Combined Algorithmic and Hardware Acceleration for Ultra-Fast Backprojection

Jeffrey Brokish and Yoram Bresler

***Abstract*–We describe the first implementation and performance of a fast O(N^2logN) hierarchical backprojection (FHBP) algorithm on a field programmable gate array (FPGA) platform. The resulting prototype tomographic backprojection system for 2D fan-beam geometry combines speedup through algorithmic improvements provided by the FHBP algorithm, with speedup of a specialized hardware platform.**

For data parameters typical in diagnostic CT, and using an off-the-shelf FPGA evaluation board, we report normalized reconstruction speeds of 160 frames a second, and relative speedup of 25x compared to conventional backprojection on the same hardware. These results demonstrate that the reduction in operation counts demonstrated previously for the FHBP algorithm can be translated to a comparable actual run time improvement in a massively parallel hardware implementation, while preserving stringent diagnostic image quality. The dramatic speedup and throughput numbers achieved indicate the feasibility of systems based on this technology, which achieve real-time 3D reconstruction for state-of-the art diagnostic CT scanners at low power, small footprint, high-reliability, and affordable cost.

I. INTRODUCTION

BACKPROJECTION is the computational bottleneck in the filtered backprojection (FBP) algorithm used in CT reconstruction. FBP algorithms are O(N^3) algorithms for 2D reconstruction. They have a very simple form, where each pixel in the image is determined by an accumulation of contributions from each projection in the acquired scan data. FBP falls into the class of 'embarrassingly parallel' algorithms, where the work can be readily divided among multiple execution units. This results in significant speedup when porting the algorithm to a parallel hardware platform.

Fast reconstruction algorithms reduce the complexity of the backprojection operation to O(N^2logN), offering the potential for tremendous speedups in computation. Fast hierarchical backprojection (FHBP) [1] operates by recursively dividing the image into smaller and smaller subregions, reducing the amount of projection data with each division. When reaching a prescribed size (in the limit a single pixel), standard backprojection is invoked to form the chunk of the image from a drastically reduced set of projections. Analysis has shown [2] that combining the work of decomposing the image space and the final backprojection step reduces the complexity of the reconstruction algorithm.

This work has been supported by an NIH NIBIB SBIR Phase I grant No. 1 R43 EB005067-01 to InstaRecon, Inc.

Jeffrey Brokish is with InstaRecon, Inc., Urbana, Illinois, USA.

Yoram Bresler is with the Coordinated Science Lab and the Department of Electrical and Computer Engineering at the Univerisity of Illinois at Urbana-Champaign, USA.

Specialized hardware platforms are particularly attractive in that the inherent reduction of computation of FHBP can be combined with the speedup potential of a parallel platform to deliver unprecedented reconstruction rates. Our goal was to map the FHBP algorithm to an FPGA platform, maintaining speedup over conventional FBP and combining algorithmic and hardware speedup to yield an ultra-fast backprojection engine.

Challenges encountered in the FPGA design:

- FHBP more complex than FBP
- Limited temporary memory space on the development board (1 MB SRAM)
- Limited memory bandwidth on the development board (1 Bank DRAM, 1 Bank SRAM)
- Fixed Point representation
- Layout and Design to maximize throughput and utilization of the FPGA chip

II. FAST HIERARCHICAL BACKPROJECTION

The FHBP algorithm for fan beam CT [2] operate using two main concepts. The *first concept*, is that the number of projections needed to accurately reconstruct a bandlimited subimage at the center of the source of rotation, is proportional to the size of the subimage. It follows that for the reconstruction of a half-size subimage, the projection data set can be angularly decimated by a factor of 2, from P to P/2 projections without any loss in reconstructed image quality. This property is extended to apply to subregions located at any position in the image. First, projections of the subimage are "centered" by shifting in the projection plane to the position of the footprint of the center of the source rotation, as shown in Fig. 1. This reduces their angular bandwidth, so that centered projections may be decimated without information loss. The decimated projections are then shifted back to their correct positions. The combined operation of centering, angular decimation, and de-centering is denoted by the acronym SDSB (Shift-Decimate-Shift Back), summarized in Fig. 2. Applying the SDSB operation results in a factor of 2 savings in backprojection of that subimage.

The *second concept* is "divide and conquer" in which the reconstructed volume is successively divided into smaller non-overlapping volumes, with the SDSB operation applied to each subimage. At every stage in this hierarchical decomposition, the number of projections is reduced by another factor of 2. Applying it logN times yields N^2 single pixel subimages, which require backprojection of P/N projections each, for a total of O(NP) work in the last stage. It turns out that the work for the SDSB operations in each

step is also O(NP), yielding a total cost of $O(N^2 \log N)$ for the hierarchical algorithm, since P = O(N) for good image quality.

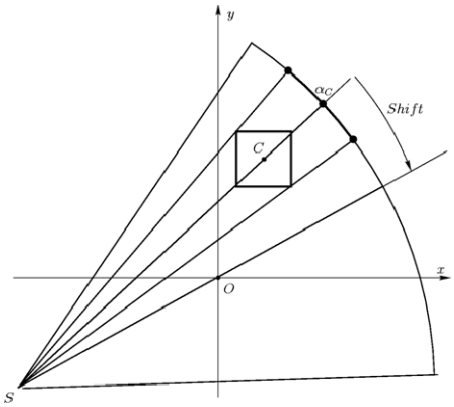

Fig. 1. "Centering" Shift for FHBP

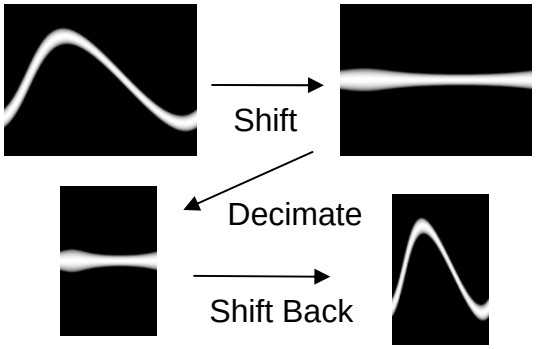

Fig. 2. Shift-Decimate-Shift Back Operation

III. FPGA Design

Our selected FPGA testbed was the Xilinx ML402 evaluation board, which consists of the Virtex 4 SX35 FPGA, 64 MB of DDR DRAM and 1 MB of ZBT SRAM. While the board provides fairly modest memory bandwidth, the SX35 supports substantial computational bandwidth having 192 hardware multipliers.

A block diagram of the on-board implementations for the FBP and FHBP algorithm are shown in Figs 3-4. The ML402 operates as a stand-alone board, so the flash memory is used to transfer projection and image data to and from the board. Initially, projection data is loaded from the flash memory into the DRAM. The design is initialized and begins fetching projection data from the DRAM. The ZBT SRAM is used to store temporary intermediate calculations. As the image reconstruction process nears completion, image pixels are written to the DRAM. Once the design signals completion, the image data is transferred from the DRAM back to the flash memory for verification and analysis. Transfers to/from flash memory were not counted in performance calculations.

The implementation of FBP exploits the projection-parallelism of the backprojection operation. Here, 8 backprojection pipelines operate on the same pixel, using neighboring projections. The results are accumulated together via an adder tree, and added to a working copy of the image in the SRAM.

While FHBP has more complex control flow and different operations than found in standard FBP, it still contains a significant amount of parallel work. As the algorithm proceeds, the reconstruction of each subimage becomes independent from one another. The Decomposition Unit performs this breakdown of the projection data, and is able to process multiple subregions simultaneously. Once the final subregion size is reached, data is sent to the backprojection unit (BPU) for reconstruction of that subregion. In this case the BPU can be specially tailored to the small image sizes present in the hierarchical decomposition.

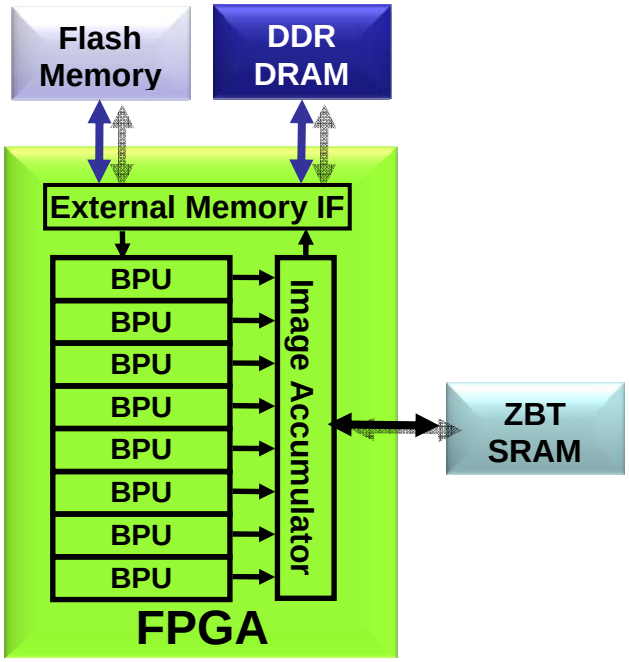

Fig. 3. Block diagram of FPGA system organization for FBP

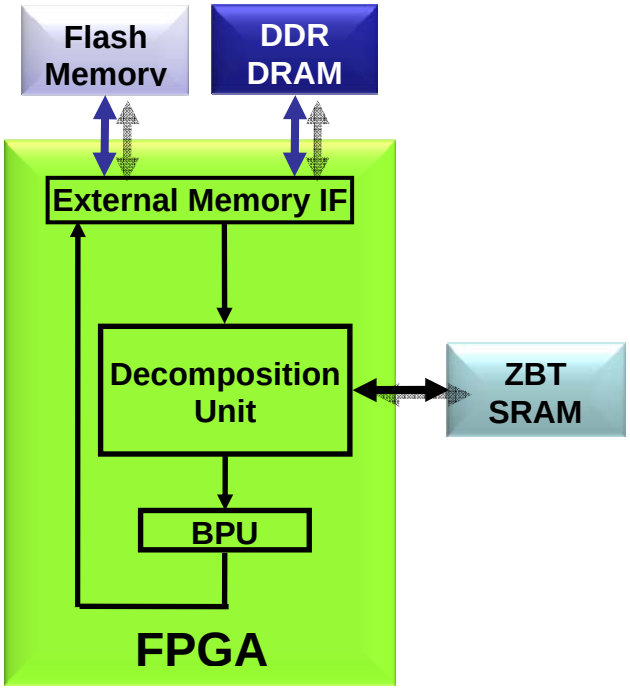

Fig. 4. Block diagram of FPGA system organization for FHBP

IV. FHBP Performance Analysis

The goal of the FHBP algorithm is to increase reconstruction rates while preserving image quality relative to the conventional backprojection operation. A suite of test images was assembled to demonstrate comparable image quality between images produced by FHBP and conventional reconstruction methods. This suite is comprised of a number of synthetic phantoms and clinical images, as shown in Fig 5. Synthetic tests included a water cylinder (for CT number uniformity and noise levels), a low contrast phantom, a high contrast phantom, and a set of point targets for PSF/MTF analysis. Results of the MTF test are shown in Fig 7. A clinical test suite, shown in Fig 6, was assembled from representative slices of various parts of the body. Projection data was simulated for a set of clinical image slices for qualitative image quality assessment. Results were supplied to radiologists in DICOM format so the images could be compared for different window levels and widths. In a blind IQ comparison, images were found to have comparable clinical diagnostic quality.

The ultimate aim of an FPGA-based FHBP algorithm is delivering high reconstruction rates. The FHBP design outlined in Figure 1 was implemented on the FPGA and optimized for speed. The current design attains a clock rate of 200 MHz. A limiting factor is the memory bandwidth provided by the local ZBT SRAM on the particular development board chosen (the Xilinx ML-402), which is rated at a maximum speed of 200 MHz. By using a different board with more banks of SRAM, which would increase the available memory bandwidth, the design could be adapted to reconstruct image slices up to twice as fast on the same FPGA chip. A hardware-based implementation of the conventional algorithm was designed for the same hardware platform to provide a benchmark for measuring speedup. This implementation consisted of eight parallel backprojection pipelines and was also clocked at 200 MHz.

Image sizes were 512x512 pixels reconstruction from 1024 projection. The 16-bit projection data was preloaded into the on-board DRAM prior to the execution of each algorithm. The reconstructed 16-bit image is written to the DRAM and then downloaded to the host computer for analysis. The reconstruction rates reported are based on the DRAM-to-DRAM execution time -- device to host transfer times are not included. Reconstruction rates are 10.7 ms/slice for FHBP and 168 ms/slice for FBP, providing throughputs of 93.3 and 6.0 frames per second, respectively, resulting in a speedup of over 15x. With regard to the FPGA resources, the FHBP design uses approximately 57% of the hardware multipliers, whereas conventional FBP utilizes 91% of the hardware multipliers of the SX35. Therefore, for an appropriate comparison we normalize the reconstruction rate by the resource utilization, which provides an upper bound on the performance of each algorithm. Here the FHBP delivers over 160 frames per second, speeding up reconstruction rates by almost 25x.

Xilinx's emerging Virtex 5 product line has even more computational power, with their top of the line DSP FPGA, the SX95T, having 640 hardware multipliers. A version of the FHBP algorithm for this chip, having full multiplier utilization and an increased clock speed of 300 MHz, would have a throughput of 800 frames per second! A high-end CT scanner with a 128 row detector, operating with a pitch of 1.5 at 3 revolutions per second, is able to scan a volume at a rate of about 600 slices per second. An FHBP implementation on the SX95T would have sufficient throughput to deliver real-time 3D reconstruction for such a machine.

Performance Metric	**Conventional FBP (8-way parallel)**	**FHBP**
Multiplier Utilization	91%	57%
Reconstruction Speed (frames/sec)	6.0	**93.3**
Speedup Factor		**15.5x**
Normalized Reconstruction Speed (frames/sec)	6.6	**165**
Normalized Speedup Factor		**25x**

Table 1. Performance Metric Comparison Conventional FBP vs FHBP

V. Conclusion

These results indicate an unprecedented speedup of the FHBP algorithm compared to the conventional algorithm on the same platform. Additionally, these impressive reconstruction rates were achieved on the off-the-shelf reconfigurable hardware development board used in these experiments. These speedups, coupled with the positive results of the image quality evaluation, demonstrate the feasibility of an ultra-fast, high quality hardware image reconstruction engine based on FHBP algorithms. Performance projections based on the next generation of FPGAs indicate the potential of a real-time, low-cost 3D reconstruction system.

Acknowledgment

The authors would like to thank Drs. Mike Vannier (University of Chicago) and Benoit Desjardins (University of Michigan) for supplying clinical images and image quality evaluations.

References

[1] S. Basu and Y. Bresler, "An $O(N^2 \log N)$ Filtered Backprojection Reconstruction Algorithm for Tomography," *IEEE Transactions on Image Processing*, pp. 1760–1773, Oct. 2000.

[2] S. Xiao, Y. Bresler, and D. Munson, "$O(N^2 \log N)$ Native Fan-Beam Tomographic Reconstruction," in *Proceedings 1st IEEE International Symposium on Biomedical Imaging, ISBI-2002*, (Washington, DC), pp. 824–827, July 2002.

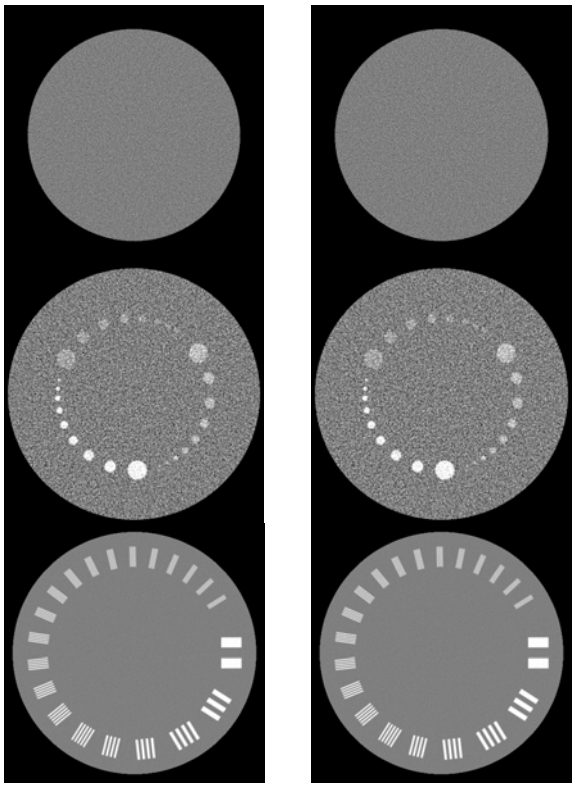

Fig. 5. Quantitative Test Suite comparing Conventional FBP (left column) with Hierarchical Backprojection (right column). Tests include CT Number Uniformity and Noise (top row), Low Contrast Detectability (middle row), and High Contrast Resolution (bottom row).

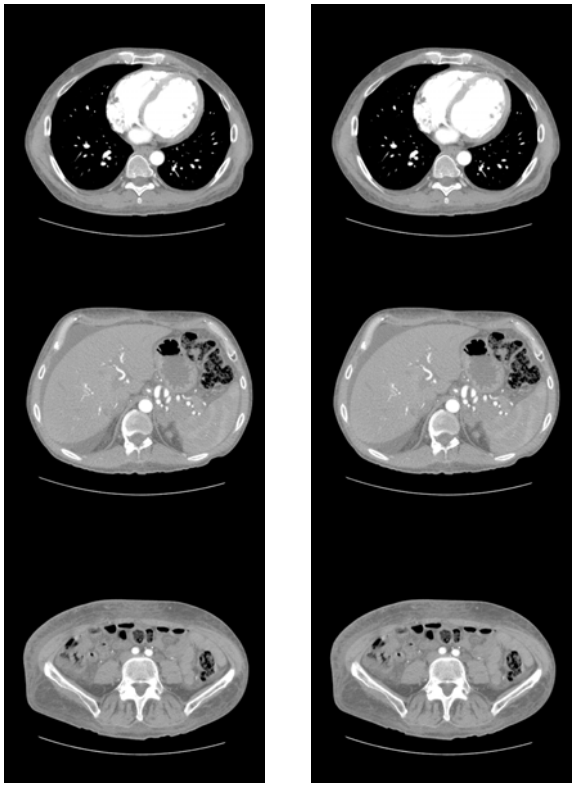

Fig. 6. Qualitative Test Suite comparing Conventional FBP (left column) with Hierarchical Backprojection (right column). Tests include Chest (top row), Abdomen (middle row), and Pelvis (bottom row) slices from a clinical scan.

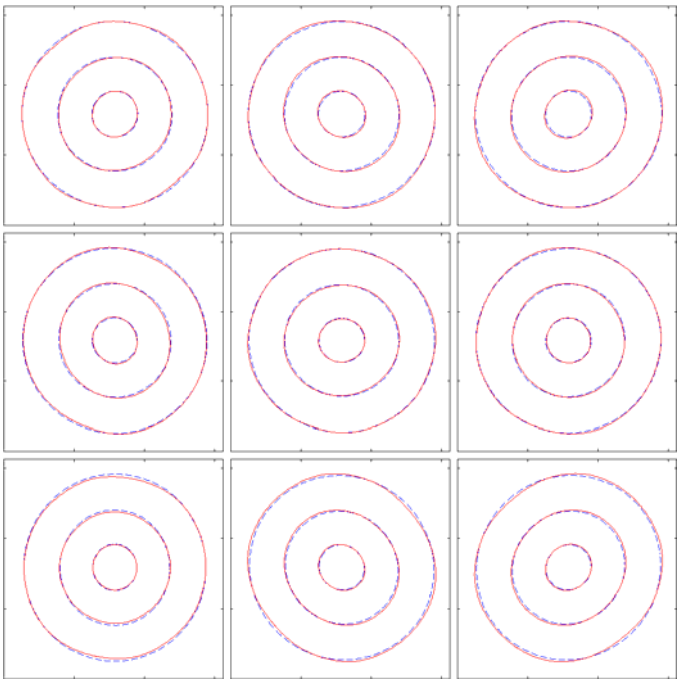

Fig. 7. MTF Comparison for an test arrangement of point targets. Red, solid lines are conventional FBP; blue, dashed lines are hierarchical FHBP. Contours are drawn at 90%, 50%, and 10% of maximum.