STATISTICAL LIST-MODE IMAGE RECONSTRUCTION AND MOTION COMPENSATION TECHNIQUES IN HIGH-RESOLUTION POSITRON EMISSION TOMOGRAPHY (PET)

by

ARMAN RAHMIM

B.Sc. (Comb Hons), The University of British Columbia, 1999 M.Sc., The University of British Columbia, 2001

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

(Physics)

THE UNIVERSITY OF BRITISH COLUMBIA

April 2005

© Arman Rahmim, 2005

Abstract

The work presented here is devoted to the proposal and investigation of 3D image reconstruction algorithms suitable for high resolution positron emission tomography (PET). In particular, we have studied imaging techniques applicable to the high resolution research tomograph (HRRT): a 3D-only state-of-the-art dedicated brain tomograph. The HRRT poses a number of unique challenges, most significant of which include presence of gaps in-between the detector heads, as well as the very large number of lines-of-response (LORs) which it is able to measure ($\sim 4.5 \times 10^9$), exceeding most modern PET scanners by 2-3 orders of magnitude.

To address the existing issues, we have developed and implemented *statistical list-mode* image reconstruction as a powerful technique applicable to the high resolution data obtained by the HRRT. We have furthermore verified applicability of this technique to dynamic (4D) PET imaging, thus qualifying the technique as viable and accurate for the research intended to be performed on the scanner. We have paid particular attention to the study of *convergent* algorithms; i.e. iterative algorithms which (with further iterations) consistently improve such figures of merit as resolution and contrast, relevant to research and clinical tasks.

With the spatial resolution in modern high resolution tomographs (including the HRRT) reaching the 2-3mm FWHM range, small patient movements during PET imaging can become a significant source of resolution degradation. We have thus devoted a portion of this dissertation to the proposal of new, accurate and practical motion-compensation techniques, and Abstract

studied them on the HRRT. We have theoretically proposed and experimentally validated the benefits of modeling the motion into the reconstruction task, thus signaling the way beyond the existing purely event-driven motioncompensation techniques.

Table of Contents

Abstra	ct		ii
Table o	of Conte	entsi	v
List of	Tables	and Figures	Х
Glossa	ry	· · · · · · · · · · · · · · · · · · ·	x
Acknow	vledgem	nents	i
1. Intr	oductio	on to PET Imaging	1
1.1	Nuclea	ar Medicine Imaging	1
	1.1.1	Single Photon Imaging	1
	1.1.2	Positron Annihilation Photon Imaging	3
1.2	Positro	on Emission Tomographs	- j
	1.2.1	Applications of PET Imaging	7
	1.2.2	Annihilation Coincidence Detection	3
	1.2.3	Scintillator Detectors)
	1.2.4	Detector to PMT Coupling	,
	1.2.5	Avalanche Photodiodes (APDs)	
1.3	2D vs.	3D PET Imaging	
1.4	Sinogra	ams in the Projection-space	
•			

, ,

				Table of Contents	V	
		1.5	Dvnar	nic PET Imaging	18	
		1.6	Cause	s of Degradation in Image quality and Quantitative Ac-		
			curacy		19	
			1.6.1	Attenuation of Annihilation Photons	21	
			1.6.2	Detection of Scattered Events	24	
			1.6.3	Detection of Random (or Accidental) Coincidences	26	
			1.6.4	Variations in Detector Pair Sensitivity	28	
			1.6.5	Detection Deadtime	30	
			1.6.6	Decay of Radioactivity	30	
1			1.6.7	Detector Blurring	31	
			1.6.8	Positron Range and Photon Non-collinearity	32	
			1.6.9	Patient Motion	32	
		1.7	The H	igh Resolution Research Tomograph	33	
			1.7.1	Challenges	33	
			1.7.2	More details for the Scanner	35	
		1.8	The A	im of This Work and Contributions	37	
	2.	3D .	Image I	Reconstruction	39	
		2.1	Introd	uction	39	
		2.2	Analy	tic Image Reconstruction	40	
			2.2.1	Reconstruction from Non-Truncated, 2D Parallel Pro-		
			,	jections	42	
			2.2.2	Central Section Theorem	42	
			2.2.3	Orlov's Sufficiency Condition	43	
			2.2.4	Image Reconstruction from Redundant Data	45	
			2.2.5	Direct Fourier Reconstruction	45	
			2.2.6	Filtered Back-Projection (FBP)	47	
			2.2.7	Backprojection followed by Filtering (BPF)	49	
,			2.2.8	Reconstruction from Truncated, 2D Parallel Projections	51	

-

.

		· · · ·	
	· .	Table of Contents	vi
	_		
	2.3	Rebinning Algorithms	. 51
		2.3.1 The SSRB and MSRB Algorithms	. 53
		2.3.2 The FOREX Algorithm	. 54
		2.3.3 The FORE Algorithm	. 56
		2.3.4 The FOREJ Algorithm	. 58
	2.4	Statistical Image Reconstruction	. 59
		2.4.1 Object Parametrization	. 60
		2.4.2 The System Matrix	. 61
		2.4.3 Statistical Modeling and Objective Functions	. 63
		2.4.4 Reconstruction Algorithms	. 69
	2.5	Problems with Expectation Maximization Algorithms	. 73
		2.5.1 Importance of Convergent Reconstruction	. 73
	,	2.5.2 Tackling Noise Artifacts: Regularization	. 74
		2.5.3 Accelerated EM Reconstruction Algorithms	. 77
		2.5.4 Summary	. 80
3.	List	mode Image Reconstruction	. 82
	3.1	Introduction	. 82
	3.2	List-mode Acquisition	. 83
	3.3	List-mode Reconstruction	. 84
		3.3.1 Benefits of List-mode Image Reconstruction	. 85
		3.3.2 Analytic List-mode Image Reconstruction	. 88
		3.3.3 Statistical List-mode Image Reconstruction	. 89
	3.4	Accelerated List-mode Reconstruction Algorithms	. 91
		3.4.1 Ordinary Subsetized List-mode EM Algorithm	. 91
		3.4.2 Convergent OSEM Reconstruction	. 92
		3.4.3 Convergent Subsetized List-mode EM Algorithm	. 94
		3.4.4 Hybrid S/CS List-mode EM Algorithm	. 96

i

}

Table of Coments	Table	of	Contents
------------------	-------	----	----------

		3.5.1 Ordinary Poisson EM Reconstruction
		3.5.2 The Delayed Events Subtraction Technique 100
	3.6	Scatter Correction in List-mode Reconstruction
	3.7	Summary
4.	Rec	onstruction Implementation and Comparison
	4.1	List-mode Geometric Back- and Forward-Projection Techniques104
		4.1.1 Projection Techniques
	4.2	Subtraction of Delayed Events and Non-negativity Constraints 107
	4.3	Methods
	4.4	Results and Discussion
	4.5	Dynamic PET List-mode Image Reconstruction
	4.6	Methods and Results
	4.7	Conclusion
5.	Mot	ion Compensation: Beyond the Event-Driven Approach 135
	5.1	Introduction
	5.2	Motion Correction in Histogram-mode EM Reconstruction $~.~.~137$
		5.2.1 Modification of the Histogram-mode EM Algorithm \therefore 138
		5.2.2 A Note on Convenience and Accuracy
		5.2.3 An Alternate, Fast Method for Calculation of Sensitiv-
		ity Factors
	5.3	Motion Correction in List-mode EM Reconstruction 147
	5.4	Methods
	5.5	Results and Discussion
	5.6	Conclusion
6.	Sum	nmary, Opinion and Conclusions
	6.1	Summary
	6.2	Opinion and Suggested Future Directions

vii

•	Table of Contents	viii
6.3 Conclusion .	· · · · · · · · · · · · · · · · · · ·	164
Bibliography		165

/ mm om o	11000
Апен	III PS
TUDUUIU	

A. Analytic Image Reconstruction from Truncated Projections 182
B. Space-variant and Anisotropic Resolution Modeling
C. Printed Sources for PET
D. Derivation of the List-Mode EM Reconstruction Algorithm with Mo- tion Compensation

List of Tables

List of Tables

1.1	Examples of Biomedical Applications of Positron Emitting
	Tracers
1.2	Table of Commonly Used Positron Emitting Isotopes 6
1.3	Table of Commonly Used Scintillation Materials (BGO: bis-
	muth germanate; GSO: gadolinium oxyorthosilicate; LSO: lutetium
	oxyorthosilicate; NaI: sodium iodide). BGO, GSO and LSO
	are not hygroscopic (unlike NaI), making packaging of the de-
	tectors easier
1.4	Specifications for Two PET Scanners
4.1	Table of Statistics for Frames used in the analysis
4.2	Image Bias for 3D-OSEM reconstruction of Frames 1 and 5 in
	table (4.1) (10M counts)
4.3	Table of random fractions
B.1	Characterization of Space-variance and Anisotropicity 189

ix

1 1	An example of a gamma camera used in single photon imaging	2
1.1	An example of a gamma camera used in single photon imaging.	2
1.2	Reconstruction of tomographic data allow non-invasive imag-	
	ing of three-dimensional objects	3
1.3	An emitted positron interacts with the surrounding medium	
	until it reaches thermal energies, after which it annihilates into	
	gamma rays	5
1.4	In single photon imaging, collimators are used in order to limit	
	the incident angles at which events will enter the detectors. This	
	lowers the sensitivity of the camera, but is necessary for image	
	formation. Use of physical collimators may however be avoided in	
	positron annihilation photon imaging systems, as described in text	
	(courtesy of Barry Pointon, BC Institute of Technology).	9
-1.5	In positron annihilation photon imaging systems, coincidence de-	
	tection may be utilized, in which two events arriving within a cer-	
	tain coincidence time window are used to trace the path along	
	which the annihilation occurred.	10
1.6	The energy distribution for an LSO (lutetium oxyorthosilicate) de-	
	tector is shown. The energy resolution of a detector (also depicted)	
	determines the extent and accuracy to which the incident gamma	
	rays may be filtered so as to reject events that do not correspond	
	to 511 keV gamma rays	11
1.7	An 8x8 crystal block is coupled to four PMTs.	14

х

1.8	3D PET imaging can be achieved by using scanners with retractable	
	septa or no septa at all. It can greatly increase the sensitivity of	• .
	the scanner, thus reducing noise in the final reconstructed images,	
	while posing new challenges in computation as well as data inver-	
	sion (courtesy of Dr. C-H. Chen, with modifications)	15
1.9	Geometry of a cylindrical PET scanner. Coordinates (s,ϕ,z,θ)	
	are used to parametrize the LOR connecting detectors A and	
	B. The direction along z corresponds to the axis of the scanner.	17
1.10^{-1}	(a) A 2D plane across an object (in image-space) containing four	
	fixed points results in (b) sinogram data containing four different	
	curves (projection-space).	18
1.11	Plots indicating annihilated photons from an event detected with	
	(a) true coincidence, (b) scattering, and (c) random coincidence	
	The last two clearly result in mis-positioning the LOR along which	
	the event is detected. \ldots	24
1.12	Three possible ways, as described in text, by which only a single	-
	event will be detected. \ldots	26
1.13	An open display of the high resolution research tomograph (HRRT).	
	33	
1.14	Sinogram data for a uniform cylinder (direct plane $\#110$ is	
	shown): the gaps comprise over 10% of the sinogram space.	
	This sinogram can be thought of as a superposition of corre-	
	sponding sinusoidal curves (Fig. 1.10) for the points which the	
	cylinder consists of. The observed non-uniformities are due to	
	noticeable variations in detector pair sensitivities (Sec. 1.6.4).	34
0.1		40
Z.1	A PEI event consists of detecting two coincident gamma rays.	40

xi

.

2.2	(a) A spherical PET scanner, with an LOR parametrized by	
	$(\hat{\mathbf{u}}, \mathbf{s})$ passing through it. (b) A cylindrical PET scanner, with	
	two LORs $(\hat{\mathbf{u}}, \mathbf{s_1})$ and $(\hat{\mathbf{u}}, \mathbf{s_2})$ passing through it, one measured	
	and one unmeasured. Therefore the measured data are trun-	
	cated since for certain directions, e.g. along $\hat{\mathbf{u}}$ shown, the line	
3	integrals are not measured for all values of \mathbf{s}	41
2.3	Each point on the sphere S^2 represents one 2D parallel projec-	
	tion. The set of measured projections Ω is in-between the two	
	parallels shown. The set Ω is intersected by the great circle	
	normal to any frequency \mathbf{v}	44
2.4	The backprojection of a 1D parallel projection onto a 2D image.	48
2.5	The rebinning procedure is summarized (courtesy of Dr. M.	
	Defrise, Vrije Universiteit Brussel).	52
2.6	(a) Single-slice rebinning (SSRB): Each oblique ($\theta > 0$) event	
	is assigned to the direct $(\theta = 0)$ sinogram centrally located be-	
	tween the two detectors. (b) Multiple-slice rebinning (MSRB):	
	Each oblique event is fractionally allocated to direct sinograms	
	that it traverses.	54
2.7	Fourier rebinning (FORE): In the frequency domain, data	÷
	from an oblique sinogram are allocated to a direct sinogram	
	axially located $k\delta/w$ below the central z value, as described	
	by Eq. (2.39).	58
3.1	Moore's Law vs. Nutt's Law, as described in text. The ver-	
	sion of the HRRT depicted is the single layer (~ 60 K crystals),	
	whereas most existing HRRT scanners are dual layer ($\sim 120 \mathrm{K}$	
	crystals) (courtesy of Dr. R. Nutt).	83

ſ

3.2	With conventional reconstruction (left) voxels along the LOR are incremented regardless of position along the LOR. With TOF reconstruction (right), each voxel is incremented by the probability (as measured by the TOF measurement) that the source originated at that voxel.	87
4.1	Projection algorithms employing (a) the Siddon method, (b)	
	trilinear, and (c) bilinear interpolation techniques are drawn,	
	as elaborated in text.	106
4.2	Results of back-projecting constant values along a projection	
	direction onto a 64x64 image at axial $\theta=0$ and transaxial	
	$\phi=30^{\circ}$.	106
4.3	(a) Plots of horizontal profiles across the back-projection im-	
	ages depicted in Fig. (4.2). (b) Projection vector obtained with	
	forward-projection of a uniform circle along $\theta=0$ and $\phi=15^{\circ}$.	107
4.4	A sample reconstruction of the radioactive paper source (3	
	iterations of the S-LMEM algorithm). The lower row of points	
	sources as well as the rectangular box were utilized for analysis	
	of resolution and noise properties.	114

xiii

4.5Calculated ratios between total counts for images reconstructed using 3D-OSEM and those obtained using FORE-2D-FBP, for the hot (top), background (middle) and cold (bottom) regions of the phantom, when (a) no random correction, (b) the sinogram non-negativity constraint and (c) the image nonnegativity constraint have been imposed. For better visibility, the plots have been shifted by ten units with respect to one another. The case of reading 10M counts is shown, and qualitatively similar results have been observed in the 40M and 80M cases. The error bars indicate statistical variation for the five transaxial planes selected. 116 Calculated ratios between total counts for images reconstructed 4.6using S-LMEM and those obtained using FORE-2D-FBP, for the hot (top), background (middle) and cold (bottom) regions of the phantom. For better visibility, the plots have been shifted by ten units with respect to one another. The case of 10M counts is shown, and qualitatively similar results have been observed in the 40M and 80M cases. The error bars indicate statistical variation for the five transaxial planes selected. 118 Contrast vs. noise plots for images reconstructed with S-LMEM 4.7and OSEM. Three iterations are performed with 16 subsets, with results shown every 4 subsets. Results are shown for (a) 20M and (b) 40M total counts. Presence of limit cycles is noticeable in both algorithms. Contrast vs. noise plots for images reconstructed using the 4.8S-LMEM (dotted), CS-LMEM (--) and hybrid S/CS-LMEM (solid) algorithms with scan durations containing (a) 10M and (b) 5M total counts. Three iterations are shown. 121

xiv

4.9	Plots of measured resolution vs. radial position for the S-
	LMEM algorithm implemented using the Siddon algorithm as
	well as bilinear and trilinear interpolation techniques. For
	comparison, results of output-driven (as described in text)
	histogram-mode OSEM reconstruction are also shown. Three
	iterations were performed
4.10	Plots of noise vs. iteration for the reconstruction schemes
	described in the caption of Fig. (4.9)
4.11	Plots of reconstructed FWHM width vs. Iteration are shown
	for the 3D-OSEM, S-LMEM, and the CS-LMEM and hybrid
	algorithms for the point source located at (a) 1 cm and (b)
	$5~\mathrm{cm}$ from center of the FOV. (c) Final FWHM values are
	shown for all the reconstructed points sources using the afore-
	mentioned four schemes (after 3 iterations).
4.12	Plots of reconstructed FWHM width vs. Iteration are shown
	for the 3D-OSEM, S-LMEM, and the CS-LMEM and hybrid
	algorithms for the point source located at (a) 1 cm and (b)
	5 cm from center of the FOV. (c) Final FWHM values are
	shown for all the reconstructed points sources using the afore-
	mentioned four schemes (after 3 iterations).
4.13	Plots of reconstructed FWHM width vs. Iteration are shown
	for the 3D-OSEM, S-LMEM, and the CS-LMEM and hybrid
	algorithms for the point source located at (a) 1 cm and (b)
	$5~\mathrm{cm}$ from center of the FOV. (c) Final FWHM values are
	shown for all the reconstructed points sources using the afore-
	mentioned four schemes (after 3 iterations).

ζ

	1	
4.14	(a) Overall FWHM vs. iteration for images reconstructed using	
	the hybrid S/CS-LMEM algorithm with axial spans of 3 and	
	9: point sources located at 3 cm and 5 cm from center of FOV	
	are shown. (b) Degradation in FWHM (along Z-direction)	
	vs. radial position as one switches from axial span 3 to 9:	
	all the point sources reconstructed using the S-LMEM, CS-	
	LMEM and hybrid schemes (after 3 iterations) are shown	126
4.15	Time activity curves (TACs) for 18 frames each 300 s in du-	
•	ration. First frame has a random fraction of 25%, last frame	
	5%. One iteration and 16 subsets are used in list-mode and	
	OSEM reconstructions.	130
4.16	Axial uniformity comparisons: profiles were drawn through	
	the hot (top), cold (bottom) and background (middle) regions	
	in each of the reconstructed images.	131
4.17	Plots of percentage contrast for the various dynamic frames.	133
5.1	An event detected at an LOR i' generated in voxel j , which	
	has been translated to voxel $j' = \mathcal{M}(j)$ at time of detection,	
	would have been detected at LOR $i = \mathcal{L}^{-1}(i')$ if the object	
	had not moved.	139
5.2	An event detected at time t_2 along an LOR i' , which would	
	have been detected along an LOR i if the object had not	
	moved, has the <i>same</i> attenuation correction factor as mea-	
	sured for A_i at time t_1 . The normalization correction factor	
	however is LOR-specific and changes with motion.	142

xvi

- (Top) Sinograms of a moving line source (a) without and (b) 5.3with motion-compensated histogramming. Clearly, upon appropriate histogramming, some counts are histogrammed into bins corresponding to detector gaps (shown by arrows), signifying that they would not have been detected had the object not moved. (*Middle*) Images reconstructed using histogrammode algorithm applied to (c) data with no motion and (d) data for motion when reconstructed without motion correction. (Bottom) (e) Image resulting from the proposed reconstruction algorithm given by Eq. (5.13). Radial profiles through this image as well as the one resulting from no motion data are shown in (f). Three iterations were applied to the data. Direct plane #110 is shown (randomly selected). . 152 Images reconstructed using list-mode algorithm applied to a) 5.4data with no motion, as well as data for motion when re-
- 5.5 Images reconstructed for a radioactive printed "W"-sign (coronal view) when (a) no motion is introduced; and with motion (three axial positions: 1 min 3 min 3 min) when (b) no motion compensation, (c) the purely event-driven approach (conventional motion correction) and (d) the proposed algorithm given by Eq. (5.20) are applied to the data. Plots of uniformity profiles drawn through images in (c) and (d) are shown in (e). Five iterations were applied to the data. . . . 157

ызь ог гідше	List	of Figu	res
--------------	------	---------	-----

B.1	Resolution relations implied by the common case of 90°-transaxial-
	rotation symmetry.
B.2	Observed transaxial (x/y) -resolution values as functions of
	space-invariant, isotropic kernel σ values. Points at which the
	curves are each minimized are indicated using circles 192
B.3	Variation of Δ as a function of modeled space-invariant, isotropic
	σ values, as described in text. Values achieved by using expo-
	nential and inverse-Gaussian modeling of degradation of reso-
	lution are also shown
C.1	Left: Side View, cut lines on printer ink cartridge. Right: Side
	View, section and assembly of modified ink cartridge 197
C.2	Axial (squares) and radial (diamonds) resolution measured for
	the HRRT
C.3	An example of a source profile imaged with (solid) and without
	(dashed) one layer of Al foil. The two profiles nearly coincide
	with one another.

2

xviii

Glossary

Glossary

Abbreviations and Acronyms

3DRP	3D Reprojection	:	reconstruction algorithm
APD	Avalanche photodiode	:	hardware
ART	Algebraic Reconstruction Technique	:	reconstruction algorithm
BGO	Bismuth Germanate	:	scintillator
BPF	Backprojection followed by filtering	:	reconstruction algorithm
CS-LMEM	Convergent Subsetized LMEM	:	reconstruction algorithm
DOI	Depth of Interaction	:	hardware
EM	Expectation Maximization	:	reconstruction algorithm
FBP	Filtered backprojection	:	reconstruction algorithm
FORE	Fourier Rebinning	:	rebinning algorithm
FWHM	Full width at half maximum	:	resolution measure
GSO	Gadolinium Oxyorthosilicate	:	scintillator
HRRT	High resolution research tomograph	:	scanner
LMEM	List-Mode Expectation Maximization	:	reconstruction algorithm
LOR	Line of Response	:	geometry
LSO	Lutetium Oxyorthosilicate	:	scintillator
LYSO	Lutetium Yttrium Oxyorthosilicate	:	scintillator
MAP	Maximum a posteriori	:	objective function
ML	Maximum likelihood	:	objective function
MSRB	Multi-Slice Rebinning	:	rebinning algorithm
NaI	Sodium Iodide	:	scintillator
NC	Normalization coefficient	:	correction factor
OSEM	Ordinary Subset EM	:	reconstruction algorithm

xix

Glossary

PET	Positron Emission Tomography	: imaging technique
PMT	Photomultiplier tube	: hardware
ROI	Region of Interest	: image analysis
SIR	Statistical image reconstruction	: reconstruction technique
S-LMEM	Subsetized LMEM	: reconstruction algorithm
SPECT	Single Photon Emission Computed	: imaging technique
	Tomography	
SSRB	Single Slice Rebinning	: rebinning algorithm
TTR	True three-dimensional reconstruction	: reconstruction filter
	•	

XX

Acknowledgements

He who does not express gratitude to creatures, has not expressed gratitude to the Creator. -Prophet Mohammad $\,$

Many individuals have contributed to my success in the course of completing this work. I would like to deeply thank my supervisor, Dr. Vesna Sossi, whose contributions to my work are many and diverse. Her very thorough and broad knowledge of PET imaging combined with her extraordinary enthusiasm have been truly inspirational. She has granted me much room for exploring my own ideas, while providing me with extremely valuable guidance throughout this work. I would also like to thank my supervisory committee members, Drs. Ruth, MacKay and Michal for their support and constructive feedback. In particular, I am indebted to Dr. Ruth for instructing me on various topics, particularly in the field of nuclear chemistry. I would also like to thank Drs. Kinahan, Kiefl, McKeown and Schrack for taking the time to review my dissertation and for their valuable comments during and following my oral examination. A very special thanks is due to Dr. Stephan Blinder, Kevin Cheng, Marie-Laure Camborde, Ken Buckley and Paul Piccioni for their much needed help in completing this work, and also to Barry Pointon, Eric Vandervoort, Siobhan McCormick and Dr. Doris Doudet for many interesting and vibrant discussions. I would also like to thank Christian Michel, Mark Lenox, Ziad Burbar, Merence Sibomana, Andrew Reader and Peter Bloomfield for their collaboration and feedback.

I am simply unable to recite my family's contributions to my life and work. My accomplishments are direct results of their love and support. My cousin and 'carpool buddy', Amir, been very supportive, and continues being a role model of kindness for me. My brother and friend, Iman, has also been incredibly kind to me. I am grateful for his words of advice in the past many years, and particularly during my job interview process. My wife, Zahra, has been truly supportive of my academic and extracurricular involvements, and has made the last year of my graduate studies one to remember for the rest of my life. My parents have sacrificed many things throughout their lives for my sake and I owe them all that I have. As a humblest gesture of appreciation, this dissertation is dedicated to them.

1. Introduction to PET Imaging

1.1 Nuclear Medicine Imaging

Nuclear medicine imaging is a non-invasive technique which allows dynamic examination of large areas of the body in a scanning session, producing images of human body functions unobtainable by other imaging techniques. It involves the use of radioactively labeled pharmaceuticals (radiopharmaceuticals) for functional (as compared to structural) imaging of the human body. The choice of pharmaceuticals, which can be intravenously injected, inhaled or ingested, depends on the function under investigation. The obtained images capture biochemical processes, such as early cancer tumor activity, that cannot be revealed by anatomical imaging with conventional X-ray, CT or MRI, and can therefore provide unique information on metabolic processes in healthy and diseased states. As further elaborated in Sec. (1.5), nuclear medicine imaging, combined with the application of appropriate mathematical models, can be used to investigate and quantify physiologic or metabolic processes *in vivo*.

Modern nuclear medicine imaging mainly consists of two main approaches: i) single photon imaging, and ii) positron annihilation photon imaging in which two photons are detected in coincidence.

1.1.1 Single Photon Imaging

Single photon imaging modality is based on detection of two-dimensional projections of the three dimensional radiopharmaceutical distribution. The



Fig. 1.1: An example of a gamma camera used in single photon imaging.

radiopharmaceuticals are formed using radioisotopes that emit *single* photons upon emission. A typical imaging system suitable for this task is the gamma camera, as shown in Fig. (1.1).

Upon rotating the gamma camera around the patient, a series of 2D projections can be obtained which can be combined to determine depth information. This technique is commonly referred to as *single photon emission* computed tomography (SPECT).

The word 'tomography' is derived from the Greek words 'tomo' meaning to slice and 'graph' meaning image. The basic idea is that if a sample is imaged several times in different orientations, three-dimensional (volume) information on the sample structure can be obtained using mathematical algorithms. This is called a tomographic reconstruction or tomography¹. It

¹ Tomography is often perceived as an imaging tool for medical examination purposes. It has to be emphasized, however, that the concept of tomography and its non-invasive way of imaging are not restricted to the medical field. Tomography has been developed, over the



Fig. 1.2: Reconstruction of tomographic data allow non-invasive imaging of threedimensional objects.

enables one to look at slices of the investigated object without physically cutting it, as depicted in Fig. (1.2).

Modern SPECT systems consist of two or three camera heads mounted on a single rotating gantry, thus allowing detection of a larger fraction of emitted photons (i.e. increased *sensitivity*) and therefore increasing the signal-tonoise ratio.

1.1.2 Positron Annihilation Photon Imaging

Positron-emitting isotopes can be also used for labeling of pharmaceuticals. As elaborated below, upon annihilation of an emitted positron, two gamma rays are produced which can be subsequently detected by the photon imaging system. One advantage of this technique (over SPECT) is that a number of elements which are fundamentally used by compounds of biological interest (C, N, O, and F) can be positron-emitters, allowing such positron-emitters to be more readily incorporated into a wide variety of useful radiopharmaceuticals (than is the case with single photon emitters). Some examples of

last decade, into a reliable tool for imaging numerous industrial applications (e.g. electrical capacitance or impedance tomography). This field of application is commonly known as Industrial Process Tomography (IPT) or simply Process Tomography (PT).

1. Introduction to PET Imaging

Radiotracer Radiopharmacutical		Examples of Biomedical Applications		
O-15 oxygen		oxygen metabolism		
O-15	carbon monoxide	blood volume		
O-15	carbon dioxide	blood flow		
O-15	water	blood flow		
N-13 ammonia blood flow		blood flow		
F-18 FDG glucose metabo		glucose metabolism		
F-18 FDOPA		pre-synaptic dopaminergic activity		
F-18 FMISO hypoxic cell tracer		hypoxic cell tracer		
C-11	DTBZ	vesicular monoamine transporter VMAT2		
C-11	Raclopride	dopamine D2 receptor		
C-11	methylphenidate (MP)	dopamine membrane transporter DAT		
C-11	SCH23390	dopamine D1 receptor		
C-11	flumazenil	benzodiazepine receptor		

Tab. 1.1: Examples of Biomedical Applications of Positron Emitting Tracers

biomedical application for radiopharmaceuticals with positron-emitting radiotracers are shown in table (1.1).

Positron Emission: A neutron-deficient nucleus can become stable via electron capture, whereby an atomic electron is captured by a proton (thus transforming the proton into a neutron), or alternatively via positron emission which arises from proton (p) decay, according to

$$p \to n + \beta^+ + \nu_e \tag{1.1}$$

where n is a neutron, β^+ is a positron and ν_e is an electron neutrino. For a general radionuclide, X, the following process occurs:

$${}^{M}_{Z}X \rightarrow {}^{M}_{Z-1}X + \beta^{+} + \nu_{e} \tag{1.2}$$

Due to emission of a neutrino along with the positron, the positron can be emitted with a continuous range of energies (up to a maximum energy;

4



Fig. 1.3: An emitted positron interacts with the surrounding medium until it reaches thermal energies, after which it annihilates into gamma rays.

see table 1.2). Positrons give up their kinetic energy principally by Coulomb interactions with electrons within the surrounding medium. As the rest mass of the positron is the same as that of the electron, the positrons may undergo large deviations in direction with each Coulomb interaction, as is depicted in Fig. (1.3).

When the positrons reach thermal energies, they start to interact with electrons: either (i) by annihilation, which produces two anti-parallel 511 keV photons, or (ii) by the formation of a hydrogen-like orbiting couple called positronium. In its ground-state, positronium has two forms: (i) orthopositronium (o-Ps), where the spins of the electron and positron are parallel, and (ii) para-positronium (p-Ps), where the spins are anti-parallel. Parapositronium again decays by self-annihilation, generating two anti-parallel 511 keV photons. Ortho-positronium, on the other hand, self-annihilates by the emission of three photons, but composes a negligible fraction of total

	$T_{1/2}$	E_{ave}	E_{max}	FWHM Range FWHM Ra	
	(min)	(MeV)	(MeV)	in water (mm)	in water (mm)
F-18	109.8	0.24	0.64	0.6	1.8
C-11	20.3	0.39	0.96	0.9	2.8
N-13	9.97	0.49	1.19	1.0	3.5
O-15	2.03	0.73	1.70	1.6	5.2

Tab. 1.2: Table of Commonly Used Positron Emitting Isotopes

positron annihilations² and can be safely ignored.

Higher energy positrons, upon being emitted, require to traverse a larger distance, on the average, in the surrounding medium before they can reach thermal energies in order to be annihilated: this distance is referred to as *positron range*. One notes that in positron annihilation photon imaging, the presence of positron range imposes a fundamental limit on spatial localization of the points from which positrons are emitted. We also note that since different positron-emitting isotopes exhibit distinct energy distributions, different isotopes also exhibit distinct positron range values. Table (1.2) summarizes such properties of commonly employed positron-emitting isotopes, as extensively elaborated in [2].

1.2 Positron Emission Tomographs

In what follows we provide a brief overview of the important and unique clinical applications of positron emission tomograpy (PET), followed by an explanation of the main components of a modern PET scanner and the prin-

² This is because, despite the larger amount (3:1) of o-Ps to p-Ps initially formed, the two gamma lifetime for p-Ps is 125 ps (in vacuum or in liquids). On the other hand, o-Ps has a three gamma lifetime of 142 ns (vacuum), yet in liquids, due to a "pick-off" process in which a second electron with opposed spin reacts with the positron in the o-Ps atom resulting in a *two photon annihilation*, the lifetime of o-Ps in liquids is considerably shorter (1800 ps in water) [1].

ciples employed to achieve high sensitivity and resolution.

1.2.1 Applications of PET Imaging

PET is a camera that produces powerful images of the human body's biological functions and entire organ system with one image. It reveals metastatic diseases other imaging techniques may not be able to detect, enabling the physicians to potentially diagnose disease before it shows up in other imaging modalities. PET can also diagnose cancer, cardiac disease, neurological brain disorders, as outlined below, and help guide physicians to the most beneficial therapies.

Tumors: PET imaging is very accurate in differentiating malignant from benign growths, as well as showing the spread of malignant tumors. PET imaging can help detect recurrent brain tumors and tumors of the lung, colon, breast, lymph nodes, skin, and other organs. Information from PET imaging can be used to determine what combination of treatment is most likely to be successful in managing a patient's tumor.

Cardiac PET: PET allows quantitative assessment of myocardial perfusion and metabolism. It can be used not only to diagnose coronary artery disease in patients with equivocal studies from conventional diagnostic techniques, but also to evaluate myocardial viability (the ability of the heart to recover after revascularization) in patients with heart failure.

Diseases of the Brain: PET imaging can provide information to pinpoint and evaluate diseases of the brain. PET imaging can show the region of the brain that is causing a patient's seizures and is useful in evaluating degenerative brain diseases such as Alzheimer's, Huntington's, and Parkinson's. Within the first few hours of a stroke, PET imaging may be useful in determining treatment therapies.

PET is an important *research* tool for the assessment of cerebral func-

tion, metabolism and receptor ligand systems. Dynamic PET imaging (see Sec. 1.5 for more detail) is particularly suited for research into physiologic and metabolic processes in the healthy and diseased states. In particular, in our PET centre at the University of British Columbia³, the main clinical research focus is the investigation of the neurodegeneration as it manifests itself in Parkinson's disease. The causes, origin and detailed disease progression mechanisms are still unknown. At present there is no cure, and only symptomatic treatment is available, which often causes disabling side-effects. PET has however been an invaluable tool in providing information on some aspects of the disease such as providing some evidence for early compensatory changes [3, 4], and possible mechanisms contributing to treatment induced motor complications [5, 6].

Furthermore, PET imaging has also proven to be a powerful tool for understanding the in vivo kinetics of new drugs, and thus is actively employed in drug discovery and development tasks.

1.2.2 Annihilation Coincidence Detection

In single photon imaging, collimators (as shown in Fig. 1.1) are used in order to enable the scanner to have knowledge of the angle at which photons are incident on the detectors. In other words, in such systems, without use of a collimator, the detected photon could have arrived from *any* region in the field of view, and therefore use of collimators is necessary for image formation (Fig. 1.4).

In the imaging of positron emitters, gamma cameras may again be utilized for measurement of events. Nevertheless, much greater sensitivity can be gained by noting that collimation can instead be performed *electronically*, as illustrated in Fig. (1.5): if two events arrive within a certain coincidence

³ http://www.pet.ubc.ca



Fig. 1.4: In single photon imaging, collimators are used in order to limit the incident angles at which events will enter the detectors. This lowers the sensitivity of the camera, but is necessary for image formation. Use of physical collimators may however be avoided in positron annihilation photon imaging systems, as described in text (courtesy of Barry Pointon, BC Institute of Technology).

time window, they are recorded as dually emitted gamma rays from a single positron annihilation. The location at which the positron annihilation has taken place can subsequently be traced to within the line-of-response (LOR) in between the two detectors. Positron emission scanners (commonly known as positron emission tomographs or PET scanners) therefore consist of a series of detector rings, each ring having a series of crystals, such that imagedomain information may be extracted from tomographic measurement of data.

This technique introduces two further considerations: (i) problem of random coincidences, in which the two events detected within the coincidence timing window originate actually from different positron annihilations, wherein their duals have not been detected. This is an important issue in PET imaging, and is further elaborated in Sec. 1.6.3; (ii) photon non-collinearity effect: since the positron and electron are never exactly at rest prior annihilation, conservation of momentum dictates a deviation from 180° between the trajectories of the two annihilation photons. This deviation is



Fig. 1.5: In positron annihilation photon imaging systems, coincidence detection may be utilized, in which two events arriving within a certain coincidence time window are used to trace the path along which the annihilation occurred.

around 0.5° full width at half maximum, which corresponds to a resolution blurring of ~1.1 mm for two detectors separated by 0.5 m. Nevertheless, the absence of the need for physical collimation in PET (replaced by electronic collimation) results in an approximately three-orders-of-magnitude increased detection sensitivity compared to SPECT.

1.2.3 Scintillator Detectors

In positron emission tomography, scintillator detectors are most commonly used to detect incoming gamma rays. The incoming radiation excites electrons in the scintillating material (molecular excitations) through Compton scattering and/or photoelectric absorption. The scintillator material subsequently "glows" as molecules return to their ground state, emitting scintillator photons. The number of emitted photons is (ideally) proportional to the energy deposited in the crystal. The scintillator photons next create electrons in the photocathode of the photomultiplier tubes (PMTs), which are subse-



Fig. 1.6: The energy distribution for an LSO (lutetium oxyorthosilicate) detector is shown. The energy resolution of a detector (also depicted) determines the extent and accuracy to which the incident gamma rays may be filtered so as to reject events that do not correspond to 511 keV gamma rays.

quently amplified by about 6-8 orders of magnitude and finally measured at the exit nodes of the PMTs.

Each scintillator material has its own characteristic properties. Generally, these can be categorized as:

i) Stopping Power (linear attenuation coefficient): the density and effective atomic number of a scintillator determine the relative probability by which Compton scattering and/or photoelectric absorption will occur as the processes by which incoming radiation is attenuated. Presence of Compton scattering is less preferred as it can result in the incoming radiation to be deflected to neighboring detectors, thus blurring the detection process, as elaborated in Sec. (1.6.7).

ii) Decay Time: this represents the time it takes for the electrons excited by the incoming radiation to return to their ground state, after which they are available to be excited by more incoming photons. Scintillators with shorter decay times allow higher count rates to be processed.

iii) Spectral Distribution: for good detection efficiency, a scintillator must give off light at wavelengths that are efficiently detected by the PMTs. It affects the energy resolution, as defined in Fig. (1.6), which will subsequently affect the ability to reject detected events that do not correspond to 511 keV gamma rays (e.g. scattered gamma rays as discussed in Sec. 1.6.2).

iv) Linearity: the amount of light produced by the scintillating material should be proportional to the energy deposited by the radiation.

v) Conversion Efficiency: this is the fraction of the radiation energy converted to detectable scintillator light. It is directly related to the photon yield (which is the number of scintillator photons produced per keV of radiation energy), and thus to energy resolution.

Table (1.3) compares properties of some scintillators commonly used in PET. BGO scanners have been very common in the past. However, with the advent of LSO (Lutetium Oxyorthosilicate) in 1991 [7] as a scintillator material which provided notable improvements in terms of decay time, conversion efficiency (therefore light output) and energy resolution, LSO (or LYSO⁴) scintillators have gradually become more popular.

1.2.4 Detector to PMT Coupling

PET scanners were originally designed for a one-to-one coupling of detector crystals and PMTs. In 1985, the scintillating crystal block technology was introduced by Casey and Nutt [8], followed nowadays in most PET scanners. An example of this implementation is shown in Fig. (1.7). In this case, an 8x8 block of crystals is coupled to four PMTs, with the basic idea being that an event detected at a particular crystal would dispense light in the four PMTs according to its distance from them, and the particular crystal geometric pattern. The crystal position can therefore be identified by linear

⁴ LSO doped with Yttrium.

	NaI	BGO	GSO	LSO
Density (g/cm^3)	3.67	7.13	6.7	7.4
Effective Z	51	74	61	66
Attenuation Length (cm)	2.88	1.05	1.43	1.16
Decay Time (nsec)	230	300	30-60	35-45
Conversion Efficiency (%NaI)	100	15	25	75
$\mu~({\rm cm^{-1}})$ at 511 keV	0.34	0.91	0.72	0.79
Energy Resolution	12	23	7.6	11.4
Hygroscopic	Yes	No	No	No

Tab. 1.3: Table of Commonly Used Scintillation Materials (BGO: bismuth germanate; GSO: gadolinium oxyorthosilicate; LSO: lutetium oxyorthosilicate; NaI: sodium iodide). BGO, GSO and LSO are not hygroscopic (unlike NaI), making packaging of the detectors easier.

averaging of light output in the four PMTs. With the considerable expense of PMTs and limitations on how small they can be made, the aforementioned scheme has allowed cost-effective manufacturing of PET scanners with small crystal sizes (few millimeters), thereby notably increasing resolution of PET scanners in the past two decades.

1.2.5 Avalanche Photodiodes (APDs)

In recent years, there has been a growing interest in exploring the possibility of replacing the PMTs with light-sensitive semiconductor detectors, such as Si photodiodes. These detectors exhibit a higher quantum efficiency (QE) compared to PMTs, yet in their original form, do not internally amplify the signal, unlike PMTs. Avalanche photodiodes (APDs), which are composed of Si photodiodes with internal gain, have therefore been proposed combining the two advantages. This has enabled detection of low light levels with good signal-to-noise ratios [9,10]. APDs have the additional advantage that they can be made very small in area and that they are typically only a



Fig. 1.7: An 8x8 crystal block is coupled to four PMTs.

few millimeters thick, including the packaging⁵. It must be noted that one major concern using APDs for PET applications is the limited timing resolution [12]. Furthermore, amidst a higher QE, they still provide 2-3 orders of magnitude less gain compared to PMTs. Nevertheless, these limitations are currently under continuous investigation and improvement, and APDs are likely going to be increasingly popular in the upcoming years.

1.3 2D vs. 3D PET Imaging

Originally, PET data were acquired in so-called septa-in or 2D mode. Metal septa were positioned between each detector ring in order to effectively absorb photons not being detected in the same ring, as shown in Fig. (1.8). For a scanner with N detector rings, the LORs were assigned to one of 2N-1 unique planes. With N of these planes corresponding to the detector ring planes, the other N-1 rings were taken to lie between the direct planes, corresponding to

⁵ Two types of APDs are commonly under investigation: (i) pixelated APDs, and (ii) position-sensitive (PS) APDs. The latter works by collecting signals from four contacts placed at the corners of the anode side of the PS-APDs, and using them to determine the position of the photon's interaction within the APD. The main advantage of this latter approach is the reduction in the number of channels that need to be read out, which results in a cost reduction and more compact design [11].



Fig. 1.8: 3D PET imaging can be achieved by using scanners with retractable septa or no septa at all. It can greatly increase the sensitivity of the scanner, thus reducing noise in the final reconstructed images, while posing new challenges in computation as well as data inversion (courtesy of Dr. C-H. Chen, with modifications)

events detected with a ring difference $\Delta = 1$ (the septa did not absorb photons at this minimal ring difference).

Later developments of 3D PET were motivated by a need to increase sensitivity of PET scanners, which would enable enhancement of image signalto-noise ratios. This was achieved by use of retractable septa (or no septa at all) and thus allowing the acquisition of those events where the gamma rays were emitted at an oblique angle relative to the tomograph axis. Typically, this leads to a fourfold to eightfold improvement in sensitivity [13]. However, 3D PET imaging poses new challenges: an important consideration has been that of additional computation burden caused by a large increase in the number of possible LORs. As an example, in the case of the high resolution research tomograph (Sec. 1.7), a system with 104 detector rings, one would have 2x104-1=207 direct (2D) data planes, whereas in fully 3D mode, interaction between all planes would be considered and thus one would have
$104^2 = 10816$ possible planes, a 50-fold increase in the size of the data. One method to tackle this computational burden is to use *rebinning algorithms* to convert the 3D-acquired data into a stack of ordinary 2D data sets, which can subsequently be reconstructed using 2D reconstruction algorithms. Sec. (2.3) discusses this approach in detail.

Another consideration is the application of suitable 3D reconstruction algorithms. This is especially complicated due to presence of incomplete data: exact analytic 3D reconstruction algorithms require access to data in all angles covering the field-of-view (FOV), yet practical scanners do not detect events along *all* possible directions. This has been discussed in more detail in Sec. (2.2.8).

A third difficulty with 3D PET imaging is the substantial increase in the scatter fraction (typically by a factor of \sim 3), as explained in Sec. (1.6.2), which renders application of accurate and suitable scatter correction algorithms very important.

1.4 Sinograms in the Projection-space

We shall restrict our attention to the general case of 3D PET imaging. As elaborated in Sec. (1.2.2), detection of two photons in coincidence defines an LOR joining the corresponding detector-pairs. The acquired LORs, corresponding to 1D lines in 3D space, can be parametrized using four parameters. We shall do so using the coordinates (s, ϕ, z, θ) where s and ϕ are the radial and azimuthal angular coordinates, z is the axial coordinate of the point midway between the two detectors, and θ is the copolar angle between the LOR and the transaxial planes. This is illustrated in Fig. (1.9). Defining $p(s, \phi, z, \theta)$ as the number of measured events along an LOR defined by the coordinates, it measures the *projection* of the image-space distribution onto the particular LOR (i.e. it represents those events whose annihilation gamma



Fig. 1.9: Geometry of a cylindrical PET scanner. Coordinates (s,ϕ,z,θ) are used to parametrize the LOR connecting detectors A and B. The direction along z corresponds to the axis of the scanner.



(a) Imaged points (b) Corresponding sinogram

Fig. 1.10: (a) A 2D plane across an object (in image-space) containing four fixed points results in (b) sinogram data containing four different curves (projection-space).

pairs are emitted along the 1D line-of-response passing through the object). 3D PET imaging of an object (in image-space), therefore, produces 4D data in the so-called *projection-space*).

In conventional PET imaging, the measured LORs are binned (or *his-togrammed*) onto 2D data sets, each defined by a particular z and θ . A given 2D data stack $p(s, \phi, ., .)$ (fixed z and θ) is referred to as a *sinogram* and is typically shown using a 2D plot with the vertical column representing ϕ and the horizontal column s. The sinogram data corresponding to a fixed point describes a sinusoidal curve, hence the origin of the term sinogram. An example of this is shown in Fig. (1.10).

1.5 Dynamic PET Imaging

The quantity that is measured in PET is the *in vivo* regional or local tissue concentration of the positron-emitting radiotracer. This quantity can be related to a physiologic or metabolic process through the application of an appropriate mathematical model⁶ of the process to dynamic PET data from the patient. As compared to static imaging, used for instance in tumor detection, dynamic imaging necessarily requires quantification (i.e. reconstructed images based on which meaningful quantitative measures can be made, as compared to, for instance, a simply qualitative detection of tumor contrast).

One major issue that renders PET an inherently quantitative imaging method (allowing the measurement of regional concentrations of the radiopharmaceutical injected after proper calibration) is the fact that the probability of survival of a pair of annihilation gamma rays is independent of the position of the annihilation along the LOR (more discussion in Sec. 1.6). To achieve quantitative detection, several problems still have to be overcome, which we review next.

The work presented in this dissertation has been focused on the need for practical quantitative image reconstruction procedures compatible with the high resolution research tomograph (HRRT; see Sec. 1.7 for details), and its requirements (e.g. a very large number of LORs, presence of gaps in-between the heads, etc.). In what follows, we first review general issues that need to be considered in PET in order to yield accurate and quantitative PET images, followed by a description of the design and properties of the HRRT.

1.6 Causes of Degradation in Image quality and Quantitative Accuracy

There are several physical phenomena that complicate PET imaging. It is important to understand these issues since lack of correction for them can re-

⁶ There exist various kinds of mathematical models, commonly referred to as tracer kinetic modeling techniques, of widely different mathematical characteristics - deterministic vs. stochastic, distributed vs. non-distributed, compartmental vs. non-compartmental, and linear vs. non-linear. In biomedical applications, linear compartmental models are most frequently used, because of their attractive mathematical properties. The reader is referred to Refs. [13–15] for more discussion.

sult in degradation of the image quality and/or quantitative accuracy. They are:

i) Attenuation of Annihilation Photons.

ii) Detection of Scattered events.

iii) Detection of Random (or accidental) coincidences.

iv) Variations in detector pair sensitivity.

v) Detection Deadtime.

vi) Decay of radioactivity.

vii) Detector blurring.

viii) Positron range

ix) Photon non-collinearity.

x) Patient Motion.

Of these, random coincidences, positron range and photon non-collinearity are specific to PET. The other effects also appear in SPECT. Furthermore, (corrections for) detector blurring, positron range and photon non-collinearity have been traditionally ignored since the camera resolution was not sufficient to be sensitive to these effects. However, with the arrival of high-resolution scanners, it has become more important to account for these factors in the reconstruction tasks, as we discuss later.

Compared to SPECT imaging, attenuation fractions are larger in PET yet corrections for them are much easier, as described next (thus relatively easing the task of quantitative imaging using PET). However, detection of scattered events is much more prevalent in 3D PET as discussed in Sec. (1.6.2), thus complicating quantitative reconstruction using PET. We discuss these phenomena in the following subsections and briefly review correction techniques used to compensate for their presence.

1.6.1 Attenuation of Annihilation Photons

One or both of the 511 keV annihilation photons may undergo photoelectric or Compton scattering prior to being detected. The incidence of photoelectric absorption is negligible (less than 1%) for 511-keV photons in body tissues. In a Compton interaction, the photon interacts with a free or outer shell electron. Subsequently, in every Compton scatter the direction of the gamma photon changes and its energy is decreased. These interaction can result in the scattered photon (i) to be deflected out of the field of view and never be detected, (ii) or to be deflected such that it is detected at another detector. Either of these cases results in non-detection of the LOR which would have been detected had no scattering occurred: this phenomenon is referred to as *attenuation*⁷. Defining the survival probability A_a as the probability of a photon not interacting as it propagates along a path a, it is given by

$$A_a = e^{-\int_a \mu(\mathbf{x}) d\mathbf{x}} \tag{1.3}$$

where $\mu(\mathbf{x})$ is the linear attenuation coefficient (which provides an indication of how effective a given material is, per unit thickness, in promoting photon interactions at position \mathbf{x}), increasing with higher matter densities, and decreasing with higher photon energies⁸. Similarly, the probability of attenuation, often referred to as the *attenuation factor* (AF), is easily given by $1 - A_a$.

At first thought, one might then conclude that since PET is based on higher energy photons compared to SPECT, the attenuation effect is less pronounced with PET. However, this is not the case due to the dual-photon nature of this modality. In other words, the survival probability A_l of an

 $^{^{7}}$ Case (ii) also results in incorrect LOR attribution, i.e. detection of scattered events, as we discuss in Sec. (1.6.2)

⁸ This is the case for energy levels in which the photoelectric effect and/or Compton scattering are dominant [16], which is the case in nuclear medicine imaging.

LOR is the product of the the probabilities that each of the two photons do not interact as they propagate along their paths a and b:

$$A_l = e^{-\int_a \mu(\mathbf{x})d\mathbf{x}} \times e^{-\int_b \mu(\mathbf{x})d\mathbf{x}} = e^{-\int_l \mu(\mathbf{x})d\mathbf{x}}$$
(1.4)

where l refers to the union of a and b, or in other words, the entire LOR. In PET imaging, typically more than 60% of all emitted photons interact with tissue [17]. However, a very significant conclusion may be drawn from Eq. (1.4): the attenuation along an LOR in PET is *independent* of the position along the LOR at which the annihilation event is generated. This is unlike SPECT in which attenuation exhibits depth-dependence (i.e. attenuation of gamma rays is dependent on the distance-to-the-detector of the location at which they are generated). This key observation renders corrections for attenuation in PET simpler and more quantitatively correct than in SPECT.

The Determination of the Attenuation Factors

The AFs need to be known in order to appropriately model the detection process and to produce quantitatively accurate reconstructed images. There are three main methods of obtaining the attenuation factors (when one makes use of the PET modality only⁹).

(i) Calculated attenuation factors: In cases when the object being scanned consists of uniformly attenuating material, and the location of the object is known, one can simply calculate the AFs using Eq. (1.4) applied to a body contour of constant μ values (recall: AF_l = 1 - A_l). This technique is more

⁹ In cases when PET scanning is combined with the CT (computed tomography) modality, the exact attenuation map $\mu(\mathbf{x})$ (i.e. attenuation values as a function of location) can be obtained using CT, and subsequently, one can calculate the AFs from the known attenuation map. This technique is complicated by the fact that μ values have a strong energy dependence, and the various components in the attenuation maps obtained using x-rays in CT need to be rescaled to correspond to μ values for positron-annihilation gamma rays.

applicable in imaging of the brain, but fails in whole body imaging, due to notable mixing of different attenuating material (e.g. bones, lungs).

(ii) Use of blank and transmission scans: Since the AFs are independent of the position of the annihilation along the LOR, one can consider a transmission source located outside of the object and performing measurements without and with the object in the FOV. The former is referred to as a blank scan and the latter, a transmission scan. Denoting B_l and T_l as the measured blank and transmission counts along an LOR l, the ratio between them is therefore a measure of the probability that generated annihilation photons are not attenuated along the LOR l. We therefore have:

$$A_l = \frac{T_l}{B_l} \tag{1.5}$$

In other words, without knowledge or reconstruction of attenuation maps, one is able to measure the AFs by performing a transmission source along with a reference scan, namely the blank, combination of which allows a measure of the probability of attenuation. The major problem with this approach is that presence of noise in the AFs can further amplify the present noise in the emission scans. This can be reduced by acquiring longer attenuation scans (i.e. more statistics), but this adds to the total scan time. One approach to reduce the noise is described next.

(iii) Transmission image segmentation: This is a technique, combining the first two approaches, intended to reduce the noise in the measured AFs. Combining Eqs. (1.4) and (1.5), we arrive at

$$-\ln\left(\frac{T_l}{B_l}\right) = \int_l \mu(\mathbf{x}) d\mathbf{x}$$
(1.6)

which is the projection of the attenuation map (defined earlier) onto the LOR l. One can then reconstruct $\mu(\mathbf{x})$ values from their projections using well-known analytic reconstruction techniques, e.g. filtered back-projection (described in detail in the next chapter). Alternatively, it is possible to



Fig. 1.11: Plots indicating annihilated photons from an event detected with (a) true coincidence, (b) scattering, and (c) random coincidence. The last two clearly result in mis-positioning the LOR along which the event is detected.

reconstruct $\mu(\mathbf{x})$ values using statistical techniques, to further suppress noise effects, as investigated in Refs. [18–20].

Once the typically-noisy attenuation map is reconstructed, it can be used to determine the boundaries of anatomical structures. Subsequently, the μ map can be segmented by assigning to regions within the boundaries the corresponding values of μ (which are known *a priori* for anatomical structures). The segmented μ -map may subsequently be used to calculate the AFs as given by Eq. (1.4). In cases when very long transmission scans are not practical, this approach allows an improved estimate of the AFs.

1.6.2 Detection of Scattered Events

Those annihilations for which one or both photons are scattered, but both are still detected, are termed *scattered* events. Scattering can result in assigning the detected photons to an incorrect LOR, as depicted in Fig. (1.11b).

Corrections for attenuated events as compared to those for scattered events differ in that for a given LOR, the former corrects the LOR for counts which it should have received had no Compton interaction taken place, whereas the latter seeks to correct the LOR for counts incorrectly attributed to it due to scattering. In 2D PET, the scatter fractions (ratio of scattered events to all measured counts) are in the 10-20% range (depending on scanner geometry, object size and the energy threshold used). This increases by about a factor of three to 30-55% in 3D PET imaging [17,21,22]. This renders use of scatter correction techniques even more critical in 3D imaging, and has been an important challenge in the transition from 2D to 3D PET imaging.

Since photons lose a fraction of their energy when they undergo a Compton interaction, a significant portion of the scattered events can be rejected by narrowing the range of photon energies accepted by the coincidence detection system (termed *energy discrimination*). Nevertheless, given the approximately 10% energy resolution on most current scanners, the threshold cannot be made too narrow so as not to reject true coincidences, and additional scatter correction techniques need to be employed for accurate quantitative reconstruction.

Various scatter correction techniques have been proposed: (i) use of convolution-subtraction with exponential functions [23, 24], (ii) comparison of 2D and 3D distributions [25], (iii) dual energy-window acquisition methods [26], (iv) direct calculation of the scatter distribution using the Klein-Nishina formula [27, 28]. The latter type of calculation is attractive since it treats the scatter using basic physical principles, and is gradually becoming a practical alternative due to the increasing computing power.



26

Fig. 1.12: Three possible ways, as described in text, by which only a single event will be detected.

1.6.3 Detection of Random (or Accidental) Coincidences

A detected annihilation photon is termed a *single* if the other photon is not detected. This can take place in three ways:

(a) The orientation of the annihilation is such that the other photon does not pass through the detectors.

(b) The other photon is scattered out of the FOV.

(c) The other photon passes through the detectors but is not detected¹⁰. These are depicted in Fig. (1.12).

Now, as explained in Sec. (1.2.2), annihilation coincidence detection works by recording two events arriving within a certain coincidence time window as dually emitted gamma rays from a single positron annihilation. However, it is possible for two singles arising from separate annihilations to be detected within the same coincidence timing window. An example of this has been shown in Fig. (1.11c).

The rate of random coincidences along a certain LOR connecting detectors i and j is given by

$$R_r = 2\tau S_i S_j \tag{1.7}$$

¹⁰ There is actually a fourth negligible possibility, and that is when the other photon undergoes photoelectric absorption. All of the three aforementioned mechanisms (but not this latter one) will be eliminated in a theoretical 4π scanner (completely covering the FOV) with 100%-sensitive detectors.

where τ is the coincidence timing window, and S_i and S_j refer to the singles rates at the two detectors¹¹. Examination of this relation shows that: (i) reducing the coincidence timing window reduces the count rate of random coincidences, and (ii) since singles rates are proportional to the activity in the scanned object, random rates vary as square of the activity. Subsequently, the fraction of detected random events (unlike the attenuation and scatter counter-parts) depends on the injected activity, and can therefore be an important factor in determination of the maximum injected activity.

Contribution of random coincidences to the measurement process has in the past been estimated and taken into account in a number of different approaches depending on the capabilities of the PET scanners being used. Such methods include direct use of the emission data for random correction, an instance of which is the image-based convolution-subtraction technique [24], which estimates the randoms in the object by fitting a function to the random tail outside the object. Work is also currently in progress to incorporate randoms into the system matrix of the EM algorithm by *estimation* of the *spatial* distribution of randoms contribution [29].

For PET systems in which singles rates can be measured during the coincidence measurement process, the singles count rate at the detectors can be used to estimate the randoms distribution [74]. For scanners without the capability to measure singles rates along with the coincidences, work is currently in progress to analytically calculate the singles rates from the reconstructed emission and transmission data [30]. This latter approach may require iterations since the initial emission image, using which singles rates are calculated, is not random-corrected.

Alternatively, a technique to correct for random coincidences is to acquire

¹¹ The number 2 in the equation can be attributed to the observation that the annihilation photon at detector i may be detected up to τ seconds before or after detection of the other photon at detector j in order for the coincidence detection system to accept and label the random pair as a coincidence event.

events arriving within a delayed coincidence window, delayed such that the probability of a true coincidence is zero. This modality is available in most modern PET scanners. The spatial distribution of the events within the delayed window is the same as those acquired in the true coincidence window. Thus, subtraction of the events acquired within the delayed window from the events acquired in the coincidence window (prompts) effectively corrects for the bias introduced by random coincidences. However, this subtraction technique can be a source of potential problems, as discussed in Sec. (2.4.3).

1.6.4 Variations in Detector Pair Sensitivity

LORs in PET scanners exhibit varying degrees of sensitivity due to a number of reasons: with changing incident angles to a crystal, the cross section (or effective surface area) of the crystal as well as its effective depth-of-interaction change, thus altering the efficiency. Furthermore, crystal imperfections, light guide variations, differences in PMT gains and variations in the electronics used to detect PMT signals also result in variations in efficiency, not only from block detector to block detector, but also across the elements of a block detector.

In order to account for these variations, current scanners typically acquire a normalization scan by illuminating all LORs with a uniform source. The normalization coefficients (NCs) are then proportional to the reciprocal of the number of counts obtained in each LOR. This process is known as *direct normalization*. from which the detector efficiencies can be calculated.

However, direct normalization can be confounded by three principal factors:

i) Very long acquisition times are required to obtain sufficient counts in high-sensitivity scanner-modes (i.e. non-compression of data) since lowactivity sources must be used to avoid problems with system dead-time¹².

ii) A further complication is that for any given LOR along which the events are attributed, the sensitivity is different depending on whether the events are trues or scattered. This is the case because:

(1) for a given LOR, the scattered photons may have arrived from a wide range of angles.

(2a) scattered events have different photon energies compared to unscattered events

(2b) detector efficiencies are energy-dependent since interactions and energy depositions inside the crystals depend on the initial photon energy. Lack of consideration of this issue can lead to both high- and low-frequency artifacts in the images [31].

iii) When using the block-detector technology, the degree of event mispositioning due to *pulse pile-up* (i.e. interference amongst pulses generated from separate gamma rays detected within the same block and within the same coincidence timing window) changes with count-rate, and this variation is not accounted for by direct normalization which is performed at a particular count-rate only [32].

To this end, component-based normalization approaches have been proposed that can take the above factors into consideration [33]. These techniques are based on separately measuring and/or calculating the various normalization components (e.g. geometric and detector efficiency factors). *Self-normalization* of emission data, without acquiring normalization data, has also been proposed [34]: but it is an approximate technique (it hinges on a limiting assumption) and does not address the second aforementioned complication.

¹² In fact, this is one important reason why direct normalization is currently not being utilized for span zero (i.e. no compression at all) in the HRRT scanner.

1.6.5 Detection Deadtime

The time required to process an event limits the counting rate of a PET scanner. With higher count-rates, a larger portion of the incoming counts are lost. The degree to which the deadtime effect is uniform within the tomograph is determined by the tomograph design and the activity distribution. This can be another limiting factor in determination of the maximum injected activity.

Depending on whether the bottleneck of the processing is in the scintillator decay time (Sec. 1.2.3), crystal identification (Sec. 1.2.4) and/or energy discrimination (Sec. 1.6.2), or the overall coincidence detection (Sec. 1.2.2), the deadtime effect will be at the level of the crystal, block or the entire system. In a system where the coincidence detection is the limiting factor, one can use the singles rates (which are not as much affected by the deadtime) to measure the global deadtime correction factors. However, other non-global deadtime correction schemes would be needed for other tomographs.

1.6.6 Decay of Radioactivity

Positron-emitting isotopes decompose and release radioactivity, enabling emission of gamma rays upon annihilation of the decayed positrons. As radioactive isotopes decay with time, the number of decays per second also decrease. Subsequently, scanning of the same object at a later time will result in fewer annihilation photons, and therefore, fewer reconstructed counts in the image. In order to enable the dynamic PET modality to be only sensitive to the biological process, one must correct the reconstructed images for decay.

Relative to the other corrections required in PET imaging, this one is the most trivial. One is simply required to scale the reconstructed image counts, depending on the radionuclide used in the scan. One complication can occur: scintillation detectors can themselves exhibit very low radioactive background counts, which are constant in time and unrelated to the injected activity¹³. In cases when the activity is close to background counts, scaling the entire image will amplify the background effect. One method to take this issue into account is to subtract an estimate of the background counts from the data prior to application of reconstruction and decay correction.

1.6.7 Detector Blurring

As described in Sec. (1.2.3), the incoming gamma rays excite electrons in the scintillating crystals through Compton scattering and/or photoelectric absorption. If these processes all occur in the crystal on which the radiation was incident, the event will be properly positioned. However, the photon detection process can exhibit two complications:

(i) It is quite likely for an incident photon to be scattered off to adjacent crystals, with this especially being the case in the detector-block technology. This is referred to as *inter-crystal scattering*. This explains the fact that the edge crystals typically receive less counts relative to the center crystals: this is because a photon scattering in a central element, may scatter to adjacent crystals but will eventually be detected (though it will be mis-positioned), however, it is quite likely for a photon entering an edge crystal to be scattered out of the entire block and not being detected at all.

(ii) It is also possible for gamma rays entering a crystal at an angle to simply pass through the crystal undetected and to be only detected in an adjacent crystal. As radiation occurs in voxels increasingly distant from the center of the FOV, it is more likely for the radiation to reach crystal fronts at higher angles of incidence, and to subsequently penetrate and be recorded in nearby crystals, degrading image resolution for such voxels. This effect

¹³ For instance, LSO has an intrinsic radioactivity of about 280 Bq/mL with singlephoton emissions in the range of 88 keV to 400 keV. On the HRRT scanner, one observes \sim 2K counts/sec (in coincidence) of LSO background using a 400-650 keV energy window.

is commonly referred to as the *parallax* effect). Measurement of depth-ofinteraction (DOI) within the crystals is known to minimize this problem, but its implementation has not achieved complete spatial invariance for resolution (see for e.g. [35]).

The above considerations are not commonly taken into account in reconstruction tasks due to computation difficulties. Furthermore, they cannot be accurately accounted for in analytic reconstruction algorithms. However, it is possible to incorporate these effects in statistical reconstruction tasks, as described in Sec. (2.4.2), though they can complicate the computation.

1.6.8 Positron Range and Photon Non-collinearity

These two effects were elaborated in sections (1.1.2) and (1.2.2). They are not often discussed as physical phenomenon that can be corrected for, rather they are often seen as limitations of PET imaging. However, with the arrival of statistical reconstruction algorithms (Sec. 2.4), it is now possible to model these effects. This is because, while it is not possible to determine, for a *particular* detected event, the positron range and photon non-collinearity, it is possible to calculate and incorporate the probability distributions for these effects into the system matrix, as elaborated in Sec. (2.4.2).

1.6.9 Patient Motion

Recent developments in 3D positron emission tomography (PET) systems have enabled the spatial resolution to reach the 2-3mm FWHM range. With such improvements in spatial resolution, small patient movements during PET imaging have become a significant source of resolution degradation. Chapter 5 provides an overview of existing motion compensation techniques. Accurate and practical motion correction algorithms, proposed in the course of this work, have also been described and tested experimentally.





1.7 The High Resolution Research Tomograph

The Vancouver high resolution research tomograph (HRRT), as shown in Fig. (1.13), is a state-of-the-art high sensitivity, high resolution scanner. It is a 3D-only dedicated brain tomograph employing the scintillators LSO/LYSO and using depth-of-interaction (DOI) information (see below).

1.7.1 Challenges

The HRRT is built of eight panel detector heads arranged in an octagon. Each head is separated by 46.9 cm from the opposing head and by 1.7 cm from the two neighboring heads. This presence of gaps in-between the detectors is clearly manifested in the sinogram data, as is shown in Fig. (1.14), and poses new challenges to image reconstruction from the HRRT data. To see this, one notes that in analytic reconstruction algorithms (see Sec. 2.2), one requires access to the complete data (i.e. data measured along all possible



Fig. 1.14: Sinogram data for a uniform cylinder (direct plane #110 is shown): the gaps comprise over 10% of the sinogram space. This sinogram can be thought of as a superposition of corresponding sinusoidal curves (Fig. 1.10) for the points which the cylinder consists of. The observed non-uniformities are due to noticeable variations in detector pair sensitivities (Sec. 1.6.4).

LORs passing through the FOV). This means that the gaps manifested in the sinograms need to be filled using approximate techniques¹⁴. On the other hand, gap-filling is not required in statistical reconstruction algorithms (see Sec. 2.4), which has been one key motivation for employing such methods in this work.

Another challenging property of the HRRT is the very large (namely 4.49×10^9) number of LORs which it measures, exceeding most modern PET scanners by 2-3 orders of magnitude. For data stored and/or reconstructed in histogram-mode, one is expected to commonly measure (especially in dynamic studies wherein a single study is divided into multiple frames), the total counts per frame to be significantly less than the total number of sino-

¹⁴ Such gap-filling techniques used in analytic reconstruction include (i) direct interpolation of the sinogram data, (ii) forward-projection of preliminary image estimates , or (iii) a constrained Fourier space method [36].

gram bins. The LOR data are acquired in list-mode (see Sec. 3.2), i.e. they are stored as they are detected one-by-one in the form of a list. The data may subsequently be histogrammed into appropriate sinogram frames. In this work, we have explored taking the list-mode approach one step further, namely skipping the histogramming step, and reconstructing the data directly from the list-mode data. This is elaborated in chapter 3.

The DOI information in the HRRT is obtained by pulse shape discrimination (PSD) to distinguish between events in the front layer from those in the back layer [37, 38]. This DOI modality was introduced for the HRRT to reduce the *parallax* effect: space-variant degradation in resolution of the scanner caused by increasing inter-crystal penetration of gamma rays with increasing angles of incidence. Nevertheless, the DOI information (due to its discrete nature) is not able to completely eliminate the parallax effect. In Sec. (B), we have investigated modeling of this effect into feasible statistical reconstruction of the HRRT data.

1.7.2 More details for the Scanner

A detector head consists of 9(transaxial)x13(axial) crystal blocks, which are viewed by 10x14 PM tubes. The HRRT therefore contains 8x9=72 blocks per ring and 8x9x13=936 total blocks. Each block consists of two 10-mm-thick layers (LSO and LYSO). Each double-layer block is cut into an 8x8 matrix of crystals: each crystal is $(2.438 \text{ mm})^2$. In total therefore, the HRRT contains $936x8x8x2\sim120$ K total crystal elements. Each head is in coincidence with five opposing heads, which results in 4.49B possible LORs. These properties are summarized in table (1.4), along with comparisons with another scanner currently in use by our group, namely the Concorde microPET R4 scanner for rodents [39].

The events are initially stored as 64-bit list-mode data, containing the in-

Specifications	microPET R4	HRRT
Detector Diameter	14.8 cm	46.9 cm
Radial Field-of-View	10.0 cm	$31.2~\mathrm{cm}$
Axial Field-of-View	· 7.5 cm	$25.1~\mathrm{cm}$
No. of Blocks per Ring	24	72
Total No. of Blocks	96	936
No. of Crystal Rings	32	104
Depth-of-Interaction Encoding	NO	YES
Total No. of Crystal Elements	6,144	119,808(!!)
No. of Lines-of-Response	8.26M	4.49B(!!)
Absolute System Sensitivity	$\sim 2\%$	$\sim 6\%$
Resolution at Center of FOV	1.8 mm	2.4 mm

Tab. 1.4: Specifications for Two PET Scanners

formation for the coordinates of the crystal-pairs along which the annihilation photons are detected. A 72 Gbyte RAID disk system allows data acquisition with a maximum storage rate of 40Mbytes/sec for list-mode data. The events can subsequently be converted into 32-bit data, wherein the coordinates of the lines joining the crystal pairs (and not the crystals themselves) are stored. Alternatively, the events can be histogrammed into sinogram data. Presently one uses 288 angular projections and 256 radial elements on each sinogram plane. Flexible axial compression modes, characterized by *span* (axial mashing of neighboring LORs) and *ring difference* (*RD*) [17] are available to further reduce the data. With no data compression, one obtains 10816 sinogram planes, corresponding to ~1.5GB of data in sinogram-mode. However routinely, the data are stored in compressed format (using a socalled span 3 and a maximum RD of 67), resulting in 6367 sinograms with a total size of ~0.9GB.

Transmission measurements are done with a Cs-137 point source, which is

driven by two servo motors and can sample the entire FOV in 2 min. Direct normalization is currently being commonly employed for the HRRT, with the normalization correction factors obtained from a long scan using a rotating rod source. The coincidence timing window can be set as low as 2 ns and can be incremented in steps of 2 ns. Currently, a coincidence time window of 6 ns is used.

1.8 The Aim of This Work and Contributions

The work presented in this dissertaion has resulted in a number of publications [40,41] and presentations [42–46]. As already mentioned in Secs. (1.5) and (1.7.1), our central aim has been to propose, implement and verify practical image reconstruction algorithms compatible with quantitative imaging with the HRRT. To address some of the challenging properties of the HRRT (namely, the very large number of LORs as well as the presence of gaps between the detector heads), we focused our attention to statistical list-mode reconstruction algorithms.

Chapter 2 provides a general summary of analytic and statistical reconstruction algorithms, while chapter 3 presents an overview of motivations for list-mode image reconstruction. Various accelerated list-mode reconstruction algorithms were proposed, with particular emphasis on a novel *convergent* reconstruction method. A practical random correction technique was also proposed for use in list-mode reconstruction from the HRRT data.

Chapters 4 and 5 represent the experimental core of this thesis work. The former elaborates upon the details of implementation and verification of statistical list-mode reconstruction on the HRRT. Furthermore, we extended the list-mode approach to dynamic (4D) image reconstruction, and demonstrated its quantitative accuracy and robustness compared to a number of conventional techniques. In addition, we proposed a modification of conventional data non-negativity constraints to remove the resulting overestimation biases, which were elaborately studies and quantified in this work.

Chapter 5 is devoted to the proposal and study of novel motion compensation techniques, with validation on the HRRT. We developed an accurate scheme, applicable to both the histogram-mode and list-mode image reconstruction methods, which incorporated patient motion into the reconstruction task itself (unlike the existing purely event-driven approach). Subsequently, clear improvements in image quality (particularly in terms of axial uniformity) were demonstrated. The practicality of the technique was derived from the observation that an otherwise time-consuming motion averaging step was only required to be performed in the image space (and not in the projection space). A more detailed summary of each chapter, as well as concluding remarks and opinions have been presented in chapter 6.

2. 3D Image Reconstruction

This chapter is intended to present an overview of 3D image reconstruction techniques applicable to PET imaging. Both analytic and statistical reconstruction algorithms are discussed, with particular emphasis and elaboration on the latter, since statistical list-mode reconstruction is the central theme of this dissertation (for more discussion see chapter 3).

2.1 Introduction

As discussed in the previous chapter, an event is recorded in the PET measurement process if two photons are measured in coincidence. Denoting the coordinates for the measured photons using (x_1,y_1,z_1) and (x_2,y_2,z_2) , a lineof-response (LOR) may be traced in-between the two coordinates, as shown in Fig. (2.1).

The image inversion or reconstruction task in positron emission tomography is aimed at finding the number of annihilation events f which occurred at a position \mathbf{x} in the image space given the number of events in the measured LORs¹. Reconstruction techniques can be classified into two main categories: analytic and statistical.

 1 With knowledge of scan duration, one may then calculate the activity at the given position.



Fig. 2.1: A PET event consists of detecting two coincident gamma rays.

2.2 Analytic Image Reconstruction

In analytic reconstruction, one aims to reconstruct f from its projection line integral

$$p(\hat{\mathbf{u}}, \mathbf{s}) = \int_{\infty}^{\infty} f(\mathbf{s} + l\hat{\mathbf{u}}) dl$$
 (2.1)

where the pair $(\hat{\mathbf{u}}, \mathbf{s})$ parametrizes the straight line along the unit vector $\hat{\mathbf{u}}$ and passing through \mathbf{s} , a point in 3D space. To avoid redundancy, the vector \mathbf{s} is restricted to a 2D plane

$$\hat{\mathbf{u}}^{\perp} = \{ \mathbf{s} \in R^3 | \ \mathbf{s}. \hat{\mathbf{u}} = 0 \}$$

$$(2.2)$$

since otherwise there exist (infinitely) many points, which would define the same parallel projection along a projection direction $\hat{\mathbf{u}}$. Therefore, for a given projection direction $\hat{\mathbf{u}}$, the vector \mathbf{s} is restricted to a 2D plane perpendicular to $\hat{\mathbf{u}}$. This is depicted in Fig. (2.2a) for a spherical scanner. The integral transform mapping the function $f(\mathbf{x})$ onto its line integrals is called the three-dimensional X-ray transform.



Fig. 2.2: (a) A spherical PET scanner, with an LOR parametrized by (û, s) passing through it. (b) A cylindrical PET scanner, with two LORs (û, s₁) and (û, s₂) passing through it, one measured and one unmeasured. Therefore the measured data are truncated since for certain directions, e.g. along û shown, the line integrals are not measured for all values of s.

2.2.1 Reconstruction from Non-Truncated, 2D Parallel Projections

Introducing S^2 as the unit sphere defining all possible directions for \hat{u} , the projection function p will in general be defined on the tangent bundle to S^2 :

$$T(S^2) = \{ (\hat{\mathbf{u}}, \mathbf{s}) | \ \hat{\mathbf{u}} \in S^2, \mathbf{s} \in \hat{\mathbf{u}}^\perp \}$$

$$(2.3)$$

In practice, however, 2D parallel projections are not completely measured in all directions. Rather, projections are measured along only a subset $\Omega \subset S^2$ of the unit sphere (called the aperture of the scanner). Furthermore, in 3D analytic PET reconstruction, the data are often *truncated*. In other words, in most applications, for a given parallel projection $\hat{\mathbf{u}}$, contributions from all coordinates \mathbf{s} in the image are not necessarily measured, and only a subregion $\hat{\mathbf{u}}_m^{\perp}$ of the projection plane $\hat{\mathbf{u}}^{\perp}$ (Eq. 2.2) is measured. An example of this is encountered in the cylindrical geometry as shown in Fig. (2.2b).

As it turns out [47], the inversion formula for such truncated geometries are all based on results obtained for the simpler case of a complete/nontruncated geometry

$$T(\Omega) = \{ (\hat{\mathbf{u}}, \mathbf{s}) | \ \hat{\mathbf{u}} \in \Omega, \mathbf{s} \in \hat{\mathbf{u}}^{\perp} \}$$
(2.4)

wherein for any projection direction $\hat{\mathbf{u}} \in \Omega$, measurements are made for all s in the projection plane $\hat{\mathbf{u}}^{\perp}$. In what follows, we shall first present an overview of analytic inversion techniques for the non-truncated case. In Sec. (2.2.8), methods of handling truncated projection data will be discussed.

2.2.2 Central Section Theorem

The central section theorem is a fundamental theorem used in analytical reconstruction in nuclear medicine imaging. We first define a rotated² Cartesian system (x', y', z') such that for a given projection direction $\hat{\mathbf{u}}$, the axis

² This can, for instance, be achieved by rotating the coordinate frame through the azimuthal angle ϕ , followed by a rotation through the copolar angle θ . It can be shown [48]

x' lies along $\hat{\mathbf{u}}$, and subsequently, the coordinate axes y' and z' span the 2D plane $\hat{\mathbf{u}}^{\perp}$. The 2D Fourier transform of Eq. (2.1) with respect to the variable $\mathbf{s} \in \hat{\mathbf{u}}^{\perp}$ is then given by

$$P(\hat{\mathbf{u}}, \mathbf{v}) = \iint_{\hat{\mathbf{u}}^{\perp}} \int_{-\infty}^{\infty} f(\mathbf{s} + l\hat{\mathbf{u}}) \exp(-2\pi i \mathbf{v} \cdot \mathbf{s}) dl d\mathbf{s}$$
(2.6)

where

$$\mathbf{v} = (v_{y'}, v_{z'}) \tag{2.7}$$

On the other hand, the 3D Fourier transform of the unknown distribution $f(\mathbf{x})$ is given by

$$F(\mathbf{V}) = \iint_{\hat{\mathbf{u}}^{\perp}} \int_{-\infty}^{\infty} f(\mathbf{s} + l\hat{\mathbf{u}}) \exp(-2\pi i \mathbf{v}.\mathbf{s}) \exp(-2\pi i v_{x'} l) dl d\mathbf{s}$$
(2.8)

where we have decomposed the 3D vector \mathbf{V} into $v_{x'}$ and \mathbf{v} . When \mathbf{V} lies along the central plane (passing through the origin, and equivalent to the projection plane $\hat{\mathbf{u}}^{\perp}$ defined earlier) which occurs when $v_{x'}=0$ and $\mathbf{V}=\mathbf{v}$, Eqs. (2.6) and (2.8) become equivalent; in other words:

$$F(\mathbf{v}) = P(\hat{\mathbf{u}}, \mathbf{v}) \tag{2.9}$$

This is the central slice theorem; namely: for a given projection direction $\hat{\mathbf{u}}$, the central plane of the 3D Fourier transform $F(\mathbf{V})$ of the 3D true activity $f(\mathbf{x})$ is equal to the 2D Fourier transform of the 2D projection data along the same direction.

2.2.3 Orlov's Sufficiency Condition

Orlov's sufficiency condition [49] states that any great circle on the unit sphere S^2 must have a non-empty intersection with the set Ω of directions

that this change in the coordinate frame is obtained mathematically via the transformation:

$$\begin{pmatrix} x'\\y'\\z' \end{pmatrix} = \begin{pmatrix} \cos\phi\cos\theta & \sin\phi\cos\theta & \sin\theta\\ -\sin\phi & \cos\phi & 0\\ -\cos\phi\sin\theta & -\sin\phi\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x\\y\\z \end{pmatrix}$$
(2.5)



Fig. 2.3: Each point on the sphere S^2 represents one 2D parallel projection. The set of measured projections Ω is in-between the two parallels shown. The set Ω is intersected by the great circle normal to any frequency **v**.

 $\hat{\mathbf{u}}$ along which complete measurements are made. To prove that Orlov's condition allows for reconstruction of the unknown activity $f(\mathbf{x})$ from parallel projections, we shall work in the Fourier domain while sampling the space with \mathbf{v} , and let us assume that Orlov's condition is satisfied. An example of this is shown for the set Ω subtended between two parallels, as shown in Fig. (2.3). Then, for any vector \mathbf{v} , it is possible to find at least one $\hat{\mathbf{u}} \in \Omega$ which is intersected by the great circle orthogonal to the vector \mathbf{v} , i.e., there is at least one projection direction $\hat{\mathbf{u}}$ along which measurements $p(\hat{\mathbf{u}}, \mathbf{s})$ (see Eq. 2.1) are made, and therefore from which the 2D Fourier component $P(\hat{\mathbf{u}}, \mathbf{v})$ can be recovered. Subsequently, one will arrive at the value of $F(\mathbf{v})$ using the central section theorem (2.9), and will completely recover $F(\mathbf{v})$ by repeating this procedure for all frequencies \mathbf{v} . The image activity distribution can finally be obtained by an inverse 3D Fourier transform of $F(\mathbf{v})$. On the other hand, if Orlov's condition is not satisfied, there exists at least one frequency \mathbf{v} such that the great circle orthogonal to it does not intersect Ω ,

hence $F(\mathbf{v})$ cannot be recovered for that particular frequency.

2.2.4 Image Reconstruction from Redundant Data

In Fig. (2.3), for a given frequency \mathbf{v} , there exists more than one projection direction $\hat{\mathbf{u}} \in \Omega$ (such that $\hat{\mathbf{u}}.\mathbf{v} = 0$) which may subsequently be used to determine $F(\mathbf{v})$. This demonstrated the non-uniqueness inherent in inverting the 3D X-ray transforms, whereas in the 2D case, the inversion is *not* overdetermined³. The over-determination of the inverse 3D X-ray transform can also be interestingly observed from a simple dimensional analysis [50]: the unknown function $f(\mathbf{x})$ depends on three scalar variables, whereas the data $p(\hat{\mathbf{u}}, \mathbf{s})$ depend on the four scalar variables required to parametrize a straight line in \mathbb{R}^3 . On the contrary, in the 2D case, a point $x \in \mathbb{R}^2$ and a straight line are both parametrized with two scalar variables!

The simplest approach to the issue of redundancy is to select only one of the estimates. In the absence of noise (i.e. consistent data), all $P(\hat{\mathbf{u}}, \mathbf{v})$ values calculated along any projection directions $\hat{\mathbf{u}}$ perpendicular to \mathbf{v} are equal to $F(\mathbf{v})$. Nevertheless, in the practical case in which noise is present, collective use of all the estimates can be used to enhance the signal-to-noise ratio, which has been the motivation in switching from 2D PET to 3D PET (Sec. 1.3). In what follows, we describe three image reconstruction techniques that make use of the entire 3D data set: (i) direct Fourier reconstruction, (ii) filtered-backprojection, and (iii) backprojection followed by filtering.

2.2.5 Direct Fourier Reconstruction

Assuming Orlov's condition to be satisfied, for any given frequency \mathbf{v} , one may obtain an estimate for the value of $F(\mathbf{v})$ by simply averaging values of

³ This can be more clearly seen by noting that the set Ω in the 2D case is simply an equilateral circle on the unit sphere.

 $P(\hat{\mathbf{u}}, \mathbf{v})$ obtained along all projection directions $\hat{\mathbf{u}}$ such that $\hat{\mathbf{u}} \cdot \mathbf{v} = 0$; in other words:

$$F(\mathbf{v}) = \frac{\int_{\Omega} P(\hat{\mathbf{u}}, \mathbf{v}) \delta(\hat{\mathbf{u}}, \mathbf{v}) d\hat{\mathbf{u}}}{\int_{\Omega} \delta(\hat{\mathbf{u}}, \mathbf{v}) d\hat{\mathbf{u}}}$$
(2.10)

The Dirac delta function in the integrand selects the projection directions which are perpendicular to the sampled frequency \mathbf{v} : the 2D integral above is then restricted to a one-dimensional integral over the intersection of such projection directions with the set Ω , as shown in Fig. (2.3). Using this technique, and sampling the frequency domain, one may subsequently take the 3D inverse Fourier transform of $F(\mathbf{v})$ to recover the image.

A more generalized approach to this problem is by weighting (not necessarily equally, unlike above) the $P(\hat{\mathbf{u}}, \mathbf{v})$ values in order to obtain an estimate of $F(\mathbf{v})$. One therefore introduces a weighting function $H(\hat{\mathbf{u}}, \mathbf{v})$ as such

$$F(\mathbf{v}) = \int_{\Omega} H(\hat{\mathbf{u}}, \mathbf{v}) P(\hat{\mathbf{u}}, \mathbf{v}) \delta(\hat{\mathbf{u}} \cdot \mathbf{v}) d\hat{\mathbf{u}}$$
(2.11)

To better understand the form of the weighting function $H(\hat{\mathbf{u}}, \mathbf{v})$, we note that assuming an object distribution $f(\mathbf{x}) = \delta(\mathbf{x})$ (i.e. a central point source), then the 3D Fourier transform satisfies $F(\mathbf{v}) = 1$; and similarly, by the central section theorem $P(\hat{\mathbf{u}}, \mathbf{v}) = 1$. Substituting these results into Eq. (2.11), we arrive at the normalization condition:

$$\int_{\Omega} H(\hat{\mathbf{u}}, \mathbf{v}) \delta(\hat{\mathbf{u}}, \mathbf{v}) d\hat{\mathbf{u}} = 1$$
(2.12)

Weighting functions $H(\hat{\mathbf{u}}, \mathbf{v})$ satisfying the above condition may be easily constructed: to see this we note that for an arbitrary function $H(\hat{\mathbf{u}}, \mathbf{v})$ defined on $T(\Omega)$, the function

$$H(\hat{\mathbf{u}}, \mathbf{v}) = \frac{G(\hat{\mathbf{u}}, \mathbf{v})}{\int_{\Omega} G(\hat{\mathbf{u}}', \mathbf{v}) \delta(\hat{\mathbf{u}}'. \mathbf{v}) d\hat{\mathbf{u}}'}$$
(2.13)

will satisfy the normalization condition (2.12), provided G is such that the

denominator is defined and does not vanish⁴. This observation explicitly demonstrates the redundancy inherent in the 3D X-ray transform, as different filters may be used for exact inversion of the transform⁵.

2.2.6 Filtered Back-Projection (FBP)

The 3D inverse Fourier transform of Equation (2.11), which recovers the image activity distribution, can be shown [52] to be mathematically equivalent to a technique commonly referred to as *filtered-backprojection (FBP)*; it proceeds in two steps:

Step 1: The filtering step

$$p^{F}(\hat{\mathbf{u}},\mathbf{s}) = \int_{\hat{\mathbf{u}}^{\perp}} h(\hat{\mathbf{u}},\mathbf{s}-\mathbf{s}') p(\hat{\mathbf{u}},\mathbf{s}') d\mathbf{s}'$$
(2.16)

is a convolution between the 2D parallel projection data and the 2D inverse Fourier transform of the weighting function $H(\hat{\mathbf{u}}, \mathbf{v})$, which can alternatively be performed using a frequency domain multiplication of $P(\hat{\mathbf{u}}, \mathbf{v})$ and $H(\hat{\mathbf{u}}, \mathbf{v})$:

$$P^{F}(\hat{\mathbf{u}}, \mathbf{v}) = H(\hat{\mathbf{u}}, \mathbf{v})P(\hat{\mathbf{u}}, \mathbf{v})$$
(2.17)

 4 Another method by which a filter satisfying the normalization condition can be reconstructed is via additive normalization of an arbitrary function ${\cal G}$

$$H(\hat{\mathbf{u}}, \mathbf{v}) = G(\hat{\mathbf{u}}, \mathbf{v}) + \frac{w(\hat{\mathbf{u}}) \left(1 - \int_{\Omega} G(\hat{\mathbf{u}}', \mathbf{v}) \delta(\hat{\mathbf{u}}', \mathbf{v}) d\hat{\mathbf{u}}'\right)}{\int_{\Omega} G(\hat{\mathbf{u}}', \mathbf{v}) \delta(\hat{\mathbf{u}}', \mathbf{v}) d\hat{\mathbf{u}}'}$$
(2.14)

where w is an arbitrary integrable function of \hat{u} such that the denominator does not vanish. It can be easily shown that this construction also satisfies the normalization condition (2.12). As elaborated in Appendix A, this latter formulation is not a simple mathematical curiosity. Rather, it can be used to design filters suitable for the task of inverting truncated projections.

⁵ The simplest and earliest weighting function, referred to as Colsher's filter [51], is one which yields Eq. (2.10) (i.e. at any frequency \mathbf{v} , all projections $P(\hat{\mathbf{u}}, \mathbf{v})$ are weighted equally). This implies, in relation to Eq. (2.11), that Colsher's filter has the form

$$H(\hat{\mathbf{u}}, \mathbf{v}) = \frac{1}{\int_{\Omega} \delta(\hat{\mathbf{u}}.\mathbf{v}) d\hat{\mathbf{u}}}$$
(2.15)

which can also be obtained by simply setting $G(\hat{\mathbf{u}}, \mathbf{v}) = 1$ in Eq. (2.13).



Fig. 2.4: The backprojection of a 1D parallel projection onto a 2D image.

since, by the convolution theorem, a multiplication in Fourier space is equivalent to a convolution in real space. In fact, due to presence of fast Fourier transforms (e.g. see [53]), the filtering step is commonly performed via multiplication in the Fourier space, followed by an inverse Fourier transform. Step 2: The filtered projections $p^F(\hat{\mathbf{u}}, \mathbf{s})$ are then backprojected to obtain the reconstructed image $f(\mathbf{x})$

$$f(\mathbf{x}) = \int_{\Omega} p^F(\hat{\mathbf{u}}, \mathbf{x} - (\mathbf{x}, \hat{\mathbf{u}})\hat{\mathbf{u}}) d\hat{\mathbf{u}}$$
(2.18)

This is shown in Fig. (2.4). For a given position \mathbf{x} in the image domain, the backprojection step is an integral over all directions $\hat{\mathbf{u}}$ which contribute to that position in the image. Using the aforementioned steps, $f(\mathbf{x})$ may thus be reconstructed for all \mathbf{x} in the image-domain.

The weighting function, or filter $H(\hat{\mathbf{u}}, \mathbf{v})$ is referred to as *factorisable* if it can be written as the product

$$H(\hat{\mathbf{u}}, \mathbf{v}) = H'(\mathbf{v})w(\hat{\mathbf{u}}) \tag{2.19}$$

Substituting this into the normalization condition (2.12), one arrives at

$$H'(\mathbf{v}) = \frac{1}{\int_{\Omega} w(\hat{\mathbf{u}})\delta(\hat{\mathbf{u}}.\mathbf{v})d\hat{\mathbf{u}}}$$
(2.20)

from which it follows that $H(\hat{\mathbf{u}}, \mathbf{v})$ is uniquely determined by $w(\hat{\mathbf{u}})$:

$$H(\hat{\mathbf{u}}, \mathbf{v}) = \frac{w(\hat{\mathbf{u}})}{\int_{\Omega} w(\hat{\mathbf{u}}) \delta(\hat{\mathbf{u}}.\mathbf{v}) d\hat{\mathbf{u}}}$$
(2.21)

Incidentally, this is what one would obtain by using $G(\hat{\mathbf{u}}, \mathbf{v}) = w(\hat{\mathbf{u}})$ (i.e. a function $G(\hat{\mathbf{u}}, \mathbf{v})$ independent of the frequency \mathbf{v}) in Eq. (2.13). In the case of using a factorisable filter, it can be shown [52] that the filtering and backprojection steps in the FBP approach may be commuted, so that the backprojection step is performed first. Thus we have:

Step 1: Backprojection

$$b_{\Omega}(\mathbf{x}) = \int_{\Omega} p(\hat{\mathbf{u}}, \mathbf{x} - (\mathbf{x}.\hat{\mathbf{u}})\hat{\mathbf{u}})w(\hat{\mathbf{u}})d\hat{\mathbf{u}}$$
(2.22)

Step 2: 3D filtering in image-space

$$f(\mathbf{x}) = \int_{R^3} h'(\mathbf{x} - \mathbf{x}') b_{\Omega}(\mathbf{x}') d\mathbf{x}'$$
(2.23)

The aforementioned BPF approach is therefore more limited than the direct Fourier and FBP reconstruction techniques, as it is only applicable when using factorisable filters. Nevertheless, in list-mode reconstruction (see chapter 4) in which measured events are not histogrammed, and are rather recorded and processed individually as they are measured, one does not have

access to projection data $p(\hat{\mathbf{u}}, \mathbf{s})$, and it is therefore not possible to filter the projections. Alternatively, one may backproject separately the individual list-mode events, after which filtering may be performed. In analytic list-mode image reconstruction, therefore, the BPF approach is the only practical one.

Comparison of Techniques and Comment on Discretization Issues

Amongst the three aforementioned methods of inverting the 3D X-ray transform, the first method (direct Fourier reconstruction), involves the use of 2D (fast) Fourier transforms of the projection data which upon integration using Eq. (2.11) can evaluate $F(\mathbf{v})$, thus yielding the reconstructed image by a final 3D inverse (fast) Fourier transform. The two other techniques (FBP and BPF), however, involve the use of computationally expensive 3D backprojection operations. The first approach is therefore expected to be more efficient. Nevertheless, the task of sampling $F(\mathbf{v})$, at appropriate frequencies prior to taking the 3D inverse Fourier transform, may likely involve interpolation of the available data [47], and can therefore be rather complex and time-consuming.

At the end of this section, we make one additional important note. Implementing backprojection operations as employed in FBP and BPF reconstruction techniques requires⁶ some kind of discretization in the image function $f(\mathbf{x})$. Nevertheless, these techniques are derived from the continuous line integral model, which therefore does not take into consideration discretized objects. In Secs. (2.4.1) and (2.4.2), we shall discuss how, by contrast, statistical reconstruction techniques intrinsically incorporate discrete images in their formulations, and thus, can potentially regress discretization artifacts.

50

 $^{^{6}}$ An exception to this is the 'natural pixel reconstruction' approach [54] in which the continuity of the object is intrinsically embedded in the reconstruction task.

2.2.8 Reconstruction from Truncated, 2D Parallel Projections

In practice, measured 3D PET data are often truncated, as discussed at the beginning of Sec. (2.2.1). Thus, one needs to consider the reconstruction of a function from the measurement of its 3D X-ray transform within the set

$$T_{\text{trunc}}(\Omega) = \{ (\hat{\mathbf{u}}, \mathbf{s}) | \ \hat{\mathbf{u}} \in \Omega, \mathbf{s} \in \hat{\mathbf{u}}_m^{\perp} \}$$
(2.24)

where for a given direction $\hat{\mathbf{u}}$, measurements are only made on a subset $\hat{\mathbf{u}}_m^{\perp}$ of the projection plane $\hat{\mathbf{u}}^{\perp}$ (Eq. 2.2).

Three approaches to the problem of reconstructing truncated, 2D parallel projections have been considered in the literature, and are summarized in Appendix A. The first two, unlike the third, are based on the assumption that the set of measured line integrals $T_{\text{trunc}}(\Omega)$ contains a subset T_0 from which $f(\mathbf{x})$ can be reconstructed (i.e. it satisfies Orlov's condition). Of course, in the absence of noise, the reconstruction from T_0 is exact; nevertheless, due to introduction of noise into the data in practice, the challenge remains to utilize the redundant data to improve the signal-to-noise ratio.

2.3 Rebinning Algorithms

The aforementioned analytic 3D reconstruction techniques are exact, but are typically more than an order of magnitude slower than the slice-by-slice reconstruction of data acquired in the 2D mode (see Sec. 1.3). In this regard, various approximate or exact techniques, commonly referred to as *rebinning algorithms*, have been proposed which aim to sort (*rebin*) the 3D-acquired data into a stack of ordinary 2D data sets. These rebinned data are geometrically equivalent to data collected in the conventional 2D mode, and therefore can be reconstructed using 2D reconstruction algorithms. This is summarized in Fig. (2.5). In this section, we shall present an overview of various existing rebinning algorithms.


Fig. 2.5: The rebinning procedure is summarized (courtesy of Dr. M. Defrise, Vrije Universiteit Brussel).

In this part, we choose to parametrize an LOR using the coordinates $(s, \phi, z, \delta = \tan \theta)$ where s and ϕ are the radial and azimuthal angular coordinates, z is the axial coordinate of the point midway between the two detectors, and θ is the copolar angle between the LOR and the transaxial planes (see Sec. 1.4). These parameters were illustrated in Fig. (1.9). Rebinning algorithms seek a mapping from the measured line integrals $p(s, \phi, z, \delta)$ along the LORs, onto a rebinned, 2D data set consisting of $p_{\text{reb}}(s, \phi, z)$, wherein δ is omitted since θ , and therefore δ , vanish along the transaxial planes.

2.3.1 The SSRB and MSRB Algorithms

The first rebinning algorithm introduced in 1987 by Daube-Witherspoon and Muehllehner [55] is very straight-forward. It makes the following approximation:

$$p_{\text{reb}}(s,\phi,z) \simeq p(s,\phi,z,\delta)$$
 (2.25)

Let us consider the common case of having a PET scanner with data truncated in the axial direction: the δ values are then restricted to within an interval $[0,\delta_{max}]$. Noting that the left-hand side of the aforementioned approximation is independent of δ , one can improve the signal-to-noise ratio by averaging all available estimates of $p_{\text{reb}}(s,\phi,z)$; i.e.

$$p_{\rm reb}(s,\phi,z) \simeq \frac{1}{\delta_{max}} \int_0^{\delta_{max}} p(s,\phi,z,\delta) d\delta$$
 (2.26)

This is commonly referred to as the *Single-Slice ReBinning (SSRB)* approximation. The algorithm is very simple to implement: it can be performed one-by-one for the acquired sinogram bins, and simultaneous consideration of the entire projection set is not required. However, the algorithm is only exact for radiation originating from the center of the LOR along which they are detected (see Fig. 2.6a), and therefore can result in significant axial blur-



Fig. 2.6: (a) Single-slice rebinning (SSRB): Each oblique $(\theta > 0)$ event is assigned to the direct $(\theta = 0)$ sinogram centrally located between the two detectors. (b) Multiple-slice rebinning (MSRB): Each oblique event is fractionally allocated to direct sinograms that it traverses.

ring, especially for regions away from the transaxial center of the scanner, in which case use of axial deconvolution has been suggested [56].

Alternatively, the Multi-Slice ReBinning (MSRB) algorithm [57] rebins the oblique ($\delta \neq 0$) LORs into multiple direct ($\theta = 0$) LORs: this is done by allocating a fraction of the counts received along each LOR to each of the LORs (with same transaxial coordinates s and ϕ) whose z-values are traversed by the LOR. This is depicted in Fig. (2.6b). This algorithm can be more accurate than the SSRB algorithm for points distant from the transaxial center of the FOV; nevertheless, it suffers from a systematic blurring of the rebinned data (which can to some extent be compensated by use of axial filtering [57], at the expense of increased noise).

2.3.2 The FOREX Algorithm

A series of rebinning algorithms based on Fourier techniques were proposed in the late '90s, which we summarize below. Defining the 2D Fourier transform of a given sinogram corresponding to a fixed pair (z, δ) as

$$P(w,k,z,\delta) = \int_{-\infty}^{\infty} \int_{0}^{2\pi} e^{-isw} e^{-ik\phi} p(s,\phi,z,\delta) ds d\phi \qquad (2.27)$$

and the 3D Fourier transform of the data as

$$\mathcal{P}(w,k,w_z,\delta) = \int_{-\infty}^{\infty} e^{-izw_z} P(w,k,z,\delta) dz \qquad (2.28)$$

Defrise et al. were able in 1996 [58] to derive an exact rebinning relation

$$\mathcal{P}_{\text{reb}}(w,k,w_z) = e^{ik\sigma} \mathcal{P}(w\chi,k,w_z,\delta)$$
(2.29)

where

$$\sigma = \sin^{-1}\alpha \tag{2.30}$$

$$\chi = \sqrt{1 - \alpha^2} \tag{2.31}$$

and

$$\alpha = \frac{\delta w_z}{w} \tag{2.32}$$

The reader is referred to Ref. [59] for a more compact and elegant derivation of Eq. (2.29). This formulation can be directly used to rebin the 3D data into 2D data sets, referred to as the *FOurier REbinning eXact (FOREX)* algorithm. Details of the algorithm and its implementation are discussed in Refs. [58, 59]. The difficulty with this algorithm is that it deals with 3D Fourier transforms of the projection data, and therefore cannot be directly applied to axially truncated data; therefore, the truncated data must first be estimated, which can be time-consuming.

On the other hand, by considering Taylor expansions of Eq. (2.29) around $\alpha = 0$, one can arrive at a class of approximate rebinning algorithms of increasing accuracy. The algorithms incorporating α^0 and α^1 terms are discussed below.

By setting $\alpha = 0$ in Eq. (2.29), we arrive at the zeroth order approximation

$$\mathcal{P}_{\text{reb}}(w,k,w_z) \simeq \mathcal{P}(w,k,w_z,\delta) \tag{2.33}$$

for which a 3D inverse Fourier transform can be performed explicitly to arrive at:

$$p_{\text{reb}}(s,\phi,z) \simeq p(s,\phi,z,\delta)$$
 (2.34)

which incidentally reproduces the SSRB approximation (2.25).

2.3.3 The FORE Algorithm

Keeping Taylor expansion terms which are linear in α (i.e. $\sigma \simeq \alpha$ and $\chi \simeq 1$), one obtains

$$\mathcal{P}_{\text{reb}}(w,k,w_z) \simeq e^{ik\frac{\delta w_z}{w}} \mathcal{P}(w,k,w_z,\delta)$$
(2.35)

It is possible to explicitly calculate the inverse 1D Fourier transform of (2.35) with respect to w_z :

$$P_{\text{reb}}(w,k,z) \simeq \int_{-\infty}^{\infty} e^{izw_z} e^{ik\frac{\delta w_z}{w}} \mathcal{P}(w,k,w_z,\delta) dw_z$$
$$\simeq \int_{-\infty}^{\infty} e^{iw_z(z+\frac{k\delta}{w})} \mathcal{P}(w,k,w_z,\delta) dw_z \qquad (2.36)$$

from which one arrives at

$$P_{\text{reb}}(w,k,z) \simeq P(w,k,z+\frac{k\delta}{w},\delta)$$
 (2.37)

Averaging the right-hand side over all measured δ values, we arrive at the FOurier REbinning (FORE) algorithm:

$$P_{\text{reb}}(w,k,z) \simeq \frac{1}{\delta_{max}} \int_0^{\delta_{max}} P(w,k,z+\frac{k\delta}{w},\delta) d\delta$$
(2.38)

Two key properties of the FORE technique, compared to the FOREX algorithm, can be summarized as follows: 1) FORE does *not* require estimation of axially truncated data: this is because it does not require taking an axial Fourier transform with respect to zand therefore the data need no longer be known for all values of z.

2) Implementation involves a 1D interpolation in z, but does not require an interpolation in the frequency variable w (unlike FOREX).

A nice geometric interpretation of the FORE algorithm is possible [58]: making the substitution $z + \frac{k\delta}{w} \rightarrow z$ in Eq. (2.37), for a given sinogram in the frequency domain (w, k) corresponding to fixed (z, δ) , we have

$$P(w,k,z,\delta) \simeq P_{\text{reb}}(w,k,z-\frac{k\delta}{w})$$
 (2.39)

This can be interpreted by saying that the value of P at the frequency (w, k) receives contribution mainly from sources located at a distance $k\delta/w$ below the axial center of the LOR (i.e. z). An example of this is depicted in Fig. (2.7). This approximate property can be viewed as a "virtual time-offlight principle" since it provides information about the location of a source along the LORs, which is only possible in time-of-flight tomography. However, it must be noted that this technique requires the use of all measured LORs (s, ϕ) in the sinogram (z, δ) in order to take 2D Fourier transforms (unlike true time-of-flight tomography which determines the location of a source on an individual LOR basis).

Furthermore, it must still be kept in mind that FORE is an approximate rebinning method, expected to result in image distortions when the maximum angle between the LORs and the transaxial slices, the *angular aperture*, exceeds a certain threshold. This is because the technique is based on firstorder Taylor expansion of the exact relation (2.29) about $\alpha = \delta w_z/w$ which increases with increasing acceptance angles. The angular aperture above which the FORE approximation exhibits artifacts is observed to be around 25° [60] at signal-to-noise levels and spatial resolutions typical of clinical



Fig. 2.7: Fourier rebinning (FORE): In the frequency domain, data from an oblique sinogram are allocated to a direct sinogram axially located $k\delta/w$ below the central z value, as described by Eq. (2.39).

studies.

2.3.4 The FOREJ Algorithm

In 1999, Defrise and Liu [61] proposed an original approach to rebinning making use of the fact that the 3D X-ray transform of a function is a solution to the second-order partial differential equation first studied by John in 1938 [62] (thus the 'J' in 'FOREJ'). For sinogram data $p(s, \phi, z, \delta)$ obtained in multi-ring PET scanners, John's equation can be written in the form

$$\frac{\partial^2 p(s,\phi,z,\delta)}{\partial z \partial \phi} + \frac{\partial^2 p(s,\phi,z,\delta)}{\partial \delta \partial s} = -s\delta \frac{\partial^2 p(s,\phi,z,\delta)}{\partial^2 z}$$
(2.40)

Using this relation, the authors were able to arrive at the FOREJ exact rebinning relation

$$P_{\text{reb}}(w,k,z) = \frac{1}{\delta_{max}} \int_{0}^{\delta_{max}} \{P(w,k,z+\frac{k\delta}{w},\delta) + \frac{\delta(\delta_{max}-\delta)}{w} \frac{\partial^{3}P(w,k,z+\frac{k\delta}{w},\delta)}{\partial w \partial z^{2}} \} d\delta \qquad (2.41)$$

The first term of the algorithm reproduces the FORE approximation (2.38), while the addition of the second term is what makes the algorithm exact. It can thus be seen that the FOREJ algorithm combines the advantages of the FOREX and FORE algorithms in that it is exact, and yet, it can be applied directly to axially truncated data (since it does not require performance of a Fourier transform in the axial z direction). However, it must be kept in mind that the second term in the algorithm involves a second derivative with respect to z, which can strongly amplify noise artifacts in the reconstruction task.

2.4 Statistical Image Reconstruction

Analytic image reconstruction techniques have a number of attractive properties: they (i) are able to perform the reconstruction task in few steps and efficiently, (ii) exhibit linearity, and (iii) do not require parameter specification, except for the degree of smoothing which can be optimized easily for a given desired statistical level.

However, (i) such techniques suffer from the approximations implicit in the line integral model on which the reconstruction formulas are based (see Sec. 2.2). For instance, one may observe that the line integral formulation (2.1) makes the fundamental assumption that the annihilation events occur along the LOR defined by the measurement, thus ignoring positron range and photon non-collinearity.

(ii) Furthermore, analytic techniques are based on the assumption that the data are *consistent*: i.e. sum of all projection line integrals in any direction is constant:

$$\int p(\hat{\mathbf{u}}, \mathbf{s}) d\mathbf{s} = \text{constant} \quad \text{for all} \hat{\mathbf{u}}$$
 (2.42)

Nevertheless, this assumption is not true in practice since presence of statistical noise can easily render the data inconsistent (this also being the case when noise artifacts are introduced in data correction steps, such as compensating for attenuation).

(iv) Moreover, analytic techniques are applicable on an assumption of shift-invariance, whereas imaging systems commonly exhibit shift-variant properties.

(v) Finally, analytic reconstruction techniques require careful consideration of data truncation, such as axial trucation and presence of gaps (see footnote 14 of chapter 1 and Sec. 2.2.8). On the contrary, this is not an issue in statistical techniques, due to the intrinsic modeling of the relation between the image and projection space.

In this regard, statistical image reconstruction (SIR) techniques aim to incorporate a more complex modeling of the measurement process, so as to improve image quality in the reconstruction tasks. In general, SIR methods require five components [63]:

1) A finite parametrization of the imaged object.

2) A system matrix that models and relates the elements in the unknown image to the expected counts along each LOR.

3) A statistical model for how the LOR measurements vary around their expectation.

4) An objective function to be maximized in order to find the image estimate.

5) An algorithm for maximizing the objective function.

Below, we present an overview of these components as found in the literature.

2.4.1 Object Parametrization

As noted in Sec. (2.2.7), backprojection operations employed in analytic reconstruction techniques require, in order to be implemented, a discretization of the image space. However, these techniques are derived from the continuous line integral model, which does not take into consideration presence of voxels and thus cannot attribute differing values to voxels that are not entirely within the tube connecting a particular detector-pair: this can ultimately produce severe artifacts. However, statistical reconstruction techniques, as originally observed by Shepp and Vardi [64], take as their starting position discrete models such as Eqs. (2.45) and (2.46), and can therefore use the discrete system matrix formulation designed to model relation of voxels to LORs to regress such artifacts, as we explain next.

In order to finitely parametrize the object, most typically, the image domain is divided into rectangular *voxels*. Use of smoother bases (e.g. blobs) have also been suggested [65,66]. Blobs are spherically symmetric functions which handle interpolation issues more effectively than voxels. Nevertheless, voxels have the practical advantage of containing minimal support (i.e. no overlap), resulting in maximal sparseness in the system models, which in turn, relaxes computation times. Throughout this work, the voxel approach has been chosen.

For the continuous distribution function $f(\mathbf{x})$, one may introduce the parametrization

$$f(\mathbf{x}) = \sum_{j=1}^{J} \lambda_j b_j(\mathbf{x})$$
(2.43)

where the image has now been divided into J voxels, λ_j is the mean activity in the *j*th voxel, and b_j represents the *j*th voxel.

2.4.2 The System Matrix

First, let us consider a model for coincidence detection. Assuming a radiotracer distribution given by function $f(\mathbf{x})$, the expectation of the measured events along a certain LOR can be written as:

$$\bar{n}_i = \int \tilde{p}_i(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}$$
(2.44)

where $\tilde{p}_i(\mathbf{x})$ represents the probability distribution of a positron generated at a position \mathbf{x} resulting in an annihilation detected along the LOR *i*. Note that this model does not assume a line integral formulation, and therefore can potentially avoid some of the pitfalls of analytic image reconstruction.

Incorporating object parametrization (2.43) into Eq. (2.44), one arrives at

$$\bar{n}_i = \sum_{j=1}^J p_{ij} \lambda_j \tag{2.45}$$

where

$$p_{ij} = \int \tilde{p}_i(\mathbf{x}) b_j(\mathbf{x}) d\mathbf{x}$$
 (2.46)

Therefore, p_{ij} is the probability of detecting a photon pair along LOR *i* produced from a positron generated at voxel *j*. The matrix $\hat{P}=(p_{ij})_{I\times J}$, often referred to as the system matrix, has the ability to model a wide variety of effects.

A variety of physical processes which are not considered in the line integral model, can instead be incorporated into the system matrix, as we describe below. One may decompose the system matrix $\hat{P}=(p_{ij})_{I\times J}$ as such [67,70]:

$$\hat{P} = \hat{W}\hat{G}\hat{B} \tag{2.47}$$

The matrix $\hat{B}=(b_{ij})_{J\times J}$ is introduced to account for voxel-to-voxel blurring, or spatial resolution, effects. It is therefore best suited for modeling of positron range. The operator $\hat{G}=(g_{ij})_{I\times J}$ contains the elements g_{ij} used to calculate the geometric probability that a voxel j is detected along an LOR i, and as described above, can assign partial values to voxels that are not entirely within a given LOR. Furthermore, this operator can be used to model photon non-collinearity; that is, the uncertainties in the angular separation of the photon pair can be incorporated into the probability that an event generated in a voxel j is detected along an LOR i. The matrix $\hat{W} = (w_{ij})_{I \times I}$, as elaborated by Ref. [68], allows modeling of LOR interactions:

(i) In the original algorithms, the measured data n_i were first corrected for attenuation and normalization effects prior to use in the reconstruction task. However, it was later proposed by Hebert and Leahy [69] that these corrections could also be included as weighting factors in the system matrix. This can be accomplished by noting that attenuation and normalization factors can be inserted in the diagonal elements of the matrix $\hat{W}=(w_{ij})_{I\times I}$, thereby allowing a weight to be assigned to each LOR to account for these effects.

(ii) It is also possible [68] to incorporate into the matrix \hat{W} the probability that a photon incident on a particular crystal is detected in another crystal. In other words, this formulation can be used as a natural and accurate framework to model presence of detector blurring (Sec. 1.6.7), though this can highly complicate the calculation (as it renders the system matrix highly non-sparse). In Appendix B, we describe an *ad hoc* method in which detector blurring (including the parallax effect) are modeled in the image domain, so as to yield practical and feasible system matrices.

As already mentioned, the photon non-collinearity effect should be modeled into the matrix \hat{P} , however this leads to an increase in the number of sinogram elements to which each voxel contributes, and hence renders the system matrix highly non-sparse. As an approximation, one can assume that this effect is depth-independent and include it in the matrix \hat{W} [68], or even more simply to include it as a finite resolution effect in the spatial matrix \hat{B} [70].

2.4.3 Statistical Modeling and Objective Functions

The statistical model describes the distribution of each measurement about its mean, and therefore provides a measure of similarity between the actual measurements n_i and the expected counts \bar{n}_i , which can be calculated using Eq. (2.45) given a current estimate for the voxel intensities. We first describe the formulation for an idealized PET system in which effects of random coincidences and scattered events are neglected. This is evident in the original derivation of the ML-EM algorithm by Shepp and Vardi, wherein they emphasize: "We shall simply be forced to ignore these problems and to assume that the only source of difficulty is in the statistical fluctuations in the counting statistics" [64].

We start by noting that the counts emitted in a voxel j such that they are detected along an LOR i follow a Poisson distribution with mean $p_{ij}\lambda_j$. Since a variable obtained by summing a number of Poisson variables is also Poissondistributed, it then follows that the distribution of total counts along any LOR i (which is the quantity we measure) must also be Poisson-distributed with mean

$$\bar{n}_i = \sum_{j=1}^J p_{ij} \lambda_j \tag{2.48}$$

Defining $\vec{\lambda} = [\lambda_1 ... \lambda_J]^T$ and $\vec{n} = [n_1 ... n_J]^T$ as vectors of image intensity and measured LOR counts, respectively, we note that the task is to reconstruct the former from the latter. One can define the Poisson likelihood function [64,71,72]:

$$L(\vec{n} \mid \vec{\lambda}) = \prod_{i=1}^{I} e^{-\bar{n}_i} \frac{\bar{n}_i^{n_i}}{n_i!}$$
(2.49)

where \bar{n}_i is given by Eq. (2.48). The task is then to find a vector $\vec{\lambda}$ producing \bar{n}_i values such that they maximize the Poisson likelihood function $L(\vec{n} \mid \vec{\lambda})$. Alternatively, one can maximize the log likelihood function $l(\vec{\lambda}) = \ln L(\vec{n} \mid \vec{\lambda})$ which can be written in the form:

$$l(\vec{\lambda}) = \sum_{i=1}^{I} \{ -\bar{n}_i + n_i \ln(\bar{n}_i) \}$$
(2.50)

where the term $\ln(n_i!)$ has been dropped since it has no dependence on

 λ_j , and therefore has no effect on the maximization task. This approach to image reconstruction is often referred to as *maximum likelihood* (ML) reconstruction.

The Ordinary Poisson Objective Function

Turning our attention to corrections for randoms and scattered events in the reconstruction task, we note that as mentioned in Sec. (1.6.3), for PET scanners with the delayed-coincidence measurement capability, distribution of random counts can be directly measured along with the prompts. In this regard, the measured delayed-coincidences are often subtracted from the measured prompts to obtain an estimate of the number of true coincidence events.

A problem with this delayed-coincidence subtraction technique is that the pre-corrected data do not follow Poisson statistics. This is because subtraction of delayed coincidences from measured prompts compensates for the coincidence events in terms of the mean but increases the variance. In this regard, the objective function (2.50) applied to data pre-corrected for randoms is not accurate. Defining \bar{r}_i and \bar{s}_i as the *expected* randoms and scattered events contributions along any LOR *i*, one notes that since the *prompt* events (not pre-corrected for randoms) are Poisson by nature, one can consider the log-likelihood function [73]

$$l(\vec{\lambda}) = \sum_{i=1}^{I} \left\{ -(\bar{n}_i + \bar{r}_i + \bar{s}_i) + n_i^p \ln(\bar{n}_i + \bar{r}_i + \bar{s}_i) \right\}$$
(2.51)

where n_i^p is the number of measured prompt events along an LOR *i*. To arrive at this objective function, we have noted that the expected number of prompt counts along an LOR *i* is given by $\bar{n}_i + \bar{r}_i + \bar{s}_i$, where \bar{n}_i is the expected number of true coincidences along the LOR given by Eq. (2.48). Eq. (2.51) is often referred to as the ordinary Poisson (OP) log-likelihood function. It must be emphasized here that \bar{r}_i in Eq. (2.51) is an *expected* value which can *not* be appropriately replaced by the *measured* delayed coincidences, rather estimates of mean random counts along the LORs must be used. To address this issue, two approaches are possible: i) using singles measurements at the detectors [74] to calculate the expected randoms contribution, or ii) variance reduction (smoothing) for the measured delayed events [75–77]. Advantages and limitations of these approaches will be discussed in Sec. (3.5.1).

Randoms Pre-corrected PET

It is the case in many PET scanners that the prompt data are pre-corrected for the detection of randoms by *real-time* subtraction of the delayed coincidences, intended to minimize data transfer and processing times. Let us refer to the output data as n'_i . In this case, the OP log-likelihood can be written in the form:

$$l(\vec{\lambda}) = \sum_{i=1}^{I} \{ -(\bar{n}_i + \bar{r}_i + \bar{s}_i) + (n'_i + \bar{r}_i) \ln(\bar{n}_i + \bar{r}_i + \bar{s}_i) \}$$
(2.52)

Here, the term $(n'_i + \bar{r}_i)$ is used as a measure of the prompts data (which, for these scanners, are only output after subtraction of delayed coincidences). However, the problem with this approach can be seen as follows. Let us refer to the prompts and delayed coincidence counts as n_i^p and n_i^d , which are Poisson variables. The output data are given by

$$n_i' = n_i^p - n_i^d \tag{2.53}$$

We now recall that for a Poisson variable X, the mean is equal to the variance (i.e. E(X)=Var(X)). Furthermore, we note that

$$\mathcal{E}(n_i^p) = \bar{n}_i + \bar{r}_i + \bar{s}_i = \operatorname{Var}(n_i^p) \tag{2.54}$$

and

$$\mathbf{E}(n_i^d) = \bar{r}_i = \operatorname{Var}(n_i^d) \tag{2.55}$$

It then follows that

$$E(n'_{i} + \bar{r}_{i}) = E(n^{p}_{i}) - E(n^{d}_{i}) + \bar{r}_{i} = \bar{n}_{i} + \bar{r}_{i} + \bar{s}_{i}$$
(2.56)

and

$$\operatorname{Var}(n'_{i} + \bar{r}_{i}) = \operatorname{Var}(n^{p}_{i}) + \operatorname{Var}(n^{d}_{i}) + 0 = \bar{n}_{i} + 2\bar{r}_{i} + \bar{s}_{i}$$
(2.57)

where we have noted that $E(\bar{r}_i)=\bar{r}_i$ and $Var(\bar{r}_i)=0$ since \bar{r}_i is a constant random variable. It is therefore clear that the term $(n'_i + \bar{r}_i)$ matches the prompts counts n^p_i in mean, but is broader in variance, and therefore does not follow Poisson statistics.

Due to the non-Poisson nature of the output data, other general approaches have also been employed:

(i) The algebraic reconstruction technique (ART) and its variations: Proposed first in 1970 by Gordon *et al.* [78] for 3D electron microscopy and x-ray photography, Herman and Meyer [79] were first to apply this technique to the PET problem. Mathematically, ART aims at solving the set of linear equations determined by the projection data (if consistent). If the system of equations is not consistent (the practical case), then, by introducing appropriate relaxation parameters, the algorithm approximates a weighted least squares solution [80]. Several variations of the ART algorithm, such as the multiplicative ART (MART), have also been proposed [78,81] (see also [115] for other variations.

(ii) Gaussian approximation of the data: these techniques results in (weighted) least squares penalty terms [63].

For scans that are *not* pre-corrected for randoms, the aforementioned methods are suboptimal because they do not fully accommodate the Poisson nature of the data. The ART algorithm, and its variations, do not at all incorporate any modeling of the statistical distribution of the data: an exception to this is the RAMLA algorithm, which is derived from the ART approach to reconstruction, but can be shown to converge to the maximum of the Poisson likelihood (2.49). However, to guarantee convergence, a relaxation parameter is needed which must be controlled and changed with every iteration to satisfy certain criteria. Similarly, especially in high resolution PET in which individual LORs often receive sufficiently low counts, the Gaussian approximation to the Poisson distribution is not accurate.

(iii) For scans that *are* pre-corrected for randoms, another approach is the saddle-point (SD) approximation technique (obtained by a second-order Taylor series expansion in the Poisson-modeling). The reader is referred to Refs. [82, 83] for more detail. According to [84], to date, the SD approach has been able to out-perform other approaches to reconstruction of randomsprecorrected data, but is relatively more difficult to implement.

(iv) Another approach, easy to implement, which still makes use of Poissonmodeling has been proposed by Yuvaz and Fessler [85]. One may notice that the variable $(n'_i + 2\bar{r}_i)$ exhibits a mean that matches its variance; i.e.

$$E(n'_{i} + 2\bar{r}_{i}) = E(n^{p}_{i}) - E(n^{d}_{i}) + 2\bar{r}_{i} = \bar{n}_{i} + 2\bar{r}_{i} + \bar{s}_{i}$$
(2.58)

and

$$\operatorname{Var}(n'_{i} + 2\bar{r}_{i}) = \operatorname{Var}(n^{p}_{i}) + \operatorname{Var}(n^{d}_{i}) + 0 = \bar{n}_{i} + 2\bar{r}_{i} + \bar{s}_{i}$$
(2.59)

where we recall again that \bar{r}_i is a *constant* random variable, and therefore has zero variance. One may then attempt to maximize the so-called *shifted Poisson* (SP) log-likelihood function defined about the variable $(n'_i + 2\bar{r}_i)$:

$$l(\vec{\lambda}) = \sum_{i=1}^{I} \left\{ -(\bar{n}_i + 2\bar{r}_i + \bar{s}_i) + (n'_i + 2\bar{r}_i) \ln(\bar{n}_i + 2\bar{r}_i + \bar{s}_i) \right\}$$
(2.60)

Nevertheless, it must be noted that while the term $(n'_i + 2\tilde{r}_i)$ matches the true Poisson variable $(n^p_i + n^d_i)$ in mean and variance, it is still not exactly a true Poisson variable, and this approach, while superior to the OP approach for the aforementioned scanners, is still an approximation [85]. In this regard, designing PET scanners with the capability to separately store prompt and delayed events can result in more accurate reconstructions of the true image activity. The HRRT is an instance of modern PET scanners capable of such measurements. In this case, one may use the OP approach, which does not make any approximations in the statistical modeling.

2.4.4 Reconstruction Algorithms

Let us start by the Poisson log-likelihood function (2.50). Unfortunately, there is no closed-form expression for the estimate $\vec{\lambda}$ that maximizes the likelihood [86], which therefore necessitates iterative algorithms. There exist a number of algorithms that maximize the Poisson likelihood, such as the rowaction maximum likelihood algorithm (RAMLA) [87] and the paraboloidal surrogates (PS) methods [88–90]. To date, however, the most popular algorithm to maximize Poisson likelihoods remains to be the *expectation maximization* (EM) algorithm, as we further elaborate.

The EM algorithm as presented by Dempster, Laird and Rubin [91] is a general approach to iterative optimization of likelihood functions when the data can be formulated in a complete/incomplete framework. Such a framework is applicable when the problem has a more natural formulation in terms of a set of unobserved data. At each iteration, the EM approach requires two steps: an expectation step (E-step) followed by a maximization step (M-step). Often, these two steps are combined into one.

The EM algorithm was first applied to Poisson likelihood maximization in emission tomography by Shepp and Vardi [64], followed by Lange and Carson [71]. In [92], Parra and Barrett were able to demonstrate applicability of the EM algorithm to *list-mode* likelihood maximization; list-mode EM reconstruction algorithms shall be discussed in chapter 4.

The combination of the E- and M-steps in the case of the objective function (2.50) yields the iterative algorithm:

$$\lambda_{j}^{m+1} = \frac{\lambda_{j}^{m}}{\sum_{i=1}^{I} p_{ij}} \sum_{i=1}^{I} p_{ij} \left[\frac{n_{i}}{\bar{n}_{i}^{m}} \right]$$
(2.61)

where

$$\bar{n}_i^m = \sum_{j=1}^J p_{ij} \lambda_j^m \tag{2.62}$$

Here λ_j^m denotes the image intensity in voxel j (j=1...J) at the *m*th iteration. The sensitivity correction factor $s_j = \sum_{i=1}^{I} p_{ij}$ is a summation over all possible measurable LORs (i=1...I) and calculates the probability of an emission from voxel j being detected anywhere.

The algorithm is iterative in that it starts from an initial estimate of the image vector $\vec{\lambda}^m$, performs a forward projection step along all LORs to calculate the expected counts \bar{n}_i^m based on the current estimate for image intensity, takes the ratio between the measurement n_i and the expected counts \bar{n}_i^m for all the LORs, followed by a back-projection summation of these ratios onto the given voxel j, subsequently divided by the sensitivity correction factor s_j , and followed by a final multiplication with the old estimate λ_j^m from which a new estimate λ_j^{m+1} is calculated⁷.

It can be shown [71,93,94] that the estimates λ_j^m monotonically converge to the maximum of the Poisson likelihood function with increasing iterations. In other words, $l(\vec{\lambda}^{m+1}) > l(\vec{\lambda}^m)$ unless $\vec{\lambda}^m = \vec{\lambda}^{m+1}$ in which case we have converged to a solution that maximizes the Poisson likelihood function.

As described in Sec. (2.4.3), the original derivations did not directly incorporate presence of randoms and scattered events into the likelihood functions.

⁷ Notice that if λ_j^m is such that the expected counts \bar{n}_i^m are equal to the measurements n_i for all *i*, one then arrives at $\lambda_j^{m+1} = \lambda_j^m$ so that the new update is the same as the old, as it completely matches the data. However, this is very unlikely to take place due to the presence of Poisson noise in the data, which renders it inconsistent.

For instance, using the delayed-coincidence subtraction technique, data were (and are) often simply pre-corrected for randoms. However, the subtraction technique faces two important issues: First, data pre-corrected for randoms are no longer Poisson and therefore the modeling is not accurate, as elaborated in Sec. (2.4.3). This also means that the algorithm (2.61) is not truly EM when applied to data pre-corrected for randoms, and therefore, convergence is not guaranteed.

Second, the subtraction of delayed-events from the prompts (i.e. $n'_i = n^p_i - n^d_i$) can result in negative sinogram bins, which can be a source of potential problems. This can occur since the delayed-coincidences are random variables which are themselves separately measured and therefore are *not* exactly equal to the number of random events measured along with true coincidences⁸. Consequently, a particular instance of the delayed-coincidence measurement can yield values that are larger than the actual randoms, and can also result in negative histogram bins upon subtraction from the prompts. This issue has typically been dealt with by setting negative sinogram values to zero, due to the nonnegative nature of Poisson random variables and since negative sinogram values can cause reconstruction algorithms to diverge [84]. This zero-thresholding can however result in positive bias in the final reconstructed image, especially in low statistics studies.

For scanners that allow separate storage of prompts and delayed-coincidence events, the aforementioned issues can altogether be avoided via directly incorporating the presence of randoms (and scattered events) directly into the OP objective functions as described by Eq. (2.51). Defining $\bar{n}_i^m = \sum_{j=1}^J p_{ij} \lambda_{j}^m$ as before, application of the EM algorithm yields [73]

$$\lambda_j^{m+1} = \frac{\lambda_j^m}{\sum_{i=1}^I p_{ij}} \sum_{i=1}^I p_{ij} \left[\frac{n_i^p}{\bar{n}_i^m + \bar{r}_i + \bar{s}_i} \right]$$
(2.63)

⁸ This is why one may *not* use interchangeably the terms "randoms" and "delayedcoincidences": the two are different realizations (or instances) of the same Poisson random variable, and therefore are not necessarily equal.

which we shall refer to as the OP-EM algorithms.

For scanners in which the data n'_i are only output after real-time subtraction of delayed-coincidences n^d_i from the measured prompts n^p_i (i.e. $n'_i = n^p_i - n^d_i$), the data are no longer Poisson, and application of (2.61) (wherein $n_i = n'_i$) is therefore not suitable. It is instead possible to use the shifted-Poisson (SP) approximation (as described in the previous section). Application of the EM algorithm to the SP objective function (2.60) yields the SP-EM algorithm:

$$\lambda_j^{m+1} = \frac{\lambda_j^m}{\sum_{i=1}^I p_{ij}} \sum_{i=1}^I p_{ij} \left[\frac{n_i' + 2\bar{r}_i}{\bar{n}_i^m + 2\bar{r}_i + \bar{s}_i} \right]$$
(2.64)

It must be noted that it is still possible that $n'_i + 2\bar{r}_i < 0$, though it is less likely than $n'_i < 0$. This means that the SP-EM algorithm has the added advantage that upon introduction of zero-thresholding, it is less likely to introduce positive bias in the SP-EM algorithm as compared to when the regular EM algorithm (2.61) is used (while this problem is non-existent in OP-EM reconstruction since n^p_i is always non-negative).

In order to altogether eliminate any introduced positive bias for data from scanners that only output n'_i , Ahn and Fessler [84] have proposed likelihood approximations that allow negative sinogram values without requiring any zero-thresholding. The proposed techniques, however, can result in a very considerable computational demand. For instance, the modified ML-EM algorithm has sensitivity correction factors, normally computed only once, that depend on the current image estimate at every iteration! Alternatively, as elaborated in Sec. (4.2), we have implemented another delayed-subtraction approach based on relaxing the sinogram non-negativity condition.

2.5 Problems with Expectation Maximization Algorithms

Expectation maximization algorithms suffer from two main drawbacks: First, the problem of determining $\vec{\lambda}$ such that it maximizes $l(\vec{\lambda})$ is generally very ill-conditioned and results in image intensities which are usually extremely noisy and exhibit edge artifacts [95]. For a nice theoretical treatment of noise properties of the EM algorithm, Ref. [96] should be consulted. Second, a very large number of iterations may be needed before they converge. Below, we elaborate upon techniques used to address these two issues.

2.5.1 Importance of Convergent Reconstruction

Prior to discussing techniques to address (i) noise artifact and (ii) slow convergence issues encountered with regular EM reconstruction, we focus our attention on the issue of convergence itself. Monotonic convergence, in the strict statistical sense, was defined Sec. (2.4.4): the Poisson-likelihood function must satisfy $l(\vec{\lambda}^{m+1}) > l(\vec{\lambda}^m)$ unless $\vec{\lambda}^m = \vec{\lambda}^{m+1}$ in which case one has converged to the solution that maximizes the Poisson likelihood function. In a more practical sense, convergence can be defined as a constant improvement (with iterations) in figures of merit, such as resolution and contrast, which are more relevant to the clinical task. In our experimental studies on the HRRT, we have taken the latter approach in order to quantify convergent reconstruction techniques (see Sec. 4.3).

One can argue that convergence properties can be very relevant to clinical medical imaging, since algorithm divergence could have unfortunate consequences [97]. For instance, the theoretical analysis of Qi and Huesman [98] in lesion detectability (using statistical reconstruction vs. FBP) applies *only* to convergent algorithms, whereas no such justification has been provided for any non-convergent algorithms. In this regard, in the course of completing this dissertation, we have paid particular attention to the study of convergent.

gent reconstruction algorithms (applicable to the HRRT), as elaborated in Sec. (3.4). In the next two sections, we provide an overview of existing literature, and the various challenges, in addressing the aforementioned existing issues in regular EM reconstruction, with emphasis on limitations/strengths of the various proposed techniques in convergent reconstruction.

2.5.2 Tackling Noise Artifacts: Regularization

Several *regularization* techniques have been proposed to remedy the problem of noise artifacts of the EM algorithm. All techniques, categorized into three main headings below, effectively alleviate the noise problem with some degradation in resolution, with some techniques even resulting in non-uniform resolution as we described below:

1) Early termination: this approach uses the EM method alone and terminates the iteration at a stage prior to deterioration of the image quality using proposed criteria. Veklerov and Llacer [99], for instance, designed a stopping rule based on statistical hypothesis testing that determined the degree of contradiction between the task of likelihood maximization and the model assumption of Poisson-distributed data. Later, they introduced the concept of "feasibility" which indicated the physical consistency between the reconstructed images and the initial data [100]. A serious problem with stopping rules is that since it is known that different parts of an image converge at different rates⁹, one is therefore likely to obtain object-dependent (and hence non-uniform) resolution and noise characteristics.

2) Filtering techniques: Initially developed by Snyder and his colleagues as

⁹ There is existing theoretical literature on dependence of image quality on position within object (object-dependency) for the *final* image using the regularization techniques discussed in the next two parts (filtering and Bayesian techniques). Nevertheless, to the best of our knowledge, the object-dependency of resolution as a function of iteration has not been treated theoretically in the past. Empirically, of course, this has been demonstrated, e.g. see Ref. ([101]).

the method of sieves [95, 102], these techniques are used to obtain a blurred version of the (unregularized) image intensity. This is often implemented in the form of post-smoothing the final reconstructed image using Gaussian kernels. The real practical problem with this approach is that the (unregularized) algorithm has to be performed until convergence is reached prior to application of post-filtering. However, hundreds of iterations may be needed in order to reach convergence. One alternate approach is to use one of the accelerated reconstruction techniques, most of which exhibit uncertain convergence properties, prior to post-filtering.

An alternate approach to post-smoothing is to perform the filtering inbetween the iterations [70, 103, 104]. However, it can be shown (theoretically and experimentally) that use of spatially invariant inter-filtering (e.g. convolution of images obtained after every iteration with Gaussian kernels of fixed width) in many iterative algorithms, and in particular in EM reconstruction, results in non-uniform resolution that would depend on the radioactivity distribution in the image: more sophisticated filters are therefore required to allow for shift-invariant and object-independent image characteristics [105]. 3) Bayesian methods: Also referred to as penalized likelihood (PL) or maximum a posteriori probability (MAP) methods, these techniques are originally derived from a simple application of Bayes' rule to image reconstruction. The probability function of the data (or measurement) vector \vec{n} conditioned on the image vector $\vec{\lambda}$ can be formulated using Bayes' equation:

$$L(\vec{n} \mid \vec{\lambda}) = \frac{L(\vec{\lambda} \mid \vec{n})L(\vec{\lambda})}{L(\vec{n})}$$
(2.65)

Here $L(\vec{\lambda} \mid \vec{n})$ is the *likelihood* function, given by Eq. (2.49) for Poissondistributed data, whereas $L(\vec{n} \mid \vec{\lambda})$ is the *a posteriori* probability distribution of the image vector. In the maximum likelihood (ML) approach, the desired image vector is defined as the maximizer of $L(\vec{\lambda} \mid \vec{n})$, whereas in the MAP approach, one maximizes $L(\vec{n} \mid \vec{\lambda})$. In the Bayesian framework, specification of the prior $L(\bar{\lambda})$ allows some *a priori* information to be incorporated into the image reconstruction task.

Geman and McClure were first to apply this technique to regularized image reconstruction in nuclear medicine [106]. They used Gibbs priors [107] to ensure that values at nearby voxels did not differ too much from each other. To see how this works, we note that by taking the log of both sides of Eq. (2.65), and omitting $L(\vec{n})$, since it is a constant for a given data set, MAP reconstruction requires finding the image vector $\vec{\lambda}$ which maximizes

$$B(\vec{n} \mid \vec{\lambda}) = \ln L(\vec{\lambda} \mid \vec{n}) + \ln L(\vec{\lambda})$$
(2.66)

One can think of the last term as a penalty term which can be defined so as to decrease when nearby voxels differ too much from each other. Examples of this approach are illustrated in Refs. [108, 109]. Furthermore, it is possible to extend the definition of the prior such that separations are penalized only if the neighboring voxels are thought to be in the same region, without unduly penalizing the larger separations which one foresees occurring at the boundary between two different regions of the image (e.g. see [110, 111])

In contrast to the use of the EM algorithm with early termination, the Bayesian approach to regularized image reconstruction has the property that the objective function (through incorporation of the penalty term) is solely responsible for the final image quality, while the particular algorithm used to maximize the objective function has no effect on the obtained image quality¹⁰ (and thus can be varied to yield faster reconstruction techniques). Furthermore, there is evidence that MAP (or equivalently penalized likelihood) methods are able to outperform the method of sieves [112].

However, again similar to the inter-filtering approach, it can be shown that non-sophisticated penalty terms can lead to non-uniform and object-

¹⁰ In other words, one is able to separate the objective function from the algorithm used to maximize it: see the opinion page of Dr. J. Fessler for further details: http://www.eecs.umich.edu/ fessler/misc/opinion,1998.html

dependent resolution properties in the image reconstruction task, even for space invariant tomographs (i.e. tomographs that exhibit uniform resolution across the field-of-view) [113].

2.5.3 Accelerated EM Reconstruction Algorithms

One major issue with EM algorithms is that many iterations may be required for the reconstructed images to converge such that the likelihood objective function is maximized. In fact, soon after the introduction of the EM algorithm for PET reconstruction, attempts to accelerate the algorithm were proposed so as to render EM reconstructions practically feasible. In 1987, Kaufman [72] treated some of the numerical aspects of the EM algorithm and proposed ways of speedings up the numerical algorithm using some of the traditional techniques in numerical analysis.

It was the introduction by Hudson *et al.* [114] of the ordered subset expectation maximization (OSEM) algorithm for histogram-mode emission tomography which generated considerable interest in accelerated reconstruction techniques. In the OSEM approach, the data are divided into LOR-based subsets. Using S_l to denote the *l*th subset (l=1...L), and m (m=1, 2, ...) as the iteration number which is only completed after a thorough loop through all the *L* subsets in the data, the algorithm can be written as:

$$\lambda_{j}^{m,l} = \frac{\lambda_{j}^{m,l-1}}{s_{j}^{l}} \sum_{i \in S_{l}} p_{ij} \frac{n_{i}}{\sum_{b=1}^{J} p_{ib} \lambda_{b}^{m,l-1}}$$
(2.67)

where

$$s_j^l = \sum_{i \in S_l} p_{ij} \tag{2.68}$$

which can be similarly applied to OP- and SP-EM algorithms. Here, we note that the back-projection summation is performed over only a subset of the data at a time, allowing the image to be updated L times for every iteration through the data. OSEM algorithms are typically able to achieve an order of magnitude speed-up of the EM algorithm. The only (easy-to-handle) increased requirement of this algorithm is that the sensitivity correction factors s_j^l must be stored for all the subsets. This is in comparison to non-subsetized EM reconstruction wherein only the overall sensitivity factor $s_j = \sum_i^I p_{ij}$ is utilized.

Nevertheless, the OSEM algorithm is not a convergent algorithm in general. Hudson *et al.* [114] were able to prove convergence *only* for a very special case, referred to as the "subset balance" condition, in which case values of s_j^l were independent of the particular subset *l*. Alternately, since $\sum_{l=1}^{L} s_j^l = s_j$, this condition can be written as:

$$s_j^l = \frac{s_j}{L} \tag{2.69}$$

However, this assumption is very restrictive, and in practice, one often observes *limit cycles*, which are non-convergent and oscillatory variations in image likelihood as well as figures of merit (e.g. contrast, resolution) with further subsets and iterations into the data, as we also report in chapter 4.

Alternatively, Byrne [115, 116] has suggested a rescaled block¹¹-iterative EM maximum likelihood (RBI-EMML) algorithm:

$$\lambda_j^{m,l} = \left(1 - \left[\frac{s_j^l}{\rho^l s_j}\right]\right)\lambda_j^{m,l-1} + \left[\frac{s_j^l}{\rho^l s_j}\right]\frac{\lambda_j^{m,l-1}}{s_j^l}\sum_{i\in S_l}p_{ij}\frac{n_i}{\sum_{b=1}^J p_{ib}\lambda_b^{m,l-1}} \quad (2.70)$$

where

$$o^{l} = \max_{j} \left\{ \begin{array}{c} \frac{s_{j}^{l}}{s_{j}} \end{array} \right\}$$
(2.71)

In the particular case when the subset balance condition (2.69) is satisfied, s_j^l/s_j , and therefore ρ^l , equal 1/L. Subsequently, Eq. (2.70) would reduce to the OSEM algorithm (2.67) since $s_j^l/(\rho^l s_j)$ equals unity in this case. The RBI-EMML algorithm, similar to the OSEM algorithm, is fairly straightforward to implement, and simply requires an additional calculation of ρ^l values (which are independent of the emission data) for all the subsets.

¹¹ The terms *block* and *subset* are interchangeably used in the literature

However, the algorithm is convergent only in the feasible case: i.e. when there is a non-negative vector $\vec{\lambda}$ such that $\sum_{j=1}^{J} p_{ij}\lambda_j = n_i$ for all measured LORs. Nevertheless in typical cases, due to presence of Poisson noise, we do not expect there to be such a non-negative vector (and this is why maximum likelihood approaches are attractive in the first place), and one therefore arrives at limit cycles. "Feedback" techniques have been suggested [117] to extract information from the limit cycles to construct suitable approximate solutions.

It must also be noted that the RBI-EMML algorithm bears some similarity to reconstructions with a relaxation parameter as used by Lewitt and Muehllehner [118] and Reader *et al.* [119] to accelerate histogram-mode and list-mode EM reconstructions, respectively. In subsetized histogram-mode, this can be written as

$$\lambda_j^{m,l} = (1-\mu)\,\lambda_j^{m,l-1} + \mu \frac{\lambda_j^{m,l-1}}{s_j^l} \sum_{i \in S_l} p_{ij} \frac{n_i}{\sum_{b=1}^J p_{ib} \lambda_b^{m,l-1}} \tag{2.72}$$

Here, unlike the term $s_j^l/\rho^l s_j$ in Eq. (2.70), the relaxation parameter μ needs to be determined on an *ad hoc* basis, and is independent of the particular voxel j and subset l. We also note that this algorithm is not guaranteed to be convergent, even in the feasible case.

There has been interest by other research groups in deriving provably convergent versions of the fast OS methods. In [87], an alternate algorithm termed row-action maximum likelihood algorithm (RAMLA) was proposed along with a convergence proof. The authors have also introduced the block sequential regularized EM (BSREM) algorithm, which extends the RAMLA approach to the case of maximum a posteriori (MAP) reconstruction [174]. Two types of globally convergent relaxed ordered subsets algorithms were also presented in [90]: one by modifying the BSREM algorithm to yield an algorithm convergent under more realistic assumptions, and the other by relaxing the OS-SPS method (first introduced in [120]). One problem with these formulations is that they are controlled by a relaxation schedule to ensure convergence and that it can be difficult to determine these schedules such that they lead to fast algorithms while simultaneously satisfying theoretical criteria to ensure convergence.

In [121,122], Hsiao *et al.* derived a new convergent complete data ordered subsets algorithm for histogram-mode EM reconstruction (C-OSEM). They have shown that the proposed algorithm, which does *not* involve the use of relaxation parameters, monotonically decreases the complete data objective function, and furthermore demonstrated that while increase in log likelihood with the iterations is not guaranteed to be monotonic (though the authors have always seen this to be the case), nevertheless, the solution *does* converge to the maximum of the log-likelihood objective function. Details of this algorithm are presented in Sec. (3.4.2), wherein we have extended the approach to list-mode reconstruction of the emission data.

2.5.4 Summary

Numerous techniques have in the past, and in recent years, been proposed to solve the 3D image reconstruction problem in PET We have provided an overview of these algorithms under two main headings: analytic and statistical image reconstruction. Analytic techniques have been presented in terms of direct 3D inversion techniques (such as filtered-backprojection), in cases when the measured data are/not truncated, as well as rebinning techniques, which aim to sort the 3D data into stacks of 2D sinograms, to be reconstructed subsequently using fast 2D reconstruction algorithms.

Motivations for switching from the analytic approach to statistical image reconstruction (SIR) were also explained in this chapter, and various SIR techniques were elaborated, with particular emphasis on maximum likelihood expectation maximization (ML-EM) algorithms. Two major problems with these algorithms (noise amplification and slow convergence) were also discussed and a brief overview of suggested solutions to these issues was presented. The next chapter discusses concepts of list-mode acquisition as well as list-mode reconstruction of PET data.

3. List-mode Image Reconstruction

In this chapter, we introduce and investigate list-mode image reconstruction algorithms as a class of imaging techniques suitable for high resolution tomographs.

3.1 Introduction

In positron emission tomography (PET) systems, there is a continuous effort to increase sensitivity and to improve spatial resolution. This, in turn, brings about the need for different approaches to data collection and image reconstruction in order to make use of the increasingly high sampling capabilities of modern PET systems. To better see this, one may make the following interesting observation [123]: Moore's Law states that the number of transistors in an integrated circuit doubles every 18 months. By comparison, the number of individual crystals in PET tomographs has doubled approximately every two years for the past 25 years (*Nutt's Law*). The corresponding number of LORs, which is measured by the square of the number of detector elements, has therefore grown faster than the available computing power. These are depicted in Fig. (3.1).

An important general conclusion can be drawn from this observation: assuming a continuation of the aforementioned trend in modern PET scanners, novel image reconstruction techniques must be considered in order to efficiently reconstruct images from the full information available from these scanners. One important step in such a direction has been the introduction



Fig. 3.1: Moore's Law vs. Nutt's Law, as described in text. The version of the HRRT depicted is the single layer (~60K crystals), whereas most existing HRRT scanners are dual layer (~120K crystals) (courtesy of Dr. R. Nutt).

of the list-mode acquisition capability in a number of modern PET scanners.

3.2 List-mode Acquisition

Conventional tomographs inherently histogram the collected data into sinogram bins as the data are acquired. This is often referred to as *histogrammode acquisition*. A limitation of this imaging modality is that in dynamic imaging, the number of frames, and durations of the individual frames, have to be specified prior to performing the scan. Furthermore, considering the increasing large number of LORs in modern scanners, it is quite likely in dynamic PET studies to have the total number of LORs exceed the actual number of events measured per frame. Subsequently, storing the data in histogram-mode can be expensive. The main motivation for the introduction of *list-mode acquisition* capability into a number of modern scanners, including the high resolution research tomographs (HRRT), has therefore been to enable data-acquisition while allowing the data to be *later* histogrammed into the desired dynamic frames, as well as potentially reducing the size of data-sets.

List-mode acquisition is achieved by storing information regarding the acquired events as they are detected one-by-one in the form of a list. The recorded information, or *attributes* of the list-mode events, typically consist of coordinates of the two detectors along which the photons are detected in coincidence, along with the time-of-detection¹. It is also possible to store further attributes for each event in the list-mode data, such as information about the time-of-flight, depth-of-interaction, and energy. Once the list-mode data are acquired, they can be binned and sub-divided into multiple frames of desired durations, and therefore greater utility is gained by removing the need to specify the frames in advance prior to the scan².

3.3 List-mode Reconstruction

An important possibility with list-mode acquisition is that it enables the use of list-mode algorithms in the inversion of the acquired data into the imaged object. *List-mode reconstruction* is referred to those class of image reconstruction algorithm that do not bin, or histogram, the acquired list-mode events into sinogram bins, and rather incorporate the *processing* of the individual list-mode events directly into the reconstruction task. Thus, we emphasize that list-mode acquisition is not identical to list-mode reconstruction; rather, the former is a pre-requisite for the latter. In fact nowadays, list-mode acquired data are often subsequently histogrammed prior to image

¹ In the case of the HRRT, actually, a more simplifying approach has been taken: time tags are inserted every 1 ms in-between the recorded events, thus removing the need to incorporate the time-of-detection information inside each event. In other words, since the detected events are sorted by time-of-arrival, introduction of time tags is a very practical method to collectively label all events recorded within every 1 ms.

 $^{^2}$ This can be very important in research, wherein the practical timing resolution of the scanner, given the particular patient, can be determined by fitting models to dynamic data obtained using a variety of framing sequences, thus allowing optimization of frame selection from the particular data set.

reconstruction. In this work, we have thus concentrated on taking the listmode approach one step further: from mere list-mode acquisition to list-mode reconstruction.

3.3.1 Benefits of List-mode Image Reconstruction

A large portion of this dissertation is focused on investigating benefits and applicabilities of list-mode reconstruction techniques for scanners with listmode acquisition capability. The high resolution research tomographs (HRRT) (see Sec. 1.7) is an instance of such scanners, and experimental investigations reported in this work have all been performed on this scanner. Several benefits can potentially be gained by performing list-mode image reconstruction, instead of the conventional histogram-mode approach:

1) Faster reconstruction: In 3D PET studies in which a low number of counts are acquired (most notably in dynamic PET scanning), list-mode reconstruction can in principle be performed more quickly and efficiently than histogram-mode reconstruction. This is because when the reconstruction of low-count frames is required, the number of events acquired may be in-fact (much) less than the number of lines of response (LORs) in a full sinogram set, especially with high resolution scanners in which the number of LORs is typically very large (e.g. $\sim 4.5B$ in the HRRT). Use of list-mode reconstruction techniques implicitly ignores LORs along which counts were not recorded [124], and can therefore considerably improve the reconstruction speed especially for low-statistics frames.

2) Preservation of maximum sampling frequency: When histogram-mode reconstruction methods are used, the data are often compressed in the axial or radial directions in order to accelerate the reconstruction tasks. This is sometimes referred to as *data mashing* by which it is meant that certain 'nearby' LORs are histogrammed into the same sinogram bin in order to reduce the size of the sinogram data. For instance, in the case of the HRRT, with no data compression, the sinogram size is 1.5G bytes. Application of axial/radial compression schemes, however, has been shown to adversely affect axial/transaxial resolution of images reconstructed using 3D-OSEM, especially as one moves away from the center of the field-of-view (FOV) [35]. In the case of list-mode reconstruction, since events are considered one-byone, sinogram data compression is in principle not needed, thus resulting in preservation of maximum sampling frequency at no extra cost in terms of time and data size.

3) Time-of-flight Positron Emission Tomography: It has been known since the early 1980s that PET scanners capable of encoding time-of-flight (TOF) information would potentially reduce the statistical noise variance in PET reconstruction [125, 126]. TOF PET is based on the observation that by measuring the difference of the arrival times of the 511 keV photons, a PET camera could restrict the position of the positron emission to a subsection of the line segment joining the detector pair. This is depicted in figure (3.2). TOF PET, especially in whole body scanning and for scans with high random fractions, is expected to considerably improve image noise behavior compared to conventional schemes in which TOF information is not incorporated [127]. Nevertheless, technological difficulties, namely slow electronics and the need for fast scintillators which were at the same time effective absorbing of annihilation photons, had limited development of TOF PET until recently. With the continuous improvements in the technology of PET imaging (e.g. faster electronics), and especially since the discovery of the scintillator LSO³, time-

³ PET scanners based on LSO have the potential of achieving significantly better coincidence timing resolution than the 6-12 ns FWHM typically achieved with BGO.



Fig. 3.2: With conventional reconstruction (left) voxels along the LOR are incremented regardless of position along the LOR. With TOF reconstruction (right), each voxel is incremented by the probability (as measured by the TOF measurement) that the source originated at that voxel.

of-flight PET is now being actively reconsidered [128, 129].

The important observation is that with the added attribute of time-offlight measured for the acquired events, the one-to-one mapping from the LORs to the sinogram bins will no longer be the case; rather, many more sinogram bins (depending on the precision of the TOF modality) may be required to take into account the measured TOF information along each LOR. On the contrary, with list-mode reconstruction, one does not require the use of sinogram bins and instead processes the events one-by-one (and reads the TOF information meanwhile). Subsequently, the data generated from such scanners are most efficiently reconstructed using the list-mode approach if the TOF information is to be made full use of.

4) Accuracy and convenience in motion compensation: An additional advantage of list-mode reconstruction is the increased accuracy and convenience with which motion correction can be implemented into the reconstruction procedure, as we shall elaborate in Sec. (5.2.2). This is the case since in histogram-mode reconstruction: (i) a motion-corrected LOR will not typically correspond exactly to the center of a sinogram bin, and therefore
an interpolation needs to be performed upon histogramming the event into motion-corrected sinograms. On the contrary, motion-corrected list-mode event coordinates can be maintained as continuous variables in list-mode reconstruction, thus potentially preserving a higher degree of accuracy in the reconstruction task resulting directly from better sampling of the measurement. Furthermore, (ii) the regularly employed sinogram-space used in histogram-mode reconstruction needs to be extended, in order to account for all measured events (including those that, upon being motion-corrected, do not correspond to actual detector pairs). On the contrary, list-mode processing of such events is convenient.

3.3.2 Analytic List-mode Image Reconstruction

Of the analytic imaging techniques discussed in chapter (2), one is particularly suitable for list-mode reconstruction: the BPF algorithm (Sec. 2.2.7). This algorithm allows backprojection of list-mode events, which may subsequently be filtered in image-space to yield reconstructed images. Interestingly, prior to the popularity of the list-mode acquisition scheme, this approach was used in such scanners as the HIDAC (II) animal imaging system [130] and the PETRRA whole-body scanner [131] as a data storage scheme [132]: that is, the individual data, as they were being measured, were backprojected onto the image space. In other words, instead of storing the data in 4D projection data format, they were stored as 3D backprojected images⁴. Image-based scatter and random correction steps were subsequently implemented to address data acquired and stored using this modality [24].

Other than the general shortcomings of the analytic approach compared to statistical reconstruction (as reviewed in Sec. 2.4), for a scanner with gaps in-between the detectors (particularly the HRRT), gap-filling techniques (see

⁴ This does require a pre-specification of framing sequences, so that images in each frame are backprojected onto the image voxels for that frame.

chapter 1, footnote 14) all require operating in the projection-space, thus making application of the purely list-mode BPF technique (without gapfilling) inaccurate. Therefore, in what follows, we shall restrict our attention to statistical list-mode image reconstruction algorithms.

3.3.3 Statistical List-mode Image Reconstruction

In this work, we have particularly concentrated on implementation and study of EM-based algorithms for reconstruction of PET data in list-mode. The list-mode expectation maximization (LMEM) reconstruction algorithm has been formulated from first principles [92]: denoting λ_j^m as the image intensity in voxel j (j=1...J) at the *m*th iteration, and p_{ij} as the probability of an emission from voxel j being detected along LOR i (often referred to as the "system matrix"), the LMEM algorithm is given by

$$\lambda_j^{m+1} = \frac{\lambda_j^m}{\sum_{i=1}^I p_{ij}} \sum_{k=1}^N p_{i_k j} \frac{1}{\sum_{b=1}^J p_{i_k b} \lambda_b^m}$$
(3.1)

where i_k refers to the LOR along which the kth list-mode event is detected and N is the number of measured events. The sensitivity correction factor $s_j = \sum_{i=1}^{I} p_{ij}$ is a summation over all possible measurable LORs (i=1...I)and calculates the probability of an emission from voxel j being detected anywhere.

Comparing this algorithm with the histogram-mode version (Eq. 2.61), one notes that the two are different in that in the list-mode algorithm, the backprojection summation over the LORs is instead performed over the measured events, while replacing n_i (the number of events measured along LOR i) with the numeral 1. Running through the entire data, the two algorithms are thus *exactly* equivalent in mathematical terms. Differences arise in: (i) calculation times, wherein the histogram-mode algorithm passes through all the LORs, while the list-mode algorithm runs through the events (see the first point in Sec. 3.3.1); (ii) in accelerated schemes, the data subsets in histogrammode and list-mode reconstruction are fundamentally different (see Sec. 3.4)

Incorporation of Attenuation and Normalization Effects

Following discussion in Sec. (2.4.2), we shall directly include the attenuation and normalization effects as weighting factors in the EM algorithm. We shall therefore decompose the system matrix $P=(p_{ij})_{I\times J}$ into P=WG where $G=(g_{ij})_{I\times J}$ is the geometric probability of an event generated at voxel j to be detected at LOR i, and the diagonal⁵ matrix $W=(w_i)_{I\times I}$ allows a weight to be assigned to each LOR, to account for variations due to attenuation and normalization. Possible schemes incorporating the parallax effect have been discussed in Appendix B.

Such matrix factorization results in the emergence of a cancellation in the algorithm, thus giving:

$$\lambda_{j}^{m+1} = \frac{\lambda_{j}^{m}}{\sum_{i=1}^{I} w_{i} g_{ij}} \sum_{k=1}^{N} g_{i_{k}j} \frac{1}{\sum_{b=1}^{J} g_{i_{k}b} \lambda_{b}^{m}}$$
(3.2)

The cancellation of w_i everywhere, except in the sensitivity factor, which is only calculated once, considerably eases the direct inclusion of normalization and/or attenuation correction. In the case of histogram-mode EM, it has been observed [133] that appropriate modeling of normalization and attenuation as such, results in better image quality when compared to the unweighted case, in which the emission scan is simply pre-corrected by these factors.

⁵ This matrix will not be diagonal if one also includes effects of crystal penetration and inter-crystal scattering. Including such effects results in a high reduction in the sparseness of the matrix and increases the computational demand [68].

3.4 Accelerated List-mode Reconstruction Algorithms

In Sec. (2.5.3), we provided an overview of existing accelerated histogrammode reconstruction techniques. In the ordered subset EM (OSEM) approach, the data are divided into LOR-based subsets, and the image estimate is updated every time the algorithm passes through a data subset.

3.4.1 Ordinary Subsetized List-mode EM Algorithm

The histogram-mode data-subset approach can be conceptually applied to list-mode reconstruction as well [70, 124]. One may consider event-based (instead of LOR-based) subsets, obtained by sub-dividing the list-mode data into segments that span a fraction of the total duration of the data⁶.

Dividing the data space into L subsets, we use S_l to denote the *l*th listmode subset (l=1...L). We shall maintain the use of m (m=1, 2, ...) as the iteration number which is only completed after a thorough loop through all the L subsets in the data. We also use $\lambda_j^{m,l}$ to denote the image estimate at the *m*th iteration and *l*th subset. The subsetized list-mode expectation maximization algorithm (which we shall refer to as the S-LMEM algorithm) is then given by:

$$\lambda_j^{m,l} = \frac{\lambda_j^{m,l-1}}{\sum_{i=1}^I w_i g_{ij}} \sum_{k \in S_l} g_{i_k j} \frac{1}{\sum_{b=1}^J g_{i_k b} \lambda_b^{m,l-1}}$$
(3.3)

One must additionally note here that list-mode subsets exhibit a fundamental difference compared to sinogram-based subsets. This is because each list-

⁶ For dynamic list-mode image reconstruction, each single frame being reconstructed should in principle be static. With a continually changing object, issues of scanner sensitivity and time resolution limit how short-in-duration a single frame may be in order to obtain sufficient counts. Consequently in practice, one is likely to encounter a non-static object for a given frame. To deal with this, therefore, it is best (as we have done in our implementation), to have each list-mode subset contain portions of data obtained at *various intervals throughout the frame*, such that all list-mode subsets represent nearly the same object.

mode subset can be thought of as a lower-statistics scan in its own right. This observation may point to another advantage of list-mode reconstruction. It has been shown [134] that the way in which sinogram-data subsets are chosen and ordered has an effect on the resulting reconstructions. The requirement of minimum variation in-between the data subsets is inherently fulfilled with list-mode subsets.

The aforementioned subsetized algorithm, however, is *not* a convergent algorithm, and instead results in *limit cycles*: oscillatory alternations in image likelihood as well as figures of merit (e.g. contrast, resolution) with further subsets and iterations into the data. Starting from first principles and using the complete data approach as in [121,122], Khurd and Gindi [135] have been able to derive a convergent list-mode EM reconstruction algorithm. The authors have subsequently tested the convergence and speed-up achieved by the algorithm using simulated SPECT data.

In what follows, (i) we show in detail a derivation of the same algorithm using an approach based on re-visiting the histogram-mode technique. (ii) We go on to present an intuitive picture of how the algorithm proceeds using additive updates in image space. (iii) Furthermore, we propose a hybrid algorithm employing both the regular and convergent subsetized list-mode algorithms. (iv) Results of implementing the regular, convergent and hybrid algorithms applied to *experimental* PET data are subsequently presented in Sec. 4.4. We first discuss issues of convergence as observed and tackled in histogram-mode reconstruction.

3.4.2 Convergent OSEM Reconstruction

In the histogram-mode approach, Hudson et al. [114] were able to prove convergence of the OSEM algorithm *only* for an impractical special case, in which the subsets chosen corresponded to a restrictive "subset balance" condition in the matrix. In practice, while the algorithm is seen to perform considerably faster than the regular EM algorithm, it is often seen not to converge to a fixed point and instead results in limit cycles. In other words, the OSEM algorithm in itself does not maximize likelihood.

Subsequently, there has been interest in deriving provably convergent versions of the fast OS methods, as reviewed in Sec. (2.5.3). In [121,122], Hsiao et al. derived a new convergent complete data ordered subsets algorithm for histogram-mode EM reconstruction (C-OSEM). They have shown that the proposed algorithm monotonically decreases the complete data objective function, and furthermore demonstrated that the solution converges to the maximum of the log-likelihood objective function. Below we elaborate upon the algorithm. This algorithm has the advantage that it does not involve a relaxation schedule (as discussed in Sec. 2.5.3), converges to the fixed point of the maximum likelihood objective function, and can be extended to list-mode reconstruction, as we demonstrate shortly.

To see how the C-OSEM algorithm works, we begin by defining C_{ij} as the complete data, as used in statistical derivations of ML-EM, representing the number of counts detected along an LOR i (i=1...I) that have originated from voxel j (j=1...J). However, one only measures and has knowledge of the incomplete data n_i : total number of counts detected along a given LOR i (regardless of the voxel(s) from which the events have originated); i.e.

$$n_i = \sum_j C_{ij} \tag{3.4}$$

Dividing the data space into L LOR-based subsets, S_l is used to denote the lth histogram-mode subset (l=1...L). We shall use m (m=1,2,...) as the iteration number which is only completed after a thorough loop through all the L subsets in the data. We also use $\lambda_j^{m,l}$ to denote the image estimate at the mth iteration and lth subset, while as before g_{ij} is the probability of an emission from voxel j being detected along LOR i).

The C-OSEM can then be written in the form [121, 122]:

$$C_{ij}^{m,l} = n_i \frac{g_{ij} \lambda_j^{m,l-1}}{\sum_{b=1}^J g_{ib} \lambda_b^{m,l-1}}, \forall i \in S_l$$

$$(3.5)$$

$$\lambda_j^{m,l} = \frac{1}{\sum_{i=1}^{I} w_i g_{ij}} \left[\sum_{s=1}^{l} \sum_{i \in S_s} C_{ij}^{m,s} + \sum_{s=l+1}^{L} \sum_{i \in S_s} C_{ij}^{m-1,s} \right]$$
(3.6)

where $C_{ij}^{m,l}$ is the current estimate for the complete data C_{ij} . Here one sets $\lambda_j^{m,0} = \lambda_j^{m-1,L}$ at the beginning of each iteration, while using some initialization for $C_{ij}^{0,s}$ values.

In place of the update Eq. (3.6), the ordinary OSEM algorithm performs the following

$$\lambda_{j}^{m,l} = \frac{1}{\sum_{i \in S_{l}} w_{i} g_{ij}} \sum_{i \in S_{l}} C_{ij}^{m,l}$$
(3.7)

Thus, we see that the C-OSEM algorithm is different from the ordinary OSEM in that calculation of image updates at every subset (numerator of Eq. 3.6) is *not* limited to C_{ij} values for LORs in that subset only. Meanwhile, at any subset l, values of C_{ij} are *updated only* for LORs $i \in S_l$: this explains why the update image at each subset can be computed nearly as fast as that of regular OSEM.

3.4.3 Convergent Subsetized List-mode EM Algorithm

Similar issues, as in the OSEM algorithm, are present in subsetized listmode reconstruction, and the proposed S-LMEM algorithm results in nonconverging (e.g. limit cycles) behavior, as reported in Sec. (4.4). Using transformations as in [124], we are able to extend the histogram-mode formulation presented in Sec. (3.4.2) into list-mode reconstruction. By defining list-mode subsets as event-based subsets, as compared to LOR-based subsets in histogram-mode reconstruction, and replacing the summations over the LORs by summations over the events, while replacing n_i in Eq. (3.5) by the numeral 1, we arrive at the following list-mode reconstruction update equations:

$$\tilde{\lambda}_{j}^{m,l} = \frac{\lambda_{j}^{m,l-1}}{\sum_{i=1}^{I} w_{i}g_{ij}} \sum_{k \in S_{l}} g_{i_{k}j} \frac{1}{\sum_{b=1}^{J} g_{i_{k}b} \lambda_{b}^{m,l-1}}$$
(3.8)

$$\lambda_{j}^{m,l} = \sum_{s=1}^{l} \tilde{\lambda}_{j}^{m,s} + \sum_{s=l+1}^{L} \tilde{\lambda}_{j}^{m-1,s}$$
(3.9)

where $\tilde{\lambda}_{j}^{m,l}$ is an *intermediate* image vector produced by the first update Eq. (3.8), subsequently used by Eq. (3.9) to arrive at the overall image estimate $\lambda_{j}^{m,l}$. Here, we shall set $\lambda_{j}^{m,0} = \lambda_{j}^{m-1,L}$ at the beginning of each iteration.

One may readily note here that the first step of the convergent approach is similar to the regular subsetized list-mode (S-LMEM) algorithm (Eq. 3.3). That is, if step two of the above approach is replaced by $\lambda_j^{m,l} = \tilde{\lambda}_j^{m,s}$, one arrives at the S-LMEM algorithm. This is *not* the case in the histogrammode approach, as used by others, since in the OSEM algorithm, at each subset of the data, the sensitivity correction factors depend on the particular subset (given by $\sum_{i \in S_l} w_i g_{ij}$), whereas this term does not appear in the update Eq. (3.6). On the other hand, since the list-mode data subsets are *event*based and not LOR-based, the sensitivity correction factors are always given by $\sum_{i=1}^{I} w_i g_{ij}$.

As shown in Eq. (3.9), the algorithm takes the form of additive updates in image-space, in that upon arriving at any subset l, the intermediate image updates which have been previously calculated for other subsets $\{\forall s | s \neq l\}$ are added to the update $\tilde{\lambda}_{j}^{m,l}$ calculated for the current subset l. We shall refer to this approach as the convergent subsetized list-mode EM (CS-LMEM) algorithm. We also note that it is easy to show that:

$$\lambda_j^{m,l} = \lambda_j^{m,l-1} + \tilde{\lambda}_j^{m,l} - \tilde{\lambda}_j^{m-1,l} \tag{3.10}$$

From this observation, it follows that by keeping track of the values of $\lambda_j^{m,l-1}$

and the values of $\tilde{\lambda}_{j}^{m,l}$ for all subsets $S_{l}, l=1...L$, values of $\lambda_{j}^{m,l-1}$ can be recursively updated according to the above relation. This makes the calculation of image updates using the CS-LMEM algorithm nearly as fast as the regular S-LMEM algorithm.

Nevertheless, as we demonstrate in Sec. (4.4), improvements in image quality achieved by the CS-LMEM algorithm are slower than those obtained using S-LMEM. This has encouraged us to propose and investigate a hybrid algorithm, as we discuss below.

3.4.4 Hybrid S/CS List-mode EM Algorithm

We have found it very useful to investigate the possibility of combining the advantages of the S-LMEM and CS-LMEM algorithms into a hybrid algorithm. Namely, one typically notices, as also shown in Sec. (4.4), that the regular S-LMEM algorithm, in the first few subsets, is able to produce images of higher quality (e.g. contrast, resolution) relative to the CS-LMEM algorithm, whereas the latter is able to exhibit convergent resolution and contrast behavior as the iterations proceed. The hybrid approach we have taken uses S-LMEM for the entire or part of the first iteration, followed by CS-LMEM in the rest of the calculation.

3.5 Random Correction in List-mode Reconstruction

Various correction techniques for the contribution of random coincidences to the measurement process were discussed in Sec. (1.6.3). The HRRT is an instance of modern PET scanners with the delayed-coincidence measurement capability, and the distribution of random counts can therefore be directly measured and recorded along with the prompts events.

In most histogram-mode reconstruction tasks, the prompt data are *pre*corrected for the detection of randoms by subtraction of the delayed coincidences. However, as discussed in Sec. (2.4.4), the subtraction can be a source of potential problems: First, the randoms-precorrected data do not follow Poisson statistics. Second, it can result in negative histogram bins, which can be a source of potential problems. The first issue has been tackled by means of practical approximations, such as the shifted Poisson model, as discussed in Sec. (2.4.3), while the second issue has typically been dealt with by zeroing negative sinogram values. This latter can in turn result in positive bias in the final reconstructed image [84], as also reported and discussed in this chapter.

3.5.1 Ordinary Poisson EM Reconstruction

The ability of the HRRT to separately store prompt and delayed events can therefore allow a more accurate reconstruction of the true image activity. Here, we summarize the case for histogram-mode reconstruction, which is easily extendable to the list-mode case. As also noted at the end of Sec. (2.4.3), the ordinary Poisson (OP) EM algorithm is well-suited for reconstruction of the HRRT data, since it incorporate accurate modeling of the Poisson statistics of the measured prompts and delayed events.

As elaborated in Secs. (2.4.3) and (2.4.4), defining \bar{r}_i and \bar{s}_i as the *expected* randoms and scattered events contributions along any LOR *i*, one notes that since the prompt events are Poisson by nature, one can consider the log-likelihood function

$$l(\vec{\lambda}) = \sum_{i} [-\bar{n}_{i}^{p} + n_{i}^{p} \ln(\bar{n}_{i}^{p})]$$
(3.11)

where n_i^p is the number of measured prompt events along an LOR *i*, and \bar{n}_i^p is the expected number of prompts along the LOR, given by

$$\bar{n}_i^p = \left\{ \sum_{j=1}^J w_i g_{ij} \lambda_j \right\} + \bar{r}_i + \bar{s}_i \tag{3.12}$$

Application of expectation maximization to the likelihood functions then yields the following OP-EM algorithm [73]:

$$\lambda_j^{m+1} = \frac{\lambda_j^m}{\sum_{i=1}^I w_i g_{ij}} \sum_{i=1}^I \frac{w_i g_{ij} n_i^p}{\sum_{b=1}^J w_i g_{ib} \lambda_b^m + \bar{r}_i + \bar{s}_i}$$
(3.13)

This algorithm poses computational difficulties which need to be addressed. First, we note that a cancellation of w_i in the forward- and backprojection steps is no longer valid, requiring constant look-up of attenuation and normalization factors for LORs along which the events are being read. Furthermore, it must be noted that \bar{r}_i in Eq. (3.13) is an *expected* value which can *not* be appropriately replaced by the *measured* delayed coincidences, rather estimates of mean random counts along the LORs must be used.

To address the latter, two approaches are possible: i) using singles measurements at the detectors [74] to calculate the expected randoms contribution, or ii) variance reduction (smoothing) for the measured delayed events [75–77]. The first approach has the advantage of utilizing high statistics singles measurements. At the same time, it increases the 'bandwidth' of the scanner, since delayed time windows need not be imposed for delayed coincidence measurements, and less saturation of counts would occur especially for studies involving high count rates.

There are certain difficulties that render the application of either of the aforementioned approaches difficult. Both approaches do require constant look-up of calculated sinograms for the expected random events along the LORs, and as such will introduce additional time-costs to the reconstruction task. As for the first approach, in the HRRT, the singles rates can *only* be measured for the crystal block and *not* the individual crystals. This can result in a relatively coarse estimation of random rates at the individual crystals. This effect is further amplified by noting that while singles are only available for the crystal blocks, each crystal has a different efficiency due to intrinsic as

well as object-dependent factors. This is because the contributions of singles events along the various lines passing through a particular crystal depend on the size and shape of the patient. Therefore, one cannot correctly use the trues normalization to compensate for variations in singles rate.

In PET brain imaging, compared to other PET modalities, in which a less, though still existent, degree of object variation across the studies is encountered, an approximate technique has been suggested [74]. It consists of scanning a reference phantom (e.g. a homogeneous cylindrical phantom) and measuring the singles as well as delayed-coincidence rates. Next, comparing the calculated expected delayed-coincidence rates (from the singles) and comparing with the measured delayed-coincidences, calibration factors can be obtained. These factors are then applied to subsequent patient scans in which only the singles rates are measured. This is clearly an approximate technique and remains to be tested and improved further. The above considerations therefore render the application of the singles approach difficult, conceptually and computation-wise, and the possibilities and advantages of this technique remain to be fully explored and addressed⁷.

Second approach (random smoothing) is currently quite difficult to implement on the HRRT. Due to the non-cylindrical shape of the HRRT as well as the depth-of-interaction measurement capability, the relation between the detector pairs and allocated sinogram bins for the LORs is not 1-to-1, i.e. some projection bins correspond to more than one physical LORs. This problem renders application of variation reduction (smoothing) difficult. In addition, this approach is computationally intensive. The HRRT contains 936 (8x8

⁷ One may note here that current scanners actually measure *hits* and not just singles (a hit is a detected photon, independent of whether the other photon is detected or not). For all existing scanners, the vast majority of hits are singles. However, as scanners become more and more sensitive to radioactivity (both in terms of coverage of FOV as well as detector sensitivities) the proportion of singles to hits would become less and less, and corrections (hardware or software) need to be implemented.

dual-crystal) blocks, corresponding to nearly 120k crystals. With no data compression (i.e. mashing), this corresponds to about 3GB of (floating) random sinograms which need to be processed for each frame if this approach is to be taken. Alternatively, it is possible to take a non-analytic approach to improve variance on randoms measurement; namely, one may acquire them with a noticeably wider time window, but this technique will correspond to a potentially considerable increase in resource allocation and 'bandwidth' consumption, and will result in saturation effects for higher count-rate studies.

3.5.2 The Delayed Events Subtraction Technique

Due to the aforementioned issues as well as the computational difficulties they pose, we have alternatively considered the delayed events subtraction method, adapted to list-mode reconstruction as elaborated in Sec. (4.2), as a practical random correction technique in the implementation of a feasible list-mode reconstruction algorithm applicable to the expected workload of the HRRT. Our approach will involve passing *only* through the list-mode data and does *not* require access to histogrammed data nor processing of them. Furthermore, as shown in Sec. (4.4), the algorithm does not exhibit the intrinsic bias observed in OSEM when zero-thresholding is applied to negative sinogram bins. Due to its practical implementation, using ~16 processors we have been able to achieve a processing rate of ~1.0GB/hour, approaching our feasibility criteria as discussed at the beginning of chapter 5.

3.6 Scatter Correction in List-mode Reconstruction

In this dissertation, along with the development of statistical list-mode reconstruction algorithms, we have incorporated, out of the ten complicating factors in PET imaging as introduced in Sec. (1.6), correction techniques for nine. Scatter correction remains to be implemented in the list-mode imaging method for the HRRT. This requires much dedicated research, to be built upon previous work on the subject.

Of the various scatter correction techniques (summarized in Sec. 1.6.2), ones employing the Klein-Nishina formula are most attractive due to treatment of the scatter using basic physical principles. A number of such schemes are currently under investigation by our group. One possible approach has already been developed for the HRRT, namely the Watson scatter correction technique [28]. The original implementation of the algorithm was purely image-based (i.e. would compute the scatter contribution using *only* the reconstructed image). More recent versions [136] require the histogrammed data, in order to scale the calculated scattering components to compensate for external scatter. This current scatter correction algorithm takes under one minute to compute and is under constant modification and speed-up [137].

Another possibility is to use the Watson technique to estimate \bar{s}_i , the scattered events contributions along any LOR *i*, and to analytically include this term in the denominator of the EM algorithm, as shown in Eq. (3.13). Finally, we would also be interested in including the effect of scatter *directly* into the system matrix of the HRRT. Thus, one would be computing the *probability* that an annihilation generated in a voxel *j* would be detected along an LOR *i*, including the effect of scatter, which only requires access to the attenuation data (is independent of the emission image) [138,139]. This method, in our opinion, is potentially the most accurate scatter correction technique (amidst the most computationally intense), since it is not based on an initial estimate of the true activity distribution. This is in contrast to the other aforementioned techniques which require an initial estimate for the image, somehow calculated without true knowledge of the scatter distribution.

3.7 Summary

In this chapter, we have introduced the concept of list-mode acquisition, and elaborated upon the list-mode image reconstruction approach, as an appropriate consideration in current state-of-the-art and future (e.g. time-of-flight) PET scanners. With particular emphasis on the statistical techniques, we have provided an outline of various methods for accelerated list-mode reconstruction. In particular, we have provided a framework, based on previous work [121,122,135], for a convergent list-mode EM reconstruction technique. Experimental studies of the various approaches, along with implementation issues (such as interpolation techniques in geometric projection), are presented in the next chapter.

4. Reconstruction Implementation and Comparison

Objective of this Work

This chapter represents the experimental core of this thesis work in statistical list-mode image reconstruction. The results, along with the algorithms proposed in the previous chapter, have been published [40] and presented [45]. In this work, we have been interested in implementing practical list-mode reconstruction algorithms capable of coping with the considerable size of list-mode data (\sim 200-300GB/week) expected to be generated from the HRRT once in fully operational mode. The issue of computation efficiency is therefore a very important one.

Accelerated variations of statistical list-mode reconstruction have therefore been investigated, with the aim of reaching and exceeding processing rates of 1.0-1.5GB/hour, to enable round-the-clock reconstruction of the data expected to be generated by the HRRT. The list-mode expectation maximization (LMEM) reconstruction algorithm was already introduced in Sec. (3.3.3). Several accelerated versions of the list-mode algorithm (i.e. subsetized, convergent-subsetized and hybrid) were also elaborated in Sec. (3.4). In Sec. (4.1), we introduce the various projection techniques investigated (for accuracy and feasibility) in this work, followed by experimental methods and results presented in sections (4.3) and (4.4).

4.1 List-mode Geometric Back- and Forward-Projection Techniques

In general, forward- and back-projection schemes can be performed using two main approaches: voxel-driven and LOR-driven. It has been suggested [48,140,141] that best results may be obtained when the back- and forwardprojection operations are *output driven*: i.e. if back-projecting, the process should be voxel-driven and if forward-projecting, the process should be LORdriven. A hand-waiving argument has been put forth [141]: if a projection is output driven, the operation would be a "many to one" operation rather than a "one to many" value operation. For instance, when back-projecting, an LOR-driven approach measures how each LOR contributes to all the voxels ("one to many"), whereas in the voxel-driven approach, one would be measuring the contributions of all the LORs to a given voxel at a time ("many to one"). However, this is not a concrete argument and we have in fact verified it not to be the case in our imaging task, as elaborated shortly.

Due to the intrinsically LOR-based nature of list-mode reconstruction, only the LOR-driven projection operations may be utilized. This is one limitation of list-mode reconstruction, and is in a sense acquired due to the fact that one does not need to access the entire projection-space in list-mode reconstruction (which is one important potential advantage of the technique in the first place, especially for low-statistic frames). In our reconstructions, as shown later, we have *not* observed a degradation in image quality when switching from voxel-driven back-projection (used in histogram-mode) to LOR-driven back-projection.

4.1.1 Projection Techniques

Three LOR-driven projection techniques were explored for use in list-mode reconstruction:

1) The Siddon method: Following work of Siddon [142], this technique is based on calculating the path length of intersection of a given LOR along each voxel, as depicted in Fig. (4.1a).

2) Trilinear interpolation: To understand this technique, let us consider an image grid with voxels of unit length in all directions. Trilinear interpolation works by stepping through a given LOR with increments of unit length, as shown in Fig. (4.1b) for the 2D case. For back-projection then, one distributes the LOR value between the nearest voxels (four in 2D, eight in 3D). In forward projection similarly, the LOR value is obtained from the nearest eight (four) voxels in 3D (2D).

3) Bilinear interpolation: In bilinear interpolation, the length of increments on a given LOR is chosen so as to ensure that the sampled points lie along the centers of the voxels in one direction, in order to eliminate interpolation in that direction. In our case, the dimension along which interpolation is eliminated is the transaxial (X or Y) direction along which the given LOR increases faster; e.g. Y direction for the LOR shown in Fig. (4.1c). This method is potentially less accurate (which we have not observed to be the case, as shown later) than the trilinear counter-part since it makes the lengths of the increments LOR-dependent. However, the technique is faster since: (i) interpolation along one direction is eliminated; i.e. for each point on the LOR, an interpolation is performed over only four (two) nearest voxels in 3D (2D), and (ii) for oblique LORs, fewer samples per LOR are considered.

Fig. (4.2) shows images obtained by back-projection of constant values along a projection direction at an oblique transaxial angle $\phi=30$. Results are shown for the three aforementioned projection techniques. Horizontal profiles across these images can be used to compare the back-projection artifacts, as shown in Fig. (4.3a). We have observed that the artifacts vary according to the projection angle, yet typically, the artifacts are seen to be stronger in the Siddon case as compared to the bilinear and trilinear interpolation

105



Fig. 4.1: Projection algorithms employing (a) the Siddon method, (b) trilinear, and (c) bilinear interpolation techniques are drawn, as elaborated in text.



Fig. 4.2: Results of back-projecting constant values along a projection direction onto a 64x64 image at axial $\theta=0$ and transaxial $\phi=30^{\circ}$.

 $techniques^1$.

We shall also demonstrate in Sec. (4.4) that: (i) *reconstructed* image quality is manifestly superior when performing trilinear and bilinear interpolations, compared to the Siddon technique, while the two former techniques perform comparably; and (ii) output-driven histogram-mode reconstruction yields nearly similar image qualities compared to LOR-driven list-mode reconstruction.

¹ The three methods are seen in Fig. (4.3b) to perform nearly similarly in the forward-projection operation.



Fig. 4.3: (a) Plots of horizontal profiles across the back-projection images depicted in Fig.(4.2). (b) Projection vector obtained with forward-projection of a uniform circle along $\theta=0$ and $\phi=15^{\circ}$.

4.2 Subtraction of Delayed Events and Non-negativity Constraints

As elaborated and argued for in Sec. (3.5.2), for the HRRT, we have considered the delayed events subtraction method as a practical technique to compensate for presence of randoms in the data. In this work, as explained later, we have imposed an *image non-negativity* constraint in the list-mode reconstruction algorithms. To better understand the constraint, we have performed studies of whether its incorporation into the reconstruction tasks can improve the overestimation bias that is likely to arise from the *sinogram non-negativity* constraint often used in histogram-mode reconstructions, as we describe below.

We first consider the histogram-mode delayed coincidence subtraction

technique:

$$\lambda_{j}^{m+1} = \frac{\lambda_{j}^{m}}{\sum_{i=1}^{I} w_{i} g_{ij}} \sum_{i=1}^{I} g_{ij} \frac{n_{i}^{trues}}{\sum_{b=1}^{J} g_{ib} \lambda_{b}^{m}}$$
(4.1)

with $n_i^{trues} = n_i^p - n_i^d$ where the measured delayed coincidences are subtracted from the measured prompts. Due to the Poisson nature of measured prompts and delayed coincidences, it is possible for any bin *i* to record more randoms than prompts. The constraint $n_i \ge 0$ has normally been imposed on the data in previous work [133,143], such that the value of n_i is set to zero for any LOR along which a negative value is obtained (i.e. the *sinogram* non-negativity constraint).

Nevertheless, in low statistics scans, especially those with large random fractions, it is commonly observed that a noticeable number of sinogram bins exhibit negative trues counts after random correction. Imposing the sinogram non-negativity constraint would in turn introduce an *overestimation* bias in the reconstructed images. In this regard, we have instead considered using a weaker condition, namely the *image* non-negativity constraint, such that if upon processing a data subset, the correction to an image voxel is calculated to be negative (due to presence of considerable contribution from negative sinogram bins), the image voxel is not updated for that particular subset².

From Eq. (4.1), one makes the intuitive observation that by replacing the summation over the LORs with a summation over the number N of measured data (i.e. effectively reading the sinogram counts one by one), and setting n_i to 1, if the histogrammed event was a prompt event, and to -1, if it was a delayed-coincidence event, one arrives at the list-mode reconstruction algorithm:

$$\lambda_{j}^{m+1} = \frac{\lambda_{j}^{m}}{\sum_{i=1}^{I} w_{i} g_{ij}} \sum_{k=1}^{N} g_{i_{k}j} \frac{\delta_{k}}{\sum_{b=1}^{J} g_{i_{k}b} \lambda_{b}^{m}}$$
(4.2)

 2 Due to the multiplicative nature of the iterative EM algorithm, if image voxels were allowed to have negative values, then the subsequent updates would be adversely affected!

where

$$\delta_k = \begin{cases} 1 & k \text{ is a prompt event} \\ -1 & k \text{ is a delayed coincidence event} \end{cases}$$
(4.3)

In this study, we implemented the statistical list-mode reconstruction scheme described by Eqs. (4.2) and (4.3) for the HRRT. We note that unlike *truly* EM algorithms such as Eq. (3.13) for which there is never any "divide by zero" error, it is possible in the above delayed events subtraction technique that the denominator of Eq. (4.2) would be zero (i.e. forward projection of the current image estimate along the LOR coordinates of a measured event is zero). In such cases, our algorithm neglects the particular events for which this occurs.

The algorithm was accelerated using the three schemes presented in Sec. 3.4, namely, the S-LMEM, CS-LMEM and hybrid S/CS-LMEM variations. The *image* non-negativity constraint, described earlier, was also imposed. We also investigated the parallelization of the list-mode reconstruction code on a Linux cluster.

The algorithms were tested with point source measurements as well as data sets spanning a wide range of count rates in order to investigate resolution, bias, contrast and noise properties for different data acquisition conditions. Convergence properties of the various schemes were also compared in contrast as well as resolution studies.

4.3 Methods

Phantoms used and measurements performed: On the HRRT, the data can be axially spanned in various modes. In list-mode reconstruction, however, as discussed in Sec. (3.1), since the individual LORs are being read one-byone, rate of processing the emission data in list-mode is independent of the axial span. One would thus be motivated to explore use of no spanning at all (i.e. span 0). However, performing the normalization measurement in span 0 is extremely time-consuming (as sufficient counts are needed in the LORs), has not yet been implemented on the HRRT, and remains to be studied. Therefore, we have instead investigated the effect of switching in-between axial spans of 3 and 9. In histogram-mode reconstruction, processing the data with an axial span of 3 is more expensive than a span of 9 (by nearly three times), whereas this is not the case in list-mode reconstruction.

Two separate experiments were performed as we outline below. In order to allow direct comparison between the various algorithms, consistent axial spans (9 for first experiment and 3 for second experiment) were imposed, while a maximum ring difference of 67 was also used. In the second experiment, which involved resolution measurements, the effect of resolution degradation by switching from a span of 3 to a span of 9 was also investigated. In all the reconstructions, 16 subsets were used for the accelerated algorithms. The hybrid S/CS-LMEM algorithm consisted of having the first 8 subsets being iterated using the S-LMEM approach and subsequently switching to the CS-LMEM counterpart³. The following experiments and analyses were performed:

Experiment 1 – Contrast Phantom

This experiment was performed in order to study the robustness of the random correction technique as well as contrast vs. noise properties of the list-mode technique. A 20 cm long, 10 cm radius phantom was used. The phantom had three 5 cm diameter cylindrical inserts: one was solid plastic, one was filled with water (cold insert) and one was filled with a ¹⁸F radioactivity concentration of 3.39 μ Ci/ml (hot insert). The phantom itself was filled with a ¹⁸F concentration of 0.622 μ Ci/ml ('background'), yielding a

 $^{^{3}}$ This is because we have often observed that the image quality improves rapidly for the first half of the subsets in the *first* iteration of the S-LMEM algorithm, while it begins to exhibit oscillating behavior for the remaining subsets.

Index	Total Activity(mCi)	Trues Rate(kcps)	Random Fraction(%)
1	2.7	814	70.9
2	1.8	591	46.8
. 3	1.3	425	32.0
4	0.95	208	16.2
5	0.19	69	7.73

Tab. 4.1: Table of Statistics for Frames used in the analysis

hot insert to background ratio of 5.45. The total amount of radioactivity in the scanner field of view (FOV) at the beginning of the scanning procedure was 4.16 mCi.

A series of sixteen 20 min long scans was acquired in list-mode, one hour apart. The measurement thus covered 8.7 radioisotope half-lives, yielding a final amount of radioactivity in the FOV of 0.00972 mCi. The subset of the scans listed in table (4.1) was used in the analysis. The listed scans were chosen for analysis so as to cover a random fraction (randoms/trues) range between 8% to 71%, which covers most clinically encountered situations. The following figures of merit were used in evaluating the reconstruction and random correction methods:

i) Quantitative Accuracy of the Random Correction: Reconstructed images were compared with those obtained using the FORE+2D-FBP scheme, since the latter exhibits linearity with measured counts and can be used as a standard for quantitative accuracy. ROIs were selected on the hot and cold as well as the background regions of the reconstructed images for each of the studied frames. Five transaxial planes were selected for the calculation of mean and standard deviation of total counts in the ROIs. Tests of quantitative accuracy were performed for the 3D-OSEM algorithm (with (a) no random correction, (b) the sinogram non-negativity constraint, and (c) the image non-negativity constraint) as well as the S-LMEM algorithm. ii) Image Random Fractions: For each list-mode reconstructed frame from table (4.1), an image random fraction (RF = randoms/trues) was measured using:

$$imageRF \equiv \left[\sum_{j=1}^{J} \lambda_j^{nrc} / \sum_{j=1}^{J} \lambda_j^{rc}\right] - 1$$
(4.4)

where λ_j^{rc} and λ_j^{nrc} are reconstructed image intensities at a voxel j with/without random-correction being performed, respectively. As such, the image random fractions were calculated and compared for images reconstructed using the S-LMEM algorithm. Image RF values are not expected to be numerically equivalent to the random fraction in the acquired data: the true and the random events have a different spatial distribution and will be therefore differently affected by attenuation and sensitivity corrections. However, the ratio between the image RF and the acquired events RF must not vary as a function of acquisition condition if no count rate or number of counts dependent bias is introduced in the data by the random corrections.

iii) Contrast vs. Noise Comparisons: Contrast vs. noise studies were performed for images reconstructed up to three iterations using the 3D-OSEM, S-LMEM, CS-LMEM and hybrid S/CS-LMEM algorithms. The contrast and the noise were estimated following approximately the NEMA NU 2001 protocol. The percent contrast Q_H for the hot cylinder was calculated by:

$$Q_H = \frac{C_H/C_B - 1}{A_H/A_B - 1} \times 100\%$$
(4.5)

where C_H and C_B are the average counts in regions of interest (ROIs) placed on the reconstructed images of the hot insert and the background region, respectively, and A_H/A_B is the actual concentration ratio between the two regions (measured to be 5.45). The percentage noise (standard deviation/mean) was calculated by placing eight ROIs on different parts of the background image and averaging their values to yield a background mean value and its standard deviation.

Experiment 2 – Radioactive Paper Source

This experiment was performed to study resolution and noise properties of the various interpolation techniques, as well as the ordinary, convergent and hybrid subsetized list-mode algorithms. Using the technique presented in Appendix C, which allows printing of radioactive patterns using a modified standard ink-jet printer, we imaged radioactive (^{18}F) point sources of size 0.7 mm placed at X=0,1,2,3,4,5 and 6 cm radially away from the center of the FOV. The sample also included a 1x7 cm rectangular area of uniform activity created for the purpose of monitoring noise behavior. For better visualization, a sample reconstructed image of the radioactive paper source is shown in Fig. (4.4). The middle row of point sources (which were printed over a background were not utilized for analysis in this work.

The overall FWHM for any given point was measured by calculation of the root mean squared value of the measured point widths in the transaxial (X,Y) and axial (Z) directions. The percentage noise (standard deviation/mean) for a given reconstructed image was calculated in two ways:

1) *Voxel noise:* in which percentage variation of the individual voxels along the entire rectangle was measured.

2) *ROI noise:* in which the activity rectangle was sub-divided into eight small rectangular ROIs, and the percentage variation of the sum of counts in the ROIs was measured.

i) Comparison of Projection Techniques: As elaborated in Sec. (4.1.1), three LOR-driven projection methods were considered and implemented for performance of back- and forward-projection on the HRRT: (a) the Siddon method as well as (b) trilinear and (c) bilinear interpolation techniques. The data were subsequently reconstructed using the aforementioned projection methods (three iterations of the S-LMEM algorithm were applied). Histogram-mode EM reconstruction was also applied to the data, wherein



Fig. 4.4: A sample reconstruction of the radioactive paper source (3 iterations of the S-LMEM algorithm). The lower row of points sources as well as the rectangular box were utilized for analysis of resolution and noise properties.

projection algorithms used bilinear interpolation and were output-driven (i.e. voxel-driven back-projection and LOR-driven forward-projection) as discussed in Sec. (4.1.1). Subsequently, the resolution (FWHM) and noise properties of the reconstructed images were compared.

ii) Convergent List-mode Reconstruction: In order to study convergent list-mode reconstruction algorithms, we performed the following comparisons:

a) plots of measured FWHM vs. iteration (for two selected point sources 1 and 5 cm away from the center of the FOV) were calculated and shown for three iterations of the 3D-OSEM, S-LMEM, CS-LMEM and hybrid S/CS-LMEM reconstruction schemes.

b) Plots of measured FWHM vs. radial position were also depicted for these four schemes to compare the final reconstructed FWHM values across the FOV.

Procedures (a) and (b) was performed for cases when i) Siddon and ii) bilin-

ear interpolation techniques were used.

c) Plots of noise vs. iteration were also calculated for the 3D-OSEM, S-LMEM, CS-LMEM and hybrid S/CS-LMEM reconstruction schemes.

iii) Effect of Axial Spanning: We have also studied effects of degradation in resolution as one switches from (axial) span 3 reconstruction to span 9. The latter is expected to result in poorer FWHM values along the axial (Z) direction and is expected to degrade further with increasing distances from the center of the FOV. Values of overall FWHM vs. iteration were plotted for two point sources at X=3 cm and 5 cm from the center of the FOV: The measured degradation in FWHM (along Z direction) vs. radial position were also depicted for the S-LMEM, CS-LMEM and hybrid S/CS-LMEM reconstruction schemes.

Code parallelization: The code parallelization was implemented on the Western Canada Research Grid⁴, University of British Columbia. It consists of 345 x IBM eServers, using dual 3.0 GHz processors (2GB RAM). The message passing interface (MPI) software was utilized to parallelize the list-mode reconstruction code. The algorithm essentially consists of having several slave nodes to perform the actual forward and backward projections, results of which are passed to the master node for updating the current image estimate after every subset.

4.4 Results and Discussion

Experiment 1 – Contrast Phantom

Quantitative Accuracy of the Random Correction: Figs. (4.5) show plots of ratios between ROI counts in images reconstructed using 3D-OSEM and those obtained using FORE+2D-FBP. In each figure, three plots are shown corresponding to ROIs in the hot, background and cold regions. Figs. (4.5a,b,c)

⁴ http://www.westgrid.ca/



Fig. 4.5: Calculated ratios between total counts for images reconstructed using 3D-OSEM and those obtained using FORE-2D-FBP, for the hot (top), background (middle) and cold (bottom) regions of the phantom, when (a) no random correction, (b) the sinogram non-negativity constraint and (c) the image non-negativity constraint have been imposed. For better visibility, the plots have been shifted by ten units with respect to one another. The case of reading 10M counts is shown, and qualitatively similar results have been observed in the 40M and 80M cases. The error bars indicate statistical variation for the five transaxial planes selected.

Frame 1	No Random Correction	Sinogram Posit. Constr.	
Hot	3.8%	1.8%	
Background	21.7%	17.1%	
Cold	74.0%	65.2%	
Frame 5	No Random Correction	Sinogram Posit. Constr.	
Hot	0.9%	0.4%	
Hot Background	0.9%	0.4%	

Tab. 4.2: Image Bias for 3D-OSEM reconstruction of Frames 1 and 5 in table (4.1)(10M counts)

correspond to cases where 3D-OSEM was performed: (a) without any random correction, (b) with the sinogram non-negativity constraint and (c) with the image non-negativity constraint. The data reconstructed contained 10M counts. The reason this comparison was performed for the 3D-OSEM algorithm was because a similar comparison is not possible in list-mode reconstruction, for which the sinogram non-negativity constraint cannot be imposed.

In Fig. (4.5c), the values are seen to be consistent (within statistical error) for a wide range of event random fractions (8%-71%). This is seen *not* to be the case in the former two schemes, especially for the background and cold regions, which contain less trues and therefore are more sensitive to insufficient random correction. One is therefore clearly able to verify that, as predicted, imposing the sinogram non-negativity constraint can introduce an *overestimation* bias in the reconstructed images, especially for low-statistic scans.

Table (4.2) shows percentage increase in ROI counts density for 3D-OSEM reconstructed images of frames 1 and 5 (71% and 8% random fractions) when reconstructed using schemes (a) and (b), when measured in comparison to



Fig. 4.6: Calculated ratios between total counts for images reconstructed using S-LMEM and those obtained using FORE-2D-FBP, for the hot (top), background (middle) and cold (bottom) regions of the phantom. For better visibility, the plots have been shifted by ten units with respect to one another. The case of 10M counts is shown, and qualitatively similar results have been observed in the 40M and 80M cases. The error bars indicate statistical variation for the five transaxial planes selected.

the case of reconstructing with the *image* non-negativity constraint only. Schemes (a) and (b) are seen to exhibit close values for overestimation bias percentages in each of the hot, background and cold regions, which can be explained by the low-statistic nature of the scans (10M counts) resulting in a notable fraction of histogram bins to measure more randoms than prompts, which are subsequently neglected in scheme (b). Similar qualitative patterns have also been observed in the cases of having 40M and 80M counts, with the difference that bias is seen to become less significant in scheme (b) (e.g. 54% for 40M counts and 35% for 80M counts, in cold region).

Testing quantitative accuracy for the list-mode algorithm, Fig. (4.6) shows similar plots of *ratios between* images reconstructed using S-LMEM and FORE+2D-FBP. The plots are also seen to be consistent (within statistical error) for the hot, cold and background regions. Consequently, the S-LMEM

Index	Events RF	Image RF	Image RF/Events RF(%)
1	70.9	45.8	65.7
2	46.8	31.0	66.7
3	32.0	21.2	66.8
4	16.2	10.7	65.7
5	7.73	5.19	67.2

Tab. 4.3: Table of random fractions

algorithm is seen to preserve reconstructed counts in various parts in the image for a wide range of random fractions.

Ratios of Random Fractions: Image random fractions, calculated as described in previous section, along with random fraction of the acquired events are shown in table (4.3). The two values are not expected to be equal, as explained in the methods section, but for an accurate list-mode random correction technique, they are expected to have same ratios independent of particular frame. As can be seen from the table, there is only a very small change in the ratios over a wide range of random fractions.

Contrast vs. Noise Comparison: Monitoring progression of image quality with iteration, Fig. (4.7) shows contrast vs. noise plots for 3 iterations (16 subsets - with results after every 4 subsets shown) of S-LMEM as well as 3D-OSEM on frame 4 (with random fraction of 16%). We have verified, for the wide range of random fractions and total counts considered, that the S-LMEM algorithm performs at least as effectively (in terms of contrast vs. noise) as 3D-OSEM. We have also seen, as shown in Fig. (4.7), that both algorithms exhibit limit cycles, commonly reported in the literature to arise from subsetization of the measured data.

Figs. (4.8a,b) show plots of contrast vs. noise for images reconstructed using three iterations of the S-LMEM, CS-LMEM and hybrid S/CS-LMEM algorithms. One is able to clearly observe limit cycles in the ordinary S-



Fig. 4.7: Contrast vs. noise plots for images reconstructed with S-LMEM and OSEM. Three iterations are performed with 16 subsets, with results shown every 4 subsets. Results are shown for (a) 20M and (b) 40M total counts. Presence of limit cycles is noticeable in both algorithms.

LMEM approach, especially with lower statistics. The CS-LMEM approach, on the other hand, is able to eliminate the observed limit cycles. Using the hybrid technique, in order to combine the advantages of the S-LMEM and CS-LMEM algorithms (as explained in Sec. 3.4.4), the plots show that one is able to attain higher contrast values in fewer iterations, and yet maintain the non-cyclical behavior as the reconstruction proceeds.

Experiment 2 – Radioactive Paper Source

i) Comparison of Projection Techniques:

Fig. (4.9) shows measured resolution values for the various points across the FOV, upon application of three iterations of the S-LMEM algorithm to the data. For comparison, results of application of histogram-mode EM reconstruction are also shown, wherein voxel-driven back-projection and LORdriven forward-projection (referred to as output-driven projection) were used,



Fig. 4.8: Contrast vs. noise plots for images reconstructed using the S-LMEM (dotted), CS-LMEM (- -) and hybrid S/CS-LMEM (solid) algorithms with scan durations containing (a) 10M and (b) 5M total counts. Three iterations are shown.



Fig. 4.9: Plots of measured resolution vs. radial position for the S-LMEM algorithm implemented using the Siddon algorithm as well as bilinear and trilinear interpolation techniques. For comparison, results of outputdriven (as described in text) histogram-mode OSEM reconstruction are also shown. Three iterations were performed.



Fig. 4.10: Plots of noise vs. iteration for the reconstruction schemes described in the caption of Fig. (4.9).

as recommended by previous investigators (see discussion in Sec. 4.1.1).

As a side note: one is also able to observe *space-variance* of the point spread function, manifesting itself as a degradation in resolution as one moves away from the center of the FOV (seen in all the reconstruction tasks). This effect occurs due to a higher probability of inter-crystal penetration with higher angles of radiation incident on crystal fronts. Depth-of-Interaction (DOI) encoding is known to improve this problem, but has not reached complete space-invariance. In Appendix B, we present an approach to model the space-variance and anisotropicity of the point-spread function into the system matrix of the EM algorithm for the HRRT.

Similarly, noise vs. iteration plots are shown in Fig. (4.10) for the aforementioned reconstruction algorithms, wherein (a) voxel-noise and (b) ROInoise were considered, as described in the methods section. Three main observations can be made with respect to these figures:

1) Clearly, the Siddon technique performs noticeably poorly compared to the other interpolation methods (especially in terms of resolution).

2) Trilinear and bilinear interpolation techniques perform nearly similarly.

3) Histogram-mode reconstruction with output-driven projections does not perform better than list-mode reconstruction with LOR-driven projections.

The current implementation of the trilinear interpolation technique is >3-4 times slower than the Siddon method, while Bilinear interpolation is comparable to the latter (only around 20% slower). This has therefore given us sufficient motivation to perform our reconstructions using projection algorithms which employ bilinear interpolation.

ii) Convergent List-mode Reconstruction: Figs. (4.11a,b) show plots of reconstructed FWHM width vs. iteration for point sources located at X=1 cm and 5 cm from center of FOV, with the data reconstructed using the 3D-OSEM, S-LMEM, CS-LMEM and hybrid S/CS-LMEM algorithms. Bilinear interpolation was used in the projection algorithms.

The values of FWHM resolution are seen to change in a cyclical manner for the 3D-OSEM and S-LMEM algorithms. In Fig. (4.11a), for instance, the FWHM width reconstructed using the S-LMEM approach is seen to oscillate between a low of 3.17 and a high of 3.23mm. Nevertheless, one clearly observes that in the CS-LMEM approach, due to its converging behavior, the FWHM widths improve with further iterations in a systematic and predictable manner.

One is also able to observe that the hybrid approach results in a faster decrease in reconstructed FWHM width with less iterations while maintaining the non-cyclical behavior. For comparison purposes, similar plots are shown when using the Siddon method, as seen in Figs. (4.11c,d). Clearly, bilinear interpolation is seen to result in superior image qualities.

Fig. (4.12) shows plots of measured FWHM values after four iterations for all the seven points located at X=0,1,2,3,4,5 and 6 cm from the center of FOV. We note from the plots that the histogram-mode and list-mode algorithms are able to achieve nearly similar FWHM values for a given point. Results


Fig. 4.11: Plots of reconstructed FWHM width vs. Iteration are shown for the 3D-OSEM, S-LMEM, and the CS-LMEM and hybrid algorithms for the point source located at (a) 1 cm and (b) 5 cm from center of the FOV. (c) Final FWHM values are shown for all the reconstructed points sources using the aforementioned four schemes (after 3 iterations).



Fig. 4.12: Plots of reconstructed FWHM width vs. Iteration are shown for the 3D-OSEM, S-LMEM, and the CS-LMEM and hybrid algorithms for the point source located at (a) 1 cm and (b) 5 cm from center of the FOV. (c) Final FWHM values are shown for all the reconstructed points sources using the aforementioned four schemes (after 3 iterations).



Fig. 4.13: Plots of reconstructed FWHM width vs. Iteration are shown for the 3D-OSEM, S-LMEM, and the CS-LMEM and hybrid algorithms for the point source located at (a) 1 cm and (b) 5 cm from center of the FOV. (c) Final FWHM values are shown for all the reconstructed points sources using the aforementioned four schemes (after 3 iterations).



Fig. 4.14: (a) Overall FWHM vs. iteration for images reconstructed using the hybrid S/CS-LMEM algorithm with axial spans of 3 and 9: point sources located at 3 cm and 5 cm from center of FOV are shown. (b) Degradation in FWHM (along Z-direction) vs. radial position as one switches from axial span 3 to 9: all the point sources reconstructed using the S-LMEM, CS-LMEM and hybrid schemes (after 3 iterations) are shown.

when applying the (a) bilinear and (b) Siddon methods are shown. The bilinear interpolation method is seen to outperform the Siddon technique for all the point sources. Furthermore, we note that the convergent and hybrid techniques perform at least as well as (if not better than) the regularly subsetized list-mode and histogram-mode EM algorithms.

Noise vs. iteration plots are shown in Fig. (4.13) for the aforementioned reconstruction algorithms (using the bilinear technique), wherein (a) voxelnoise and (b) ROI-noise were considered. The convergent CS-LMEM algorithm converges very slowly, compared to the other algorithms, and after the first iteration exhibits poorer FWHM values. On the other hand, the hybrid algorithm is seen to be suitable for reconstructions where early termination is used (for time-cost considerations) as it reaches noise levels comparable to (and possibly, voxel-wise, better than) other non-convergent algorithms. *iii)* Effect of Axial Spanning: Plots of FWHM vs. iteration for point sources at X=3 cm and 5 cm from the center of the FOV are shown in Fig. (4.14a). The image has been reconstructed using the hybrid algorithm with axial spans of 3 and 9. It is seen, as expected, that the effect of axial spanning is more significant for the point source more distant from the center of the FOV. To see this more clearly, Fig. (4.14b) shows plots of final FWHM values (along the Z-direction) for the various point sources reconstructed using the S-LMEM, CS-LMEM and hybrid algorithms. The resolution along the Z-direction is seen to degrade further as one moves away from the center of the FOV for the various schemes. We have also checked the resolution along the transaxial direction and have observed no change to occur upon switching in-between the axial spans, as expected, since this is only expected to affect the resolution in the Z-direction.

Parallelization of the Reconstruction Task: Parallelization was successfully implemented using MPI for the list-mode algorithms. It was found that having the slave nodes indirectly accessing list-mode data *did* improve the efficiency compared to the case of the master node distributing the actual workload to all the processors. This required multiple read access to a single file as supported by the cluster. With this data on the cluster, using 16 subsets and 16 processors, we are able to reach a processing rate of 0.5-1.0GB/hour, depending on the number of counts per subset⁵.

4.5 Dynamic PET List-mode Image Reconstruction

In this section, we present an investigation into the extension of the aforementioned list-mode image reconstruction techniques developed for the HRRT to dynamic PET imaging (see discussion in Sec. 1.5). Our dynamic scheme consists of independently reconstructing the events obtained withing each in-

⁵ With too few counts per subset (i.e. too many subsets in the data), the task is not as efficiently parallelized.

dividual timing sequence (i.e. dynamic frame), with quantitative corrections applied to each generated image. The extension from 3D imaging to dynamic (or 4D) imaging necessarily requires corrections for the following factors:

i) Detector deadtime (see also Sec. 1.6.5): Currently, the deadtime correction scheme used on the HRRT involves a global scaling of the reconstructed images within each frame as determined by the average singles rates within each frame⁶. The scheme involves application of the following deadtime multiplicative factor to the reconstructed images:

deadtime factor =
$$\exp(\alpha S)$$
 (4.6)

where S indicated the measured singles rates per block (the HRRT measures the singles rates only at the block level, see Sec. 3.5.1 for implications), and α is a factor determined experimentally to be 7.7004×10⁻⁶ when LLD=350 ke-Vand 8.94×10⁻⁶ when LLD=400 keV(LLD is the Lower Level energy Discriminator, events with energies below which are rejected by the scanner⁷).

ii) Decay of radioactivity: As mentioned in Sec. (1.6.6), decay correction is relatively the most trivial of all corrections to be applied to reconstructed images. It requires a global scaling of the final reconstructed image by a factor determined by the start and end times within which the events are acquired⁸.

⁶ It has been determined by the manufacturing company that singles rates are not noticeably affected by deadtime effects, compared to the deadtime saturation effect encountered by the coincidence events. Thus, the major bottleneck in the processing can be attributed to the coincidence detection system which saturates the acquired coincidences globally, as mentioned in Sec. (1.6.5).

⁷ Most HRRT scanners, including the Vancouver HRRT, are currently setting the LLD to 400 eV.

⁸ For a given dynamic frame (counts within which are reconstructed into a single image) in ordinary S-LMEM reconstruction, all-but-the-last subsets in the frame merely contribute to improved image estimates, and it is only the counts in the very last subset which determine (quantitatively) the final image. This implies that we need only implement deadtime and decay corrections for the counts occurring within the durations spanned by the very last subset. In other words, scaling the previous image estimates for

4.6 Methods and Results

In what follows, we study the quantitative accuracy of our dynamic list-mode image reconstruction technique, and compare it to the ones readily available for the HRRT: namely (i) FORE+2D-FBP, and (ii) 3D-OSEM.

Contrast Phantom: We used a phantom similar to the one described in experiment 1 of Sec. (4.3). The phantom was this time filled with a ¹¹C concentration of 0.311 μ Ci/ml ('background'), with a 1.61 μ Ci/ml activity in the hot insert. The 'cold insert' was filled with only water. An initial trues rate of 1.6 Mcps were observed. Eighteen dynamic frames (each 5 minutes in duration) were considered and 16 subsets were used in each reconstruction. The measurement thus covered 4.4 radioisotope half-lives. The random fraction was 25% in the first frame and 5% in the last frame. The following comparisons were performed:

i) Time activity curve (TAC) comparisons: Plots of mean reconstructed voxel intensity (for the three hot, cold and background regions) were obtained for all the reconstructed frames. For a quantitatively accurate reconstruction algorithm, the TAC curves are expected to be constant.

ii) Axial profile comparisons: Mean reconstructed voxel intensities within each axial plane (for the hot, cold and background regions) were plotted as a function of the axial plane. The corresponding curves for the various frames were overlayed for a visual comparison.

iii) Contrast recovery comparisons: The percent contrast for the hot cylinder was calculated using Eq. (4.5). The plots were depicted as a function of the number of acquired counts per frame.

deadtime and decay effects has no impact, since λ^{i+1} is not affected by scaling of λ^i in the EM algorithm. In convergent CS-LMEM reconstruction, however, images reconstructed from each subset contribute (additively) to the final image, and therefore correction factors need to be applied to all the intermediate images.



Fig. 4.15: Time activity curves (TACs) for 18 frames each 300 s in duration. First frame has a random fraction of 25%, last frame 5%. One iteration and 16 subsets are used in list-mode and OSEM reconstructions.



Fig. 4.16: Axial uniformity comparisons: profiles were drawn through the hot (top), cold (bottom) and background (middle) regions in each of the reconstructed images.

Results: Fig. (4.15) shows images of time activity curves obtained for the three reconstruction schemes: (a) S-LMEM (b) FORE+2D-FBP and (c) 3D-OSEM. The list-mode algorithm is seen to yield a relatively flat TAC, similar to the FORE+2D-FBP case. However, the histogram-mode scheme yields time activity curves which (especially for the cold and background regions) increases with higher count-rate frames. This is because the current implementation of dynamic histogram-mode reconstruction imposes a sinogram non-negativity constraint, which as we have shown in previous sec-

tions, results in a positive bias with frames with higher activities (thus higher random fractions).

Axial profiles of the mean ROI activity have been depicted in Fig. (4.16) for the hot (top), cold (bottom) and background (middle) regions. The profiles for the 18 frames have been drawn overlaying one another. The list-mode algorithm is clearly seen to outperform the FORE+2D-FBP algorithm in terms of axial uniformity. We note here that the power of statistical Poisson modeling of the measurement process clearly exhibits itself on the HRRT, wherein, due to the enormous number of LORs, typical dynamic frames, rarely exhibit counts per LOR to be more than 1 or 2, thus rendering the EM algorithm (as compared to Gaussian techniques, and simple analytic methods) a very powerful solution to the task of image reconstruction. This exhibits itself particularly in terms of the reconstructed noise, axial uniformity, etc. as we observe here. Furthermore, due to the aforementioned zero-thresholding bias, the OSEM algorithm (distributed by the manufacturing company) does not yield overlapping axial profiles for the various frames, and performs poorly.

Calculated percentage contrasts are depicted in Fig. (4.17) for the various frames. Best uniformity is observed for the list-mode technique, while the OSEM approach (with zero-thresholding of negative sinogram bins) performs poorly, as we also observed in Fig. (4.5b) and commented in Sec. (4.4).

4.7 Conclusion

Practical statistical list-mode reconstruction algorithms were developed for the high resolution research tomograph (HRRT). An *image* non-negativity constraint was employed in the algorithm. This constraint was shown, using comparisons in histogram-mode 3D-OSEM reconstruction, to effectively remove the overestimation bias typically encountered when a *sinogram* non-



Fig. 4.17: Plots of percentage contrast for the various dynamic frames.

negativity constraint is used. Results indicated that images were comparable in terms of statistical properties, contrast vs. noise as well as resolution with images generated by 3D-OSEM. Furthermore, robustness of the list-mode technique was demonstrated in dynamic PET image reconstruction, upon the inclusion of appropriate deadtime and decay correction factors (demonstrating the quantitative accuracy of the technique).

Three LOR-driven projection techniques were considered: (a) the Siddon method as well as (b) trilinear and (c) bilinear interpolation methods. We were able to demonstrate that the Siddon approach performs poorly compared to the other two algorithms (especially in terms of FWHM). Our method of choice was therefore the bilinear approach as it was considerably faster than the trilinear interpolation technique and produced comparable image qualities.

Using regular subsetized list-mode reconstruction, we were able to observe limit cycles: oscillatory alternations in image quality parameters (such as contrast and resolution) with further subsets and iterations into the data, similar to what is commonly encountered in ordered subset histogram-mode reconstruction. To address this issue, we implemented a convergent list-mode EM reconstruction algorithm, based on previous work [121,122,135], and investigated its properties using experimental PET data. It was demonstrated that the algorithm is robust and does not result in limit cycles.

A proposed hybrid algorithm was shown to combine the advantages of the ordinary and the convergent list-mode algorithms (i.e. it was able to reach a higher image quality in fewer iterations while maintaining a convergent behavior), making it a good alternative to the ordinary subsetized list-mode EM algorithm.

5. Motion Compensation: Beyond the Event-Driven Approach

This chapter represents the core of this thesis work in the proposal and validation of accurate and practical motion compensation techniques applicable to high resolution 3D PET. The proposed techniques along with the experimental results have been published in [41].

5.1 Introduction

Recent developments in 3D positron emission tomography (PET) systems have enabled the spatial resolution to reach the 2-3mm FWHM range. With such improvements in spatial resolution, small patient movements during PET imaging become a significant source of resolution degradation. One method to correct for patient movement involves gating of detected events into multiple acquisition frames (MAF), with the use of an external monitoring system, followed by spatial registration and then summation of reconstructions from the acquired frames [145, 146].

However, the major limitation of the MAF approach is that it is unable to correct for motion during a frame. Subsequently, in the presence of considerable movement, one may need to acquire many *low-count* frames, and subsequently spatially align them, to better correct for patient motion. Lack of an adequate number of acquired events in the individual frames can in turn adversely affect the quality of the final reconstructed image, and an increased number of frames will also lead to increased (histogram-mode) reconstruction times.

136

Correction of individual lines-of-response (LORs) for motion has alternatively been suggested to achieve optimal motion correction [147]. This is what we shall refer to as the *event-driven approach*, by which it is meant that in either of the histogram-mode or list-mode reconstruction techniques, motion correction is performed by transforming the LORs along which the events *are* measured to where they *would* have been measured had the object not moved. To this end, motion-tracking systems have been used for accurate real-time measurements of position and orientation of the patient (see, for instance, Ref. [148] for *Polaris*, a system based on opto-electronic position sensitive detectors). In the histogram-mode approach, this introduces an event-by-event rebinning technique, resulting in motion-compensated sinograms, which are are subsequently reconstructed using any of the common reconstruction algorithms [149].

However, it can be argued that the purely event-driven approach can produce image artifacts as compared to a more comprehensive modeling of the image-data relation, as we also demonstrate experimentally in this work. This is because one can observe [150,151] that certain events that would have been detected in some LORs may have exited the PET scanner (e.g. axially or through detector gaps) undetected due to object motion. Regular reconstruction methods do not employ knowledge of missing data, and therefore would assume simply that nothing was detected.

One further notes that the converse also requires consideration: upon transformation of certain LORs along which the events are detected to LORs along which the events would have been detected, had the object not moved, the calculated LORs may correspond to no actual detector pairs. In other words, motion can result in some LORs which correspond to undetectable regions to exhibit non-zero probabilities of detection, which also needs to be properly modeled.

Taking these effects into consideration, it is therefore not sufficient to

merely correct *the events* for motion, rather one must also modify the algorithms themselves. Sections 5.2 and 5.3 discuss incorporation of a comprehensive modeling of motion into the histogram-mode and list-mode EM reconstruction. Both sections also describe (subsection B) an alternative, imagespace based method that has the potential of being considerably faster in presence of frequent motion especially in high resolution tomographs. Methods and results for experiments on transaxial and axial motion are presented in sections 5.4 and 5.5.

5.2 Motion Correction in Histogram-mode EM Reconstruction

Similar to the previous chapter, denoting λ_j^m as the image intensity estimate in voxel j (j=1...J) at the m^{th} iteration, and n_i as the number of events detected along LOR i (i=1...I), the histogram-mode EM algorithm can be written as

$$\lambda_{j}^{m+1} = \frac{\lambda_{j}^{m}}{\sum_{i=1}^{I} p_{ij}} \sum_{i=1}^{I} p_{ij} \frac{n_{i}}{\sum_{b=1}^{J} p_{ib} \lambda_{b}^{m}}$$
(5.1)

where p_{ij} is the probability of an emission from voxel j being detected along LOR i, and constructive summation is performed over those LORs for which $p_{ij} \neq 0$.

In the weighted scheme, the system matrix $P=(p_{ij})_{I\times J}$ can be written as P=WG where $G=(g_{ij})_{I\times J}$ denotes the geometric probability of an event generated at voxel j to be detected at LOR i, and the diagonal matrix $W=(w_i)_{I\times I}$ allows a weight to be assigned to each LOR, to incorporate normalization and attenuation effects. One then arrives at:

$$\lambda_{j}^{m+1} = \frac{\lambda_{j}^{m}}{\sum_{i=1}^{I} w_{i} g_{ij}} \sum_{i=1}^{I} g_{ij} \frac{n_{i}}{\sum_{b=1}^{J} g_{ib} \lambda_{b}^{m}}$$
(5.2)

wherein the sensitivity correction factor $s_j = \sum_{i=1}^{I} p_{ij}$ is a summation over all possible measurable LORs (i=1...I) and calculates the probability of an emission from voxel j being detected anywhere.

Here we make an observation: For any LOR i not corresponding to actual detector pairs, w_i is zero and therefore may not be canceled in the numerator and denominator of the EM algorithm. However, this is not a point of concern since such an LOR would not receive any counts, i.e. $n_i=0$, and therefore, the back-projection summation is performed only over those LORs that can correspond to actual detector pairs. This will not be the case in presence of motion, as we discuss later

5.2.1 Modification of the Histogram-mode EM Algorithm

The event-driven histogram-mode motion correction EM scheme (as for instance elaborated in Ref. [152]) consists of reading each list-mode event, producing correction factors (CFs) for the particular event (such as normalization), transforming the LOR in which the event is detected into the LOR in which the event would have been detected had the object not moved, and finally histogramming the event at this calculated LOR along with the application of the CFs. The motion-corrected sinograms, which have also been pre-corrected for the CFs, are subsequently reconstructed using the unweighted EM algorithm given by Eq. (5.1).

Once a motion-corrected sinogram is obtained, some LORs may be transformed into undetectable LORs (i.e. LORs for which no actual pair of detectors exist). On the other hand, some detectable LORs exhibit fewer counts than they would have if the object had not moved, since corresponding events have passed through undetectable LORs. These missing LORs can in principle be located: i) radially out of the field of view. ii) axially out of the scanner. iii) exceeding the maximum allowed ring difference in the scanner's acquisition mode, or iv) in regions corresponding to gaps in between the detectors.



Fig. 5.1: An event detected at an LOR i' generated in voxel j, which has been translated to voxel $j' = \mathcal{M}(j)$ at time of detection, would have been detected at LOR $i = \mathcal{L}^{-1}(i')$ if the object had not moved.

The first case may be ignored as it can be safely assumed that the object stays in the radial field of view of the scanner all the time, whereas the second and third cases are especially important in 3D PET imaging in which small rotations in the object can result in many LORs to exit the FOV axially or exceed the maximum allowed ring difference, and vice versa. The last possibility is expected to be significant in non-cylindrical designs. In the octagonal design of the high resolution tomograph (HRRT) [35], for instance, gaps existing between the eight detector heads of the scanner occupy over 10% of the sinogram space, as shown for a typical sinogram in Fig. (1.14).

One recently suggested method by Thielemans et al. [151] involves scaling of counts recorded in the motion corrected sinogram bins, where the scale factors are computed using averaging of LOR weighting factors. This can be thought of as a "motion pre-correction" technique, which requires consideration of noise enhancement due to small scale factors division. The authors have then suggested use of forward-projection of an initial estimate of the image, in combination with or replacing the calculated scale factors (by which the data are divided) when these factors are below a certain threshold.

We now note that such consideration can alternatively be fully addressed by modifying the system matrix itself so as to take motion into account ("motion weighting" of the system matrix as compared to "motion pre-correction" of sinograms). This, as we shall demonstrate, results in modification of the sensitivity correction factors, and can potentially result in less noise artifacts. This is because, compared to individual scaling of sinogram bins, only the summation of motion-compensation factors over *all* directions is used¹. Furthermore, as we demonstrate in Sec. 5.2.3, calculation of such correction can be performed in the image-domain, compared to the LOR-domain, and can thus potentially result in considerable speed-up of the algorithm.

In this regard, we first introduce an invertible operator \mathcal{L} which models the motion of the object by transforming the LOR *i* along which an event would have been detected in absence of any motion, to the LOR *i'* along which the event *is* detected at time *t*. An instance of this is shown in Fig. (5.1), where an event generated from a voxel *j*, currently located at a position *j'* due to some transformation $\mathcal{M}(j)$, has been detected along an LOR *i'*, and therefore must be histogrammed along the motion-corrected LOR $i=\mathcal{L}^{-1}(i')$, as is shown.

We next introduce $\tilde{g}_{i'j}^t$ as the motion-dependent geometric probability of detecting an event generated in voxel j along an LOR i' at a given time t. The superscript t indicates knowledge of object orientation and position with

¹ This observation is conceptually very similar to relative noise reduction in "weighted" schemes (*attenuation/normalization* correction) as compared to pre-correction schemes. Instead in this work, we are concerned with *motion* correction.

respect to the origin of the time $axis^2$. Also, defining

$$\delta_i = \begin{cases} 1 & \text{if LOR } i \text{ corresponds to a detector pair} \\ 0 & \text{otherwise} \end{cases}$$
(5.3)

such that $\{i|\delta_i = 0\}$ is the set of all LORs that correspond to no actual detector pairs, we write

$$\tilde{g}_{ij}^t \equiv g_{ij}^t \delta_i \tag{5.4}$$

where we use g_{ij}^t to calculate the purely geometric overlap between a voxel j and an LOR i, and δ_i is used to incorporate whether/not the LOR i can be detected by the scanner.

Next we note that according to the frame of reference of a given voxel j, from a purely geometric point of view, an LOR i' at time t is equivalent to a calculated LOR $i=\mathcal{L}_t^{-1}(i')$ at time t=0, since all LORs at time t=0 have now been transformed by the motion operator \mathcal{L}_t }. We thus note that the geometric overlap between a voxel j and an LOR i' at time t, is equal to the overlap it would have had with the calculated LOR i had the object not moved (which is the case at t=0); that is

$$g_{i'j}^t = g_{ij}^0 \quad \text{where} \quad i = \mathcal{L}_t^{-1}(i') \tag{5.5}$$

Denoting N_i and A_i as the attenuation and detector normalization factors for an LOR i, incorporation of these factors in presence of motion must now be addressed. In the unweighted scheme, these factors are applied onto the acquired LOR events that are histogrammed into appropriate motioncompensated sinogram bins. Alternatively, in weighted schemes, one or both of these factors are instead used inside the reconstruction algorithm itself. The normalization factor for an LOR i' along which an event is detected

 $^{^{2}}$ It is often assumed that the patient does not move between the attenuation scan and the start of the emission scan. Nevertheless, the more general case of having the patient move between the two scans can be treated by motion-correcting with respect to the time at which the attenuation scan was performed.



Fig. 5.2: An event detected at time t_2 along an LOR i', which would have been detected along an LOR i if the object had not moved, has the same attenuation correction factor as measured for A_i at time t_1 . The normalization correction factor however is LOR-specific and changes with motion.

is given by the value of $N_{i'}$ for the LOR itself, independent of any motion. However, this is *not* the case for attenuation correction, in which case the factor is given by the value of the attenuation factor at LOR *i* along which the event would have been detected had the object not moved. This is depicted in Fig. (5.2).

It follows that the weighting factor $w_{i'}$ for any LOR $i' = \mathcal{L}(i)$ can be written as

$$w_{i'} = \begin{cases} N_{i'} & \text{N-weighted} \\ A_i N_{i'} & \text{AN-weighted} \end{cases}$$
(5.6)

Introducing $\overline{P} = (\overline{p}_{ij})_{I \times J}$ as the motion-compensated system matrix, \overline{p}_{ij} must indicate the probability, for the course of the *entire* scan, that an event generated in object voxel j is finally binned into LOR i. It must therefore incorporate time-weighted probability-of-detection $(w_{i'}\tilde{g}_{i'j}^t)$ contributions from any LOR i' that could record events that would have been detected in LOR i had the object not moved; i.e.

$$\overline{p}_{ij} = \frac{1}{T} \int_0^T w_{i'} \tilde{g}_{i'j}^t dt$$
(5.7)

It then follows that Eq. (5.7) combined with Eqs. (5.4) and (5.5) can be written as

$$\overline{p}_{ij} = \frac{1}{T} g_{ij}^0 \int_0^T w_{i'} \delta_{i'} dt$$
(5.8)

In the AN-weighted scheme, for instance, upon replacing p_{ij} in Eq. (5.1) with \overline{p}_{ij} and dropping the superscript in g_{ij}^0 for convenience, one obtains, after cancellations, the following AN-weighted reconstruction algorithm:

$$\lambda_j^{m+1} = \frac{\lambda_j^m}{\overline{s}_j} \sum_{i=1}^I g_{ij} \frac{n_i}{\sum_{b=1}^J g_{ib} \lambda_b^m}$$
(5.9)

where

$$\overline{s}_{j} = \frac{1}{T} \sum_{i=1}^{I} g_{ij} A_{i} \int_{0}^{T} N_{i'} \delta_{i'} dt$$
(5.10)

5.2.2 A Note on Convenience and Accuracy

It must be noted, in histogram-mode reconstruction, that in order to make use of all measured events, the regularly employed sinogram-space has to be extended in order to allow histogramming of all motion-compensated LORs, including those that do not correspond to existing detector pairs. In other words, the back-projection summation in the numerator of Eq. (5.9) needs to be performed over *all* observed counts, and not merely those that correspond to detector pairs³. List-mode reconstruction of the data, therefore, is somewhat more convenient, in that such sinogram extension is not required since the individually measured events are motion-compensated and processed one-by-one.

³ Whenever $n_i \neq 0$, it is necessarily the case that $\overline{p}_{ij} \neq 0$ and therefore cancellation of weighting components of \overline{p}_{ij} in the back- and forward-projection steps is valid

Another observation may also be made: the LOR i (calculated from the measured LOR i') does not typically correspond exactly to the center of a sinogram bin and an interpolation needs to be performed upon histogramming the event into motion-corrected sinograms. However, as described in Sec. (5.3), list-mode event coordinates, upon being motion-corrected, can be maintained as continuous variables in list-mode reconstruction, therefore pointing to a potential advantage in terms of inherent accuracy in list-mode reconstruction.

5.2.3 An Alternate, Fast Method for Calculation of Sensitivity Factors

For a given voxel j, the task of calculating the motion-weighted sensitivity correction factor \bar{s}_j (Eq. 5.10) requires integrating over the entire duration of the scan to derive the properly motion-corrected normalization factors, which are subsequently back-projected along with the attenuation factors. This calculation can therefore be very time-consuming for frequent motion in high resolution scanners. In what follows we propose an alternative approach in the calculation of motion-weighted sensitivity factors, which can potentially result in very efficient calculation times for high resolution scanners, as we explain. The technique is applicable to the case where *all* measured LORs are histogrammed and considered in the reconstruction task, including those that do not correspond to actual detector pairs after being corrected for motion.

We make the following very useful observation: at any given time t, calculation of the time-dependent geometric factor $g_{i'j}^t$ can be performed in another way: instead of mapping LOR i' into a motion-corrected LOR i (the approach taken in Eq. 5.5), one can map the object voxel j to the new voxel j' it has moved to at time t; i.e.

$$j' \equiv \mathcal{M}_t(j) \tag{5.11}$$

where \mathcal{M}_t is an image-space based motion-tracking operator, as depicted in Fig. 5.1. Mathematically, we have the following identity:

$$g_{ij} = g_{i'j'} \tag{5.12}$$

The above relation can lead to considerable speed increase in calculation of s_j for schemes in which data are pre-corrected for attenuation (i.e. unweighted and N-weighted schemes), as we show here. In the N-weighted scheme, for instance, we have

$$\lambda_j^{m+1} = \frac{\lambda_j^m}{\overline{s}_j} \sum_{i=1}^I g_{ij} \frac{n_i / A_i}{\sum_{b=1}^J g_{ib} \lambda_b^m}$$
(5.13)

where \overline{s}_j is given by

$$\overline{s}_{j} = \frac{1}{T} \sum_{i=1}^{I} g_{ij} \int_{0}^{T} N_{i'} \delta_{i'} dt = \frac{1}{T} \int_{0}^{T} \sum_{i=1}^{I} g_{ij} N_{i'} \delta_{i'} dt$$
(5.14)

We next note that the summation over all *i* in the above equation involves consideration of all $i' = \mathcal{L}(i)$, and thus the summation index *i* can as easily be replaced by *i'*. Combining this observation with Eq. (5.12) gives

$$\overline{s}_j = \frac{1}{T} \int_0^T \sum_{i'} g_{i'j'} N_{i'} \delta_{i'} dt$$
(5.15)

and therefore

$$\overline{s}_j = \frac{1}{T} \int_0^T s_{j'} dt \text{ where } s_j = \sum_{i'} g_{i'j} N_{i'} \delta_{i'} = \sum_i g_{ij} N_i \delta_i \tag{5.16}$$

while noting that s_j is simply the standard sensitivity correction factor which needs to be calculated *only once* for all voxels.

In other words, in the calculation of \overline{s}_j for any voxel j, instead of having to perform time-averaging in the LOR domain, one can do so in the image domain, by evaluating sensitivity factors (which are calculated once for the object at t=0) at voxels j' over time. This corresponds to tracing the motion of voxel j with time, as measured by the motion-tracking system. Comparing Eqs. (5.9) and (5.13), we make the following observations: (I) The first proposed algorithm does allow for incorporation of attenuation into the system matrix, whereas the latter requires pre-correction of the emission data for attenuation. This can result in an increase in image noise (e.g. see Ref. [133]), though less increase is expected with better accuracy and statistics in the attenuation and emission measurements. On the other hand, the gain in computation efficiency, especially in full motion compensation, can be very significant, as we shall argue for.

(II) The sensitivity correction factor in the first algorithm (Eq. 5.10) involves, for each LOR *i*, a time-integral over $N_{i'}\delta_{i'}$ for all LORs *i'* along which LOR *i* has come to be aligned due to motion at some point during the scan. In other words, the motion-tracking operator \mathcal{L}_t must be applied for all the times *t* considered to every LOR *i*. Following this time-averaging, the resulting factors are back-projected. In the second algorithm, on the other hand, the sensitivity correction factor (Eq. 5.16) is given by a time-integral over the standard factors s_j (which is computed once by back-projection of $N_i\delta_i$) for all voxels j' to which voxel j had moved at some point during the scan.

Since with current, high-resolution scanners, the LOR domain is typically much larger than the image domain (for instance, in the HRRT, with span 3 and maximum ring difference of 67, one deals with \sim 470M LORs whereas one has 256x256x207 \sim 14M image voxels), the second approach implies a considerable potential speed-up in full motion compensation of motion-corrected sinograms. This speed advantage increases with the number of motionmonitoring intervals (i.e. with degree of "exact-ness" in monitoring of motion) used in the computation. Nevertheless, scanner, motion and task-dependent studies remain to be performed in order to investigate the trade-off between image noise (first observation) and computation speed (second observation). As we shall show, this technique will also be applicable to list-mode recon-

147

struction.

5.3 Motion Correction in List-mode EM Reconstruction

One must observe that while motion-compensated histogramming involves interpolating the transformation $\mathcal{L}_t^{-1}(i')$ into an actual sinogram bin *i*, in the case of list-mode reconstruction one can work with LOR coordinates, as compared to rebinned LOR positions, thus potentially preserving a higher degree of accuracy in the reconstruction task. In other words, this improvement results directly from better sampling of the measurement. In this regard, the probability element g_{ij} can be replaced by the operation $\mathcal{G}_j(\mathbf{i})$, which calculates the geometric probability of an event generated at voxel j to occur at LOR $\mathbf{i}=\mathcal{L}_t^{-1}(i')$, a continuous variable, holding the exact coordinates of LOR i' after being motion-corrected.

The list-mode expectation maximization (LM-EM) reconstruction algorithm has been previously formulated by Parra and Barrett [92]. Investigation of the applicability of the algorithm and its accelerated versions are nowadays common in the literature (e.g. see Refs. [70, 124, 153]). Following the list-mode approach, we are able to show in Appendix D that full consideration of motion results in the following EM algorithm:

$$\lambda_j^{m+1} = \frac{\lambda_j^m}{\frac{1}{T} \int_0^T s_j^t dt} \sum_{k=1}^N \mathcal{G}_j(\mathbf{i}_k) \frac{1}{\sum_{b=1}^J \mathcal{G}_b(\mathbf{i}_k) \lambda_b^m}$$
(5.17)

where $\mathbf{i}_k = \mathcal{L}_t^{-1}(i'_k)$ with i'_k denoting the LOR along which the *k*th event is detected (k=1...N), and s^t_j is a time-dependent sensitivity correction factor: the probability at time *t* that an emission from voxel *j* is detected anywhere.

The overall time-averaged sensitivity correction factor $\overline{s}_j \equiv \frac{1}{T} \int_0^T s_j^t dt$ can be calculated on an LOR-domain approach such that any LOR i' is transformed to the corresponding motion-corrected LOR $i=\mathcal{L}_t^{-1}(i')$ for the calculation of time-dependent attenuation and geometric factors; i.e.:

$$\overline{s}_{j} = \frac{1}{T} \int_{0}^{T} s_{j}^{t} dt = \frac{1}{T} \int_{0}^{T} \sum_{i'} g_{ij} A_{i} N_{i'} dt$$
(5.18)

which has also been suggested by Qi and Huesman [154] in order to maximize the log-likelihood function of list-mode data.

Nevertheless, similar to the approach in Sec. (5.2.3), we notice that in the N-weighted scheme, again using Eq. (5.12), one can write

$$\overline{s}_{j} = \frac{1}{T} \int_{0}^{T} \sum_{i'} g_{ij} N_{i'} dt = \frac{1}{T} \int_{0}^{T} \sum_{i'} g_{i'j'} N_{i'} dt$$
(5.19)

which reproduces Eq. (5.15). Therefore, we propose the following N-weighted algorithm:

$$\lambda_{j}^{m+1} = \frac{\lambda_{j}^{m}}{\overline{s}_{j}} \sum_{k=1}^{N} g_{j}(\mathbf{i}_{k}) \frac{1/A_{i_{k}}}{\sum_{b=1}^{J} g_{b}(\mathbf{i}_{k})\lambda_{b}^{m}}$$
(5.20)

where the overall sensitivity correction factors \overline{s}_j are given by Eq. (5.16).

5.4 Methods

Phantoms used and measurements performed

The studies performed in this work involve phantom studies with manual repositioning of the source. Two separate studies were performed, one involving transaxial rotation, and the other, axial translation. The technique described in this work, however, is *fully applicable to 3D motion*, which is the case commonly encountered in practice. Upon the availability of a motion tracking system to us in the near future, the experimental studies will be extended to the general 3D case. The two experiments are elaborated below.

Experiment 1 – Transaxial Rotation

An F-18 line source was inserted axially into a 70-cm long NEMA phantom (20 cm diameter), placed 4cm away from and parallel to the central axis of

the cylinder. The measured true count rate was 124k/s with a random fraction of 13%. Three separate frames each of duration 4 minutes were acquired with the cylinder manually rotated each time by approximately 45 degrees around the z-axis.

Experiment 2 – Axial Translation of an Extended Source

Via the technique presented in Appendix C, which allows printing of radioactive patterns using a modified standard ink-jet printer, we imaged an F-18 "W"-sign (\sim 4x4cm, 4.81mCi) printed on regular paper. Scans were performed with the source positioned (i) nearly at center of the scanner (1 minute), and also shifted axially by (ii) 5cm (3 minutes) and (iii) 9cm (3 minutes) with respect to the first frame.

In both studies, the separate motion frames were then combined into one single frame to study the proposed motion correction techniques. Detector normalization correction factors were obtained from a 12 hour scan using a rotating rod source.

Motion Measurement and Compensation

The center of the cylinder used in the first study was expected to have undergone small translations during the manual rotation procedure. In this regard, exact inter-frame rotations and translations of the phantom were determined by comparison of separate reconstructions for the three frames. The center of the line source reconstructed in each frame was found by means of fitting a Gaussian profile. The relative translational and rotational shifts in-between the three centers were then found by means of matching one with another. The motion coordinates (with respect to the first frame) for the second and third frames were found to be: $\Delta X=3, \Delta Y=7, \ \Delta \theta=40^{\circ}$ and $\Delta X=9, \Delta Y=3, \Delta \theta=85^{\circ}$, respectively, where the displacement units (i.e. voxel dimensions) are 1.2mm.

We denote the LOR coordinate system using (r, θ, z, δ) where r and θ contain the transaxial radial and angular coordinates, while z and δ define the axial position of the center of the LOR and the ring difference between the end-points. In the aforementioned case of transaxial motion, the two elements of the LOR affected by such motion are r and θ . These two follow the relation

$$r = X\cos(\theta) + Y\sin(\theta) \tag{5.21}$$

for an LOR passing through an image voxel with coordinates (X, Y, Z).

It follows that for an event detected at a given time for an object rotated (counterclockwise) by an angle $\Delta\theta$ with respect to its position at t=0, the appropriate LOR-domain motion operator becomes

$$\mathcal{L}^{R}(r,\theta,z,\delta) = (r,\theta + \Delta\theta, z,\delta)$$
(5.22)

whereas for an object translated by $(\Delta X, \Delta Y)$, we would have

$$\mathcal{L}^{T}(r,\theta,z,\delta) = (r + \Delta X \cos(\theta) + \Delta Y \sin(\theta), \theta, z, \delta)$$
(5.23)

where the superscripts R and T represent the transaxial rotation and translation operations. Similarly, the equivalent image-domain motion operators are

$$\mathcal{M}^{R}(X, Y, Z) = (\cos(\Delta\theta)X - \sin(\Delta\theta)Y, \\ \sin(\Delta\theta)X + \cos(\Delta\theta)Y)$$
(5.24)

and

$$\mathcal{M}^{T}(X,Y) = (X + \Delta X, Y + \Delta Y)$$
(5.25)

In the axial motion study, the motion operator for an axial shift by ΔZ is given by

$$\mathcal{L}^{A}(r,\theta,z,\delta) = (r,\theta,z+\Delta Z,\delta)$$
(5.26)

where the superscripts A represent the axial translation operations, while the equivalent image-domain motion operator is

$$\mathcal{M}^A(X,Y,Z) = (X,Y,Z + \Delta Z) \tag{5.27}$$

The aforementioned LOR-domain and image-domain operators were subsequently used for incorporation into the proposed histogram-mode and listmode reconstruction methods.

Reconstruction Schemes

Throughout the studies presented in this work, a span of 3 and maximum ring difference of 67 were used. The motion-compensated histogram-mode (Eq. 5.13) and list-mode (Eq. 5.20) algorithms proposed in this work were implemented on the HRRT. In chapter 4, details of implementation and validation of the regular list-mode EM algorithm with random events correction for the HRRT were discussed, and the statistical properties were compared with histogram-mode (3D-OSEM) reconstruction.

Four schemes were considered in each of the histogram-mode and listmode tasks: (I) Data for the non-moving phantoms (single phantom position) were first reconstructed. In addition, data for the motion studies were reconstructed (II) without any motion correction, (III) using the conventional purely event-driven technique (i.e. without modification of the sensitivity correction factors), and (IV) with the proposed motion correction algorithms. In histogram-mode reconstruction, two instances of scheme III were considered: (a) one in which data allocated to the HRRT detector gaps, after motion corrected histogramming, were ignored and (b) when gap data were not dropped and were used in the reconstruction. The results were subsequently compared using measurement of relative count loss and resolution (line-source study) as well as radial/axial profile comparisons.



Fig. 5.3: (Top) Sinograms of a moving line source (a) without and (b) with motion-compensated histogramming. Clearly, upon appropriate histogramming, some counts are histogrammed into bins corresponding to detector gaps (shown by arrows), signifying that they would not have been detected had the object not moved. (Middle) Images reconstructed using histogrammode algorithm applied to (c) data with no motion and (d) data for motion when reconstructed without motion correction. (Bottom) (e) Image resulting from the proposed reconstruction algorithm given by Eq. (5.13). Radial profiles through this image as well as the one resulting from no motion data are shown in (f). Three iterations were applied to the data. Direct plane #110 is shown (randomly selected).



Fig. 5.4: Images reconstructed using list-mode algorithm applied to a) data with no motion, as well as data for motion when reconstructed b) without and c) with motion correction. (d) Radial profiles through imaged line source in cases a) and c) are shown. Three iterations were applied to the data. Plane 110 is shown (randomly selected).

5.5 Results and Discussion

Histogram-mode Reconstruction

Experiment 1: Fig. (5.3a) shows a typical sinogram acquired without any motion-correction applied to the line source transaxial motion data. The presence of three distinct sinogram patterns as well as the detector gaps are clearly observable.

Fig. (5.3b) shows the resulting sinogram when motion-compensated histogramming is performed. The figure clearly exhibits non-zero counts in certain histogram bins corresponding to detector gaps. Such events, not corresponding to *any* existing detector pairs, are detected due to object motion, as predicted in Sec. 5.1.

Reconstruction of the motion-compensated sinograms, while ignoring data allocated to detector gaps, using the conventional event-driven technique (i.e. scheme III while ignoring gap data) was seen to considerably underestimate the image intensity (by 18%), compared to the reference image (scheme I). On the other hand, for the reconstruction schemes III (conventional, this time *not* ignoring gap data) and IV (proposed), count losses of 2% and 1% were observed, respectively. Furthermore, scheme III resulted in a reconstructed image resolution of 4.5mm, while scheme IV showed a resolution of 4.4mm. In comparison, scheme I produced a resolution of 4.1mm.

Thus, it was observed that ignoring gap data resulted in a reconstructed count bias (plus clearly, discarding of a portion of measured data corresponds to a loss in statistics). Furthermore, schemes III and IV did not result in noticeably different image qualities. However, as we shall demonstrate, in the axial motion study with an extended object, the two schemes result in significantly different final images.

Figs. (5.3c,d,e) show final images reconstructed using schemes I, II and IV. Radial profiles through Figs. (5.3c&e) (i.e. no motion study vs. proposed algorithm) are also shown in Fig. (5.3f). Comparing the two profiles, a slight broadening of resolution, from 4.1mm in scheme I to 4.4mm in scheme IV, was measured which can likely be attributed to the nature of motion measurement we have used (since an accurate motion tracking system is not yet available to us).

List-mode Reconstruction

Experiment 1: Similar to the previous analysis, images reconstructed from

the line source (transaxial motion study) using schemes (I), (II) and (IV) are shown in Figs. (5.4a,b,c). The total number of counts in images reconstructed using the three schemes were confirmed to be within 0.5%. Radial profiles were subsequently drawn through resulting images from schemes (I) and (IV), as shown in Fig. (5.4d). Scheme I resulted in an imaged resolution of 4.0mm, while scheme IV showed a slightly-broadened resolution of 4.3mm. This, we attribute to the nature of motion measurement we have used, as previously mentioned.

However, it was again observed that the conventional scheme (III) did not result in a noticeable decline in image quality. That is, a count loss of 1.5% and resolution of 4.4mm were measured compared to the proposed scheme (IV), which resulted in a count loss of 0.5% and a resolution of 4.3mm. This, as in histogram-mode study, is likely due to the transaxial nature of motion in the line source, and as we demonstrate below, the difference is very significant in the next study.

Experiment 2: For the extended source motion study, resulting final images reconstructed using the four schemes (I), (II), (III) and (IV) are shown in Figs. (5.5a,b,c and d). One can clearly observe a non-uniformity in the reconstructed radioactive "W"-sign, when using the conventional scheme (III), whereas this is not the case in the scheme (IV) proposed in this work. Furthermore, upon drawing profiles through images from schemes (III) and (IV) as shown in the figures, one clearly observes, in addition to the axial non-uniformity artifact visible in the conventional scheme, an overall quantitative loss, measured to be 17% at axial plane 90 (corresponding to the left edge of the sign). Combined with the non-uniformity artifact, the image voxels are underestimated by nearly 50% at axial plane 150 (the right edge of the sign).

These artifacts can be explained by the fact that in the conventional scheme, the sensitivity factors are not corrected for in accordance with object motion, and therefore result in overall count underestimation as well as nonuniformity upon motion towards less sensitive regions in the FOV. We have thus shown in the new experimental data, that the nature of the inaccuracy is *not simply a global scaling* and can result in loss of uniformity, whereas the proposed algorithm, achieves better uniformity.

On a last note, we point out that scatter correction is yet to be incorporated into the proposed algorithms in this work. We are currently evaluating different schemes. One possible approach has already been developed for the HRRT, namely the Watson scatter correction technique [28]. The algorithm is image-based (i.e. computes the scatter contribution using the reconstructed image) and therefore has the practical advantage of being directly applicable to images generated using our proposed methods.

5.6 Conclusion

In this chapter, we have argued for and experimentally demonstrated that the event-driven approach to motion correction, in which one *merely* transforms LORs along which events are detected to LORs along which they would have been detected, can result in image artifacts. The nature of the inaccuracy was shown not to be necessarily restricted to a global scaling, and could manifest itself, for instance, as a loss in uniformity. This inaccuracy was attributed to the existence of LORs corresponding to no actual pairs of detectors (e.g. LORs axially out of the FOV) and their "interaction" with the detectable LORs due to motion.

To address the issue, a more comprehensive modeling of the image-data relation was considered. It was subsequently shown that appropriate system matrix modeling of the aforementioned effects into the histogram-mode as well as list-mode reconstructions introduced time-weighted sensitivity correction factors. These factors needed to be implemented *in addition* to compen-



Fig. 5.5: Images reconstructed for a radioactive printed "W"-sign (coronal view) when (a) no motion is introduced; and with motion (three axial positions: 1 min - 3 min - 3 min) when (b) no motion compensation, (c) the purely event-driven approach (conventional motion correction) and (d) the proposed algorithm given by Eq. (5.20) are applied to the data. Plots of uniformity profiles drawn through images in (c) and (d) are shown in (e). Five iterations were applied to the data.

158

sation of the measured events for motion, with the conventional event-driven approach *only* performing the latter.

The resulting motion-compensated sensitivity factors, at first instance, appeared to be potentially very time-consuming to compute for the general case of frequent motion throughout the scan. However, closer inspection of the sensitivity factors revealed that use of attenuation pre-correction for the data would allow computation of time-averaging of sensitivity correction factors to be performed in the image-domain. This was shown to potentially reduce reconstruction time significantly when correcting for frequent motion especially in high resolution scanners.

6. Summary, Opinion and Conclusions

6.1 Summary

This section briefly summarized the main points from the chapters in this thesis.

Introduction to PET Imaging (Chapter 1):

In this chapter, we introduced the PET imaging technique. Main applications of PET imaging, along with a summary of the main characteristics of modern PET scanners were presented. Motivations for and challenges in switching from 2D PET to 3D PET imaging were also discussed. The importance of dynamic PET imaging was explained. The chapter also elaborated upon the causes of degradation in image quality and quantitative accuracy. Finally, details of the high resolution research tomograph (HRRT) were explained and the challenges in imaging with this scanner were mentioned.

3D Image Reconstruction (Chapter 2):

In this chapter, we elaborated upon the two main approaches to 3D image reconstruction: analytic and statistical. Analytic techniques (3D reconstruction as well as rebinning algorithms followed by 2D reconstruction) were introduced. We focused our attention, however, on statistical image reconstruction techniques, which were seen to be particularly suitable for imaging using the HRRT. Advantages of these techniques, their important properties, and the various approaches within statistical image reconstruction were
discussed. We paid particular attention to the various classes of expectation maximization (EM) algorithms, and introduced the importance of convergent image reconstruction, and of dealing with the various issues in the EM reconstruction technique.

List-mode Image Reconstruction (Chapter 3):

In this chapter, we introduced the concept of list-mode acquisition, and the possibility of list-mode reconstruction. Various advantage for the list-mode reconstruction techniques were propounded. We argued that the statistical list-mode reconstruction technique is very much suitable for reconstruction using the state-of-the-art HRRT, which contains a very large number of possible lines-of-response (LORs). Various accelerated list-mode reconstruction algorithms were proposed (experimentally tested in the next chapter), with particular emphasis on convergent accelerated image reconstruction. A practical random correction technique was also proposed for use in list-mode reconstruction from the HRRT data.

Reconstruction Implementation and Comparison (Chapter 4):

In this chapter, implementation details necessary for the task of list-mode image reconstruction were depicted. Various geometric back- and forwardprojection techniques were proposed, and amongst them, the bilinear interpolation technique was selected as the method of choice, due to its relatively powerful combination of speed and accuracy. The method was compared to histogram-mode OSEM reconstruction, commonly used with a sinogram zero-thresholding criterion. We demonstrated that a milder non-negativity constraint can relax the overestimation bias encountered with OSEM reconstruction of low-count frames. We also extended the list-mode approach to dynamic (4D) approach, and demonstrated its robustness (uniformity in time-activity-curves, axial profiles, percentage contrast) compared to other

techniques.

Motion Compensation: Beyond the Event-Driven Approach (Chapter 5): In this chapter, we reviewed existing motion compensation techniques, and argued for the implementation of techniques that move beyond the purely event-driven motion compensation techniques. We proposed efficient and accurate motion compensation techniques (both in list-mode and histogrammode imaging), and demonstrated the relative improvements encountered in terms of image quality figures of merit.

6.2 Opinion and Suggested Future Directions

Our main challenge in the course of this work was the presence of numerous areas of further investigation applicable to our technique (and the obvious necessity not to pursue completely all areas of interest!). In what follows, we summarize main areas of further research that could yield important improvements to the list-mode image reconstruction technique proposed, implemented and studies in this thesis.

Use of splines in interpolation tasks:

We have already demonstrated how significant the effect of the interpolation method of choice in the forward and backprojection tasks can be on final reconstructed image qualities. In particular, we showed in Sec. (4.1.1) that switching from the Siddon method to bilinear interpolation improves the FWHM at the center of the FOV from 3.1 mm to 2.7 mm (by ~15%). One area of research of much potential is the use of more suitable interpolation techniques. Use of splines, along with their very useful properties, seems to be a promising area of research, with regenerated interest in the recent years [155, 156].

Ordinary Poisson List-mode Image Reconstruction:

We identify that the delayed events subtraction technique can be replaced by the ordinary Poisson technique for more accurate modeling of the Poisson process. The difficulties with this approach, as elaborated in Sec. (3.5.1), mainly involve the fact that in the HRRT the singles events are acquired at the block level, and not the crystal level, and thus estimation of the mean random rates at the crystal level is difficult. To counter this problem, a recently proposed approach¹ involves the use of delayed coincidences (which unlike the singles data are acquired at the crystal level) to estimate the singles rates at the crystal level. This technique, however, will result in increasing computational demands (in the calculation of the smoothed delayed coincidences, and constant look-up of calculated mean random coincidence rates).

Improved 4D Image Reconstruction:

The current approach we have chosen to imaging of the dynamic list-mode data involves *independent* reconstruction of the dynamic frames as specified by the user. However, it is worth noting that an alternate technique would involve using of the image reconstructed from a frame as an initial estimates for the image to be reconstructed for the subsequent frames: one may note that since iterative reconstruction techniques are often not run to convergence (see discussion in Sec. 2.5.2), the starting image estimate can play an important role in determining the accuracy of the final image [157].

In this alternate approach some type of filtering of each reconstructed image should be performed prior to use as an initial estimate for the subsequent frame, so as to avoid increasingly noisy images as further frames are reconstructed [158]. Nevertheless, this approach suffers from a systematic problem. The final image produced in a frame will be potentially biased by

¹ Internal documentation of the manufacturing company.

the initial estimate, and this, in our opinion, would produce problems when studying fast dynamics in which adjacent frames differ noticeably.

The ultimate approach, in our opinion, to reconstructing dynamic listmode data involves truly 4D spatio-temporal reconstruction. This technique would *not* involve framing of the data into dynamic frames (activities within which are assumed to be static), thus making maximal use of the temporal resolution of the scanner. Rather, it would involve the use of temporal basis function to represent the activity in each voxel, and estimating the coefficients of the basis functions by the *collective* use of the data. A possible approach to this problem is presented by Nichols *et al.* [159].

Scatter Correction:

Scatter correction remains to be implemented in the list-mode image reconstruction method for the HRRT. This requires much dedicated research, to be built upon previous work on the subject. We have already presented our views and suggested future directions in Sec. (3.6).

Modeling of the Detector Blurring:

In Appendix B, we described a method in which detector blurring (including the parallax effect) were modeled in the image domain, in order to avoid a highly non-sparse system matrix. However, computational issues are only temporary, and long-term research should not view them as permanent obstacles. In our view, a better approach to modeling of detector blurring would involve its effective modeling into the forward and back-projection operations. This method would involve extension of the strip-integral approach [160], which currently accounts for the finite detector width rather than using overly idealized line integrals, to account for inter-crystal blurring as well. In this way, the inter-crystal penetration (which results in the parallax effect), and the asymmetry associated with it (which is highly dependent on the individual LOR's angle of incidence), can also be incorporated into the system matrix.

6.3 Conclusion

The work presented in this dissertation has been a result of the need for 3D image reconstruction algorithms suitable for high resolution positron emission tomography (PET). This has been particularly the case for the high resolution research tomograph (HRRT): a 3D-only state-of-the-art dedicated brain tomograph, with a very large number of lines-of-response (LORs) which it is able to measure ($\sim 4.5 \times 10^9$), exceeding most modern PET scanners by 2-3 orders of magnitude. We proposed, studied and implemented statistical list-mode EM image reconstruction algorithms. In particular, we considered convergent accelerated algorithms, and demonstrated their applicability, especially when combined with the ordinary accelerated algorithms (i.e. hybrid ordinary-convergent) in terms of convergence properties and speed. Importantly, we also validated the technique as robust, efficient and powerful in dynamic PET image reconstruction.

With the spatial resolution in modern high resolution tomographs (including the HRRT) reaching the 2-3mm FWHM range, small patient movements during PET imaging can become a significant source of resolution degradation. We thus devoted a portion of this dissertation to the proposal of new, accurate and practical motion-compensation techniques, and studied them on the HRRT. We proposed and validated the benefits of modeling the motion into the reconstruction task, thus demonstrating that greater image quality can be gained by motion compensation techniques that move beyond the existing purely event-driven approach.

Bibliography

- [1] M. D. Harpen, "Positronium: Review of symmetry, conserved quantities and decay for the radiological physicist", *Med. Phys.*, vol. 31, pp. 57-61, 2004.
- [2] C. S. Levin and E. J. Hoffman, "Calculation of positron range and its effect on the fundamental limit of positron emission tomography system spatial resolution", *Phys. Med. Biol.*, vol. 44, pp. 781-799, 1999.
- [3] C. S. Lee, A. Samii, V. Sossi, T. J. Ruth, M. Schulzer, J. E. Holden, J. Wudel, P. Pal, R. de la Fuente-Fernadez, D. B. Calne, and A. J. Stoessl, "In vivo PET evidence for compensatory changes in presynaptic dopaminergic nerve terminals in Parkinson's disease", Ann. Neurol., vol. 47, pp. 493-503, 2000.
- [4] V. Sossi, R. de la Fuente-Fernndez, J. E. Holden, M. Schulzer, T. J. Ruth, and A. J. Stoessl, "Onset and progression of parkinsonian symptoms: a manifestation of impaired compensatory mechanisms", J. Cerebral Blood Flow Metab., vol. 24, pp. 869-876, 2004.
- [5] R. de la Fuente-Fernandez R, J-Q Lu, V. Sossi, S. Jivan, M. Schulzer, J. E. Holden, C. S. Lee, T. J. Ruth, D. B. Calne, and A. J. Stoessl, "Biochemical variations in the synaptic level of dopamine precede motor fluctuations in Parkinson's disease: PET evidence for increased dopamine turnover", Ann. Neurol., vol. 49, pp. 298-303, 2001.
- [6] R. de la Fuente-Fernndez, V. Sossi, Z. Huang, S. Furtado, J-Q Lu, D. B. Calne, and A. J. Stoessl, "Levodopa-Induced Changes In Synaptic Dopamine Levels Increase With Progression Of Parkinson's Disease: Implications For Dyskinesias", *Brain*, in press 2004.
- [7] C. L. Melcher and J. S. Schweitzer, "Cerium-doped lutetium orthosilicate: a fast, efficient new scintillator", *IEEE Trans. Nucl. Sci.*, vol. 39, pp. 502-505, 1992.
- [8] M. E. Casey, and R. Nutt, "A Multicrystal Two Dimensional BGO Detector System for Positron Emission Tomography", *IEEE Trans. Nucl. Sci.*, vol. 33, pp. 460-463, 1986.

- [9] I. Holl, E. Lorenz, S. Natkaniez, D. Renker, C. Schmelz, and B. Schwartz, "Some studies of avalanche photodiode readout of fast schintillators", *IEEE Trans. Nucl. Sci.*, vol. 42, pp. 351-356, 1995.
- [10] C. Schmelz, S. M. Braddury, E. Lorenz, D. Renker, and S. Ziegler, "Feasibility study of an avalanche photodiode readout for a high resolution PET with nsec time resolution", *IEEE Trans. Nucl. Sci.*, vol. 42, pp. 1080-1084, 1995.
- [11] K. Burr, A. Ivan, J. LeBlanc, S. Zelakiewicz, D. L. McDaniel, C. L. Kim, A. Ganin, K. S. Shah, R. Grazioso, R. Farrell, and J. Glodo, "Evaluation of a Position Sensitivie Avalanche Photodiode for PET", *IEEE Trans. Nucl. Sci.*, vol. 50, pp. 792, 2003.
- B. J. Pichler, B. K. Swann, J. Rochelle, R. E. Nutt, S. R. Cherry, and S. B. Siegel, "Lutetium oxyorthosilicate block detector readout by avalanche photodiode arrays for high resolution animal PET", '*PET. Med. Biol.*, vol. 49, pp. 4305-4319, 2004.
- [13] S. R. Cherry, J. A. Sorenson, and M. E. Phelps, "Physics in Nuclear Medicine", 3rd ed., Philadelphia: Saunders, 2003.
- [14] E. R. Carson, C. Cobelli, and L. Finkelstein, The Mathematical Modeling of Metabolic and Endocrine Systems, New York: Wiley, 1983.
- [15] S-ch. Huang and M. E. Phelps, "Principles of Tracer Kinetic Modeling in Positron Emission Tomography and Autoradiography" in *Positron Emission Tomography and Autoradiography: Principles and Applications for the Brain and Heart*, New York: Raven Press, 1986.
- [16] H. E. Johns and J. R. Cunningham, "The Physics of Radiology", Springfield, Illinois: Charles C Thomas Publisher, 1983.
- [17] B. Bendriem and D. Townsend, "The theory and practice of 3D PET", in Developments in Nuclear Medicine, Dordrecht, The Netherlands: Kluwer, 1998, vol. 32.
- [18] K. Lange, M. Bahn, and R. Little, "A theoretical study of some maximum likelihood algorithms for emission and transmission tomography", *IEEE Trans. Med. Imag.*, vol. 6, pp. 106-114, 1987.
- [19] H. Erdogan and J. A. Fessler, "Monotonic algorithms for transmission tomography", *IEEE Trans. Med. Imag.*, vol. 18, pp. 801-814, 1999.

- [20] A. J. Reader and C. J. Thompson, "Transmission Tomography Algorithm for 3D List-Mode or Projection Data", *IEEE NSS & MIC 2003 conf. rec.*, Portland, OR, Oct 2003.
- [21] C. J. Thompson, "The effect of collimation on scatter fraction in multi-slice PET", *IEEE Trans. Nucl. Sci.*, vol. 35, pp. 598-602, 1988.
- [22] T. J. Spinks, T. Jones, D. L. Bailey, D. W. Townsend, S. Grootoonk, M. C. Gilardi, M. E. Casey, B. Sipe, and J. Reed, "Physical performance of a positron tomograph for brain imaging with retractable septa", J. Comput. Assis. Tomogr., vol. 18, pp. 1004-1009, 1992.
- [23] D. L. Bailey and T. Jones, "A convolution-subtraction scatter correction method for 3D PET", Phys. Med. Biol., vol. 39, pp. 411-424, 1994.
- [24] A. J. Reader, Sha Zhao, P. J. Julyan, D. L. Hastings and J. Zweit, "Adaptive correction of scatter and random events for 3-D backprojected PET data", *IEEE Nucl. Sci. Trans.*, vol. 48, pp. 1350-1356, 2001.
- [25] S. R. Cherry, S. Meikle, and E. J. Hoffman, "Correction and characterization of scattered events in 3D PET using scanners with retractable septa", J. Nucl. Med., vol. 34, pp. 671-678, 1993.
- [26] S. Grootoonk, T. J. Spinks, D. Sashin, N. M. Spyrou, and T. Jones, "Correction for scatter in 3D brain PET using a dual energy window method", *Phys. Med. Biol.*, vol. 41, pp. 2757-2774, 1996.
- [27] J. M. Ollinger, "Model-based Scatter correction for fully 3-D PET", Phys. Med. Biol., vol. 41, pp. 153-176, 1996.
- [28] C. C. Watson, "New, Faster, Image-Based Scatter Correction for 3D PET", IEEE Trans. Nucl. Sci., vol. 47, pp. 1587-1594, 2000.
- [29] R. Manavaki, A. J. Reader, C. Keller, J. Missimer and R. J. Walledge, "Scatter Modeling for 3-D PET List-Mode EM Reconstruction", *IEEE NSS & MIC* 2002 conf. rec., Norfolk, VA, Nov 2002.
- [30] B. W. Pointon and V. Sossi, "Towards Model-Based Randoms Correction for 3-D Dual Head Coincidence Imaging", *IEEE NSS & MIC 2003 conf. rec.*, Portland, OR, Oct 2003.
- [31] J. M. Ollinger, "Detector efficiency and Compton scatter in fully 3D PET", *IEEE Trans. Nucl. Sci.*, vol. 42, pp. 1168-1173, 1995.

- [32] G. Germano and E. J. Hoffman, "A study of data loss and mispositioning due to pileup in 2-D detectors in PET", *IEEE Trans. Nucl. Sci.*, vol. 37, pp. 671-675, 1990.
- [33] R. D. Badawi and P. K. Marsden, "Developments in component-based normalization for 3D PET", Phys. Med. Biol., vol. 44, pp. 571-594, 1999.
- [34] R. D. Badawi and P. K. Marsden, "Self-normalization of emission data in 3D PET", *IEEE Trans. Nucl. Sci.*, vol. 46, pp. 709-712, 1999.
- [35] K. Wienhard, M. Shmand, M. E. Casey, K. Baker, J. Bao, L. Eriksson, W. F. Jones, C. Knoess, M. Lenox, M. Lercher, P. Luk, C. Michel, J. H. Reed, N. Richerzhagen, J. Treffert, S. Vollmar, J. W. Young, W. D. Heiss, and R. Nutt, "The ECAT HRRT: Performance and First Clinical Application of the New High Resolution Research Tomograph", *IEEE Trans. Nucl. Sci.* vol. 49, pp. 104-110, 2002.
- [36] J. S. Karp, G. Muehllehner, and R. M. Lewitt, "Constrained Fourier space method for compensation of missing data in emission computed tomography", *IEEE Trans. Nucl. Sci.*, vol. 7, pp. 21-25, 1988.
- [37] M. Schmand, L. Erikson, M. E. Casey, M. S. Andreaco, C. Melcher, K. Wienhard, G. Flügge, W. D. Heiss, and R. Nutt, "Performance results of a new DOI detector block for high resolution PET-LSO research tomograph HRRT", *IEEE Trans. Nucl. Sci.*, vol. 45, pp. 3000-3006, 1998.
- [38] M. Schmand, L. Eriksson, M. E. Casey, K. Wienhard, G. Flügge, and R. Nutt, "Advantages using pulse shape discrimination to assign the depth of interaction information (DOI) from a multi layer phoswich detector", *IEEE Trans. Nucl. Sci.*, vol. 46, pp. 985-990, 1999.
- [39] C. Knoess, S. Siegel, A. Smith, D. Newport, N. Richerzhagen, A. Winkeler, A. Jacobs, R. N. Goble, R. Graf, K. Wienhard, and W. D. Heiss, "Performance evaluation of the microPET R4 PET scanner for rodents", *Eur. J. Nucl. Med. Mol. Imaging*, vol. 30, pp. 737-47, 2003.
- [40] A. Rahmim, M. Lenox, A. J. Reader, C. Michel. Z. Burbar, T. J. Ruth, and V. Sossi, "Statistical list-mode image reconstruction for the high resolution research tomograph", *Phys. Med. Biol.*, vol. 49, pp. 4239-4258, 2004.
- [41] A. Rahmim, P. Bloomfield, S. Houle, M. Lenox, C. Michel, K. R. Buckley, T. J. Ruth, and V. Sossi, "Motion compensation in histogram-mode and listmode EM reconstructions: beyond the event-driven approach", *IEEE Trans. Nucl. Sci.*, vol. 51, pp. 2588-2596, 2004.

- [42] A. Rahmim, M. Lenox, A. J. Reader, C. Michel, Z. Burbar, T. J. Ruth, and V. Sossi, "Weighted Iterative List-Mode Reconstruction for the High Resolution Research Tomograph", Proc. of the VIIth Intern. Conf. on Fully 3D Reconst. in Radiol. and Nucl. Med. (Mo PM3-2), Saint-Malo, France, 2003.
- [43] A. Rahmim, M. Lenox, C. Michel, A. J. Reader, and V. Sossi, "Space-Variant and Anisotropic Resolution Modeling in List-Mode EM Reconstruction", *IEEE NSS & MIC 2003 Conf. Record*, Portland, OR, Oct 2003.
- [44] A. Rahmim, P. Bloomfield, S. Houle, M. Lenox, C. Michel and V. Sossi, "Practically Feasible Histogram-Mode and List-Mode EM Reconstructions with Full Motion Compensation", *IEEE NSS & MIC 2003 conf. rec.*, Portland, Oregon, Oct 2003.
- [45] A. Rahmim, T. J. Ruth, and V. Sossi, "Study of a Convergent Subsetized List-mode EM Reconstruction Algorithm", *IEEE NSS & MIC 2004 conf. rec.*, Rome, Oct 2004.
- [46] A. Rahmim, S. Blinder, J. C. Cheng, and V. Sossi, "Statistical List-mode Reconstruction in Quantitative Dynamic Imaging using the High Resolution Research Tomograph", Accepted to the Fully 3D Reconstr. Meeting in Radiol. and Nucl. Med., Salt Lake City, Utah, 2005.
- [47] M. Defrise, R. Clark, and D. W. Townsend, "Image reconstruction from truncated, two-dimensional, parallel projections", *Inverse Problems*, vol. 11, pp. 287-313, 1995.
- [48] A. J. Reader, "Image Reconstruction and Correction Techniques for Positron Volume Imaging with Rotating Planar Detectors", *Ph.D Thesis*, University of London, 1999.
- [49] S. S. Orlov, "Theory of three-dimensional reconstruction. 1. Conditions of a complete set of projections", Sov. Phys. Crystallography, vol 20, pp. 312, 1976.
- [50] D. W. Townsend, "Positron emission tomography with the high density avalanche chamber camera", *Ph.D Thesis*, University of Geneva, 1987.
- [51] J. G. Colsher, "Fully three-dimensional PET", Phys. Med. Biol., vol. 25, pp. 103-115, 1980.
- [52] M. Defrise, D. W. Townsend, and R. Clack, "Three-dimensional image reconstruction from complete projections", *Phys. Med. Biol.*, vol. 34, pp. 573-587, 1989.

- [53] E. O. Brightam, "The Fast Fourier Transform", Prentice-Hall, Inc., Englewood Cliffs, NJ, 1974.
- [54] E. J. Soares and S. J. Glick, "Analysis of Image Noise in Natural Pixel PET Reconstruction", IEEE NSS & MIC 2004 conf. rec., Rome, Oct 2004.
- [55] M. E. Daube-Witherspoon and G. Muehllehner, "Treatment of axial data in three-dimensional PET", J. Nucl. Med., vol. 28, pp. 1717-1724, 1987.
- [56] K. Erlandsson, P. D. Esser, S. E. Strand, and R. L. van Heertum, "3-D reconstruction for a multi-ring PET scanner by single-slice rebining and axial deconvolution", *Phys. Med. Biol.*, vol. 39, pp. 619-629, 1994.
- [57] R. M. Lewitt, G. Muehllehner, and J. S. Karp, "Three-dimensional reconstruction for PET by multi-slice rebinning and axial image filtering", *Phys. Med. Biol.*, vol. 39, pp. 321-340, 1994.
- [58] M. Defrise, P. E. Kinahan, D. W. Townsend, C. Michel, M. Sibomana, and D. Newport, "Exact and approximate rebinning algorithms for 3-D PET data", *IEEE Trans. Med. Imag.*, vol. 16, pp. 145-158, 1997.
- [59] X. Liu, M. Defrise, C. Michel, M. Sibomana, C. Comtat, P.E. Kinahan, and D. W. Townsend, "Exact Rebinning Methods for Three-Dimensional PET", *IEEE Trans. Med. Imag.*, vol. 18, pp. 657-664, 1999.
- [60] S. Matej, J. S. Karp, R. M. Lewitt, and A. J. Becher, "Performance of the Fourier rebinning algorithm for PET with large acceptance angles", *Phys. Med. Biol.*, vol. 43, pp. 787-795, 1998.
- [61] M. Defrise and X. Liu, "A fast rebinning algorithm for 3D positron mission tomography using John's equation", *Inverse Problems*, vol. 15, pp. 1047-1065.
- [62] F. John, "The ultrahyperbolic differential equation with four independent variables", *Duke Math. J.*, vol. 4, pp. 300-322, 1938.
- [63] J. A. Fessler, "Penalized Weighted Least-Squures Image Reconstruction for Positron Emission Tomography", *IEEE Trans. Med. Imag.*, vol. 13, pp. 290-300, 1994.
- [64] L. A. Shepp and Y. Vardi, "Maximum Likelihood Reconstruction for Emission Tomography", *IEEE. Trans. Med. Imag.*, vol. 1, pp. 113-123, 1982.

- [65] R. M. Lewitt, "Multidimensional digital image representations using generalizing Kaiser-Besse window functions", J. Optical Soc. Am. A, vol. 7, pp. 1834-1846, 1990.
- [66] R. M. Lewitt, "Alternatives to voxels for image representation in iterative reconstruction algorithms", *Phys. Med. Biol.*, vol. 37, pp. 707-716, 1992.
- [67] E. Ü. Mumcuoğlu, R. M. Leahy, S. R. Cherry, and Z. Zhou, "Fast Gradient-Based Methods for Baysian Reconstruction of Transmission and Emission PET Images", *IEEE Trans. Med. Imag.*, vol. 13, pp. 687-701, 1994.
- [68] E. U. Mumcuoğlu, R. M. Leahy, S. R. Cherry and E. Hoffman, "Accurate geometric and physical response modelling for statistical image reconstruction in high resolution PET", *IEEE Nucl. Sci. Symp. Conf. Record.*, vol. 3, 1569-1573, 1996.
- [69] T. J. Hebert and R. Leahy, "Fast methods for including attenuation in the EM algorithm", *IEEE Trans. Nucl. Sci.*, vol. 37, pp. 754-758, 1990.
- [70] A. J. Reader, S. Ally, F. Bakatselos, R. Manavaki, R. Walledge, A.P. Jeavons, P.J.Julyan, S.Zhao, D. L. Hastings and J. Zweit, "One-Pass List-Mode EM Algorithm for High-Resolution 3-D PET Image Reconstruction Into Large Arrays", *IEEE Trans. Nucl. Sci.* vol. 49, pp. 693-699, 2002.
- [71] K. Lange and R. Carson, "EM reconstructio for emission and transmission tomgraphy", J. Comput. Assist. Tomogr., vol. 8, pp. 306-312, 1984.
- [72] L. Kaufman, "Implementing and Accelerating the EM Algorithm for Positron Emission Tomography", *IEEE Trans. Med. Imag.*, vol. 6, pp. 37-51, 1987.
- [73] D. G. Politte and D. L. Snyder, "Corrections for Accidental Coincidences and Attenuation in Maximum-Likelihood Image Reconstruction for Positron-Emission Tomography", *IEEE Trans. Med. Imag.*, vol. 10, no. 1, pp. 82-89, 1991.
- [74] O. Rokitta, M. Casey, K. Wienhard, and U. Pictrzyk, "Random correction for positron emission tomography using singles count rates", *IEEE Nucl. Sci.* Symp. Conf. Record, vol. 3, pp. 17/37 -17/40, 2000.
- [75] M. E. Casey and E. J. Hoffman, "Quantitation in positron emission tomography: 7. A technique to reduce noise in accidental concidence meansurements and coincidence efficinecy calibration", J. Comput. Assist. Tomogr., vol. 10, no. 5, pp. 845-850, 1986.

- [76] E. Ü. Mumcuoğlu, R. M. Leahy and S. R. Cherry, "Baysian reconstruction of PET images: methodology and performance analysis", *Phys. Med. Biol.*, vol. 41, pp. 1777-1807, 1996.
- [77] R. D. Badawi, M. P. Miller, D. L. Bailey, and P. K. Marsden, "Randoms variance-reduction in 3D-PET", *Phys. Med. Biol.* vol. 44, pp. 941-954, 1999.
- [78] R. Gordon, R. Bender, and G. Herman, "Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and x-ray photography", J. *Theoret. Biol.*, vol. 29, pp. 471-481, 1970.
- [79] G. T. Herman and L. B. Meyer, "Algebraic reconstruction techniques can be made computationally efficient", *IEEE Trans. Med. Imag.*, vol. 12, pp. 600-609, 1993.
- [80] Y. Censor, P. P. B. Eggermont, and D. Gordon, "Strong underrelaxation in Kaczmarz's method for inconsistent systems", *Numer. Math*, vol. 41, pp. 83-92, 1983.
- [81] A. R. De Pierro, "Multiplicative iterative methods in computed tomography", in *Mathematical Methods in Tomography*, G. T. Herman, A. Louis, and F. Natterer, Eds., Lecture Notes in Mathematics 1497. Berlin:Springer Verlag, pp. 167-186, 1991.
- [82] D. L. Snyder, C. W. Helstrom, and A. D. Lanterman, M. Faisal, and R. L. White, "Compensation for readout noise in CCD images:, J. Opt. Soc. Amer. A, vol. 12, pp. 272-283, 1995.
- [83] M. Yavuz and J. A. Fessler, "Statistical image reconstruction methods for randoms-precorrected PET scans", *Med. Im. Anal.*, vol. 2, pp. 369-378, 1998.
- [84] S. Ahn and J. A. Fessler, "Emission image reconstruction for randomsprecorrected PET allowing negative sinogram values", *IEEE Trans. Med. Imag.*, vol. 23, pp. 591-601, 2004.
- [85] M. Yavuz and J. A. Fessler, "Objective functions for tomographic reconstruction from randoms-precorrected PET scans," in *Proc. IEEE Nuclear Science Symposium and Medical Imaging Conference*, Anaheim, CA, 1996, vol. 2, pp. 1067-1071.
- [86] J. M. Ollinger and J. A. Fessler, "Positron emission tomography", IEEE Sig. Proc. Mag., vol. 14, pp. 43-55, 1997.

- [87] J. Browne and A. De Pierro, "A row-action alternative to the EM algorithm for maximizing likelihoods in emission tomography", *IEEE Trans. Med. Imag.*, vol. 15, no. 5, pp. 687-699, 1996
- [88] J. A. Fessler and H. Erdogan, "A paraboloidal surrogates algorithm for convergent penalized-likelihood emission image reconstruction", in *Proc. IEEE NuclNuclear Science Symp. Medical Imaging Conf.*, vol. 2, pp. 1132-1135, 1998.
- [89] H. Erdoğan and J. A. Fessler, "Monotonic algorithms for transmission tomography", *IEEE Trans. Med. Imag.*, vol. 18, pp. 801-814, 1999.
- [90] S. Ahn and J. Fessler, "Globally convergent ordered ubsets algorithms for emission tomography using relaxed oredred subsets algorithms", *IEEE Trans. Med. Imag.*, vol. 22, no. 5, pp. 613-626, 2003.
- [91] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm", J. Roy. Statist. Soc., ser. B, vol. 39, pp. 1-38, 1977.
- [92] L. Parra and H. H. Barrett, "List-mode Likelihood: EM algorithm and image quality estimation demostrated on 2-D PET", *IEEE Trans. Med. Imag.*, vol. 17, no. 2, pp. 228-235, 1998.
- [93] A. N. Iusem, "A short convergence proof of the EM algorithm for a specific Poisson model", *Revista Brasileira de Probabilidade e Estatistica*, vol. 6, pp. 57-67, 1992.
- [94] C. Byrne, "Iterative reconstruction algorithms based on cross-entropy minimization", in *Image Models (and their Speech Model Cousins)*. ser. The IMA Volumes in Mathematics and its Applications, S. E. Levinson and L. Shepp, Eds. New York: Springer-Verlag, vol. 80, pp. 1-11, 1995.
- [95] D. L. Snyder, M. I. Miller, L. J. Thomas, and D. G. Politte, "Noise and Edge Artifacts in Maximum-likelihood Reconstruction for Emission Tomography", *IEEE Trans. Med. Imag.*, vol. 6, pp. 228-238, 1987.
- [96] H. H. Barrett, D. W. Wilson, and B. M. W. Tsui, "Noise properties of the EM algorithm", Phys. Med. Biol., vol. 39, pp. 833-846, 1994.
- [97] J. A. Fessler and A. O. Hero, "Penalized Maximum-Likelihood Image Reconstruction Using Space-Alternating Generalized EM Algorithms", *IEEE Trans. Image Process.*, vol. 4, pp. 1417-1429, 2004.

- [98] J. Qi and R. H. Huesman, "Theoretical study of lesion detectability of MAP reconstruction using computer observers", *IEEE Trans. Med. Imag.*, vol. 20, pp. 815-822, 2001.
- [99] E. Veklerov and J. Llacer, "Stopping rule for the MLE algorithm based on statistical hypothesis testing", *IEEE Trans. Med. Imag.*, vol. 6, pp. 313-319, 1987.
- [100] J. Llacer and E. Veklerov, "Feasible images and practical stopping rules for iterative algorithms in emission tomography", *IEEE Trans. Med. Imag.*, vol. 8, pp. 186-193, 1989.
- [101] D. W. Wilson and B. M. W. Tsui, "Spatial Resolution Properties of FP and ML-EM Reconstruction Methods", *IEEE Nuclear Science Symposium and Medical Imaging Conference*, pp. 1189-1193, 1993.
- [102] D. L. Snyder, M. I. Miller, "The use of sieves to stabilize images produced with the EM algorithm for emission tomography", *IEEE Trans. Nucl. Sci.*, vol. 32, pp. 3864-3872, 1985.
- [103] B. W. Silverman, M. C. Jones, J. D. Wilson, and D. W. Nychka, "A smoothed EM approach to indirect estimation problems, with particular reference to stereology and emission tomography", J. R. Stat. Soc. B., vol. 52, pp. 271-324, 1990.
- [104] M. Jacobson, R. Levkovitz, A. Ben-Tal, K. Thielemans, T. Spinks, D. Belluzzo, E. Pagani, V. Bettinardi, M. Gilardi, A. Zverovich, and G. Mitra, "Enhanced 3D PET OSEM reconstruction using inter-update Metz filtering", *Phys. Med. Biol.*, vol. 45, pp. 2417-2439, 2000.
- [105] S. Mustafovic and K. Thielemans, "Object dependency of resolution in reconstruction algorithms with interiteration filtering applied to PET data", *IEEE Trans. Med. Imag.*, vol. 23, pp. 433-446, 2004.
- [106] S. Geman and D. McClure, "Baysian image analysis: An application to single photon emission tomography", in Proc. Statist. Comput. Sect., Amer. Statist. Accoc., Washington, DC, pp. 12-18, 1985.
- [107] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images", *IEEE Trans. Pattern Anal. Mach. Intel.*, vol. PAMI-6, pp. 721-741, 1984.

- [108] E. Levitan and G. T. Herman, "A maximum *a posteriori* probability expectation maximization algorithm for image reconstruction in emission tomography", *IEEE Trans. Med. Imag.*, vol. 6, pp. 185-192, 1987.
- [109] T. J. Hebert and R. Leahy, "A Generalized EM Algorithm for 3-D Baysian Reconstruction from Poisson Data Using Gibbs Priors", *IEEE Trans. Med. Imag.*, vol. 8, pp. 194-202, 1989.
- [110] C-T. Chen, V. E. Johnson, W. H. Wong, X. Hu, C. E. Metz, "Baysian image reconstruction in positron emission tomography", *IEEE Trans. Nucl. Sci.*, vol. 37, pp. 636-641, 1990.
- [111] V. E. Johnson, W. H. Wong, X. Hu, C-T. Chen, "Image Restoration Using Gibbs Priors: Bounday Modeling, Treatment of Blurring, and Selection of Hyperparameter", *IEEE Trans. Pattern Analy. Mach. Intel.*, vol. 13, pp. 413-425, 1991.
- [112] C. S. Butler and M. I. Miller, "Maximum A Posteriori Estimation for SPECT Using Regularization Techniques on Massively Parallel Computers", *IEEE Trans. Med. Imag.*, vol. 12, pp. 84-89, 1993.
- [113] J. A. Fessler and W. L. Rogers, "Spatial resolution properties of the penalized-likelihood image reconstruction methods: Space invariant tomographs", *IEEE Trans. Image Processing*, vol. 5, pp. 1346-1358, 1996.
- [114] H. M. Hudson and R. S. Larkin, "Accelerated image reconstruction using ordered subsets of projection data", *IEEE Trans. Med. Imag.*, vol. 13, no. 4, pp. 601-609, 1994.
- [115] C. Byrne, "Accelerating the EMML Algorithm and Related Iterative Algorithms by Rescaled Block-Iterative Methods", *IEEE Trans. Imag. Process.*, vol. 7, pp. 100-109, 1998.
- [116] C. Byrne, "Likelihood Maximization for List-Mode Emission Tomographic Image Reconstruction", *IEEE Trans. Med. Imag.*, vol. 20, pp. 1084-1092, 2001.
- [117] C. Byrne, "Convergent Block-Iterative Algorithms for Image Reconstruction from Inconsistent Data", *IEEE Trans. Imag. Process.*, vol. 6, pp. 1296-1304, 1997.
- [118] R. M. Lewitt and G. Muehllehner, "Accelerated iterative reconstruction for positron emission tomography based on the EM algorithm for maximum likelihodd estimation", *IEEE Trans. Med. Imag.*, vol. 5, pp. 16-22, 1986.

- [119] A. J. Reader, R. Manavaki, S. Zhao, P. J. Julyan, D. L. Hastings, and J. Zweit, "Accelerated List-Mode EM Algorithm", *IEEE Trans. Nucl. Sci.*, vol. 49, pp. 42-49, 2002.
- [120] H. Erdoğan and J. A. Fessler, "Ordered subsets algorithms for transmission tomography", *Phys. Med. Biol.*, vol. 44, pp. 2835-2851, 1999.
- [121] I. T. Hsiao, A. Rangarajan, and G. Gindi, "A provably convergent OS-EM like reconstruction algorithm for emission tomography", *Conf. Rec. SPIE Med. Imaging*, vol. 4684, pp. 10-19, 2002.
- [122] I. T. Hsiao, A. Rangarajan, and G. Gindi, "A new convergent MAP reconstruction algorithm for emission tomography using ordered subsets and separable surrogates", *Conf. Rec. IEEE Int. Symp. Biomed. Imaging*, pp. 409-412, 2002.
- [123] R. Nutt, "History of PET", www.cpspet.com/our_company/history_o_pet.shtml The reader is recommended to also consult: G. L. Brownell, "A History of Positron Imaging", www.mit.edu/~glb/alb.html, for a different perspective on the early history of PET.
- [124] A. J. Reader, K. Erlandsson, M. A. Flower and R. J. Ott, "Fast accurate iterative reconstruction for low-statistics positron volume imaging", *Phys. Med. Biol.*, vol. 43, pp. 835-846, 1998.
- [125] D. L. Snyder, L. J. Thomas, and M. M. Ter-Pogossian, "A mathematical model for positron-emission tomography systems having time-of-flight measurements:, *IEEE Trans. Nucl. Sci.*, vol. 28, pp. 3575-3583, 1981.
- [126] T. Tomitani, "Image reconstruction and noise evaluation in photon time-offlight assisted positron emission tomography", *IEEE Trans. Nucl. Sci.*, vol. 28, pp. 4582-4589, 1981.
- [127] J. A. Kimdon, J. Qi, and W. W. Moses, "Effect of Random and Scatter Fractions in Variance Reduction using Time-of-Flight Information", *IEEE NSS & MIC 2003 conf. rec.*, Portland, OR, Oct 2003.
- [128] W. W. Moses and S. E. Derenzo, "Prospects for Time-of-Flight PET using LSO Scintillators", *IEEE Trans. Nucl. Sci.*, vol. 46, pp. 474-478, 1999.
- [129] W. W. Moses, "Time of Flight in PET Revisited", IEEE Trans. Nucl. Sci., vol. 50, pp. 1325-1330, 2003.

- [130] D. Townsend, P. Frey, A. Jeavons, H. J. Reich, H. J. Tochon-Donguy, A. Donath, A. Christin, and G. Schaller, "High density avalanche chamber (HI-DAC) positron camera", J. Nucl. Med., vol. 28, pp. 1554-1562, 1987.
- [131] D. M. Duxbury, R. J. Ott, M. A. Flower, K. Erlandsson, A. J. Reader, J. E. Bateman, R. Stephenson, and E. J. Spill, "Preliminary results from the new large-area PETRRA positron camera:, *IEEE Trans. Nucl. Sci.*, vol. 46, pp. 1050-1054, 1999.
- [132] K. Erlandsson, A. J. Reader, M. A. Flower and R. J. Ott, "A New 3D Backprojection and Filtering Method for PET Using All Detected Events", *IEEE Trans. Nucl. Sci.*, vol. 45, pp. 1183-1188, 1998.
- [133] C. Michel, M. Sibomana, A. Boi, X. Bernard, M. Lonneux, M. Defrise, C. Comtat, P. E. Kinahan and D. W. Townsend, "Preserving Poisson characteristics of PET data with weighted OSEM reconstruction", *IEEE Nucl. Sci. Symp. Conf. Record*, vol. 2, pp. 1323-1329, 1998.
- [134] M. Takahashi and K. Ogawa, "Selection of projection set and the order of calculation in ordered subsets expectation maximization method", *IEEE Nucl. Sci. Symp.*, vol. 2, pp. 1408 -1412, 1997.
- [135] P. K. Khurd and G. R. Gindi, "A Globally Convergent Ordered-Subset Algorithm for List-Mode Reconstruction", *IEEE NSS & MIC 2003 conf. rec.*, Portland, OR, Oct 2003.
- [136] C. C. Watson, "The Ecat HRRT: An Example of NEMA Scatter Estimation Issues for LSO Based PET Systems", IEEE NSS & MIC 2003 conf. rec., Portland, OR, Oct 2003.
- [137] C. C. Watson, "Advances in Scatter Correction for 3D PET/CT", IEEE NSS & MIC 2004 conf. rec., Rome, Oct 2004.
- [138] P. J. Markiewicz, A. J. Reader, M. Tamal, P. J. Julyan, and D. L. Hastings, "Towards an Analytic Unified Scatter and Attenuation System Model for 3D Whole Body PET Imaging", *IEEE NSS & MIC 2004 conf. rec.*, Rome, Oct 2004.
- [139] M. Tamal, A. J. Reader, P. J. Markiewicz, D. L. Hastings, and P. J. Julyan, "Noise Properties of Four Strategies for Incorporation of Scatter and Attenuation Information in 3D Whole Body PET", *IEEE NSS & MIC 2004 conf. rec.*, Rome, Oct 2004.

- [140] M. Egger, "Fast volume reconstruction in positron emission tomography: implementation of four algorithms on a high performance scalable parallel platform", *Ph.D. thesis*, University of Lausanne, Switzerland, 1996.
- [141] J. W. Wallis and T. R. Miller, "An Optimal Rotator for Iterative Reconstruction", *IEEE Trans. Med. Imag.*, vol. 16, pp. 118-123, 1997.
- [142] R. L. Siddon, "Fast calculation of the exact radiological path for a threedimensional CT array", Med. Phys., vol, 12, pp. 252-255, 1985.
- [143] C. Michel, X. Liu, S. Sanabria, M. Lonneux, M. Sibomana, A. Bol, C. Comtat, P. E. Kinahan, D. W. Townsend, and M. Defrise, "Weighted schemes applied to 3D-OSEM reconstruction in PET", *IEEE Nucl. Sci. Symp. Conf. Record*, vol. 3, pp. 1152 -1157, 1999.
- [144] National Electrical Manufacturers Association, "NEMA NU-2 Standards Publication NU-2-2001: performance measurements of positron emission tomography", Rosslyn, VA: National Electrical Manufacturers Association, 2001.
- [145] Y. Picard and C. J. Thompson, "Motion correction of PET images using multiple acquisition frames", *IEEE Trans. Med. Imag.*, vol. 16, pp. 137-144, 1997.
- [146] R. R. Fulton, S. R. Meikle, S. Eberl, J. Pfeiffer, C. J. Constable, "Correction for head movement in positron emission tomography using an optical motion tracking system", *IEEE Trans. Nucl. Sci.*, vol. 49, pp. 116-123, 2002.
- [147] M. E. Daube-Witherspoon, Y. C. Yan, M. V. Green, R. E. Carson, K. M. Kempner, and P. Herscovitch, "Correction for motion distortion in PET by dynamic monitoring of patient position (Abstract)", J. Nucl. Med., vol. 31, pp. 816, 1990.
- [148] B. J. Lopresti, A. Russo, W. F. Jones, T. Fisher, D. G. Crouch, D. E. Altenburger, D. W. Townsend, "Implementation and Performance of an Optical Motion Tracking System for High Resolution Brain PET Imaging", *IEEE Trans. Nucl. Sci.*, vol. 46, pp. 2059-2067, 1999.
- [149] P. M. Bloomfield, T. J. Spinks, J. Reed, L. Schnorr, A. M. Westrip, L. Livieratos, R. Fulton, and T. Jones, "The design and implementation of a motion correction scheme for neurological PET", *Phys. Med. Biol.*, vol. 48, pp. 959-978, 2003.

- [150] J. Qi and R. H. Huesman, "Correction of Motion in PET using Event-Based Rebinning Method: Pitfall and Solution (Abstract)", J. Nucl. Med., vol. 43, pp. 146P, 2002.
- [151] K. Thielemans, S. Mustafovic, and L. Schnorr, "Image Reconstruction of Motion Corrected Sinograms", *IEEE NSS & MIC 2003 Conf. Record*, Portland, OR, Oct 2003.
- [152] W. F. Jones, "Real-time Event Stream Correction for Patient Motion in Clinical 3-D PET", *IEEE Nuc. Sci. Symp. Conf. Record*, vol. 4, pp. 2062-2064, 2001.
- [153] R. H. Huesman, G. J. Klein, W. W. Moses, J. Qi, B. W. Reutter and P. R. G. Virador, "List mode maximum likelihood reconstruction applied to positron emission mammography with irregular sampling", *IEEE Trans. Med. Imag.*, vol. 19, pp. 532-537, 2000.
- [154] J. Qi and R. H. Huesman, "List mode reconstruction for PET with motion compensation: a simulation study", Proc. IEEE Inter. Symp. Bio. Imag. 413-416, 2002.
- [155] M. Unser, "Splines: A Perfect Fit for Signal and Image Processing", IEEE Signal Proc. Mag., vol. 16, pp. 22-38, 1999.
- [156] Consult the following useful website: http://bigwww.epfl.ch/index.html
- [157] R. J. Walledge, A. J. Reader, E. O. Aboagye, T. J. Spinks, J. Missimer, M. Honer, and A. P. Jeavons, "4D PET with the quad-HIDAC: Development of dynamic list-mode image reconstruction", *IEEE NSS & MIC 2002 conf. rec.*, Norfolk, Virginia, 2002.
- [158] R. J. Walledge, R. Manavaki, A. J. Reader, "Inter-Frame Filtering for List-Mode EM Reconstruction in High Resolution 4D PET", *IEEE NSS & MIC* 2003 conf. rec., Portland, OR, Oct 2003.
- [159] T. E. Nichols, J. Qi, E. Asma, and R. M. Leahy, "Spatiotemporal Reconstruction of List-Mode PET Data", *IEEE Trans. Med. Imag.*, vol. 21, pp. 396-404, 2002.
- [160] J. A. Fessler, "ASPIRE 3.0 user's guide: A sparse iterative reconstruction library", *Tech. Rep. 293*, Comm. and Sign. Proc. Lab., Dept. of EECS, Univ. of Michigan, Ann Arbor, MI, 48109-2122, July 1995, Available from http://www.eecs.umich.edu/ fessler

- [161] G. Chu and K. C. Tam, "Three dimensional imaging in the positron camera using Fourier techniques", Phys. Med. Biol., vol. 22, pp. 245-265, 1977.
- [162] P. E. Kinahan and J. G. Rogers, "Analytic three-dimensional image reconstruction using all detected events", *IEEE Trans. Nucl. Sci.*, vol. 35, pp. 635-638, 1990.
- [163] B. J. Ra, C. B. Lim, Z. H. Cho, S. K. Hilal, and J. Correll, "A true 3D reconstruction algorithm for the spherical positron tomograph", *Phys. Med. Biol.*, vol. 27, pp. 37-50.
- [164] B. Karuta and R. Lecomte, "Effect of Detector Weighting Functions on the Point Spread Function of High-Resolution PET Tomographs: A Simulation Study", *IEEE Trans. Med. Imag.*, vol. 11, pp. 379-385, 1992.
- [165] J. W. Stayman and J. A. Fessler, "Regularization for uniform spatial resolution properties in penalized-likelihood image reconstruction", *IEEE Trans. Med. Imag.*, vol. 19, pp. 601-615, 2000.
- [166] J. W. Stayman and J. A. Fessler, "Compensation for Nonuniform Resolution using Penalized-Likelihood Reconstruction in Space-Variant Imaging Systems", *IEEE Trans. Med. Imag.*, vol. 23, pp. 269-284, 2004.
- [167] A. J. Reader, P. J. Julyan, H. Williams, D. L. Hastings, and J. Zweit, "EM algorithm system modeling by image-space techniques for PET reconstruction", *IEEE Trans. Nucl. Sci.*, vol. 50, pp. 1392-1397, 2003.
- [168] J. Qi and R. H. Huesman, "Scatter correction for positron emission mammography", Phys. Med. Biol., vol. 47, pp. 2759-2771, 2002.
- [169] S. A. Larsson and C. Jonsson, "2D sampling of 3D object properties-a novel principal approach for image analysis and quantitative assessments of SPECT (and PET) Nuclear Science", *IEEE Trans. Nucl. Sci.*, vol. 47, pp. 1130 -1135, 2000.
- [170] H. El-Ali, M. Ljungberg, S. E. Strand et al., "Calibration of a radioactive ink-based stack phantom and its applications in nuclear medicine", *Cancer Biother Radiopharm*, vol. 18, pp. 201-207, 2003.
- [171] K. J. Van Laere, J. Versijpt, M. Koole et al., "Experimental performance assessment of SPM for SPECT neuroactivation studies using a subresolution sandwich phantom design", *Neuroimage*, vol. 16, pp. 200-216, 2002.

- [172] S. A. Larsson, C. Jonsson, M. Pagani, et al., "A novel phantom design for emission tomography enabling scatter- and attenuation-"free" single-photon emission tomography imaging", *Eur. J. Nucl. Med.*, vol. 27, pp. 131-139, 2000.
- [173] V. Sossi, K. R. Buckley, P. Piccioni, A. Rahmim, M-L. Camborde, E. Strome, S. Lapi, and T. J. Ruth, "Printed sources for positron emission tomography (PET)", *Trans. Nucl. Sci.*, in Print, 2005.
- [174] A. R. De Pierro and M. E. B. Yamagishi, "Fast EM-like methods for maximum a poteriori estimates in emission tomography", *IEEE Trans. Med. Imag.*, vol. 20, no. 4, pp. 280-288, 2001

A. Analytic Image Reconstruction from Truncated Projections

In this appendix, we shall present a brief overview of the three exising analytic reconstruction techniques used on truncated data.

The Universal Aperture Approach

The universal aperture of a region of space B (for a particular scanner geometry) is the set of all projection directions along which projections of B are measured (i.e. not truncated):

$$\Omega_{\text{univ}}(B) = \{ \hat{\mathbf{u}} \in \Omega | \forall \mathbf{s} \in B, p(\hat{\mathbf{u}}, \mathbf{s}) \text{ is measured} \}$$
(A.1)

With the assumption that the universal aperture of the support B of the unknown function satisfies Orlov's condition (so that it is sufficient for inversion of the data), then one may consider set of projections $T_0 = T(\Omega_{\text{univ}}(B))$ and reconstruct an image estimate using one of the methods described in the previous sections (for non-truncated data). This simple solution however will only be acceptable for some applications where $\Omega_{\text{univ}}(B)$ is not much smaller than Ω (e.g. see [161]), as it does not attempt to utilize the data measured along directions outside the universal aperture. An extreme example of this approach is to merely use the data obtained in the 2D-PET modality.

The 3DRP Scheme: Forward Projection of the Unmeasured Line Integrals

This method, commonly referred to as the 3D reprojection (3DRP) algorithm, proposes to recover the truncated data by using a first estimate $n_0(\mathbf{x})$ of the image, which has been reconstructed from the data subset T_0 as discussed above, to calculate line integrals along the truncated directions which are not meaured. Subsequently, combining the measured data with the calculated projection line integrals, one obtains the non-truncated set $T(\Omega)$ from which the solution may be reconstructed. An example of this approach used with the cylindrical geometry is discussed by Kinahan and Rogers [162].

Use of Suitable Filters for Inversion of Truncated Data

We have discussed already that due to the inherent redundancy in the 3D X-ray transform, different filters may be utilized to reconstruct the image, as long as the normalization condition (2.12) is satisfied. Let us consider the filtered-backprojection algorithm (Sec. 2.2.6). With truncated data, the filtering step (2.16) can not be appropriately performed, since for a given projection direction $\hat{\mathbf{u}}$, one does not have access to $p(\hat{\mathbf{u}}, \mathbf{s})$ for all $\mathbf{s} \in \hat{\mathbf{u}}^{\perp}$. However, it is possible to design the convolution kernel such that it vanishes in specific regions of $\hat{\mathbf{u}}^{\perp}$, thereby allowing the calculation of the convolution (2.16) even when the projections are truncated. In other words, assume one is able to to design a convolution kernel g satisfying

$$\int_{\hat{\mathbf{u}}^{\perp}} g(\hat{\mathbf{u}}, \mathbf{s} - \mathbf{s}') p(\hat{\mathbf{u}}, \mathbf{s}') d\mathbf{s}' = \int_{\hat{\mathbf{u}}_{\mathbf{m}}^{\perp}} g(\hat{\mathbf{u}}, \mathbf{s} - \mathbf{s}') p(\hat{\mathbf{u}}, \mathbf{s}') d\mathbf{s}'$$
(A.2)

where $\hat{\mathbf{u}}_m^{\perp}$ is the subset of the projection plane for which measurements are made, then for any direction $\hat{\mathbf{u}}$, values of $p(\hat{\mathbf{u}}, \mathbf{s})$ for $\mathbf{s} \notin \hat{\mathbf{u}}_m^{\perp}$ are not required in the calculation (2.12). Ref. [47] should be consulted for a discussion of how such filters may be constructed. However, once the convolution kernel g is constructed, it will not necessarily satisfy the normalization condition (2.12): an exception to this is the *true three-dimensional reconstruction' (TTR)* filter proposed by Ra *et al.* [163]. Otherwise, one can use the additive normalization formulation (2.14) which we write here as

$$H(\hat{\mathbf{u}}, \mathbf{v}) = G(\hat{\mathbf{u}}, \mathbf{v}) + L(\hat{\mathbf{u}}, \mathbf{v})$$
(A.3)

where G is the Fourier transform of the convolution kernel g (as discussed above) and

$$L(\hat{\mathbf{u}}, \mathbf{v}) = \frac{w(\hat{\mathbf{u}}) \left(1 - \int_{\Omega} G(\hat{\mathbf{u}}', \mathbf{v}) \delta(\hat{\mathbf{u}}', \mathbf{v}) d\hat{\mathbf{u}}'\right)}{\int_{\Omega} G(\hat{\mathbf{u}}', \mathbf{v}) \delta(\hat{\mathbf{u}}', \mathbf{v}) d\hat{\mathbf{u}}'}$$
(A.4)

184

One may then define $w(\hat{\mathbf{u}})$ as the characteristic function of the universal aperture

$$w(\hat{\mathbf{u}}) = \begin{cases} 1 & \text{if} \quad \hat{\mathbf{u}} \in \Omega_{\text{univ}}(B) \\ 0 & \text{if} \quad \hat{\mathbf{u}} \notin \Omega_{\text{univ}}(B) \end{cases}$$
(A.5)

With this choise of w, the correction filter L will only be applied to completely measured projections, and G is already designed to work in the presence of truncated data. It must be noted [47] that G must still not be too different from an exact filter (i.e. the term $\int_{\Omega} G(\hat{\mathbf{u}}, \mathbf{v}) \delta(\hat{\mathbf{u}}.\mathbf{v}) d\hat{\mathbf{u}}$ must be close to unity). Otherwise, the contribution of the correction filter L, which only deals with a subset of the data $\Omega_{\text{univ}}(B)$, would be large, and signal-to-noise ratio might be noticeably degraded¹.

¹ The TTR filter already satisfies the normalization condition, and therefore L = 0 in this case. The TTR algorithm *does* make a mild assumption: that for any given point **x** in image-space, one can find a plane passing through **x** along which all line integrals are measured (an assumption that is satisfied for most existing scanners; notable exceptions to this are scanners with partial rings, as well as scanners with gaps in-between the detectors, such as our very own HRRT!).

B. Space-variant and Anisotropic Resolution Modeling

In this part, we have proposed and investigated a method of incorporating the effect of inter crystal penetration (resulting in the parallax effect as discussed in Sec. 1.6.7). We first presented this work in [43].

Introduction

One issue common to PET scanners is the *space-variance* of the point spread function (PSF): manifesting itself as resolution degradation as one moves away from the center of the field-of-view (FOV). This effect is mainly caused by the dependence of inter-crystal penetration on angle of incidence of radiation [164]. As radiation occurs in voxels increasingly distant from the center of the FOV, it is more likely for the radiation to reach crystal fronts at higher angles of incidence, and to subsequently penetrate and be recorded in nearby crystals, ultimately degrading image resolution for such voxels (commonly referred to as the *parallax* effect).

Measurement of depth-of-interaction (DOI) within the crystals is known to minimize this problem, but its implementation has not achieved complete spatial invariance for resolution (see for e.g. [35]). In order to reconstruct images with uniform resolution properties in such systems a number of approaches are possible. The regularization approach (see Sec. 2.5.2), initially introduced to suppress noise artifacts in statistical image reconstruction, can be used to yield images with uniform resolution properties: appropriately space-variant smoothing filters or penalty terms have previously been designed for use in inter-filtering [105], post-filtering or penalized-likelihood reconstruction for space-variant imaging systems [165, 166]. In expectation maximization (EM) algorithms, the system matrix itself may be utilized to incorporate space-variance of the system: an attempt to do so is elaborated in this part.

Furthermore, spatial resolution for various scanners is commonly reported in the literature to be *anisotropic* (i.e. the PSF at a given position in the FOV exhibits distinct values along the axial and the two transaxial directions). This should be critical in correct modeling of the parallax effect, which exhibits anisotropic resolution at positions away from the center of the FOV. Simultaneous modeling of these two factors can allow more accurate characterization of degradation of each of the axial and the two transaxial resolution widths as one moves away from the center of the FOV.

Modeling of Space-Variance and Anisotropicity

In what follows, we describe how modeling of image resolution may be incorporated into the list-mode EM reconstruction technique. Since modeling is included into the system matrix, which is similarly utilized in histogrammode EM reconstruction, extension of such space-variant and/or anisotropic modeling to histogram-mode EM algorithm is straightforward, as pursued in [167] with space-invariant and isotropic resolution kernels.

Denoting λ_j^m as the image intensity in voxel j (j=1...J) at the *m*th iteration, and p_{ij} as the probability of an emission from voxel j being detected along LOR i, the list-mode expectation maximization (LM-EM) reconstruction algorithm is given by [92]

$$\lambda_{j}^{m+1} = \frac{\lambda_{j}^{m}}{s_{j}} \sum_{k=1}^{N} p_{i_{k}j} \frac{1}{\sum_{a=1}^{J} p_{i_{k}a} \lambda_{a}^{m}}$$
(B.1)

where i_k refers to the LOR along which the kth list-mode event is detected,

N is the number of measured events, and the sensitivity correction factor $s_j = \sum_{i=1}^{I} p_{ij}$ is a summation over all possible measurable LORs (i=1...I)

We next note that the system matrix or operator $\hat{P}=(p_{ij})_{I\times J}$ may be decomposed as [67, 70]:

$$\hat{P} = \hat{W}\hat{G}\hat{B} \tag{B.2}$$

where, as elaboreated in Sec. (2.4.2), $\hat{B}=(b_{ij})_{J\times J}$, $\hat{G}=(g_{ij})_{I\times J}$ and $\hat{W}=(w_{ij})_{I\times I}$ model voxel-voxel, voxel-LOR and LOR-LOR interactions, respectively. These matrices shall be referred to as spatial blurring, projection and sensitivity operators.

Sensitivity variations due to attenuation and normalization can be taken into account using the diagonal elements of the sensitivity matrix \hat{W} . Ideally, inter-crystal scattering and the parallex effect should also be incorporated into \hat{W} , an example of which has been shown in [68]. However, including these effects can result in a high reduction in the sparseness of the matrix and can significantly increase the computational demand, especially for high resolution tomographs.

For a scanner capable of measuing DOI information and achieving nearly isotropic resolution, Qi and Huesman [168] have argued using symmetry arguments that crystal scattering can instead be performed using space-invariant spherically symmetric blurring functions in image space (i.e. in the spatial blurring matrix \hat{B})¹.

As explained above, we shall therefore be using a diagonal matrix $\hat{W} = (w_i)_{I \times I}$. Upon substituting Eq. (B.2) into Eq. (B.1), and upon cancellation of the diagonal elements w_i in the forward and back-projection steps, and using

¹ It has been suggested by Reader et al. [29] that image scatter may also be modeled by this blurring matrix. However, as also noted by Qi and Huesman [168], since soft issue has a long attenuation length, object scatters do not have the local property compared to crystal scatters, and therefore, modeling crystal scatters in the spatial blurring matrix \hat{B} seems more suitable than for object scatters. Moreover, image scatter is object-dependent, and therefore, the system matrix needs to be object-dependent to correctly account for this effect, which can only complicate the matter!

 $\vec{S} = [s_1...s_J]^T$ and $\vec{\Lambda}^m = [\lambda_1^m...\lambda_J^m]^T$ to denote *J*-dimensional vectors of image sensitivity and image intensity (at iteration *m*), Eq. (B.1) can be written in the compact form [70]:

$$\vec{\Lambda}^{m+1} = \frac{\vec{\Lambda}^m}{\vec{S}} \times \hat{B}^T \left[\sum_{k=1}^N \mathrm{BP}_{i_k} \left\{ \frac{1}{\mathrm{FP}_{i_k}(\hat{B}\vec{\Lambda}^m)} \right\} \right]$$
(B.3)

where vectorial multiplication and division operations are performed on an element-by-element basis, and FP_{i_k} and BP_{i_k} denote geometric operators which perform forward- and back-projection along LOR i_k along which the kth list-mode event is detected.

We find it intuitively helpful to think of \hat{B} as an operator that models the PSF response of the system by effectively blurring image intensities of all voxels into the neighboring voxels, which are then subsequently forward-projected. Upon application of the back-projection operation, the "de-blurring" operation \hat{B}^T is then invoked in order to position back-projected values from neighboring voxels back to the appropriate voxels.

In previous works, the spatial-blurring component \hat{B} of the system matrix has been represented as a set of *space-invariant* and *isotropic* convolution kernels ρ . In this work, we have explored additional incorporation of the parallex effect into \hat{B} , and have subsequently considered using *position-dependent* and *anisotropic* kernel dimensions. Defining $\vec{R} = (X, Y, Z)$ as the position vector, each space-variant kernel $\rho(\vec{R})$ has been modeled as an anisotropic Gaussian, with the *distinct* widths $(\sigma_x(\vec{R}), \sigma_y(\vec{R}), \sigma_z(\vec{R}))$ along the transaxial (X,Y) and axial (Z) directions dependent on *position-inside-FOV* of any voxel which is being blurred.

Various functional forms may be considered for the modeling of spatial distribution of the three kernel widths. We have considered exponential and inverse-Gaussian forms of degradation of image resolution, as given by

$$\sigma_i(\vec{R}) = \sigma_i(0)e^{\left(\frac{X}{L_{iX}}\right)}e^{\left(\frac{Y}{L_{iY}}\right)}e^{\left(\frac{Z}{L_{iZ}}\right)}$$
(B.4)

$\sigma_x(0)$	L_{xX}	L_{xY}	L_{xZ}	
$\sigma_y(0)$	L_{yX}	L_{yY}	L_{yZ}	
$\sigma_z(0)$	L_{zX}	L_{zY}	L_{zZ}	

Tab.	B.1:	Characterization	of S ₁	pace-variance	and	Anisot	trop	icit	y
									•

and

$$\sigma_i(\vec{R}) = \sigma_i(0)e^{\left(\frac{X^2}{2L_{iX}^2}\right)}e^{\left(\frac{Y^2}{2L_{iY}^2}\right)}e^{\left(\frac{Z^2}{2L_{iZ}^2}\right)}$$
(B.5)

where i=x, y or z.

We therefore note that twelve parameters would be needed for this general parametrization of the blurring operator $\rho(\vec{R})$, as shown in table B.1. The horizontal extension of the table models XYZ-variant degradation of image resolution, and the vertical takes its anisotropicity into consideration.

Measurement of Finite Resolution Kernels

In principle, object-independent components of blurring can be collectively modeled by scanning point sources positioned at various transaxial and axial positions in the FOV and reconstructing the scan data *without* any resolution modeling. The resulting spatial profiles can then be fit to yield the resolution kernel characteristic lengths. If it is desired to inherently include modeling of positron range, the point sources need to be inserted inside a medium of the same density as the regions in patients/animals in which emissions occur. Moreover, since magnitude of positron range depends on the positron energy, which varies widely among isotopes, such measurements will be isotope-specific.

It must be noted that for the common case of 90°-transaxial-rotation symmetry (encountered, for instance, in circular or octagonal designs), it



Fig. B.1: Resolution relations implied by the common case of 90°-transaxialrotation symmetry.

must be the case that

$$\sigma_x(0) = \sigma_y(0) \tag{B.6}$$

(i.e. (x/y)-resolution values at the center of the FOV are equal), and that the characteristic lengths satisfy

$$\begin{cases}
L_{xX} = L_{yY} \\
L_{yX} = L_{xY} \\
L_{zX} = L_{zY}
\end{cases}$$
(B.7)

by symmetry. These identities have been illustrated in Fig. (B.1).

It therefore follows that characteristic lengths obtained from the fitting of spatial distribution of resolution along one of the transaxial directions can be determined from the corresponding one in the other transaxial direction.

Methods

Tomograph: Data were taken on the high resolution research tomograph

(HRRT) [35]. This scanner has an octagonal design, with the detector heads consisting of a double 1 cm layer of LSO/LYSO crystal layers for a total of 119,808 detector crystals (crystal size $2.1 \times 2.1 \times 10 \text{ mm}^3$). The total number of possible LORs is 4.486×10^9 .

Phantom used and measurement performed: Four C-11 line sources, oriented axially, were simultaneously positioned in the horizontal (i.e. Y=0) direction of the transaxial plane by distances of X=0, 4, 8 and 12 cm from the center of the scanner. A scan duration of 7 minutes was considered resulting in 60.0M trues and 4.9M randoms. This study has been performed for the transaxial directions and is currently also being extended to the axial direction. Detector normalization correction factors were obtained from a 12 hour scan using a rotating rod source.

Data analysis: Details of implementation of the list-mode EM algorithm on the HRRT are explained in chapter (4). In this work, in order to determine variation of image resolution for the scanner, we have used the following technique: the simultaneously-scanned line-sources have been reconstructed by space-invariant and isotropic Gaussian kernels for a number of times with different σ values ranging from 0 to 5 mm. Reconstructions were also performed using space-variant and anisotropic Gaussian kernels modeled with exponential as well as inverse-Gaussian forms of degradation of image resolution. Subsequently, transaxial widths ($w_x(X)$ and $w_y(X)$) of the four lines sources in the source were measured and compared (further elaborated in next section) at X=0,4,8,12 cm.

Results

Space-invariant and isotropic resolution modeling: Fig. (B.2) illustrates measured transaxial widths $(w_x(X) \text{ and } w_y(X))$ of the line sources as reconstructed for a wide range of Gaussian kernel widths. The following observa-



Fig. B.2: Observed transaxial (x/y)-resolution values as functions of spaceinvariant, isotropic kernel σ values. Points at which the curves are each minimized are indicated using circles.

tions can be made:

(i) Improvement in reconstructed line widths upon performing space-invariant and isotropic modeling is significant compared to the case of performing no modeling whatsoever, i.e. $(\sigma_x = \sigma_y = 0)$.

(ii) For each imaged position, there exist unique intermediate modeling widths (referred to as $\tilde{\sigma x}$ and $\tilde{\sigma y}$) that minimize the reconstructed line widths. These points are marked using circles on each plot in the figure and are interpreted as modeling widths that best model PSF at the particular X-position.

(iii) $\widetilde{\sigma x}$ and $\widetilde{\sigma y}$ are not necessarily equal to one another for any imaged positions (anisotropicity).

(iv) Increasingly larger values of $\widetilde{\sigma x}$ and $\widetilde{\sigma y}$ are observed as one moves away from the center of the FOV (space-variance).

Space-variant and anisotropic resolution modeling: The spatial distribution of $\widetilde{\sigma x}$ and $\widetilde{\sigma y}$ (as obtained at X=0,4,8,12 cm) were then separately fit using exponential and inverse-Gaussian functions, from which values of L_{xX} and L_{yX} (and therefore L_{yY} and L_{xY} as given by symmetry Eqns. (B.7a) and (B.7b)) were determined. The resulting estimated characteristic lengths were then incorporated into the space-variant and anisotropic system matrix, and the data were subsequently reconstructed with the new system matrix.

For any reconstructed image, we have introduced an error measure:

$$\Delta = \frac{1}{N} \sum_{i} \sum_{X} \left[\frac{(w_i(X) - w_i^{min}(X))}{w_i^{min}(X)} \times 100\% \right]$$
(B.8)

where i=(x, y), X=(0, 4, 8, 12 cm) (and therefore N=8 summations are performed), $w_i(X)$ indicates the line widths obtained using the particular modeling under consideration and $w_i^{min}(X)$ indicates the minimum width obtained using the procedure described in *data analysis* and illustrated in Fig. (B.2). The above metric therefore measures the percentage difference between the reconstructed line widths and the *minimum* reconstructed line width, and averages it over the two transaxial directions and the four positions across the FOV. Value of Δ would therefore approach zero when a particular type of modeling achieves minimum observed widths *simultaneously* for *all* the lines across the FOV.

Fig. (B.3) plots $\Delta(\sigma)$ as a function of the various space-invariant, isotropic σ values. Note that the two straight lines in the figure are for space-variant kernels and are plotted for comparison (and are not related to space-invariant models). The error measure has a minimum value of $\Delta=0.914$ obtained at $\sigma=3.1$ mm. By contrast, Δ values of 0.4169 and 0.5348 are obtained for exponential and inverse-Gaussian modeling of degradation of resolution across the FOV, as also shown in the figure. It is clearly seen that the latter error values are less than any Δ value obtained using regular modeling of σ values.



Fig. B.3: Variation of Δ as a function of modeled space-invariant, isotropic σ values, as described in text. Values achieved by using exponential and inverse-Gaussian modeling of degradation of resolution are also shown.

Conclusion

In conclusion, we have shown that the introduction of a space variant and anisotropic point spread function (PSF), compared to a space-invariant and isotropic PSF, can improve the resolution across the FOV. However, we must recognize this approach as what it is: an *ad hoc* method intended to render the otherwise-highly-non-sparse system matrices feasible and practical. Furthermore, symmetric (though anisotropic) PSFs were employed in this work. However, blurring caused by the parallex effect is asymmetric and the degree of this asymetry is highly dependent on the individual LORs (since it increases with higher angles of incidence). For space-variant tomographs, therefore, the ideal approach consists of modeling inter-crystal scattering and the parallex effects into \hat{W} , the operator which models LOR interactions in data-space.

C. Printed Sources for PET

Introduction

In order to characterize the spatial resolution of a particular positron emission tomograph, one often images a point source and estimates the full width at half maximum (FWHM) of the point spread function (PSF). This allows a measurement of the response of the tomograph. In order to accurately assess the system PSF, the source dimensions must be much smaller compared to the PSF FWHM. The spatial resolution achievable with modern tomographs is of the order of few millimeters (\sim 1-3 mm). It is thus becoming increasingly difficult to manufacture practical point or line sources that are sufficiently small so as not to influence the determination of the PSF FWHM.

In this section, we describe a novel technique to print radioactive point sources on paper using high concentrations of F-18 and using a modified standard ink-jet printer (HP DeskJet). This technique requires minimal human intervention, thus allowing to safely deal with relatively high concentrations of radioactivity. Printed point sources have been previously developed and successfully used in SPECT applications where gamma emitters are used as radioisotopes [169–172]. Their feasibility for PET imaging, although anticipated in principle, has never been practically demonstrated [171]. PET imaging of point sources is complicated by the fact that the positron must annihilate to produce 511 keV gamma rays, as discussed in Sec. (1.1.2).

Therefore, due to the finite positron range, it is not clear whether ordinary paper would provide enough material to enable sufficient number of annihi-
lations in a reasonable length of time. Use of further attenuating material on top of the paper sources, while allowing improved rate of annihilations, might noticeably degrade the measured resolution.

Resolution in PET is typically measured using F-18 since this radioisotope emits the lowest energy positron amongst those commonly used in PET (see table 1.2), and therefore positron range of F-18 (approximately 1 mm in water or tissue) renders it most suitable for resolution measurements. This also implied that positrons generated by F-18 require (on the average) the least amount of attenuating material to annihilate. Consequently, the proposed approach uses an F-18 ink solution to print the point sources.

The results of testing the printed sources on the high resolution research tomograph (HRRT) are reported in what follows. Ref. [173] should be consulted for results of testing this technique performed by our group on the Siemens/CTI ECAT 953B as well as the Concorde microPET R4. The resolution of these tomographs spans quite a wide range (approximately 1.8 mm for the microPET, 2.8 mm for the HRRT and 5.5 mm for the ECAT 953B) and therefore has offered a large range of source testing conditions.

Manufacturing of the Printed Sources

An important consideration when manufacturing the radioactive sources was the requirement of no human intervention since relatively high amounts of radioactivity are required to produce a very high radioactivity concentration ink solution. This was achieved by removing the original inkjet cartridge from the Hewlett Packard Deskjet printer (model C2170A) and installing a modified cartridge. An HP 51626A cartridge was disassembled to drain the ink, to remove all other components, and to insert an adapter which was custom machined to fit an extant round port in the bottom of the cartridge. A swagelok (Columbia Valve & Fitting) fitting coupled the adapter to a



Fig. C.1: Left: Side View, cut lines on printer ink cartridge. Right: Side View, section and assembly of modified ink cartridge.

22 gauge needle allowing the use of standard luer fittings for fluid delivery. These modifications minimized the volume of liquid which needed to be in the cartridge to allow printing.

A 3mL vial was mounted on the printer and connected through remote controlled solenoid valves to the printer cartridge and a supply of low pressure (1-2 psig) helium. F-18 was produced in a conventional niobium bodied water target by irradiating 18O-H2O. For these tests the vial was preloaded with 0.2mL of ink previously removed from the cartridge and the irradiated water was added directly to this vial with no pre-treatment. Specific activity was not measured. This solution was then transferred to the previously modified cartridge (Figure C.1) using a low pressure helium system. The pattern of radioactivity to be printed was created as a drawing in AutoCAD drafting software or Microsoft Word where the diameter of the dots or thickness and length of the line were specified. The initial page to be printed started with a large block of solid ink approximately 4 cm by 16 cm to ensure a uniform flow of ink before the printing of the sources of interest. The source distributions imaged in the tomographs and in the phosphor imager were all printed on multi-purpose 20lb paper (Econosource).

Effect of the Radioactive Ink and Paper Choice on Point Source Size

Two additional aspects of the printed source manufacturing were investigated: affect of the addition of radioactivity to the ink and impact of the choice of paper. To determine if the addition of ink would degrade the source size, ten 0.64 mm radioactive and ten non-radioactive sources were printed with the same printer, and subsequently the points were enlarged with a photocopier to determine their diameter. Furthermore, taking into account the outcome of this experiment, the influences of five kinds of paper on point size degradation were investigated using cold ink. The paper types studied were: HP Premium Photo paper, Econosource paper, Lexmark Premium InkJet paper and heavy bond paper.

Results: The mean value of the size of the sources produced with the radioactive ink was 0.67 ± 0.03 mm, while for those produced with the cold ink was 0.68 ± 0.04 mm, demonstrating that the procedure of introducing radioactivity into the ink did not affect printing quality. Both the HP Premium Photo paper and the Econosource photocopy paper gave spots that had diameters within 5% of the intended diameter as determined by the photocopy enlargement method. These two types of paper also gave the most homogeneously spherical points. The other papers tested gave spots that were on average 16% larger than the intended diameter (range 5-32% increase). The Econosource paper was therefore chosen for all the imaging experiments.

Point Source Manufacturing Reproducibility

Nine 2 mm diameter point sources were printed and their individual radioactivity was measured in a dose calibrator of known accuracy. The manufacturing reproducibility in terms of radioactivity was assessed by evaluating the radioactivity mean and standard deviation between the sources. The same procedure was performed with sixteen 0.5 mm point sources. Typical radioactivity levels for the 0.5 mm and 2 mm diameter point sources were approximately 1 mCi and 9 mCi, respectively.

0.5 mm sources: The radioactivity mean value and standard deviation between the 16 sources were 1.01 ± 0.08 mCi indicating a variability of approximately 8%.

2 mm sources: The radioactivity mean value and standard deviation between the nine sources were 9.27 \pm 0.3 mCi, indicating a variability of approximately 3%.

Imaging of Point Sources

Sixteen 0.5 mm diameter point sources were printed on a 4x4 grid with 3.5 cm separation between the points. This way the grid coordinates spanned approximately one quarter of the FOV: one edge of the grid was located at the centre of the FOV (X=Y=Z=0) while the other edge (diagonally away from the center of the FOV) was placed at X = Z = 10.5 cm, y=0 cm. The sources were scanned in air and sandwiched between two Al layers. Radial and axial profile FWHMs were obtained by averaging over the FWHM of the profiles of the sources located at different Z values but at the same X value. Data were reconstructed using one iterations and 16 subsets of the subsetized list-mode EM (S-LMEM) reconstruction algorithm (using the Siddon projection technique) as described in chapter (4).



Fig. C.2: Axial (squares) and radial (diamonds) resolution measured for the HRRT.

Data were acquired for 10 minutes. The sheet of paper with the printed sources was taped to a lateral styrofoam support (placed away from the printed sources). The FWHM values were estimated following the NEMA NU2-2001 protocol [144].

results: The FWHM of the axial and radial resolution as a function of radial position measured with paper alone on the HRRT is shown in Fig. (C.2). The resolution follows the expected pattern dictated by the parallax effect, as discussed in Appendix B, manifesting itself as a space-variant point spread function.

It must also be noted that the obtained FWHM values are dependent on the particular reconstruction algorithm used, parameters in the algorithm (e.g. number of iterations) as well as the particular interpolation techniques employed in projection operations in the reconstruction task as we have elaborately investigated in Sec. (4.4) and clearly shown, for instance in Fig. (4.11).



Fig. C.3: An example of a source profile imaged with (solid) and without (dashed) one layer of Al foil. The two profiles nearly coincide with one another.

In addition, an example of a reconstructed source profile with and without one layer of Al foil is shown in Fig. (C.3). The FWHM values were seen to be consistent in both cases for the HRRT. This effect was more elaborately tested on the microPET scanner, as we report next.

Effect of Additional Attenuating Medium

The effect of the additional attenuating material on both the resolution evaluation and the number of counts originating from the source region was extensively tested on the microPET scanner as reported in [173]. This scanner was selected since it has the best resolution amongst those used and was thus deemed to provide the most sensitive test to possible changes in the resolution evaluation due to the presence of the attenuating medium. The effect of the attenuating medium on the source profiles was investigated by comparing the FWHM and FWTM of the PSF with and without the additional Al foils. The effect of additional attenuating medium on the number of acquired counts was investigated by comparing the number of counts in a region of interest placed around the source image obtained from the scans with paper alone after scanning with and without the Al foil.

Results: No overall consistent difference in the FWHM values were obtained with and without the Al layers. In particular the results obtained in air and with a single Al layer agree within approximately 0.1 mm which was found to be the measurement reproducibility. For the FWTM the values with and without one Al layer coincided almost completely.

The effect of the attenuating material on the number of detected events originating from the source position could not be estimated directly from number of total acquired counts; many annihilation in-fact occur in the scanner material such as the gantry when the source is placed in air.

The effect was thus estimated by evaluating the count density in a region of interest placed around the source in the various scanning situations. A 3.4, 5.3, 6.1 and 6.7 fold increase in the count density was observed when adding 1,2,3 and 4 Al layers on each side respectively.

Discussion

It has been shown that it is possible to manufacture printed positron emitting point sources by developing an automated method to print radioactive point sources that does not require human handling of high amounts of radioactivity. Ordinary paper was found to be of sufficient quality to provide highly reproducible and reliable printed sources as determined by an independent radioactivity measurement. We have also shown that ordinary paper thickness provides enough material for a significant fraction of the positrons to annihilate - however if counting statistics need to be maximized, a thin layer of attenuating material can be added without negatively affecting the resolution evaluation.

We have described a novel technique to manufacture printed positron-

emitting point sources which does not require human handling of radioactivity. Ordinary paper was found to be of sufficient quality to provide highly reproducible and reliable printed sources as determined by an independent radioactivity measurement. We have also shown that ordinary paper thickness provides enough material for a significant fraction of the positrons to annihilate. However, if higher counting statistics are required in a particular task, a thin layer of attenuating material (e.g. Al foil) can be added without adversely affecting the imaged resolution¹. Consequently, we have made use of this practical technique in the course of this work for characterization of resolution achieved via various reconstruction (chapter 4) and motioncompensation (chapter 5) techniques. We have also utilized this method for printing and study of extended sources (e.g. see Figs. 4.4 and 5.5) used to investigate noise and uniformity properties, as elaborated in the subsequent chapters.

¹ For instance, it was found for the microPET scanner, that even a single 0.016 mm thick Al layer placed on each side of the paper increases the number of counts originating from the source position by a factor of 3.4. In future studies, it might be possible and desirable to laminate the paper sheet within clear plastic, which would have the added advantage of allowing visual determination of the source location while providing rigidity and attenuation.

D. Derivation of the List-Mode EM Reconstruction Algorithm with Motion Compensation

In what follows, time-of-detection will be treated as a continuous random variable, and therefore probability *density* elements, denoted using lower case p, will be utilized in addition to regular probability elements, denoted here using upper case P. We define $p(\mathbf{A}|D_j, j)$ as the probability density that an event generated in voxel j leads to a measurement \mathbf{A} in the detector, *assuming* the event will be detected (denoted by D_j). The list-mode acquired measurement $\mathbf{A} \in \mathbb{R}^d$, where d denotes the dimensionality of the measurement space, can contain such information about the event as, for instance, the LOR i' along which it has been detected, time-of-detection t and energy.

Parra and Barrett [92] have previously derived the list-mode EM reconstruction algorithm from first principles, which (written in our own notation¹) is given by:

$$\lambda_j^{m+1} = \lambda^m \sum_{k=1}^N p(\mathbf{A}_k | D_j, j) \frac{1}{\sum_{b=1}^J p(\mathbf{A}_k | D_b, b) \overline{s}_b \lambda_b^m}$$
(D.1)

where \mathbf{A}_k refers to the coordinates of the kth list-mode event (k=1...N) and \overline{s}_j is defined to be the probability that an event generated from voxel j leads to a measurement anywhere (as we note later, this term must incorporate presence of motion, to correspond to its definition).

¹ One needs to carefully note that Ref. [92] *does* assume, in the definition of the probability density elements, that the emitted events will be detected (i.e. D_j) but does not explicitly write this out in the derivations.

We next invoke the relation:

$$p(\mathbf{A}|D_j, j) = \frac{p(\mathbf{A}, D_j|j)}{P(D_j|j)} = \frac{p(\mathbf{A}|j)}{P(D_j|j)}$$
(D.2)

where we have dropped D_j in the numerator of the last term because an event which leads to a measurement **A** is necessarily detected. Furthermore, we note that $p(D_j|j)$, the probability that an event generated from voxel jis detected (along any LOR and at any time), corresponds exactly to the overall sensitivity correction factor \overline{s}_j ; i.e.

$$P(D_j|j) \equiv \overline{s}_j \tag{D.3}$$

This term, which has no time dependence, must take into account presence of motion throughout the scan, as we show below.

Let us consider any measurement $\mathbf{A} = \{i', t\}$ to indicate the LOR i' and time t at which the measured event is detected. We then have

$$p(\mathbf{A}|j) \equiv p(i',t|j) = P(i'|t,j)p(t|j) \equiv P(i'|t,j)\frac{1}{T}$$
(D.4)

where under the assumption that image intensity λ_j at any voxel j does not change over duration T of the scan, one notes that p(t|j), the probability density that an emission at voxel j has occurred at time t, is a constant (i.e. $p(t|j)=\frac{1}{T}$). The term P(i'|t,j) indicates the probability that an event generated in voxel j at time t will be detected at LOR i'. This corresponds exactly to the concept of a time-dependent system matrix. It then follows that the probability of detecting (along any LOR and at any time) an event emitted at voxel j is given by

$$\overline{s}_{j} \equiv \int_{0}^{T} \sum_{i'} p(i',t|j) dt = \frac{1}{T} \int_{0}^{T} \sum_{l} P(i'|t,j) dt$$
(D.5)

In other words, the probability that a photon emitted at voxel j is detected should include a time-average of the time-dependent sensitivity correction factors s_j^t for the course of the scan, i.e.

$$\overline{s}_j = \frac{1}{T} \int_0^T s_j^t dt \quad \text{where} \quad s_j^t \equiv \sum_{i'} P(i'|t,j) \tag{D.6}$$

Now, similar to the approach of Sec. 5.2.1, the attenuation and normalizationweighted probability elements at time t are given by:

$$P(i'|t,j) = A_i N_{i'} \mathcal{G}_j(\mathbf{i}) \quad \text{where} \quad i = \mathcal{L}_t^{-1}(i') \tag{D.7}$$

Combining Eqs. (D.2), (D.3), (D.4) and (D.7), one has

$$p(\mathbf{A}|D_j, j) = \frac{A_i N_{i'} \mathcal{G}_j(\mathbf{i})}{T\overline{s}_j}$$
(D.8)

which upon being substituted into Eq. (D.1) results in

$$\lambda_j^{m+1} = \frac{\lambda_j^m}{\overline{s}_j} \sum_{k=1}^N \mathcal{G}_j(\mathbf{i}_k) \frac{1}{\sum_{b=1}^J \mathcal{G}_b(\mathbf{i}_k) \lambda_b^m}$$
(D.9)

where \overline{s}_j , given by Eq. (D.6), is the overall motion-compensated sensitivity correction factor.