Title
High-Resolution Optogenetic Functional Magnetic Resonance Imaging Powered by Compressed Sensing and Parallel Processing

Author
Le, Nguyen Van

Publication Date
2012

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

High-Resolution Optogenetic Functional Magnetic Resonance Imaging Powered by Compressed Sensing and Parallel Processing

A thesis submitted in partial satisfaction

of the requirements for the degree Master of Science

in Electrical Engineering

by

Nguyen Van Le

2012

ABSTRACT OF THE THESIS

High-Resolution Optogenetic Functional Magnetic Resonance Imaging

Powered by Compressed Sensing and Parallel Processing

by

Nguyen Van Le

Master of Science in Electrical Engineering

University of California, Los Angeles

Professor Jin Hyung Lee, Chair

Optogenetic functional magnetic resonance imaging (ofMRI) [1] is a powerful new technology that enables precise control of brain circuit elements while monitoring their causal outputs. To bring ofMRI to its full potential, it is essential to achieve high-spatial resolution with minimal distortions. With our proposed compressed sensing (CS) enabled method, high-spatial resolution ofMRI images can be obtained with a large field of view (FOV) without increasing spatial distortions and the amount of acquired data. The ofMRI data were sampled with passband balanced steady-state free precession (b-SSFP) [8, 17] fast stack-of-spiral sequence in order to achieve ultra-high-spatial resolution images in a short amount of time. Interleaves of data were randomly collected. The images were recovered from the undersampled k-space data by solving an unconstrained convex optimization problem, which balances the trade-off between data consistency and sparsity. The optimization problem can be solved by gradient descent combined

with backtracking line search algorithms. Discrete cosine transform (DCT) were chosen as a sparsifying transform. The ofMRI image reconstruction was processed in parallel on a graphics processing unit (GPU) using C/C++ language supported by NVIDIA CUDA engine in order to achieve short reconstruction time. An existing nonequispaced fast Fourier transform (NFFT) algorithm [13, 14] was modified for our GPU parallel processing purpose. The results demonstrate that the compressed sensing reconstructed image has higher resolution while maintaining a precise activation map, compared to a fully sampled low-resolution image with the same amount of data and scan time. A 4-D image can be reconstructed in less than fifteen minutes, which allows compressed sensing ofMRI to become a practical application.

The thesis of Nguyen Van Le is approved.

Alan Laub

Kung Yao

Jin Hyung Lee, Committee Chair

University of California, Los Angeles

2012

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First of all, I would like to express my gratitude to my academic advisor, Professor Jin Hyung Lee, who has enthusiastically guided me through the entire process of this thesis writing. She has provided me a professional working environment and helpful advice whenever I need it.

I am truly indebted to Professor Alan Laub and Professor Kung Yao, who spend their precious time to serve as my committee's members. I greatly appreciate their time, effort, and interest in this thesis.

I am thankful for Jin Hyung Lee's group members for their help and discussion. They are my co-workers and friends, with whom I have shared great time working together.

Last but not least, I would like to present my deepest appreciation to my family for the invaluable understanding, encouragement, and support.

CHAPTER 1 – INTRODUCTION

Compressed sensing algorithm has been proposed for magnetic resonance imaging (MRI) applications [4]. Many algorithms have been suggested to improve MRI image quality through reconstruction. However, the quality of a compressed sensing MRI image also depends on other factors such as experimental set up and data acquisition method. ofMRI, first introduced by my advisor, is a new powerful MRI based technology for brain circuitry study. ofMRI activation map, which is very sensitive to the image quality and scan noise, requires not only accurate image reconstruction but also enhancements from stimulation method and data acquisition. Instead of reducing scan time, we apply compressed sensing to increase resolution of ofMRI images, which results in much lower SNR of our acquired data. Therefore, we proposed our compressed sensing ofMRI method, which allows for high-SNR data acquisition and rapid reconstruction of high-resolution ofMRI images.

1.1 Functional Magnetic Resonance Imaging

Magnetic resonance imaging is a powerful noninvasive medical imaging technique that uses the principles of nuclear magnetic resonance (NMR), the spectroscopy study of the magnetic properties of the nucleus, to visualize a subject's internal anatomy. MR signals can be generated if a strong static magnetic field, magnetic gradients, and radiofrequency excitations are applied to a subject with high proton density (water, fat). Unlike other common imaging techniques such as x-ray computed tomography (CT) or positron emission tomography (PET), MRI does not involve ionizing radiation, which is known to cause potential health risks. In addition, MRI has the advantage of superior soft-tissue contrast compared to other imaging modalities, facilitating its widespread adoption by the medical and scientific community.

Functional magnetic resonance imaging (fMRI) is a technique that detects the associated changes in the flow of blood around the brain by utilizing the magnetic resonance imaging modality. It is thought that active regions of the brain require more oxygen supply to generate neural signals, which results in the increase of blood flow to that region. Therefore, by measuring the blood oxygenation level-dependent (BOLD) signals in a region of the brain, researchers can detect neuronal activity in that region.

1.2 Optogenetic Functional Magnetic Resonance Imaging

fMRI is widely used in brain circuit research because of the ability to detect neuronal activity. The brain can be thought of as an electronic circuit, which has many components connected by wires. In order to study circuit connectivity, a probe should be applied at different locations in the circuit to generate various input signals and the output signals of the circuit are recorded for analysis. By studying the responses of the circuit to different input signals, researchers are able to analyze and debug the brain circuitry. Similar to electronic circuitry, brain circuitry can be studied if the brain's neurons are excited to generate nervous impulses and the corresponding responses of the brain are captured. Since fMRI can accurately detect brain activity, it can be utilized to capture the responses of the brain to an input stimulation. Electrical brain stimulation (EBS) is a widely used method to stimulate neurons. However, EBS is non-selective since it stimulates all cells near the site of current injection. In 2005, Karl Deisseroth's group at Stanford University introduced optogenetics, a technique to allow for cell-type specific stimulation based on the expression of bacterial rhodopsin transgenes [18, 19, and 20]. When a specific wavelength of light delivered via a laser is shined onto genetically targeted cells carrying the rhodopsin protein, these cells can be stimulated or inhibited. Therefore, unlike EBS, optogenetics allows for highly selective stimulation of specific cells, which is suitable to be used as a probe for brain circuitry debugging. In 2010, Jin Hyung Lee's group invented a new technique to study brain circuitry *in vivo*, optogenetic fMRI (ofMRI), which combines optogenetics and fMRI to noninvasively monitor the response of the brain to selective stimulations of each brain circuit element [1].

1.3 Passband Balanced Steady State Free Precession

Compared to the conventional gradient-echo (GRE) blood oxygenation level-dependent (BOLD) fMRI utilizing echo-planer imaging (EPI), balanced-steady-state free precession (b-SSFP) imaging possesses the advantages of distortion-free 3D imaging, high-resolution isotropic voxel acquisition, and minimal signal dropouts [7, 8]. b-SSFP fMRI uses balanced gradient pulses to spatially encode the magnetic resonance (MR) signals generated by fast radiofrequency (RF) excitation pulses during each rapid excitation repetition interval (T_R). There are two b-SSFP imaging techniques, namely transition-band b-SSFP and passband b-SSFP. Transition-band b-SSFP is the original technique, first proposed by Scheffler et al. [22]. With the use of the steep transitional portion of the b-SSFP off-resonance spectrum, this method produces high oxygen contrast due to the shift in resonance frequency induced by deoxyhemoglobin. Since transition-band b-SSFP fMRI generates oxygenation sensitive contrast in a narrow range of frequencies near resonance, even for small volume coverage, it requires multi-frequency acquisitions [23]. Unlike transition-band b-SSFP, the passband b-SSFP approach uses the flat portion, instead of the steep transitional portion, of the b-SSFP off-resonance spectrum. The use of the flat portion of the off-resonance spectrum allows for high-resolution imaging of the whole brain with only two acquisitions [7, 8].

The short T_R characteristic of b-SSFP (few milliseconds) is compatible with 3D imaging. Moreover, 3D imaging produces more stable steady-state in the presence of blood flow and motion, compared to 2D multi-slice acquisition. 3D acquisition is often combined with interleaved stack-of-EPI or interleaved stack-of-spiral trajectories in order to achieve high-resolution imaging with large field-of-view (FOV).

In b-SSFP fMRI, data acquisitions usually focus on a region of interest (ROI) in the brain. When large volume coverage is required, it is necessary to adjust the phase-cycling angles to shift the oxygen-sensitive region of the b-SSFP [7, 8]. For full-brain imaging, while the transition-band approach requires multiple acquisitions with different adjusted phase-cycling angles, two acquisitions at 0° and 180° phase-cycling angles are sufficient to cover the entire off-resonance spectrum in passband b-SSFP fMRI. To combine the two acquisitions, maximum intensity projection (MIP) is utilized. With this method, the combination can avoid the mixing contrast from the passband and transition-band regions.

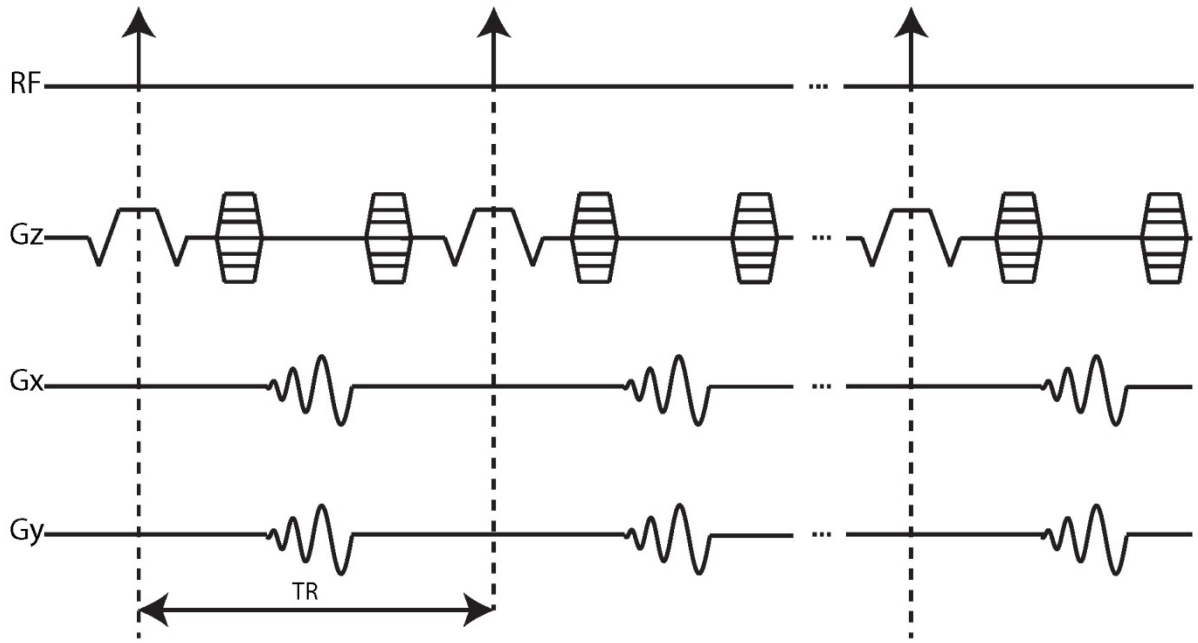


Figure 1.1: Balanced-steady-state free precession (b-SSFP) pulse sequence

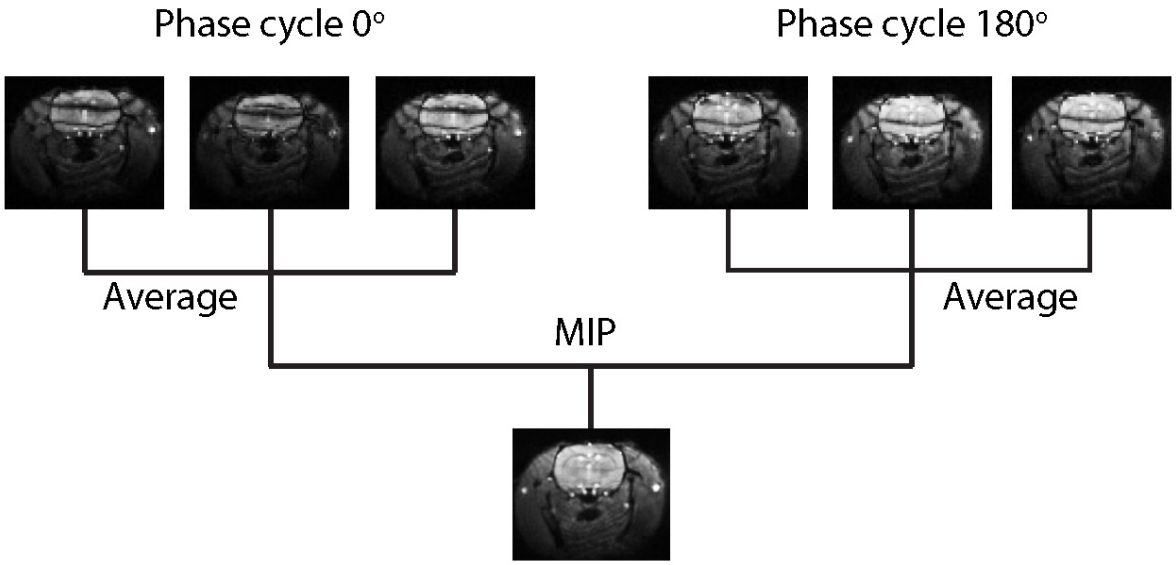


Figure 1.2: Two-acquisition method. The black strips from the brain regions are from the banding effect of passband b-SSFP.

1.4 Compressed Sensing Algorithm

With the MRI technique, high-contrast soft-tissue images can be obtained without known health risks. However, the disadvantage of MRI is long acquisition time due to physical limitations such as slew-rate. In ofMRI, which monitors dynamic brain activities, it is crucial to acquire data in a short interval of time to achieve high temporal resolution with minimal spatial distortion. Within 3-second temporal resolution, researchers can acquire only 500 x 500 x 500 μm^3 spatial resolution 3-D ofMRI images with fast stack-of-spiral sequences and passband bSSFP method. Therefore, compressed sensing theory is necessary in order to obtain higher spatial resolution ofMRI images with an equivalent amount of acquired data and temporal resolution.

Compressed sensing theory allows for the reconstruction of signals from undersampled data without aliasing artifacts. In other words, with a minimal amount of acquired data to reconstruct the fully sampled low-resolution images, we can reconstruct higher-resolution images with compressed sensing theory. In Fourier domain, if the spectrum of an image is undersampled, aliasing artifacts will occur in image domain. ofMRI data are sampled in Fourier domain called k-space. Therefore, undersampled ofMRI k-space data result in aliasing artifacts in spatial domains if the images are reconstructed by adjoint Fourier transform.

In order to be reconstructed from undersampled data, ofMRI images must have a sparse representation in a transform domain and the aliasing artifacts have to be incoherent in that domain. Most ofMRI images have a sparse representation, which is a matrix with very few non-zero coefficients. In other words, they are compressible. Sparsifying transforms can be found in various compression algorithms. The most popular image compression techniques are JPEG and JPEG-2000, underlined by discrete cosine transform (DCT) and wavelet transform, respectively.

In terms of energy distributions, the sparsifying transforms compact most of the energy and information of the image into very few coefficients. Other coefficients have very little or no energy distribution. With high-energy coefficients, we can reconstruct the image with minimal distortion, depending on compression techniques. This is the mechanism of lossy compression algorithms.

To obtain the incoherent aliasing artifacts in the transform domain, the k-space has to be sampled randomly. The aliasing artifacts generated by randomly undersampled data have the properties of additive noise, whose spectrum spreads out in the transform domain, as illustrated in figure 1.3. Therefore, the MR signals are recoverable by a thresholding method, as discussed in [4]. Unlike random sampling, traditional equispaced undersampling does not produce noise-like aliasing which greatly distorts the MR signals and prevents them from recovery because the aliasing artifacts cannot be distinguished from the signals. However, truly random sampling is impractical since it does not reduce the acquisition time.

Our ofMRI data sampling follows the pattern of stack-of-spiral interleaves. Each RF pulse excites one interleaf and data readout is performed for that interleaf. The random sampling within an interleaf does not reduce the scan time. In addition, with passband b-SSFP 3D imaging, most of the energy concentrates around the center of the k_z-axis. In practice, the numbers of interleaves should be denser in the slides that are closer to the center of k_z-axis for energy distribution matching purpose. As a result, we perform the data readout for the entire interleaf and excite more interleaves in the slides that are closer to the center of k_z-axis while we randomize the interleaves selection in each slide in order to obtain the best subset of k-space sampling. With this sampling scheme, we cannot achieve truly incoherent aliasing artifacts.

However, the artifacts are partially incoherent, which still allows for the recovery of the ofMRI images.

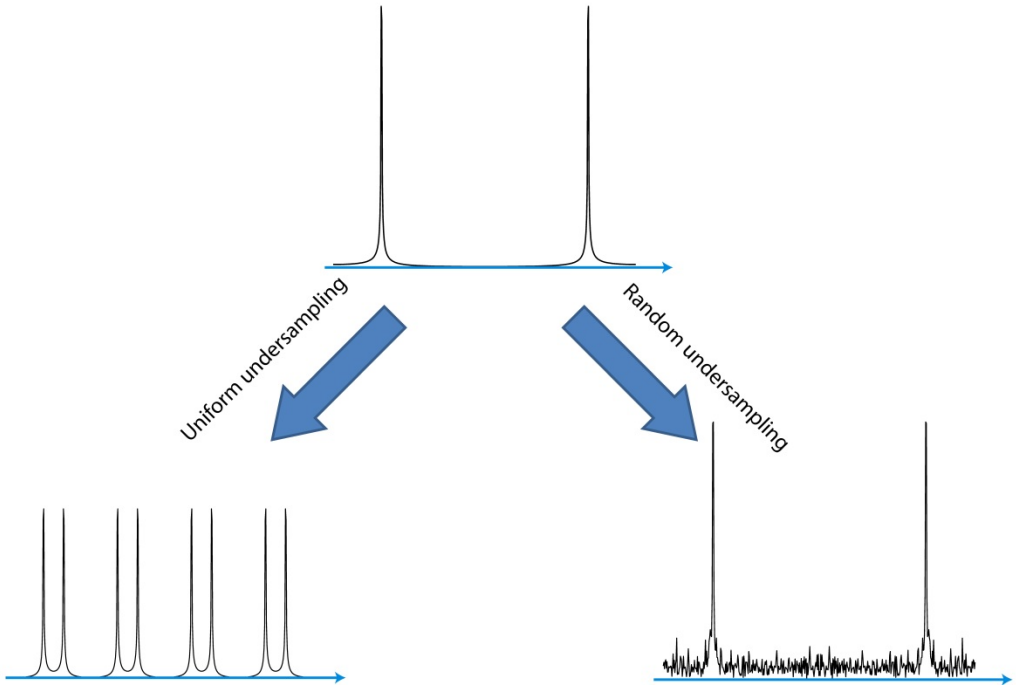


Figure 1.3: Uniform undersampling versus random undersampling. Top figure is sparse representation of cosine function in Fourier domain. Left figure is the aliased spectrum of undersampled cosin function. Random undersampling result in noise-like aliasing artifacts in right figure.

CHAPTER 2 – PROPOSED COMPRESSED SENSING ENABLED ofMRI

2.1 Compressed Sensing Data Acquisition

For activation map verification, low-resolution data without CS enabled were collected to determine the activation map of a specific stimulation on an animal. The low-resolution data were sampled with passband b-SSFP fMRI method with fast 10-interleaf stack-of-spiral sequences, 3.5 x 3.5 x 1.6 cm^3 volume coverage. The image was reconstructed on a 70 x 70 x 32 grid, which has a spatial resolution of 500 x 500 x 500 μm^3. 2-dimensional (2D) adjoint nonequispaced fast Fourier transform (NFFT) and fast Fourier transform were utilized for image reconstruction. The repetition time, echo time (T_E), and scan time was 9.375 ms, 2 ms, and 3 s, respectively. The flip angle was set to be 30° and the readout duration was 1.7 ms. The first 10 scans (30 s) were used for functional baseline estimation and magnetic field stabilization. In the next 120 scans (360 s), 6 cycles of 20-second stimulation were performed. The activation maps were determined with 0.35 threshold level.

In order to achieve higher resolution images with the same amount of acquired data, scan time, and temporal resolution as the low-resolution images, we developed a compressed sensing enabled ofMRI technique, which allows us to achieve an ultra-high spatial resolution of 210 x 210 x 500 μm^3 with the same volume coverage and temporal resolution, while avoiding any significant image distortion or signal dropout. In practice, it is impossible to achieve both this spatiotemporal resolution and field of view (FOV) with the traditional ofMRI method due to the MRI scanner's physical limitations. Similar to the low-resolution data, high-resolution data were also sampled with passband b-SSFP fMRI with stack-of-spiral trajectory. Except for the spatial resolution, all other parameters were maintained the same as the low-resolution data acquisition.

High-resolution images were recovered on 167 x 167 x 32 grid with 130 frames. In order to support the aforementioned high-spatial resolution and FOV, a fully sampled dataset requires 30-interleaf stack-of-spiral trajectory. With compressed sensing, one third of the total number of interleaves was sampled, meaning that the amount of acquired data reduces by a factor of three, compared to the fully sampled data. As discussed above, in order to attain a realistic sampling scheme matching the energy distribution, undersampling rates are not the same across slides on the k_z-axis. The number of interleaves is denser closer to the center and decreases toward the two sides of the k_z-axis. The sampled interleaves are also varied from each frame, producing incoherent aliasing artifacts in the transform domains of temporal dimension. The activation maps were also determined with the same threshold level.

Low-resolution and compressed sensing data were acquired from 7T Bruker with 39.6 G/cm maximum gradient amplitude and 457 G/cm/s maximum slew-rate.

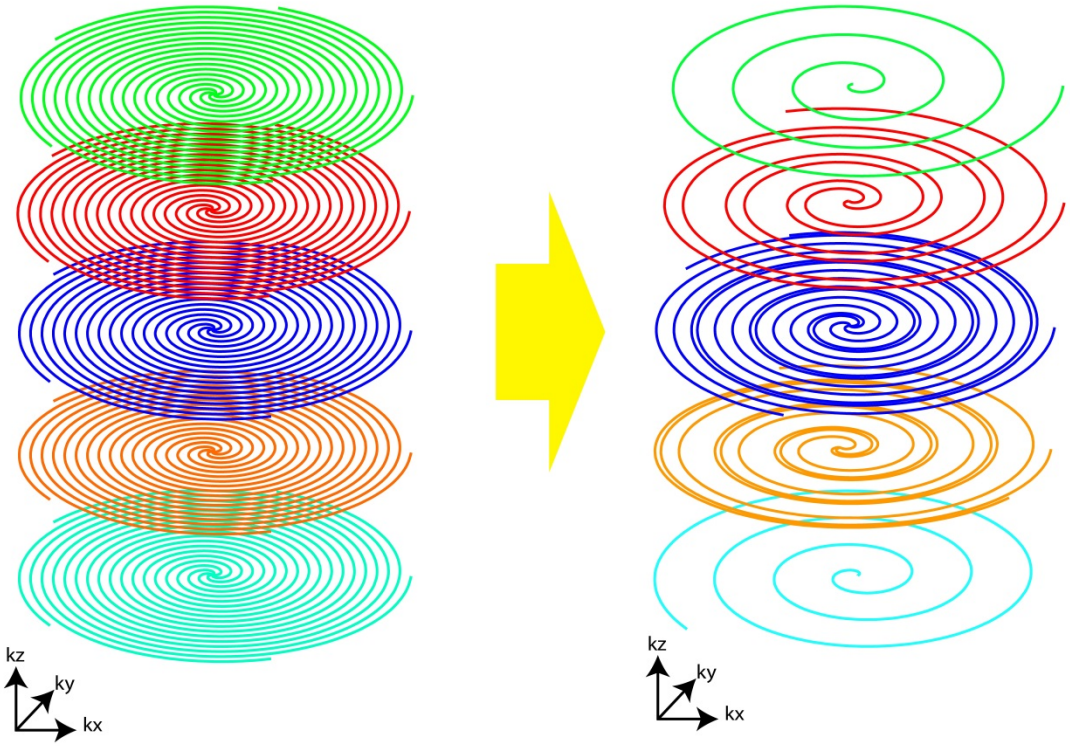


Figure 2.1: Fully sampled stack-of-spiral trajectory (left) and compressed sensing stack-of-spiral trajectory (right).

2.2 Compressed Sensing Image Reconstruction

Let $f(m)$ be the cost function, which is,

$$f(m) = \frac{1}{2}\|F_N m - y\|_2^2 + \sum_{k=1}^{2} \lambda_k \|\Psi_k m\|_1, \quad (1)$$

The two terms on the right-hand side of Eq. 1 represent data fidelity and sparsity over temporal and spatial dimensions of a 4-D image *m*, respectively. The goal of the reconstruction process is to determine a 4-D image *m* which minimizes the cost function $f(m)$:

$$\underset{m}{\operatorname{argmin}} \frac{1}{2}\|F_N m - y\|_2^2 + \sum_{k=1}^{2} \lambda_k \|\Psi_k m\|_1, \quad (2)$$

where λs are weighting parameters that determine the tradeoff between the data consistency and the sparsity while Ψs are sparsifying transforms. Ψ_1 is DCT of temporal dimension while Ψ_2 is DCT of all three spatial dimensions. $F_N \in \mathbb{C}^{n \times p}$ is the nonequispaced fast Fourier transform operation. $y \in \mathbb{C}^n$ is the acquired k-space data from the scanner and $m \in \mathbb{C}^p$ is the reconstructed 4D image. We analyzed temporal and spatial dimensions separately because we are interested in the variation of the BOLD signal over time. The spatial quality can be traded for the temporal quality in order to attain precise activation maps.

In this work, DCT is utilized as sparsifying transforms. DCT is widely used in many lossy compression techniques such as audio compression (MP3) and image compression (JPEG). DCT is chosen because of its fast computation and high compression ratio. For a given transform size, the DCT coefficients can be pre-computed and reused. The DCT of one dimension can be computed by matrix multiplication of pre-computed coefficient matrix and the data matrix. Most

of the ofMRI images exhibit extremely sparse representations in DCT domain, where the energy concentrates in the low-frequency portion.

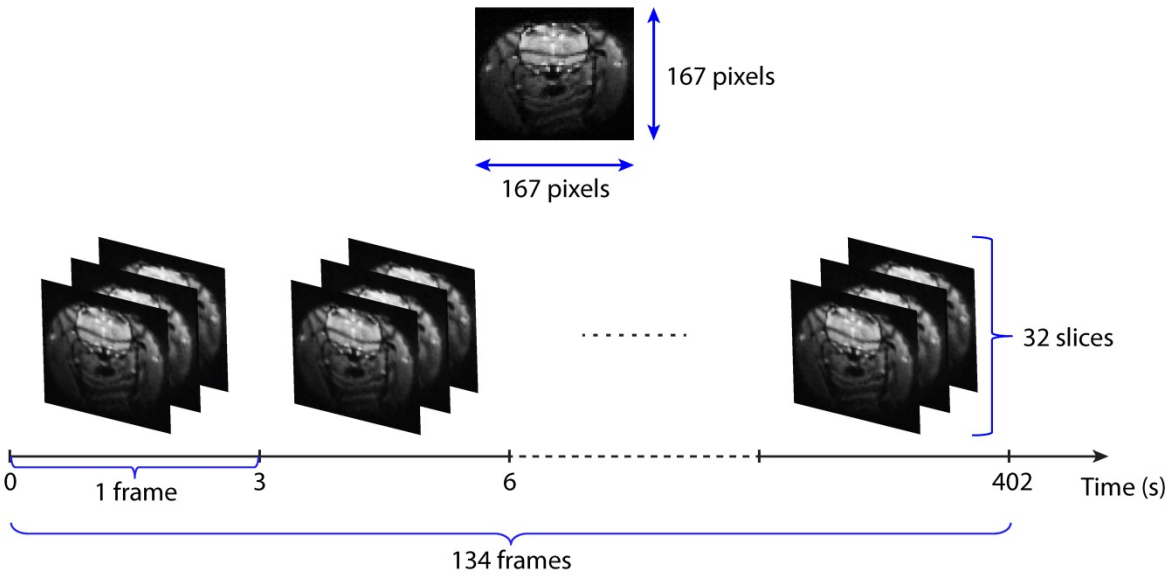


Figure 2.2: Reconstruction grid. For visualization, our ofMRI high-resolution 4D images are reconstructed on a 167 x 167 x 32 x 134 grid.

2.3 Accelerated Gradient Descent Algorithm

The gradient descent algorithm requires the computation of the gradient $\nabla f(m)$, which is described in the below equation:

$$g = \nabla f(m) = F_N^H(F_N m - y) + \sum_{k=1}^{2} \lambda_k \nabla \|\Psi_k m\|_1, \tag{3}$$

where F_N^H is the adjoint nonequispaced fast Fourier transform (iNFFT). L1-norm is a non-differentiable function, which should be approximated by a smooth function: $\|x\|_1 \approx \sum_i \left(\sqrt{x_i^* x_i + \mu^2} - \mu\right)$, where μ is the smoothing parameter.

Since we used the fast spiral sequences for data acquisition, nonequispaced fast Fourier transform (NFFT) was required to transform the data from the spiral k-space domain to the Cartesian spatial domain. The computation of NFFT, which is very expensive, is required in order to evaluate $f(m)$. However, the cost function $f(m - tg)$ has to be evaluated in each backtracking line search loop when *t* is updated. Therefore, to reduce the computational cost, we avoided computing NFFT in every backtracking line search loop by pre-computing some parts of the L2-norm with some simplification from the below equation:

$$f(m - tg) = \frac{1}{2}\|F_N(m - tg) - y\|_2^2 + \sum_{k=1}^{2} \lambda_k \|\Psi_k(m - tg)\|_1, \tag{4}$$

The first term can be written as:

$$\begin{aligned} \frac{1}{2}\|F_N(m - tg) - y\|_2^2 &= \frac{1}{2}\|(F_N m - y) - tF_N g\|_2^2 \\ &= \frac{1}{2}[(F_N m - y) - tF_N g]^H[(F_N m - y) - tF_N g] \end{aligned}$$

$$= \frac{1}{2}(F_N m - y)^H (F_N m - y) + \frac{1}{2}t^2 (F_N g)^H (F_N g)$$

$$- \frac{1}{2}t[(F_N m - y)^H (F_N g) + (F_N g)^H (F_N m - y)]$$

$$= \frac{1}{2}\|F_N m - y\|_2^2 + \frac{1}{2}t^2 \|F_N g\|_2^2 - t\Re\{(F_N m - y)^H (F_N g)\}$$

$$= a + t^2 b - tc \quad (5)$$

where $a = \frac{1}{2}\|F_N m - y\|_2^2$, $b = \frac{1}{2}\|F_N g\|_2^2$, and $c = \Re\{(F_N m - y)^H (F_N g)\}$. They are independent from t and can be pre-computed prior to the execution of the backtracking line search. In the backtracking line search loop, when t is updated, the value of the L2-norm is evaluated by simple addition and multiplication (Eq. 5), instead of computationally expensive NFFT and FFT. Evaluating DCT is much less computationally intensive. As a result, they are not necessarily pre-computed. Gradient descent and accelerate gradient descent algorithms are described in detail in appendix III and IV.

2.4 Parallel Reconstruction

Recently, demand for parallel computing has rapidly increased. Applications such as real-time and high-definition 3D graphics rely greatly on parallel computing. Both central processing units (CPU) and graphics processing units (GPU) have evolved into multi-core, multi-thread processors, attaining highly parallel structures. The parallel architectures of CPU and GPU are quite different. Therefore, CPUs and GPUs have different parallel characteristics. CPUs can support a limited number of concurrent threads (up to 48 threads with Hyper-Threading Technology). Threads on CPUs are generally used for heavy loaded computations. In contrast, GPUs can have a massive number of concurrent threads (a 16-multiprocessor can have 512 cores with more than 24,000 threads). However, threads on GPUs are extremely lightweight. Therefore, GPUs are suitable for highly parallel tasks with light computations for each thread. For instance, large matrix addition can be performed on GPUs with high efficiency while it would take much longer time to be processed on a CPU. Each element in the matrices can be mapped to a parallel processing thread, which performs only one addition. In 2006, NVIDIA introduced the CUDA engine, which is a very powerful and simple tool to implement parallel computing on GPUs to solve many parallelizable complex problems more efficiently than CPUs.

In this paper, GPU parallelization is implemented to speed up the reconstruction process to bring ofMRI to practice. Without parallelization, the reconstruction process would take a few days. However, with GPU parallelization, it takes less than 15 minutes to reconstruct and analyze a full 4D ofMRI image. The reconstruction algorithms were written in C/C++ language supported by the NVIDIA CUDA engine. Due to graphics card memory limitation and data size (the size of our ofMRI image is about 900 MB), the parallelization was implemented for small

processes such as computing matrix multiplication, calculating L1-norm, or performing fast Fourier transform for one dimension, etc.. Moreover, as mentioned above, in order to optimize the performance of GPU computing, the GPU threads should be lightweight. Therefore, unlike CPU parallel processing, it is inefficient to assign each reconstruction process to a GPU thread, or block. CPU and its physical memory, RAM, are considered to be a host system, while the GPU and its memory are treated as the device. The host and the device are separated by the PCI Express (PCIe) bus. The bandwidth of the PCIe bus is much smaller compared to the bandwidth between the CPU and RAM or the bandwidth between the GPU and its memory. To achieve optimal throughput, data should not be frequently transmitted through the PCIe bus. Therefore, the acquired data and the intermediate data for the gradient descent algorithm were alternatively stored on either device's memory or host's memory to balance the device's memory demand and throughput. Frequently used data such as the 4D image has higher priority to be stored on device's memory than other rarely used data. CUDA built-in library such as cuBLAS and cuFFT were utilized to optimize the performance of GPU processing. The nonequispaced fast Fourier transform algorithm was rewritten with the CUDA engine enabled for GPU processing. Table 1, 2, 3, and 4 in chapter 3 demonstrate the time improvement of our parallelized processes.

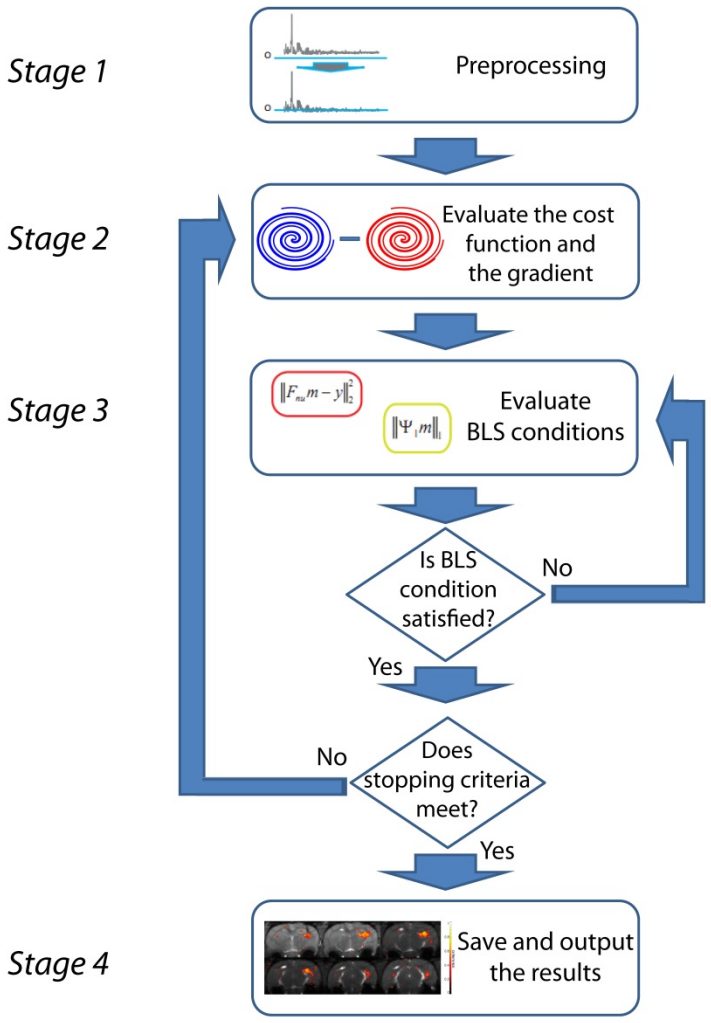


Figure 2.3: Reconstruction process. Parallel reconstruction is performed within each stage.

2.5 Parallelized NFFT for GPU Processing

We applied GPU parallel processing to the NFFT algorithm developed by Daniel Potts' group [13, 14]. The main idea behind this algorithm is to approximate the NFFT of a function by a linear combination of shifted periodic window functions.

Let $\mathbb{T}^d$ be the domain of the nonequispaced nodes x_j

$$\mathbb{T}^d := \left\{ x = x(t)_{t=0,\dots,d-1} \in \mathbb{R}^d : -\frac{1}{2} \le x_t \le \frac{1}{2}, t = 0, \dots, d-1 \right\}$$

where $d \in \mathbb{N}$ is the number of dimensions. Also, let

$$I_N := \left\{ k = (k_t)_{t=0,\dots,d-1} \in \mathbb{Z}^d : -\frac{N_t}{2} \le k_t \le \frac{N_t}{2}, t = 0, \dots, d-1 \right\}$$

where N_t is the size of the Cartesian grid in a specific dimension.

Without loss of generality, we consider 1D NFFT

$$f(x_j) := \sum_{k \in I_N} \hat{f}_k e^{-j2\pi k x_j} \tag{6}$$

where $\hat{f}_k$ is a 1D equispaced function and f is its NFFT with nonequispaced nodes $x_j, j = 0, 1, \dots, M-1$. Let s be an approximation of f by a linear combination of shifted periodic window functions $\tilde{\varphi}$

$$s(x) := \sum_{l \in I_n} g_l \tilde{\varphi}\left(x - \frac{l}{n}\right) \tag{7}$$

where $n := \sigma N$. $\sigma > 1$ is the oversampling factor. Assume that the periodic window function $\tilde{\varphi}$ can be represented by its Fourier series:

$$\tilde{\varphi}(x) = \sum_{k\in\mathbb{Z}} c_k e^{-j2\pi kx} \tag{8}$$

and $\hat{g}_k$ are the Fourier coefficients of the weighting parameters g_l:

$$\hat{g}_k := \sum_{l\in I_n} g_l e^{j2\pi\frac{kl}{n}} \tag{9}$$

Combine (7), (8), and (9), we have:

$$s(x) = \sum_{k\in I_N} \hat{g}_k c_k e^{-j2\pi kx} + \sum_{r\in\mathbb{Z}\setminus\{0\}} \sum_{k\in I_n} \hat{g}_k c_{k+nr} e^{-j2\pi(k+nr)x} \tag{10}$$

Compare (6) and (10) with an assumption that c_k is small for $|k| > n - \frac{N}{2}$, we have:

$$\hat{g}_k := \begin{cases} \dfrac{\hat{f}_k}{c_k}, & k \in I_N, \\ 0, & k \in I_n \setminus I_N \end{cases} \tag{11}$$

Hence, g_l can be obtained by

$$g_l = \frac{1}{n} \sum_{k\in I_N} \hat{g}_k e^{-j2\pi\frac{kl}{n}} \qquad (l \in I_n) \tag{12}$$

The window function φ can be also approximated by a finite-length window function ψ:

$$\psi(x) = \varphi(x)\Pi\left(\frac{n}{2m}x\right), \qquad m \ll n, m \in \mathbb{N} \tag{13}$$

Let $\tilde{\psi}$ be the periodic version of ψ. The new index set is

$$I_{n,m}(x_j) := \left\{l \in I_N : nx_j - m \le l \le nx_j + m\right\} \tag{14}$$

Then, s can be defined as:

$$s(x_j) := \sum_{l \in I_{n,m}(x_j)} g_l \tilde{\psi}\left(x_j - \frac{l}{n}\right) \tag{15}$$

Gaussian, cardinal central B-splines, and Kaiser-Bessel functions are widely used as window function because of their small aliasing and truncation errors.

In summary, NFFT f of an equispaced function $\hat{f}_k$ can be approximated by the following steps:

- Compute c_k which are the Fourier series coefficients of the window function $\tilde{\varphi}$.
- Evaluate $\hat{g}_k$ from the equispaced function $\hat{f}_k$ and c_k.
- Perform FFT to $\hat{g}_k$ to obtain g_l.
- Obtain the approximation of f from g_l and the truncated window function $\tilde{\psi}$ as Eqn. 15.

Since the grid size of the image and the k-space nodes are fixed, the window function $\tilde{\varphi}$, its Fourier coefficients c_k, and its truncated version $\tilde{\psi}$ can be pre-computed and reused for all NFFT and adjoint NFFT evaluations. The computation of multidimensional NFFT and adjoint NFFT are demonstrated in appendix I and II.

We modified the NFFT library with CUDA engine to empower GPU parallel processing. When the NFFT plan is initialized, c_k are computed in parallel on the GPU and stored on the graphics card global memory. For every 2D slice, $\hat{g}_k$ is calculated in parallel from Cartesian image $\hat{f}_k$ and the Fourier coefficients c_k. The calculation of an element of $\hat{g}_k$ is performed by a GPU thread. We utilize CUDA cuFFT built-in library to evaluate g_l, which is simply the Fourier transform of $\hat{g}_k$. The resulting NFFT is determined from ψ and g_l, which are cached in texture memory for rapid data retrieval.

The parallelization of the adjoint NFFT is very similar to the forward NFFT, except for the inverted procedure and the inverse FFT in step two. The performance of our forward NFFT and adjoint NFFT are demonstrated in table 1 in chapter 3.

2.6 ofMRI Stimulation and Activation Map Analysis

A living animal's brains always exhibits neuronal activity. In order to distinguish the activity caused by the optogenetic stimulation from the normal brain's activity, we use a low-frequency periodic stimulation. If some regions of the brain respond to that stimulation; the brain's activity in those regions should have approximately the same frequency with the stimulation. The strength of the responsive signals should be much higher than the normal brain signals and the noise level so that the active voxels from the stimulation can be distinguished from other active voxels or noise at the same frequency by a threshold.

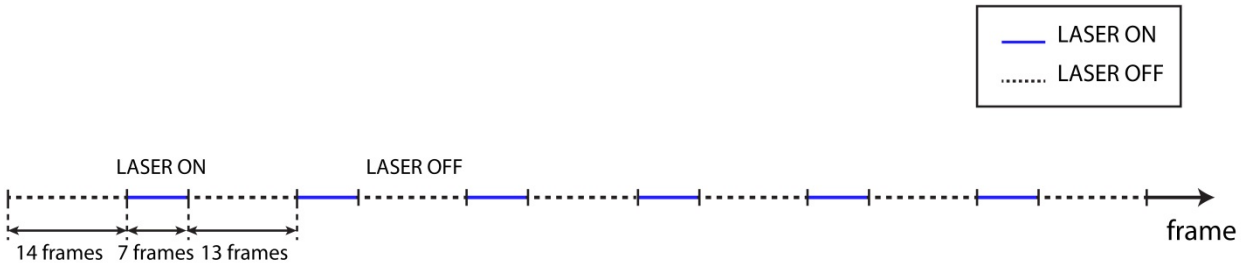


Figure 2.4: Stimulation procedure. The stimulation starts after 14 frames with 6 cycles. There are 20 frames for each cycle, where the laser is turned on during the first 7 frames.

We have 134 frames for one data set with 3-second temporal resolution. The first 4 frames are dummy frames, which are used to bring the MRI scanner to its steady state. These frames are removed from the reconstruction. The next 10 frames are used for baseline analysis. The optogenetic stimulation starts at the 14^{th} frame. The laser is on for 7 frames, and it is turned off during the next 13 frames. This procedure is repeated six times during one scan for one data set, as demonstrated in figure 2.4.

The activation maps of ofMRI images can be determined by Fourier analysis of 3D images across temporal dimension. Let f_π be the frequency of the brain signal generated by the optogenetic stimulation at one voxel, which can be fully controlled by the stimulation cycle. Also, let *f* be the spectrum of that voxel, which can be determined by Fourier transform of the intensity of the voxel across 120 frames, from frame 15 to frame 134. Let R be the square root of the ratio of the energy of the responsive frequency and the spectrum. We have:

$$R = \sqrt{\frac{f_\pi^2 + f_{-\pi}^2}{\sum_{i=1}^{119} f_i^2}} = \frac{\sqrt{2}|f_\pi|}{\sqrt{\sum_{i=1}^{119} f_i^2}} \tag{16}$$

In Eqn. 16, *i* starts from *1* instead of *0* since we ignore the energy of DC component. If the value of R of a voxel is higher than a threshold; then that voxel is considered to be active to the stimulation. Empirically, a 0.35 threshold level is widely used in fMRI. The coherence maps measure the strength of the activity.

2.7 Anesthesia and Motion Artifacts

Our ofMRI data were collected from living animals under anesthesia. The amount of isoflurane being administered is very important to the quality of the ofMRI images. For example, animals under deep anesthesia exhibit no response to the laser stimulations. Hence, no activity can be detected. Conversely, if the isoflurane level is too low, motion from the animals can occur and interfere with data acquistion. Therefore, the amount of isoflurane should be at an appropriate level so that the animals generate minimal motion while still exhibiting strong responses to the laser stimulations. Motions from breathing are unavoidable.

In practice, with our anesthesia level, the motions of the anesthetic animal are very little compared to the FOV. However, since the activation maps are calculated pixel-wise, small displacements of the brain regions between frames can greatly weaken the activation maps. As a result, it is necessary to have reconstructed images corrected for motion for accurate activation maps.

We use an inverse compositional algorithm which is based on Lucas-Kanade algorithm to align our images from motion. It is one of the most reliable image alignment techniques, which is widely used in computer vision. The goal of this algorithm is to minimize the sum of squared error between an input image and a reference image. Our motion correction tool can align 3D images from displacements and rotations between time frames in image domain. We have 6 parameters which measure the relative motions between images. They are the linear displacements in x, y, and z directions and the rotation around x, y, and z-axis. The relative motion can be described by the combination of these 6 parameters.

2.8 Signal-to-Noise Ratio

The compressed sensing technique allows for the reconstruction of 4D ofMRI images from undersampled data. However, undersampled data results in signal-to-noise ratio (SNR) loss due to fewer data readouts. Therefore, the compressed sensing reconstructed images are expected to have weaker activation maps, compared to the fully sampled high-resolution images and low-resolution images. The SNR of an ofMRI image is directly proportional to the voxel size, the square root of total readout interval, and pulse-sequence-dependent function $f(\rho, T_1, T_2)$:

$$SNR \propto (Voxel\ Size) \times \sqrt{Total\ Readout\ Interval} \times f(\rho, T_1, T_2), \tag{17}$$

Compared to the SNR of the low-resolution image with 500 x 500 x 500 μm^3 resolution, the SNR of 210 x 210 x 500 μm^3 3X undersampled high-resolution image is about 3.27 times less since it has the same readout interval but higher spatial resolution. In order to reduce the SNR gap, six compressed sensing high-resolution data sets were acquired, three for each phase-cycling angle. Then, three reconstructed images of the same phase were averaged to bring the SNR $\sqrt{3}$ times higher. The final compressed sensing reconstructed image was obtained from the MIP of the averages of the two phases, due to the use of passband b-SSFP full-brain coverage (figure 1.2). As a result, the SNR of the final compressed sensing image is equivalent to the SNR of the fully sampled image, but it is still theoretically 1.73 times less than the SNR of the low-resolution image.

2.9 Measurement of Reconstruction Quality

In ofMRI, we are interested in the accuracy of the activation maps. In other words, we are interested in the location of the brain's activity. Therefore, we measure the distortion between active voxels of the activation maps of high-resolution CS reconstructed image and phantoms or interpolated low-resolution images to determine how accurate the reconstruction is. There are two types of activation map errors, missed voxels and leaked voxels. If a voxel is active in the activation map of a low-resolution image or phantom, but not in the CS reconstructed image, then it is considered to be a missed voxel. In contrast, if a voxel in the activation map of a CS reconstructed image doesn't exist in the activation map of the corresponding low-resolution image or phantom, it is a leaked voxel. The sum of the total number of missed and leaked voxels is the total error, which measures the reconstruction quality. Our goal is reconstructing the compressed sensing ofMRI images with the most recoverable active voxels or minimal total error. The correlation between the number of recoverable active voxels and the total error is discussed in the next chapter.

Empirically, we have observed that the strength of the activation maps is very sensitive to the value of the weighting parameters in the cost function. Moreover, the optimal parameter set varies with the acquired data. For instance, the optimal parameter set that produces the strongest activation map for the hippocampus stimulation dataset is not the same as the optimal parameter set for the thalamus stimulation or phantom images dataset. However, the optimal parameter set is almost the same for the same phantom with different additive noise levels (different SNR).

Summary of Chapter 2

In this chapter, we have proposed our methods to improve many procedures, from data acquisition, stimulation to image reconstruction and analysis. Experimental set up, including anesthesia, stimulation, and pulse sequence design, has important consequences for image quality. Since activation maps are very sensitive to image quality, the experimental set up has to be carefully performed. Our parallel image reconstruction and analysis methods can minimize artifacts from our experiment and enhance the visualization of the activation maps.

Data acquisition with passband b-SSFP stack-of-spiral trajectory process is very important to quickly attain high-SNR data. Energy-matching random sampling results in incoherent aliasing artifacts and preserves more information in undersampled data. Stimulation and anesthesia processes have to be performed carefully to achieve good activation maps with minimal motion artifacts. Parallelized reconstruction should be designed carefully to optimize the performance of GPU processing. Motion correction aligns 2D slices to strengthen the activation map. The combination of these methods allows for achieving high-quality compressed sensing ofMRI images with rapid scan time.

CHAPTER 3 – RESULTS

3.1 Performance of GPU Processing

Our reconstruction process was performed on a Linux operating system platform, powered by Intel Core i7 2600k quad-core, 16 GB DDR3 dual-channel, NVIDIA GeForce GTX 580 with 512 CUDA cores.

GPU parallelization offers much faster image processing capability, compared to traditional sequential CPU processing. Table 1 compares the execution time of the modified NFFT and its adjoint between parallelized and non-parallelized versions. As expected, the time to execute forward NFFT and adjoint NFFT are the same, since the calculation of the adjoint NFFT is just the inverted procedure of the forward NFFT. We first tried to modify NFFT with “pthread” library for CPU parallelization. We utilized all 8 threads of the Intel Core i7 CPU; each thread performs NFFT to a slide of the image. This method can cut the execution time into half, compared to the non-parallelization version. As expected, the GPU parallelization has the best performance, about 7 times faster than the CPU parallelization.

In table 2, we compare the execution time of some matrix operations. Matlab has great performance for some matrix operations, compared to traditional C/C++ algorithms. Therefore, we also include it in our comparison. All matrices are complex with single precision. Matrix addition is performed between two matrices, whose total number of elements is 116,018,240. This is also the total size of our 4D images. Similarly, a matrix with the same size is used in scalar multiplication. For matrix-matrix multiplication, we measure the time needed to compute the product of 167-by-167 and 167-by-694,720 matrices. We also measure the time to transpose every 167 x 167 slices of our 4D image. Finally, the time to evaluate the dot product used in our

accelerated gradient descent algorithm is recorded. Each array has the size of 17,971,200 elements, which is the total number of samples of our acquired k-space data. Table 2 shows that Matlab has fast algorithms for matrix operations. However, our GPU parallel processing has superior performance over Matlab, which has better performance than the non-parallelization approach. It greatly improves the performance of the computationally expensive matrix-matrix multiplication operation. As a result, as shown in Table 3, with GPU parallelization, we can significantly reduce the reconstruction time, which is one of the most important factors in bringing ofMRI to practice.

As discussed in section 2.4, one of the challenging problems in GPU processing design is the balance between device's memory allocation and data transmission through the PCIe bus. We perform matrix addition, matrix scalar multiplication, and matrix-matrix multiplication with built-in cuBLAS library to matrices that allocated either on host's or device's memory. We used mapped page-locked memory allocation for host's memory allocation, which has the fastest transfer rate between host and device. For device allocation, data were stored in device's global memory. From table 4, although mapped memory has faster processing rate compared to traditional sequential processing and Matlab (matrix addition and matrix scalar multiplication), it is still much slower than the device's memory allocation approach because of the bottleneck at PCIe bus.

	Without Parallelization	With CPU Parallelization (pthread)	With GPU Parallelization (CUDA)
Forward NFFT	40 seconds	20 seconds	3 seconds
Adjoint NFFT	40 seconds	20 seconds	3 seconds

Table 1: Timing table of the modified nonequispaced fast Fourier transform.

	Without Parallelization (C/C++)	Matlab	With GPU Parallelization (CUDA)
Matrix addition (subtraction)	0.33 s	0.49 s	0.011 s
Matrix scalar multiplication	0.55 s	0.64 s	0.011 s
Matrix transpose	1.9 s	0.87 s	0.22 s
Matrix-matrix multiplication	200 s	1.72 s	0.14 s
Dot product	0.21 s	0.10 s	0.04 s

Table 2: Timing table of matrix operations used in the reconstruction.

	Without Parallelization	With GPU Parallelization (CUDA)
DCT for a 4D image	812 s	0.86 s
Evaluate L1-norm	0.66 s	0.099 s
Evaluate L2-norm	2.21 s	0.28 s
Reconstruction time for 1 data set	~2 days	~ 15 minutes

Table 3: Timing table compares the performance of parallelized and non-parallelized versions of main processes in the reconstruction.

	Memory Allocation on Host	Memory Allocation on Device
Matrix addition (subtraction)	0.32 s	0.011 s
Matrix scalar multiplication	0.32 s	0.011 s
Matrix-matrix multiplication	3.0 s	0.14 s

Table 4: Performance of GPU processing between host's and device's memory allocation.

3.2 Hippocampus and Thalamus Stimulations

Figure 3.1 and figure 3.2 illustrate the activation maps of the hippocampus and thalamus stimulations, respectively. The top image in each figure is the activation map of the fully-sampled low-resolution image. The bottom image is the compressed sensing reconstructed high-resolution image, corresponding to the low-resolution image in each figure. The bar on the right-hand-side presents the strength of the active voxels mapped by colors.

It is obvious that the CS reconstructed images look sharper than the fully-sampled images since they have higher resolution. We also notice the accurate positions of the activation maps in the high-resolution images, compared to their low-resolution versions. However, the activation maps in the high-resolution images are weaker compared to the corresponding activation maps in the low-resolution images. It is expected because of the SNR reduction from increasing spatial resolution, as discussed. However, compressed sensing possesses SNR boosting ability, which reduces the SNR gap between the low-resolution fully sampled image and the compressed sensing reconstructed undersampled image. Another source of error we have to consider is the non-identical response of the brain. The low-resolution data were acquired right before the compressed sensing data. Some regions of the brain may need time to recover for the next stimulation. Unfortunately, we cannot determine what regions would need recovery time and their recovery time interval. Hence, using the activation maps of the low-resolution images may not be accurate to measure the reconstruction quality. We need pre-determined activation maps to compare with our reconstructed images. Therefore, we need to create phantoms where the activation maps are pre-determined.

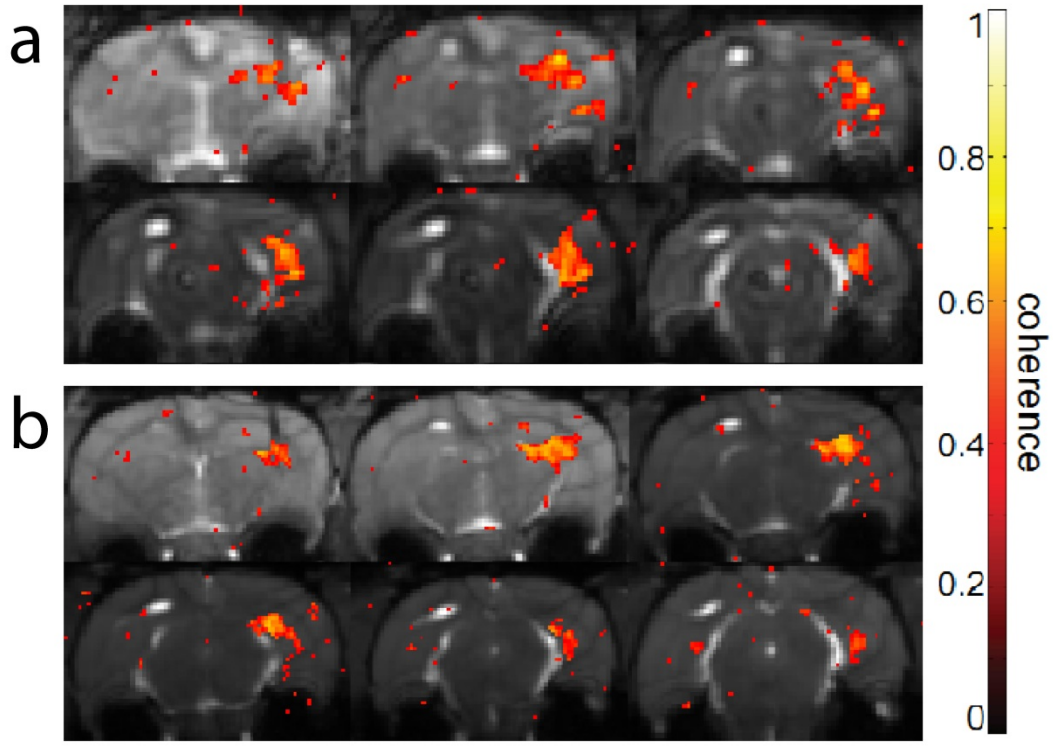


Figure 3.1: Hippocampus stimulation. (a) Coherence map of low-resolution fully sampled image. (b) Coherence map of compressed sensing reconstructed image.

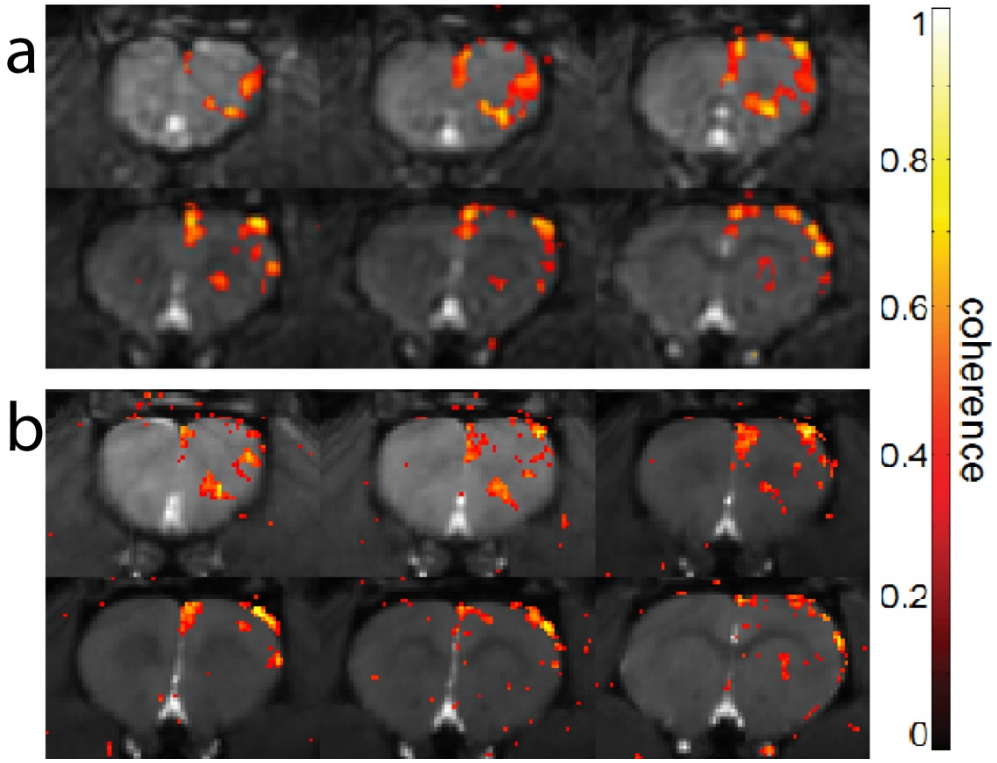


Figure 3.2: Thalamus stimulation. (a) Coherence map of low-resolution fully sampled image. (b) Coherence map of compressed sensing reconstructed image.

3.3 Phantom Experiment

We also created phantom images with various activation maps and SNR to determine the reliability of our reconstruction algorithms. Unlike the low-resolution images, phantom images provide pre-determined activation maps for the evaluation of the reconstruction's reliability. The reconstructed images with noise-free background are used as bases for our phantoms. Pre-determined activation maps and Rician noise, which is typical additive noise in MRI, are added to the basis images. Those images are transformed to stack-of-spiral Fourier domain and the interleaves are removed with the same sequence we used for compressed sensing data acquisition. In other words, we use the same k-space stack-of-spiral trajectory and interleaf selection as what we used to acquired undersampled data.

The activation maps of the phantoms are letters, which have sharp edges and uniform intensities. The letters were added on consecutive slices on z-axis, which results in sharp transitions of activation map between slices as well. Different noise levels were also added to the phantoms to determine the noise-robustness of our compressed sensing algorithm. The compressed sensing images were reconstructed with different weighting parameter sets in the cost function to study the relationship between those parameters and the reconstructed activation maps.

Figure 3.3 illustrates the phantom image in (a) and the reconstructed image in (b) without activation map analysis. Without any noise added to the phantom, the images in (a) and (b) look very similar. The distortion is not obvious and cannot be recognized without computer analyses. However, in figure 3.4, when the activation map analyses are added, we can now see the differences between (a) and (b). Because of the sensitivity of the activation maps to the image

quality, the compressed sensing ofMRI is much more challenging than other compressed sensing applications, such as conventional MRI. It requires high-quality and high-accuracy reconstruction in order to minimize the distortion and achieve precise activation maps.

In figure 3.4, 3.5, and 3.6, activation maps are manually added to the phantoms without noise. They are reconstructed with three different parameter sets, which results in strong, medium, and weak activation maps of the reconstructed images. In the strong activation map, all letters are recovered from the undersampled data. We also experience leaked active voxels from the adjacent slices, as shown in figure 3.4 (b). There is activation leakage on xy-plane, but it is not obvious since the resolution of that plane is very high. In addition, there is no other activation on the same xy-plane to see the interference. Therefore, if the activation leakage occurs in the voxels adjacent to the edges of the letters, it is hard to be recognized. However, it is obvious if the activation leakage occurs on the adjacent slices, which is next to each other on the z-axis of a 4D grid, where there is another activation whose shape is different from its neighbor. We can also observe that without added noise, the image reconstructed with $8x10^{-3}$ temporal penalty and 10^{-4} spatial penalty (figure 3.4) has a stronger activation map than image reconstructed with $3x10^{-3}$ temporal penalty and 10^{-4} spatial penalty (figure 3.5). The parameter set in figure 3.4 results in stronger activation map and more recoverable voxels, but the percentage error is also higher than activation map reconstructed by the parameter set in figure 3.5, because of the activation leakage. With $8x10^{-4}$ temporal penalty and 10^{-4} spatial penalty, the reconstructed image doesn't show obvious activation leakage, but its activation map is much weaker than the images reconstructed with the other parameter sets. Again, all background images (images without activation maps) look almost the same in spite of the differences between those activation maps.

From figure 3.7 to 3.9, 40-dB Rician noise is added to the phantom. We can easily see the distortion in the original activation map caused by the noise. With low-level noise added, the reconstructed activation maps are weaker than the activation maps of the noise-free phantoms. In addition, the activation leakage is also decreased, which significantly lower the percentage errors in figure 3.7 and 3.8. Most of the active voxels are recovered but they are weaker in general because of the SNR reduction from interleaf removal, as discussed in section 2.8.

We also add 30-dB noise (figure 3.10 to 3.12) and 25-dB noise (figure 3.13 to 3.15) to our phantoms. The original activation maps are highly distorted at these noise levels. Looking at the "F" and "O" characters in the original activation maps (top images), we can see that the missing parts of these characters have relatively weaker activation or irrecoverable in the compressed sensing reconstructed images of noise-free and 40-dB phantoms. These parts experience more distortion from noise than other parts of the activation map in the original phantoms. Therefore, the probability of recovering these parts is lower than the others. It is obvious that the reconstructed activation maps are stronger than the activation maps of the corresponding phantoms, which demonstrates the noise-suppression ability of our compressed sensing algorithm. Moreover, at 30-dB and 25-dB noise-levels, the activation leakage is minimal. Figure 3.13 and 3.14 illustrate that we can still recover many active voxels, which are corrupted in the original phantom by 25-dB noise. At 30-dB and 25-dB, $3x10^{-3}$ temporal penalty and 10^{-4} spatial penalty parameter set (figure 3.11 and 3.14) exhibits slightly stronger activation map with lower total error compared to $8x10^{-3}$ temporal penalty and 10^{-4} spatial penalty parameter set (figure 3.10 and 3.13).

The existence of Rician noise significantly reduces the activation leakage. Activation maps that have smooth gradient in 3D space also exhibit extremely less activation leakage than the above letter activation maps. Therefore, stronger activation maps, which have more active voxels, also have lower reconstruction errors since the missed errors are lower while the leaked errors are small or negligible. In practice, the optimal parameter set produces both strongest activation map and lower reconstruction error since noise always exists and the activation maps are smooth in the real ofMRI images. The impact of the smoothness of the activation maps to the activation leakage are discussed later in this section.

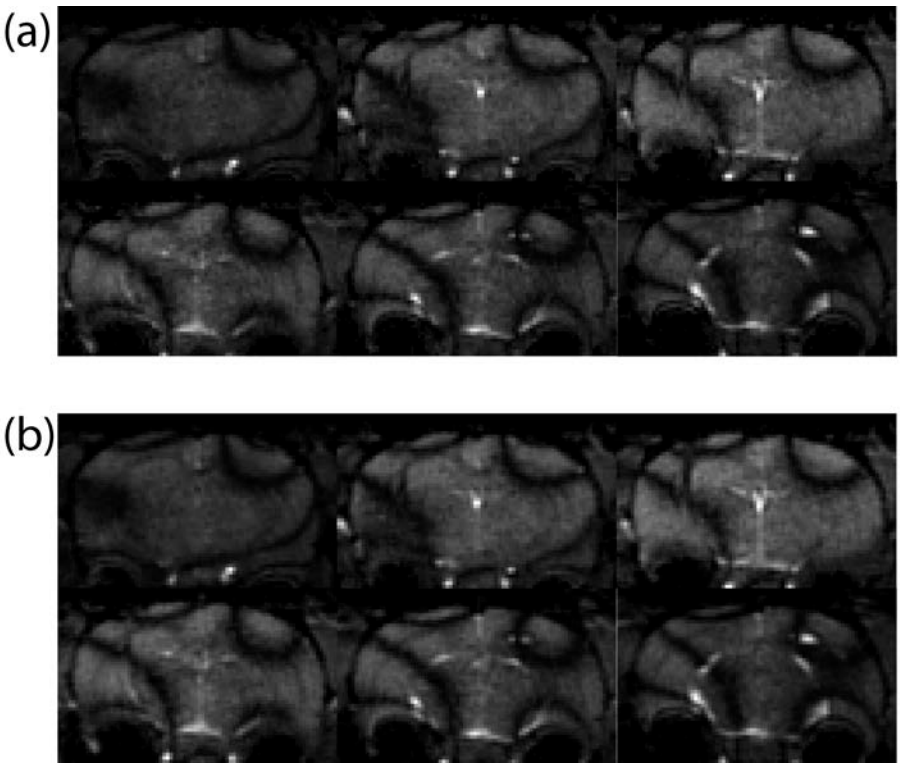
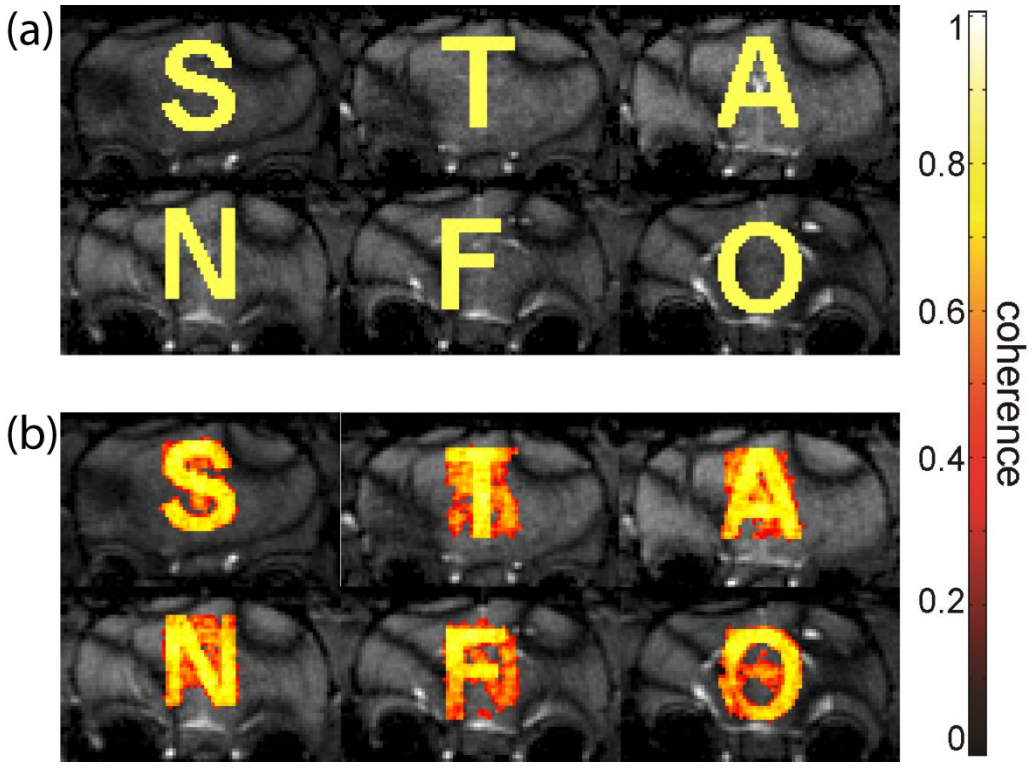
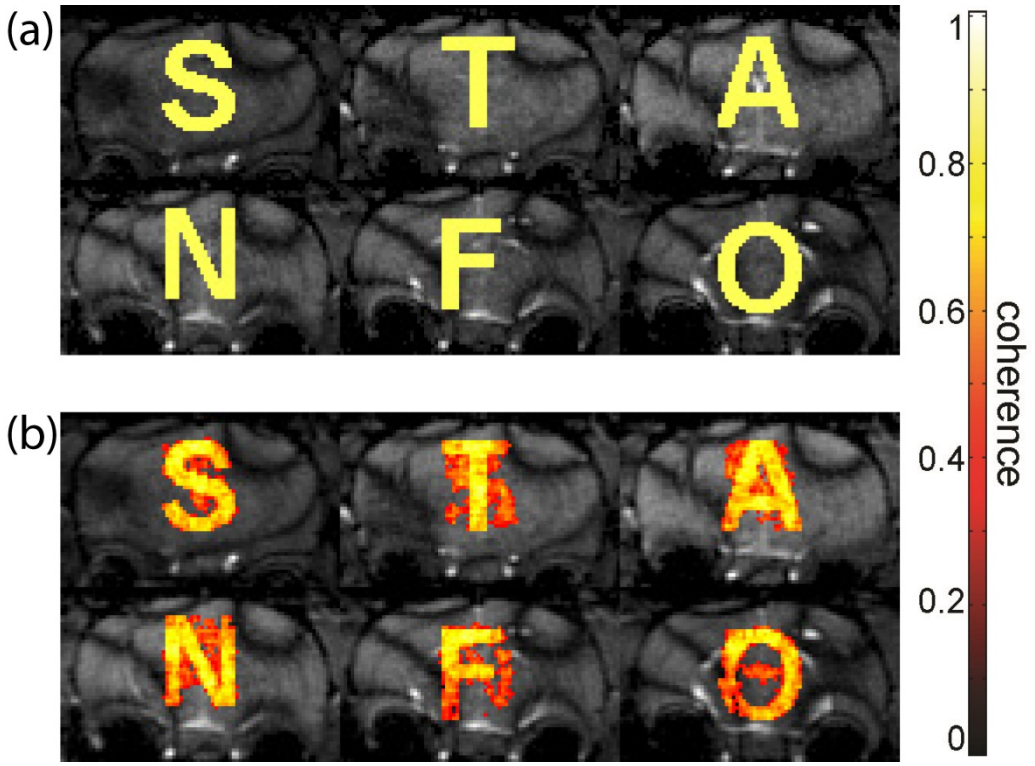


Figure 3.3: ofMRI images without activation map analysis. (a) The phantom images without additive noise. (b) Corresponding reconstructed images by our compressed sensing algorithm.



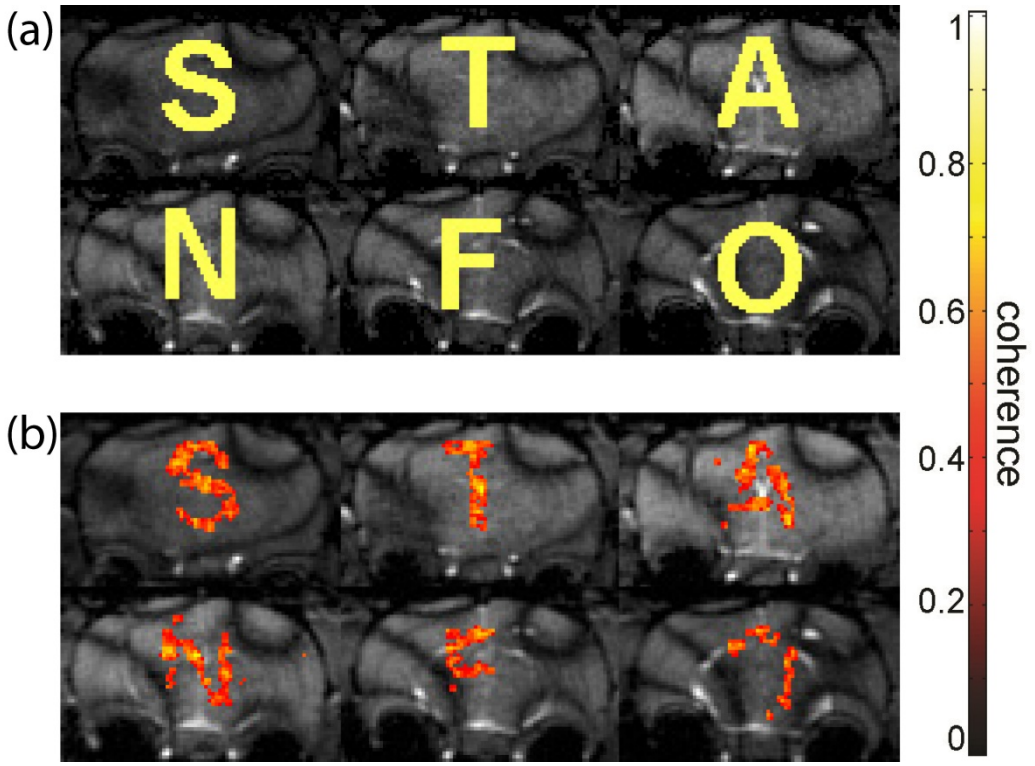
	# of Missed Voxels	# of Leaked Voxels	# of Recovered Voxels	% of Recoverable Voxels	Error Percentage
Phantom	0	0	2261	100%	0%
Reconstruction	25	1899	2236	98.9%	85.1%

Figure 3.4: No additive noise. (a) Activation map of the original phantom. (b) Activation map of the reconstructed image with $8x10^{-3}$ temporal penalty and 10^{-4} spatial penalty.



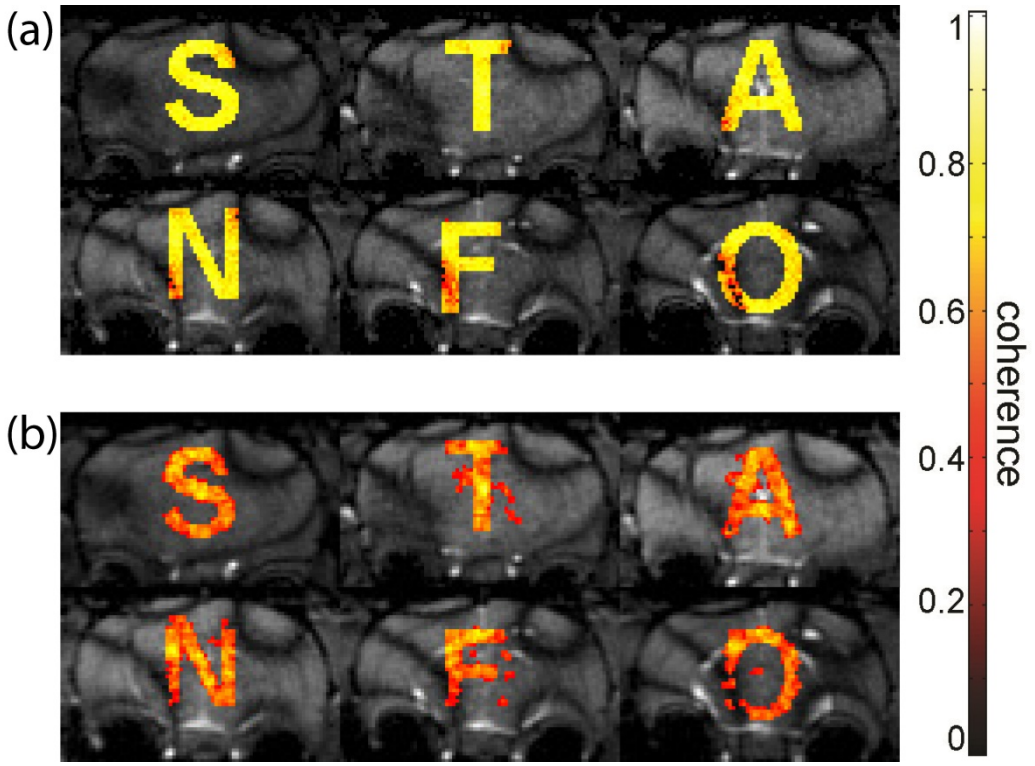
	# of Missed Voxels	# of Leaked Voxels	# of Recovered Voxels	% of Recoverable Voxels	Error Percentage
Phantom	0	0	2261	100%	0%
Reconstruction	67	1193	2194	97.0%	55.7%

Figure 3.5: No additive noise. (a) Activation map of the original phantom. (b) Activation map of the reconstructed image with $3x10^{-3}$ temporal penalty and 10^{-4} spatial penalty.



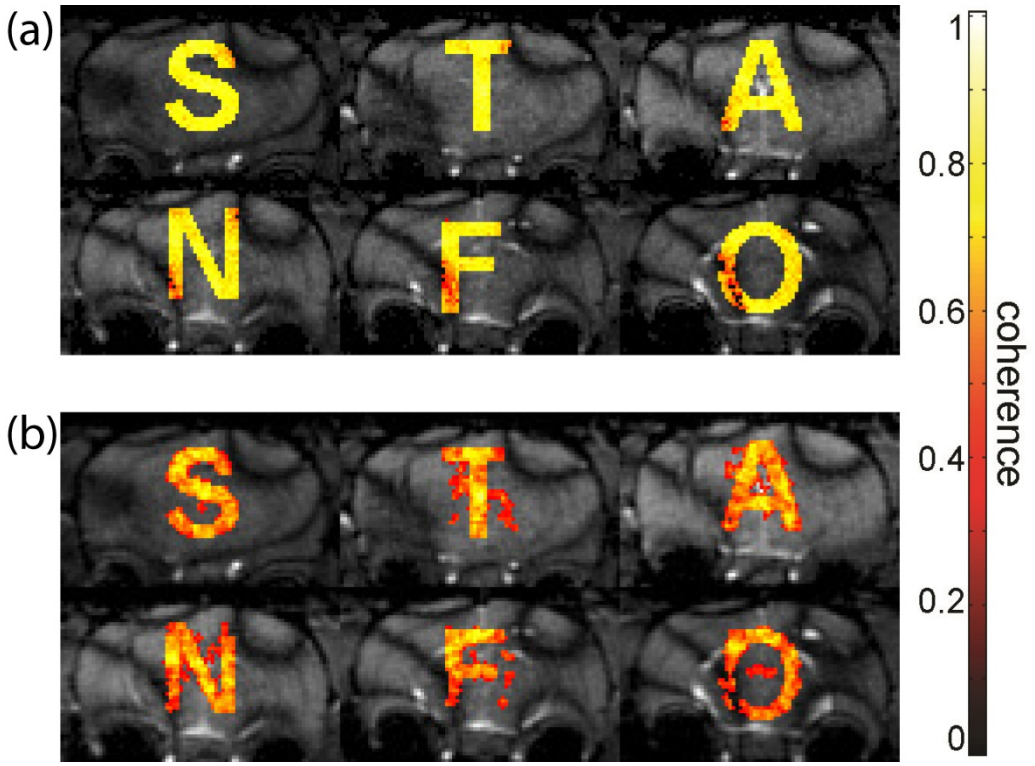
	# of Missed Voxels	# of Leaked Voxels	# of Recovered Voxels	% of Recoverable Voxels	Error Percentage
Phantom	0	0	2261	100%	0%
Reconstruction	1055	109	1206	53.3%	51.5%

Figure 3.6: No additive noise. (a) Activation map of the original phantom. (b) Activation map of the reconstructed image with $8x10^{-4}$ temporal penalty and 10^{-4} spatial penalty.



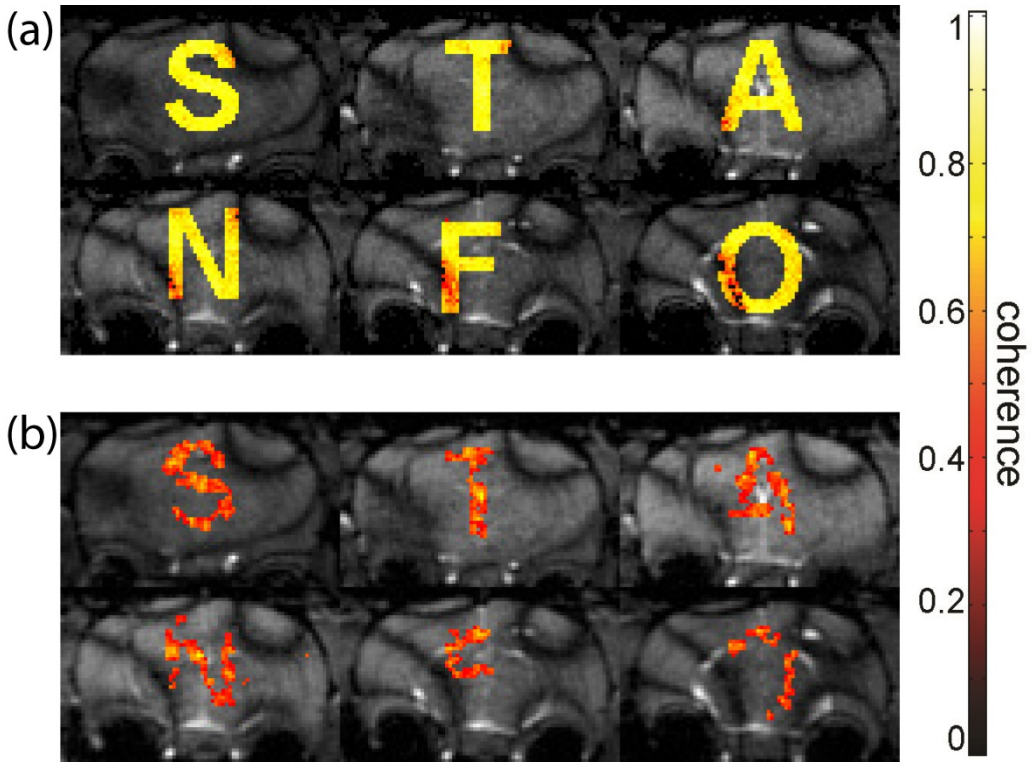
	# of Missed Voxels	# of Leaked Voxels	# of Recovered Voxels	% of Recoverable Voxels	Error Percentage
Phantom	43	2	2218	98.1%	2.0%
Reconstruction	343	337	1918	84.8%	30.1%

Figure 3.7: Rician additive noise, 40 dB SNR. (a) Activation map of the original phantom. (b) Activation map of the reconstructed image with $8x10^{-3}$ temporal penalty and 10^{-4} spatial penalty.



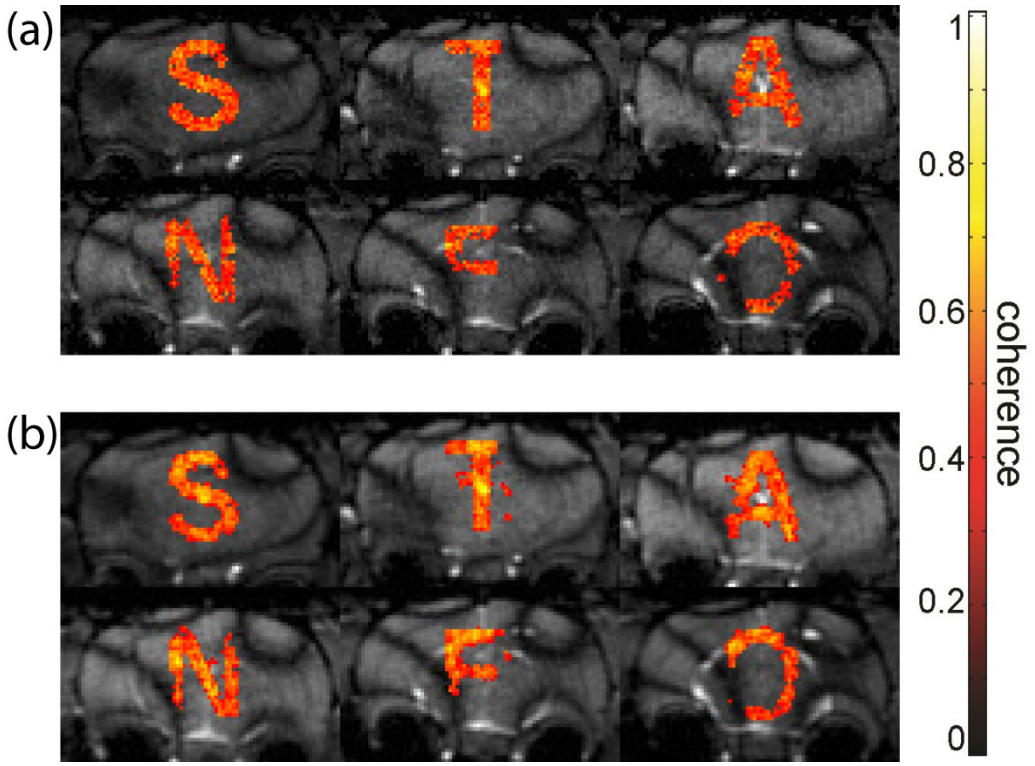
	# of Missed Voxels	# of Leaked Voxels	# of Recovered Voxels	% of Recoverable Voxels	Error Percentage
Phantom	43	2	2218	98.1%	2.0%
Reconstruction	501	238	1760	77.8%	32.7%

Figure 3.8: Rician additive noise, 40 dB SNR. (a) Activation map of the original phantom. (b) Activation map of the reconstructed image with $3x10^{-3}$ temporal penalty and 10^{-4} spatial penalty.



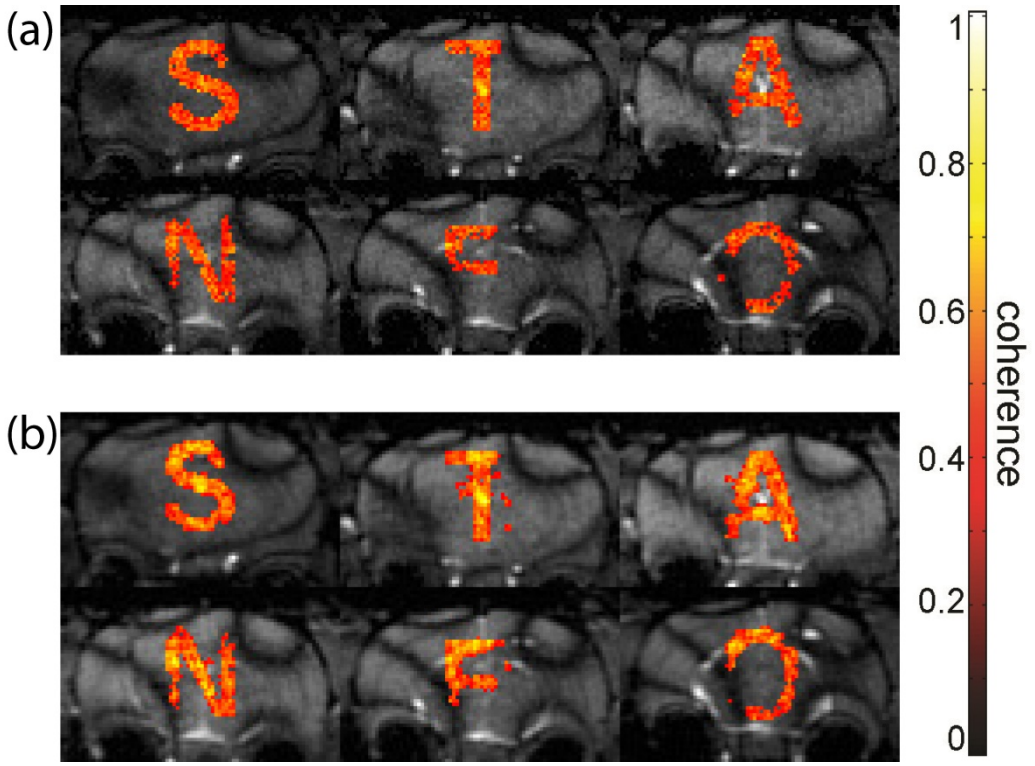
	# of Missed Voxels	# of Leaked Voxels	# of Recovered Voxels	% of Recoverable Voxels	Error Percentage
Phantom	43	2	2218	98.1%	2.0%
Reconstruction	1137	93	1124	49.7%	54.4%

Figure 3.9: Rician additive noise, 40 dB SNR. (a) Activation map of the original phantom. (b) Activation map of the reconstructed image with $8x10^{-4}$ temporal penalty and 10^{-4} spatial penalty.



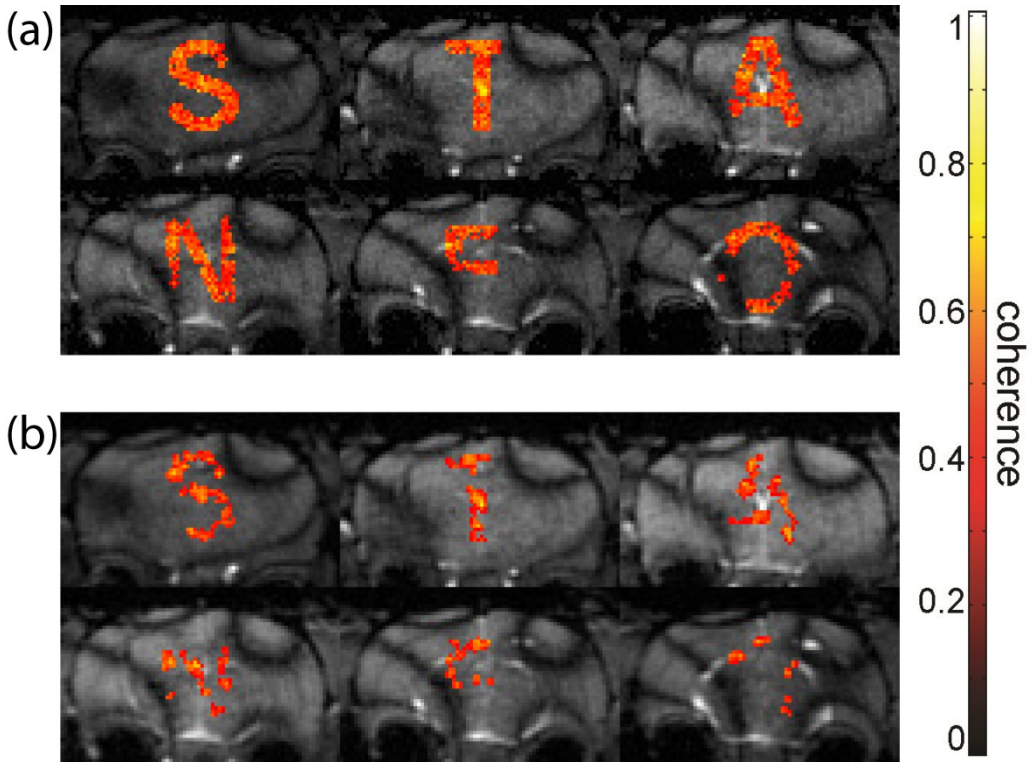
	# of Missed Voxels	# of Leaked Voxels	# of Recovered Voxels	% of Recoverable Voxels	Error Percentage
Phantom	677	1	1584	70.1%	30.0%
Reconstruction	568	202	1693	74.9%	34.1%

Figure 3.10: Rician additive noise, 30 dB SNR. (a) Activation map of the original phantom. (b) Activation map of the reconstructed image with $8x10^{-3}$ temporal penalty and 10^{-4} spatial penalty.



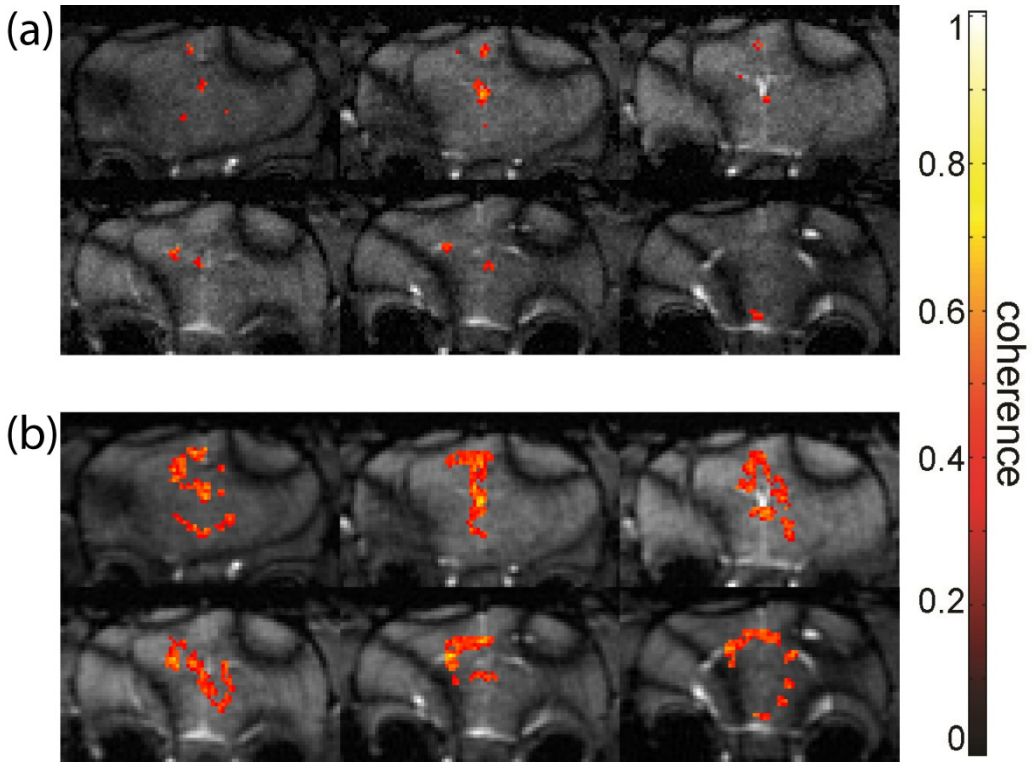
	# of Missed Voxels	# of Leaked Voxels	# of Recovered Voxels	% of Recoverable Voxels	Error Percentage
Phantom	677	1	1584	70.1%	30.0%
Reconstruction	537	232	1724	76.2%	34.0%

Figure 3.11: Rician additive noise, 30 dB SNR. (a) Activation map of the original phantom. (b) Activation map of the reconstructed image with $3x10^{-3}$ temporal penalty and 10^{-4} spatial penalty.



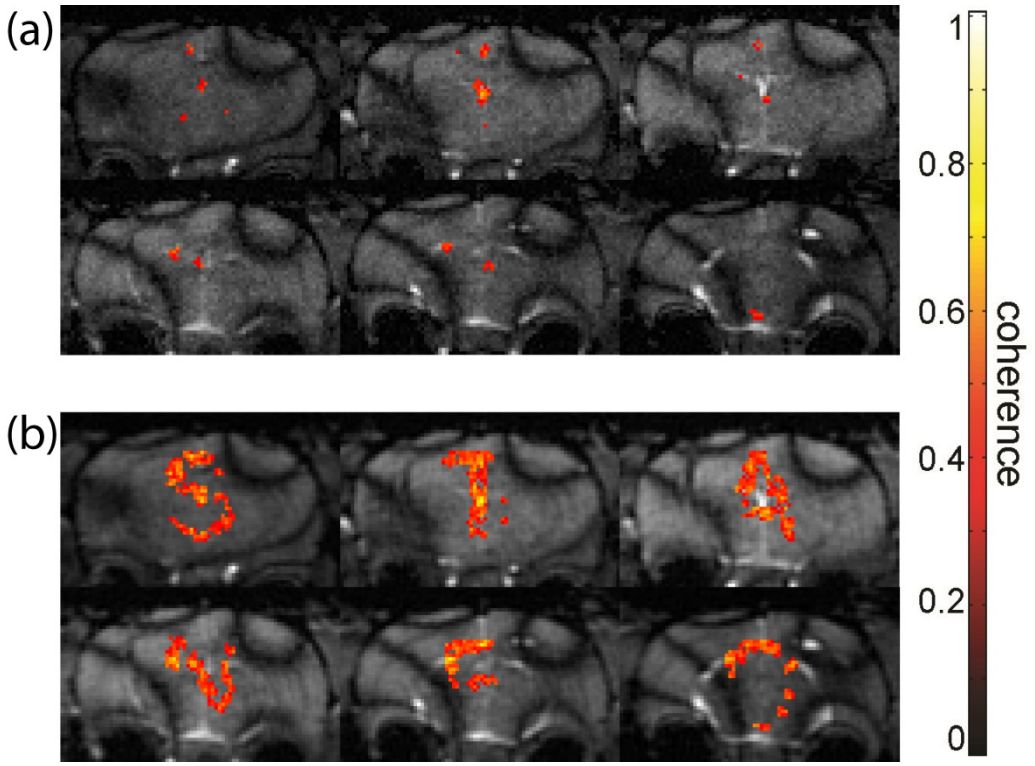
	# of Missed Voxels	# of Leaked Voxels	# of Recovered Voxels	% of Recoverable Voxels	Error Percentage
Phantom	677	1	1584	70.1%	30.0%
Reconstruction	1427	74	834	36.9%	66.4%

Figure 3.12: Rician additive noise, 30 dB SNR. (a) Activation map of the original phantom. (b) Activation map of the reconstructed image with $8x10^{-4}$ temporal penalty and 10^{-4} spatial penalty.



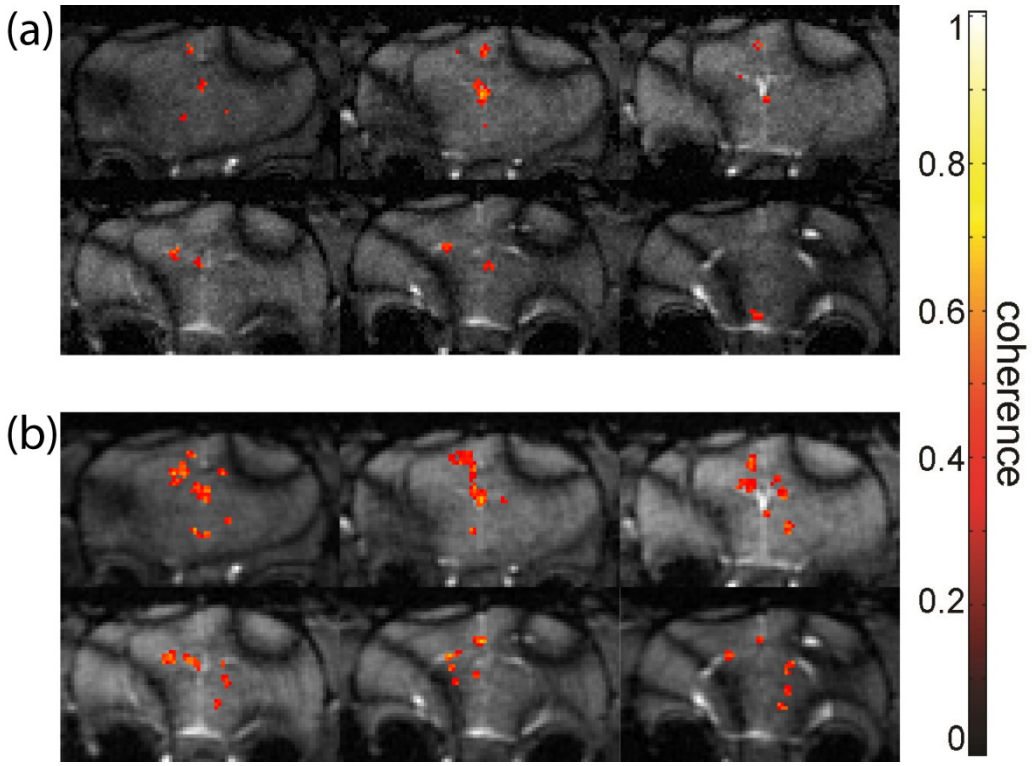
	# of Missed Voxels	# of Leaked Voxels	# of Recovered Voxels	% of Recoverable Voxels	Error Percentage
Phantom	1801	2	460	20.3%	79.7%
Reconstruction	1247	56	1014	44.8%	57.6%

Figure 3.13: Rician additive noise, 25 dB SNR. (a) Activation map of the original phantom. (b) Activation map of the reconstructed image with $8x10^{-3}$ temporal penalty and 10^{-4} spatial penalty.



	# of Missed Voxels	# of Leaked Voxels	# of Recovered Voxels	% of Recoverable Voxels	Error Percentage
Phantom	1801	2	460	20.3%	79.7%
Reconstruction	1129	85	1134	50.2%	53.7%

Figure 3.14: Rician additive noise, 25 dB SNR. (a) Activation map of the original phantom. (b) Activation map of the reconstructed image with $3x10^{-3}$ temporal penalty and 10^{-4} spatial penalty.



	# of Missed Voxels	# of Leaked Voxels	# of Recovered Voxels	% of Recoverable Voxels	Error Percentage
Phantom	1801	2	460	20.3%	79.7%
Reconstruction	1749	61	512	22.6%	80.1%

Figure 3.15: Rician additive noise, 25 dB SNR. (a) Activation map of the original phantom. (b) Activation map of the reconstructed image with $8x10^{-4}$ temporal penalty and 10^{-4} spatial penalty.

	Desired SNR (dB)			
	25 dB	30 dB	40 dB	Noise-Free
Real SNR of the Phantoms	24.3	29.1	38.8	10^7
$\lambda_1 = 8x10^{-3}$, $\lambda_2 = 10^{-4}$	24.6	28.1	32.2	36.6
$\lambda_1 = 3x10^{-3}$, $\lambda_2 = 10^{-4}$	24.0	27.3	31.5	34.4
$\lambda_1 = 8x10^{-4}$, $\lambda_2 = 10^{-4}$	23.3	26.4	29.9	31.3
Naïve Reconstruction with Zero-Filling	18.7	19.6	20.1	20.2

Table 5: SNR of reconstructed images compared to their corresponding phantoms at different noise-level

Table 5 shows the SNRs of reconstructed images. Compared to the real SNR of the phantoms, our reconstructed images with strong activation maps at 30 dB and above have slightly lower SNRs. Below 30 dB, with $\lambda_1 = 8x10^{-3}$, $\lambda_2 = 10^{-4}$, our reconstructed image has a slightly higher SNR than the corresponding phantom. With zero-filling reconstruction, the missing interleaves are filled with zeros and reconstructed with adjoint NFFT to obtain an aliased image. Compared to the SNRs of the aliased phantoms, all reconstructed images have significantly higher SNRs. It shows the ability of signal boosting and noise or noise-like alias suppressing of the compressed sensing method.

Figure 3.16 demonstrates the reconstructed image of noise-free phantom with $\lambda_1 = 8x10^{-3}$, $\lambda_2 = 10^{-4}$ in (a) and its activation leakage map in (b). We can easily see that most of the activation leakage in one slice is caused by its adjacent slices on z-axis. The letter “A” doesn’t cause the leakage to the slices contain letter “S” and “F”, which are two slices away from “A”. The leakage on xy-plane is only one voxel next to the edge of the letter, as clearly illustrated in the error map (figure 3.16 (b)) at the bottom of “S” letter, or at the two sides of the “N” letter, or at the top right corner of the “O” letter. We also notice that the energy of the leaked voxels is lower

than the energy of recovered voxels. At 40-dB noise level, the activation leakage is significantly lower than the noise-free case, as in figure 3.17 (b).

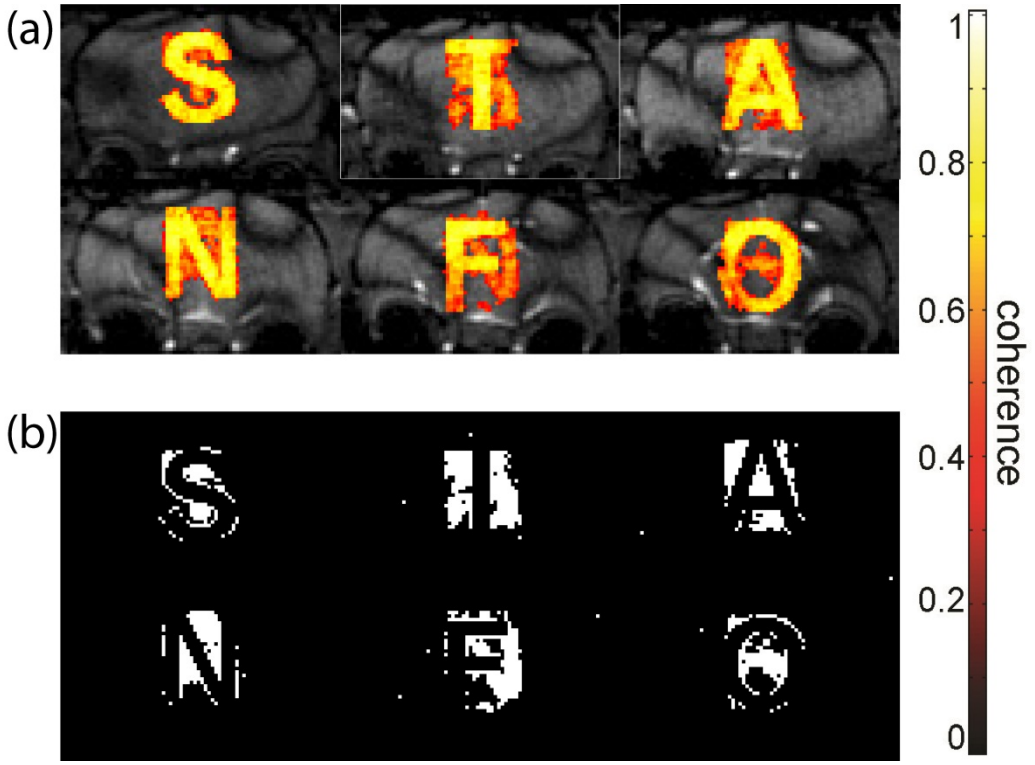


Figure 3.16: (a) Strong activation map of the reconstructed image of noise-free phantom. (b) The activation leakage error map of (a) compared to its original phantom.

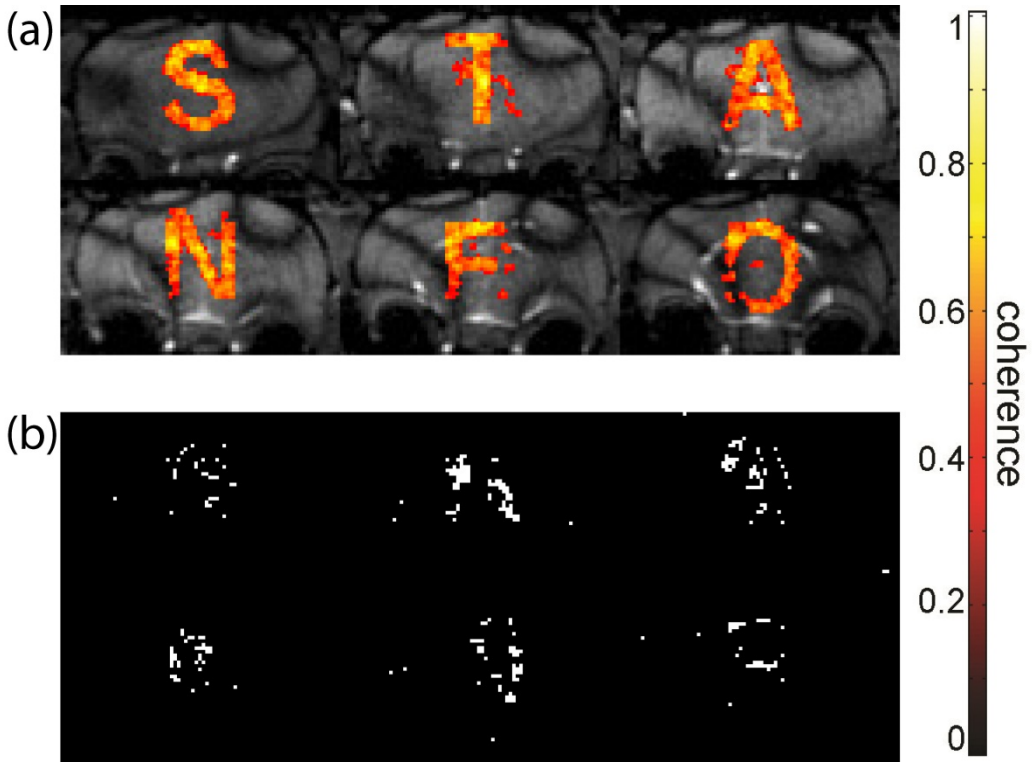


Figure 3.17: (a) Strong activation map of the reconstructed image of 40-dB phantom. (b) The activation leakage error map of (a) compared to the original noise-free phantom.

In practice, most of the activation maps of ofMRI images do not have sharp transitions. In other words, the ofMRI activation maps smoothly change in space. In figure 3.18 (a), we applied smooth gradient to the letter “S” in all three directions (x, y, and z), together with 40-dB additive Rician noise. Figure 3.18 (b) is the strongest reconstructed activation map we obtained. The majority of active voxels are recovered. Figure 3.19 (b) shows the activation leakage error of the reconstructed image. It is obvious that the number of leaked voxels in figure 3.19 (b) is much less than the number of leaked voxels in figure 3.17 (b), which has the same 40-dB SNR. In addition, unlike figure 3.17 (b), the leaked voxels in figure 3.19 (b) are located at the edges of the

real activation maps. Both 3.17 (a) and 3.19 (a) were the strongest reconstructed activation maps we obtained.

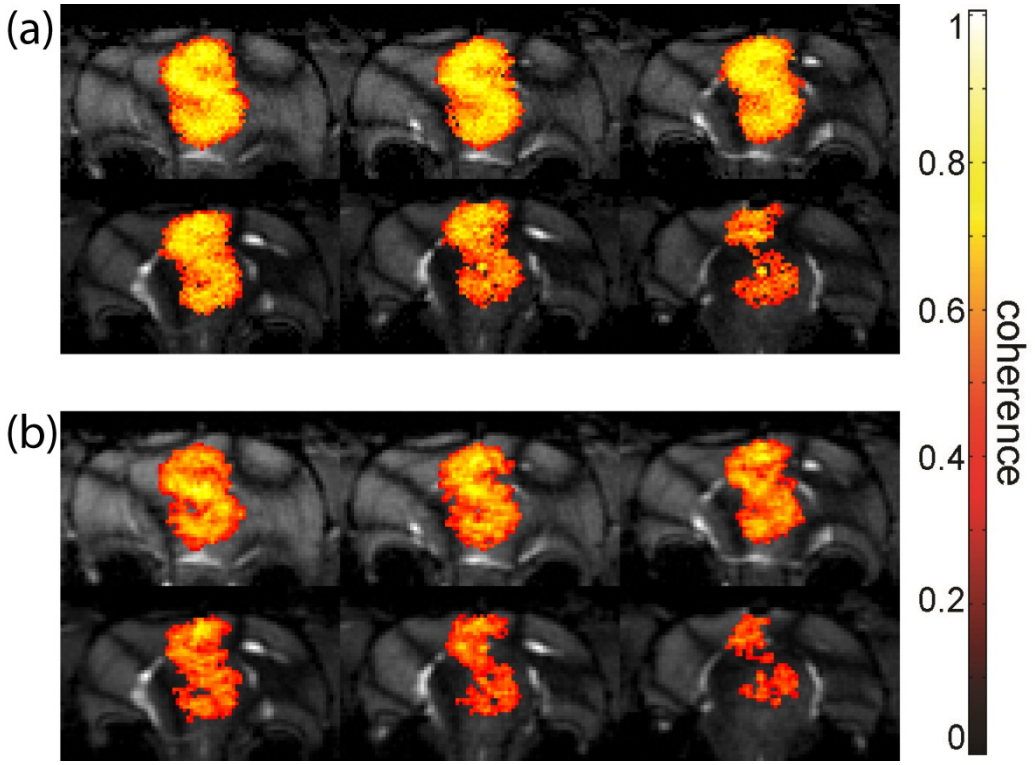


Figure 3.18: Rician additive noise, 40 dB SNR. The activation map is smoothed out in all directions in 3D space. (a) Activation map of the original phantom. (b) Activation map of the reconstructed image with $8x10^{-3}$ temporal penalty and 10^{-4} spatial penalty.

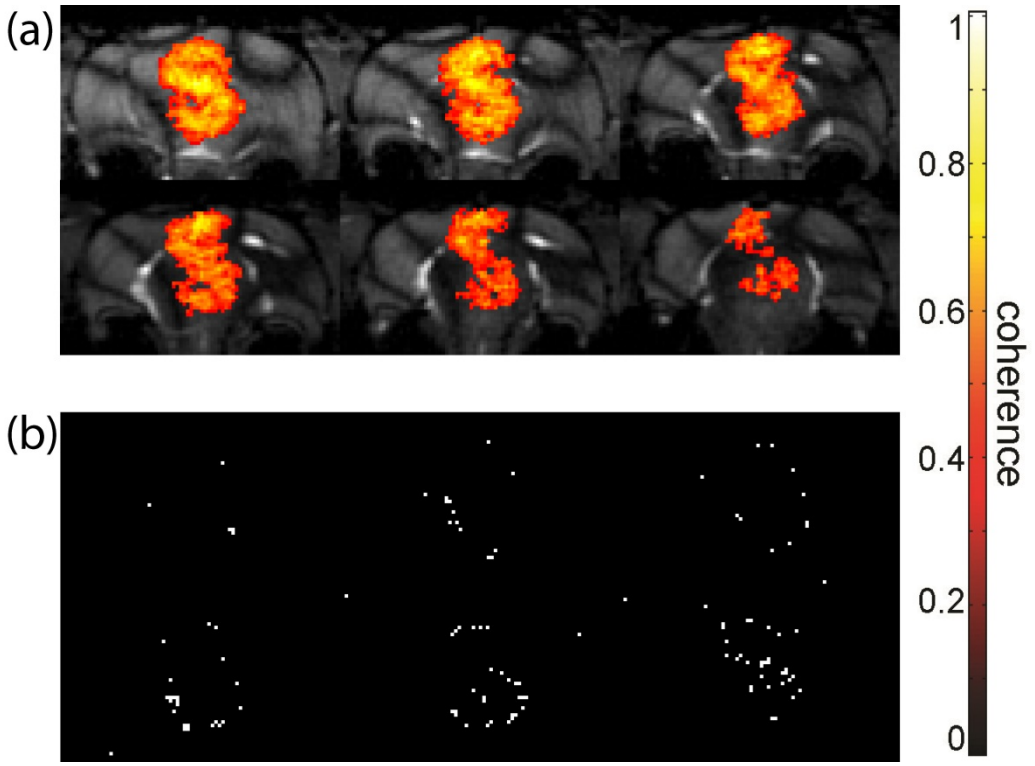


Figure 3.19: (a) Strong activation map of the reconstructed image of 40-dB phantom with smooth activation map. (b) The activation leakage error map of (a) compared to the original noise-free phantom.

CHAPTER 4 – CONCLUSION AND FUTURE WORK

In this thesis, we proposed the combination of our methods in different ofMRI processes to attain high-quality ofMRI images. Our stimulation method allows for precise baseline correction, steady-state scan, and easy activity detection. Our anesthesia level minimizes motion artifacts from data acquisition. The advantage of passband b-SSFP stack-of-spiral trajectory over the conventional GRE-BOLD offers high-SNR images and large brain coverage with only two acquisitions. Our random undersampling sequence promotes incoherent aliasing artifacts while maintaining short scan time and energy-matching purpose. Parallel GPU processing combined with accelerated gradient descent reduces the image reconstruction process from two days to a few minutes. Finally, our motion correction aligns slices in our 4D images to enhance the strength of the activation maps.

With our proposed compressed sensing ofMRI methods, high-resolution and accurate activation maps can be obtained for ofMRI and neurology study. Short reconstruction time, achieved by GPU processing, is necessary to allow our compressed sensing ofMRI method to become a practical application. Because the activation map is very sensitive to the image quality, all processes from experimental set up to image reconstruction and analysis should be performed carefully to minimize artifacts and errors. In real ofMRI, Rician noise always exists and the activation map has smooth changes in spatial domain, which allows for the negligibility of the activation leakage problem. Currently, the optimal weighting parameter sets, which produce strongest activation maps and smallest activation map errors, are determined empirically. For the hippocampus and thalamus stimulation dataset, as well as the phantom experiments' dataset, our compressed sensing ofMRI method exhibits superior noise suppression ability. The activation

maps reconstructed with our compressed sensing method recovers most of the active voxels with accurate locations.

In the future, we can improve the energy matching of our data by designing new sampling trajectory such as non-uniform stack-of-spiral, in which the spiral nodes are denser at the center of the k_{xy}-plane. We can also improve the rate of convergence by using non-linear conjugate gradient instead of our accelerated gradient descent method. Undersampling factor can be further increased to achieve faster scan time or higher-resolution images. The reconstruction process can be performed on two or three graphics cards at the same time to attain faster image reconstruction.

APPENDICES

Appendix I: Forward Nonequispaced Fast Fourier Transform

Input:

- $d, M \in \mathbb{N}, N \in 2\mathbb{N}^d$.
- $x_j \in \left[-\frac{1}{2}, \frac{1}{2}\right]^d, j = 0, 1, \ldots, M-1,$
- $\hat{f}_k \in \mathbb{C}, k \in I_N$

Procedure:

1. For $k \in I_N$, compute
$$\hat{g}_k := \frac{\hat{f}_k}{|I_n| c_k}$$
2. For $l \in I_n$, compute g_l by d-dimensional FFT
$$g_l := \sum_{k \in I_N} \hat{g}_k e^{-2\pi i k (n^{-1} \odot l)}$$
3. For $j = 0, 1, \ldots, M-1$, compute f
$$f_j := \sum_{l \in I_{n,m}(x_j)} g_l \tilde{\psi}\left(x_j - n^{-1} \odot l\right)$$

Output: approximate values $f_j, j = 0, \ldots, M-1$.

Complexity: $\mathcal{O}(|N| \log |N| + M)$

Source: D. Potts' group

Appendix II: Adjoint Nonequispaced Fast Fourier Transform

Input:

- $d, M \in \mathbb{N}, N \in 2\mathbb{N}^d$.
- $x_j \in \left[-\frac{1}{2}, \frac{1}{2}\right]^d, j = 0, 1, \ldots, M-1,$
- $f_j \in \mathbb{C}, j = 0, 1, \ldots, M-1$

Procedure:

1. For $l \in I_n$, compute
$$g_l \coloneqq \sum_{j \in I_{n,m}^T(l)} f_j \tilde{\psi}\left(x_j - n^{-1} \odot l\right)$$
2. For $k \in I_N$, compute by d-dimensional inverse FFT
$$\hat{g}_k \coloneqq \sum_{l \in I_n} g_l \, e^{+2\pi i k \left(n^{-1} \odot l\right)}$$
3. For $k \in I_N$, compute
$$\hat{h}_k \coloneqq \frac{\hat{g}_k}{|I_n| c_k}$$

Output: approximate values $\hat{h}_k, \ k \in I_N$.

Complexity: $\mathcal{O}(|N| \log |N| + M)$

Source: D. Potts' group

Appendix III: Gradient Descent Algorithm

α, β: backtracking line search parameters $(0 < \alpha < 0.5, 0 < \beta < 1)$

m: the numerical approximation of Eq. TBD

f(m): the cost function in Eq. TBD

g(m): the gradient of the cost function $\left(g(m) = \nabla f(m)\right)$

maxIter: number of iterations for stopping criteria

ϵ: stopping criteria by the change in cost function

$k = 0; m = 0;$

while $(k < maxIter)\{$

$g_k = \nabla f(m_k)$; (requires NFFT and iNFFT)

$t = 1;$

while $(f(m_k - tg_k) > f(m_k) - \alpha t g_k^T g_k)\{$

$t = t\beta;$

} (requires NFFT for each loop)

if $\left(\left(\frac{1}{4}\sum_{n=k-3}^{k} f(m_n) - f(m_k - tg_k)\right)/f(m_k - tg_k) < \epsilon\right)$

break;

$m_{k+1} = m_k - tg_k;$

}

Appendix IV: Accelerated Gradient Descent Algorithm

Gradient Descent Algorithm with Pre-computed L2-norm

α, β: backtracking line search parameters $(0 < \alpha < 0.5, 0 < \beta < 1)$

m: the numerical approximation of Eq. TBD

f(m): the cost function in Eq. TBD

g(m): the gradient of the cost function $\left(g(m) = \nabla f(m)\right)$

maxIter: number of iterations for stopping criteria

ϵ: stopping criteria by the change in cost function

$k = 0; m = 0;$

while $(k < maxIter)\{$

$g_k = \nabla f(m_k)$; (requires NFFT and iNFFT)

$fxmy = F_N m_k - y$; ($F_N m_k$ is obtained from the calculation of g_k. No NFFT needed)

$fg = F_N g_k$; (requires NFFT)

$a = \frac{1}{2}\|fxmy\|_2^2;$

$b = \frac{1}{2}\|fg\|_2^2;$

$c = \Re\{\langle fxmy, fg\rangle\};$

$t = 1;$

while $(f(m_k - tg_k) > f(m_k) - \alpha t g_k^T g_k)\{$

$t = t\beta;$

} (requires *a*, *b*, and *c* instead of NFFT for each loop)

if $\left(\left(\frac{1}{4}\sum_{n=k-3}^{k} f(m_n) - f(m_k - tg_k)\right)/f(m_k - tg_k) < \epsilon\right)$

break;

$m_{k+1} = m_k - tg_k;\}$

REFERENCES

[1] Lee et al., “Global and local fMRI signals driven by neurons defined optogentically by type and wiring,” Nature, Vol 465, 10 June 2010.

[2] Lee, J.H., “Tracing activity across the whole brain neural network with optogenetic functional magnetic resonance imaging,” Frontiers in Neuroinformatics, Volume 5, October 2011.

[3] Candes, E.J., Wakin, M.B., “An Introduction To Compressive Sampling,” IEEE Signal Processing Magazine, March 2008,

[4] Lustig, M., Donoho, D., Pauly, J.M., “Sparse MRI: The Application of Compressed Sensing for Rapid MR Imaging,” Magnetic Resonance in Medicine, 2007.

[5] Lustig, M., Donoho, D., Santos, J.M., Pauly, J.M., “Compressed Sensing MRI,” IEEE Signal Processing Magazine, March 2008.

[6] Lustig, M., Santos, J.M., Lee, J.H., Donoho, D., and Pauly, J. “Application of “Compressed Sensing” for Rapid MR Imaging,”

[7] Lee et al., “Full-Brain Coverage and High-Resolution Imaging Capacities of Passband b-SSFP fMRI at 3T, Magnetic Resonance in Medicine, 2008.

[8] Lee, J.H., “Balanced steady state free precession fMRI,” International Journal of Imaging System and Technology, 2010.

[9] Candes, E.J., Tao, T., “Near-Optimal Signal Recovery from Random Projections: Universal Encoding Strategies?” IEEE Transaction on Information Theory, Vol. 52, No. 12, December 2006.

[10] Donoho, D.L., “Compressed Sensing.” IEEE Transaction on Information Theory, Vol. 52, No. 4, April 2006.

[11] Andrei, N., “An acceleration of gradient descent algorithm with backtracking for unconstrained optimization.” Springer Numerical Algorithm, Vol. 42, Number 1, 2006.

[12] Yizhar et al., “Optogenetics in Neural Systems,” Neuron 71, July 14, 2011.

[13] Keiner, J., Kunis, S., and Potts, D. (2008). “Using NFFT 3 – a software library for various nonequispaced fast Fourier transforms.” ACM Transactions on Mathematical Software, Vol. V, No. N, M 2008, Page 1-23.

[14] Keiner, J., Kunis, S., and Potts, D. (2008). NFFT. http://www-user.tu-hemnitz.de/~potts/nfft/.

[15] NVIDIA CUDA C Programming Guide, Version 4.1, http://developer.download.nvidia.com/compute/DevZone/docs/html/C/doc/CUDA_C_Programming_Guide.pdf

[16] CUDA C Best Practice Guide, Version 4.1, http://developer.download.nvidia.com/compute/DevZone/docs/html/C/doc/CUDA_C_Best_Practices_Guide.pdf

[17] Kim, T.S., Lee, J., Lee, J.H., Glover, G.H., Pauly, J.M., “Analysis of the BOLD Characteristics in Pass-Band bSSFP fMRI.

[18] Boyden, E.S., Zhang, F., Bamberg, E., Nagel, G., Deisseroth, K., “Millisecond-timescale, genetically targeted optical control of neural activities,” Nature Neuroscience 8, 1263-1268, 2005.

[19] Zhang, F., Wang, L.P., Boyden, E.S., Deisseroth, K., “Channelrhodopsin-2 and optical control of excitable cells,” Natural Method 3, 785-792, 2006.

[20] Deisseroth et al., “Next-generation optical interrogation of neural circuitry,” Nature 446, 633-639, 2007.

[21] Ogawa et al., “Intrinsic signal changes accompanying sensory stimulation: Functional brain mapping with magnetic resonance imaging,” *Proc. Nalt Acad. Sci. USA* 89, 5951-5955 (1992).

[22] Scheffler et al., “Detection of BOLD changes by means of a frequency-sensitive trueFISP technique: preliminary results,” NMR Biomed, 2001.

[23] Miller et al., “Functional brain imaging using a blood oxygenation sensitive steady state,” Magnetic Resonance in Medicine, 2003.

[24] Candes, E.J., Romberg, J., “Robust Uncertainty Principles: Exact Signal Reconstruction From Highly Incomplete Frequency Information,” IEEE transactions on information theory, Volume 52, No. 2, February 2006.