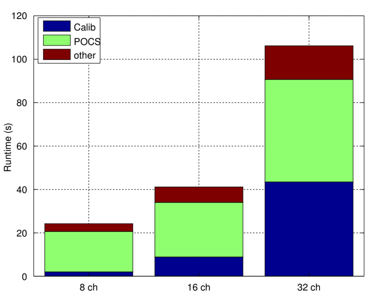
Fast l1-SPIRiT Compressed Sensing Parallel Imaging MRI: Scalable Parallel Implementation and Clinically Feasible Runtime

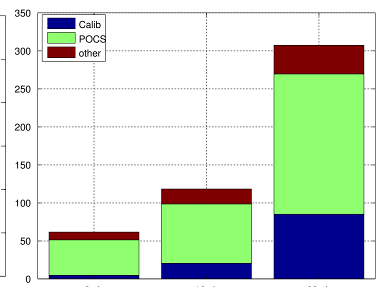
Mark Murphy[1], Miki Lustig
[1]*EECS UC Berkeley, Berkeley, CA, United States*

Introduction: We present a detailed performance analysis of our clinically-deployed GPGPU implementation of ℓ1-SPIRiT [1], a combined Parallel Imaging (PI) and Compressed Sensing (CS) reconstruction. Preliminary results have been presented in a previous abstract [2]. This work extends the analysis to consider alternate parallelizations of the algorithm to suit different data and processor sizes. Additionally, we describe an algorithmic improvements for SPIRiT autocalibration. Our ℓ1-SPIRiT implementation achieves sub-minute runtimes for clinical 3D reconstructions, and has been available to our clinical collaborators for on-line reconstructions continuously for over a year.

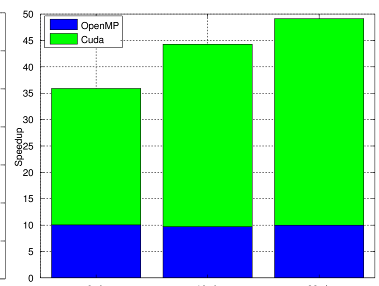
Methods: Achieving efficient parallelization on modern parallel processors requires an efficient exploitation of the hierarchical sources of parallelism in MRI computations. For small data sizes, or for processors with large numbers of cores, coarse-grained parallelizations (multiple slices in flight simultaneously per GPU) will incur less synchronization overhead. For image matrix sizes a fine-grained parallelization (1 slice per GPU) will require more frequent syncrhonization, but achieve better caching performance. We also compare algorithms for performing SPIRiT autocalibration. With n_c channels, the naïve approach solves the calibration independently for each channel and has complexity $O(n_c^4)$. Our approach that re-uses a single Cholesky factorization for all channels has complexity $O(n_c^3)$. Our modified SPGR 3DFT pulse sequence undersamples y and z phase encodes in a variable-density poisson-disc distribution. All data were acquired with 32-channel torso coils, and we present reconstruction times for the 32-channel data and coil-compressed 8- and 16-channel data. Performance results were gathered on a dual-socket, six-core (12 cores total) 2.67 GHz Intel Westmere system with four Nvidia GTX 580 graphics processors (GPUS).



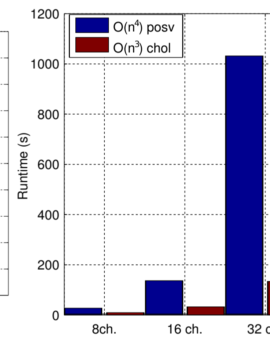
Avg. Runtime, readout 192



Avg. Runtime, readout 320

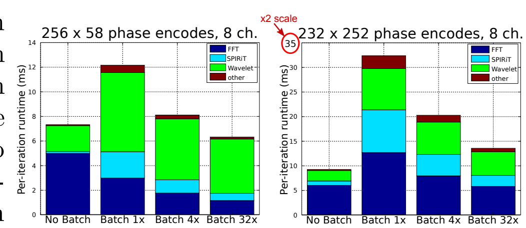


Avg. Speedup over sequential



Calib. Runtimes

Results: Runtime is proportional to readout length, so we present 8,16, and 32 channel reconstruction times for readout lengths 192 and 320 separately. An OpenMP-parallel CPU implementation achieves 10× speedup over sequential on average, while the Cuda achieves 35-50× speedup. The Cholesky-based algorithm reduces calibration runtime by over 5×, and is crucial to efficient runtime for large coil arrays. The sub operations within the reconstruction respond differently to the alternate parallelizations. Fourier transforms require frequent synchronization, and thus prefer a coarse parallelization (Bached), while the wavelet transform implementation can take advantage of caching in a fine-grained parallelization (No batch).



Alt. Parallelization

References: [1] M. Lustig and J.M. Pauly, *SPIRiT: Iterative self-consistent parallel imaging reconstruction from arbitrary k-space*, MRM 2010 [2] M. Murphy *et al.*, *Clinically Feasible Reconstruction Time for L1-SPIRiT Parallel Imaging and Compressed Sensing MRI*, Proc. ISMRM 2010