Clinically Feasible Reconstruction Time for L1-SPIRiT Parallel Imaging and Compressed Sensing MRI

M. Murphy[1], K. Keutzer[1], S. Vasanawala[2], and M. Lustig[1]
[1]EECS, UC Berkeley, Berkeley, CA, United States, [2]Radiology, Stanford University, Stanford, CA, United States

Introduction: It has been shown that L1-SPIRiT [1], a method that combines parallel imaging and compressed sensing, has potential for reducing MRI scan times and improving image quality [2]. However, the clinical applicability of the technique is limited by the computational difficulty of the reconstruction. Reconstruction must be nearly instantaneous, but a straightforward implementation of L1-SPIRiT's non-linear minimization requires hours to complete. However, recent GPGPUs and multi-core CPUs have shown great potential for accelerating computations in many application areas, including in MRI reconstruction [3,4,5]. In this work, we demonstrate a fast implementation of L1-SPIRiT reconstruction that achieves clinical runtime; reconstruction of the 4-fold poisson disc accelerated 3D abdominal MRI in Figure 1 completes in only 97 seconds.

Theory: L1-SPIRiT reconstruction requires solving a non-linear constrained optimization problem: minimize $||\mathbf{W}x||_1$ subject to $||\mathbf{G}x - x||_2 < \varepsilon$ and $\mathbf{F}_\Omega x = y$. The solution x has a maximally sparse wavelet representation, is a fixed point of the calibration operator **G**, and is consistent with the undersampled k-space measurement. We solve the minimization via Projections Onto Convex Sets (POCS). We iteratively project a solution onto three sets: images which are (a) wavelet-sparse, (b) calibration-consistent, and (c) measurement-consistent. Runtime is dominated by the least-squares computation of the calibration operator **G**, the wavelet transforms and soft-thresholding in (a), the filtering operations for computing **G**x for (b), and the Fourier transforms for (c).

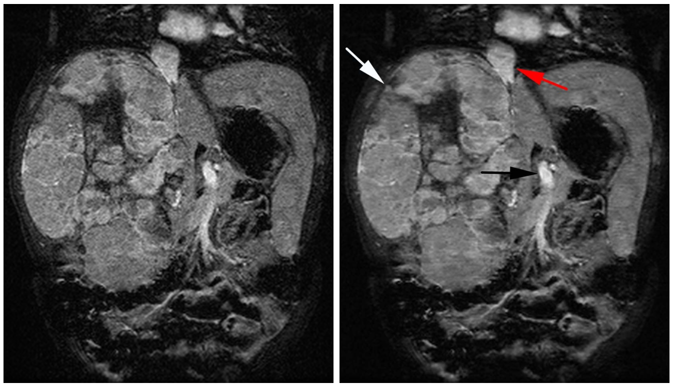

Figure 1: Slice from an 8-channel abdominal MRI of an 8-month old patient, reconstructed using both a GE ARC PI product (left) and our L1-SPIRiT algorithm (right), which ran in 97 seconds. A hepatoblastoma (white arrow) abuts and compresses the inferior vena cava (red arrow). Note the superior mesenteric vessels (black arrow).

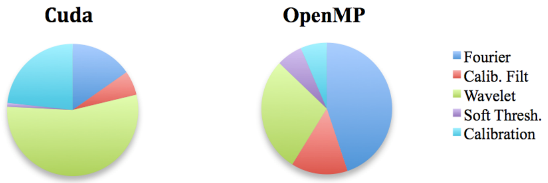

Figure 2: Breakdown of the recon runtime by sub-computation. Total Cuda per-slice runtime is 140 ms calibration + 480 ms L1-SPIRiT iterations; OpenMP runtime is 140ms + 2.1s. Fourier transforms, Calibration Filtering, and Thresholding are much faster on the GPU, where Wavelets dominate the runtime. The CPU cache greatly benefits Wavelet transforms, for which the GPU is only 1.8x faster.

Methods: We have implemented L1-SPIRiT using two sets of extensions to C++, OpenMP [6] and Nvidia's Cuda [7]. For convenience, Matlab scripts handle file I/O and some minor computations with negligible overhead. We use fast FFT libraries: FFTW 3.2.2 [8] and CuFFT 2.3 [7]. Our evaluation platform is an 8-core 2.8 GHz Intel Xeon E5462 machine with a 30-core, 8-SIMD 1.48 GHz GTX285 GPGPU. On both the CPU and GPGPU, efficiency requires identifying data parallelism matched to available cache resources. Assigning each of 8 cores a different slice of an 8-channel 3D scan produces a working set of nearly 40 MB. Our CPU has 12 MB of cache, and the GPU only 480 kB; better cache utilization is achieved by parallelizing the reconstruction of a single slice. For the data-parallel soft-thresholding and calibration operators we partition the data equally among cores. In our CPU solver, each core performs the Fourier and Wavelet transforms for a single channel. In our GPU solver's Wavelet transform, each core computes convolutions for a block of rows or columns of a channel, and each core's SIMD lanes share the work of each convolution.

Results: Figure 1 shows a slice of a 3D, 4-fold accelerated (poisson disc sampling, 8 channel array) post-contrast acquisition of an 8 month old patient using an SPGR sequence implemented on a GE 750 3T scanner. The runtime of our OpenMP calibration is 22 seconds (140 ms per slice). 40 iterations of our OpenMP POCS solver run in 334 seconds (2.1 seconds per slice). Our Cuda POCS solver runs in 75 seconds (480 ms per slice) - 4.5x faster. Our GPU wavelet implementation is bandwidth-inefficient: a more highly optimized implementation will be up to 50% faster. Also, multi-GPU parallelization will provide additional 3-4X speedup. Using our OpenMP calibration and our Cuda POCS solver results in 97-second reconstruction time.

Conclusion: In this work we have implmented the L1-SPIRiT algorithm for two parallel processing platforms, tailoring the implementation to each architecture. Our solvers are scalable to larger image sizes, more channels, and to larger processing platforms. We found that GPGPUs are more efficient than CPUs for this type of reconstruction. We have demonstrated the first clinically feasible reconstruction time of a combined autocalibrating parallel imaging and compressed sensing reconstruction.

References: [1] M. Lustig et al., ISMRM '09 [2] Vasanawala et al. SPR '09 [3] S. Stone et al, CF '08 [4] F. Knoll et al., ISMRM '09 [5] Hansen et al., MRM 2008 [6] www.openmp.org [7] www.nvidia.com [8] www.fftw.org