(x,y)\) will contribute to the line integral specified by \((r,\theta)\) are given by:
\[ r = x \cos \theta + y \sin \theta \]
Each line integral
\[ p(r_i, \theta_i) = \int f(x,y) \, ds \]
will generate one value of the function \(p(r,\theta)\) that represents the projection data (fig. 4.4).

Ideally, \(f(x,y)\) is a continuous two-dimensional function and an infinite number of projections are required for reconstruction. In practice, \(f(x,y)\) is calculated at a finite number of points from a finite number of projections. Reconstruction is carried out using a square region of dimensions \(D \times D\) subdivided into small square cells, pixels, each of dimension \(d \times d\).

The methods that allow for the calculation of \(f(x,y)\), the unknown distribution, can be classified into four groups: simple backprojection, iterative reconstructions, analytic reconstructions and event by event reconstructions.
Simple backprojection: This approach was used in early experiments. It is the simplest in concept and the easiest to understand, but it produces reconstructions with
Figure 4.3: Relation between the fixed coordinate system \((x,y)\) and the rotated coordinate system \((r,s)\).

\(\theta\) is the angle of rotation with respect to the \(x\) axis.

Figure 4.4: Polar representation of the projection data.

In the projection at angle \(\theta_j\), the point \(p(r_i,\theta_j)\) represents the value of the line integral at \(r_i\).
substantial artifacts.

Iterative reconstructions: Iterative methods may be thought of as brute force ways of solving the image equation.

Analytic reconstructions: Analytic methods are based on exact mathematical solutions to the image equation and hence are faster.

Event by event reconstructions: These methods are limited to radioisotopic techniques where each detected event can be processed individually. They may have an application in monitoring the image while it is being accumulated.

4.2.2. Simple backprojection

Reconstruction is performed by backprojecting each profile across the plane, i.e., the magnitude of each ray sum is applied to all points that make up the ray. The backprojected density at each point is simply the sum of all ray sums passing through the point.

The mathematical equation describing simple backprojection is:

\[ f(x,y) = \int p(r,\theta) \, d\theta \]

\[ = \int p(x \cos \theta_i + y \sin \theta_i, \theta_i) \, d\theta \]

The integration extends over all projection angles. The argument

\[ (x \cos \theta_i + y \sin \theta_i) \]
Figure 4.5: Cylindrical source (2 cm in diameter) containing a solution of $^{18}\text{F}$ reconstructed using the simple backprojection method (a) and the filtered backprojection method (b). Note the extended halo around the object in (a), the star artifact.
selects only those rays which pass through the point \((x,y)\).

Simple backprojection does not produce a reliable reconstruction because each ray sum is applied to all points along the ray. This means that points outside the original object receive some contribution from the ray sum. This defect shows up most strikingly as discrete areas of high density, or hot spots, producing a star like artifact (fig. 4.5). In addition, points within the object receive contributions from neighbouring points so that subtle differences if \(f(x,y)\) cannot be resolved. This defect shows as a blurring of the image.

4.2.3. Iterative reconstructions

4.2.3.1. General principles

The basic strategy of iterative methods is to apply corrections to arbitrary initial cell densities in an attempt to match the measured ray projections. Since former matchings are lost as new corrections are made, the procedure is repeated until the calculated projections agree with the measured ones to within the desired accuracy.

To start out, the object to be reconstructed is approximated by an array of cells with density values \(f_1(x,y)\). The projections are broken up into strips whose widths is generally chosen as equal to that of the cell. The ray sums are made up of contributions from each cell intersected by the ray, i.e., for the \(j^{th}\) ray
where the w's are weighting factors that represent each cell's contribution. This equation represents a set of matrix equations for the density values \( f_i \). Since their direct solution is generally not practical because of the large number of cells, the aim is to adjust the density values \( f_i \) iteratively until the calculated projections agree with the measured projections. First, a starting set of values is chosen for each cell density. Then projections are calculated from these starting values. If, for example, a calculated ray sum is too small when compared with the measured value, each cell that contributes to this ray is increased in density by an appropriate amount, according to certain formulae. When this has been done for all cells and all rays, the first iteration is complete. This procedure is then repeated until the desired accuracy in the difference between measured and calculated ray projections is achieved.

The first iteration is equivalent to simple backprojection if the starting distribution is zero everywhere, since the calculated projections in this case are zero. In succeeding iterations, a correction factor is backprojected. A single iteration fails to be exact for the same reasons that simple backprojection is not exact.
Corrections are not applied selectively to cells that need them, because this is not known. Instead, a correction is applied to all cells along the ray.

4.2.3.2. Simultaneous corrections: ILTS

(Iterative Least Squares Technique)

This is the simplest approach. All projections are calculated at the beginning of the iteration, and all corrections are made simultaneously (fig. 4.6). This, however, leads to an overcorrection, as each cell is re-corrected for every ray passing through it, with the result that the iterations oscillate about the correct solution. A damping factor can be applied to all corrections so as to produce the best least squares fit after each iteration. The exact choice of damping factor is, however, not critical for the method to work.

4.2.3.3. Ray by ray corrections: ART

(Algebraic Reconstruction Technique)

In this approach, at the start of each iteration, one ray sum is calculated and corrections are applied to all points that contribute to the ray. The procedure is then repeated for a second ray, a third ray, etc., always embodying previous corrections in each new calculation, until all rays in all projections have been treated, thus completing one iteration (fig. 4.7).
Figure 4.6: Iterative least squares technique: note that all corrections are made simultaneously. After the ninth iteration, the original distribution is still not obtained.
Figure 4.7: Algebraic reconstruction technique: note that the corrections are introduced one projection at a time. The original distribution is recovered after only one iteration for this simple example.
The method works provided the corrections are made one projection at a time, with large angles between consecutive projections. The correction is divided among the cells in proportion to their weighting factor. This ensures that successive corrections are independent of each other so that errors do not accumulate.

4.2.3.4. **Point by point corrections: SIRT**

(Simultaneous Iterative Reconstruction Technique)

In this variation, each iteration begins with a particular point, which is corrected for all rays that pass through it. Other points are then treated the same way, except that corrections made during the iteration are embodied in succeeding calculations. Each cell receives a correction in proportion to its current density. This is done by multiplying the current density by the ratio of the measured to calculated ray sum.

4.2.4. **Analytic reconstructions**

Analytic reconstructions are based on direct solutions of the image equation. They are subdivided into two groups that can be shown to be equivalent: filtered backprojection and 2-D Fourier reconstructions.

The problem posed is to obtain a good estimate, or reconstruction of \( f(x,y) \) from a knowledge of \( p(r,\theta) \).
The one dimensional Fourier transform of the projection at an angle $\theta_i$ is:

$$P(\omega, \theta_i) = \int p(r, \theta_i) \ e^{-i\omega r} \ dr$$

$$= \int \int f(x, y) \ e^{-i\omega r} \ ds \ dr$$

The two dimensional Fourier transform of the original distribution is:

$$F(\omega_x, \omega_y) = \int \int f(x, y) \ e^{-i(x \omega_x + y \omega_y)} \ dx \ dy$$

The point specified in polar coordinates by $(\omega, \theta_i)$ can be specified in an orthogonal coordinate system (fig. 4.8) by

$$\omega_x = \omega \cos \theta_i$$

$$\omega_y = \omega \sin \theta_i$$

so that:

$$F(\omega_x, \omega_y) = \int \int f(x, y) \ e^{-i\omega(x \cos \theta_i + y \sin \theta_i)} \ dx \ dy$$

Because both the $(x,y)$ and the $(r,s)$ coordinate systems are Cartesian

$$dx \ dy = ds \ dr$$

and since

$$r = x \cos \theta_i + y \sin \theta_i$$

then

$$F(\omega_x, \omega_y) = \int \int f(x, y) \ e^{-i\omega r} \ ds \ dr$$
Figure 4.8: Relation between the orthogonal coordinate system \((\omega_x, \omega_y)\) and the polar coordinate system \((\omega, \theta)\).

Figure 4.9: The function \(f(x,y)\) in the spatial domain can be represented by a 2-D Fourier transformation in the frequency domain. The vertical arrows in (a) represent the values of the line integrals in a projection at a given \(\theta\). The vertical arrows in (b) represent the values of the 1-D Fourier transform of the projection in (a). They also represent samples of the 2-D Fourier transform along the line making the same angle \(\theta\) with the \(\omega_x\) axis.
so that
\[ P(\omega, \theta_1) = F(\omega_x, \omega_y) \]

In words, the one dimensional Fourier transform of the projection at an angle corresponds to a slice taken at the same angle through the two dimensional Fourier transform of the original function (fig. 4.9).

Taking the two dimensional reverse Fourier transform of \( F(\omega_x, \omega_y) \) should therefore give the function \( f(x,y) \):
\[
f(x,y) = \frac{1}{4\pi^2} \iint F(\omega_x, \omega_y) e^{i\omega x} d\omega_x d\omega_y
\]
\[
= \frac{1}{4\pi^2} \iint P(\omega, \theta) e^{i\omega x} d\omega_x d\omega_y
\]

The function \( P(\omega, \theta) \) is in polar coordinates, whereas the function \( F(\omega_x, \omega_y) \) is in Cartesian coordinates. The relation to change an integration in an orthogonal coordinates system to one in polar coordinates is:
\[
d\omega_x d\omega_y = \omega d\omega d\theta
\]
and
\[
f(x,y) = \frac{1}{4\pi^2} \int_{0}^{2\pi} \int_{-\infty}^{\infty} P(\omega, \theta) \omega e^{i\omega x} d\omega d\theta
\]
and since the point specified by \( (\omega, \theta) \) is the same as that specified by \( (-\omega, \theta+\pi) \).
If the function \( p(r,0) \) is defined as:

\[
f(x,y) = \frac{1}{4\pi^2} \int_0^\pi \int_{-\infty}^{\infty} P(\omega,\theta) \left| \omega \right| e^{i\omega r} \, d\omega \, d\theta
\]

This means that \( f(x,y) \) can be obtained by backprojecting the function \( \hat{p}(r,\theta) \).

If the term \( |\omega| \) is neglected in the function giving \( \hat{p}(r,\theta) \), \( \hat{p}(r,\theta) \) simply becomes the projection data \( p(r,\theta) \).

If a function \( \phi(r) \) is defined such that its Fourier transform is \( |\omega| \)

\[
|\omega| = \int_{-\infty}^{\infty} \phi(r) e^{-i\omega r} \, dr
\]

then, the convolution of the projection data, \( p(r,\theta) \), with this function gives:

\[
p(r,\theta) * \phi(r) = \frac{1}{2\pi} \int_0^\pi \int_{-\infty}^{\infty} P(\omega,\theta) \left| \omega \right| e^{i\omega r} \, d\omega \, d\theta
\]

\[
= \hat{p}(r,\theta)
\]
In words, the original projection data has to be convolved with the function $\phi(r)$, and then backprojected.

Because $|\omega|$ is a real and even function, the function $\phi(r)$ whose Fourier transform is $|\omega|$ will have to be real and even. The filter function in the spatial domain is given by:

$$\phi(r) = \int_{-\infty}^{\infty} |\omega| e^{i\omega r} d\omega$$

This integral diverges, unless the limits of integration are restricted. Restricting the limits of integration is equivalent to multiplying the ramp function $R(\omega) = |\omega|$ with a rectangular window (fig. 4.10)

$$\pi(\omega) = \begin{cases} 
1 & -\Omega < \Omega \\
0 & \text{elsewhere}
\end{cases}$$

so that

$$\phi(r) = \int_{-\Omega}^{\Omega} |\omega| e^{i\omega r} d\omega$$

Since the object is sampled at intervals equal to the distance between adjacent points in a projection, $\Delta r$ cm, the highest meaningful spatial frequency is:

$$\Omega = \frac{1}{2\Delta r} \text{ cm}^{-1}$$

As long as it does not exceed $2\Omega$, a cutoff frequency other than $\Omega$ may be used. A cutoff frequency higher than $\Omega$ will produce higher resolution but it will also be more sensitive
Figure 4.10: The windows illustrated in (a) are multiplied by the ramp function to give the frequency domain filters shown in (b). The inverse Fourier transform of those give the spatial domain convolution functions shown in (c).
to statistical fluctuations in the measurements.

The sharp frequency cutoff at $\Omega$ introduced by the rectangular window will give rise to oscillations at sharp boundaries in the object. This is known as the Gibbs phenomenon and arises because of the finite limits imposed on $\phi(r)$ by its definition. The effect of the Gibbs phenomenon can be reduced by using other types of frequency windows in which the frequency cutoff is rolled off rather than cutoff sharply (fig. 4.10). These windows will result in smoother images, which are less susceptible to noise, but will exhibit poorer spatial resolution.

The rectangular window that was introduced by Bracewell and Riddle gives the best resolution when perfect data is reconstructed, but it amplifies noise for data with statistical fluctuations.

The window introduced by Shepp and Logan

\[
\begin{bmatrix}
|\text{sinc}(\omega)| & -\Omega < \omega < \Omega \\
0 & \text{elsewhere}
\end{bmatrix}
\]

is very similar to the rectangular window. It gives comparable resolution to that obtained with a rectangular window and it has the effect of reducing the noise amplification for data with statistical fluctuation.

The Hamming window

\[
\begin{bmatrix}
|a - [(1 - a) \cos(2\pi\omega)]| & -\Omega < \omega < \Omega \\
0 & \text{elsewhere}
\end{bmatrix}
\]

where $0 \leq a \leq 1$
is a general window. Setting $\alpha = 1$, the window becomes the rectangular window. Setting $\alpha = 0$, the window becomes a cosine window, a low pass window commonly used in signal processing. The judicious choice of $\alpha$ will depend on the noise level in the data, and on the nature of the object being reconstructed.

The convolution backprojection method is therefore the convolution of the projection data with a suitable filter function in the spatial domain prior to backprojection. The 2-D Fourier transform method involves taking the one dimensional Fourier transform of the projection data, multiplying it in the frequency domain with the ramp function, or a suitable bounded function similar to it, and then performing a two dimensional reverse Fourier transformation.

The method has been derived in its continuous form. In the real world, the projection data are obtained at discrete intervals imposed by the physical size of the detectors, and the collimators, and by the finite number of projections obtained around the object. The data are imperfect. They contain random fluctuations, they depend on the stability and relative efficiency of the detection system, and are affected by absorption and scattering in the object. The backprojection operation introduces an added problem, that of matching the discrete reconstruction matrix to the discrete projection data. A function giving the fractional
contribution that each ray sum makes to a pixel is therefore needed. This is done through the choice of an interpolating function that will allow uniform filling of the matrix. Because of its discrete implementation, the method cannot perfectly reconstruct the distribution of activity in the object. It can only provide a good estimate of it. The choice of filter function and of interpolating function remain as degrees of freedom and the final image will be very dependent on the choice made.

4.2.5. Event by event reconstructions

4.2.5.1. General principles

These methods take advantage of the fact that for radioisotopic procedures, the projection data are accumulated one event at a time. The advantage of these techniques would be that the reconstructed image is formed during the course of the data collection. Consequently, the image can be monitored as it is being accumulated so that a study could be terminated when a satisfactory image had been obtained, or adjustments could be made without delay if an unsatisfactory image was being formed on the display monitor. However, these methods cannot take into account any of the corrections that are required in radioisotope tomography such as efficiency correction or attenuation corrections.
4.2.5.2. **Event by event filtered backprojection**

In this scheme, instead of accumulating all of the projection data and then filtering and backprojecting, the filter function is backprojected for each event recorded. Therefore, the two mathematical operations of filtering and backprojecting are identical for each event so that a simple hardwired backprojector could be constructed and used.

4.2.5.3. **ARTIST**

(Algebraic Reconstruction Technique Intended for Storage Tube)

This method also treats each collected event individually. It uses previously accumulated and stored events in deciding to which point in the image to assign the latest event. A random number is used to locate a single point along the coincidence line so that the density along a ray is increased in proportion to its local value.

The advantage of such a method would be the speed and the simplicity of the image forming process.

4.3. **Image reconstruction in positron tomography**

4.3.1. **Introduction**

The first positron tomograph to be designed, the MGH PT I, consisted of two opposed rectangular arrays of detectors. These two arrays had to be rotated around the object in order to obtain the required angular sampling.
Obviously, this method made very poor use of the available photons and the spatial resolution was limited by the fixed distance between detector centres. The next generation of tomographs used detectors placed in a hexagonal array. The banks of detectors could be translated, thereby achieving any desired linear sampling. They could also be rotated around the object, thereby achieving the required angular sampling. Although making more efficient use of the available photons, the rotation and translation motions required to gather the data limited the possible applications because of the time it took to complete an examination. The next step was to arrange the detectors along the circumference of a circle thereby eliminating the need for the rotate/translate motion since all possible coincidence pairs through the object would be measured at once. However, a rotation by half a detector spacing was still believed to be needed for maximum sampling and hence maximum spatial resolution. With the opposed arrays of detectors and with the hexagonal array of detectors, the only reconstruction geometry that could be used was the parallel geometry. This is the geometry that was applied to the ring tomographs. However, another way of reordering the data, into divergent projections, makes more efficient use of the gathered information.

4.3.2. Parallel geometry

The spatial resolution of any ring positron tomographic device is primarily determined by the linear
sampling, provided adequate angular sampling has been obtained. The linear sampling is fixed by the distance between sample points in a projection. In order to improve the linear sampling, it has been proposed to rotate the ring about its axis by an amount equal to half a detector spacing. While this motion does not increase the angular sampling, it does halve the distance between adjacent points in a projection and hence doubles the linear sampling. If one considers a parallel projection to be the set of sample points, each representing the sum of radioactivity along a line (sampling ray) normal to the direction of projection, all sampling rays in this projection are parallel to each other and separated in the middle of the reconstruction field by a distance equal to the detector spacing (fig. 4.11). At the edge of the reconstruction field, this distance is less. If the device is rotated about its axis by an amount equal to half a detector spacing, a new set of parallel rays, normal to the same direction of projection, is obtained. Combining these two sets, all the rays are normal to the direction of projection and separated from each other by one half a detector spacing (fig. 4.11). This means that if an image is reconstructed using this parallel ray geometry, the ray spacing will be equal to the detector spacing in the first case and equal to half the detector spacing in the second case. Therefore, half rotation should
Figure 4.11: Parallel geometry: One projection consists of the line integrals along the paths (solid lines) joining the faces of opposed detectors. The spacing between these rays is one detector spacing. If the ring is rotated by half a detector spacing, a new set of paths is generated (dotted lines). Those, combined with the original set, give a projection in which the spacing between adjacent rays is half a detector spacing.
improve the spatial resolution of the system.

This parallel ray geometry is the most natural way of considering the data accumulated by positron tomographs. Historically, it was the only geometry that could be applied to the MGH positron camera that used two diametrically opposed arrays of sodium iodide detectors, or the positron tomographic devices that use a hexagonal, or octagonal, array of detectors.

4.3.3. Divergent geometry

If one examines the circular geometry of the newer ring devices, and considers all possible sampling rays, one realizes that every ray that connects pairs of detectors is tangent to one of a finite set of concentric circles (fig. 4.12). If one considers a divergent projection to be the set of sample points, each representing the sum of radioactivity along the line joining one detector to any opposed detector, all sampling rays in the projection originate from the same detector and each is tangent to one of these concentric circles. As one rotates the ring about its centre, no matter by how much or how little, no new circles are generated since the rays connecting the centre of the faces of the detectors remain tangent to the same circles (fig. 4.12). This means that although the angular sampling has increased, the linear sampling has remained the same and the spatial resolution cannot be
Figure 4.12: Divergent geometry: One projection consists of the line integrals along the paths joining the face of one detector to the faces of detectors opposed to it. All possible paths are represented. Note that they are tangent to one of a finite set of concentric circles.
A parallel projection (dark lines) is shown for comparison. Note that these rays miss every other circle.
Figure 4.12 (continued): If the ring is rotated by half a detector spacing, a new set of paths is generated. However, the rays in this new set are tangent to the same circles as before. The space is not sampled more finely. A parallel projection (dark lines) is shown for comparison. Note that, in this case, each ray is tangent to one of the circles.
improved by half rotation. It is clear from figure 4.12 that the separation between adjacent rays varies along the ray. Because projections are collected through 360°, a region near the edge of the reconstruction field that had been sampled coarsely in a projection will be sampled more finely in the diametrically opposed projection. The net effect is an average sampling distance of half a detector spacing. This is exactly the spacing that is achieved by half rotation in parallel ray geometry (fig. 4.12). Half rotation is therefore an unnecessary, time consuming step, and the full resolution of the instrument can be obtained at once by a more efficient use of the information at hand, i.e., using divergent ray algorithms in the reconstructions.

Other algorithms have been proposed to recover the maximum spatial resolution of the tomograph without the need for half rotation. For example, the group at the Donner laboratory merges even and odd parallel projections thereby halving the number of projections, but doubling the number of data points per projection. Within each projection, alternate data points differ by an angle \( \pi / \text{number of detectors} \) from its nearest neighbour. A parallel ray reconstruction algorithm is then used for reconstruction. Even though spatial resolution is improved by this strategy, the more exact mathematical solution is to use divergent geometry with the initial set of data points.
4.4. Corrections in positron tomography

4.4.1. Introduction

In positron tomography, many corrections must be applied to the gathered data in order to obtain an accurate image of the object being studied. Each detector has an intrinsic detection efficiency and since the detection efficiency along any line integral depends on the product of the detection efficiencies of the detector pair involved in its measurement, a correction for the variations in detection efficiency has to be applied. Because the 511 keV photons resulting from the annihilation of the positron interact with the object, a correction for the attenuation of the rays through the object must also be applied.

In interactions with matter, the photons are scattered away from their original paths. Three conditions can be identified: -i- The two photons detected in coincidence are correlated and are a result of the annihilation of a single positron located in the field of view. However, one or both photons were Compton scattered before reaching the detectors (fig. 4.13). In this case, even though the coincidence is a true event, the annihilation may have taken place some distance away from the line joining the two detectors that have recorded the event.

-ii- The two photons detected in coincidence are correlated and resulted from the annihilation of a single positron.
Figure 4.13: Scattered true coincidences: In (a), the two photons, resulting from the annihilation of a positron in the field of view, are detected, but one or both photons are Compton scattered before reaching the detectors. In (b), the positron undergoes annihilation outside the field of view, but one or both photons are scattered back into the detectors, therefore registering a true event. Random coincidences: In (c), the two photons detected within the resolving time of the system resulted from the annihilation of two separate positrons, either within the field of view, or outside it.
However, this positron was located outside the field of view and one or both photons were Compton scattered back into the detectors, thereby contributing an event that does not belong in the image. In these two cases, the event is called a scattered event and corrections can be applied to try to minimize their effect in the reconstructed image. The two photons detected in coincidence are not correlated and result from the annihilation of two, independent, positrons either located in the field of view, or outside it (fig. 4.13). These events are called random events and corrections can be applied to minimize their effect.

4.4.2. Efficiency corrections

Even though all the detectors in a tomograph are matched for detection efficiency, each detector will have a detection efficiency that differs from all other detectors. Because coincidence detection efficiency is proportional to the product of the efficiencies of the two detectors involved in the measurement, each line integral measurement will differ from all others even when viewing a uniformly distributed source. Therefore a correction has to be applied to each line integral measurement to correct for these deviations.

If a point source filled with a solution of positron emitting radionuclides is placed at the geometric centre
of the ring, the individual efficiencies of all detectors can be determined. However, this information is not sufficient to correct the data collected from an extended source because many more detector pair combinations are involved in imaging it. To correct for the efficiency of all detector pairs used to image such an extended object, a cylindrical object, uniformly filled with a solution of positron emitting radionuclides, is placed so that its centre corresponds with the centre of the ring, and the diameter of the cylinder is chosen so that it extends beyond the extension of the largest object to be examined by the tomograph. Each individual ray path through the object can be calculated geometrically and this information can be used to calculate and correct for differences in efficiencies. Thus, each line integral is corrected for the relative efficiency of the detector pair involved in its measurement using the information gathered in imaging that cylindrical phantom.

4.4.3. Attenuation corrections

Because gamma rays travelling through an object will interact with the constituents of the object, each line integral has to be weighted by a factor proportional to its path length through the object, and representing the absorption of the gamma rays in the object. Since attenuation lengths for each detector pair can be
uniquely defined, attenuation correction is relatively easy to accomplish. The most accurate way for making this correction is by using complementary transmission data. Another method is to use a water bath of a size approximately equal to that of the object and to assume that the average attenuation coefficient of the object is equivalent to that of water. Alternatively, a ring of activity can be used to determine the boundaries of the object and correction is performed by assuming an average \( \mu \) value throughout the object.

Both the boundary as well as the density along each line integral can be measured using transmission data. They require two additional measurements. The first measurement is made by recording the projection data with a ring of activity large enough to enclose the entire object. The object is then introduced within the tomograph and another set of data is collected. Without disturbing the position of the object, the ring is then removed, activity injected into the object, and the emission data are collected.

For a detector pair \((ij)\), the coincidence counts collected for a given time for the ring with the object is \((I_{ro})_{ij}\). This can be written as:

\[
(I_{ro})_{ij} = (N_{ro})_{ij} \cdot n_i \cdot n_j \cdot \Omega_{ij} \cdot e^{-\mu ij} \cdot L_{ij}
\]
where \((N_{ro})_{ij}\) is the fractional ring activity seen by the detector pair, \(n_i\) and \(n_j\) are the efficiencies of the two detectors, \(\Omega_{ij}\) is the solid angle subtended by the detector pair combination, \(\mu_{ij}\) is the average value of the linear attenuation coefficient for the length \(L_{ij}\) between the detectors.

Without the object, if the counts collected for the same interval of time are given by \((I_r)_{ij}\), and the half life of the activity is long compared to the collection time, then:

\[
(I_r)_{ij} = (N_{ro})_{ij} \cdot n_i \cdot n_j \cdot \Omega_{ij}
\]

and:

\[
e^{-\mu_{ij} \cdot L_{ij}} = \frac{(I_{ro})_{ij}}{(I_r)_{ij}}
\]

With the ring removed, and radioactivity introduced in the object, the fractional activity within the object, \(N_{ij}\), seen by the detector pair can then be calculated from the coincidence counts collected for a given time interval, \(I_{ij}\):

\[
I_{ij} = N_{ij} / [(I_{ro})_{ij} / (I_r)_{ij}]
\]

Alternatively, a predetermined geometry can be generated by using a water bath around the object. By using a circular water bath, all \(L_{ij}\) values have a simple geometrical
relation and can easily be calculated. If an average value of \(\mu\) can be assumed for the entire object, the term \(e^{-\mu L_{ij}}\) can be obtained without requiring any additional measurements.

The actual boundaries of the object can also be directly determined by surrounding it with a weak line source of activity. The values of \(L_{ij}\) are computed from the boundaries as defined by the line source and an average value for \(\mu\) is assumed.

4.4.4. Random corrections

Because radioactive decay is a random process, there is a small, but finite probability that more than one positron will decay within a very short time interval. A random coincidence is defined as a coincident event resulting from the detection of two uncorrelated photons. These events can only contribute to the background noise as they are due to the accidental combination of two separate events. Practically, random coincidences will occur if a second pulse arrives within the resolving time \(\tau_r\) following a typical signal pulse. Because the time intervals separating adjacent events are randomly distributed, some will be less than the resolving time of the system. These random coincidences increase rapidly with increasing count rate. For a random pulse rate of \(r_s\) and \(r_s\tau_r \ll 1\), the rate at which coincidences occur should be the fraction of all times that lies within \(\tau_r\) of a preceding pulse (given
by $r_1 \cdot r_r$ multiplied by the rate of pulse arrival ($r_2$):

$$r_{\text{random}} = r_1 \cdot r_2 \cdot r_r$$

Therefore, if the individual singles rates in all detectors are measured and the resolving time of the system is known, the random coincident count rate in each detector pair can be calculated provided the dead time losses of the system are known.

Another approach is to assign a unique time delay to each detector and to measure the coincident count rate using these delays. These events can only be random coincidences since none of the events are correlated in time. However, this approach introduces a major complexity in the electronics. Furthermore, in order not to miss any true coincidences, the sum of all individual delays must be kept smaller than the time it takes the electronics to decide whether a true coincidence had occurred. This method allows the reconstruction of a random coincidences "image" that can be compared with the true coincidences image.

4.4.5. **Scattered corrections**

Scattered coincidences are those events that result from the detection of two correlated photons, but where one or both photons have been Compton scattered before reaching the detectors.
One method of determining the scattered component is to image a line source in a scattering medium making a series of measurements with the line source positioned at varying places in the field of view. This should give a good measure of the average scatter in the object. The scatter fraction is obtained from the data profiles before reconstruction. The number of events seen in the region of the source are assumed to be unscattered true coincidences, and all events outside that region are assumed to be scattered events. Performing this measurement in a series of positions in the field of view and weighting the data appropriately will give the average scatter fraction in the field of view.

Attempts at measuring scatter from the reconstructed image can only give ambiguous results. The scatter value observed in a zero activity region of a phantom image is the result of the convolution of the reconstruction filter and the attenuation correction on the scatter. This result is highly dependent on the choice of attenuation correction technique and is not necessarily related to the actual scatter fraction.
4.5. Discussion

4.5.1. Linear sampling

In general, the image distribution is just a noisy version of the convolution of the object distribution with the point spread function of the imaging system. If the point spread function is narrow compared with the details in the object distribution, the image will look very much like a noisy version of the object. If the object distribution contains details that are smaller in scale than the width of the point spread function, however, the details will be seriously smoothed out and may even appear to be lost altogether.

The phenomenon of aliasing is basic to the sampling of the data at equally spaced intervals. If the sampling interval is too large, the image will be distorted because some of its closely spaced components will overlap. From Fourier analysis, it can be shown that if the continuous range of frequencies are limited to a symmetrical band about the origin

\[-F < f < F\]

of width 2F and the sampling is done at an interval \( \Delta r \) between the samples, then, the relation

\[2f\Delta r < 1\]

applies. The maximum sampling interval (1/2F) is related to the folding, or Nyquist, frequency. The common way of
expressing this relation is to say that at least two samples must be obtained in the highest frequency present in order to avoid aliasing.

In medical imaging, there is no upper limit of frequencies for the structures that are imaged. Hence, the sampling theorem can never be satisfied. However, it gives an indication as to the size of the structures that can be imaged with a high degree of fidelity and sets an upper limit to the structures that may be identified, but not quantified.

4.5.2. Angular sampling

In theory, an infinite number of views must be available in order to reconstruct an image exactly. However, in any practical implementation of a tomographic device, only a finite number of views is available. The implication is that the density function cannot be reconstructed exactly in practice, even in the absence of any other source of errors. If too few views are obtained, substantial distortions will be introduced in the reconstructed image.

Intuitively, the angular sampling distance should be approximately equal to the linear sampling distance. If the diameter of the ring of detectors is D, its circumference is \( \pi D \), then:

\[
\text{number of angles} = \frac{\pi D}{2} \cdot \frac{1}{\Delta r}
\]
where $\Delta r$ is the linear sampling distance. This is stating that the length of arc around the object is divided into equally spaced increments of length $\Delta r$, the linear sampling distance.

The preceding discussion has bearings on the design of tomographs in which the detectors are not arranged along the circumference of a circle, and would determine the optimum angular rotation and number of views required to image an object. In ring tomographs, an angular projection is obtained at every detector position, and hence, the requirement is automatically satisfied.

4.5.3. Wobble motion

As has already been discussed, the lines connecting the centre of the faces of equally spaced detectors are all tangent to one of a finite set of concentric circles. The location of these circles is determined by the location of the centre of the ring, the size of the ring and the number of detectors. If the centre of the detector ring is fixed in space, the distance between circles limits the resolution that can be achieved by reconstructing the data collected by such a device. This limitation cannot be removed by rotating the ring about its centre, since the lines connecting the centre of the faces of the detectors remain tangential to the same circles as the ring rotates (fig. 4.12).
Figure 4.14: Wobble motion: sixteen detectors are placed along the circumference of a circle. The centre of the circle is successively located at each of four points, describing a small circle. Data is collected at each centre location and organized into parallel projections. The data from the four wobble points are combined together. Note that the sampling of the space is much finer than that obtained at an individual wobble point. Note also that the rays in a combined projection are not equally spaced and that there are redundant rays. For example, in the projection illustrated, the parallel rays obtained at location 1 and 3 sample the same space.
However, if a series of translations of the detector system is performed relative to the patient in the transaxial plane, this limitation can be overcome. If the ring is moved, so that its centre traces out a circle, data can be accumulated at any number of positions along this circle and reorganized so that the rays along which data have been collected are rebinned into sets of equally spaced parallel rays (fig. 4.14), spatial resolution can be substantially improved.

The mechanical implementation of such a scheme is simple. However, the time taken to accumulate a complete set of data is considerably increased, imposing limitations on the performance of dynamic studies or on the duration of examination of patients.

4.5.4. Dichotomic motion

Another scheme has been proposed in an effort to improve linear sampling. In this scheme, the detector array is hinged and is moved so as to open up a space the size of one detector at the point opposite the hinge. This mechanical displacement transforms the detector array from one with an even number of detectors to one with an odd number of detectors, halving the sampling distance in the process (fig. 4.15). This leads to an improved linear sampling and hence an improvement in spatial resolution with only a simple mechanical motion to two positions.
Figure 4.15: Dichotomic motion: the detector ring is hinged so as to open up a space the size of one detector at a location opposite the hinge. With the hinge closed, the ring contains an even number of detector locations; with the hinge open, it contains an odd number of detector locations. Note that in the first case, the geometric centre of the ring is sampled whereas in the second case, it is not. Combining the data obtained using the two geometries halves the sampling distance.
4.5.5. Continuous rotation

Yet another scheme has been proposed to satisfy the sampling requirement "with a simple motion" (!). The basis of this method is the use of a circular array of detectors in which the detectors are unequally spaced and the array continuously rotated at any desired speed; this provides suitable linear sampling during a complete rotation of the detector ring about its centre. The linear sampling density distribution depends on the arrangement of the detectors on the circle. The search for a suitable detector array configuration, providing the required fineness and density uniformity in the linear sampling, is difficult using analytical methods, but simple iterative methods have been developed for this end.

4.5.6. Conclusion

Many solutions to the problem of image reconstruction from projections have been proposed. The one we personally favour is that of convolution backprojection, because of the ease with which it can be implemented on a computer and because of the ease with which the corrections to the original data that are necessary in positron tomography can be performed.

Of the various motions of the detectors that have been proposed in an effort to improve the linear sampling,
and hence the spatial resolution, the dichotomic motion is the most attractive because of the ease with which it can be implemented, and because it requires minimal data reorganization. The continuous rotating positron camera with detectors of different dimensions is an interesting one from the theoretical point of view, but should prove difficult to implement in a clinical setting.
CHAPTER 5

DESIGN AND CONSTRUCTION OF THE TOMOGRAPH.

5.1. Introduction 142

5.2. The detector assembly 142
   5.2.1. Choice of scintillation crystal 142
   5.2.2. Detector size 145
   5.2.3. Septa 151
   5.2.4. Depth of shielding 155

5.3. The gantry 157

5.4. The electronics 160
   5.4.1. Pulse height discrimination 161
   5.4.2. Timing 162
   5.4.3. Coincidence logic 162

5.5. The data handling 170
   5.5.1. Reconstruction 170
   5.5.2. Display and manipulation 171

5.6. Performance of the tomograph 176
   5.6.1. Detection efficiency 176
   5.6.2. Linearity 178
   5.6.3. Scattered fraction 178
   5.6.4. Spatial resolution 180
   5.6.5. Axial resolution 183

5.7. Summary 183
5.1. **Introduction**

Many factors must be considered in the design of a tomograph. Two variables that have to be optimized, however, for a clinical instrument, are sensitivity and spatial resolution. These will be discussed in relation to the choice of scintillation crystal and their dimensions. Any tomograph can be subdivided into elementary building blocks: the scintillation crystal-photomultiplier, or detector, assembly; the gantry that contains the detector head and supports the lead shielding; the electronics that comprise analogue components, such as preamplifiers and amplifiers, and logic components, such as discriminators, gates and address formers; and the computer that stores the information, reconstructs the images and allows their display and manipulation. Each building block will be discussed separately, and the final design and performance of the tomograph that has been built will be presented.

5.2. **The detector assembly**

5.2.1. **Choice of scintillation crystal**

Three scintillators are currently being used in positron tomography; bismuth germanate offers two advantages over sodium iodide and caesium fluoride. Its stopping power, or linear attenuation coefficient, at 500 keV is more than
twice that of either sodium iodide or caesium fluoride (table 5.1). It is also nonhygroscopic and therefore does not need to be encapsulated in an aluminium can, thereby allowing for a higher packing fraction.

The one disadvantage of bismuth germanate is that its light decay time is relatively long (table 5.1). This limits the timing resolution attainable with pairs of detectors operating in coincidence. In that respect, caesium fluoride, with its much shorter decay time, allows for a narrower coincidence timing window. This narrow timing window would decrease the number of random coincidences accepted. The rejection of random coincidences becomes important only in very high count rate applications. However, in the majority of applications, the restrictions imposed by the radiation that can be given to patients, and the time taken for chemical synthesis limit the number of photons available so that maximum detection efficiency is much more important than random coincidences rejection.

Another possible use of caesium fluoride is time-of-flight tomography. However, the fastest coincidence timing possible with present day electronics is still not fast enough to improve substantially spatial resolution. Those timings are of the order of 500 to 600 psec (FWHM) which translate into a position uncertainty of 70 to 95 mm. Therefore, purely from a detection efficiency point of view, bismuth
Table 5.1: Comparison of some of the physical properties of sodium iodide (NaI(Tl)), caesium fluoride (CsF) and bismuth germanate (BGO).

<table>
<thead>
<tr>
<th></th>
<th>Density (g cm(^{-3}))</th>
<th>Linear attenuation coefficient at 500 keV (cm(^{-1}))</th>
<th>Scintillation decay time (ns)</th>
<th>Coincident timing resolution (FWHM) (ns)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NaI(Tl)</td>
<td>3.67</td>
<td>0.34</td>
<td>230</td>
<td>2-3</td>
</tr>
<tr>
<td>CsF</td>
<td>4.61</td>
<td>0.39</td>
<td>5</td>
<td>0.5</td>
</tr>
<tr>
<td>BGO</td>
<td>7.13</td>
<td>0.91</td>
<td>300</td>
<td>5-9</td>
</tr>
</tbody>
</table>
germanate is the scintillator of choice for most photon limited applications as is the case in studies of the brain.

5.2.2. Detector size

In the circular geometry, the detector to detector spacing ultimately determines the spatial resolution. Whereas a sodium iodide or caesium fluoride detector narrower than 20 mm is impractical because of the deterioration in spatial resolution resulting from crosstalk between adjacent detectors, bismuth germanate should allow narrower crystal widths.

In order to determine the most suitable detector size, a block of bismuth germanate was cut sequentially into smaller and smaller pieces using a rotating diamond wheel in a kerosene atmosphere. Each crystal was optically coupled to a 75 mm photomultiplier tube operated at 1000 volts. The energy spectrum obtained was displayed on a multichannel analyser. It was calibrated using sources of $^{51}$Cr, $^{22}$Na and $^{137}$Cs. Two windows were set, one from 360 to 700 keV, the other from 50 to 1000 keV. The photomultiplier tube and crystal assembly was well shielded with lead on all sides except for a long narrow aperture extending 15 cm in front of it. The cross section of this aperture was adjusted to correspond to the face section of the crystal being studied. A broad source of $^{18}$F was placed 40 cm in front of the crystal. Background
Figure 5.1: $^{18}$F spectrum obtained with a 8 x 30 x 30 mm$^3$ bismuth germanate crystal. The lower spectrum is the background spectrum obtained with the same crystal.
radioactivity was measured after a lead block, 10 cm thick, had been placed between the source and the detector (fig. 5.1). This was done for each crystal size investigated.

The photopeak efficiency is defined as the number of counts that fall in the photopeak window (360 to 700 keV) of a crystal expressed as a fraction of the number of counts in the photopeak of a reference crystal. The photopeak fraction is defined as the number of counts that fall in the photopeak window (360 to 700 keV) expressed as a fraction of the total number of counts registered (window: 50 to 1000 keV).

When the area of the exposed face of bismuth germanate crystals was reduced while maintaining the length of the crystal constant, the photopeak efficiency was found to be proportional to the area of the exposed face (fig. 5.2). However, as the area of the exposed face is made smaller, there is a degradation in both the photopeak fraction and the energy resolution (table 5.2). This means that when the crystal is long and narrow, there is incomplete light collection and this is translated into the effect observed.

When the length of the bismuth germanate crystal was reduced while maintaining the exposed face constant, there was a loss in photopeak efficiency (fig. 5.3). This means that as the crystal was made shorter, more gamma photons were escaping from the crystal without depositing all of their energy in the crystal. The 25 % increase in
Figure 5.2: Variation in photopeak efficiency as the crystal face is reduced while maintaining the length constant at 5 cm. The solid line represents the line of identity.

<table>
<thead>
<tr>
<th>Crystal face cm x cm</th>
<th>Photopeak fraction (a)</th>
<th>Energy resolution (b) percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.0 x 3</td>
<td>0.62</td>
<td>27.7</td>
</tr>
<tr>
<td>1.0 x 3</td>
<td>0.64</td>
<td>27.8</td>
</tr>
<tr>
<td>0.8 x 3</td>
<td>0.70</td>
<td>28.9</td>
</tr>
<tr>
<td>0.8 x 2</td>
<td>0.63</td>
<td>28.8</td>
</tr>
<tr>
<td>0.8 x 1</td>
<td>0.55</td>
<td>55.8</td>
</tr>
<tr>
<td>length = 3 cm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0 x 3</td>
<td>0.73</td>
<td>27.8</td>
</tr>
<tr>
<td>0.8 x 3</td>
<td>0.69</td>
<td>25.4</td>
</tr>
<tr>
<td>0.8 x 2</td>
<td>0.68</td>
<td>25.4</td>
</tr>
<tr>
<td>0.5 x 3</td>
<td>0.64</td>
<td>32.1</td>
</tr>
<tr>
<td>0.5 x 2</td>
<td>0.65</td>
<td>30.9</td>
</tr>
<tr>
<td>0.3 x 2</td>
<td>0.52</td>
<td>58.7</td>
</tr>
</tbody>
</table>

(a) fraction of counts in window 50-1000 keV that fall in window 360-700 keV  
(b) FWHM at 511 keV

Table 5.2: Effects of reducing the area of the exposed face of a bismuth germanate crystal while maintaining its length constant.
photopeak efficiency measured when the crystal length was increased from 30 mm to 50 mm is in apparent contradiction with the prediction that 24.4 mm are sufficient to stop 90% of all incident 500 keV photons. However, the predicted value is calculated using the total absorption coefficient, and assumes perfect geometry, i.e., a point source of radioactivity placed some distance away from an infinitely large crystal. Our measurements were made using a distributed source of dimensions similar to those of the crystal, an experimental situation approximating that encountered in positron tomography. Our results can be interpreted in the sense that most events falling in the photopeak were the result of primary photoelectric interactions, and that photons that were initially Compton scattered in the detector escaped from it without further interactions.

For the crystal dimensions investigated, there was no degradation in photopeak fraction or energy resolution (table 5.3), implying good light collection under this set of conditions. Actually, a slight improvement in energy resolution was noted as the crystal was made shorter, as this improves the geometry of light collection by the photomultiplier tube.

In order to increase the detection efficiency while maintaining the thickness of the section of the object being imaged constant, the effects of making the crystal
Figure 5.3: Variation in the photopeak efficiency as the crystal length is shortened from 5 to 3 cm.

<table>
<thead>
<tr>
<th>Crystal length cm</th>
<th>Crystal face cm x cm</th>
<th>Photopeak fraction</th>
<th>Energy resolution percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>1.0 x 3</td>
<td>0.64</td>
<td>27.7</td>
</tr>
<tr>
<td>3</td>
<td>0.8 x 3</td>
<td>0.73</td>
<td>27.8</td>
</tr>
<tr>
<td>5</td>
<td>0.8 x 3</td>
<td>0.70</td>
<td>28.9</td>
</tr>
<tr>
<td>4</td>
<td>0.69</td>
<td>28.7</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.68</td>
<td>25.4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.8 x 2</td>
<td>0.63</td>
<td>28.8</td>
</tr>
<tr>
<td>4</td>
<td>0.67</td>
<td>28.6</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.68</td>
<td>25.4</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.3: Effects of reducing the length of bismuth germanate crystal while maintaining the exposed face constant.
taller than the aperture was investigated. This stratagem should increase the probability of secondary interactions in the crystal and hence improve the photopeak efficiency. The experimental results show that when a 30 mm high crystal was combined with a 20 mm aperture, the observed detection efficiency was 10% greater than that of a 20 mm high crystal in the same geometry. As would be expected, an increase in the photopeak fraction was also noted (table 5.4).

In summary, there is a definite advantage in increasing the length of the crystal to 50 mm. This increase in length represents a 25% increase in detection efficiency over a crystal 30 mm long. Crystals of width smaller than 8 mm and of height smaller than 30 mm exhibit an appreciable degradation in photopeak efficiency. Finally, there is an advantage in having the crystal taller than the aperture.

5.2.3. Septa

In systems in which sodium iodide was used as scintillator, tungsten septa were required between adjacent crystals. This was done in order to minimize crosstalk and so improve spatial resolution. Because the stopping power of bismuth germanate for 500 keV photons is three times that of sodium iodide, and therefore closer to that of tungsten, we investigated the change in spatial resolution when tungsten septa are inserted between bismuth germanate crystals.
<table>
<thead>
<tr>
<th>Crystal dimensions cm x cm x cm</th>
<th>Counts in photopeak percent</th>
<th>Photopeak fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8 x 3 x 5</td>
<td>113</td>
<td>0.70</td>
</tr>
<tr>
<td>0.8 x 2 x 5</td>
<td>100</td>
<td>0.63</td>
</tr>
<tr>
<td>0.8 x 3 x 4</td>
<td>111</td>
<td>0.69</td>
</tr>
<tr>
<td>0.8 x 2 x 4</td>
<td>100</td>
<td>0.67</td>
</tr>
<tr>
<td>0.8 x 3 x 3</td>
<td>109</td>
<td>0.70</td>
</tr>
<tr>
<td>0.8 x 2 x 3</td>
<td>100</td>
<td>0.68</td>
</tr>
</tbody>
</table>

Table 5.4: Effects of increasing the crystal height while maintaining the opening at 0.8 cm x 2 cm.
Spatial resolution in this context is defined as the line response of a pair of detectors operated in coincidence.

Experiments were done with two banks of crystals operating in coincidence. Each bank consisted of four bismuth germanate crystals, 8 mm wide. The banks were placed along the circumference of a circle, 21 cm in radius. Measurements were done at 10° intervals from an initial angle between the banks of 180° (banks opposite each other) to a final angle of 90°. A 1 mm wide source containing $^{18}\text{F}$ was driven, in 2 mm increments, along a line normal to the line joining the two banks (fig. 5.4). From the measurements at each angle, the FWHM and the FW1/10M of the line response was estimated (fig. 5.5). With the two banks at 180°, the FWHM was 5.2 mm. When tungsten septa (2 mm wide) were inserted between each crystal, the FWHM was 5.0 mm. This small difference in response was maintained down to an angle of 140°. At 90°, the FWHM was 7.6 mm and 6.7 mm respectively. The FW1/10M showed a similar pattern, being 11.1 mm and 10.0 mm at 180°, and 20.7 mm and 15.5 mm at 90°. We concluded that in a system in which crystals, 8 mm wide, are arranged along the circumference of a circle, and the volume occupied by the septa is replaced by additional crystals, the overall efficiency of the system would improve by about 20% while adequate spatial resolution would be maintained.
Figure 5.4: Geometry used to measure the line response of the two banks of detectors operating in coincidence.

Figure 5.5: Line response of the two banks of detectors operating in coincidence. The solid curve is that obtained with septa between adjacent crystals; the dashed curve is that obtained with no septa.
5.2.4. **Depth of shielding**

In a tomographic device designed to look at horizontal slices of brain, the line of sight of opposed detectors defines the volume from which true events can originate. Photons that originate beyond the slice of interest, but are detected, contribute to the total count rate in each crystal, yet cannot contribute useful information. They also increase the probability of random coincidences. Thus, shielding is used on either side of the detector plane to block activity external to the transverse section being imaged.

We investigated the length of collimator required to reduce the detection of gamma photons originating from outside the plane of interest. A bismuth germanate crystal, $10 \times 25 \times 50 \text{ mm}^3$, was placed behind a lead collimator. The length of the collimator was 5 cm or 10 cm. A point source, 2 mm in diameter, containing a solution of a positron emitter ($^{68}\text{Ge/}^{68}\text{Ga}$) was moved in a direction perpendicular to the axis of the collimator in 5 mm increments at distances varying from 10 cm to 25 cm from the crystal face. Isoresponse curves were computed and plotted (fig. 5.6). It can be appreciated that a 5 cm long lead collimator allows a considerable amount of radiation from outside the plane of interest to be detected by the crystal. A 10 cm long collimator reduces substantially this amount of radiation.
Figure 5.6: Isoresponse curves obtained with a depth of shielding of 5 cm (a) and 10 cm (b).
Obviously, making the collimator even longer will reduce unwanted events even more. However, two points have to be considered. First, the patient port diameter has to be large enough to accommodate most heads at a variety of angles, and therefore cannot be reduced much further than 30 cm. Second, to increase the depth of shielding would then require placing the detectors along the circumference of circles of increasing diameters, with the concomitant loss in detection efficiency due to the reduced solid angle subtended by the detectors.

A compromise was adopted that consisted of a patient port of 30 cm in diameter and a depth of shielding of 10 cm.

5.3. The gantry

The gantry comprises the detector assembly and the lead shielding (fig. 5.7). The detector assembly consists of 160 bismuth germanate crystals closely packed on a ring, 53.5 cm in diameter; adjacent crystals are not separated by inserts. Each crystal is trapezoidal in shape; the face dimensions are 8 X 25 mm$^2$, the length is 45 mm. The packing fraction is better than 0.95. Each crystal is directly coupled to a 12.7 mm photomultiplier tube (fig. 5.8). In addition, the detector assembly contains voltage dividers and a preamplifier for each detector (fig.5.9). The analogue signals from the gantry are sent to a separate electronics rack for further processing.
Figure 5.7: Photograph of the gantry. The collimator located towards the patient's head has been removed.

Figure 5.8: Photograph of the crystals fitted with their photomultiplier tube.
Figure 5.9: Photograph of a portion of the detector assembly.

Figure 5.10: Photograph of the tomograph and the patient's couch.
An annular lead collimator, 32 cm I.D., 52 cm O.D., is placed between the face of the crystals and the patient's head. This opening of 32 cm allows the comfortable positioning of most patients. The collimator is 10 cm thick towards the body of the patient and 5 cm thick in the other direction. The slice thickness can be fixed at either 20 mm or 10 mm.

The whole assembly can be rotated 20° in either direction in order that the axis of the tomographic plane be fixed in relation to the orbito-meatal line. A patient couch is appended to the gantry. It provides support for the head of the patient, and can be accurately indexed in millimeter increments (fig. 5.10).

5.4. The electronics

The output of the detector is a pulse that contains information about the energy that the incident gamma photon has deposited in the crystal, and about the precise instant in time the photon interacted with the scintillator. This information has to be extracted from the characteristic shape of the pulse by the electronics. Decisions as to the occurrence of valid coincidences, and the formation of an address that records the pair of detectors involved in a true coincident event are also performed by the electronics. This address is then sent to the computer for storage and subsequent manipulation.
5.4.1. Pulse height discrimination

The two principal interactions of 511 keV photons with matter are the photoelectric interaction and Compton scattering. Because in a photoelectric interaction the energy of the photon is completely transferred to a bound electron, this photon will never be detected. In Compton scattering interactions, the photon will be scattered by an electron, losing only part of its energy in the process, and may reach a detector. Photons may be scattered in the plane of the detector ring, and no amount of collimation will discriminate against them. Although due to a true coincident event, the line joining the detectors in which these events have been detected will not be the line along which the annihilation of the positron has taken place, and hence will result in mispositioning of the event (fig. 4.13). Photons that are scattered from outside the plane of the detectors may still reach the detectors in spite of adequate shielding. In these two cases, because a photon loses energy when it is scattered and because this loss is proportional to the scattering angle, a narrow energy window will help reduce this inherent problem. In practice, only a lower level discriminator is needed, as the isotopes of biological interest, $^{15}$O, $^{13}$N, $^{11}$C, $^{18}$F, are pure positron emitters and hence no gamma ray of energy higher than 511 keV can be present.
5.4.2. Timing

The timing window serves two purposes. The first is to flag those events that were detected within a very short time interval and therefore could be assumed to be correlated in time. The second is to reject events that are not correlated in time. However, some events that are detected within the resolving time of the system may result from the annihilation of two separate positrons. These random coincidences act as a background above which the true coincidence peak is located (fig. 5.11). The number of random coincidences can be reduced by collimation, as discussed above. Another way of reducing random coincidences is to shorten the timing window of the system, since the number of random coincidences is the product of the singles rates and the resolving time.

5.4.3. Coincidence logic

The coincidence logic comprises a sector controller module, five sector logic modules, each handling one sector, and twenty discriminator modules, each handling eight detectors (fig. 5.12).

When a photon is stopped in one of the detectors, a leading edge timing discriminator generates a fast timing pulse. If the energy of the photon is greater than a set threshold, this pulse passes through the appropriate input gates and the sector logic module forms an address that

\[ \text{With this, there are no advantages in using constant fraction discriminators.} \]
Figure 5.11: Time spectrum obtained using two channels of the tomograph. Note that the time window includes random coincidences in addition to the true coincidences.
Figure 5.12: Block diagram of the coincidence logic: Each sector logic module handles 32 detectors. They communicate with the sector controller module. When the sector controller detects a valid coincidence event, it forms a 14 bits address that represents the detector pair involved and sends it to the NPR interface which communicates with the computer via the UNIBUS.
represents the number of the detector (fig. 5.13). A preset number of nanoseconds later, all input gates are closed (fig. 5.14). Two situations can arise: either no other photon was detected within the set resolving time of the system, or one or more photons were detected. In the first case, the event was a 'singles' event, and all the input gates are reopened with no further processing of the initiating event (fig. 5.15). In the second case, two or more events were detected 'in coincidence'. If more than two sectors were involved, this cannot represent a valid coincidence; hence all input gates are reopened with no further processing of the data (fig. 5.16). If exactly two sectors had detected an event, the sector controller will form a 14 bit binary address which represents the addresses of the detector pair, will send that address to a non-processor request (NPR) interface and will reopen the input gates when that address has been accepted by the computer (fig. 5.17). Meanwhile, the NPR interface generates an 18 bit address by adding a preset base address to the 14 bits received from the sector controller, and then increments that particular word of PDP 11/45 memory.

Raw data is accumulated in memory for a preset number of seconds, and then written onto an RK 05 removable disk cartridge for storage and subsequent analysis; thus allowing the collection of data for rapid sequential images.
Figure 5.13: Detecting a 511 keV photon: If the pulse appearing at the input of channel $n$ is greater in energy than a set threshold, and if GATE $x$-$xx$ is asserted (gates are open), the READY $n$ signal will be asserted, until it is reset by the pulse INIT $x$-$xx$. The READY $n$ signal will pass on to the sector logic module where it will form an address that represents the number of the channel, and will assert the EVENT $s$ signal that will assert REQUEST and be passed on to the sector controller module.
Figure 5.14: Closing the input gates: The EVENT s signal, passed on from the sector logic module s will set the RS flip flop. In so doing, all input gates are disabled. An INIT pulse will eventually reset the flip flop and in so doing enable the gates.
Figure 5.15: Rejecting a 'singles' event: With only one of the EVENT signals asserted, COINC will be asserted. COINC will combine with EVENT to generate an INIT pulse that will reopen the gates through the RS flip flop and will reset READY n.
Figure 5.16: Rejecting a coincidence event involving more than two sectors: With more than one EVENT's signal asserted during the propagation time delay in the three inverters in series, COINC will be asserted. COINC will combine with EVENT to generate a GRANT pulse that is passed on to the first sector logic module. If this sector logic module was not asserting REQUEST, GRANT will pass on to the next sector logic module. If a sector logic module is asserting REQUEST, GRANT will transfer the address of the detector involved to the sector controller and will clear REQUEST. A second GRANT pulse is then issued by the sector controller. It will be passed on to the following sector logic module asserting REQUEST. The address of this second detector will be passed on to the sector controller and REQUEST will be cleared. If REQUEST is still asserted at the sector controller, this means that a third sector had also detected an event. In this case, an INIT pulse will be generated and the processing aborted.
Figure 5.17: Processing of a valid coincidence event: If two, and only two, sectors were asserting REQUEST, the address of the two detectors involved will be combined into a 14 bit address and DATA READY asserted. When the NPR interface has processed this coincidence event, it asserts DATA ACCEPTED. This will generate an INIT pulse that will reopen the input gates and there the two READY signals at the discriminator modules.
5.5. The data handling

5.5.1. Reconstruction

The 10,240 possible coincident rays, or line integrals of the activity along the path joining two opposed detectors, are arranged into 160 divergent projections, each containing 91 rays; the other rays are rejected as they fall well outside the field of view. Each line integral is corrected for the relative efficiency of the detector pair involved in its measurement using the information gathered in imaging a cylindrical phantom, 25 cm in diameter, uniformly filled with a solution of $^{68}$Ge/$^{68}$Ga. Each line integral is also weighted by a factor proportional to its path length in the object and representing the absorption of gamma photons in the object. A filtered backprojection algorithm developed for divergent geometry is then used to reconstruct the distribution of activity in the plane being studied. If the rays were arranged into parallel projections, either rotation of the ring by half a detector spacing, or interpolation between rays would be required to achieve the same reconstructed resolution. The filter used in the reconstructions is the Hamming window; the choice of its parameter depends on the statistical accuracy of the data being reconstructed. The data is backprojected onto a 128 X 128 matrix, each pixel representing 2 mm. The reconstructed images are stored on an RK 05 removable disk cartridge.
5.5.2. Display and manipulation

The reconstructed image consists of a two dimensional array of numbers, each representing the extent of accumulation of the tracer at a particular location in the transverse plane. However, for the ease of interpretation of the image, it is necessary to display it in a fashion that is familiar to the human mind, i.e., as a topographical representation in which picture elements (pixels) of similar value are assigned the same visual symbol, either a solid colour, the shade of a colour, or a distinctive pattern. It is also necessary to consider the spatial organization of the data on the display in order to avoid the introduction of artifacts, the creation of detail finer than that that can be obtained by the imaging device, or the loss of detail that can be recorded by it.

The use of distinctive patterns to identify areas of equal accumulation of radioactivity is useful only in applications where the image is to be printed on paper. The use of video monitors and photography is a much more versatile way of manipulating the image. These monitors are either monochrome displays, or colour displays. Each type of display has advantages and disadvantages. For a continuous distribution of shades of colour from violet to red, or of shades of grey (for example) from black to white, the just-noticeable-differences, the distance between
two colours or two shades of grey that can be unequivocably discerned, will be smaller in the latter case. This means that the resolution of the human eye for colour differences is less than for monochrome differences. This implies that the dynamic range of a monochrome display, the range of differences that can be displayed with some certainty, will be greater and, consequently, that this type of display should be preferred. However, displayed images will invariably contain structure, and it is this structure that is most important in the final analysis of the image; the eye's perception of a monotonically varying colour function is irrelevant. In that respect, the eye is most sensitive to sharp edges. In a black and white display, adjacent grey levels will tend to merge together, thus creating the illusion of edges. In a monotonically varying colour display, no edges are created by adjacent colours. Hence, it might be easier to distinguish an object in black and white rather than in colour. However, there exist other colour schemes different from the monotonically varying one, schemes that will introduce edges in the displayed image. One such colour scheme is obtained by varying the intensity of each of a number of distinct colours from zero to 100%. Therefore, theoretically at least, there is no best display, and the perception of the image by the observer will strongly depend on his degree of training in looking at, and interpreting, the images (fig. 5.18). In practice,
Figure 5.18: Illustration of a black and white display and two different colour displays.
however, black and white displays might be more appropriate for the display of radiographic images. For example, in x-ray computerized tomography of the brain, the lower and upper ends of the scale are usually ignored, and the middle portion stretched. This is done because structure is not sought either in high attenuation, or low attenuation, tissues; most of the dynamic range of the display is restricted to a narrow range of attenuation coefficients usually centered around that of grey matter. In positron tomography, the situation is not similar, and no loss of contrast is acceptable at either ends of the scale: each value of the local accumulation of radioactivity has an intrinsic content that is important in the analysis of the image.

The other consideration is that of the topographical arrangement of the data. From a purely aesthetic point of view, it is important that the observer not see individual pixels. This means that a few display pixels should be used for each stored pixel. This requires the values of the stored pixels to be interpolated to the locations of the display pixels. It is worth noting that in this respect the noise in a noisy image will mask the digital appearance of the displayed image. The choice of interpolating function is important as it may lead to resolution losses. Other factors have also to be considered such as the dimensions of the video display screen and the distance at which it is viewed.
For a fixed video display screen size and a fixed viewing distance, the larger the image is made on the screen, the more imaging pixels are needed to represent the stored image. In our case, the stored image consists of a 128 X 128 array; one 512 X 512 image, or four 256 X 256 images, or sixteen 128 X 128 images can be displayed simultaneously.

Once the image is displayed in a particular format using a particular colour scheme, it is useful to be able to threshold it, i.e., to assign the full dynamic range of the display to only a portion of the range spanned by the values for the stored image, in order to enhance small differences in accumulation of radioactivity. Any decisions as to the significance of these differences is based, however, on the comparison of the actual values for the accumulation of isotope in those areas. These can be easily obtained using subroutines that allow the definition of regions of interest in the stored image, that integrate the number of counts in those regions, and that output those numbers. This capability is also used to obtain the time course of radioactivity in predefined regions.
5.6. **Performance of the tomograph**

5.6.1. **Detection efficiency**

The detection efficiency of the tomograph was determined using a cylindrical phantom, 20 cm in diameter and 20 cm long, uniformly filled with a 1 µCi/ml solution of $^{18}$F. The total number of coincidence events recorded was 46,500 cps for a 20 mm wide slice and 18,200 cps for a 10 mm wide slice. The number of random coincidences was calculated using the total number of singles events per detector and a resolving time of 13 ns. These were typically 9,100 cps for the 20 mm slice and 2,200 cps for the 10 mm slice, resulting in rates of 37,400 cps and 16,000 cps respectively for true coincidences. No effort was made to estimate the proportion of scattered events in these measurements.

A comparison of these numbers with those reported by users of other tomographs is difficult as they depend strongly on the slice thickness and the geometry of the detector assembly. Of the tomographs with a smaller ring diameter, the detection efficiency of the Neuro PET (National Institute of Neurological and Communicative Disorders and Stroke, Bethesda MD) is reported as 44,000 cps/µCi/ml. It is a tomograph in which 128 bismuth germanate detectors are arranged along the circumference of a circle, 34 cm in diameter; the slice thickness is 20 mm and the depth
of shielding is 4.5 cm. The detection efficiency of the Positome III (Montreal Neurological Institute, Montreal, PQ) is reported as 110,000 cps/μCi/ml. It is a tomograph in which 64 bismuth germanate detectors are arranged along the circumference of a circle, 42 cm in diameter; the slice thickness is 30 mm and the depth of shielding is 5.0 cm.

Of the tomographs with a larger ring diameter, the detection efficiency of the Neuro ECAT (U.C.L.A., Los Angeles CA) is reported to be 30,000 cps/μCi/ml. It is a tomograph in which 88 bismuth germanate detectors are arranged along the perimeter of an octagon, 65 cm to a side; the slice thickness is 27 mm and the depth of shielding is 17 cm. The detection efficiency of the Donner Laboraroty tomograph (Berkley CA) is reported to be 11,800 cps/μCi/ml. It is a tomograph in which 280 bismuth germanate detectors are arranged along the circumference of a circle, 94 cm in diameter; the slice thickness is 20 mm and the depth of shielding is 22 cm.

The detection efficiency of PET VI (Mallinckrodt Institute of Radiology, St. Louis MO) is reported to be 32,000 cps/μCi/ml. It is a tomograph in which 72 caesium fluoride detectors are arranged along the circumference of a circle, 57 cm in diameter; the slice thickness is 29 mm and the depth of shielding is 14 cm. Normalising for a standard slice thickness of 10 mm, these numbers are 11,000 cps/μCi/ml for the Neuro PET, 12,200 cps/μCi/ml for the Positome III, 4,100 cps/μCi/ml for

*normalisation is performed by assuming that the efficiency is proportional to the square of the slice thickness.
the Neuro ECAT, 3,800 cps/μCi/ml for the PET VI, and 2,900 cps/μCi/ml for the Donner Laboratory tomograph. Thus, our detection efficiency of 16,000 cps/μCi/ml, measured using a 10 mm wide slit, compares favorably with that of other tomographs presently in use.

5.6.2. **Linearity**

The linearity of the tomograph in response to decreasing amounts of radioactivity in the field of view was determined using the 20 cm cylindrical phantom and the 10 mm wide aperture. The phantom was uniformly filled with a 350 nCi/ml solution of $^{18}$F and repeatedly imaged over the following nine hours; the final concentration of radioactivity in the field of view was 11 nCi/ml. The response of the tomograph was linear over that range of concentrations of radioactivity (fig. 5.19).

5.6.3. **Scattered fraction**

The number of true coincidence events scattered in the field of view of the tomograph, and thus contributing erroneous position information, was measured using a cylindrical phantom, 25 cm in diameter and 10 cm long, fitted with a cylinder, 5 cm in diameter and 10 cm long. The larger cylinder was filled with a solution of $^{18}$F and the smaller cylinder was filled with water. The counts in a region of interest placed within the area containing radioactivity were compared with those obtained in a similar region of
Figure 5.19: Response of the tomograph to decreasing amounts of radioactivity in the field of view. Each point represents counts accumulated over a 20 minutes period. The statistical error associated with the last point on the graph is of the order of 0.1%.
interest placed within the area containing no radioactivity. The counts originating from the latter region could only have been due to mispositioned events. They represented 14% of the counts in the region containing radioactivity. This phantom was imaged repeatedly over a period of nine hours during which the concentration of $^{18}$F radioactivity would have fallen to about 3% of its original value. The scattered fraction did not change over the period of observation (fig. 5.20). This represents an expected result since the scattered fraction should not be dependent on the amount of radioactivity, provided the geometry of the measurement remains the same. It should be noted, however, that the distribution of scattered coincidences will depend on the local distribution of radioactivity in the plane, so that the value of this measurement will change from object to object.

5.6.4. Spatial resolution

The inplane resolution was measured using cylinders of varying diameters filled with a solution of $^{18}$F. The results were expressed as the FWHM of the line response measured in an image. Cylinders of 2 mm and 4 mm (inside diameter) gave a FWHM of 6 mm; cylinders of 8 mm, 10 mm, 12 mm, 16 mm and 20 mm (inside diameter) gave a FWHM equal to the inside diameter of the cylinder being imaged (fig. 5.21). It can be concluded that the spatial resolution is, at best, 6 mm (FWHM). These values were obtained using a sharp filter
Figure 5.20: True coincidences count rate (upper curve) and scattered coincidences count rate (lower curve) as a function of radioactivity in the field of view. (10 mm 51I).

cps
10 000
1 000
100

hours
1 2 3 4 5 6 7 8 9
Figure 5.21. Observed spatial resolution of the tomograph as a function of object size. Note that objects smaller than 8 mm give the same observed resolution. The minimum error on the determination of the observed resolution is ± 1 mm. Measurements made at the center of the tomograph.
in the reconstruction algorithm and it should be remembered that the observed spatial resolution depends strongly on the choice of filter function. In clinical routine use, the spatial resolution is probably closer to 8 mm (FWHM). Fifteen centimeters away from the axis of the tomograph, the spatial resolution is 10 mm (FWHM).

5.6.5. Axial resolution

The axial resolution of the tomograph was measured using a 1.2 mm diameter, 1 mm long $^{68}$Ga source. This point source was moved in a direction parallel to the axis of the tomograph at various distances from the centre. For a 20 mm wide slice, the FWHM and FW1/10M at the center of the tomograph were 12 mm and 25 mm respectively. Fifteen centimeters away from the axis of the tomograph, these were 14 mm and 26 mm respectively (fig. 5.22). For a 10 mm wide slice, the FWHM and FW1/10M at the centre of the tomograph were 10 mm and 20 mm respectively. Fifteen centimeters away from the axis, these were 11 mm and 20 mm respectively (fig. 5.22).

5.7. Summary

Based on a series of experiments designed to maximize detection efficiency while maintaining adequate spatial resolution, a positron tomograph has been constructed. It comprises 160 bismuth germanate detectors, $10 \times 25 \times 45$ mm$^3$ closely packed on a ring, 53.5 cm in diameter. The tomographic
Figure 5.22: Response of the tomograph to a point source of radioactivity moved along the axis of the tomograph for a 10 mm wide (a) and a 20 mm wide (b) aperture.
slices examined are either 10 mm or 20 mm thick. The spatial resolution is 8 mm (FWHM). The sensitivity is 16 000 cps/\( \mu \text{Ci/ml} \) for a 10 mm slice. Its measured spatial resolution is comparable to the best spatial resolution reported so far for other tomographs. Its detection efficiency outperforms that of all other tomographs presently in use.

The tomograph allows the quantitation of structures in the brain of the order of 1 cm\(^3\). It has been in routine clinical use for more than eighteen months. Its proper operation is checked daily using a cylindrical phantom, uniformly filled with a solution of \(^{68}\text{Ge}/^{68}\text{Ga}\); \(^{68}\text{Ge}\) decays by electron capture to \(^{68}\text{Ga}\) with a half life of 250 days. So far, no degradation in detection efficiency or field uniformity has been noted. The electronics have been stable and problem free. Examples of its use in the measurement of regional metabolic rates in the brain will be presented in the following chapter.
CHAPTER 6

THE STUDY OF REGIONAL CEREBRAL METABOLISM DURING LIFE.

6.1. Introduction

6.2. The tracer technique
   6.2.1. Assumptions in tracer techniques
   6.2.2. Advantages of substrate analogues in tracer techniques

6.3. The measurement of energy requirements in the brain
   6.3.1. Introduction
   6.3.2. The deoxyglucose method
      6.3.2.1. Rationale of the method
      6.3.2.2. Compartmental analysis
      6.3.2.3. Discussion

6.4. The measurement of intracerebral neurotransmitter metabolism
   6.4.1. Introduction
   6.4.2. The fluorodopa method
      6.4.2.1. Rationale of the method
      6.4.2.2. Compartmental analysis
      6.4.2.3. Discussion

6.5. Correlation between CT, NMR and PT findings in the brain
   6.5.1. Introduction
   6.5.2. Examination of nigro-striatal pathway
   6.5.3. Examination of the hypoxic brain
   6.5.4. Discussion

6.6. Conclusions
6.1. Introduction

Even though positron tomography has been used to observe static processes, such as regional blood volume, the main advantage of the technique is that it allows for the regional measurements of dynamic processes in the brain. For example, regional blood flow and energy requirements have been measured using \(^{15}\text{O}\) water and \(^{18}\text{F}\) 2-deoxyglucose or \(^{15}\text{O}\) oxygen respectively. More recently, the distribution of a neurotransmitter and its fate have been obtained using \(^{18}\text{F}\) fluorodopa. These studies provide information that could not be obtained previously.

Apart from blood flow, these dynamic processes are attempts at measuring, in vivo, rates of chemical reactions. A chemical reaction is the conversion of one species of molecule to another; the rate at which this conversion proceeds is determined by measuring the disappearance rate of one or more of the reactants, or, alternatively, the formation rate of one or more of the products. The addition of a positron emitting label to one of the reactants in molecular concentrations small enough not to alter the kinetics of the overall reaction allows the measurement of the chemical transformation rate of the labelled species using externally placed radiation detectors. To derive
the total reaction rate from the measurement of the reaction rate of the labelled species, it is necessary to know the integrated specific activity (the ratio of labelled to total molecules at all times) of the precursor pool.

In assays of biochemical reactions in vivo, it is not possible to measure the integrated specific activity of the precursor pool directly as this calculation would require measuring the complete time course of the concentrations of the labelled and unlabelled precursor molecules in the tissue itself, at the enzyme site. It is therefore necessary to determine the specific activity of the precursor pool indirectly from measurements made in the blood that supplies the tissue. The specific activity in the arterial blood can be readily measured directly, and the specific activity of the precursor pool can then be calculated from it by correcting for the lag in the equilibration of the precursor pool in the tissue behind that of the blood; to apply this correction it is necessary to know the kinetics of the equilibration process between the precursor pool in the tissue and in the blood.

Furthermore, emission tomography measures only the total concentration of isotope and cannot distinguish among the various chemical species that may be labelled, including the precursor and any of the possible labelled products. Strategies must therefore be developed to ensure that the
radioactivity is contained exclusively in the precursor, or in one or more of the products specific to the chemical reaction to be assayed. The labelled precursor should also be selected so that the radioactive label is confined only to the specific reaction or pathway under study.

6.2. The tracer technique

6.2.1. Assumptions in tracer techniques

Several important assumptions must be addressed in the application of any tracer technique. The first assumption is that the tracer is transported and metabolized in the same manner and at the same rate as the compound being traced. This requirement is clearly met when a substrate such as glucose is labelled with $^{11}$C. $[^{11}\text{C}]$ glucose is biologically indistinguishable from native, unlabelled, glucose. This requirement is not met in the case of substrate analogue tracers, such as $[^{18}\text{F}]$ 2-deoxyglucose or $[^{18}\text{F}]$ fluorodopa, and appropriate correction factors must be applied.

The second assumption is that the metabolized tracer is retained within the area of interest during the measurement period. Alternatively, the products of further metabolism are removed from the field of view and can be accounted for. For example, this is true when fluorodopa is used to trace dopamine metabolism. In this case, the label is associated predominantly with either fluorodopa, the precursor molecule,
fluorodopamine, the molecule being traced, or fluorohomovanilic acid, the result of further metabolism of the dopamine molecule; when formed, fluorohomovanilic acid is efficiently removed from the field of view.

The third assumption is that the amount of tracer not metabolized, i.e., free tracer in the blood and the extracellular fluid, is either negligibly small, or can be accounted for at the time of the measurement. This requirement is most easily met when labelled analogues that are irreversibly trapped within the tissue are used. It allows for the delay of the actual measurement until free tracer in tissue and blood has fallen to insignificant levels, thus minimizing the errors in estimating the free tracer in the tissue.

The fourth assumption is that of linearity and stationarity, i.e., the system being investigated can be regarded as being in a steady state; metabolic rates, blood flow, blood volume, pool sizes, fractional exchange rates, concentrations of unlabelled species are all assumed to be time invariant.

6.2.2. Advantages of substrate analogues in tracer techniques

Radiolabelled substrate analogues have a number of unique advantages. First, some of these compounds can be designed so as to remain irreversibly trapped within the tissue once they have been metabolized. Thus the egress of
metabolites is of no concern when they are used to measure metabolic rates. Because these compounds remain irreversibly trapped within tissue, it is possible to minimize the corrections for free tracer in tissue and blood. By reducing or eliminating the necessity for making such corrections, the accuracy of the measurement of metabolism is obviously enhanced. Second, other substrate analogues can be labelled with longer lived isotopes so as to allow the measurement of slower metabolic rates.

However, radiolabelled substrate analogues do have a number of noteworthy disadvantages. First and most importantly, they are not biochemically identical to the compounds being traced. When radiolabelled substrate analogues are used, it is necessary to correct for differences in transport properties and enzyme affinities which have been shown to vary among species. Such corrections may present an added difficulty when the organ being studied is diseased. Second, because the strategy used relies on the time for unmetabolized tracer to clear from the blood and tissues, measurement times tend to preclude the examination of transient phenomena, and repeat measurements. This time requirement also necessitates the administration of somewhat larger quantities of the radionuclide and/or the use of radionuclides with longer half lives in order to achieve adequate counting statistics. Third, substrate
analogues are capable of disrupting normal metabolism if they enter the cell in sufficiently high concentration. Although this is not a problem when tracer quantities of radiolabelled substrate analogues are used, it is important that synthesis techniques employed in the production of such tracers assure very low concentrations of unlabelled carrier, i.e., very high specific activity.

6.3. The measurement of energy utilization in the brain
6.3.1. Introduction

Although most tissues can oxidize various substrates, such as amino acids, sugars and fatty acids, to fulfill their energy requirements, glucose is the predominant energy source available to the brain. Glycolysis is the sequence of reactions that convert glucose to pyruvate with the concomittant production of adenosine triphosphate, the central molecule in energy exchanges in biological systems (figure 6.1). Under aerobic conditions, pyruvate is eventually completely oxidized to carbon dioxide and water.

Satisfactory measurements of regional cerebral glucose metabolism can be achieved using \(^{11}\text{C} \) glucose but the measurement must be performed within the first few minutes following the administration of the tracer. Beyond about five minutes some of the labelled products of glucose metabolism are lost too rapidly from the tissue
Figure 6.1: The glycolytic pathway
and many other labelled products that are dependent on additional chemical reactions other than glucose metabolism are retained or appear in the venous effluent.

6.3.2. The deoxyglucose method

6.3.2.1. Rationale of the method

The deoxyglucose method was developed to measure the rates of glucose utilization simultaneously in all structural and functional components of the central nervous system in conscious individuals.

The requirement for metabolic trapping is achieved when substrate analogues, such as 2-deoxyglucose or 2-fluoro-2-deoxyglucose, are used. 2-deoxyglucose is transported bidirectionally between blood and brain by the same carrier that transports glucose across the blood brain barrier. In the cerebral tissues, it is phosphorylated, like glucose, by hexokinase to produce 2-deoxyglucose-6-phosphate. Deoxyglucose and glucose are competitive substrates for both blood brain transport and hexokinase-catalysed phosphorylation. Unlike glucose-6-phosphate, however, which is eventually metabolized further to pyruvate, 2-deoxyglucose-6-phosphate cannot be converted to fructose-6-phosphate (fig. 6.2), and it is also a poor substrate for glucose-6-phosphate dehydrogenase. There is relatively little glucose-6-phosphatase activity in the brain, and even less 2-deoxyglucose-6-phosphatase activity.
Figure 6.2: Comparison of the pathways for glucose and 2-deoxyglucose. The latter compound cannot proceed through the isomerization step.
Therefore the label is retained in the tissues in either of two chemical species: the unmetabolized precursor molecule, or the immediate product of its metabolism.

6.3.2.2. **Compartmental analysis**

The quantity of 2-deoxyglucose-6-phosphate accumulated in any cerebral tissue at any time following the introduction of 2-deoxyglucose into the circulation is equal to the integral of the rate of 2-deoxyglucose phosphorylation by hexokinase in that time during that interval of time. This integral is, in turn, related to the amount of glucose that has been phosphorylated over the same interval, depending on the time course of the relative concentrations of 2-deoxyglucose and glucose in the precursor pool and the Michaelis-Menten kinetic constant for hexokinase with respect to both 2-deoxyglucose and glucose. With cerebral glucose consumption in a steady state, the amount of glucose phosphorylated during the interval of time equals the steady flux of glucose through the hexokinase-catalysed step times the duration of the interval, and the net rate of flux of glucose through this step equals the rate of glucose utilization.

These relations can be rigorously combined into a model (fig. 6.3) which can be mathematically analysed to derive an operational equation (figs. 6.4, 6.5).
**Figure 6.3:** Compartmental model for the 2-deoxyglucose method.

<table>
<thead>
<tr>
<th>PLASMA</th>
<th>PRECURSOR POOL</th>
<th>METABOLIC PRODUCTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>glucose</td>
<td>glucose</td>
<td>glucose-6-phosphate</td>
</tr>
<tr>
<td>2-deoxyglucose</td>
<td>2-deoxyglucose</td>
<td>2-deoxyglucose-6-phosphate</td>
</tr>
</tbody>
</table>
Rate of change of free tracer in tissue:

\[
\frac{dC_2}{dt} = k_{21} C_1 - (k_{12} + k_{32}) C_2 + k_{23} C_3 \quad ; \quad C_2(0) = 0
\]

Rate of change of metabolized tracer in tissue:

\[
\frac{dC_3}{dt} = k_{32} C_2 - k_{23} C_3 \quad ; \quad C_3(0) = 0
\]

Amount of free tracer in tissue:

\[
C_2(t) = \left(\frac{k_{21}}{\alpha_2 - \alpha_1}\right) \left[ (k_{23} - \alpha_1) e^{-\alpha_1 t} + (\alpha_2 - k_{23}) e^{-\alpha_2 t} \right] * C_1(t)
\]

Amount of metabolized tracer in tissue:

\[
C_3(t) = \left(\frac{k_{21} \cdot k_{32}}{\alpha_2 - \alpha_1}\right) \left[ e^{-\alpha_1 t} - e^{-\alpha_2 t} \right] * C_1(t)
\]

Total amount of tracer in tissue:

\[
C_1(t) = C_2(t) + C_3(t)
\]

\[
= \left(\frac{k_{21}}{\alpha_2 - \alpha_1}\right) \left[ (k_{32} + k_{23} - \alpha_1) e^{-\alpha_1 t} + (\alpha_2 - k_{32} - k_{23}) e^{-\alpha_2 t} \right] * C_1(t)
\]

where:

\[
\alpha_1 + \alpha_2 = k_{12} + k_{32} + k_{23}
\]

\[
\alpha_1 \cdot \alpha_2 = k_{12} \cdot k_{23}
\]

and the symbol * denotes the convolution operation.

**Figure 6.4:** Mathematical analysis of the 2-deoxyglucose model. Solution of the differential equations describing the time course of unmetabolized and metabolized tracer in cerebral tissue.
In a steady state, the local cerebral metabolic rate for glucose is equal to the net phosphorylation rate of glucose:

\[ L_{CMRg} = k_{32} C_2 - k_{23} C_3 \]
\[ = \phi k_{32} C_2 \]

where \( \phi = 1 - \frac{k_{23} C_3}{k_{32} C_2} \) and represents the fraction of glucose that is metabolized after it is phosphorylated.

Also in a steady state, there is no net accumulation, or depletion, of glucose:

\[ k_{21} C_1 + k_{23} C_3 = k_{12} C_2 + k_{32} C_2 \]

\[ C_2 = \frac{k_{21}}{k_{12} + \phi k_{32}} C_1 \]

Thus:

\[ L_{CMRg} = \frac{k_{21} k_{32} \phi}{k_{12} + \phi k_{32}} C_1 \]

---

**Figure 6.4:** (continued)

Results derived from the steady state condition for glucose metabolism.
Multiplying the numerator and denominator of the last equation by:

\[
\frac{(k_{21}^* \cdot k_{32}^*)}{(k_{12}^* + k_{32}^*)}
\]

where the superscript * denotes the rate constants for the label, the equation becomes:

\[
L_{CMR_g} = \frac{k_{21}^* \cdot k_{32}^*}{k_{12}^* + k_{32}^*} \frac{\phi}{\lambda f} C_1
\]

where \( f = \frac{k_{32}^*}{k_{32}} \); \( \lambda = \frac{k_{21}^* / (k_{12}^* + k_{32}^*)}{k_{21} / (k_{12}^* + k_{32}^*)} \)

Since 2-deoxyglucose and glucose are competitive substrates for hexokinase in the phosphorylation process and their rates follow the Michaelis-Menten relation:

\[
k_{32} = \frac{V_m / K_m}{1 + (C_2 / K_m) + (C_2^* / K_m)} \quad ; \quad k_{32}^* = \frac{V_{m}^* / K_{m}^*}{1 + (C_2 / K_m) + (C_2^* / K_m)}
\]

Thus:

\[
f = \frac{V_m / K_m}{V_m^* / K_m^*}
\]

and:

\[
\frac{\lambda f}{\phi} = \frac{\lambda}{\phi} \frac{V_m K_m}{V_m^* K_{m}^*} = LC, \text{ the lumped constant.}
\]

where \( \lambda \) is the ratio of the distribution volumes for 2-deoxyglucose and glucose in tissue, and \( \phi \) represents the fraction of glucose that is metabolized after it is phosphorylated.

Figure 6.4: (continued)

Derivation of the lumped constant.
Therefore:

\[ \text{L}_{\text{CMRg}} = \frac{C_1}{L \text{C}} \frac{k_{21}^* \cdot k_{32}^*}{k_{12}^* + k_{32}^*} \]

Multiplying the numerator and denominator of this expression by \( C_3^*(t) = C_1^*(t) - C_2^*(t) \), and replacing \( C_3^*(t) \) and \( C_2^*(t) \) by their respective expression, the equation becomes:

\[
\text{L}_{\text{CMRg}} = \frac{C_1(T)}{L \text{C}} \frac{C_1^*(t) - \left[ \frac{k_{21}^*}{a_2 - a_1} \right] \left[ (k_{23}^* - a_1^*) e^{-a_1^* t} + (a_2^* - k_{23}^*) e^{-a_2^* t} \right] * C_1^*(t)}{\left[ \frac{k_{12}^* + k_{31}^*}{a_2 - a_1} \right] e^{-a_1^* t} - e^{-a_2^* t} * C_1^*(t)}
\]

Thus, \( \text{L}_{\text{CMRg}} \) can be estimated from the measured quantities \( C_1 \), the plasma glucose concentration, \( C_1^*(t) \), the time course of 2-deoxyglucose in the plasma, \( C_1^*(t) \), the time course of the label in any pixel of the tomographic image, and published values for the rates constants \( k_{21}^*, k_{12}^*, k_{32}^* \) and \( k_{23}^* \) and for the lumped constant.

Figure 6.4: (continued)
An expression for the calculation of local cerebral metabolic rates of glucose.
The first compartment represents glucose in the plasma, whereas compartment II represents metabolic precursors in tissue. The boundary between compartments I and II is the capillary membrane. Compartment III represents metabolic products, in the case of 2-deoxyglucose, mostly 2-deoxyglucose-6-phosphate. The boundary between compartments II and III is not a physical barrier, but the phosphorylation reaction catalysed by hexokinase.

The rate constants linking compartments I and II represent the forward and reverse transport of 2-deoxyglucose across the capillary membrane. The rates constants linking compartments II and III represent the rate of phosphorylation of 2-deoxyglucose and dephosphorylation of 2-deoxyglucose-6-phosphate respectively. The rate of dephosphorylation of 2-deoxyglucose-6-phosphate is the key to the successful application of the method. If it is very slow, then the requirement for metabolic trapping would be achieved. In the first demonstration of the method, Sokoloff ignored this backflux, and in a refinement of the method, Phelps et al considered it, but found it to be significant only in very long studies. However, more recently, this assumption has come under considerable criticism.

The numerator of the operational equation (fig. 6.5) represents the amount of radioactive product formed in a given interval of time; it is equal to the combined
rate of reaction = \left[ \frac{\text{labelled product formed in time interval (0 to T)}}{\text{isotope effect correction factor \times \text{integrated specific activity of precursor}}} \right]

= \left[ \frac{\text{total activity in tissue at time T}}{\text{precursor remaining in tissue at time T}}} \right] - \left[ \text{isotope effect \times \text{integrated plasma specific activity \times correction for lag in tissue \times equilibration with plasma}} \right]

\[ C_i(t) = k_{21} e^{-(k_{12}^* + k_{32}^*)T} \int_0^T C_p^* e^{(k_{12}^* + k_{32}^*)t} \, dt \]

\[ L_{CMRg} = \frac{\lambda V_m^* K_m}{\phi V_m^* K_m^*} \left[ \int_0^T \frac{C_p^*}{C_p} \, dt - e^{-(k_{12}^* + k_{32}^*)T} \int_0^T \frac{C_p^*}{C_p} e^{(k_{12}^* + k_{32}^*)t} \, dt \right] \]

Figure 6.5: The operational equation for the calculation of regional cerebral metabolic rates of glucose. The rate of dephosphorylation of 2-deoxyglucose-6-phosphate has been ignored.
concentrations of 2-deoxyglucose and 2-deoxyglucose-6-phosphate in the tissue at time T, less a term that represents the free, unmetabolized, 2-deoxyglucose still remaining in the tissue. The denominator represents the integrated specific activity of the precursor pool times a factor that is equivalent to a correction factor for an isotope effect. The term with the exponential factor in the denominator takes into account the lag in the equilibration of the tissue precursor pool with that of plasma.

6.3.2.3. Discussion

By the use of labelled 2-deoxyglucose as a probe, a single biochemical reaction, which is the first step in the pathway of glucose metabolism, has been isolated. This step is the hexokinase-catalysed phosphorylation of glucose. The total amount of radioactive product formed and the integrated specific activity of the precursor at the enzyme site can be determined. From these data and the use of a correction factor for the difference in the kinetic behaviour of 2-deoxyglucose and glucose, the net rate of glucose phosphorylation can be calculated by the operational equation. In a steady state, the net rate through any step in a pathway equals the net rate through the overall pathway. The deoxyglucose method therefore measures the net rate of glucose phosphorylation in vivo and the net rate of the entire pathway of oxidative glucose metabolism in a steady state.
The 2-fluoro-2-deoxyglucose method has been used to study normal brain function. It has been shown that predictable changes in energy consumption of the occipital lobe could be elicited in response to simple and complex visual stimuli. In response to a reduction in auditory as well as visual sensory stimuli inputs, a reduction in the metabolism of the parietal and occipital cortices as well as a left/right asymmetry in the distribution of the label were observed. A decline in cerebral glucose metabolism was demonstrated in older subjects.

The method has also been applied to the study of a variety of disease states such as epilepsy, Huntington's chorea, Parkinson's disease and schizophrenia. The results have shown a consistent pattern of abnormal glucose metabolism characteristic of the disease.

At McMaster University, we have repeated most of these studies. Furthermore, we have demonstrated that a reduction in striatal glucose consumption and a prolonged time to initiate a movement are early features in Huntington's disease. We are using these findings in the study of individuals at risk for this disease. We have studied the effects of electrical stimulation of the thalamus on cerebral glucose consumption; these stimulations are performed in an effort to control epilepsy. We have also studied the effects of electrical stimulations of the
cerebellum that are performed to alleviate the symptoms of cerebral palsy. We are comparing the patterns of glucose metabolism of a number of young schizophrenics who have never been medicated with those of aged matched, psychologically normal, controls.

A typical study starts with the iv injection into a subject of about 50 μCi/kg of $[^{18}\text{F}] 2$-fluoro-2-deoxyglucose. While the subject lies with his eyes shut and the background noise kept to a minimum, blood samples are collected at various time intervals over the first 45 minutes following the administration of the label to define its fate in the blood. Fifty minutes after injection, the regional concentration of $[^{18}\text{F}]$ is measured in sequential slices of the brain. The slices are usually horizontal and parallel to a plane 5° above the orbito-meatal line. Each slice overlaps its predecessor by 5 mm and a series of sixteen slices are routinely obtained (fig. 6.6). More than one million counts are accumulated for each slice and an examination usually lasts for about 45 minutes. The values for the regional cerebral metabolic rate of glucose are derived from the knowledge of the time course of the label in the blood, the plasma glucose concentration and the number of counts in each pixel of the tomographic image, using the operational equation and published values for the different constants.
Figure 6.6: Tomographic slices of normal brain from a single individual obtained one hour after the intravenous injection of $^{[18}F\) 2-fluoro-2-deoxyglucose. Each slice is separated from its predecessor by 5 mm. All the figures are colour coded with purple representing the highest local cerebral metabolic rate of glucose, followed by the reds, yellows, greens and blues. [10 mm slice; reconstruction filter: Hamming 0.5]. The regions that can be identified in this set of figures are the cerebellum, the temporal lobes, and the hippocampal gyri.
Figure 6.6: (continued)

The region of the thalami and the basal ganglia.
Figure 6.6: (continued)

The region of the corpus callosum.
Figure 6.6: (continued)
The region of the cingulate gyrus and the top of the brain.
It can be appreciated from figure 6.6 that the spatial resolution of the tomograph allows for the delineation of many important structures in the adult brain. We have used the tomograph to study the regional cerebral glucose metabolism of newborn infants weighing as little as 750 g and have also found the spatial resolution to be adequate for the positive identification of structures such as the thalami and the various gyri.

The counting efficiency of the tomograph has allowed us to perform studies in newborn infants using individual doses of less than 100 μCi. A study in an adult was actually successfully performed after the administration of 0.450 mCi to an 82 kg woman.

6.4. The measurement of intracerebral neurotransmitter metabolism

6.4.1. Introduction

Nerve cells, like those of other body tissues, carry on many metabolic activities: they generate energy, mainly from glucose, using aerobic metabolic pathways and they synthesize proteins and lipids that make up the structural component of the cell. All these activities are important for maintaining the integrity of the cell; but these metabolic processes are only indirectly involved in performing the major function of the nervous system, that of transmitting and processing information.
A neuron is composed of a main body, the soma, a single axon which may branch more or less profusely and conducts away from the cell body, and many dendrites which are thin projections of the soma and conduct towards the cell body. Neurons interact with each other at junctions called synapses; the communication across most synapses is mediated through chemical transmitters. A neuron will usually synthesize, store within vesicles, and release only one type of transmitter, but different transmitters may act on one cell.

Along a neuron, the information is transmitted in the form of electricity, the action potential, an all-or-none change in the membrane potential of the body or axon of a neuron. When a presynaptic action potential reaches the synaptic region, the depolarization of the nerve terminal causes the release of neurotransmitters from their storage vesicles into the synaptic cleft in discrete quantities directly proportional to the frequency of the action potential. These released transmitters reach, are recognized by, and bind to specific receptors on the surface of a postsynaptic neuron, thereby generating a synaptic potential at a particular membrane locus. This synaptic potential is a local and graded electrical response to the chemical stimulus that is proportional to the amount of transmitter released. These synaptic potential changes produce, in turn,
electrotonic potentials, a depolarization affecting the surrounding zones of the membrane with a decreasing amplitude at increasing distances from the initial locus. Thus, the larger the initial depolarization, the greater the spread of the effect to surrounding areas of the membrane. This mechanism allows the summation of effects of multiple excitatory stimuli applied at numerous points on the receiving surface of the neuron. If this potential reaches a critical threshold, the whole of the surrounding membrane will depolarize and a postsynaptic action potential will be generated. The electrotonic postsynaptic potentials can be either excitatory or inhibitory. An excitatory stimulus will result in the depolarization of the membrane whereas an inhibitory stimulus will cause the hyperpolarization of the membrane. The elicited effect will depend on the nature of the postsynaptic cell's receptor for the particular transmitter molecule released.

About thirty neurotransmitters have been identified so far. These are acetylcholine, serotonin, catecholamines, amino acid transmitters, neuroactive peptides, and others. The neurophysiological activity of acetylcholine has been known since the turn of the century and its neurotransmitter role since the mid-1920s. It is found in many areas of the brain and has an excitatory effect. Serotonin (5-hydroxytryptamine) has also been known for a relatively long time and is found in neurons that project to many brain areas.
It is thought to act as an inhibitor of pain pathways in the spinal cord and it is also believed to help in the control of mood and sleep. The catecholamines are a class of neurotransmitters that comprises dopamine, norepinephrine and epinephrine. Dopamine is mostly found in neurons that originate in the substantia nigra and end in the striatum. Its effect is usually inhibitory. Norepinephrine is found in many neurons whose cell bodies are located especially in the locus caeruleus and also in the hypothalamus. These neurons send nerve fibers to widespread areas of the brain and probably have an inhibitory effect. Epinephrine is found in a small number of neurons whose distribution parallels the adrenergic system. One of the amino acid transmitters is gamma aminobutyric acid (GABA). It is present in large amounts in specific regions of the central nervous system such as the spinal cord, the cerebellum and the basal ganglia. It has an inhibitory effect. Neuroactive peptides form a large class of molecules some of which act as proper neurotransmitters while others perform modulatory or regulatory roles that do not fit easily into the conceptual mold of a transmitter. Substance P is one such peptide, which is probably released by pain fiber terminals in the spinal cord.

Specific pathways in the brain, such as the reciprocal connections in spinal cord, the thalamus, the striatum and
the motor cortex employ a variety of neurotransmitters to control the flow of information (fig. 6.7). Although only a few of the chemical transmitters of the brain neurons have been fully characterized so far, their role in behaviour, mood and learning, and in diseases such as depression, schizophrenia and motor disorders is beginning to evolve. From a practical point of view, dopamine is a neurotransmitter which is well localized in only a few areas of the brain and which plays a central role in the control of mood and locomotion, making it the neurotransmitter of choice to be studied in man, during life, by positron tomography.

6.4.2. The fluorodopa method

6.4.2.1. Rationale of the method

Dopamine is a major neurotransmitter in the mammalian central nervous system. The bulk of the intracerebral dopamine is contained in the nigro-striatal pathway with lesser quantities in the mesolimbic and mesocortical projections. Disturbances of dopamine metabolism are a key feature of locomotor disorders such as Parkinson's disease and reduced striatal dopamine concentration has recently been described in progressive supranuclear palsy. Central dopaminergic pathways may also malfunction in schizophrenia even to the extent that the plasma concentration of the major metabolite of dopamine, homovanillic acid, is elevated in that condition.
Neurotransmitter:
- dopamine
- GABA
- glutamate
- acetylcholine
- not known

Figure 6.7: Diagram of some of the main connections of the basal ganglia.
Dopamine is formed within neurons by the action of L-aromatic acid decarboxylase on its immediate precursor, dihydroxyphenylalanine (dopa) (fig. 6.8). In life, dopa is formed from tyrosine by the action of tyrosine hydroxylase, the rate limiting step in dopamine synthesis. Aromatic acid decarboxylase is normally present in excess and consequently only insignificant quantities of dopa itself can be found in either the brain, the cerebrospinal fluid or the blood.

Dopamine is stored in intraneuronal vesicles where it can be demonstrated by immunohistofluorescence techniques. Ordinarily, low concentrations of dopamine remain free in and unprotected in the cytosol. The vesicles play a dual role: they maintain a ready supply of dopamine at the terminal, and they mediate the process of their release. A cytoplasmic pool of dopamine that turns over rapidly has also been suggested.

When a dopaminergic neuron is stimulated electrically, calcium channels open, allowing an influx of the cation into the terminal; increased intracellular Ca$^{++}$ promotes the fusion of vesicles near the synaptic specialization with the neuronal membrane. The vesicles then discharge dopamine into the extraneuronal space, where it can bind reversibly with its specific receptor on the post synaptic neuron.
Two enzymes are primarily responsible for the catabolic inactivation of dopamine: monoamine oxidase and catechol-0-methyl transferase. Monoamine oxidase oxidatively deaminates catecholamines to the corresponding aldehydes; in turn, these can be converted by aldehyde dehydrogenase to the analogous acids. Thus, dopamine is deaminated to its corresponding aldehyde which, in turn, is converted to 3,4 dihydroxyphenylacetic acid. This acid, like dopamine, is a substrate for catechol-0-methyl transferase and therefore can be converted to homovanilic acid (fig. 6.8). Because of its intracellular localization, monoamine oxidase plays a strategic role in inactivating catecholamines that are free within the nerve terminal and not protected by storage vesicles. Catechol-0-methyl transferase transfers a methyl group to the 3-hydroxy group on the catechol ring, forming for example 3-methoxytyramine from dopamine. In turn, 3-methoxytyramine can be oxidized to homovanilic acid by the action of monoamine oxidase and aldehyde dehydrogenase (fig. 6.8). Catechol-0-methyl transferase is located primarily on the outer plasma membrane of nearly all cells, including erythrocytes; thus, this enzyme acts on neuronal as well as extraneuronal catecholamines.

Some of the dopamine that is released into the extraneuronal space is deactivated extracellularly by the action of catechol-0-methyl transferase. The bulk of the
Figure 6.8: Pathways of dopamine biosynthesis and metabolism.
synaptic dopamine is reaccumulated by the presynaptic terminal and some of this material is deactivated by the action of intracellular monoamine oxidase. The exact proportions of dopamine that are deactivated through catechol-0-methyl transferase and through monoamine oxidase are not known. However, homovanilic acid, the consequence of deamination and methylation, is the major metabolite of dopamine found in the brain of primates at autopsy.

6.4.2.2. **Compartmental analysis**

An explanatory model of intracerebral dopamine should, therefore, have compartments that represent tyrosine, dopa, dopamine and its metabolites. The dopamine compartment would be further subdivided into stored dopamine and dopamine that is rapidly metabolized. The metabolic compartment would account for dihydroxyphenylacetic acid as well as for homovanilic acid.

Because $^{18}$F fluorodopa is used as the tracer for dopamine metabolism, we do not consider a compartment for tyrosine. Consequently, the model does not take into consideration the natural rate limiting step. It does, however, provide information about the formation, storage and degradation of dopamine itself (fig. 6.9).

The first compartment in the model is a dopamine precursor pool. Anatomically, this pool probably comprises endothelial cytosol and cerebrospinal fluid. It is separated
Figure 6.9: Compartamental model for the fluorodopa method.

<table>
<thead>
<tr>
<th>PLASMA</th>
<th>BRAIN TISSUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>dopa ← → dopa —→ dopamine —→ (storage)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>↓</td>
</tr>
<tr>
<td></td>
<td>(degradation)</td>
</tr>
</tbody>
</table>
from blood at the luminal surface of the endothelial cells of the brain capillaries. Dopa and fluorodopa cross this membrane with equal facility. The second compartment should represent newly formed dopamine that is destined to be stored, or metabolized further. However, because the capacity of the aromatic acid decarboxylase system is large, it is difficult to separate mathematically the precursor compartment from that representing newly formed dopamine. Consequently, compartment I comprises all dopa and dopamine that is not protected in the storage vesicles. Fluorodopa is a substrate for aromatic acid decarboxylase with kinetics similar to those of native dopa.

Compartment II represents stored dopamine. Evidence for the storage of fluorodopamine comes from the observation that when fluorodopa is given intravenously, striatal $^{18}$F radioactivity can be discharged by giving reserpine. Compartment III represents homovanilic acid. In primates, the product of dopamine metabolism is almost exclusively all unconjugated homovanilic acid. This simplification would not be tenable in lower orders where dihydroxyphenylacetic acid makes up a sizable fraction of the metabolites and would therefore have to be considered as a separate compartment.

Compartment I is fed from the blood. Dopa is actively transported across the blood brain barrier along
with the other large neutral amino acids. Circumstantial evidence also suggest that there is a backflux of amino acids from interstitial fluid to blood. The rate constant to compartment II from compartment I represents the speed with which a fraction of newly formed dopamine is sequestered in the storage vesicles. Compartment III is fed from the pool of newly formed dopamine (compartment I) and empties according to a rate constant that represents the rate of removal of homovanillic acid from the field of view of the tomograph. Over 90% of all homovanillic acid formed in the brain is removed through the circulation; it does not travel through the cerebrospinal fluid.

Ideally, the feed to the model would be given by the time activity curve of fluorodopa in the blood measured in the intracerebral capillaries. The single pass extraction of fluorodopa through the brain is relatively low and therefore it is reasonable to use arterial blood to approximate capillary blood. At the onset, all the $^{18}$F activity in the blood is associated with fluorodopa. This may not be true at later times because red blood cells, liver and peripheral muscles all contain catechol-0-methyl transferase so that some $^{18}$F activity will be associated with fluromethyldopa. Fluorohomovanillic acid will also be present in the blood. However, if the methylated fractions are small, they can be neglected in the mathematical analysis (fig. 6.10).
Rate of change of free dopa and dopamine in tissue:

\[
\frac{dC_1}{dt} = k_{10} C_p - (k_{01} + k_{21} + k_{31}) C_1 \quad ; \quad C_1(0) = 0
\]

Rate of accumulation of dopamine in storage vesicles:

\[
\frac{dC_2}{dt} = k_{21} C_1 \quad ; \quad C_2(0) = 0
\]

Rate of change of metabolic products in tissue:

\[
\frac{dC_3}{dt} = k_{31} C_1 - k_{03} C_3 \quad ; \quad C_3(0) = 0
\]

---

**Figure 6.10**: Mathematical analysis of the fluorodopa model. The differential equations.
Amount of free tracer in tissue:

\[ C_1(t) = k_{10} \left[ e^{-\left(k_{01} + k_{21} + k_{31}\right)t} \right] \ast C_p(t) \]

Amount of stored tracer in vesicles:

\[ C_2(t) = k_{10} \frac{k_{31}}{k_{01} + k_{21} + k_{31}} \left[ 1 - e^{-\left(k_{01} + k_{21} + k_{31}\right)t} \right] \ast C_p(t) \]

Amount of metabolized tracer remaining in tissue:

\[ C_3(t) = k_{10} \frac{k_{31}}{k_{01} + k_{21} + k_{31} - k_{03}} \left[ e^{-k_{03}t} - e^{-\left(k_{01} + k_{21} + k_{31}\right)t} \right] \ast C_p(t) \]

Total amount of tracer in tissue:

\[ C_i(t) = C_1(t) + C_2(t) + C_3(t) \]

\[ = k_{10} \left[ 1 - \frac{k_{31}}{\beta} + \frac{k_{21}}{\alpha} \right] e^{-\alpha t} + \left[ \frac{k_{31}}{\beta} e^{-k_{03}t} \right] \ast C_p(t) \]

where: \( \alpha = k_{01} + k_{21} + k_{31} \); \( \beta = k_{01} + k_{21} + k_{31} - k_{03} \).

Figure 6.10: (continued)
The response of the model.
where $C_i(t)$ represents the time course of $^{18}$F in any pixel of the tomographic image, $C_p(t)$ the time course of $^{18}$F in the plasma and $C_r(t)$ the unit impulse response of the model.

The result of the convolution of the unit impulse response with the plasma feed is compared to the observed response, and the parameters $k_{01}, k_{10}, k_{21}, k_{31}$ and $k_{03}$ are adjusted so as to minimize the square of the differences between the two curves.

Figure 6.10: (continued)
Strategy to arrive at the rates of formation, storage and degradation of dopamine.
6.4.2.3. Discussion

By the use of labelled dopa, the distribution of a neurotransmitter in the brain has been obtained. Unlike 2-deoxyglucose, fluorodopa, or its immediate product, is not metabolically trapped in the cerebral tissues, so that a simple operational equation cannot be used to derive rates of dopamine metabolism. However, a compartmental analysis can be applied to the data and rate constants derived. These rate constants, determining the rates of storage and degradation of dopamine, may prove useful in the study of normal and diseased states.

The fluorodopa method has been used to examine the distribution of dopamine in normal subjects and in subjects suffering from a variety of motor disorders such as Huntington's chorea and dystonia. The most striking findings are in the study of patients suffering from Parkinson's disease in whom the abnormality is clearly shown to be confined to the putamen, at least in the early stages of the disease (fig. 6.15). The determination of the rates of formation, storage and degradation of dopamine is also being carried out in the study of parkinsonian patients who no longer respond to therapy. The rate constants determined before and after such patients follow a drug holiday are providing insight into the effects of the treatment and its mode of action.
Figure 6.11: Observed rate of change of $^{18}$F activity in striatal region (open circles). The solid curve represents the line of best fit resulting from the convolution of the plasma activity curve with the unit impulse response of the fluorodopa model.

$k_{10} = 0.0293 \text{ min}^{-1}$
$k_{01} = 0.103 \text{ min}^{-1}$
$k_{21} = 0.012 \text{ min}^{-1}$
$k_{31} = 0.103 \text{ min}^{-1}$
$k_{03} = 0.08 \text{ min}^{-1}$
Figure 6.11: (continued)
Observed rate of change of $^{18}$F activity in occipital region (open circles) and line of best fit. Note that the rate constants representing stored, or metabolized, dopamine are equal to zero. This is in keeping with the notion that there are no dopaminergic projections to the occipital lobe.
A typical study starts with the IV injection into a subject of 50 μCi/kg of $^{18}$F fluorodopa. Arterial blood samples are collected at various time intervals over the first ninety minutes following the administration of the label. During that time, the subject lies in the tomograph and the regional concentration of $^{18}$F is measured in a slice of brain that contains the basal ganglia. Forty-five sequential images, each of two minutes duration, are obtained in order to determine the time course of the label in the slice (fig. 6.11). At the end of the dynamic portion of the study, sequential slices are obtained, each slice overlapping its predecessor by 5 mm. The complete mapping of the region of the basal ganglia usually requires five slices and additional slices may be obtained, depending on the patience of the subject (fig. 6.12). For the comfort of some sick patients, the dynamic study is not performed and only the static images, depicting the distribution of the label in the slice, are obtained.
Figure 6.12: Tomographic slices of normal brain from a single individual obtained one hour after the intravenous injection of $[^{18}\text{F}]$ fluorodopa. All the figures are colour coded with purple representing the most intense accumulation of $^{18}\text{F}$, followed by the reds, yellows, greens and blues. The most peripheral activity in the figures is due to radioactivity in the temporalis muscles.
6.5. Correlation between CT, NMR and PT findings in the brain

6.5.1. Introduction

Three tools are presently available with which to study the intact brain. The first, x-ray computerized tomography, is restricted to the differentiation of tissues according to their mass attenuation coefficients. The second, nuclear magnetic resonance, also differentiates between tissues, but according to their proton content and to the environment surrounding these protons. The detail available in an NMR study is greater than that available from a CT study; although both provide the same clinical information: specific tissue characterization. The third, positron tomography, does not characterize tissue at all; it provides a map of the distribution of a labelled molecule in the cerebral tissues. It can also provide the time course of this label in any region of the brain. It therefore depicts functional anatomy of the brain. However, the three techniques are complementary and information from any one can be used to augment and clarify the interpretation of the others (fig. 6.13). The complementary metabolic information obtained using different positron emitting agents and positron tomography is invaluable for the correct understanding of the diseased brain.
Figure 6.13: The normal brain: horizontal section obtained at the level of the basal ganglia.
(a) CT scan
(b) NMR scan
(c) PT scan ([¹⁸F] 2-fluoro-2-deoxyglucose)
(d) PT scan ([¹⁸F] fluorodopa)
6.5.2. Examination of the nigro-striatal pathway

The caudate nucleus and putamen (the striatum) are large subcortical masses innervated in part by dopamine containing nerve endings that originate in the substantia nigra. In the striatum, the dopaminergic input system and the intrinsic cholinergic neurones are functionally interdependent. Normal functioning of the striatum as a whole is determined by the proper balance between these two systems; elimination of the nigral dopamine input to the striatum effectively disrupts normal function. The function of the caudate nucleus may be associated with behaviour, while the putamen is predominantly associated with movement.

Huntington's disease is an inherited, degenerative progressive disorder characterized by behavioural changes and choreiform movements. At autopsy, there is a marked loss of neurons in the caudate nucleus accompanied by some loss of glia. This loss of tissue appears on CT as an enlargement of the lateral ventricles and is evident in the later stages of the disease. In the early stages of the disease, the CT scan is essentially normal and therefore not helpful in making a diagnosis (fig. 6.14). However, the PT study done with $^{18}\text{F}$ 2-fluoro-2-deoxyglucose shows that striatal glucose consumption is markedly reduced both in the caudate and putamen (fig. 6.14). In contrast, the PT study done with $^{18}\text{F}$ fluorodopa shows an obvious accumulation.
of the label in the region of the striatum, the caudate and putamen accumulating the tracer to the same extent (fig. 6.14). These findings indicate that a profound reduction in striatal energy consumption is an early feature of Huntington's disease that antedates any definitive changes in CT. Our findings also suggest that the dopaminergic pathway continues to function even though the neurons with which this pathway synapses no longer consume glucose.

Parkinson's disease is also a degenerative, progressive disease characterized by an interference with all aspects of voluntary motor performance. The CT scan of patients suffering from Parkinson's disease is essentially normal (fig. 6.15). The PT study done with $^{18}$F 2-fluoro-2-deoxyglucose shows that glucose metabolism in the striatum is normal (fig. 6.15), at least in the early stages of the disease. However, the PT study done with $^{18}$F flurodopa shows that dopamine accumulation is markedly reduced in the region of the putamen contralateral to the side of the body exhibiting the signs of the disease (fig. 6.15). In this case, the neurons of the striatum are normal with respect to glucose consumption; however, they lack the input of the essential neurotransmitter, dopamine.

Dystonia is a disease marked by involuntary, irregular contortions of the muscles of the trunk and of the extremities. Again, the CT is usually not helpful in
Figure 6.14: CT scan (a), $[^{18}F]$ 2-fluoro-2-deoxyglucose scan (b) and $[^{18}F]$ fluorodopa scan (c) of a young patient suffering from Huntington's disease.

Figure 6.15: CT scan (a), $[^{18}F]$ 2-fluoro-2-deoxyglucose scan (b) and $[^{18}F]$ fluorodopa scan (c) of a patient suffering from unilateral Parkinson's disease.

Figure 6.16: CT scan (a), $[^{18}F]$ 2-fluoro-2-deoxyglucose scan (b) and $[^{18}F]$ fluorodopa scan (c) of a young patient suffering from secondary dystonia.
establishing a diagnosis (fig. 6.16). However, in secondary
dystonia, both the accumulation of $^{18}F$ 2-fluoro-2-deoxy-
glucose and of $^{18}F$ fluorodopa is markedly reduced in the
region of the striatum, predominantly in the putamen
(fig. 6.16).

6.5.3. Examination of the hypoxic brain

Hypoxia of the brain may result from a lack of oxygen in the blood delivered to the brain, or from a disruption in the circulation that supplies the brain with oxygen. The consequences of this lack of oxygen can be an interference with the normal activity of the brain, or a complete loss of function in a region of the brain.

We have used CT and PT performed using $^{18}F$ 2-fluoro-
2-deoxyglucose to study a group of infants who have suffered some degree of asphyxia at birth. In these newborn infants, there is a lack of information about external landmarks for the structures in the brain. A CT scan performed before the PT examination has allowed us to position the baby more properly. Our results show that, in general, the region of the brain that is abnormal on CT also shows abnormal glucose consumption compared to the same region on the opposite side of the brain (fig. 6.17). It is notable that the area of metabolic impairement demonstrated on PT extends considerably beyond the anatomical area of the abnormality seen on the CT scan.
Figure 6.17: CT scan (a) and $[^{18}\text{F}]$ 2-fluoro-2-deoxyglucose scan (b) of a six day old infant.

Figure 6.18: NMR scan (a) and $[^{18}\text{F}]$ 2-fluoro-2-deoxyglucose scan (b) of a patient who had suffered a large frontoparietal infarct five years prior to the studies.
We have used NMR and PT performed using $^{18}_F$ 2-fluoro-2-deoxyglucose to study a group of patients with cerebral infarction. In this group of patients, the region of metabolically inactive tissue as demonstrated by PT is more extensive than the anatomical lesion as depicted by NMR (fig. 6.18).

6.5.4. Discussion

In general, we have found the ability to correlate the morphological information available from a CT or an NMR study and the metabolic information available from a PT study to be invaluable in the investigation of normal and disease processes in the living brain. To be able to correlate the findings obtained by CT, NMR and PT the position at which the head is imaged must be reproduced using each technique. Similarly, the level at which each slice is obtained must be the same. Ideally, the proportions of the image of the brain obtained by each technique should be matched to allow easy superposition of the various images. Although some of these problems have not yet been satisfactorily resolved, the correlation between the three imaging modalities is proving to be very useful in providing a better understanding of the working brain in health and disease.
6.6. Conclusions

Positron tomography is providing unique new information on the biochemistry and physiology of the human nervous system in health and disease. The possibility of quantitatively following the metabolic kinetics of a labelled element or molecule associated with the function of an organ allows the determination of the functional state of that organ along with the consequences that this implies from a diagnostic and prognostic point of view. For example, positron tomography has been used as a predictor for the success, or failure, of external carotid/internal carotid by-pass surgery.

As in any imaging system, two parameters are central in the design of experiments and in the correct interpretation of the results. These are temporal resolution, or how fast can an image of good quality, in a statistical sense, be accumulated; and spatial resolution, or what are the dimensions of the smallest structure that can be unequivocally imaged.

A number of variable factors related to the regional, functional integrity of the brain, such as blood flow, metabolism and blood volume change rapidly as a function of time, probably on a time scale appreciably shorter than one minute. To investigate such variables by positron
tomography, methods that mirror these events in the time dimension as closely as is practical should be used. Yet, many positron tomographical examinations of the brain, including glucose metabolic studies carried out using $^{18}$F 2-fluro-2-deoxyglucose and oxygen metabolic studies carried out using the equilibrium inhalation technique are performed with data acquisition times lasting from many minutes to more than one hour. Such prolonged data acquisition times are dictated in part by the tracer model. For example, in the application of the 2-deoxyglucose method to the measurement of glucose metabolism, it takes about 45 minutes for the precursor pool concentration to fall to negligible levels. In other applications, such as the use of $^{18}$F fluoro-dopa to measure dopamine metabolism, the fraction of the injected dose that enters the brain and is retained is so small that long acquisition times, of the order of five minutes per image, are required to accumulate enough counts to produce a statistically satisfactory image. Prolonged data acquisition times are also dictated by the design features of the tomograph. For example, the introduction of bismuth germanate as a scintillator and the move from hexagonal arrays of detectors to circular arrays of detectors have shortened considerably the acquisition times; the introduction of caesium fluoride as a scintillator has also helped to shorten acquisition times through the use of higher
doses of radiation.

The brain is a heterogeneous organ in which most anatomical structures are small, of irregular shape and lie adjacent to each other. All imaging devices experience quantitative measurement inaccuracies as the object size is reduced below the spatial resolution of the imaging device. In this respect, it is important to differentiate signal estimation from signal detection. In signal estimation, the aim is to recover the true concentration of isotope in all structures, irrespective of their size and location in the image plane; in signal detection, the aim is to accurately measure the distribution of isotope in the plane being examined. Whereas signal estimation is a measurement independent of the tomograph, signal detection is a measurement highly dependent on the three dimensional spatial response of the tomograph used. Absolute quantitation, signal estimation, is unattainable in tomography. It is inherently a three dimensional problem, a problem that tomography almost exclusively considers in two dimensions. Relative quantitation between homologous structures may also be in error and might become more inaccurate with improved resolution in the transverse plane without a concomittent improvement in axial resolution.

These temporal and spatial limitations imposed on positron tomography must be recognized. Eventhough it is
not impossible, in theory, to design a tomograph with a resolution of about 2 mm in the three spatial dimensions and recognizing that the limitations in the temporal dimension only arise from statistical considerations, the major constraint remains the radiation dose that can be safely administered, especially if repeat studies in the same subject are contemplated. Hence, positron tomography is restricted to the study of processes that vary relatively slowly with time. Furthermore, the tomographic images must not be interpreted as providing absolute quantitative data. Nevertheless, positron tomography is providing new, although maybe imperfect, information about the working brain in health and disease.

Toward these goals, the tomograph that has been proposed and built during the course of this thesis is proving adequate. Its spatial resolution of about 8 mm in the transverse and axial planes is not at the theoretical limit imposed by the positron range, but is sufficient to image with confidence the important central grey structures of the brain. Of more consequence, its sensitivity is the highest that can be attained with present day technology, allowing for a reduction in the radiation dose that is given to subjects, and the acquisition of images with better counting statistics. The introduction of additional arrays of detectors to image more than one plane simultaneously
would obviously be an advantage, presenting no conceptual
difficulties. Any improvement in the design would come
from the discovery of a scintillation detector with a
stopping power higher than that of bismuth germanate.
| BIBLIOGRAPHY |
|--------------|-------------|
| Chapter 1    | 245         |
| Chapter 2    | 245         |
| Chapter 3    | 246         |
| Chapter 4    | 250         |
| Chapter 5    | 253         |
| Chapter 6    | 254         |
Chapter 1: Organization of the brain.


Chapter 2: Positron emission and energy losses.


Chapter 3: Detectors for positron tomography.

General:


Multiwire proportional counters:


Solid state detectors:


Scintillation detectors:

Sodium Iodide (thallium):


Caesium fluoride:


Barium Fluoride:


Bismuth Germanate:


Chapter 4: Principles of image reconstruction from projections.


Chapter 5: Design and construction of the tomograph.


Chapter 6: The study of regional cerebral metabolism during life.


