GPU-based Sparse Matrix Beamformer for High-Frame-Rate Ultrasound Imaging

Di Xiao[1], Billy Y. S. Yiu[1], Adrian J. Y. Chee[1], Alfred C. H. Yu[1], [1]Schlegel Research Institute for Aging, University of Waterloo, Waterloo, Canada

Background, Motivation and Objective

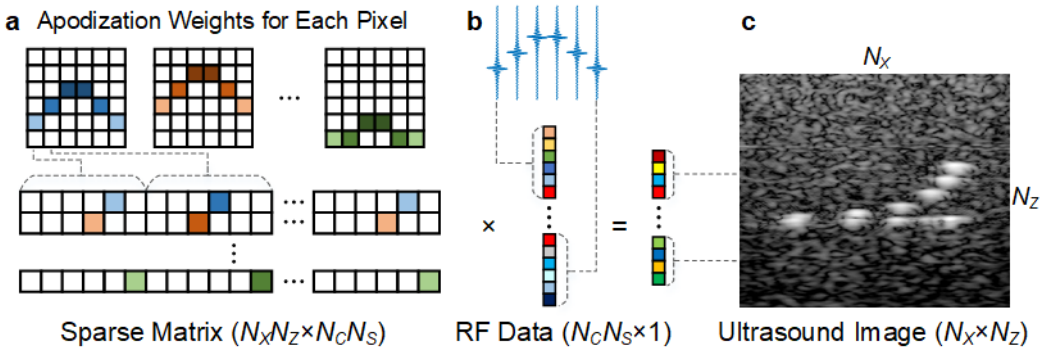
One bottleneck in achieving live high-frame-rate imaging (HiFRUS) is the computational load due to the delay-and-sum (DAS) beamforming operations required to generate each pixel for all frames (100+ million pixels in 1000 frames per second, fps) in real-time. Researchers have recently turned to parallel computing solutions such as graphics processing units (GPUs) to address these computational needs. Many of these approaches, however, require nontrivial GPU programming optimizations for parallelization and efficient memory use which are crucial to achieving real-time performance. Instead, we show that DAS beamforming can be reformulated as a sparse-matrix multiplication to attain highly optimized GPU usage by using an out-of-the-box CUDA library.

Statement of Contribution/Methods

Fig. 1 gives an overview of the presented framework. Fig. 1a (top) shows the apodization matrix to form one particular pixel. This matrix was vectorized as illustrated in Fig. 1a to form one row of the sparse matrix. This was repeated for all pixels in the $N_X \times N_Z$ image to generate the final sparse matrix. This sparse matrix was then multiplied with the vectorized RF data (Fig. 1b) using the CUDA cuSPARSE library to generate the final image in vector form which was then reshaped (Fig. 1c). The method was tested by beamforming a 512x256 image from a 128-channel 3120-sample unsteered plane wave acquisition of an imaging phantom, resulting in a matrix of size 131072x399360 at a sparsity level of 0.00036%. Execution time was measured using CUDA profiler and compared with previously reported GPU beamformer (T-UFFC, 2011, 58: 1698-1705).

Results/Discussion

Timing results showed that our sparse matrix beamformer generated an image frame in 0.668 ms, corresponding to a throughput of 1500 fps. While the performance of sparse matrix multiplication did not exceed the optimized GPU-based beamformer (0.070 ms/frame), almost no GPU programming knowledge was necessary – no threads, blocks, or kernels were exposed – to achieve 95% occupancy on the GPU and still maintain over 1000 fps. Additionally, given that DAS beamforming is data agnostic, the sparse matrix needs only to be formed once for each plane wave transmit and receive pair. Overall, our work shows that sparse matrix beamformer is a practical alternative to bypass the high technical threshold required to achieve live HiFRUS imaging.



N_X number of lateral pixels, N_Z number of axial pixels,
N_C number of channels, N_S number of samples