

Final Report for Exploratory Blue Waters Allocation

March 2016

Project title: Computational Brain Blood Flow Imaging (ILL_jsa)

Principal Investigator:

Brad Sutton, PhD

Associate Professor of Bioengineering

Technical Director, Biomedical Imaging Center, Beckman Institute

1215D Beckman Institute

405 N Mathews Ave

Urbana, IL 61801

Email: bsutton@illinois.edu

Phone: 217-244-5154

Executive summary:

Despite changes in vascular structure with age, adequate brain blood flow is critical for supporting healthy function into old age. Imaging blood flow in the brain using magnetic resonance (MR) imaging requires computational techniques to maximize the signal to noise ratio of the acquired data to compensate for the intrinsically low signal due to the small volume fraction of blood in the brain. As part of this allocation, PowerGrid, a toolkit for accelerating MR image reconstructions using GPUs and distributed computing was created in C++ and OpenACC, while also leveraging MPI to distribute across multiple GPUs. Using PowerGrid, reconstruction algorithms for field inhomogeneity and nonlinear motion induced phase correction were implemented to support our ongoing investigations into changes in brain blood flow with age.

Description of research activities and results:

Key Challenges:

The majority of our reconstruction code currently used to support active research relies on an extensive amount of code written in MATLAB. While high-level MATLAB code provides a flexible and expressive syntax for formulating model-based image reconstruction problems in MR, it does not scale across multiple machines nor is it applicable on HPC platforms like Blue Waters.

In addition, our group's previous work in parallelizing MR image reconstruction using GPUs, the Illinois Massively Parallel Acceleration Toolkit for Image reconstruction with ENhanced Throughput in MRI (IMPATIENT MRI), built in association with Wen-Mei Hwu

and Zhi-Pei Liang, maximized speed-up by formulating the entire model-based reconstruction problem in relatively low-level CUDA code. This approach was successful in speeding up the reconstruction by fully utilizing a single GPU accelerator on a single workstation at the cost of extremely difficult to maintain code, lack of flexibility of implementing novel reconstruction algorithms, and no immediately available path to increasing scale via distributed memory computations.

Diffusion-weighted imaging, used by our team to quantify changes in the microvasculature, presents multiple challenges for iterative, model-based MR image reconstruction. Using efficient, high signal to noise ratio acquisition strategies, such as multi-shot 3D imaging, requires computationally expensive techniques to correct for the inevitable motion induced phase errors in the acquired data along with accurately combining information from multiple coils¹. Additionally, the use of high signal to noise ratio acquisitions can involve the use of trajectories with long readout times, leading to extensive distortion due to magnetic field inhomogeneity². These distortions can be corrected using techniques used by our group, at the cost of significant computation time for an image. Furthermore, given that the percentage of diffusion weighted signal that cerebral blood contributes is extremely small, on the order of 5%,³ reducing computational complexity or cost by accepting increased distortion or motion induced phase error is not acceptable.

Why it Matters:

Sufficient and reactive blood flow in the brain is a critical component for the health of neurons and their supporting cells. However, advanced aging is accompanied by critical changes to the vasculature, including the microvasculature (arterioles and capillaries) that is involved in exchanging nutrients and waste between the blood and tissues. Although vascular density and blood flow changes have been observed, these may be related to architectural changes accompanying age. The coiling, looping, and degradation of microvessels perfusing brain tissue can be seen in both the gray and white matter of aged subjects⁴. Prior to our work, measuring changes and degradations in the microvascular architecture of the human brain was limited to postmortem samples by the destructive nature of microscopy and histology used to quantify the characteristics of the microvasculature. Through development of acquisition, iterative model-based image reconstruction, and estimation procedures, we are able to capture the weak signal changes due to the degradation of the vascular structure with age. Fig. 1 shows a representative vascular perfusion fraction map and pseudodiffusion coefficient for blood flow for a slice high in a young human brain. These maps require the acquisition of 14 separate diffusion weighted acquisitions in order to characterize the very small blood flow signal as reflected in the vascular perfusion map.

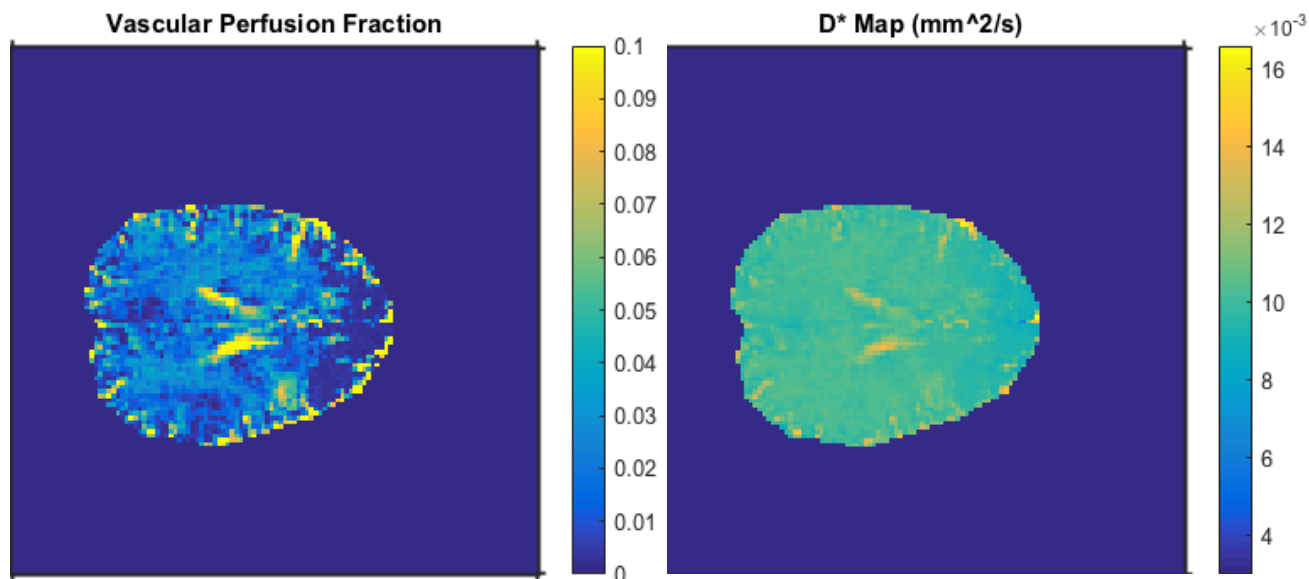


Figure 1: (left) Vascular perfusion fraction reporting amount of blood flowing in the microvasculature in each voxel. Notice that the number is small, less than 10% in all but voxels dominated by cerebrospinal fluid, and much less than 5% for most of the voxels in the brain parenchyma. (right) Pseudodiffusion coefficient (D^*) for the same slice in the brain. D^* reflects the mean square displacement of the blood flow during the MR image acquisition. Higher D^* reflects more blood motion on average, while lower D^* reflects less blood motion. Taken together, the perfusion fraction and D^* can reflect blood flow.

Why Blue Waters:

For each blood flow dataset, we need to reconstruct multiple slabs of tissue, from multiple imaging volumes, from 10-20 measurements that vary the diffusion weighting and are received from 32 parallel receiver coils arranged around the head. This creates a large amount of data that is well-suited to parallel implementation across GPU-equipped nodes. Blue Waters provides access to a large number of GPU-equipped compute nodes with high speed interconnect ideal for distributed memory computations. In addition, the OpenACC support present in the Cray Programming Environment makes Blue Waters ideal for translating MR reconstruction algorithms.

Accomplishments:

Our team developed PowerGrid, a toolkit for accelerating iterative, model-based MR image reconstructions via GPUs. Implemented in C++, PowerGrid allows us to retain the familiar structure that is a result of years of work developing algorithms for advanced MR image reconstruction in MATLAB while leveraging HPC resources, such as Blue Waters and OpenACC.

The object structure of PowerGrid was designed to combine state of the art techniques for model-based image reconstruction in MR and resources available on Blue Waters, namely MPI and OpenACC. This approach was necessary as MR physicists are not trained in parallel computing languages, but traditionally do the majority of software development for reconstructions inside an interactive high-level language like MATLAB. Fig. 2 shows how PowerGrid maintains familiarity with the code as typically written in MATLAB but enables the use of OpenACC accelerated routines. Other routines, not shown, enable the use of MPI for distributed computation.

IRT/MATLAB SENSE Operator	PowerGrid/C++ SENSE Operator
<pre> %Forward Operator if ~S.is.transpose outData = zeros(nx, ncoils); for ii = 1:ncoils outData(:, ii) = G*(SMap(:, ii). * d(:)); end else %G*d is a forward NUFFT/DFT end %Adjoint Operator d = reshape(d, [], ncoils); outData = 0; for ii = 1:ncoils outData = outData + (conj(SMap(:, ii))) * (G' * d(:, ii)); end %G'*d is an adjoint NUFFT/DFT end </pre>	<pre> //Forward Operator Col <CvTl> operator*(const Col <CvTl> & d) const { Mat <CvTl> outData = zeros<Mat<CvTl> >> (this->nx, this->ncoils); for (unsigned int ii = 0; ii<this->ncoils; ii++) outData.col(ii) = (*this->G_obj) * (d * (this->SMap.col(ii))); return vectorise(outData); //G_obj*d is a forward NUFFT/DFT } //Adjoint Operator Col <CvTl> operator/(const Col <CvTl> & d) const { Mat <CvTl> inData = reshape(d, this->nx, this->ncoils); Col <CvTl> outData = zeros<Col<CvTl> >> (this->ny); for (unsigned int ii = 0; ii<this->ncoils; ii++) outData += conj(this->SMap.col(ii)) * ((*this->G_obj) / inData.col(ii)); return vectorise(outData); //G_obj/d is an adjoint NUFFT/DFT } </pre>
IRT/MATLAB Examples	PowerGrid/C++ Examples
<pre> %Forward SENSE Operator d = S * img; %Adjoint SENSE Operation imgProj = S' * d; %Gradient Descent SENSE Recon N = floor(sqrt(length(imgIt))) img = 1/(2*N) * (A' * data) % Form image estimate for ii = 1 : num_iterations dataEst = (1/(2*N)) * (A * img); % Form scaled k-space estimate imgError = (1/(2*N)) * (A' * (data - dataEst)); % Image space error img = img + imgError; %Update Image end </pre>	<pre> //Forward SENSE Operation Col<CvTl> d = S * img; //Adjoint SENSE Operation Col<CvTl> imgProj = S / img; //Gradient Descent SENSE Recon uword N = floor(std::sqrt(xInitial.n_rows)); Col<CvTl> dataEst, imgError, img = (1/(2*N)) * (A / data); //Form image estimate for(int ii = 0; ii < num_iterations; ii++) { dataEst = (1/(2*N)) * (A * img); // Form scaled k-space estimate imgError = (1/(2*N)) * (A / (data - dataEst)); //Image space error img = img + imgError; //Update Image } </pre>

Figure 2: Diagram showing correspondence between code from the Iterative Reconstruction Toolkit (IRT) on the MATLAB platform and the code implemented in PowerGrid during the course of this allocation. PowerGrid wraps the GPU accelerated numerical transforms in syntax that approximates the widely used Iterative Reconstruction Toolkit.

Using PowerGrid, we have shown speed up factors of up to $\sim 11\times$ above the single GPU case through the use of MPI for distributed computing as shown in Fig. 3 for a small benchmark data set. We anticipate increased speed up factors as the dataset size and complexity increases. Our use of MPI is enabled both by the code written in PowerGrid and the exploitation of natural parallelism that exists in the SENSE approach to parallel imaging in MR imaging⁵. Parallel imaging uses multiple coils and receiver channels to increase the speed of acquisition. These separate streams of data each provide a natural work unit that can be assigned to a GPU, although the parallelism is not complete, requiring the use of MPI communication routines to share results before performing another global iteration of the image reconstruction.

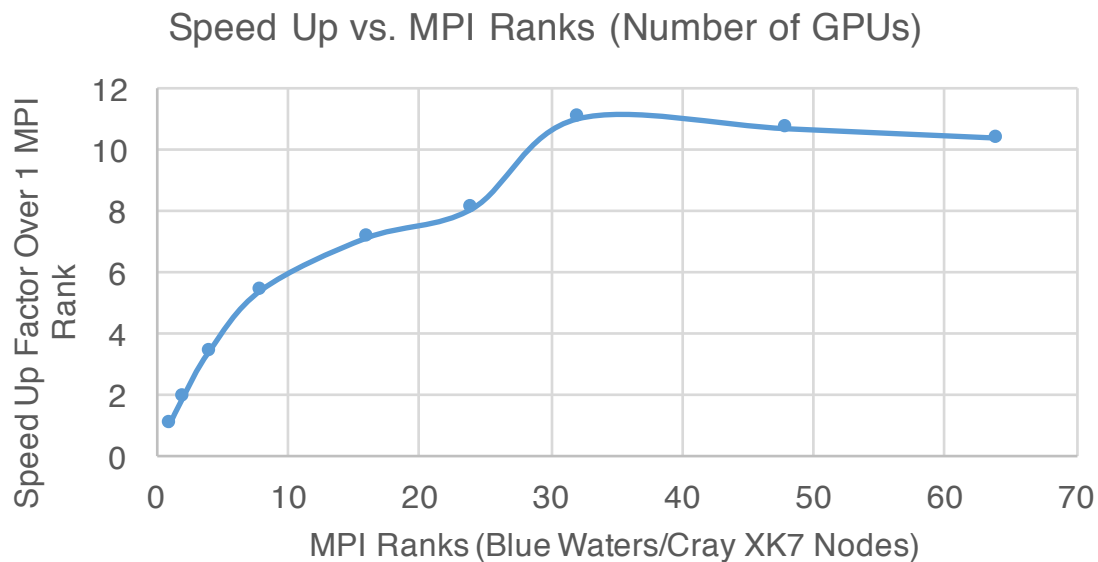


Figure 3: Results showing speed up versus number of MPI ranks (K20x GPUs) on Blue Waters showing peak speed of up ~11x with 32 nodes and saturation for additional ranks for a small benchmark case distributed with PowerGrid.

List of publications associated with this work:

Cerjanic, A., Holtrop J.L., Ngo, G.C., Leback, B., Arnold, G., Van Moer, M., LaBelle, G., Fessler, J.A., Sutton, B.P., “PowerGrid: A open source library for accelerated iterative magnetic resonance image reconstruction.” ISMRM Annual Meeting 2016, Singapore, May 7th-13th.

References:

1. Holtrop JL, Sutton BP. Higher spatial resolution in diffusion weighted imaging on clinical 3 T MRI scanners using multi-slab spiral acquisitions. SPIE Journal of Medical Imaging.
2. Sutton BP, Noll DC, Fessler JA. Fast, iterative image reconstruction for MRI in the presence of field inhomogeneities. IEEE Transactions on Medical Imaging. 2003;22(2):178–188.
3. Le Bihan D, Breton E, Lallemand D, Aubin ML, Vignaud J, Laval-Jeantet M. Separation of diffusion and perfusion in intravoxel incoherent motion MR imaging. Radiology. 1988;168(2):497–505.
4. Fang H. Observations on Aging Characteristics of Cerebral Blood Vessels, Macroscopic and Microscopic Features. In: Neurobiology of Aging. 1st Edition. New York: Raven Press; p. 155–166.

5. Pruessmann KP, Weiger M, Scheidegger MB, Boesiger P. SENSE: Sensitivity encoding for fast MRI. *Magnetic Resonance in Medicine*. 1999;42(5):952–962.