IMPATIENT MRI: Illinois Massively Parallel Acceleration Toolkit for Image reconstruction with ENhanced Throughput in MRI

X-L. Wu[1], J. Gai[2], F. Lam[1,2], M. Fu[1,2], J. P. Haldar[1,2], Y. Zhuo[2,3], Z-P. Liang[1,2], W-M. Hwu[1,2], and B. P. Sutton[2,3]
[1]Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, United States, [2]Beckman Institute, University of Illinois at Urbana-Champaign, Urbana, IL, United States, [3]Bioengineering Department, University of Illinois at Urbana-Champaign, Urbana, IL, United States

Introduction: Significant progress has been made in optimizing imaging trajectories and reconstruction approaches for a variety of MRI application areas. Often these acquisitions result in high resolution 2D or 3D data acquired with non-Cartesian trajectories that have been significantly under-sampled. Image reconstruction is then performed either in unacceptably long times or with highly optimized code that includes several approximations and interpolation steps in order to keep the reconstruction time low. In this work, we introduce the Illinois Massively Parallel Acceleration Toolkit for Image reconstruction with ENhanced Throughput in MRI (IMPATIENT MRI) package to speed image reconstruction to acceptable times while still taking advantage of a variety of advanced image acquisitions and reconstruction techniques. We provide an open source, highly optimized implementation on graphics processing units (GPUs) which allow for massively parallel computation to greatly reduce image reconstruction times.

Method: IMPATIENT MRI is currently equipped to solve reconstruction problems of the form:

$$\hat{\boldsymbol{\rho}} = \arg\min_{\boldsymbol{\rho}} \frac{1}{2}\left\|\mathbf{E}\boldsymbol{\rho} - \mathbf{d}\right\|_2^2 + \beta\left\|\mathbf{W}\mathbf{C}\boldsymbol{\rho}\right\|_2^2 \quad (1)$$

where $\boldsymbol{\rho}$ is the image to be reconstructed, $\mathbf{E}$ is a matrix that models the acquisition physics, $\mathbf{d}$ is the vector of measured data, $\boldsymbol{C}$ is a regularization matrix, $\boldsymbol{W}$ is an optional diagonal weighting matrix, and β is a regularization parameter. The current implementation of IMPATIENT MRI includes handling of non-Cartesian acquisitions, magnetic field inhomogeneity correction [3], multiple coil acquisition and SENSE reconstruction [4], and regularization (including both Tikhonov and l_1-based regularization strategies). Unlike traditional methods [3], IMPATIENT MRI makes *no approximation* in the computation involving forward ($\boldsymbol{E}$) and adjoint operators ($\boldsymbol{E}^H$) for non-Cartesian trajectories and field inhomogeneity. All computations are done using brute-force. OpenMP-optimized C++ implementation is used as a performance baseline, running on a multi-core processor system. Our reconstruction workstation is a dual-socket Intel Xeon E5520 system, with one Tesla C2050 (Fermi) GPGPU card. The GPU optimization involves algorithm level optimizations (tiled computation, loop fusion, loop distribution) and statement level optimizations (common sub-expression elimination, reuse of computation results, loop unrolling with/out increase registers) [5]. Since our dominating kernels are forward and adjoint operators, these optimizations are aiming to reduce instructions, hide memory latency, and reduce memory bandwidth. The IMPATIENT MRI code is achieving a high degree of parallelism at the instruction-level, loop-body level, warp-level, and algorithm-level. Data was simulated based on the abdominal scan (Double Vision) from the 2010 ISMRM Data Reconstruction Challenge, using the spiral acquisition with 8 channel receivers and a magnetic field inhomogeneity map [1]. In order to simulate different matrix sizes, the 320x320 image from a single slice of the data set was simulated using various k-space trajectories after adding noise to the image and interpolating to the correct image resolution. The spirals were designed according to [2] using a maximum gradient amplitude of 22 mT/m and a maximum slew rate of 140 mT/m/ms. We simulated matrix sizes of [128,256,512] using multishot spiral with [4,8,20] shot spirals, respectively. The 8-channel sensitivity maps and magnetic field inhomogeneity map was used from the double vision data set.

Results: The performance comparison is shown in Table 1. We test the four datasets on the Fermi GPU card. All reconstructions execute 20 conjugate gradient iterations. The 'Time (sec)' row shows the execution timings of both CPU-based and GPU-based iterative regularized SENSE reconstructions using single-precision, floating-point arithmetic. In summary, we observed a speedup of up to 640x in single precision mode with normalized root mean square error less than 10^{-3} compared to CPU reconstruction for 256x256 data size. Using double precision is considerably slower than single precision (e.g. 5x slower for 256x256), though still much faster than CPU (~128x speedup). Figure 1 shows three GPU reconstruction results on the 512x512 dataset under different experimental settings. Figure 1a is obtained using only one coil without field correction; Figure 1b is the result using all 8 coils (SENSE) without field correction; Figure 1c is the result using all 8 coils (SENSE) with field correction. As seen in Figure 1, through the use of the IMPATIENT MRI Toolkit, iterative field-corrected non-Cartesian SENSE reconstruction is completed significantly faster than before with dramatically improved image quality. For comparison with existing methods, we reconstructed the 256x256 matrix size dataset with a common approximation method using SENSE and the NUFFT to approximately handle non-Cartesian data and time segmentation to handle magnetic field inhomogeneity as in [3]. This reconstruction took 2098.2 s, approximately 40 times slower than the GPU reconstruction which does not require setting parameters for approximations.

Dataset	128x128 (SENSE,8coils)			256x256 (SENSE,8coils)			512x512 (SENSE,8coils)		
	CPU	GPU	Speedup	CPU	GPU	Speedup	CPU	GPU	Speedup
Time (sec)	872.27	4.644	188x	34007.28	53.142	640x	Too Long	843.67	N/A

Table 1: A comparison of the runtimes of the OpenMP implementation on the CPU side with the CUDA implementation on the GPU side (single-precision mode).

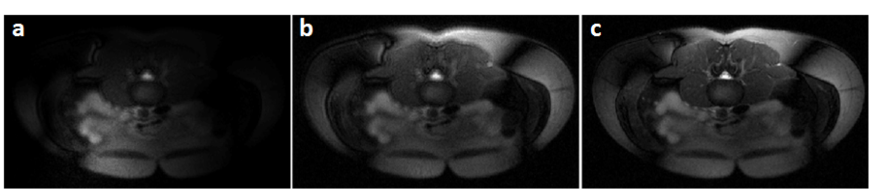

Figure 1: Image reconstruction results for a) single coil reconstruction without field inhomogeneity correction, b) SENSE reconstruction without field inhomogeneity correction, c) SENSE reconstruction with field inhomogeneity correction.

Discussion:
IMPATIENT MRI provides computation capabilities for both clinical and research imaging practitioners, enabling advanced reconstruction schemes with considerably larger data sets and more coils than typical CPU-based reconstructions. Also, it enables the flexibility and possibility to try multiple reconstruction settings for fine tuning the reconstruction quality due to the large increase in reconstruction speed. A speedup of two orders of magnitude is achieved compared to a parallel CPU implementation and a speedup by a factor of 40 is obtained compared to a common approximation technique of NUFFT and time segmentation. IMPATIENT MRI allows for clinical feasible reconstruction times while taking advantage of efficient non-Cartesian acquisitions, parallel imaging, magnetic field inhomogeneity correction, and incorporation of *a priori* information. Additionally, multiple slice acquisitions may be further accelerated by distributing reconstructions of different slices to different GPUs.

References: [1] Data Reconstruction Challenge. 2010 *Intl Soc Magn Reson Med*. (ismrm.org/mri_unbound) [2] Glover. *Magn Reson Med*. 1999 Aug;42(2):412-5. [3] Sutton, *et al*. *IEEE Trans. Med. Imaging*, vol. 22, pp. 178-188, 2003. [4] Pruessmann, et al. *Magn. Reson. Med*., vol. 42, pp. 952-962, 1999. [5] D.B. Kirk, *et al*. *Morgan Kaufmann*, 2010.

Acknowledgements: This work was supported by NIH grant 1R21EB009768-01A1.