

## ABSTRACT

MERSEREAU, BRYANT GARRISON. MRI-Guided and Compressed Sensing Reconstruction Methods for PET/MRI. (Under the direction of Dr. David S. Lalush).

Recent advances in hybrid positron emission tomography/magnetic resonance imaging (PET/MRI) have opened up a multitude of research areas in neurological, oncological, and cardiac imaging. Using a PET/MRI scanner, complementary functional information from PET and structural information from MRI can be acquired simultaneously. One technical advantage of this hybrid imaging modality is the opportunity to use MRI information to improve the quality of PET reconstruction.

There are two major challenges to MR-guided PET reconstruction. The first is the computational challenge of incorporating MR information into iterative PET reconstruction. Serial CPU-based implementations of the OSEM PET reconstruction algorithm can take up to 30 minutes to complete, which can be stressful on research and clinical workflows. To overcome this obstacle, the OSEM algorithm was parallelized for GPU processing using the CUDA programming language. The parallelized reconstruction program has been shown to reconstruct PET images 60 times faster than a serial implementation of the OSEM algorithm.

The second challenge is that acquired PET/MRI datasets can be problematic to build due to the high logistical and monetary cost associated with recruiting and scanning patients. To alleviate this problem a digital PET simulation platform was developed. This platform utilizes PET phantoms built from patient MRI and CT images to simulate realistic PET images. The platform has shown to produce results consistent with acquired PET data and provides high configurability and flexibility as a simulation tool.

While PET provides a unique diagnostic imaging utility, it suffers from a host of physical and reconstruction effects which degrade its image quality. An MRI-guided PET

reconstruction method was developed for making PET less susceptible to partial volume effects. The method incorporates an MRI prior comprised of multiple MRI images directly into the PET reconstruction algorithm and reconstructs PET images at MRI scale. Using the prior and MAP reconstruction criterion, noise was successfully reduced in a spatially-varying manner in the resulting PET images, and the anatomical guidance of the prior was demonstrated.

Compressed sensing is a relatively new concept in MRI imaging being used to accelerate acquisitions. For PET/MRI imaging this could mean better image simultaneity between the modalities or motion compensation for PET. A variety of sparse transforms were investigated as well as the spatiotemporal smoothing capabilities of compressed sensing.

© Copyright 2017 Bryant Garrison Mersereau

All Rights Reserved

MRI-Guided and Compressed Sensing Reconstruction Methods for PET/MRI

by  
Bryant Garrison Mersereau

A dissertation submitted to the Graduate Faculty of  
North Carolina State University  
in partial fulfillment of the  
requirements for the degree of  
Doctor of Philosophy

Biomedical Engineering

Raleigh, North Carolina

2017

APPROVED BY:

---

Dr. Hongyu An

---

Dr. Shawn Gomez

---

Dr. Shumin Wang

---

Dr. Terence Wong

---

Dr. David S. Lalush  
Chair of Advisory Committee

## **DEDICATION**

To my family, source of boundless love and support

## **BIOGRAPHY**

Bryant Mersereau received his Bachelor of Science in Engineering degree from Duke University in May 2007. In August 2010 he completed his Master of Science degree from Clemson University and in May 2011 completed his Master of Business Administration degree from The Citadel Military College. In August 2012 he joined the graduate program in the Joint Department of Biomedical Engineering at North Carolina State University and The University of North Carolina at Chapel Hill and began working toward his Ph.D. under the guidance of Prof. David S. Lalush.

## **ACKNOWLEDGMENTS**

First and foremost, I would like to thank my advisor Dr. David S. Lalush for his support, supervision, and inspiration over the course of my Ph.D. studies. I would also like to thank Dr. Hongyu An for her excellent tutelage and guidance these past four years.

Thank you to my laboratory colleagues Jason Brown, Cihat Eldeniz, Ph.D., and Meher Juttukonda, Ph.D. for their contributions to my research as well as my advisory committee members for contributing their valuable time and expertise. Thank you to the entire administrative staff at the NCSU/UNC Joint Department of Biomedical Engineering, especially Vilma Berg, for helping me navigate the intricacies of graduate education.

Thank you to my family and friends for all their support and encouragement. Last but far from least, I would not have made it to this milestone without my wife Alice. Thank you Alice for all you do and all you are.

## TABLE OF CONTENTS

LIST OF TABLES .....	viii
LIST OF FIGURES .....	ix
CHAPTER 1: INTRODUCTION .....	1
1.1 Preface .....	1
1.2 Scope of Research .....	1
CHAPTER 2: PRINCIPLES OF PET AND MR IMAGING .....	4
2.1 PET Physics .....	4
2.2 PET Image Formation .....	6
2.3 PET Image Correction .....	9
2.4 PET Image Quality .....	12
2.5 MRI Physics .....	15
2.6 MRI Data Acquisition and Image Formation .....	18
2.7 MRI Image Quality .....	19
2.8 Hybrid Imaging .....	20
CHAPTER 3: ACCELERATION OF PET IMAGE RECONSTRUCTION .....	23
3.1 Introduction .....	23
3.2 Overview of Methods .....	24
3.3 Parallelization of the Projection and Backprojection Models .....	25
3.4 GPU Hardware Acceleration .....	29
3.5 Memory Management .....	31
3.6 Results .....	33

3.7 Conclusions .....	34
CHAPTER 4: DIGITAL PET PHANTOM DEVELOPMENT .....	35
4.1 Introduction .....	35
4.2 Methods .....	36
4.3 Results and Analysis .....	41
CHAPTER 5: MRI-GUIDED PET RECONSTRUCTION .....	46
5.1 Partial Volume Effect .....	46
5.2 Overview of Existing Methods .....	50
5.3 Formation of the Anatomical Prior .....	55
5.4 Patient Data Acquisition .....	58
5.5 Simulation Data Acquisition .....	60
5.6 Results and Analysis .....	61
5.7 Conclusions .....	73
CHAPTER 6: COMPRESSED SENSING IN MRI .....	75
6.1 Overview .....	75
6.2 Methods .....	78
6.3 Results .....	83
6.4 Discussion and Conclusions .....	90
CHAPTER 7: SUMMARY AND CONCLUSION .....	93
7.1 Summary .....	93
7.2 Future Work .....	96
7.3 Conclusion .....	97

REFERENCES ..... 98

## LIST OF TABLES

Table 3.1	Siemens Biograph mMR PET reconstruction geometries .....	28
Table 6.1	ROI coefficient of variation for each reconstruction method and frame ....	90

## LIST OF FIGURES

Figure 2.1	Flowchart of generic iterative reconstruction algorithm .....	8
Figure 2.2	Illustration of PET attenuation .....	10
Figure 2.3	Diagram of direct (solid) and oblique (dashed) coincidence lines in PET hardware embedded in a hybrid PET/MRI system .....	13
Figure 2.4	Diagram of Hybrid PET/MRI imaging system .....	21
Figure 3.1	Illustration of program parallelization strategy .....	26
Figure 3.2	Illustration of efficient CPU and GPU ray tracing methods .....	30
Figure 3.3	Diagram of concurrent stream processing in CUDA .....	33
Figure 4.1	E-phantom creation flowchart .....	37
Figure 4.2	Inputs to the phantom creation process .....	37
Figure 4.3	Example outputs of the phantom creation process .....	38
Figure 4.4	PET simulation flowchart .....	40
Figure 4.5	Results of E-phantom PET reconstruction process validation .....	42
Figure 4.6	Scatter plot and regression of $\mu$ -map reconstruction error determined using acquired data with manufacturer reconstruction software and phantom data simulated with the E-phantom platform .....	43
Figure 4.7	Visual comparison between acquired and simulated PET reconstructions .	44
Figure 4.8	Anatomic and activity diversity in the E-phantom platform .....	45
Figure 5.1	Comparison of coefficient of variation vs. beta for $^{18}\text{F}$ -FBP reconstruction of acquired data .....	61
Figure 5.2	Comparison of coefficient of variation vs. beta for $^{18}\text{F}$ -FDG reconstruction of acquired data .....	62
Figure 5.3	$^{18}\text{F}$ -FBP PET reconstructions using real subject at various $\beta$ -values .....	63
Figure 5.4	$^{18}\text{F}$ -FDG PET reconstructions using real subject data at various $\beta$ -values .	63
Figure 5.5	Example reconstruction with overtuned hyperparameter $\beta$ (C) .....	64

Figure 5.6	Comparison of $^{18}\text{F}$ -FBP PET reconstruction regional bias .....	65
Figure 5.7	Comparison of $^{18}\text{F}$ -FDG PET reconstruction regional bias .....	66
Figure 5.8	Mean, bias, and standard deviation images for simulated noisy $^{18}\text{F}$ -FDG ..	68
Figure 5.9	Regional mean bias and standard deviation for simulated noisy $^{18}\text{F}$ -FDG .	70
Figure 5.10	Profile analysis of hot lesion in simulated noisy $^{18}\text{F}$ -FDG along with input images true activity (A), $\beta=6\text{E}-08$ MAP reconstruction (B), $\beta=0$ MLEM reconstruction with 3 mm FWHM post-reconstruction smoothing (C), T1-weighted MRI (D), and T2-FLAIR MRI (E) .....	71
Figure 5.11	Profile analysis of prior effects in simulated noisy $^{18}\text{F}$ -FDG along with input images true activity (A), $\beta=6\text{E}-08$ MAP reconstruction (B), $\beta=0$ MLEM reconstruction with 3 mm FWHM post-reconstruction smoothing (C), T1-weighted MRI (D), and T2-FLAIR MRI (E) .....	72
Figure 5.12	Comparison of lesion contrast .....	73
Figure 6.1	Flowchart of CS reconstruction pipeline .....	78
Figure 6.2	Sagittal full-angle reconstruction (A) and rebinned reconstruction at end expiration (B) .....	84
Figure 6.3	Sagittal full-angle reconstruction (A) and rebinned reconstruction at full inspiration (B) .....	84
Figure 6.4	Montage of reconstruction results from NUFFT reconstruction (NF), Karhunen-Loève iterative thresholding (KL), temporal CS (CSt), spatial and temporal CS (CSst), and post-NUFFT reconstruction Gaussian blurring (NFblur) .....	86
Figure 6.5	Full inspiration frame reconstructions (frame one) with ghosting artifact (arrow) .....	87
Figure 6.6	Artifact smoothing region of interest (blue circle) .....	89

## **CHAPTER 1: INTRODUCTION**

### **1.1 Preface**

All of the work presented herein was conducted at the Biomedical Research Imaging Center at the University of North Carolina at Chapel Hill. Clinical data was collected at Washington University in St. Louis and University of North Carolina at Chapel Hill affiliated hospitals using IRB-approved protocols. Research data was collected at the Biomedical Research Imaging Center (BRIC) using an IRB-approved protocol. All data was processed at the Biomedical Research Imaging Center at the University of North Carolina at Chapel Hill.

### **1.2 Scope of Research**

Simultaneous positron emission tomography/magnetic resonance imaging (PET/MRI) is a versatile medical imaging modality that provides anatomical information from MRI aligned to functional information from PET imaging. To date, the fusion of PET and MRI information from PET/MRI scanners has been largely superficial. The primary use of MRI in PET/MR imaging is as an indirect replacement for computed tomography (CT) in PET attenuation correction and as a direct replacement for CT as an anatomical reference in PET/MRI overlay (fusion) images. While MRI provides an adequate ionizing radiation-free substitute for attenuation correction applications and has clinical value as an anatomical overlay, there is significant untapped potential in a more holistic merging of PET/MR imaging's simultaneously acquired anatomical and functional information.

There are two common approaches to integrating MR information with PET data. The first approach uses MR anatomical information to determine regions of interest (ROIs) for standard uptake value (SUV) calculation, dynamic model fitting, or correction of partial volume effects via segmentation [1-3]. The second approach integrates MR data into the PET reconstruction process, directly influencing the resulting PET image to improve resolution and signal-to-noise ratio (SNR). This approach assumes a correspondence between functional PET data and structural MR data, either collectively and/or along tissue boundaries. Historically, partial volume correction (PVC) and anatomically-guided PET reconstruction have been separate-but-related research paths. As multi-modal imaging has become more widespread, research into PVC has collided with anatomically-guided PET. The modality of choice for both PVC and anatomically-guided PET is MRI, due to its high soft-tissue contrast and excellent spatial resolution [1].

MR-guided PET relies on spatially aligned MRI and PET datasets. Historically this has been achieved via multi-modal intra-patient registration, which introduces additional error to the reconstruction process. Simultaneous PET/MRI scanners provide not only spatially aligned datasets without the need for post-acquisition registration but spatio-temporally aligned datasets, which offer additional improvement opportunities for dynamic and motion-compensated PET. Compressed sensing (CS) is a relatively new reconstruction technique in MR imaging that corrects artifacts resulting from accelerated and dynamic MR acquisition sequences. The benefits of CS may extend to PET via MR-guided PET reconstruction.

The overarching hypothesis of the graduate research presented herein is that MR anatomical information incorporated directly into the PET reconstruction algorithm can result

in spatially varying, edge-preserving smoothing, partial volume correction, and improved SNR for PET images. Additionally, potential exists for CS to improve MR-guided PET reconstructions in novel ways.

## CHAPTER 2: PRINCIPLES OF PET AND MR IMAGING

### 2.1 PET Physics

Positron emission tomography (PET) is a nuclear medicine imaging modality that acquires data in projection space. This acquired data is then converted to transverse slice images of the subject volume via tomographic reconstruction. As a nuclear imaging modality, data are collected from the detection of discrete high-energy photons expelled by injected radionuclides as they experience radioactive decay. The decay mechanism of suitable PET imaging radionuclides is beta decay, which results in the emission of a positron during each decay event. This positron then annihilates upon contact with the surrounding tissue, producing a pair of anti-parallel annihilation gamma photons at 511 keV mass-energy. These anti-parallel gamma photons pass through the subject volume and are detected by the PET scanner's cylindrical detector rings. Each pair of detected photons forms a coincidence line. The angles at which coincidence lines can be detected is determined by hardware system geometry and form the system lines of response (LORs).

PET scanners use multiple rings of detector blocks to cover a typical axial field of view (FOV) of 25 cm. Each detector block consists of a divided scintillator crystal to convert gamma photons into light, photomultiplier tubes (PMTs) or avalanche photo diodes (APDs) to record light output, and sophisticated timing electronics to match coincidence events. Cerium-doped lutetium oxyorthosilicate (LSO) is the preferred choice of scintillator in PET systems due to its high attenuation coefficient and fast response time.

Unlike MR or CT imaging, PET is a functional imaging modality in that it measures the distribution of injected radionuclides in the body rather than subject anatomy directly. These radionuclides are bound with bioactive molecules to form radiotracers that target functional processes within the body. The most common radionuclides used in PET imaging are carbon-11 ( $^{11}\text{C}$ ), nitrogen-13 ( $^{13}\text{N}$ ), oxygen-15 ( $^{15}\text{O}$ ), and fluorine-18 ( $^{18}\text{F}$ ). Fluorine-18 is a popular radionuclide for clinical emission tomography because its relatively long half-life (109 minutes) makes it more forgiving to process in a clinical setting compared to other radionuclides [4].

$^{18}\text{F}$  is used to make the important radiotracer [18F]-fluorodeoxyglucose ( $^{18}\text{F}$ -FDG or more simply FDG). FDG is a glucose analog molecule with fluorine-18 radionuclide substituted in place of a hydroxyl group. After injection, FDG enters cells using the same mechanism as glucose, but unlike glucose, FDG is not expelled from cells. The concentration of FDG within cells therefore grows in proportion to cellular glucose metabolism, which gives an indication of cellular activity. FDG is especially important in cancer imaging, because malignant cells often consume glucose at a higher rate than normal cells, making them appear bright on PET images [5]. FDG can also be used to identify regions of reduced cellular activity, such as in brain lesions caused by stroke or trauma. Another recently approved radiotracer gaining popularity is [18F]-florbetapir (FBP). Florbetapir selectively binds to the amyloid- $\beta$  ( $\text{A}\beta$ ) peptides of amyloid plaques formed in the brains of persons suffering from Alzheimer's disease. Florbetapir imaging provides a quantitative metric to help physicians determine the risk for progressive cognitive decline [6].

## 2.2 PET Image Formation

In tomographic PET image reconstruction, projection data is converted into an axial cross-sectional image of the acquired radiotracer distribution. There are two paradigms by which to view the reconstruction process. In the first paradigm, the PET acquisition model is treated as a series of line integrals of the radiotracer distribution analogous to the Radon transform in two dimensions and the X-ray transform in three dimensions [7]. A closed-form analytical solution to this model can be found exactly using the filtered backprojection algorithm, which computes the inverse Radon transform of the projection data. The “filtered” component of the algorithm signifies the inclusion of a one dimensional ramp filter term in the equation that compensates for the high frequency attenuation inherent to standard backprojection. In the second paradigm, the PET acquisition model is treated as a general linear model, allowing for sophisticated statistical reconstruction techniques and probabilistic models to be introduced. This paradigm better approximates reality at the expense of significant increased computational cost. No closed-form solution exists for finding the inverse of the linear system matrix, which is not only too large to compute, but also poorly conditioned. Instead, a numerical solution can be found using an iterative method to progressively refine a starting estimate of the true radiotracer distribution.

Any iterative method of image reconstruction consists of two related components. The first component is the criterion. The criterion is the metric that determines which image estimate is the best representation of the true underlying image. The second component is the algorithm. The algorithm contains the operational steps to be performed to reach the solution prescribed by the criterion. Although the criterion most explicitly states what determines a

satisfactory solution, the algorithm determines the pathway taken through solution space to arrive at the desired solution. Importantly, this can also alter not only the time required to reach a solution, but also the final destination of the reconstruction process [8].

The most common PET image reconstruction algorithm is the ordered subset expectation maximization algorithm (OSEM) [9]. The OSEM algorithm is a modification of the maximum likelihood expectation maximization algorithm (MLEM). OSEM divides projection data into subsets, and iterates on each subset individually. Instead of updating the image estimate after each full iteration, the image estimate is compared against the criterion after each sub-iteration. This simple modification results in comparable solutions 8-10x faster than with MLEM [8]. This makes OSEM in effect an accelerated version of MLEM. If a single subset is used in OSEM, the algorithm performs standard MLEM image reconstruction.

The MLEM algorithm is a general solution to a variety of statistical problems, with a wide range of applications. It was first demonstrated as a solution to emission tomography in 1984, though its derivation was developed a decade earlier. Each iteration of the MLEM algorithm consists of four fundamental stages: project, compare, backproject, and update. In the project stage, the image space estimate is projected into projection space. In the compare stage, the projection estimate is compared to acquired projection data via division. In the backproject stage, the error estimate produced by the division of the projection estimate with acquired projection data is backprojected into image space. In the final update stage, the next iteration of the image estimate is produced by multiplying the initial image estimate by the backprojected error estimate.

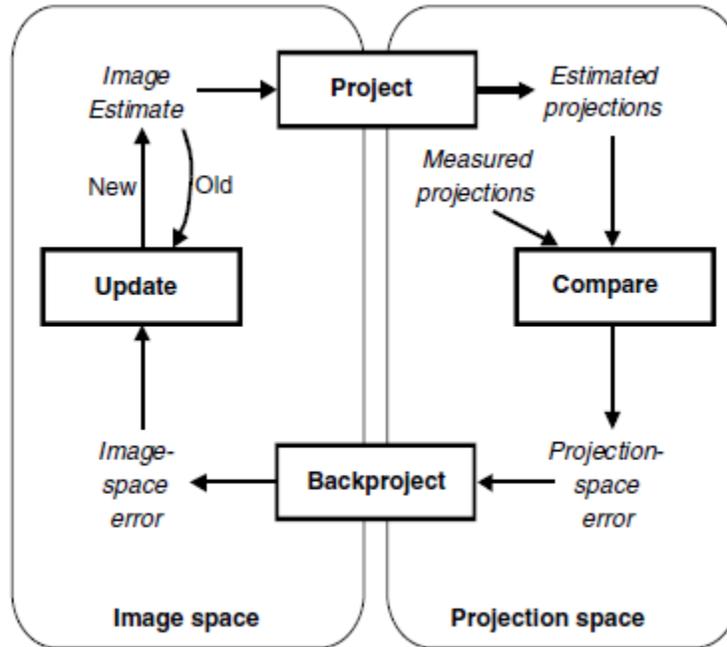


Figure 2.1: Flowchart of generic iterative reconstruction algorithm. © 2004 Elsevier Inc. Reprinted with permission from Lalush et al. [8]

MLEM is a robust algorithm with consistent and predictable convergence behavior. It also automatically imposes a non-negativity constraint, a useful property for emission tomography [8]. There are two major downsides to MLEM. The first downside, slow convergence, is remedied by the ordered subset modification used in OSEM. The second downside, noisy image results, is a result of the maximum likelihood (ML) criterion. The ML criterion, in words, is to select the solution which produces the greatest likelihood of the measured data given the image solution. Mathematically, choose the image  $f$  with greatest  $p(g|f)$ , where  $g$  is the measured data. ML estimators are unbiased and produce the minimum variance (noise) among unbiased estimators [8]. However, the amount of noise present in MLEM reconstructions of PET data taken to algorithm convergence is too high for practical use. One

solution to this noise problem is to trade variance for bias in the form of post-reconstruction smoothing. Another possibility comes from an interesting property of the MLEM/OSEM algorithm. Namely, lower frequencies tend to converge first, meaning another way to reduce noise is to halt the MLEM algorithm prior to convergence [10]. This has the added benefit of reducing reconstruction time. In practice, both premature stopping and post-reconstruction smoothing are typically applied to reconstructed PET images, the combination of which provides the greatest flexibility to tune the appearance of the reconstructed image to the desired result.

Lastly, the ML criterion requires a model for determining the probability of producing the acquired data given an image result. A common choice for this model is the Poisson distribution, which is discrete, relates the probability of a given number of events, and imposes a non-negativity constraint. Radioactive decay events have been shown to follow a Poisson distribution, and their uncorrected detection by PET system hardware also approximates a Poisson distribution if detector dead time is ignored. Other possible models include the Gaussian distribution and shifted Poisson distribution.

### **2.3 PET Image Correction**

The two anti-parallel gamma photons produced during positron annihilation each have a mass-energy of 511 keV. These photons are a form of electromagnetic radiation and undergo two main types of interactions with atomic electrons as they pass through tissue. These interactions are the Compton effect and the photoelectric effect. In Compton scattering, a photon interacts with an outer-shell (valence) electron, imparting a portion of its energy to the electron. This

interaction results in the ejection of the electron from its shell and a proportional change in the trajectory of the incident photon. In the photoelectric effect, a photon interacts with an inner-shell bound electron, depositing all of its energy to expel the electron, and disappearing in the process (Figure 2.2).

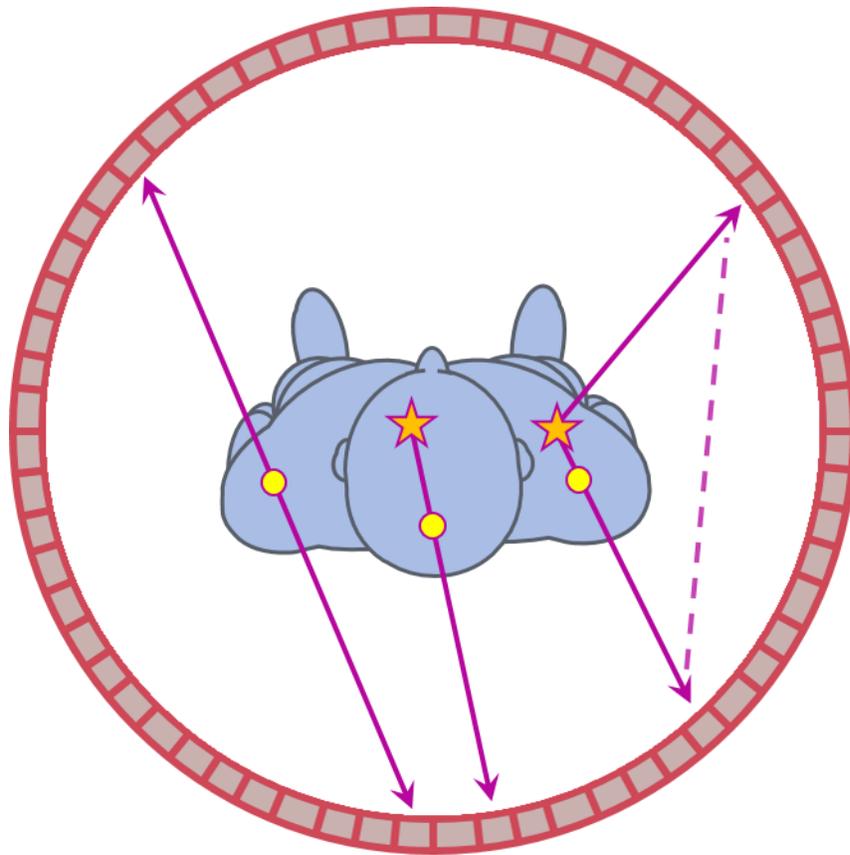


Figure 2.2: Illustration of PET attenuation. Unattenuated photons pass through the imaging volume and are detected (left). Photoelectric effect (middle). Compton effect results in scattered photon and incorrectly detected coincidence line (right).

The result of photon-electron interactions is attenuation of the measured PET count rate. Photons consumed by the photoelectric effect are always uncounted, while photons scattered by the Compton effect are still counted only if their new trajectory takes them into the path of PET detectors. Even if the scattered photons are detected, they will likely produce incorrect coincidence lines. Large angle scatter reduces image contrast by blanketing the image FOV with low level activity. Photoelectric photon absorption produces progressive radioactivity underestimation from FOV edge to center. Photon absorption, along with detector coincidence window size, and FOV activity all contribute to the production of random coincidences. Random coincidences are false coincidence lines produced by the matching of photon detections from two different positron annihilation events. Attenuation, scatter, and randoms are all corrected for during PET image reconstruction [11]. Attenuation correction is applied as a multiplicative correction factor of projection data while scatter and randoms correction are applied as additive correction factors.

Attenuation correction factors are computed prior to PET image reconstruction and are stored in the image domain as an attenuation map ( $\mu$ -map). The  $\mu$ -map is an image of linear attenuation coefficients (LAC) generated either directly from a scaled CT scan of the subject (or other rotating gamma source) or indirectly from an MRI of the subject. Before being applied, the subject  $\mu$ -map is projected into projection space. Scatter and randoms correction factors are created from probabilistic models of photon behavior in the imaging volume. These models incorporate the subject  $\mu$ -map, radiotracer characteristics, scan parameters, and hardware limits to estimate the number of scattered and random coincidences identified by each detector pair.

## 2.4 PET Image Quality

PET image quality is primarily determined by five factors. The first factor is total counts. As the total number of detected annihilation events increases, so too does the signal-to-noise ratio (SNR) of the resulting PET image. Higher total counts can be accomplished by increasing injected activity level, increasing scan duration, or by decreasing detector dead time. Unfortunately, there are significant downsides to each of these options. Increasing the amount of injected radiotracer results in greater exposure to ionizing radiation for the subject as well as increased random coincidence detection. Increasing scan duration will likely negatively impact clinical workflow, and detector dead time impacts random coincidence detection and is limited by hardware.

The second and third factors affecting PET image quality are system geometry and detector size. These factors are related and as such require mutual consideration. System geometry refers to the three dimensional layout of detectors surrounding the imaging volume. Detector size is a simplified way of addressing overall detector response, since PET detectors operate as unified blocks of scintillator crystals and light detectors and not as independent pixelated grids such as in semiconductor-based camera image sensors. The most basic goal of the PET system geometry is to meet the data sufficiency requirement of tomographic reconstruction. For two dimensional image reconstruction, this requirement is to acquire all projection angles between zero and 180 degrees. Beyond the basic data sufficiency requirement, PET system geometry determines the number of LORs available for the reconstruction process. Increasing the number of available LORs generally results in a better reconstruction image, as the imaging field of view is less likely to experience insufficiency

“streaking” artifacts resulting from the non-uniform LOR coverage of the imaging FOV. Modern PET acquisition systems can acquire both direct in-plane LORs as well as oblique LORs. Oblique projections contain LORs with detector end-points in offset axial planes. Incorporating oblique projections into the reconstruction model can greatly improve image quality, though care must be taken to reduce parallax error caused by the uncertainty in depth of interaction of incident gamma photons within a detector crystal. Parallax error is proportional the obliqueness of the LOR and its distance from center FOV.

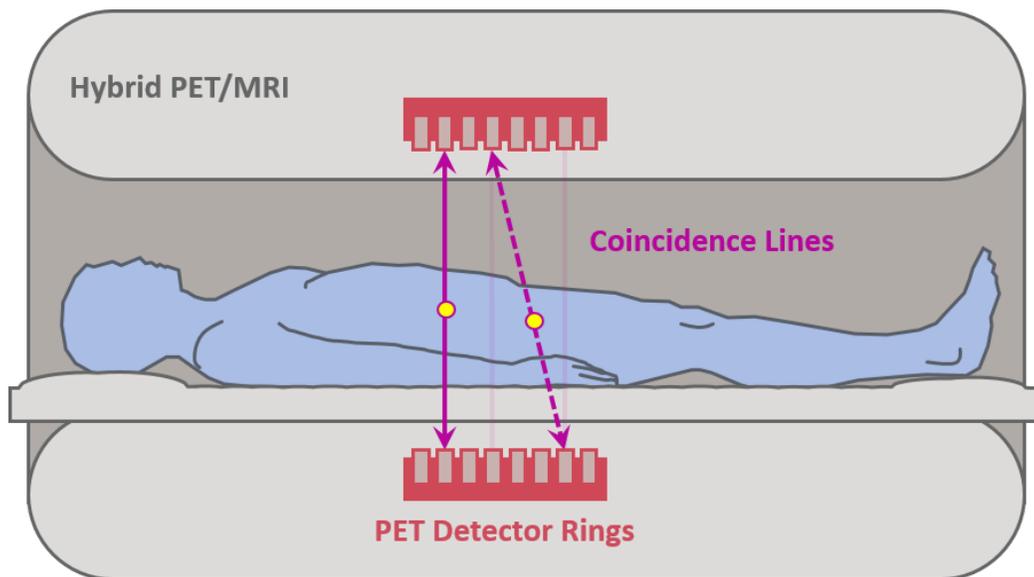


Figure 2.3: Diagram of direct (solid) and oblique (dashed) coincidence lines in PET hardware embedded in a hybrid PET/MRI system

The impact on image resolution by detector size is represented by the detector point-spread function (PSF), which is typically quantified as the full width at half-maximum (FWHM) of the photon spectrum obtained for a collimated point source placed a fixed distance from the

detector. Due to the complex nature of PET detectors, each detector has a unique response profile. This, combined with the non-uniform LOR coverage of the imaging FOV, is why PET imaging is modeled as a shift-variant system for purposes of numerical image reconstruction. For an  $^{18}\text{F}$  PET system with 60 cm bore diameter and 4 mm detector size, the FWHM spatial resolution of the detector response is roughly 2.5 mm [12].

The fourth factor affecting PET image quality is positron range, which is the distance traveled by a positron from its location of creation to its place of annihilation with an electron. Positron range is determined by the positron decay energy at creation and the attenuation characteristics of the surrounding tissue [12]. The positron range distribution for a given radionuclide can be modeled as a cusp-like function in one dimension.  $^{18}\text{F}$  has the lowest positron range among the common PET radionuclides, with a FWHM of 0.1 mm [13].

The fifth factor affecting PET image quality is photon non-collinearity. Detailed experiments have shown that a positron-electron annihilation event does not, in fact, always produce perfectly anti-parallel photons. The angular deviation of photon pairs resulting from a positron-electron annihilation event follows a Gaussian distribution with  $0.25^\circ$  FWHM [13]. This results in an image domain linear deviation of 1.32 mm FWHM for a PET system with 60 cm bore diameter [12].

The combination of these five factors, along with the effects of reconstruction result in a spatial resolution on the order of 4 mm for PET imaging systems. Finite spatial resolution results in two effects, blurred reconstruction images and an underestimation of activity in structures smaller than three times the size of the system spatial resolution (approximately 1.2 cm). This underestimation of small structures is known as the partial volume effect, and it

influences all discrete imaging systems, though its effect is particularly impactful in PET imaging due to its low spatial resolution. The most visible impact of blurring is the loss of edge delineation [11]. Measurements of PET image quality may be divided into quantitative technical measures (such as signal-to-noise ratio, contrast-to-noise ratio, noise power spectrum) and task-based measures (lesion detectability, region-of-interest quantification). Common parameters calculated for interpretation of PET images include standard uptake value (SUV), functional neuroactivity, and kinetic modeling of radiotracer activity to determine blood flow, glucose metabolism, and receptor concentration [5].

## **2.5 MRI Physics**

Magnetic resonance imaging (MRI) has been the fastest growing medical imaging modality in recent years due to its high image quality, noninvasive nature, and safety [11]. The name magnetic resonance imaging refers to the application of nuclear magnetic resonance (NMR) to radiological imaging [14]. The source of signal in MRI derives from the magnetic resonance properties of tissues, namely, macroscopic spin systems and their interaction with magnetic and electric fields. This interaction is described by the Bloch Equations, which relate the proton density, longitudinal relaxation, and transverse relaxation physical properties of spin systems to measurable MRI signal in the presence of a strong magnetic field [11].

The term “nuclear” in nuclear magnetic resonance refers only to the nucleus of an atom, and not to the presence of radioactivity, as in nuclear medicine. NMR-active nuclei are those that maintain angular momentum as a result of having an odd atomic number. These nuclei possess “spin,” and collections of identical nuclei (those of the same element) are referred to

as “nuclear spin systems.” The NMR-active nucleus of most importance to MRI is that of hydrogen ( $^1\text{H}$ ). Hydrogen atoms are abundant in the body (due to high water content), and hydrogen nuclei broadcast a strong NMR signal. On their own, nuclear spin systems do not produce a magnetic field, as the orientation of the individual nuclei spins cancel out macroscopically. However, in the presence of a magnetic field, nuclear spin systems align in the direction of the applied external field, a property termed “nuclear magnetism” [11].

There are two types of external magnetic fields used in MRI. The stronger field, usually denoted  $B_0$ , is a very strong, static magnetic field generated in the z-axis direction (head-to-foot) by the MRI scanner superconducting coil. This coil is supercooled by liquid helium to near-zero resistance, allowing it to maintain a constant circulating electric current. The static field strength of an MRI scanner is measured in units of tesla (T). Typical scanner static field strengths are 1.5T or 3T, though small animal scanners and dedicated research scanners can have static field strengths up to 12T. NMR signal strength is directly related to static field strength. Thus, scanners with high static field strength produce images with high SNR. However, as static magnetic field strength increases, so too does the difficulty of maintaining the consistency of the field across the imaging volume. Fluctuations in  $B_0$  lead to three sources of artifacts in MRI: magnetic field inhomogeneities, magnetic susceptibility, and chemical shift. Further discussion of these sources of error is beyond the scope of this manuscript, but suffice it to say that maintaining a high strength static magnetic field across an imaging volume is a challenging task. The second type of external magnetic field used in MRI is the gradient field. This magnetic field (G) is used to modify the magnetic field strength experienced by a position in the imaging volume. Unlike the static  $B_0$  field, the gradient field can be applied in

any coordinate direction and the pattern formed by the time-varying application of the gradient field forms the basis of three dimensional image formation in MRI [11].

Returning again to the source of MRI signal, when a nuclear spin system experiences an external magnetic field (the  $B_0$  field), the spin system aligns with the external field and forms a bulk magnetization vector ( $M$ ). Assuming a classical (non-quantum) view of magnetism, the orientation of this magnetization vector forms a precessional rotation about the direction of the  $B_0$  field (z-axis) due to intrinsic bulk angular momentum. The precession of the  $M$  vector about  $B_0$  is described by the Bloch equations and is analogous to the motion of the axis of a spinning top in Earth's gravity. The rate of precession is known as the Larmor frequency, and it is nuclei and external field strength dependent.

The rotating  $M$  vector can be further broken down into longitudinal and transverse components. The longitudinal component of the  $M$  vector is oriented along the axis of the  $B_0$  field ( $M_z$ ). The transverse component of the  $M$  vector is oriented orthogonal to the  $B_0$  field ( $M_{xy}$ ). The magnitude of the detectable NMR signal is dependent only on the transverse magnetization. In order to translate the existing magnetization vector into observable MRI signal, the precessing magnetization vector must be tipped into the transverse plane using a radiofrequency pulse (RF) tuned to the Larmor frequency of the magnetization vector. This RF excitation process, including the forced precession of the magnetization vector toward the transverse plane and subsequent induction of observable NMR signal in an external wire coil is described by Faraday's law of induction [11].

## 2.6 MRI Data Acquisition and Image Formation

Spatial positioning in MRI signal acquisition is encoded by the Larmor frequency and phase difference of the measured signal. The spatial coordinate system of an MRI scanner is arbitrary, though customarily the z-axis is oriented along the direction of the static  $B_0$  field, the y-axis is oriented up, and the x-axis is oriented from subject right-to-left. Frequency encoding of spatial position is used along the x- and z-axis directions to cause Larmor frequency to be spatially dependent. Phase encoding of spatial position is used along the y-axis direction. The specific timing of the application of RF pulse, gradient fields, and analog-to-digital conversion (ADC) is of paramount importance in MRI and determines not only spatial accuracy, but also the source of image contrast. The diagram of these timing instructions for MRI data acquisition is called a pulse sequence. Pulse sequences determine the length of MRI data acquisition as well as the source of contrast in the reconstructed image. As mentioned earlier, the amplitude of the NMR signal depends on proton (hydrogen nucleus) density ( $P_D$ ), longitudinal relaxation time ( $T_1$ ), and transverse relaxation time ( $T_2$ ). The weighted contribution of each of these factors in the MR image is determined by the timing of events defined in the pulse sequence.

The frequency and phase encoded spatial positioning data acquired by a pulse sequence can be interpreted as a scan of Fourier space. In the MRI community, this acquisition space is known as k-space. A standard MRI pulse sequence acquires data with a rectilinear, line-by-line sweep of Cartesian k-space. This k-space data is then decoded into an image space representation of the imaging volume by two or three dimensional Fourier transform depending on the specific imaging parameters. Two other common k-space trajectories are spherical and stack-of-stars. In a spherical k-space acquisition, k-space is sampled by three dimensional polar

lines originating from the k-space origin (isocenter). In a stack-of-stars acquisition, k-space is sampled by radial spokes in the  $k_{xy}$  plane repeated along the  $k_z$  axis to form a cylindrical stack of radially sampled lines.

## **2.7 MRI Image Quality**

One consequence of sampling in Fourier space is that aliasing can result in the image domain if care is not taken to maintain accordance with the Nyquist-Shannon sampling theory [15]. Antialiasing filtering can be applied in the frequency encoding direction to prevent signal contribution from outside the imaging FOV, but this is not possible for the phase encoding direction. The image FOV is inversely proportional to the phase encoding step size ( $\Delta k_y$ ), which means aliasing can be avoided by decreasing  $\Delta k_y$  to extend the FOV to fully cover the object being imaged. If the object to be imaged is still too large fit inside the FOV, saturation pulses can be used to dephase signal outside of the FOV [11]. The design of receiver coils also plays a role in preventing wraparound artifacts by limiting coil sensitivity to a small volume near the coil. This also aids in decreasing Johnson noise, the primary source of noise in MRI [11]. Fully reconstructed MR images are actually a combination of several images, each derived from a separate receiver coil tuned to a different part of the imaging FOV. In this manner, maximum SNR can be achieved without loss of imaging volume.

Typical MRI resolution is on the order of one mm, though this can vary considerably along each imaging axis, as they are independently programmable. Pulse sequence and total imaging time requirements may also limit apparent resolution as well as introduce ringing artifacts as a

tradeoff for reduced imaging time. Involuntary subject motion also limits imaging resolution through the production of ghosting artifacts.

## **2.8 Hybrid Imaging**

A hybrid imaging system combines separate imaging modalities into a single unified hardware system. A perfect hybrid imaging system would achieve synergistic interplay between the individual modalities as well as provide workflow and/or cost benefits to the purchaser. The classic example of hybrid imaging systems is PET/CT, which has become standard in clinical oncological applications. The PET component of PET/CT provides whole-body molecular imaging, while the CT component offers anatomical context for the PET image with minimal registration error due to the lack of need for subject repositioning between scans. Furthermore, the synergy between PET and CT extends beyond anatomical context as the CT also provides clinical-standard attenuation correction for the PET reconstruction process.

Despite the success of PET/CT hybrid imaging, the system still has a few significant downsides. First, since PET and CT both use ionizing radiation as their source of signal, the combination of the two increases the radiation exposure to the imaging subject. From a practical perspective though, this concern is mitigated by the relatively low exposure from diagnostic nuclear medicine in general, the use of low-dose CT for purposes of attenuation correction, and the regular pairing of PET and CT imaging even prior to the introduction of hybrid PET/CT scanners. Regardless, the stochastic effects of repeated exposure to ionizing radiation should not be wholly discounted. Second, the anatomical context provided by CT is sufficient for general localization in whole-body PET, but the soft tissue contrast of CT is

vastly inferior to that of MRI. CT has poor soft tissue contrast due to the similarity in linear attenuation coefficients of soft tissue types at CT gamma photon energy levels. Third, and most important, PET/CT is a mature technology likely near the apex of its possible functionality. Advancements have been made in each of the modalities that comprise PET/CT, but as a hybrid technology PET/CT has found little room for growth since its introduction in the early 2000s.

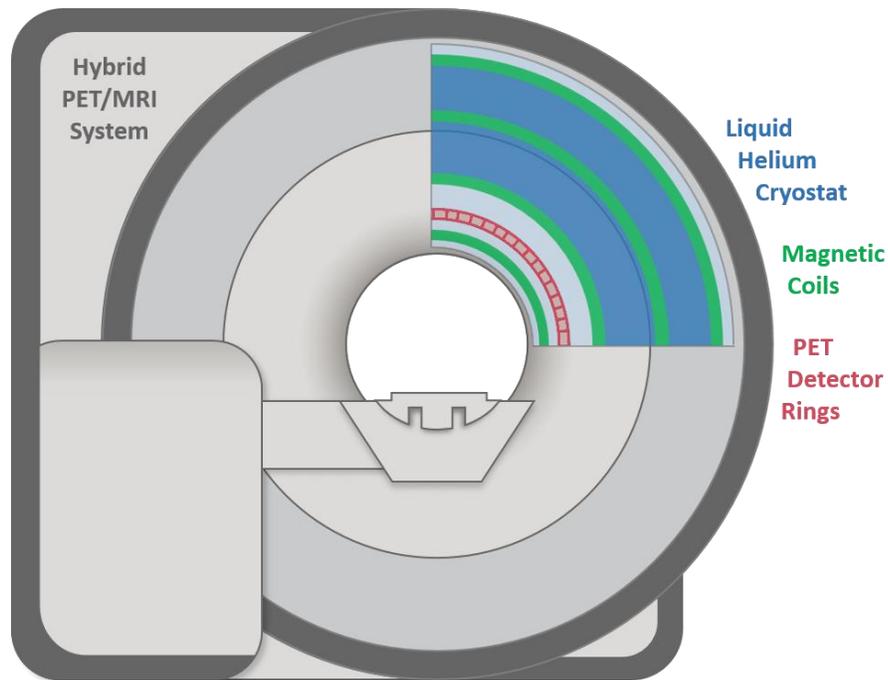


Figure 2.4: Diagram of Hybrid PET/MRI imaging system

Into this landscape has stepped the latest form of hybrid imaging, PET/MRI. Perceived initially as a “solution looking for a problem,” PET/MRI is now starting to find its foothold in quantitative molecular imaging [16, 17]. There are yet few clinical studies comparing the diagnostic/prognosis accuracy of PET/MRI compared to PET/CT, but initial findings suggest

no improvement over PET/CT [18]. This result is unsurprising, as the benefit derived from superior soft tissue localization provided by PET/MRI fusion is countered by less than ideal MR-based PET attenuation correction. PET/CT remains the consensus clinical workhorse for most oncological applications, but at least one long-standing PET/MRI site now prefers PET/MRI over PET/CT for cardiac, neurological, and pediatric imaging, all areas where MRI is considered the clinical-standard for anatomical characterization due to superior tissue contrast or radiation concerns [17]. The goal of PET/MRI imaging should therefore not be to displace PET/CT in “anatomically enhanced molecular imaging” (fusion overlay), but instead to find innovative ways to incorporate MRI information directly into PET images [17]. PET/CT is essentially a one-trick oncological imaging pony, while PET/MRI provides the necessary imaging flexibility and diversity to offer more than simply fusion overlay fusion images of anatomical and functional results. The vast range of imageable content by PET/MRI systems affords researchers the opportunity to use data mining and multi-parametric imaging techniques to find innovative solutions to the next generation of diagnostic challenges [17].

## CHAPTER 3: ACCELERATION OF PET IMAGE RECONSTRUCTION

### 3.1 Introduction

The OSEM algorithm reduces PET reconstruction time compared to MLEM, but the overall length of time required to reconstruct a PET image on a modern desktop computer is still at least 30 minutes. The proposed introduction of MRI guidance parameters into the PET reconstruction process would also increase the complexity of the reconstruction algorithm, necessitating additional computation time. This would push total reconstruction time beyond acceptability for both clinical and research application throughput. The solution to this problem was to accelerate the OSEM algorithm itself using parallel computing techniques. The goal was to accelerate only the implementation of the OSEM algorithm without appreciably affecting its reconstruction output. The following method was developed jointly by Bryant Mersereau and Dr. David S. Lalush, associate professor at North Carolina State University.

In a standard computer program, human readable code instructions are compiled into binary machine code and performed sequentially by the computer's central processing unit (CPU). In parallel computing, computational tasks are sorted into two groups, sequential and parallel. Sequential tasks are performed on the CPU, while parallel tasks can be performed on the graphical processing unit (GPU). The modern form of programmable GPUs were introduced in the late 1990s with the purpose of accelerating the graphical display of images. Unlike CPUs, which are designed to handle a wide range of computational tasks, GPU designs are tailored specifically for performing image processing and floating point math operations. Using GPUs to perform non-graphical related calculations is a relatively new area of interest,

termed general-purpose computing on graphical processing units (GPGPU). GPGPU has become an attractive tool for researchers in a variety of fields struggling to make use of the ever-increasing amount of data available in the digital world.

A successful GPGPU project relies on the principle of heterogeneous computing to increase program performance by assigning tasks based on their compatibility with processor architecture. Namely, parallel tasks should be assigned to the parallel GPU processor, while sequential tasks should be assigned to the serial CPU processor. Correctly assigning tasks to the optimal processor is of paramount importance when accelerating an algorithm. The functions and memory associated with the CPU are termed host components, and the functions and memory associated with GPU are termed device components. The program must be initially executed using the CPU, thus all memory initially resides on the host. Because there is substantial overhead involved in transferring memory to the GPU, only the most parallelizable program functions should be ported to the GPU. The remaining functions should be executed sequentially on the CPU. Assuming equivalent generation hardware, the CPU will usually perform large serial tasks faster than the GPU. This is due to the type of optimization inherent to CPU architecture for performing sequential computations not present on GPU devices (such as branch prediction), as well faster single core performance.

### **3.2 Overview of Methods**

A previously validated CPU-based OSEM image reconstruction method was adapted to function on commercially available NVIDIA-brand GPUs (Nvidia Corp., Santa Clara, CA). The algorithm was programmed using the C/C++ programming language and CUDA parallel

computing platform (Nvidia Corp., Santa Clara, CA). CUDA is an application program interface (API) containing a range of libraries that allow the programmer access to GPU hardware functionality. The reconstruction algorithm was programmed to function on all NVIDIA GPUs compatible with compute capability 2.1 or greater. Benchmarking was performed on a desktop computer system equipped with an Intel Xeon E5606 2.13 GHz CPU and a NVIDIA GTX 780 video card.

The parallel implementation of the OSEM algorithm relies on three strategies to achieve maximal acceleration: parallelization of projection and backprojection functions, hardware acceleration of FOV traversal, and intelligent memory management.

### **3.3 Parallelization of the Projection and Backprojection Models**

The processing structure for a parallel function is: 1) copy data from host to device memory, 2) load GPU program and execute, 3) copy results from device back to host memory. When source code is compiled, the CUDA compiler separates code into host and device components as designated by the `__global__` keyword. Host code is processed by the host compiler (gcc) and device code is processed by the CUDA compiler (nvcc). Parallel device functions are called from host code using triple chevrons. This call is known as a kernel launch. The kernel launch also contains the execution configuration, which designates the number of blocks and threads the GPU will use to parallelize the kernel. For example, a call to `mykernel<<<N,M>>>()` in the host code spawns N blocks times M threads (N x M) copies of the device function `mykernel()`, each of which will perform the same `mykernel()` operation. The set of parallel invocations is called a grid of thread blocks, and each invocation can be

assigned a unique index using the predefined `blockDim`, `blockIdx`, and `threadIdx` variables. This unique index can then be used in the device function to allow each invocation to access a different memory index of the array being iterated upon. Each invocation of the kernel is executed near simultaneously as allowed by GPU hardware. If the execution configuration designates more blocks than are available on the installed video card, extra blocks are executed as multiprocessors become available. Tailoring the execution configuration to the number of available processor cores can improve execution time.

Only the most computationally intensive functions need to be parallelized to realize significant performance gains (Figure 3.1). The optimal components of the OSEM algorithm to parallelize are the projection and backprojection routines. These are highly computationally intensive steps that together comprise roughly 95% of the computation time required to complete the OSEM algorithm. Thus, reducing the time spent performing these routines is key to reducing the overall time required to complete a reconstruction.

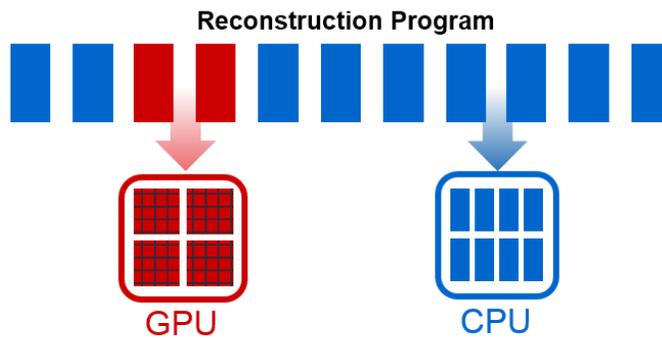


Figure 3.1: Illustration of program parallelization strategy

Fundamentally, performing a transformation from image space to projection space requires the repeated measurement of image content between a set of two points. A transformation from projection space to image space requires the distribution of projection content into image space between a set of two points. Importantly, the traversal of image space between two points does not influence the traversal of image space between a different set of points. Thus, each LOR projection or backprojection operation may be treated as an independent task. Thread independence is key to maintaining high computational throughput.

The PET imaging geometry of the Siemens Biograph mMR (Siemens AG, Munich, Germany) can be represented in PET reconstruction process in several ways. The imaging geometry determines the size of projection space, which in turn determines the complexity of the projection and backprojection processes. The standard reconstruction geometry is comprised of 72.6 million LORs. At a standard image space output size of  $344 \times 344 \times 127$  voxels, each LOR would iterate over an average of 375 voxels per traversal. In other words, each projection or backprojection operation requires iteration over approximately 27 billion voxels. A complete reconstruction with three iterations of projection and backprojection therefore requires iteration over approximately 162 billion voxels. A reconstruction at MRI-scale, which is required for true MRI-guided PET reconstruction in Chapter 5, requires iteration over approximately 252 billion voxels at an image space size of  $400 \times 400 \times 140$  voxels. The standard PET imaging geometry is actually a collapse for computational simplicity of the native imaging geometry. The collapse process combines either five or six (depending on detector position) neighboring oblique LORs into a single collapsed LOR by summing the projection intensities of each neighboring LOR. Native PET imaging geometry is comprised

of 354 million LORs, a roughly fivefold increase in computational burden. A breakdown of the three geometries employed in the Siemens Biograph mMR scanner is collected in Table 3.1.

Table 3.1: Siemens Biograph mMR PET reconstruction geometries

	Geometry		
	Native (G1)	Collapsed (G2)	Direct (G3)
Angles	252	252	252
Transverse Bins	344	344	344
Planes	4084	837	127
Maximum Ring Difference	60	120	0
Total Number of Detectors	32,256	64,008	64,008
Total Number of LORs	3.54E+08	7.26E+07	1.10E+07

The parallel projection and backprojection process follows the basic parallel function structure. A host array of LOR endpoint locations in image space is first transferred to the device, and memory space for an output array is allocated on the device. The projection/backprojection kernel is called, and each parallel thread is assigned a single LOR to trace. The result of each line trace operation is stored in the output array. The output array is then transferred back to host memory in preparation for the next algorithm iteration.

### 3.4 GPU Hardware Acceleration

Even though code written for CPU execution can be translated near identically to be performed on a GPU, solutions which take advantage of the accelerated image processing hardware in a GPU are often more efficient than code written for serial CPU execution. The projection of lines through a volume is exactly the type of operation a GPU is well suited to accelerate. From the perspective of each LOR, each projection operation is the summation of voxel intensity between two points weighted by ray length. For a serial CPU implementation, an efficient method of ray tracing identifies on a voxel-by-voxel basis each voxel plane intersected by the LOR, compares this plane to the previously intersected plane, then calculates the voxel's contribution to the projection as the product of the voxel intensity and the distance between the intersected planes. The problem with this implementation from the perspective of the GPU is the number of if-statements required to find the intersecting voxel planes. GPU threads are organized into warps of 32 threads. Each warp is computed simultaneously, except when if-statements diverge the computational path of individual threads. Thus, in general the more if-statements a device function contains, the less efficiently a GPU can process the function. CPUs handle if-statements more efficiently for a number of reasons, one important one being branch prediction circuitry.

In the parallel OSEM implementation, the imaging volume is stored on the GPU in a special structure called texture memory. Texture memory is read-only, cached memory optimized for spatial locality. Texture memory offers automatic interpolation and boundary handling, as well as dedicated hardware for processing interpolation. Texture memory is created by binding a region of device memory to a texture reference, and texture memory

locations are accessed through texture fetch commands. In order to reduce thread divergence, a fixed step size was set relative to normalized voxel size. Through experimentation it was found that a fractional step size of 0.25 voxels was sufficient for the projection and backprojection routines to adequately sample a LOR. This fractional step size corresponds to four samples per voxel traversal. By using a fixed step size in conjunction with texture interpolation, all if-statements corresponding to voxel plane detection could be removed. The two methods of ray tracing are illustrated in Figure 3.2. The CPU method (purple) and the GPU method (green) both arrive at similar projection estimates. The intensity difference between the two projection calculations further narrows as a larger voxel field is traversed. The impact of any remaining difference is nullified by the application of the same ray tracing method for both projection and backprojection.

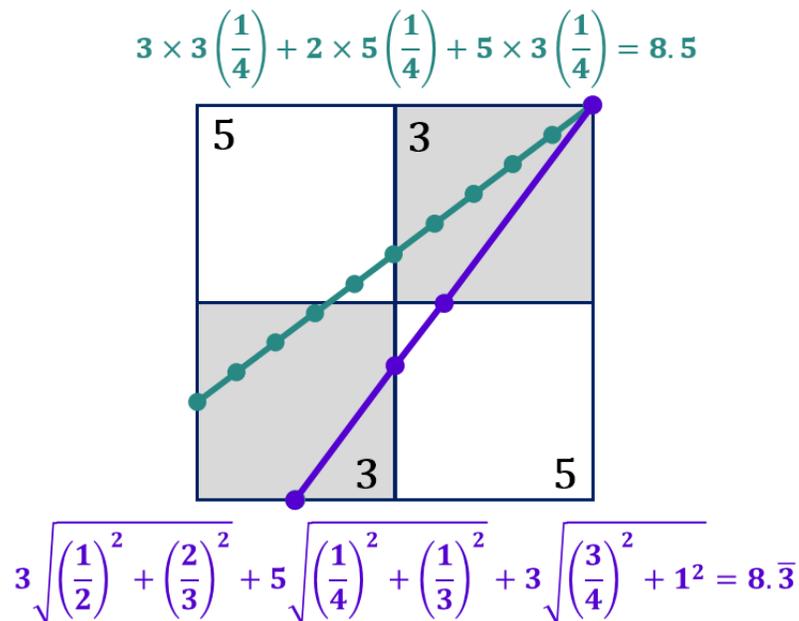


Figure 3.2: Illustration of efficient CPU (purple) and GPU (green) ray tracing methods

Since texture memory is a read-only memory structure, it was not available to use for image domain backprojection. Instead, a standard device memory array was used as the backprojection volume. In order to maintain scaling compatibility with the texture-based projection routine, the same step size was used for both routines. In backprojection, the LOR intensity (count) value is added to every visited location along the LOR. Nearest neighbor rounding was used to identify a single voxel index to receive the LOR intensity at each visited location. An important reason to require compute capability 2.1 is feature support for floating-point atomic addition. Atomic addition, accessed by the `atomicAdd()` function, prevents race conditions in the output volume caused by multiple threads attempting to change the value of an index location simultaneously. Without atomic functions, backprojection would require additional overhead to ensure intersecting LORs were never traversed concurrently. The simpler solution, `atomicAdd()`, does slow completion of backprojection compared to projection by serializing simultaneous memory accesses, but the time cost is worth the prevention of race conditions.

### **3.5 Memory Management**

There is no question parallel computing improves computational performance for a wide variety of tasks, but one downside of heterogeneous computing is the need to manage two banks of program memory, one accessible only by the host processor and one accessible only by the device processor. Device memory is allocated and assigned in much the same way as host memory, and CUDA provides analogous device memory functions to `malloc()` and `memcpy()` in the form of `cudaMalloc()` and `cudaMemcpy()` respectively. Because of the

immense size of projection space arrays and scanner geometry information, memory transfers to and from host and device across the PCI bus turns out to be the limiting acceleration factor for parallel OSEM reconstruction.

For standard PET reconstruction geometry, 1.7 GB is required to store LOR endpoints, 1.45 GB (5 x 290 MB) for projection space arrays used in the reconstruction process, as well as a variety of image space arrays and constant variables. The total memory requirements for standard geometry PET reconstruction total roughly 3.5 GB of dynamic program memory. For native geometry the memory requirements total roughly 17 GB, well above the amount of memory available to current generation video cards. Even the 3.5 GB of device memory needed for standard geometry reconstructions is only available to high-end video cards. Peak single direction bandwidth between host memory and device memory is 8 GB/s on PCIe x 16 Gen2, which in practice is never fully realized. Besides transfer rate limitations, the order of memory allocation can impact program function. Large arrays must be defined early in the code base to ensure contiguous memory blocks are available.

Some of the problems related to the large memory requirements of the PET reconstruction process may be mitigated by intelligent memory management. First, instead of trying to load the entire geometry structure into device memory each iteration, LORs are processed in batches corresponding to the amount of device memory available to the system. The number of LORs processed in each batch is calculated independently of OSEM algorithm subset size and is determined based on the hardware configuration of the installed video card. Second, large array memory transfers from host to device are made using pinned host memory. Host memory allocations are pageable by default. When data transfers from host to device memory are

requested, the pageable memory is temporarily transferred to a page-locked (pinned) host array before being transferred to device memory. By directly allocating host memory as pinned using the `cudaHostAlloc()` function, this temporary transfer step is avoided. Third, memory transfers are programmed to run concurrently with LOR processing using stream processing. CUDA streams are an organizational structure for performing CUDA operations. By default CUDA uses a single stream, which causes all CUDA functions to execute in issue-order sequence. However, by explicitly separating CUDA function calls into different streams, function calls execute asynchronously allowing multiple CUDA operations to occur simultaneously. In parallel OSEM reconstruction, large pinned memory transfers are organized into one stream, while LOR ray tracing operations are organized into another stream. This allows the program to perform look-ahead memory transfers of the next batch of LORs while the current batch of LORs is being (back)projected.

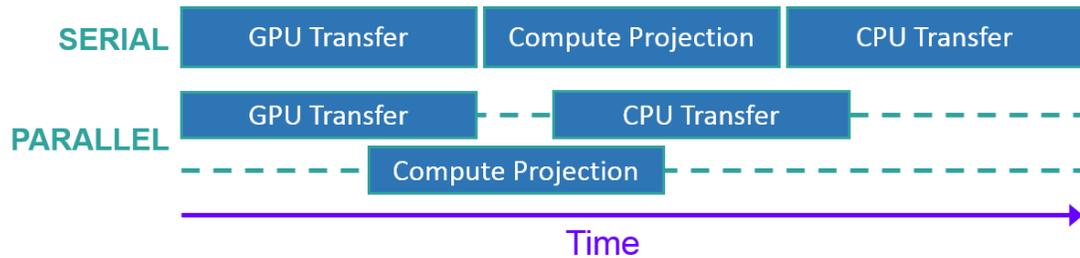


Figure 3.3: Diagram of concurrent stream processing in CUDA

### 3.6 Results

The results of benchmarking performed on a desktop computer system equipped with an Intel Xeon E5606 2.13 GHz CPU and a NVIDIA GTX 780 video card showed a 60x improvement

in OSEM reconstruction time between the serial CPU and parallel GPU implementations. CPU-based reconstructions took an average of 30 minutes to complete versus 30 seconds for GPU-based reconstructions. This improvement in program run time is consistent with other applications of CUDA parallel computing presented in NVIDIA product marketing materials. The reduced reconstruction time permits integration of additional reconstruction complexity from MRI-guided reconstruction methods, advanced system geometries, or other improvements without severe workflow restriction. Importantly, the two methods produce numerically similar reconstruction results. The mean voxel difference in a central region of interest (cylinder with 20 cm radius) between the CPU-based method and GPU-based method is -0.375% (SD  $\pm$ 9.78%) for reconstructions of a uniform sinogram. Numerical disagreement between the methods is due to ray-tracing algorithm differences and minor bug fixes.

### **3.7 Conclusions**

A GPU-based PET reconstruction software implementation was created and compared against a CPU-based implementation. The GPU version showed an improvement in reconstruction time of approximately 60 times as compared to a CPU-only implementation using Siemens standard collapsed PET geometry. Importantly, the images resulting from the CPU and GPU implementations were not appreciably different. The GPU version was also shown to be scalable to the more demanding uncollapsed native geometry. We conclude that the accelerated GPU implementation of the OSEM PET reconstruction algorithm is a better choice for PET reconstruction due to its significantly faster reconstruction processing and geometry scalability.

## CHAPTER 4: DIGITAL PET PHANTOM DEVELOPMENT

### 4.1 Introduction

The recent introduction of simultaneous PET-MR scanners has opened up a plethora of new areas of research utilizing multimodal patient data. However, due to practical, logistical, and funding limitations, the acquisition of real-world tri-modality (PET/MR/CT) data can be problematic. The developed electronic phantom (E-phantom) platform solves the PET data insufficiency problem for researchers and streamlines the research process without compromising anatomical diversity or PET realism.

The simultaneous PET/MRI imaging modality is still in its relative infancy with significant research and clinical challenges yet to conquer. Unlike PET/CT imaging, which offers few integration pathways between the two modalities, PET/MRI has the potential to provide true combination imaging. Nuclear imaging has historically been a separate research tract from MRI. This distinct separation of knowledge represents a unique challenge to researchers already proficient in one acquisition scheme. The primary goal of the E-phantom platform is to make PET imaging accessible to traditional MRI researchers now interested in PET/MRI combination imaging. In order to provide a suitable substitute for real-world acquired data, the E-phantom platform must offer substantial anatomic and disease variation, tracer distribution customization, and ease of operation.

A number of PET phantoms already exist. Three of the most popular phantoms are the Hoffman phantom, the Zubal phantom, and the XCAT phantom [19-21]. The Hoffman 3-D Brain Phantom<sup>TM</sup> (Biodex Medical Systems, New York, USA) is a physical phantom

comprised of nineteen fillable plastic plates formed in the shape of a human brain. The phantom can be filled with radioactive or contrast material and imaged using a variety of modalities. A digital reference also exists for the phantom [22]. Digital phantoms have several key advantages over physical phantoms, the first of which is cost and ease of study. Digital phantoms also offer greater control over imaging parameters and the ability to turn off physical effects not being investigated [23]. The Zubal phantom is full body digital phantom rendered in voxel form. Its main disadvantage is that it represents only a single anatomy. The XCAT phantom is a vector-based digital phantom that has greater anatomic diversity and scalability than the Zubal phantom, but it is still less anatomically diverse than the proposed E-phantom.

## **4.2 Methods**

*Phantom.* 28 Patient-based PET E-phantoms were constructed from intra-subject registered MR and CT images and the MNI ICBM 2009c nonlinear asymmetric atlas [24]. Intra-subject registrations were performed using FLIRT linear registration tool with six parameter model [25-27]. Atlas registrations were performed using the ANTs registration module for the Insight Toolkit (ITK) [28-30]. A flowchart of the phantom creation process is shown in Figure 4.1.

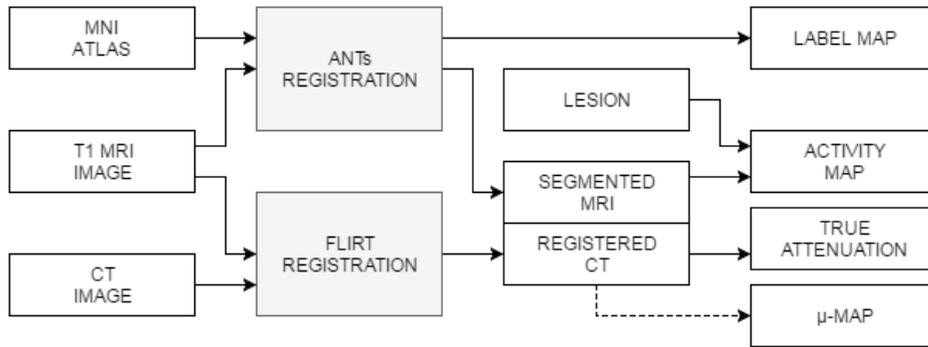


Figure 4.1: E-phantom creation flowchart

Each E-phantom consists of an activity map of desired tracer distribution and true-attenuation map. Activity maps are built from tissue-segmented MR (brain) and CT (skull/soft tissue) images with tracer-appropriate relative activity levels assigned by tissue class and literature (Figure 4.2). Activity maps may also be computed from acquired PET data when available.

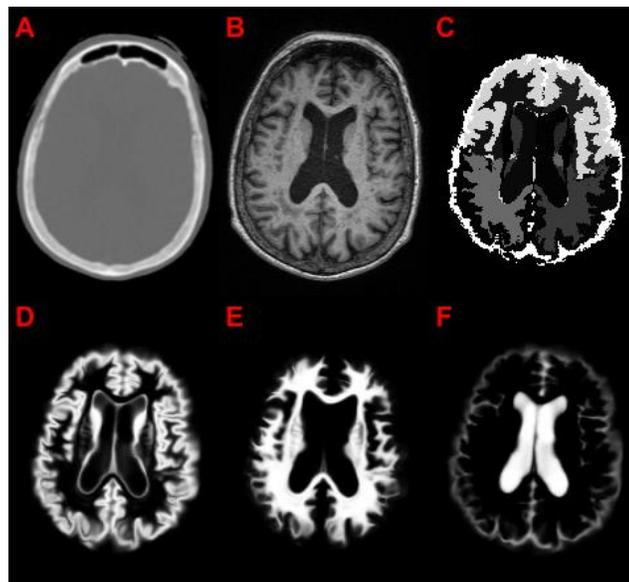


Figure 4.2: Inputs to the phantom creation process, CT (A), MRI (B), label map (C), segmented grey matter (D), segmented white matter (E), and segmented CSF (F)

Additional anatomical structures, such as hot or cold lesions, can also be added to the E-phantom. Lesions may be targeted to brain regions using atlas labels and limited by tissue class. True attenuation maps are created by conversion of CT-derived Hounsfield units into linear attenuation coefficients (LACs) at 511 keV [31]. An attenuation correction map ( $\mu$ -map) can optionally be added, using the researcher preferred derivation method (typically MRI-derived for PET/MRI applications). The result of the phantom creation process is an activity map of true relative radiotracer distribution, a true attenuation map, and a  $\mu$ -map for attenuation correction (Figure 4.3).

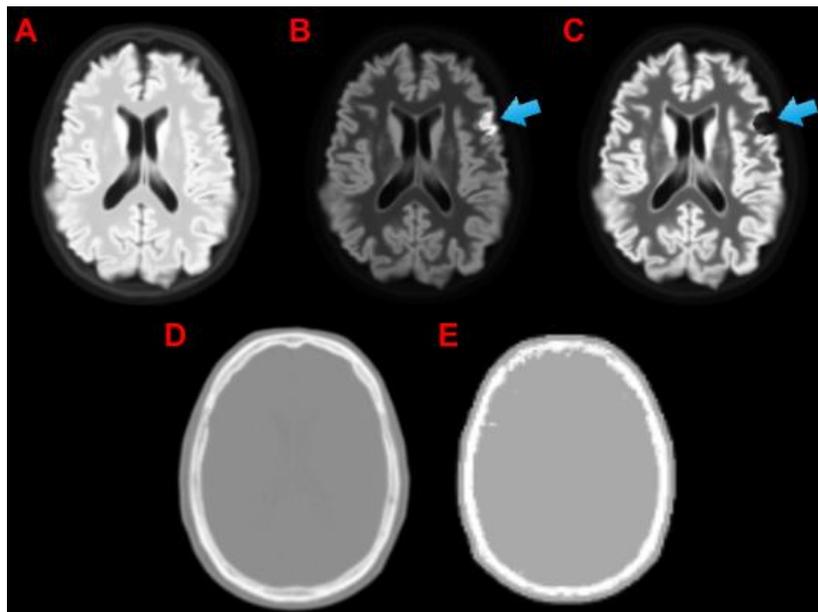


Figure 4.3: Example outputs of the phantom creation process, standard activity map (A), activity map with hot lesion (B-arrow), activity map with cold lesion (C-arrow), CT-based  $\mu$ -map (D), and MRI-based  $\mu$ -map (E)

*Simulation.* The goal of the simulated PET data acquisition process is to as closely as possible mimic the real world effects that impact PET data acquisition (Figure 4.4). The starting point for the simulation procedure is an image space map of the relative true activity distribution. The magnitude of voxel intensities in the activity map do not matter at this stage, only their relative difference compared to each other affect the reconstructed distribution. Along with the activity map is an image space map of true attenuation coefficients in the imaging field. The true attenuation map is different from the  $\mu$ -map used for attenuation correction, as no real world  $\mu$ -map perfectly represents the actual attenuation experienced by gamma photons. The activity map is first blurred slightly to represent positron range, photon non-collinearity, and detector response effects. The activity map and true attenuation map are then converted to projection space using native PET/MRI system geometry (Table 3.1). The projected activity is attenuated by dividing it by the projected true attenuation. The resulting projection represents the relative true number of events detected by the native scanner hardware during acquisition. This true projection is then collapsed to standard PET projection space by adding together native-geometry LORs in a specific format. The collapsed/standard PET projection space uses 5x fewer LORs to represent collected data than native projection space. Uniformity effects are applied to the collapsed true projection which represent the sensitivity of scanner geometry to the imaging field-of-view. The simulated data is then scaled to a target total count. Acquisition noise is applied via voxel-wise sampling from a Poisson distribution. The simulated PET data are then iteratively reconstructed using an industry-standard OSEM algorithm accelerated by GPU as described in Chapter 3. The reconstruction platform is capable of performing attenuation, randoms, and system uniformity corrections on

the simulated projected data but does not support scatter simulation or correction. Reconstructed PET images can be matched to manufacturer reconstruction space, allowing straightforward validation and comparison to clinically acquired PET images.

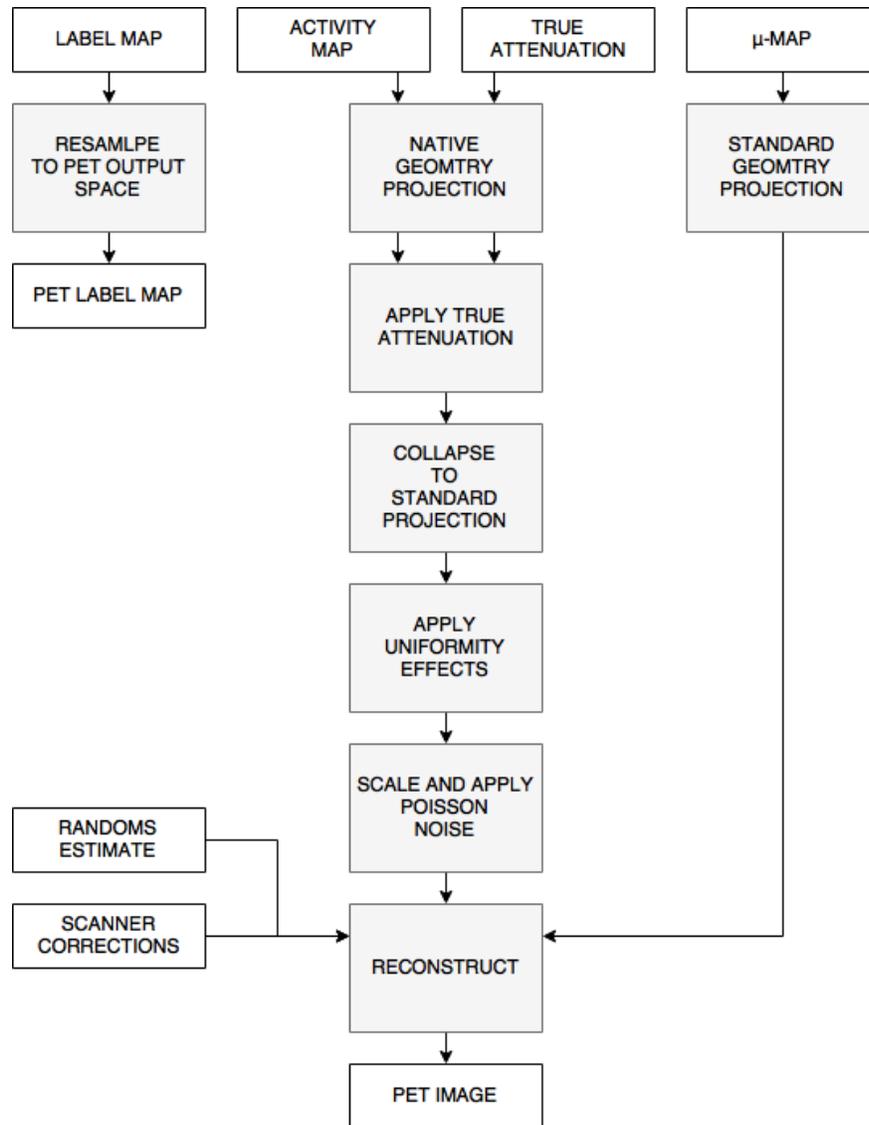


Figure 4.4: PET simulation flowchart

*Validation.* Validation and capabilities of the E-phantom platform were demonstrated using tri-modality ( $^{18}\text{F}$ -Florbetapir PET/MR/CT) data acquired from 28 subjects [32]. The data for these subjects were acquired at Washington University in St. Louis under an IRB-approved protocol and de-identified. The process was validated by regional and whole-brain error comparison between simulated and acquired PET images. Acquired PET data were reconstructed using manufacturer software with clinical-standard CT-based  $\mu$ -maps. These reconstructions were used as the activity map input to the E-phantom platform and reconstructed using the same clinical-standard CT-based  $\mu$ -maps. One application of the E-phantom platform is MRI-derived attenuation map development. To demonstrate its capability as a performance evaluation tool, error analysis between an MRI-derived ultrashort echo time (UTE)  $\mu$ -map and clinical-standard CT-derived  $\mu$ -map was performed using acquired PET data and E-phantom simulated PET data.

### **4.3 Results and Analysis**

Mean whole-brain error ( $\pm$ standard deviation) between manufacturer and E-phantom simulation platforms from the validation experiment was 1.54% ( $\pm$ 0.46%). Regional error varied from -2.14% to +3.02% (Figure 4.5).

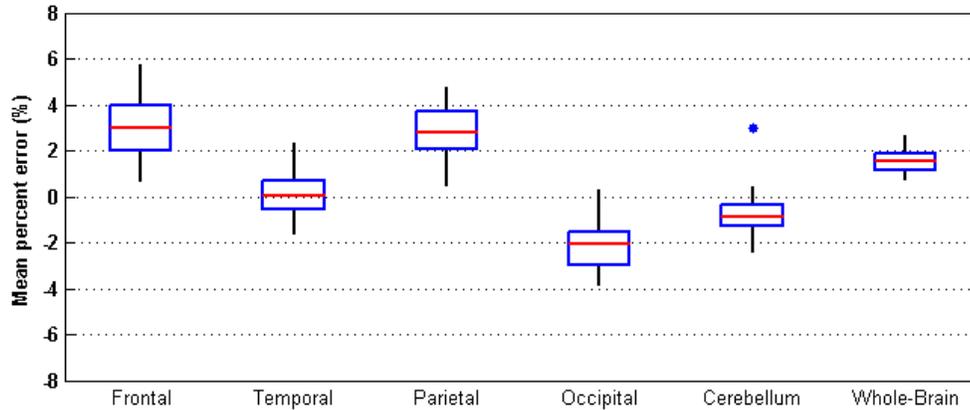


Figure 4.5: Results of E-phantom PET reconstruction process validation. Acquired PET data were reconstructed using manufacturer software with clinical-standard CT-based  $\mu$ -maps. These reconstructions were used as the activity map inputs to the E-phantom platform and reconstructed using the same clinical-standard CT-based  $\mu$ -maps.

Evaluation of MR-based  $\mu$ -map performance revealed strong linear correlation (slope = 0.90,  $r^2 = 0.93$ ) between  $\mu$ -map reconstruction error determined using acquired data with manufacturer reconstruction software and phantom data simulated with the E-phantom platform (Figure 4.6). Higher overall error in E-phantom simulations is likely due to photon scatter not being represented in the simulation model. This is analogous to perfect scatter correction by the E-phantom platform, while acquired reconstructions rely on imperfect models based on the  $\mu$ -map to correct for scatter. The result is a greater difference in overall PET distribution from the sum of applied corrections in the E-phantom simulation.

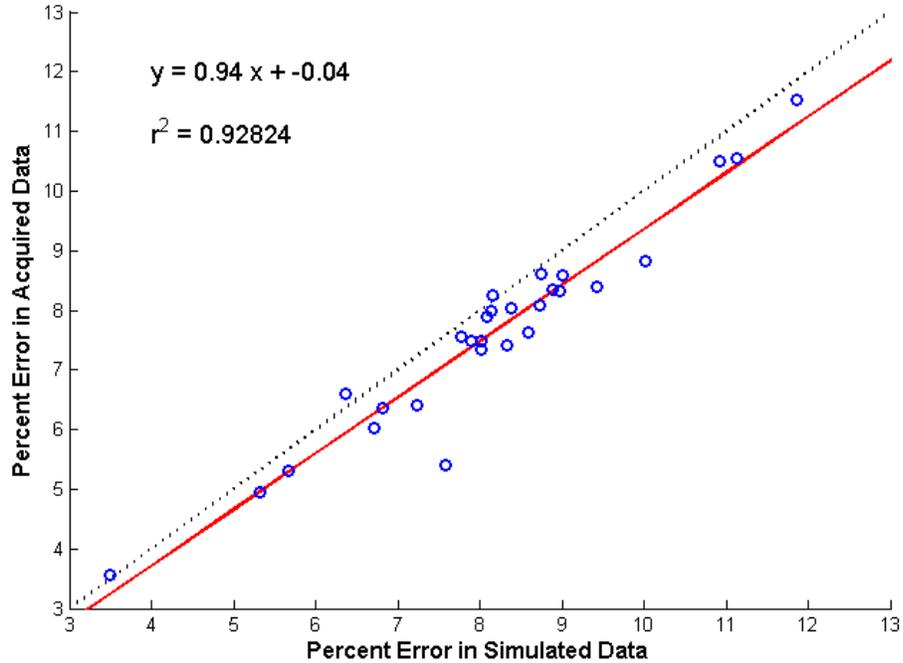


Figure 4.6: Scatter plot and regression of  $\mu$ -map reconstruction error determined using acquired data with manufacturer reconstruction software and phantom data simulated with the E-phantom platform. The regression analysis suggests a strong linear relationship between error analyses performed in E-phantom space and acquired data space.

The acquired and simulated reconstructions are visually alike, as seen in Figure 4.7, with similar tracer and error distribution. The small difference in percent error between acquired and simulated reconstructions is most apparent in regions experiencing high attenuation. In Figure 4.7, the center of the reconstructed head phantom experiences the most attenuation and also the greatest difference in percent error between the two reconstructions.

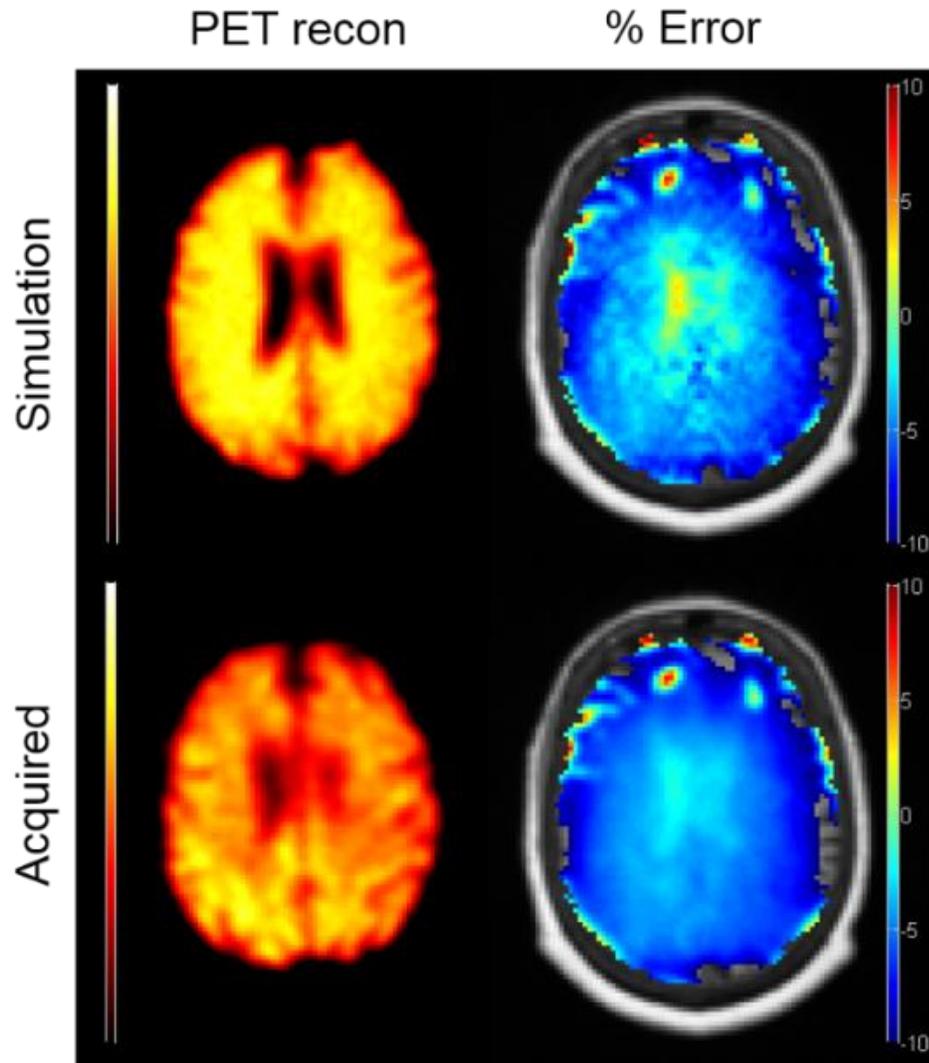


Figure 4.7: Visual comparison between acquired and simulated PET reconstructions. Error distributions from the simulated E-phantom platform were consistent with error distributions from acquired data reconstructed using manufacturer software.

The total time required to simulate acquisition of  $3.5 \times 10^8$  LORs, collapse to  $7.3 \times 10^7$  LORs, apply correction factors, and iteratively reconstruct one phantom was 15 minutes on a desktop computer system equipped with an Intel i7-860 CPU, NVIDIA GTX 780 GPU, and 8 GB RAM.

The E-phantom simulation platform provides numerous advantages over real-world acquired PET data. Since only bi-modal (MR/CT) data are required to utilize the E-phantom platform, larger research datasets can be constructed than typically available to PET-MR researchers. These large datasets can improve anatomical diversity and statistical power over acquired PET data alone. Additional capabilities of the E-phantom platform include the ability to selectively add region targeted hot and cold lesions and to model different PET tracer distributions, such as PIB, FDG, or Tau. As a demonstration of the E-phantom diversity, Figure 4.8 shows a montage of PET reconstructions carried out at different activity levels using a variety of phantoms. The platform has been demonstrated using MR-based  $\mu$ -map evaluation; other applications include study of PET dose reduction and MRI-guided reconstruction methods.

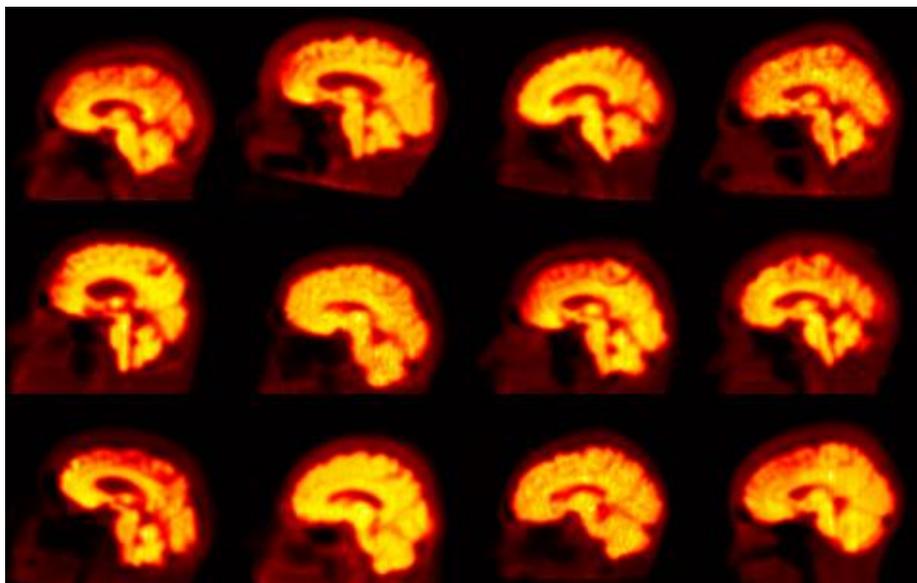


Figure 4.8: Anatomic and activity diversity in the E-phantom platform

## CHAPTER 5: MRI-GUIDED PET RECONSTRUCTION

### 5.1 Partial Volume Effect

The partial volume effect (PVE) is an inherent bias in PET imaging that affects the quantitative and qualitative interpretation of images. The effect is a combination of two phenomena that act to blur intensity values across a PET image. The first, called the finite resolution effect, is a result of the limited spatial resolution of the imaging system. Compared to other tomographic imaging modalities, PET has relatively poor resolution. The main limiting factors of PET resolution are positron range, photon noncollinearity, detector design (crystal arrangement, type, size), and the reconstruction process (number of iterations, algorithm choice, post-filtering) [33]. The result of the finite resolution effect is signal from source radiotracer is distributed over a wider volume in reconstructed images than its true position in the subject. Signal from tumor areas “spills-over” into adjacent regions and signal from adjacent regions “spills-into” tumor areas. The impact is that tumors in reconstructed PET images appear to be larger, less compact, and of lower intensity than they are in reality.

The second phenomenon, called the tissue-fraction effect, is a result of the nature of three dimensional sampling. Tissue boundaries invariably fall within single voxels, meaning multiple tissues contained in a voxel are represented by a single intensity. Put another way, voxels on source edges include source and background signal, and the resulting voxel intensity is an average of these two signal intensities. The impact is similar to the finite resolution effect, resulting in blurred boundaries and diffuse signal intensities. Tissue-fraction effect can be reduced by decreasing pixel size and slice thickness, which makes it less likely for multiple

tissue types to occupy the same voxel space. It is worth noting that the tissue fraction effect is not limited to PET systems, and it remains a problem with any imaging modality employing sampling.

PVE is a major problem for PET image reconstruction, on the same order of magnitude as other problems such as attenuation, scatter, noise, and motion-artifacts. PVE has three significant impacts on image reconstruction. First, the spilling of signal into and out of tumor regions greatly complicates the correction process by introducing compensating signal from surrounding tissue. This also affects the distribution of intra-tumor intensity, masking necrotic tumor areas and diffusing high activity regions [34]. Signal spilling results not only from tumor radiotracer uptake, but also tissue and organ uptake. This makes the impact of PVE spatially dependent on the background tissue surrounding the tumor.

Second, PVE blurring makes it difficult to determine tumor boundaries, and by extension, apparent tumor size. This is especially problematic for radiotherapy treatment planning, which may rely on PET imaging for targeting. Apparent tumor size also impacts the calculation of standard uptake value (SUV), which is widely used clinically to characterize tumor activity and monitor treatment progress. Partial volume effects lower SUV, but the magnitude of the decrease is dependent upon apparent tumor size, which causes PVE to affect SUV non-linearly as demonstrated by several clinical studies [33, 35, 36]. Smaller tumors are more impacted by partial volume effects than larger ones [37]. As a general rule, PVE severely impacts image quantification and qualification when tumors are less than two to three times the PET image resolution [38]. PVE impact also depends on tumor shape or “compactness”, as determined by

tumor surface area per volume, with spherical tumors being most compact. The less compact a tumor, the more it is affected by partial volume effects [33].

Third, PVE affects measured tumor intensity by spreading intensity over a wider volume. Unlike attenuation, PVE does affect the total activity of the tumor (in the absence of background). It merely redistributes the activity across a wider volume. The impact of PVE on maximum intensity within a region of interest (ROI) is relatively moderate at the single voxel level, but it creates an additional tradeoff for clinicians who rely on tumor activity levels to calculate SUV [33]. If activity is measured across a multi-pixel ROI, the SUV is impacted by apparent tumor size determination, and if it is calculated at the single voxel level, the activity levels are distorted by noise and may not accurately represent heterogeneous tumors.

The oldest correction method for partial volume effects uses simple, high-contrast phantoms to estimate errors caused by PVE. A series of tumor phantoms with known activity of varying size and shape are imaged, and a set of recovery coefficients (RC) are determined by comparing the known quantities with the resulting images [37]. After a clinical scan, the identified tumors are then matched with the most closely resembled phantom, and the respective RC is used to correct the calculated tumor SUV for partial volume effects. Regional background spill-in can also be accounted for using an assumed uniform activity level for the surrounding tissue [39]. The RC method is still commonly used to correct for PVE in PET tumor imaging [33]. There are, however, several limitations of this method. Tumors are assumed to be simple geometric shapes of uniform activity, RC values are specific to both individual scanners and the spatial position in the scanner at which the phantom is imaged, and the method is only viable for quantitative correction, not reconstructed image enhancement.

The RC method essentially compartmentalizes the body as either tumor or not-tumor, and uses this assumption to correct for PVE. A logical progression of the RC method is the geometric transfer matrix method (GTM), which does not limit the number of compartments [33]. Structures are selected for compartmentalization based on the region of interest. For example, the brain could be divided into white matter, grey matter, and cerebrospinal fluid. The activity contribution of each of these compartments on the others is then calculated by solving a matrix of  $n$  equations and  $n$  unknowns [40]. As with basic RC correction, activity is assumed to be uniform within each compartment. Structure identification is typically achieved using registered CT or MR data. Common sources of error for this method are incorrect compartment segmentation and anatomical registration error.

The RC and GTM correction methods apply only at a regional level and for the most part only affect intensity quantification. More sophisticated PVE correction methods apply at the voxel level and change not only quantification, but also the final qualitative image. These correction methods typically rely on registered anatomical data from CT or MR to correct reconstructed images. An exception to this is iterative deconvolution, which attempts to reconstruct the true image using various iterative techniques based on discrete Fourier transform (DFT), maximum likelihood principle (ML), or least squares (LS) criterion to name a few [41]. Deconvolution relies solely on PET data, but it suffers from heavy noise amplification. Another non-anatomical approach to voxel level correction attempts to fit a spherical tumor convolved with the system point spread function to the PET activity data by solving an objective function [42]. The major limitations of this method are it assumes a spherical tumor shape and uniform uptake for both tumor and background [33].

All of the methods discussed thus far are applied after initial image reconstruction. The most current methods of PVE correction accommodate for partial volume effect during the actual image reconstruction. These methods essentially incorporate the partition correction method directly into the reconstruction process [43]. The advantages of considering PVE during reconstruction are tissue boundaries can be better maintained, and noise levels in the partition of interest, the tumor, are less impacted [33]. The problems with anatomical image registration still apply to PVE accommodated reconstruction, and the PVE correction cannot be removed later as it occurs as part of the image creation process.

## **5.2 Overview of Existing Methods**

MRI-guided PET techniques may be divided into three broad categories: those which address issues related to the substitution of PET/MRI imaging for PET/CT imaging (attenuation correction mainly), those which address limitations of PET imaging, and those which attempt to synthesize new information from the merging of two modalities [44]. The focus of this work is the second category of methods, which aims to incorporate MRI information into the reconstruction process to improve PET imaging. Two major limitations of PET imaging compared to MRI are image degradation due to partial volume effects and image degradation due to motion blurring [44]. This chapter will address partial volume/resolution correction of PET images using MRI information.

Anatomical information incorporated directly into the PET reconstruction process is typically referred to as a “prior.” Historically derived from either CT or MR images, priors are typically used as a component of smoothing to reduce noise resulting from iterative PET

reconstruction. Unlike smoothing priors based solely on the Gibbs distribution or post-reconstruction shift-invariant filters, anatomical priors are potentially both edge preserving and spatially varying [8]. The major underlying assumption of anatomical priors is correspondence between structural information from the prior and functional information from the collected PET data.

Gindi et al. proposed an MRI anatomical edge prior based on the expected correlation between anatomical and functional boundaries [45]. The edge prior was defined by hand and incorporated into the two dimensional maximum a posteriori (MAP) reconstruction process as a Gibbs distribution [46]. Results from simulation on a piecewise defined phantom showed improved region of interest root mean squared (ROI-RMS) error compared to MLEM reconstructions. The authors noted their method was computationally intensive and lacked an automated method of edge determination. Ouyang et al. also used an anatomical Gibbs prior to promote boundaries [47]. But unlike the Gindi et al. method, the Ouyang et al. model used an automated cross correlation “weighted line site” function to control prior contribution to boundary enhancement.

Chiao et al. proposed an extension to their previously reported model-based joint maximum likelihood estimator (MJMLE) to improve delineation of myocardial boundaries in emission computed tomography (ECT) [48]. Unlike other prior methods of the time, Chiao et al. focused on the issue of alignment of low-count gated ECT reconstructions. The method was comprised of a polygonal heart model to improve alignment between ECT, MRI, and CT measurements and a regularization component that penalized the log-likelihood function to promote boundary smoothness. The method was limited to 2D translational alignment and was

designed to operate independently of available MRI/CT anatomical data, instead using only boundary information derived from exist ECT data. Boundary regularization was required to guarantee convergence of the MJMLE algorithm, and it had the added result of reducing the number of iterations to convergence. Visual comparison between MJMLE and filtered backprojection images showed good boundary agreement.

Ardekani et al. specifically addressed the problem of partial volume effect in PET imaging with their cross-entropy prior model [49]. Unlike previous Gibbs priors, the Ardekani method (MXE) used an objective function comprised of the weighted sum of two cross-entropy terms, one for similarity to measured emission data and one for similarity to prior anatomical MR data. An adaptive smoothing filter was also included. MXE reconstructions were found to be substantially less noisy than standard maximum likelihood expectation maximization (MLEM) reconstructions, but as with any prior term, image bias depended heavily on the degree of overall prior contribution to the reconstruction controlled by a scalar hyperparameter. Bowsher et al. proposed a segmentation-free smoothing Markov prior which operated on a position dependent neighborhood of voxels using a similarity metric to enforce regularity between similar ROIs based on MR anatomical information [50]. Unlike other investigators which validated methods using simulation studies, Bowsher et al. used high resolution small animal models and histology studies to provide rigorous analysis of their method.

Yan et al. recently proposed a method of MRI-guided PET using the “guided filter” described by He et al. [51, 52]. The utilized guided filter is an edge-preserving smoothing filter like the bilateral filter, but it avoids gradient reversal artifacts. It is based on the assumption of a local linear model between a guidance image (in this case a high-resolution MRI) and the

filtering output. The method cost function accounts for both anatomical similarity and partial volume effects of PET imaging. Simulation results showed substantial improvement over uncorrected PET reconstruction in mean squared voxel error and peak SNR. Quantitative analysis of neurological clinical resulted in reduced bias, and visual inspection of corrected PET images showed less MRI appropriation of the image content than is generally seen in prior-based MRI-guided PET methods.

Nuyts et al. compared maximum a posteriori (MAP) to post-processed maximum likelihood (ML) methods of incorporating anatomical information and found prior-based MAP methods superior [53]. A recent review of MRI-guided PET reconstruction strategies by Bai et al. supports this conclusion, as the majority of current methods directly incorporate MRI anatomical information into the PET reconstruction algorithm [1]. A Monte Carlo simulation-based comparison of popular partial volume correction (PVC) methods by Hutton et al. specifically aimed at simultaneous PET/MRI imaging drew a similar conclusion. In their simulation, reconstruction-based methods outperformed post-processed corrections in recovery of lesion contrast and demonstrated greater resilience to registration errors [54].

The effects of MRI registration error [3, 55-58] and segmentation error [57-59] on PET partial volume correction have been investigated by a number of researchers. Misaligned registrations of one PET voxel have been reported to cause anywhere from 3-8% error in the resulting PET images. PET region of interest (ROI) bias was found to be robust to MR registration error, while ROI standard deviation was extremely sensitive to registration error [57]. In studies investigating both segmentation and registration errors in PVC, segmentation errors were found to be more influential on PET activity accuracy than registration errors [57,

58]. Gutierrez et al. investigated the evolution of segmentation methods used by the Statistical Parametric Mapping (SPM) toolbox for the purpose of MR-guided PET and found that the differences between SPM version segmentation algorithms highly influenced the results of MRI-guided partial volume correction [59].

There have historically been three primary challenges to improving PET image quality using MRI: 1) registration of MRI and PET imaging volumes, 2) segmentation of MRI, and 3) incorporation sequence of MRI information. Registration problems are largely overcome by hybrid PET/MRI scanners (ignoring dynamic imaging), segmentation problems can be avoided by using voxel-based smoothing priors, and MRI information has been shown to be best incorporated directly within the PET reconstruction process. Thus, a voxel-based method based on data from a hybrid PET/MRI scanner that incorporates MR information directly into the PET reconstruction process is the ideal starting point for an MRI-guided PET algorithm. The proposed MRI-guided PET method fulfills these requirements. Besides using data from a PET/MRI hybrid scanner, the proposed method differs from previously studied methods in two key areas. First, the MRI information incorporated into the prior is not limited to a single image. Instead, all anatomically (and perhaps functionally) relevant MRI images can be used as input to the prior. Second, the PET reconstruction is completed at MRI-scale, which usually means the resulting image contains smaller voxels than standard PET images. This means the MRI prior information is incorporated at its near-native resolution, resulting in maximum prior contribution and minimum partial volume effect.

### 5.3 Formation of the Anatomical Prior

The developed MRI-guided PET reconstruction algorithm is built from the previously validated accelerated OSEM algorithm described in Chapter 3, which is itself an accelerated version of MLEM reconstruction. The key change to make the algorithm accept prior MRI information is a shift from a maximum likelihood (ML) criterion to a maximum *a posteriori* (MAP) criterion. Instead of choosing the image  $f$  with greatest  $p(g|f)$ , where  $g$  is the measured data, the MAP criterion selects the image  $f$  which maximizes the posterior probability density function  $p(f|g)$  [8]. The MAP criterion may be expressed symbolically as:

$$\hat{f} = \arg \max_f p(f|g) \quad (5.1)$$

By substituting  $p(f|g)$  using Bayes' rule, the MAP criterion can be expressed:

$$\hat{f} = \arg \max_f \frac{p(g|f) \cdot p(f)}{p(g)} \quad (5.2)$$

The probability  $p(g)$  may be omitted because it is not dependent on image  $f$ . The form of the MAP algorithm may be further simplified by taking the logarithm of the quantity to be maximized:

$$\hat{f} = \arg \max_f [\ln p(g|f) + \ln p(f)] \quad (5.3)$$

The quantity  $p(f)$  in Equation 5.3 is known as the prior in the MAP algorithm. If the prior term is removed or uniform distribution, Equation 5.3 reduces to the ML solution. Thus, the MAP criterion is equivalent to introducing a log-prior penalty term to the ML criterion. The prior term introduces bias to the normally unbiased ML criterion, forcing the criterion to prefer solutions which are consistent with the prior term [8].

As noted in Chapter 5.2, the prior may take many forms, many of which use a Markov random field (MRF) model to specify the joint distribution in terms of local “potentials” to describe the interaction between groups of neighboring voxels [1, 60]. It is worth noting that the defined neighborhood does not have to include adjacent voxels. For our implementation, we have chosen a prior in the form of a Gibbs probability density function, which expresses the joint density of a MRF [46]:

$$p(f) = \frac{1}{Z} \exp[-\beta U(f)] \quad (5.4)$$

The  $Z$  term in Equation 5.3 is a normalizing constant, which is unused in the MAP solution. The  $\beta$  term is a weighting parameter for  $U(f)$ , which is an energy function of weighted potentials from a predetermined neighborhood of voxels in the prior. When the Gibbs distribution is inserted into the MAP criterion, and the likelihood function is inserted for the probability  $p(g|f)$ , the MAP criterion takes the form:

$$\hat{f} = \arg \max_f [\ln L(f) - \beta U(f)] \quad (5.5)$$

where  $L(f)$  is the likelihood function. The energy term  $U(f)$  acts as a penalty to the maximum log-likelihood weighted by hyperparameter  $\beta$ . The magnitude of  $\beta$  determines the degree of smoothing imposed by the prior and acts as a balancing parameter between the maximum log-likelihood solution and the MAP solution. Determining the appropriate value of  $\beta$  is a difficult task as the degree of smoothing imposed by the prior can vary from subject-to-subject depending on the nature of the energy function. The choice of  $\beta$  is also a matter of qualitative visual preference, as MAP solutions can appear very different from ML solutions to trained PET image interpreters.

The energy function  $U(f)$  is determined for each voxel based on the prior contribution from its set of neighbors. Neighborhood  $k$  is defined as a cuboid of voxels surrounding each voxel  $j$ . Symbolically this relationship is defined by:

$$U(f) = \sum_j^N \sum_k w_{jk} M_{jk} \quad (5.6)$$

where  $w_{jk}$  is the weighting for each voxel-neighbor pair and  $M_{jk}$  is the prior contribution of the voxel-neighbor pair. The weighting for each neighbor  $k$  is defined as the inverse of one plus the center-to-center distance in centimeters from voxel  $j$  to neighbor  $k$ . By having the voxel-neighbor contribution weighted by the inverse of distance, neighbors closest to voxel  $j$  have the greatest contribution to the energy penalty function. Prior contribution  $M_{jk}$  is determined by Equation 5.7.

$$M_{jk} = \frac{(f_j - f_k)^2}{\max_n [A_j^n - A_k^n]^2 + \epsilon} \quad (5.7)$$

The prior compares the squared difference of the current PET activity voxel  $f_j$  and neighbor activity  $f_k$  to the squared maximum intensity difference of all anatomical images for voxel  $A_j$  and neighbor  $A_k$ . Anatomical images ( $A$ ) were individually normalized by maximum intensity magnitude. The prior function was tested using the combination of a normalized T<sub>1</sub>-weighted MRI image and a normalized T<sub>2</sub>-FLAIR MRI image, though any measure of anatomical similarity that can be represented as an array of dimension equal to PET reconstruction output size may be substituted as a contribution to the prior.

#### 5.4 Patient Data Acquisition

Imaging data for testing the MAP reconstruction algorithm were obtained from two studies.  $^{18}\text{F}$ -Florbetapir (Amyvid [Avid], Eli Lilly, Indianapolis, IN) PET data and MR images were acquired on a hybrid PET/MR system (Biograph mMR, Siemens, Erlangen, Germany) from subjects at Washington University Hospitals (St. Louis, MO). The study used an IRB-approved protocol and informed consent. Patients were injected with approximately 370 MBq of the  $^{18}\text{F}$ -Florbetapir tracer, and PET acquisitions were begun either immediately after injection or 50 minutes after injection. No participants had comorbidities that could interfere with testing, and participants received no additional radiotracer administrations within 24 hours. The enrollment criteria included contraindications to PET, PET/CT, or PET/MRI.

$T_1$ -weighted MR images were acquired using a sagittal MPRAGE sequence with the following imaging parameters: Repetition Time (TR)/Echo Time (TE) = 2300/2.95 ms, Inversion Time (TI) = 900 ms, flip angle =  $9^\circ$ , number of partitions = 176, field-of-view (FOV) =  $256 \times 256$  voxels, and voxel size  $1.055 \times 1.055 \times 1.200$  mm<sup>3</sup>.

$T_2$ -FLAIR MR images were acquired using an axial FLAIR sequence with the following imaging parameters: Repetition Time (TR)/Echo Time (TE) = 9000/91 ms, Inversion Time (TI) = 2500 ms, flip angle =  $150^\circ$ , number of partitions = 35, field-of-view (FOV) =  $256 \times 256$  voxels, and voxel size  $0.859 \times 0.859 \times 5.000$  mm<sup>3</sup>.

Vendor-provided  $\mu$ -maps were acquired using an axial UTE sequence with the following imaging parameters: TR/TE1/TE2 = 11.94/0.07/2.46 ms, flip angle =  $10^\circ$ , number of partitions = 192, FOV =  $192 \times 192$  voxels, and voxel size  $1.563 \times 1.563 \times 1.563$  mm<sup>3</sup>.

$^{18}\text{F}$ -fluorodeoxyglucose PET data and MR images were acquired on a hybrid PET/MR system (Biograph mMR, Siemens, Erlangen, Germany) from subjects at University of North Carolina at Chapel Hill Hospitals (Chapel Hill, NC). Patients were injected with approximately 190 MBq of the  $^{18}\text{F}$ -FDG tracer. The study used an IRB-approved protocol and informed consent.

$T_1$ -weighted MR images were acquired using a sagittal MPRAGE sequence with the following imaging parameters: Repetition Time (TR)/Echo Time (TE) = 1900/2.16 ms, Inversion Time (TI) = 900 ms, flip angle =  $9^\circ$ , number of partitions = 224, field-of-view (FOV) =  $208 \times 256$  voxels, and voxel size  $0.859 \times 0.859 \times 1 \text{ mm}^3$ .

$T_2$ -FLAIR MR images were acquired using an axial FLAIR sequence with the following imaging parameters: Repetition Time (TR)/Echo Time (TE) = 8510/100 ms, Inversion Time (TI) = 2500 ms, flip angle =  $160^\circ$ , number of partitions = 30, field-of-view (FOV) =  $208 \times 256$  voxels, and voxel size  $0.859 \times 0.859 \times 4 \text{ mm}^3$ .

Vendor-provided  $\mu$ -maps were acquired using a coronal Dixon sequence with the following imaging parameters: TR/TE1/TE2 = 3.6/1.23/2.46 ms, flip angle =  $10^\circ$ , number of partitions = 128, FOV =  $192 \times 126$  voxels, and voxel size  $2.604 \times 2.604 \times 3.120 \text{ mm}^3$ .

PET Reconstructions were performed using the GPU-accelerated MAP reconstruction algorithm with Gibbs prior described in Chapter 5.2. A variety of values for hyperparameter  $\beta$  were used. Reconstructions were performed using 3 iterations and 24 subsets with corrections for attenuation, randoms, scatter, and scanner geometry. PET image output size was set to dimension  $400 \times 400 \times 140$  with voxel size  $1.05 \times 1.05 \times 2.03 \text{ mm}$ . The neighborhood size was set to voxel dimensions of  $5 \times 5 \times 3$  (xyz), corresponding to a maximum neighbor distance

of roughly 2 mm from the center voxel. T<sub>1</sub>-weighted and T<sub>2</sub>-FLAIR MRI images registered to PET output space served as anatomical guidance for the prior. Noise and mean intensity were evaluated regionally using standard deviation and coefficient of variation metrics. Profile analysis was used to evaluate prior edge preservation.

### **5.5 Simulation Data Acquisition**

A simulation study was performed using the digital phantoms described in Chapter 4. The phantoms simulated acquisition of an <sup>18</sup>F-FDG PET study with mean total counts of  $2.7 \times 10^8$ . Counting noise, attenuation, and geometry effects were both simulated and corrected for in the reconstruction algorithm. A UTE-based MRI-derived  $\mu$ -map was used for attenuation correction. Three phantoms were simulated. The first phantom was lesion-free. The second phantom had a 10 mm cold-lesion placed in the left frontal grey matter with 10% normal activity. The third phantom had a 10 mm hot-lesion placed in the left frontal grey matter (same location as cold-lesion phantom) with 190% normal activity. Noise and mean intensity were evaluated regionally using standard deviation and coefficient of variation metrics. Profile analysis was used to evaluate prior edge preservation.

Lastly, a bias-noise analysis was performed to determine 2<sup>nd</sup>-order noise and bias characteristics of the developed prior. 50 noisy PET simulations were performed using the hot-lesion phantom for  $\beta = 0$ ,  $\beta = 6 \times 10^{-8}$ , and  $\beta = 6 \times 10^{-9}$ . Post-reconstruction Gaussian blurring with FWHM of 3 mm was applied to the  $\beta = 0$  reconstructions to compare the effects of post-reconstruction smoothing to prior contribution. A noise-free simulation served as the true activity distribution.

## 5.6 Results and Analysis

Reconstructions were performed on real subject data from  $^{18}\text{F}$ -FDG and  $^{18}\text{F}$ -FBP radiotracer acquisitions. To determine the effect of hyperparameter beta ( $\beta$ ) on the reconstructed activity distribution, the coefficient of variation (CV) was computed for cerebellum and caudate regions of interest (ROI) (Figures 5.1, 5.2). The dotted lines in each plot represent the CV of the MLEM ( $\beta=0$ ) reconstruction with Gaussian 3 mm FWHM post-reconstruction smoothing applied.

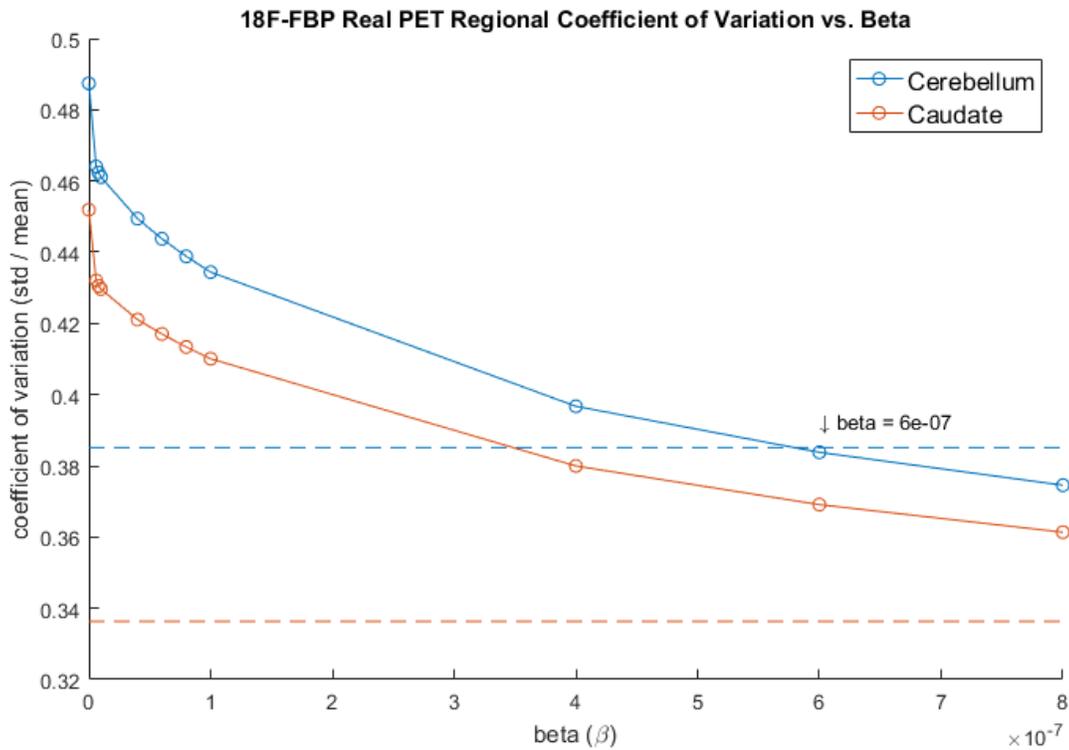


Figure 5.1: Comparison of coefficient of variation vs. beta for  $^{18}\text{F}$ -FBP reconstruction of acquired data.

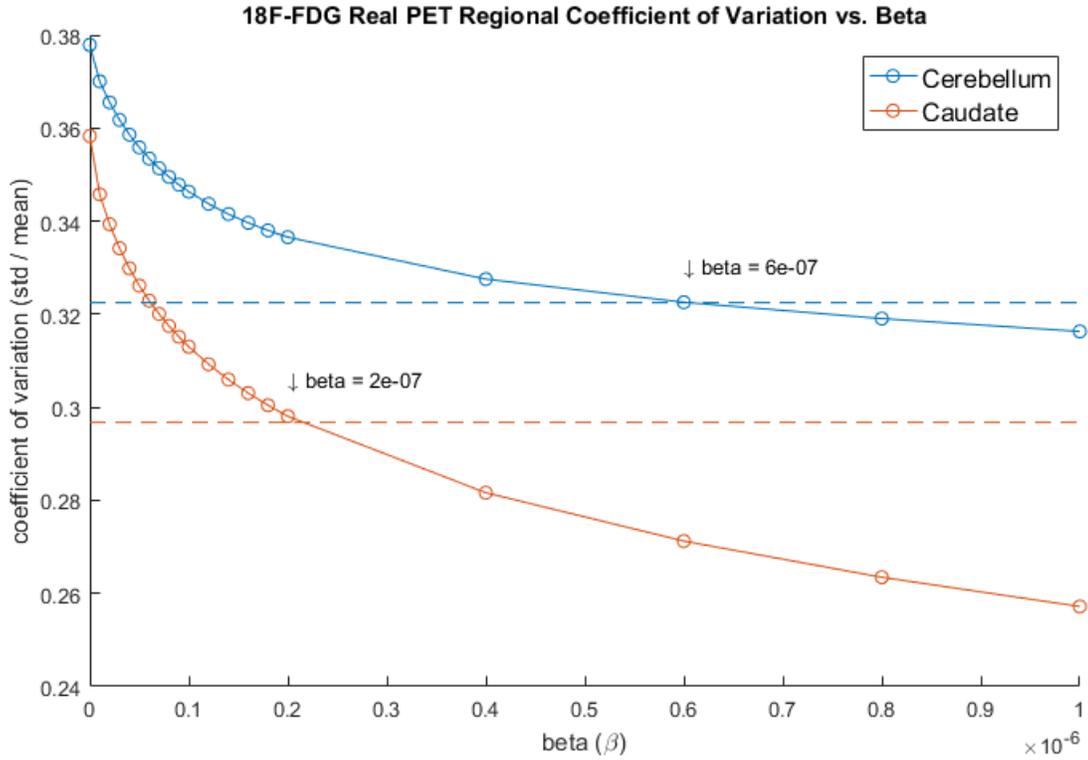


Figure 5.2: Comparison of coefficient of variation vs. beta for  $^{18}\text{F}$ -FDG reconstruction of acquired data.

The  $\beta$ -value of the MAP reconstruction with equal CV to the post-smoothed MLEM reconstruction varied by radiotracer (intersection of solid and dotted lines in Figures 5.1, 5.2). For each radiotracer the lower  $\beta$ -value corresponded with the highest intensity brain tissue for the given radiotracer. For  $^{18}\text{F}$ -FBP radiotracer, the cerebellum region (white matter) matched the post-smoothed CV with lowest  $\beta$ -value. For  $^{18}\text{F}$ -FBP radiotracer, the caudate region (grey matter) matched the post-smoothed CV with lowest  $\beta$ -value. This outcome is likely related to the PET image estimate term in Equation 5.7 ( $f_j - f_k$ ), which scales the contribution of the prior by the existing PET image intensity. It is preferable to use the smallest amount of

smoothing necessary for a given task to minimize bias due to smoothing. The  $\beta$ -values identified by CV matching were used for subsequent analysis.

Representative reconstructions of real  $^{18}\text{F}$ -FBP and  $^{18}\text{F}$ -FDG data using the MAP algorithm show the effect of increasing  $\beta$  on image smoothness (Figures 5.3, 5.4).

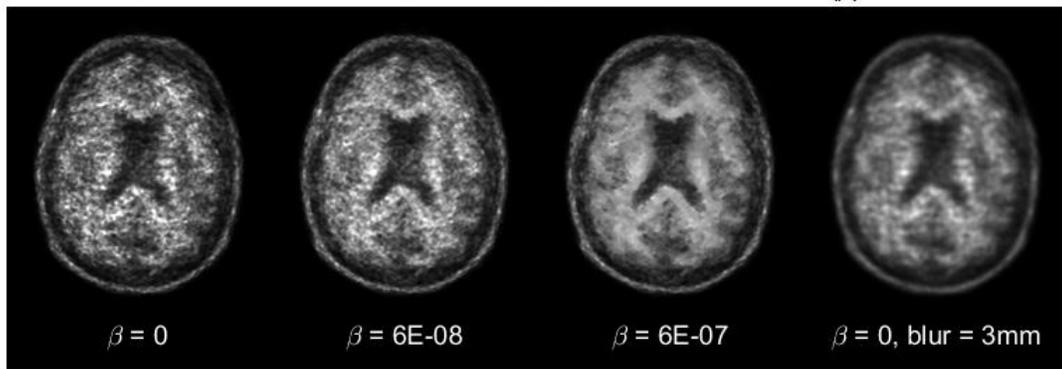


Figure 5.3:  $^{18}\text{F}$ -FBP PET reconstructions using real subject at various  $\beta$ -values

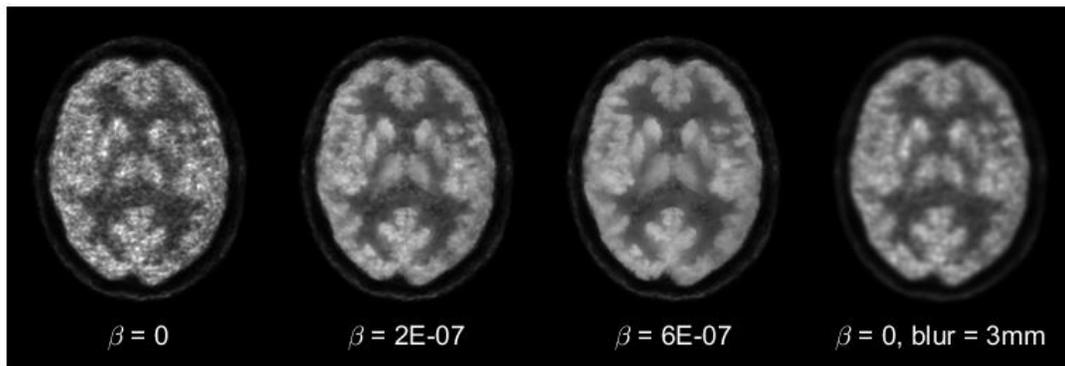


Figure 5.4:  $^{18}\text{F}$ -FDG PET reconstructions using real subject data at various  $\beta$ -values

In the  $^{18}\text{F}$ -FBP reconstructions (Figure 5.3), as  $\beta$  increases smoothing increases first in the white matter regions. In the  $^{18}\text{F}$ -FDG reconstructions (Figure 5.4), as  $\beta$  increases smoothing increases first in the grey matter regions. Overtuning  $\beta$  can cause the MAP reconstructions to take on a “paint-by-numbers” regional look to them (Figure 5.5c). The contribution of the prior becomes so large that it overwhelms the contribution of the measured data, causing the image to look overly smooth and more MRI-like than PET-like. These types of images are difficult to justify in a clinical setting.

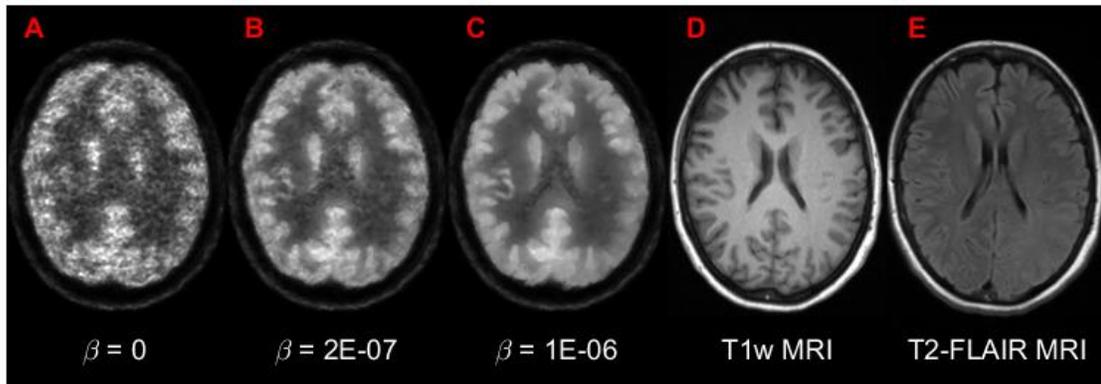


Figure 5.5: Example reconstruction with overtuned hyperparameter  $\beta$  (C)

Further increasing  $\beta$  can cause the MAP algorithm to break down. This occurs when a voxel change in the PET image estimate from one iteration to next is larger than the current estimate for that voxel, pushing the voxel intensity to zero. Because the image estimate update between iterations is multiplicative, once a voxel estimate reaches zero its value is fixed. Although the MLEM  $\beta=0$  reconstructions in Figure 5.3 and Figure 5.4 appear to contain more detail than the

larger beta reconstructions, the majority of the additional variation is due to excessive noise resulting from the MLEM algorithm as explained in Chapter 2.2.

The bias introduced by the MAP algorithm can be quantified by comparing the regional difference between MAP and MLEM reconstructions (Figures 5.6, 5.7).

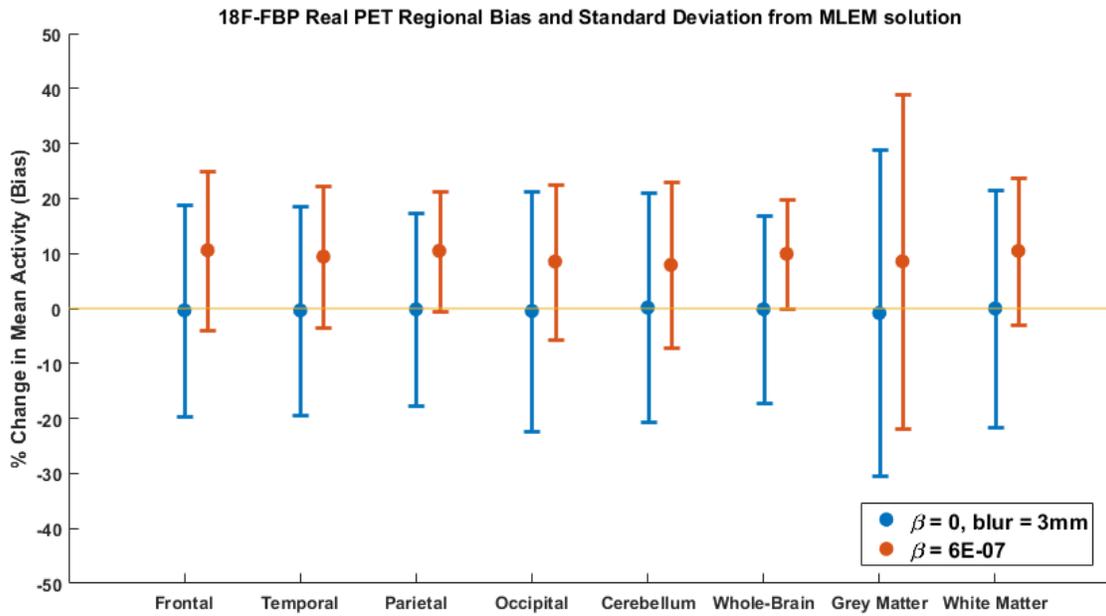


Figure 5.6: Comparison of  $^{18}\text{F}$ -FBP PET reconstruction regional bias

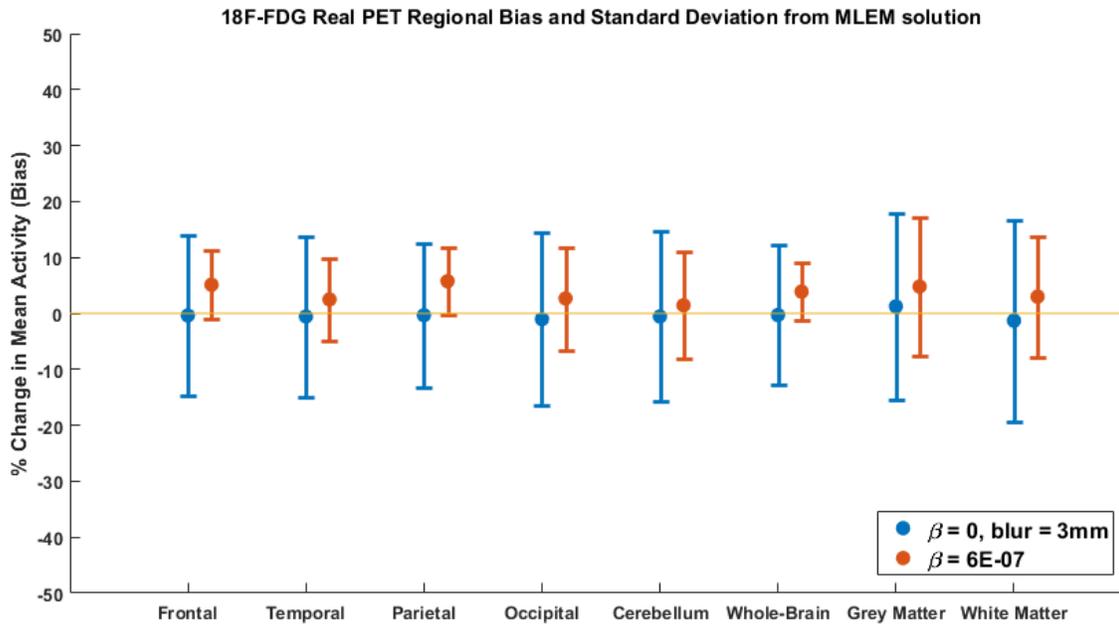


Figure 5.7: Comparison of  $^{18}\text{F}$ -FDG PET reconstruction regional bias

Post-reconstruction smoothing using a Gaussian filter has little-to-no change of mean regional intensity, while MAP smoothing does impose a bias on the reconstructed images with respect to the MLEM solution. The error bars in Figure 5.6 and Figure 5.7 represent the percent change in standard deviation of regional intensity from the MLEM solution. Imposing a bias on regional intensity results in better reduction in image variance (smaller error bars in Figures 5.6 and 5.7) for an equal level of smoothing as measured by some metric (in this case regional CV). Thus, the key to producing favorable MAP smoothing is to ensure that the bias introduced makes sense from a radiopharmacological perspective. Again, a difference in applied smoothing is evident between  $^{18}\text{F}$ -FBP and  $^{18}\text{F}$ -FDG reconstructions. The  $^{18}\text{F}$ -FBP MAP reconstruction strongly reduces variance in white matter, while the  $^{18}\text{F}$ -FDG reconstruction has more evenly applied smoothing. This difference in the regional smoothing may correspond to

the specificity of the radiotracer, as  $^{18}\text{F}$ -FBP appears to have both more specific smoothing and uptake than  $^{18}\text{F}$ -FBP.

In Figure 5.8 mean axial images of the 50-reconstruction noisy simulated data (left column) are displayed next to their corresponding bias images (middle column) and standard deviation images (right column). The true activity distribution was a noise-free reconstructed image. For this type of digital phantom, a true image space representation of the activity distribution exists only as a relative intensity map, because activity is determined in projection space by total number of counts detected. A noise-free reconstruction is therefore the best non-biased representation of the true simulated activity distribution. The MLEM reconstruction (Figure 5.8b) displays very low bias and very high noise. The MAP reconstructions (Figure 5.8c-d) have increased bias compared to the MLEM solution, but also significantly less noise. Bias is largest at the edges of grey matter. The post-smoothed MLEM reconstruction (Figure 5.8e) displays modest levels of bias and noise. Feature-wise, MAP solution e has significantly sharper edge detail than the post-smoothed solution. Based on the bias pattern, the MAP solution tends to overestimate grey matter edge intensity, while the post-smoothed MLEM solution tends to underestimate grey matter edge intensity.

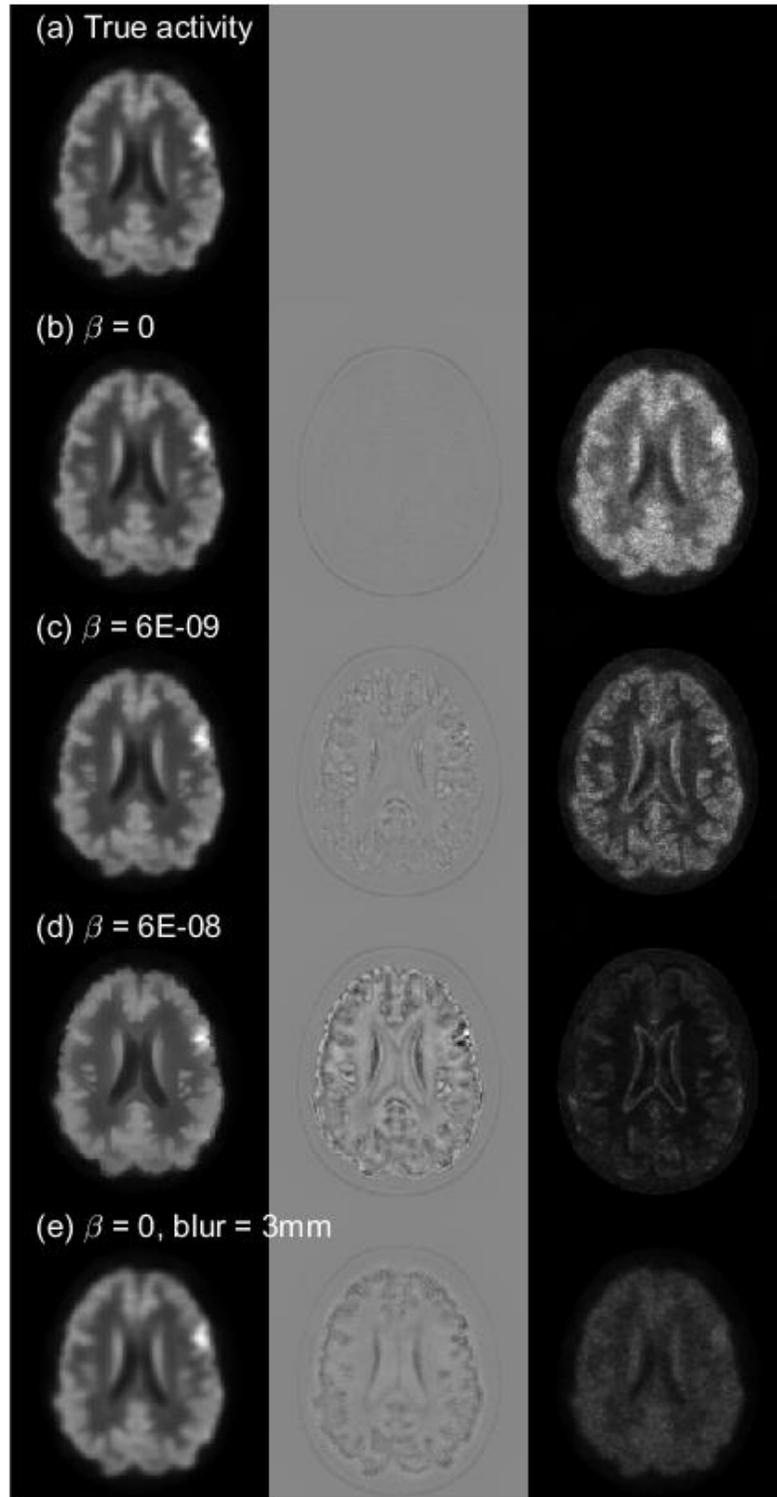


Figure 5.8: Mean, bias, and standard deviation images for simulated noisy  $^{18}\text{F}$ -FDG

Results of regional analysis of bias and variance in the 50-reconstruction simulated dataset confirm the results of Figure 5.8 (Figure 5.9). Interestingly, the MAP solution displays both lower mean bias and mean variance than the post-smoothed MLEM solution for most brain regions and whole-brain. In white matter and lesion regions, the MAP solution displays greater mean bias but lower mean variance than the post-smoothed MLEM solution. This result suggests that the bias introduced by the MAP algorithm may be beneficial to the accuracy of final reconstructed image. From a lesion detection task standpoint, the MAP solution had on average greater bias from the true distribution than the post-smoothed MLEM solution. This result may be partially explained by the lack of a true image space representation of pre-simulated phantom activity. As part of the simulation process the relative phantom activity distribution is blurred to account for a variety of real world spatial effects of positron emission. This blurring process is similar to post-reconstruction blurring and may bias the noise-free true activity in Figure 5.8a toward the post-smoothed MLEM solution. The lesion is also not accounted for by the anatomical prior.

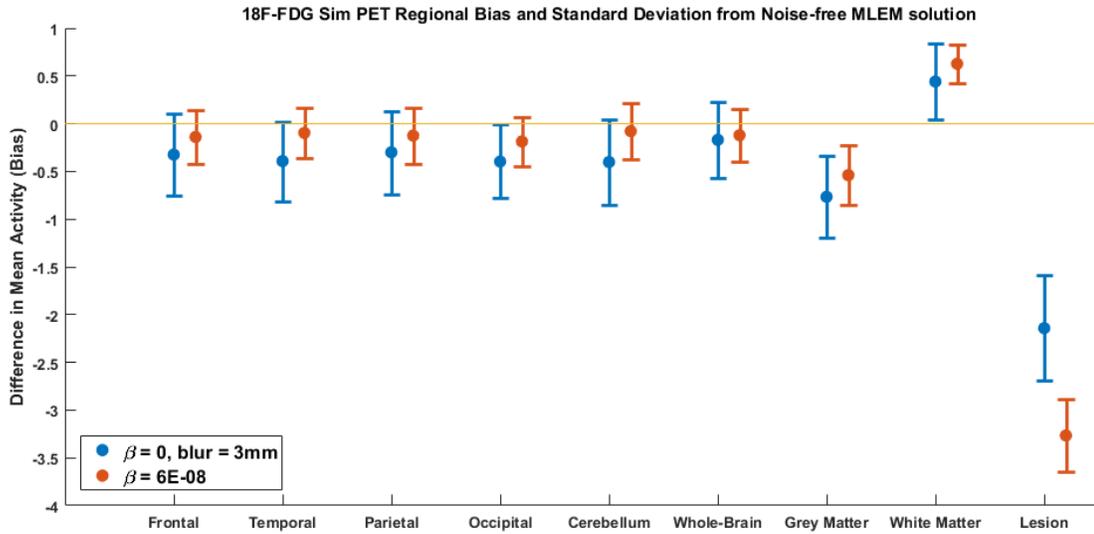


Figure 5.9: Regional mean bias and standard deviation for simulated noisy <sup>18</sup>F-FDG

Comparing lesion activity to the relative phantom activity distribution scaled to mean noise-free reconstruction activity may be a better method of examining edge response. The one caveat to this type of analysis is that because the phantom activity is scaled rather than experiencing the same simulation process as the reconstructions, there is an inherent bias to this measure of true activity distribution.

In the profile plot of mean simulation intensity across the hot lesion, some of the edge differences between the MAP and post-smoothed MLEM reconstructions can be observed (Figure 5.10). The post-smoothed MLEM solution is overall a smoother representation of the phantom activity. Neither the MAP nor post-smoothed MLEM solutions are able to fully represent the true activity distribution, with both smaller peaks and shallower troughs than the true distribution.

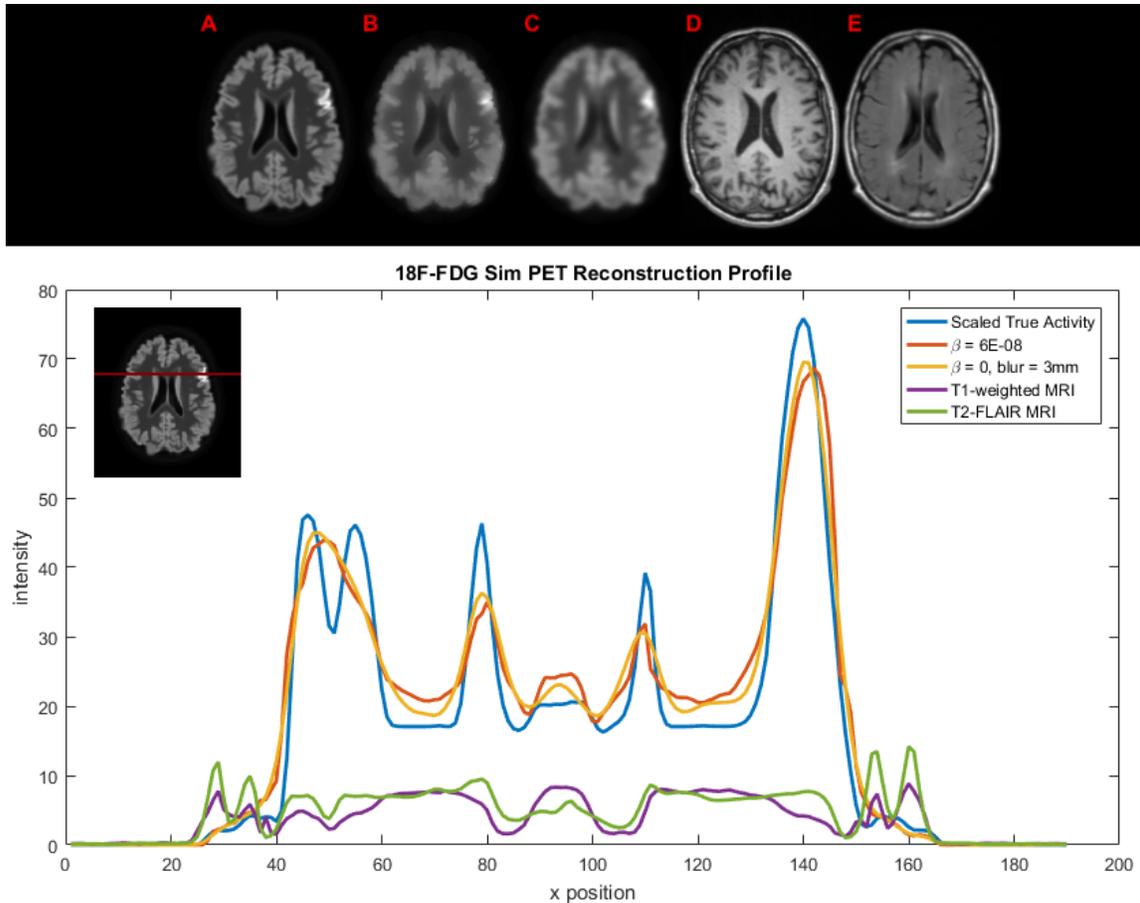


Figure 5.10: Profile analysis of hot lesion in simulated noisy  $^{18}\text{F}$ -FDG along with input images true activity (A),  $\beta=6E-08$  MAP reconstruction (B),  $\beta=0$  MLEM reconstruction with 3 mm FWHM post-reconstruction smoothing (C), T1-weighted MRI (D), and T2-FLAIR MRI (E)

In another profile plot, the effects of the MRI prior are more apparent (Figure 5.11). At x position 70, the change in T1-weighted MRI intensity corresponds with the superior true activity tracking in the MAP solution compared to the post-smoothed MLEM solution. A similar correspondence occurs at x position 130. Figure 5.11 underscores the spatially-varying nature of the MAP smoothing. The degree of smoothing and comparable filter shape change with every voxel location. The contribution of the prior is difficult to capture in a single

dimension, as information from 75 surrounding voxels and multiple prior images is used to determine the degree of smoothing at any given location.

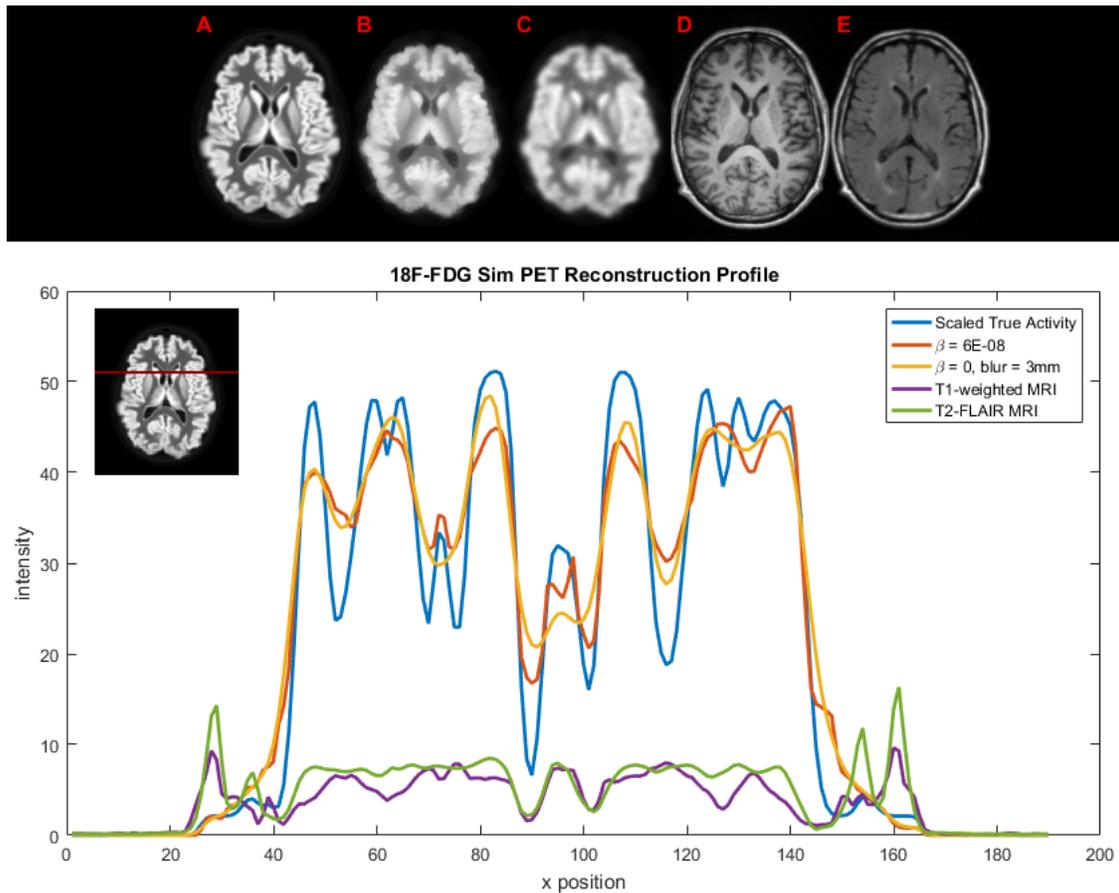


Figure 5.11: Profile analysis of prior effects in simulated noisy  $^{18}\text{F}$ -FDG along with input images true activity (A),  $\beta=6E-08$  MAP reconstruction (B),  $\beta=0$  MLEM reconstruction with 3 mm FWHM post-reconstruction smoothing (C), T1-weighted MRI (D), and T2-FLAIR MRI (E)

In Figure 5.12, simulated mean lesion activity is compared to the activity of the surrounding grey matter (radius = 7 voxels). In both the cold and hot lesion cases, the MAP solution displays less contrast than the post-smoothed MLEM solution. A decrease in contrast

is expected in this situation because the lesion is not represented by the anatomical prior. This is essentially a worst-case scenario for the MRI-guided PET algorithm. Importantly, even in this worst-case scenario the MAP reconstruction still provides substantial lesion contrast, falling only 6.5 percentage points behind the standard MLEM solution.

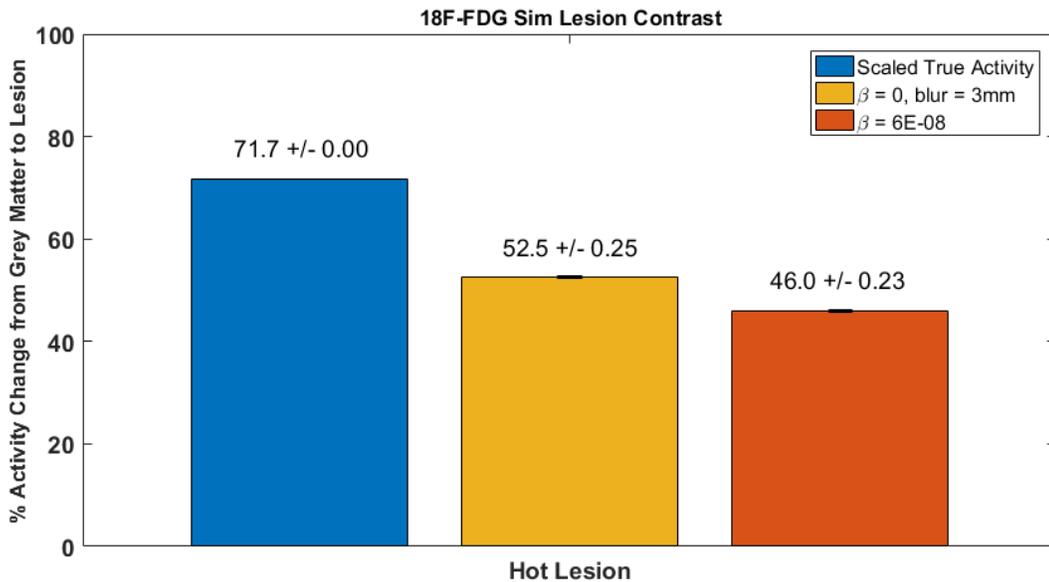


Figure 5.12: Comparison of lesion contrast

## 5.7 Conclusions

A multi-parameter, MR-scale, smoothing prior was successfully implemented for PET reconstruction. Simulation experiments revealed decreased variance in the resulting MRI-guided PET images compared to post-reconstruction smoothed MLEM images at the expense of some increased bias. This result is consistent with previously studied MRI-based PET reconstruction priors and is the most important benefit of the proposed MRI-guided PET

method. Additionally, at comparable smoothing levels, the MAP reconstruction results in better anatomical edge definition compared to MLEM solutions. This result is accentuated by the reconstruction at MR-scale, which reduces the impact of partial volume effects on the reconstructed PET image.

MAP reconstructions were responsive to anatomical input in the form of MRI images, and in cases where the prior MRI images were not consistent with the underlying true radiotracer distribution the resulting PET images exhibited reduced lesion contrast compared to the MLEM solution. This reduction in lesion contrast is the primary drawback of MAP reconstruction methods and is of great significance to the clinical applicability of MAP reconstructions in general. The incorporation of multiple MRI images of varying tissue contrast into the prior model is one way the proposed method attempts to overcome this obstacle. There exist a vast array of MRI imaging sequences, each tuned to provide maximum benefit for a given application. By incorporating many imaging sequences into the image prior, the underlying assumption of anatomical consistency with the injected functional radiotracer is most likely to be upheld.

## CHAPTER 6: COMPRESSED SENSING IN MRI

### 6.1 Overview

Compressed sensing (CS) is an iterative MR image restoration approach used in the reconstruction of undersampled data. CS is based on the principal of image compressibility and was derived from the field of information theory [61]. Image compression is the process of reducing the memory required to digitally store images. Image compression transforms an image into a sparse transform domain that contains all the necessary information for satisfactory image domain recovery in fewer basis function coefficients than pixels in the original image. In a CS optimized sequence and reconstruction, the inherent compressibility of MR images is used as a way to reduce k-space point sampling below the Nyquist-Shannon sampling threshold and accelerate image acquisition [62].

CS requires a compressible desired image, a nonlinear reconstruction approach, and a k-space sampling trajectory which produces noise-like energy leakage in the sparse transform domain [62]. As first framed by Candes et al., random sampling was key to the beneficial sensing property of CS. Subsequent studies have shown that strictly random sampling is not required, only that the sampling scheme produce noise-like artifacts in the transform domain [62-64]. The selection of the sparsity transform and k-space sampling trajectory can be linked by a transform point spread function (TPSF), which measures sidelobe energy leakage in the transform domain. The ideal sparsifying transform compresses image energy content into as few coefficients as possible, and the ideal k-space trajectory incoherently spreads

undersampling artifact energy across many transform basis functions [62]. CS reconstruction is typically framed as a constrained optimization problem, namely:

$$\begin{aligned} & \text{minimize} && \|\Psi m\|_1 \\ & \text{s. t.} && \|\mathcal{F}_u m - y\|_2 < \epsilon \end{aligned} \tag{6.1}$$

Here  $m$  is the reconstructed image,  $\Psi$  is the sparse transform operator,  $\mathcal{F}_u$  is the undersampled Fourier transform, and  $y$  is the measured k-space data. Minimizing the  $\ell_1$  norm (sum of absolute value) of the image in the sparsifying transform domain enforces sparsity by rejecting images which leak energy into a wide number of basis elements. The  $\ell_2$  norm constraint enforces data consistency [62].

The concept of CS was introduced to the MRI community by Candes et al. and Donoho [61, 64]. Lustig et al. made the approach accessible to MR researchers by summarizing CS advantages in specific MR applications and releasing publically available CS reconstruction code [62]. Gamper et al. and Lustig et al. extended CS to dynamic cardiac MRI, applying CS along the temporal dimension [65, 66]. Otazo et al. combined parallel imaging concepts with CS by distributing sparsity constraints across all available coil images [67]. Usman et al. applied CS reconstruction to a radial non-Cartesian k-space acquisition for dynamic cardiac MRI using a 2D navigator for time binning based on respiratory cycle [68]. Feng et al. proved the flexibility of CS by applying CS reconstruction techniques to a variety of contrast-enhanced dynamic MRI imaging schemes. The group combined previously published advances in parallel imaging combined CS with golden-angle radial stack-of-stars acquisition. In contrast to Usman et al., Feng et al. used a self-navigated temporal sorting scheme to group radial spokes based on their temporal imaging order [69]. Feng et al. further refined their sorting

scheme to track respiratory and cardiac motion, with CS restoration applied along each dynamic dimension [70].

CS has been applied to numerous MRI applications, including breast and prostate imaging, angiography, cardiac MRI, and liver MRI to name a few [71]. Overall, there are two primary benefits of compressed sensing in MRI: sequence acceleration and artifact removal. In sequence acceleration applications, k-space is deliberately undersampled to reduce acquisition time. Aliasing artifacts caused by k-space undersampling are then removed via CS. Several studies have reported diagnostic quality images using CS reconstructions at 2x acceleration, with mixed results up to 12.5x acceleration depending on the application [71, 72]. The artifact removal benefit of CS is important for dynamic MRI. Instead of relying on external navigators and low-resolution fast acquisition protocols, dynamic imaging performed with CS reconstruction sorts the time data post-acquisition, then removes undersampling artifacts at each time point using CS techniques.

Chapter 6.2 describes the acquisition and compressed sensing-based reconstruction methods investigated in detail. Briefly, radially acquired k-space data are sorted by k-space spoke into time bins based on respiratory motion. This results in each time frame being radially undersampled by a factor of the number of time frames. This radial undersampling is equivalent to the underdetermined problem of limited-angle tomography. In the image domain, radial undersampling results in variable intensity streaking artifacts that severely impair the utility of the reconstructed images. Compressed sensing imposes a compressibility constraint on the reconstructed image, which makes solutions with features with low compressibility (namely undersampling artifacts) less desirable to the reconstruction framework.

## 6.2 Methods

The iGRASP reconstruction pipeline developed by Feng et al. was used as the basis for our data processing procedure [69]. A flowchart of the CS reconstruction scheme is presented in Figure 6.1.

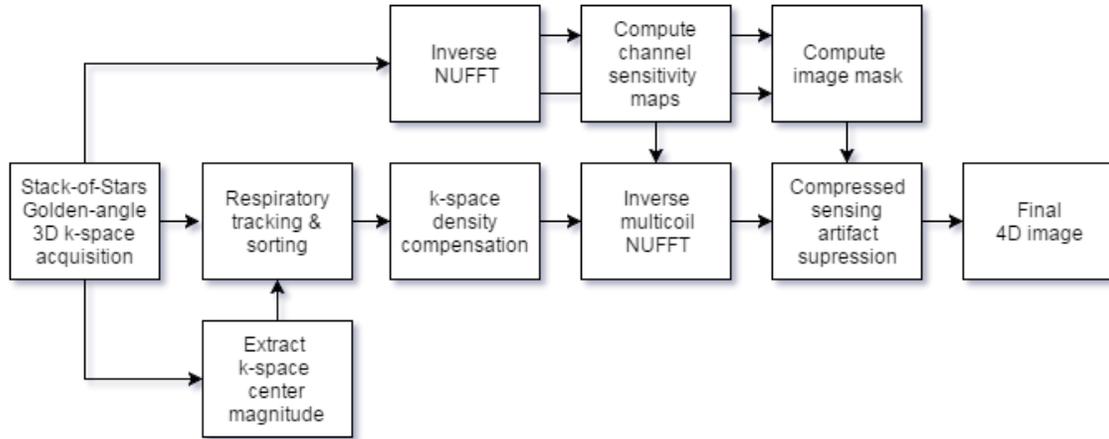


Figure 6.1: Flowchart of CS reconstruction pipeline.

*Data acquisition.*  $T_1$ -weighted MRI data were acquired using a sagittal VIBE sequence with the following imaging parameters: Repetition Time (TR)/Echo Time (TE) = 3.54/1.69 ms, flip angle =  $10^\circ$ , number of partitions = 112, field-of-view (FOV) =  $320 \times 320$  voxels, and voxel size  $1.125 \times 1.125 \times 3$  mm<sup>3</sup>. The k-space trajectory used was a radial stack-of-stars sampling pattern. The stack-of-stars trajectory is Cartesian in the  $k_z$  sampling direction and radial in the  $k_{xy}$  sampling plane. A golden-angle approach to radial spoke ordering was employed as described by Winkelmann et al [73]. The advantage of golden-angle radial sampling is that k-space coverage is maximized for a given number of radial spokes. 2000 radial spokes were acquired in each  $k_{xy}$  plane, producing high-density sampling near the k-

space origin that decreases as distance from center k-space increases. Spokes were acquired with 2x oversampling, meaning 640 data points were acquired per spoke to create a  $320 \times 320$  image. During processing, the first 20 spokes were discarded due to non-steady state conditions.

*Density compensation.* Due to the non-uniform sampling density resulting from the stack-of-stars acquisition, density compensation was required to weight k-space data evenly. Density compensation was calculated based on Voronoi areas as described by Rasche et al. [74]. Density compensation was applied to k-space data as part of the non-uniform fast Fourier transform (NUFFT) operation.

*Respiratory tracking.* The stack-of-stars acquisition is not inherently respiratory gated. Instead, gating and rebinning of acquired data was performed post-acquisition. In a stack-of-stars acquisition, each spoke (minimum  $k_{xy}$  to maximum  $k_{xy}$ ) is acquired for every  $k_z$  location before the next angular increment is selected (via golden angle ratio). There is a small time delay between each angular spoke acquisition, which makes rebinning data by spoke angle a natural sorting scheme for this type of sequence. The additional benefit is that each undersampled time bin has a radial undersampling pattern, which satisfies the sparsity transform noise-like energy-leakage requirement of compressed sensing. Since each time bin is comprised of different angular k-space spokes, the undersampling artifacts produced by each time-bin reconstruction will differ significantly along the temporal axis. The respiratory tracking signal was extracted from the k-space center magnitude of each acquired spoke arranged by order of acquisition. This produces a repetitive oscillating two-dimensional signal with time along the independent axis and magnitude along the dependent axis. The signal was

lowpass filtered to reduce noise, then each spoke was sorted by magnitude into time bins. This approach has the same end goal as the respiratory tracking process described by Feng et al., but is performed using filtering instead of principal component analysis [70].

*Sensitivity Maps.* Channel sensitivity maps represent the contribution of each data acquisition channel to the reconstructed MR image. The SNR of each channel depends heavily on the distance of the physical receiving coils that comprise the channel from the field-of-view as well as imaging subject composition. The goal of computing sensitivity maps is to find the weighted linear combination of channels for each voxel that results in the highest SNR for the image. However, without reference scans it is not possible to assess absolute sensitivity values [75]. Instead, the sensitivity maps must be computed from the individual channel reconstruction images. One caveat to the computation of sensitivity maps is that because of future applications of the CS reconstruction, the data must remain complex. This is noted because MR imaging in general tends to rely on magnitude images rather than phase or complex information. Sensitivity maps were computed based on a modified version of the procedure described Pruessmann et al. for computing sensitivity maps [75]. This includes normalization of channel images by the sum-of-squares combination image, then 2<sup>nd</sup>-order polynomial smoothing of the noisy resulting images using a square moving window with five pixel radius.

*Image Mask.* A subject mask was used to remove undersampling streak artifacts outside the subject volume. The mask was computed on a slice-by-slice basis using the unbinned combined channel reconstruction image. Initial thresholding was performed using Otsu's

Method followed by morphological dilation, fill, and erosion operations to remove stray masked voxels [76].

*Non-uniform Fast Fourier Transform (NUFFT)*. The Fourier transform is typically used for conversion from uniform Cartesian k-space sampling to image space. However, the built-in implementations of the Fourier transform in MATLAB are unable to handle non-uniform input. Instead, conversion from k-space to image space and vice versa was performed using the NUFFT toolbox for MATLAB created by Jeffrey Fessler (available online at <http://web.eecs.umich.edu/~fessler/irt/irt/nufft/>) [77]. This implementation has optimized interpolation methods to handle non-uniform input and has been used previously in other compressed sensing studies [69, 70].

*Compressed Sensing*. Compressed sensing regularization was implemented across both spatial and temporal dimensions. The purpose of the spatial penalty is to maintain slice intensity regularization, while the temporal penalty provides smooth transition between respiratory time frames. The conjugate gradient method was used to solve the constrained optimization problem presented in Equation 6.1. A modified version of the implementation written by Michael Lustig for compressed sensing applications was used to perform the reconstructions [62]. The main modification was to change the primary working domain from transform space to image space, which allows each  $\ell_1$  term to be computed identically. The number of possible  $\ell_1$  terms was also modified to be limitless. Thus, the constrained optimization problem solved was actually a modification of Equation 6.1.

$$\begin{aligned} & \text{minimize} && \sum_j^N \|\Psi_j m\|_1 \\ & \text{s. t.} && \|\mathcal{F}_u m - y\|_2 < \epsilon \end{aligned} \tag{6.2}$$

In Equation 6.2, the number of  $\ell_1$ -norm terms is equal to the number of sparsifying transform domains ( $\Psi_j$ ) used to enforce solution sparsity. A variety of transforms, including the discrete cosine transform, wavelet transforms, Karhunen-Loève (KL) transform, and total variation (TV) transform were investigated. Solutions were evaluated by degree of artifact removal and temporal smoothing using region-of-interest analysis of coefficient of variation and diaphragm movement across time.

*Comparison.* The results of compressed sensing reconstructions were compared against two additional methods of artifact removal/smoothing. The first method, Gaussian smoothing, is the post-reconstruction application of a three dimensional (two spatial plus one temporal dimension) non-uniform Gaussian smoothing filter to NUFFT-reconstructed time-binned images. The second method, iterative KL thresholding, uses the KL transform to distribute streaking artifact energy across a wide range of basis coefficients, thereby making artifact dominated coefficients less attractive to selection via intensity thresholding. Each iteration, all coefficients of intensity greater than 0.6 times the maximum intensity coefficient are selected. Unlike simple filtering in the KL domain, there is feedback introduced to the coefficient selection process by calculating and removing the undersampling interference produced by the selected coefficients. Undersampling interference is calculated by transforming the selected KL coefficients into the undersampled k-space domain and back again. This effectively recreates the undersampling artifact resulting from the selected transform coefficients in the KL domain. Removing this interference reduces the overall artifact energy in the KL domain, allowing previously hidden non-artifact coefficients to be identified by thresholding. The Hough transform was used to measure artifact content and provide the stopping criteria (25%

reduction). This process is a greedy pursuit form of compressed sensing in contrast to the convex optimization form of compressed sensing solved using the conjugate gradient algorithm.

### **6.3 Results**

The result of rebinning 1980 acquired spokes of k-space data into five time frames of 396 frames is that each time frame contains only partially sampled k-space. This radial undersampling in the Fourier (k-space) domain results in streaking artifacts in the image domain. These streak artifacts corrupt the underlying images of respiratory motion, limiting their diagnostic or functional utility. An example of the streaking artifacts caused by radial undersampling of k-space can be seen in Figures 6.2 and 6.3. In Figure 6.2, an unbinned full-angle reconstruction (Figure 6.2a) is shown beside a rebinned undersampled reconstruction at end expiration (Figure 6.2b). The streak artifacts make it difficult to discern spatial detail from noise/artifact. Regardless, the full-angle reconstruction appears blurrier than the rebinned reconstruction in areas with heavy respiratory movement, especially near the top and bottom of the live.

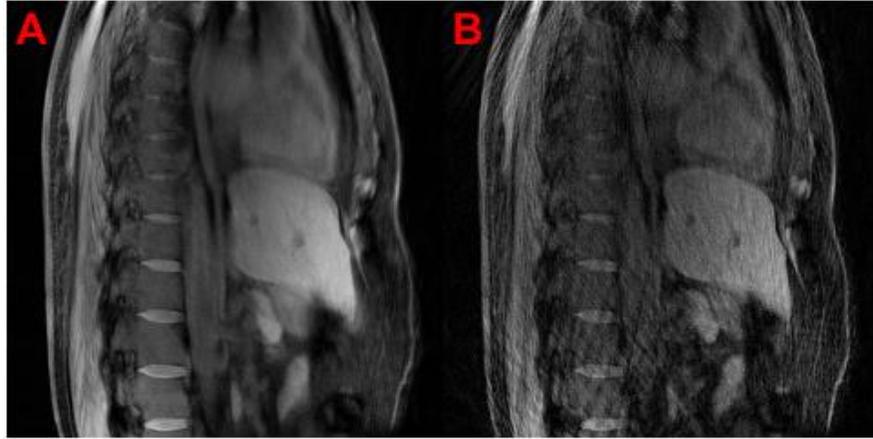


Figure 6.2: Sagittal full-angle reconstruction (A) and rebinned reconstruction at end expiration (frame = 4) (B)

Little difference in regional movement can be observed between the images in Figure 6.2. This is due to more rebinned spokes taking place during expiration than inspiration, causing the full-angle reconstruction to appear more closely related to expiration frames than inspiration frames. In Figure 6.3, the full-angle reconstruction is paired with a rebinned undersampled reconstruction at full inhalation.

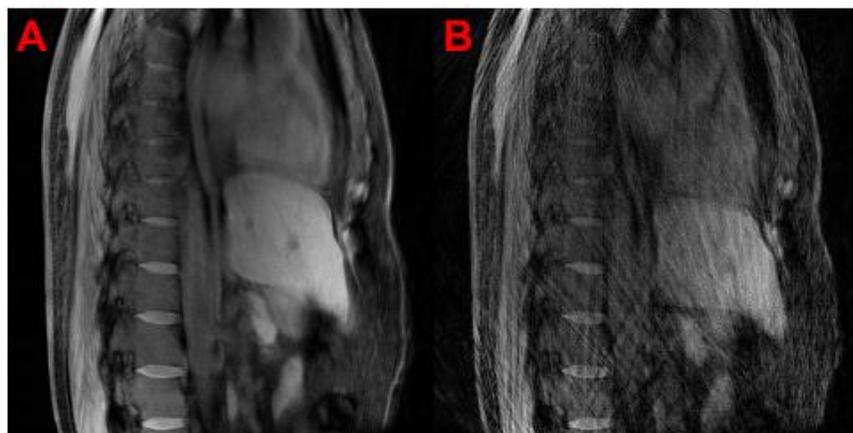


Figure 6.3: Sagittal full-angle reconstruction (A) and rebinned reconstruction at full inspiration (frame = 1) (B)

In Figure 6.3b, the effect of the undersampling artifacts is more apparent, as well as significant blurring caused by fewer available spokes to match perfectly with full inhalation.

The images in Figures 6.2 and 6.3 are uncorrected NUFFT reconstructions (NF). CS reconstructions were performed using a range of  $\ell_1$ -norm weighting parameters and sparse transform domains in an effort to improve upon these NUFFT reconstructions of rebinned data. Of the sparsity transforms investigated, image quality was found to be most consistent using a two-dimensional discrete cosine transform for spatial regularization and a total variation transform for temporal regularization (CSst). Using only a total variation temporal transform for  $\ell_1$  regularization was also found to perform well (CSt). These compressed sensing-based reconstruction solutions were optimized for parameter weighting and compared against the iterative KL transform thresholding reconstruction (KL) and post-NUFFT reconstruction Gaussian blurred solution (NFblur).

Results for each reconstruction type are displayed in Figure 6.4. Frame number one corresponds to full inhalation and frame number four corresponds to end expiration. Reconstruction differences are most apparent in the full inspiration frame (Figure 6.5).

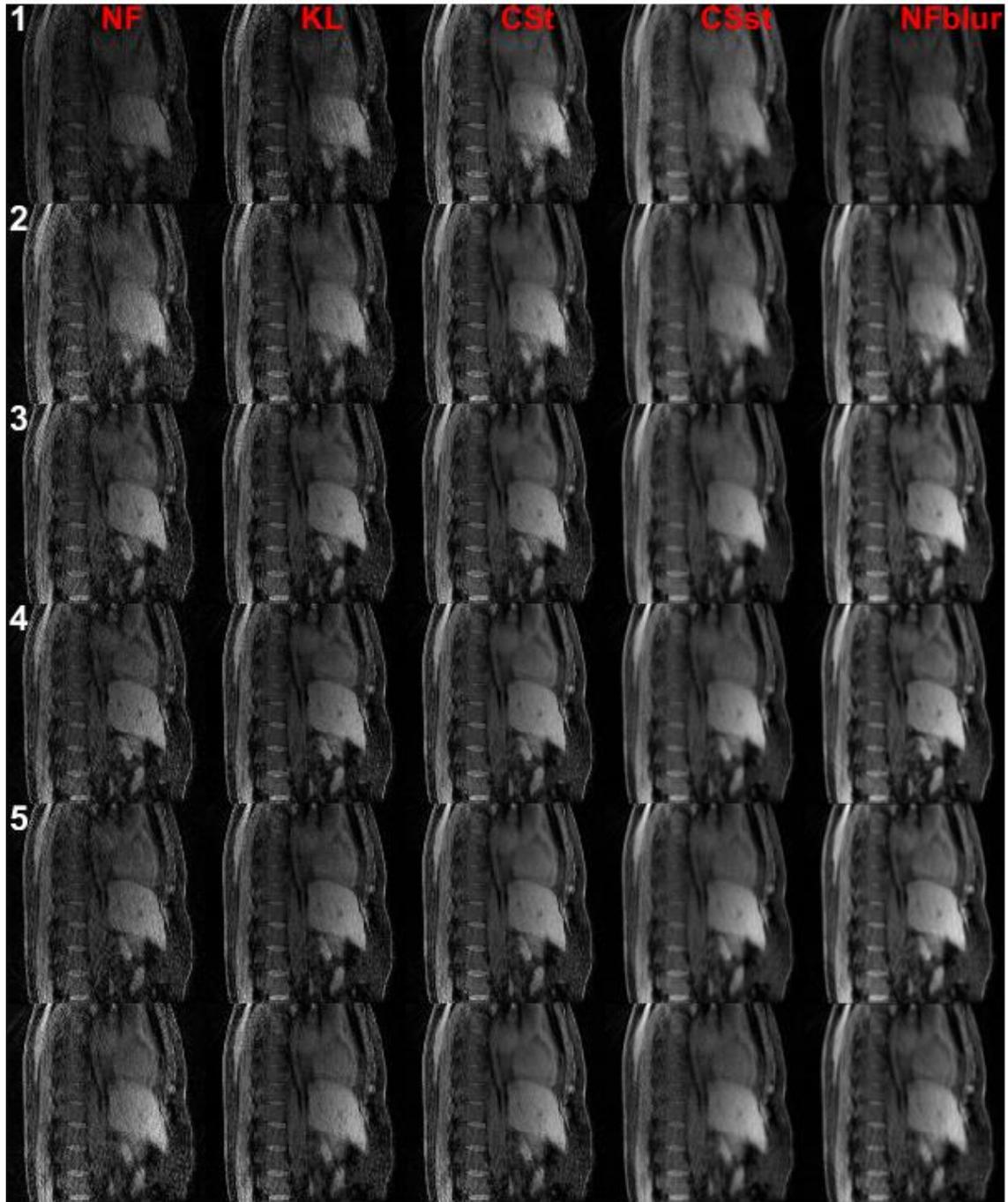


Figure 6.4: Montage of reconstruction results from NUFFT reconstruction (NF), Karhunen-Loève iterative thresholding (KL), temporal CS (CSt), spatial and temporal CS (CSst), and post-NUFFT reconstruction Gaussian blurring (NFblur).

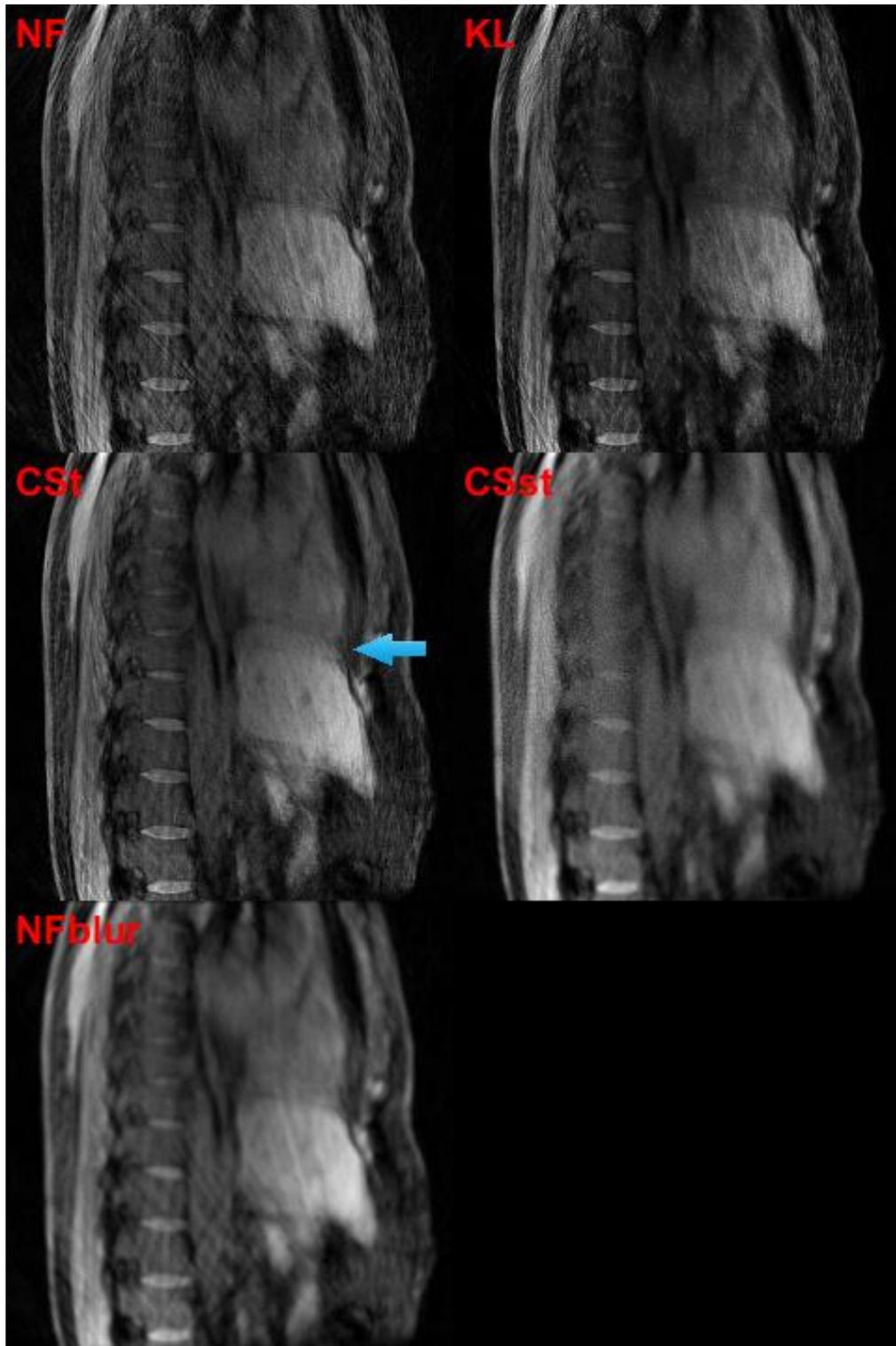


Figure 6.5: Full inspiration frame reconstructions (frame one) with ghosting artifact (arrow).

The streak artifact varies between time frames because the rebinning procedure causes the k-space sampling pattern to vary between time frames. All of the reconstruction methods in Figure 6.5 exploit this artifact temporal inconsistency to reduce the magnitude of streaking by incorporating information from other time frames. The methods reduce artifact magnitude compared to the NUFFT reconstruction (NF), but a cost of the artifact reduction is ghosting from other time frames. This ghosting artifact can be seen in Figure 6.5 near the top of the liver (arrow). The success of any smoothing method must therefore be measured by its ability to not only remove streak artifacts but also prevent frame ghosting.

The iterative KL method did the best job preserving liver movement across time frames compared to the NUFFT reconstruction (Figure 6.5). The CSt and CSst methods produced images with strong ghosting artifacts near the liver at full inspiration (frame 1). The NFblur method resulted in a moderate amount of temporal ghosting.

The second measure of method success is its ability to remove streak artifact. This was measured by coefficient of variation (CV) in a ROI defined in the liver. An interior liver ROI was chosen because its intensity was expected to be relatively uniform both spatially and temporally. The defined ROI is shown in Figure 6.8.

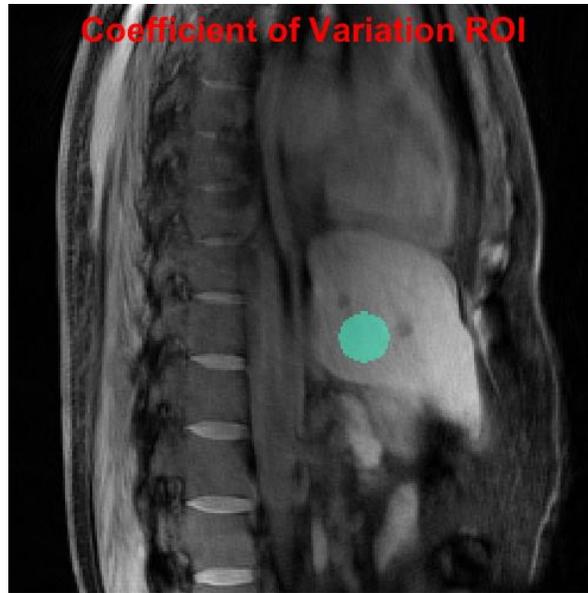


Figure 6.6: Artifact smoothing region of interest (blue circle).

The CV for the full-angle reconstruction was 0.0860, nearly equal to that of the Gaussian blur and spatiotemporal CS methods (Table 6.1). The temporal CS method had the next lowest mean CV. The iterative KL method performed poorly in this test, with a mean CV close to that of the NUFFT reconstruction.

Table 6.1: ROI coefficient of variation for each reconstruction method and frame.

Frame	Reconstruction Type				
	NF	KL	CSt	CSst	NFblur
1	0.143	0.169	0.111	0.076	0.086
2	0.127	0.118	0.106	0.071	0.085
3	0.110	0.104	0.094	0.087	0.077
4	0.116	0.108	0.091	0.091	0.081
5	0.129	0.110	0.092	0.090	0.089
6	0.134	0.122	0.098	0.090	0.092
MEAN	0.127	0.122	0.099	0.084	0.085
SD	0.011	0.022	0.007	0.008	0.005

#### 6.4 Discussion and Conclusions

The spatiotemporal CS reconstruction method provided strong artifact suppression, but did not perform well in movement preservation. This could be a result of non-optimal  $\ell_1$ -term weight tuning, but if so, it highlights the difficulty of tuning unpredictable multi-parameter reconstruction algorithms. The temporal CS method, which has only a single  $\ell_1$ -term to tune, is likely a better strategy for future investigations into compressed sensing. It also has the added benefit of being formulated as the more common  $\ell_1$ - $\ell_2$  optimization problem in Equation 6.1, for which numerous iterative solutions exist. A reversible transform may perform better than the total variation transform, or at least speed reconstruction time by providing a more optimized path through solution space. Both methods with spatial constraints, the spatiotemporal CS method and Gaussian blur method, produced overly smooth results. This

suggests that the most effective constraints for compressed sensing are those applied across the temporal dimension. Spatial constraints appear unable to adequately separate artifact from image content, resulting in solutions in which both artifact and underlying image content is made undesirably uniform.

While not a focus of this work, the processing time required for CS reconstruction is a major impediment to clinical applicability. Depending on the number of channels, time frames, sparse transform characteristics, and iterations, it can take a MATLAB implementation up to an hour on a modern desktop computer to reconstruct a single CS reconstruction slice. This computational burden was the primary reason for exploring an iterative thresholding approach using the KL transform. The iterative KL transform approach performed poorly in the CV smoothness test because it produces noisy results compared to the CS and Gaussian blurring methods, but visually it does a good job of suppressing streak artifacts. It is also a fast transform to compute, and the iterative artifact removal process requires only a of couple minutes. Another advantage is that, compared to the other methods in Figure 6.5, the KL method results in the least temporal ghosting of other frames. The downside to the method is that it very quickly erases frame variation if not tuned properly, producing a static image over all frames. Overall, the iterative KL method had the best temporal preservation of the investigated methods and good artifact removal. This, combined with its fast execution time, makes the iterative KL thresholding method the most successful compressed sensing method investigated.

As a CS constraint, the KL transform is an attractive choice because it is reversible, orthogonal, and fast to compute. Among orthogonal transforms, it is the most compressible.

The major downside to the KL transform is that it is produced from the image being operated upon. Thus it requires a preliminary reconstruction to form the transform and its applicability is tailored specifically to the image(s) from which it was derived. This may result in the KL transform being more difficult to universally tune than constraint transforms unrelated to image content, impacting its clinical applicability.

## CHAPTER 7: SUMMARY AND CONCLUSION

### 7.1 Summary

Though still in its infancy, hybrid PET/MRI imaging could be a game changer for the medical imaging community. Previously developed hybrid technologies, such as PET/CT, are limited in their integration potential. PET/MRI, on the other hand, provides simultaneously acquired, aligned imaging data from both functional and anatomical sources.

The goal of the work presented in this manuscript was to explore potential avenues for integrating MRI information with PET imaging and to explore computational methods for improving image quality in both PET and MRI. Chapter 3 presents the results of PET reconstruction acceleration using parallel computing methods and graphical processing units. Using these technologies, the PET reconstruction process was accelerated 60-fold compared to a serial implementation. In terms of hybrid integration, the new accelerated algorithm allows MRI information to be incorporated into the PET reconstruction process without comprising clinical relevance due to prolonged processing.

More broadly, the translation of the OSEM algorithm from serial to parallel architecture has made clear the applicability of parallel computing to the field of medical imaging. Parallel computing and GPGPU have brought significant changes to fields such as bioinformatics, computer vision, and statistical modeling. As the medical imaging community continues to adopt parallel computing techniques, significant opportunity exists for advancement in hybrid imaging. The amount of data output by modern medical imaging systems is immense and likely underutilized. This is almost certainly the case in PET imaging, which reduces image

processing complexity by artificially compressing projection space data prior to reconstruction. Parallel computing eliminates the need to throw away or compromise acquired data due to computational constraints.

However, one of the problems with combining technologies as different as PET and MRI is the knowledge barrier to entry for researchers previously focused on a single imaging modality. The E-phantom presented in Chapter 4 provides MRI researchers with a tool for quickly and cost effectively exploring PET/MRI imaging opportunities. The validity of the phantom and simulation procedure was tested on MRI-based PET attenuation correction techniques then utilized in the MRI-guided PET reconstruction research described in Chapter 5.

Besides providing a stepping stone for single modality researchers interested in hybrid PET/MRI imaging, the E-phantom platform equips researchers with something that real-world medical imaging can rarely supply, a known ground truth for algorithm development. This is especially important for development of partial volume effect mitigation strategies and small lesion detection, where algorithm effects are best evaluated at sub-clinical voxel sizes.

While PET provides unique diagnostic imaging utility, it suffers from a host of physical and reconstruction effects that degrade its image quality. Chapter 5 presents a method for making PET less susceptible to partial volume effects through incorporation of a multi-image MRI prior into the OSEM algorithm and MRI-scale PET reconstruction. Using the MRI prior and MAP reconstruction criterion, noise was successfully reduced in a spatially-varying manner in reconstructed PET images, and the anatomical guidance of the prior was

demonstrated. Results of the MAP reconstruction method were consistent with the results of other researchers.

The MAP reconstruction method presented in Chapter 5 is essentially a modern update of the classical Bayesian estimation criterion for anatomically-guided PET. The key difference is the recognition that in a data-rich environment such as medical imaging, the anatomical guidance need not come from a single scan nor be purely anatomical in nature. Unbridled from processing constraints by parallel computing, the question of prior incorporation into the reconstruction process is no longer limited to which hyperparameter to tune for best results but which data provide useful features. This presupposes a holistic, data-centric approach to medical imaging in which a Bayesian estimation criterion may prove outdated. Data-centric approaches to problem solving increasingly rely on machine learning algorithms to organically derive solutions to classification, regression, and detection problems. These algorithms are modeled after the connection patterns of neurons and rely heavily on truthed datasets to train neural networks to produce desired output. The E-phantom platform could prove useful as a training tool for machine learning algorithms since it offers researchers the type of varied and truthed training data required of deep learning methods.

Chapter 6 presents research into compressed sensing, a relatively new concept in MR imaging being used to accelerate acquisitions and remove reconstruction artifacts. The research focused primarily on the artifact removal aspect of compressed sensing, which has potential application in PET respiratory motion correction. A variety of sparse transforms were investigated as well as the spatiotemporal smoothing capabilities of compressed sensing. Spatial transforms were determined to add little value to the sparse domain representation of

the restoration volume. In addition, the standard  $\ell_1$ - $\ell_2$  constraint formulation and conjugate gradient optimization strategy was found to be several orders of magnitude slower and produce less desirable results than a KL domain sparse representation combined with greedy optimization strategy.

In some sense, applying sparsity as a constraint is somewhat of a misnomer for the actual artifact removal process. The key aspect of a successful sparsifying transform is not necessarily that it compresses a signal well, but instead that it more broadly redistributes artifact energy across basis components than the underlying signal. A result of this process is that sparsity (and more specifically an  $\ell_1$  constraint) can be used to attenuate or remove transform components dominated by artifact. Viewed from this perspective, compressed sensing is a feature extraction technique applied in a transform domain, which may be more intuitive to some researchers.

## 7.2 Future Work

Future work in MRI-guided PET reconstruction could focus on improving the multi-parameter prior. More complex methods of neighborhood identification could improve the anatomical guidance of the prior. There is also a need to investigate the types of MR images that correspond best with PET imaging. It is likely that the form of the MRI prior should change based on the radiotracer used for PET imaging.

For compressed sensing, it would be beneficial to identify more suitable sparse transform domains compatible with temporal undersampling artifact removal. Machine learning algorithms may be prove useful for building unique and compressible transforms.

### **7.3 Conclusion**

The work presented in this manuscript is an in depth investigation of methods for integrating MRI with PET imaging. The industry standard OSEM reconstruction algorithm was adapted for parallel computing, resulting in substantially accelerated PET image reconstruction. A digital PET phantom platform was developed to provide realistic, varied, and configurable PET acquisition datasets with known ground truth for PET/MRI researchers. A new method for integrating MRI data into PET reconstructions was produced, which combines the accelerated PET reconstruction algorithm with MAP criterion. The method integrates MRI information from multiple MRI images into the PET reconstruction process, producing MRI-guided PET reconstructions at native MRI-scale. Lastly, compressed sensing was investigated as a method of undersampling artifact removal in MRI for the purpose of improving PET respiratory motion compensation. A new method of compressed sensing was developed which utilizes the KL transform as a sparsifying operator along with a greedy optimization strategy.

## REFERENCES

1. Bai, B., Q. Li, and R.M. Leahy, *Magnetic resonance-guided positron emission tomography image reconstruction*. Semin Nucl Med, 2013. **43**(1): p. 30-44.
2. Zasadny, K.R. and R.L. Wahl, *Standardized uptake values of normal tissues at PET with 2-[fluorine-18]-fluoro-2-deoxy-D-glucose: variations with body weight and a method for correction*. Radiology, 1993. **189**(3): p. 847-50.
3. Meltzer, C.C., et al., *Correction of PET data for partial volume effects in human cerebral cortex by MR imaging*. J Comput Assist Tomogr, 1990. **14**(4): p. 561-70.
4. Pontecorvo, M.J. and M.A. Mintun, *PET amyloid imaging as a tool for early diagnosis and identifying patients at risk for progression to Alzheimer's disease*. Alzheimers Res Ther, 2011. **3**(2): p. 11.
5. Wernick, M.N. and J.N. Aarsvold, *Chapter 2 - Introduction to Emission Tomography*, in *Emission Tomography*. 2004, Academic Press: San Diego. p. 11-23.
6. Doraiswamy, P.M., et al., *Amyloid-beta assessed by florbetapir F 18 PET and 18-month cognitive decline: a multicenter study*. Neurology, 2012. **79**(16): p. 1636-44.
7. Kinahan, P.E., M. Defrise, and R. Clackdoyle, *Chapter 20 - Analytic Image Reconstruction Methods A2 - Wernick, Miles N*, in *Emission Tomography*, J.N. Aarsvold, Editor. 2004, Academic Press: San Diego. p. 421-442.
8. Lalush, D.S. and M.N. Wernick, *Chapter 21 - Iterative Image Reconstruction*, in *Emission Tomography*. 2004, Academic Press: San Diego. p. 443-472.
9. Hudson, H.M. and R.S. Larkin, *Accelerated image reconstruction using ordered subsets of projection data*. IEEE Trans Med Imaging, 1994. **13**(4): p. 601-9.
10. Wilson, D.W., B.M. Tsui, and H.H. Barrett, *Noise properties of the EM algorithm: II. Monte Carlo simulations*. Phys Med Biol, 1994. **39**(5): p. 847-71.
11. Prince, J.L. and J.M. Links, *Medical Imaging Signals and Systems*. 1st ed. 2006, Upper Saddle River, NJ: Pearson.
12. Levin, C., *Chapter 4 - Basic Physics of Radionuclide Imaging A2 - Wernick, Miles N*, in *Emission Tomography*, J.N. Aarsvold, Editor. 2004, Academic Press: San Diego. p. 53-88.
13. Levin, C.S. 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, 1999. **44**(3): p. 781-99.

14. Haacke, E.M., et al., *Magnetic Resonance Imaging Physical Principles and Sequence Design*. 1999, New York, NY: John Wiley & Sons.
15. Shannon, C.E., *Communication in the Presence of Noise*. Proceedings of the IRE, 1949. **37**(1): p. 10-21.
16. Yankeelov, T.E., et al., *Simultaneous PET-MRI in oncology: a solution looking for a problem?* Magn Reson Imaging, 2012. **30**(9): p. 1342-56.
17. Bailey, D.L., et al., *Combined PET/MRI: from Status Quo to Status Go. Summary Report of the Fifth International Workshop on PET/MR Imaging; February 15-19, 2016; Tubingen, Germany*. Mol Imaging Biol, 2016. **18**(5): p. 637-50.
18. Besson, F.L., V. Lebon, and E. Durand, *What are we expecting from PET/MRI?* Medecine Nucleaire-Imagerie Fonctionnelle Et Metabolique, 2016. **40**(1): p. 31-40.
19. Zubal, I.G., et al., *Computerized three-dimensional segmented human anatomy*. Med Phys, 1994. **21**(2): p. 299-302.
20. Hoffman, E.J., et al., *3-D phantom to simulate cerebral blood flow and metabolic images for PET*. IEEE Transactions on Nuclear Science, 1990. **37**(2): p. 616-620.
21. Segars, W.P., et al., *4D XCAT phantom for multimodality imaging research*. Med Phys, 2010. **37**(9): p. 4902-15.
22. Harrison, R.L., et al. *A digital reference object for the 3D Hoffman brain phantom for characterization of PET neuroimaging quality*. in *2013 IEEE Nuclear Science Symposium and Medical Imaging Conference (2013 NSS/MIC)*. 2013.
23. Ljungberg, M., *Chapter 25 - Simulation Techniques and Phantoms A2 - Wernick, Miles N*, in *Emission Tomography*, J.N. Aarsvold, Editor. 2004, Academic Press: San Diego. p. 551-563.
24. Fonov, V.S., et al., *Unbiased nonlinear average age-appropriate brain templates from birth to adulthood*. NeuroImage, 2009. **47**, **Supplement 1**: p. S102.
25. Greve, D.N. and B. Fischl, *Accurate and robust brain image alignment using boundary-based registration*. NeuroImage, 2009. **48**(1): p. 63-72.
26. Jenkinson, M., et al., *Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images*. NeuroImage, 2002. **17**(2): p. 825-841.
27. Jenkinson, M. and S. Smith, *A global optimisation method for robust affine registration of brain images*. Medical Image Analysis. **5**(2): p. 143-156.

28. Avants, B.B., et al., *A reproducible evaluation of ANTs similarity metric performance in brain image registration*. NeuroImage, 2011. **54**(3): p. 2033-2044.
29. Avants, B.B., et al., *A Unified Image Registration Framework for ITK*, in *Biomedical Image Registration: 5th International Workshop, WBIR 2012, Nashville, TN, USA, July 7-8, 2012. Proceedings*, B.M. Dawant, et al., Editors. 2012, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 266-275.
30. Yoo, T.S., et al., *Engineering and algorithm design for an image processing Api: a technical report on ITK--the Insight Toolkit*. Stud Health Technol Inform, 2002. **85**: p. 586-92.
31. Carney, J.P., et al., *Method for transforming CT images for attenuation correction in PET/CT imaging*. Med Phys, 2006. **33**(4): p. 976-83.
32. Juttukonda, M.R., et al., *MR-based attenuation correction for PET/MRI neurological studies with continuous-valued attenuation coefficients for bone through a conversion from R2\* to CT-Hounsfield units*. Neuroimage, 2015. **112**: p. 160-8.
33. Soret, M., S.L. Bacharach, and I. Buvat, *Partial-volume effect in PET tumor imaging*. J Nucl Med, 2007. **48**(6): p. 932-45.
34. Soret, M., et al., *Biases affecting the measurements of tumor-to-background activity ratio in PET*. Ieee Transactions on Nuclear Science, 2002. **49**(5): p. 2112-2118.
35. Hallett, W.A., et al., *Effect of corrections for blood glucose and body size on [18F]FDG PET standardised uptake values in lung cancer*. Eur J Nucl Med, 2001. **28**(7): p. 919-22.
36. Vesselle, H., et al., *Lung cancer proliferation correlates with [F-18]fluorodeoxyglucose uptake by positron emission tomography*. Clin Cancer Res, 2000. **6**(10): p. 3837-44.
37. Hoffman, E.J., S.C. Huang, and M.E. Phelps, *Quantitation in positron emission computed tomography: 1. Effect of object size*. J Comput Assist Tomogr, 1979. **3**(3): p. 299-308.
38. Kessler, R.M., J.R. Ellis, Jr., and M. Eden, *Analysis of emission tomographic scan data: limitations imposed by resolution and background*. J Comput Assist Tomogr, 1984. **8**(3): p. 514-22.
39. Zito, F., et al., *Single-photon emission tomographic quantification in spherical objects: Effects of object size and background*. Eur J Nucl Med, 1996. **23**(3): p. 263-271.

40. Rousset, O.G., Y. Ma, and A.C. Evans, *Correction for partial volume effects in PET: principle and validation*. J Nucl Med, 1998. **39**(5): p. 904-11.
41. Tohka, J. and A. Reilhac, *Deconvolution-based partial volume correction in Raclopride-PET and Monte Carlo comparison to MR-based method*. Neuroimage, 2008. **39**(4): p. 1570-84.
42. Chen, C.H., et al., *Simultaneous recovery of size and radioactivity concentration of small spheroids with PET data*. J Nucl Med, 1999. **40**(1): p. 118-30.
43. Baete, K., et al., *Evaluation of anatomy based reconstruction for partial volume correction in brain FDG-PET*. Neuroimage, 2004. **23**(1): p. 305-17.
44. Lalush, D.S., *MR-Derived Improvements in PET Imaging*. Magnetic Resonance Imaging Clinics of North America, 2017. **25**(2).
45. Gindi, G., et al., *BAYESIAN RECONSTRUCTION OF FUNCTIONAL IMAGES USING ANATOMICAL INFORMATION AS PRIORS*. Ieee Transactions on Medical Imaging, 1993. **12**(4): p. 670-680.
46. Geman, S. and D. Geman, *Stochastic relaxation, gibbs distributions, and the bayesian restoration of images*. IEEE Trans Pattern Anal Mach Intell, 1984. **6**(6): p. 721-41.
47. Ouyang, X., et al., *INCORPORATION OF CORRELATED STRUCTURAL IMAGES IN PET IMAGE-RECONSTRUCTION*. Ieee Transactions on Medical Imaging, 1994. **13**(4): p. 627-640.
48. Chiao, P.C., et al., *MODEL-BASED ESTIMATION WITH BOUNDARY SIDE INFORMATION OR BOUNDARY REGULARIZATION*. Ieee Transactions on Medical Imaging, 1994. **13**(2): p. 227-234.
49. Ardekani, B.A., et al., *Minimum cross-entropy reconstruction of PET images using prior anatomical information*. Physics in Medicine and Biology, 1996. **41**(11): p. 2497-2517.
50. Bowsher, J.E., et al. *Utilizing MRI information to estimate F18-FDG distributions in rat flank tumors*. in *Nuclear Science Symposium/Medical Imaging Conference*. 2004. Rome, ITALY: Ieee.
51. He, K.M., J. Sun, and X.O. Tang, *Guided Image Filtering*. Ieee Transactions on Pattern Analysis and Machine Intelligence, 2013. **35**(6): p. 1397-1409.
52. Yan, J.H., J.C.S. Lim, and D.W. Townsend, *MRI-guided brain PET image filtering and partial volume correction*. Physics in Medicine and Biology, 2015. **60**(3): p. 961-976.

53. Nuyts, J., et al., *Comparison between MAP and postprocessed ML for image reconstruction in emission tomography when anatomical knowledge is available*. Ieee Transactions on Medical Imaging, 2005. **24**(5): p. 667-675.
54. Hutton, B.F., et al., *What approach to brain partial volume correction is best for PET/MRI?* Nuclear Instruments & Methods in Physics Research Section a- Accelerators Spectrometers Detectors and Associated Equipment, 2013. **702**: p. 29-33.
55. Videen, T.O., et al., *Regional correction of positron emission tomography data for the effects of cerebral atrophy*. J Cereb Blood Flow Metab, 1988. **8**(5): p. 662-70.
56. Rousset, O.G., et al., *CORRECTION FOR PARTIAL VOLUME EFFECTS IN PET USING MRI-BASED 3D SIMULATIONS OF INDIVIDUAL HUMAN BRAIN METABOLISM*, in *Quantification of Brain Function: Tracer Kinetics and Image Analysis in Brain Pet*, K. Uemura, et al., Editors. 1993. p. 113-125.
57. Strul, D. and B. Bendriem, *Robustness of anatomically guided pixel-by-pixel algorithms for partial volume effect correction in positron emission tomography*. J Cereb Blood Flow Metab, 1999. **19**(5): p. 547-59.
58. Frouin, V., et al., *Correction of partial-volume effect for PET striatal imaging: fast implementation and study of robustness*. J Nucl Med, 2002. **43**(12): p. 1715-26.
59. Gutierrez, D., et al., *Anatomically guided voxel-based partial volume effect correction in brain PET: impact of MRI segmentation*. Comput Med Imaging Graph, 2012. **36**(8): p. 610-9.
60. Besag, J., *On the Statistical Analysis of Dirty Pictures*. Journal of the Royal Statistical Society. Series B (Methodological), 1986. **48**(3): p. 259-302.
61. Donoho, D.L., *Compressed sensing*. Ieee Transactions on Information Theory, 2006. **52**(4): p. 1289-1306.
62. Lustig, M., D. Donoho, and J.M. Pauly, *Sparse MRI: The application of compressed sensing for rapid MR imaging*. Magn Reson Med, 2007. **58**(6): p. 1182-95.
63. Block, K.T., M. Uecker, and J. Frahm, *Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint*. Magn Reson Med, 2007. **57**(6): p. 1086-98.
64. Candes, E.J., J. Romberg, and T. Tao, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*. Ieee Transactions on Information Theory, 2006. **52**(2): p. 489-509.

65. Gamper, U., P. Boesiger, and S. Kozerke, *Compressed sensing in dynamic MRI*. Magn Reson Med, 2008. **59**(2): p. 365-73.
66. Lustig, M., et al. *k-t SPARSE: high frame rate dynamic MRI exploiting spatio-temporal sparsity*. in *14th Annual Meeting of ISMRM*. 2006. Seattle.
67. Otazo, R., et al., *Combination of compressed sensing and parallel imaging for highly accelerated first-pass cardiac perfusion MRI*. Magn Reson Med, 2010. **64**(3): p. 767-76.
68. Usman, M., et al., *k-t Group sparse: a method for accelerating dynamic MRI*. Magn Reson Med, 2011. **66**(4): p. 1163-76.
69. Feng, L., et al., *Golden-angle radial sparse parallel MRI: combination of compressed sensing, parallel imaging, and golden-angle radial sampling for fast and flexible dynamic volumetric MRI*. Magn Reson Med, 2014. **72**(3): p. 707-17.
70. Feng, L., et al., *XD-GRASP: Golden-angle radial MRI with reconstruction of extra motion-state dimensions using compressed sensing*. Magn Reson Med, 2016. **75**(2): p. 775-88.
71. Jaspán, O.N., R. Fleysheer, and M.L. Lipton, *Compressed sensing MRI: a review of the clinical literature*. Br J Radiol, 2015. **88**(1056): p. 20150487.
72. Rapacchi, S., et al., *High spatial and temporal resolution dynamic contrast-enhanced magnetic resonance angiography using compressed sensing with magnitude image subtraction*. Magn Reson Med, 2014. **71**(5): p. 1771-83.
73. Winkelmann, S., et al., *An optimal radial profile order based on the Golden Ratio for time-resolved MRI*. IEEE Trans Med Imaging, 2007. **26**(1): p. 68-76.
74. Rasche, V., et al., *Resampling of data between arbitrary grids using convolution interpolation*. IEEE Trans Med Imaging, 1999. **18**(5): p. 385-92.
75. Pruessmann, K.P., et al., *SENSE: sensitivity encoding for fast MRI*. Magn Reson Med, 1999. **42**(5): p. 952-62.
76. Otsu, N., *A Threshold Selection Method from Gray-Level Histograms*. IEEE Transactions on Systems, Man, and Cybernetics, 1979. **9**(1): p. 62-66.
77. Fessler, J.A. and B.P. Sutton, *Nonuniform fast Fourier transforms using min-max interpolation*. IEEE Transactions on Signal Processing, 2003. **51**(2): p. 560-574.