SAND2018-1569C

Convolutional Backprojection on the S1 Reduced Precision Processor

Michael Holzrichter and Randy Spaulding

Summary
Sandia investigated reduced precision computing as a way to increase the performance of processing raw radar data in a way that does not rely on Moore's Law. If less precision can be tolerated, then more parallelism can be obtained from the same number of transistors. This means increased processing performance without increasing size, weight and power (SWaP).

A core effort was an evaluation of the S1 processor from Singular Computing. Sandia evaluated the performance of its S1 implementation of the convolutional backprojection (CBP) algorithm. Sandia produced good-quality SAR images using reduced precision computations, demonstrating reduced precision computation's utility.

Keywords
Reduced precision computation, convolutional backprojection, SAR image formation, S1 processor, Moore’s Law

Introduction

With the expectation that Moore's Law cannot drive improvements in processor performance forever, Sandia’s SAR business area looked into ways to increase performance from processors that do not rely on Moore's Law. One of the strategies considered was reduced precision computing. The concept is that if less precision can be tolerated, then more parallelism can be obtained from the same number of transistors. This means more performance without increasing size, weight and power (SWaP).

In 2013 Sandia became aware of a start-up called Singular Computing and their DARPA funded S1 processor. The S1 is a prototype chip that implements reduced precision computing. It has a large number of simple computing elements. The S1 provided Sandia the opportunity to explore reduced precision SAR image formation in a concrete manner. The exploration centered around implementing and evaluating the convolutional backprojection algorithm (CBP) [1], [2] on the S1. This presentation summarizes Sandia’s initial experience with using the S1 to form SAR images.

S1 Concept and Architecture

The S1 chip is a chip from a start-up called Singular Computing. The S1's primary target is the neural-net application space. Its design is inspired by the brain in that it strives to maximize parallelism and communication and minimize power consumption [3]. To achieve these goals, the S1 sacrifices precision in computations. In practice, the S1 increases the amount parallelism obtained from a given number of transistors by limiting the capabilities of the individual processing elements (cores).

At the heart of the S1 architecture are the Approximate Processing Elements (APEs) [4],[5]. Each S1 chip has 2112 APEs. An APE roughly corresponds to a CPU "core" but with restricted capabilities. Each APE has 512 2-byte words of memory for a total of approximately 2 megabytes of memory per chip. Each APE has 8 2-byte general-purpose registers. The S1 has three data types: Boolean, 16-bit integers and a 14-bit "approximate float" (occupying a 2-byte word) which is similar to an IEEE half-precision floating point. (For brevity, the term "approx" will be used for "approximate float" in the remainder of this document.)

The operations available for the approx datatype are Add, Sub, Mult, Div and Sqrt. The operations available for integers are Add, Sub, And, Or, ShiftLeft, and ShiftRight. The S1 does not provide machine instructions to multiply or divide integers. General-purpose integer multiply or divide operations must be synthesized from long sequences of existing operations and hence are much slower. All arithmetic machine instructions complete in a single clock cycle.

Within the S1 chip the APEs are arranged in a grid with 48 rows and 44 columns. The APEs operate in a SIMD (Single Instruction, Multiple Data) fashion with all APEs executing in lock-step the same instruction at the same time. The APEs communicate with each other over a NEWS (north, east, south, west) interconnection network.

One of the central features of the S1 architecture is the approx data type. It is similar to, but not the same as, 14-bit floating point. One of the 14 bits is a sign bit, seven bits are an exponent and the remaining 6 are a "logtissa" i.e. the log of the mantissa. The inherent granularity of the approx data type is 1 part in 64 (about 1%). The approx data type has a wide dynamic range (on the order of $10^{+/-20}$). One of the peculiarities of the approx data type is that it cannot exactly represent zero.

Evaluation of Results

Our S1 implementation of the CBP algorithm produced the SAR image in Figure 1. This is an image of Sandia's antenna test range on the south side of the Kirtland Air Force Base. The vertical dark stripe in the center is a slab of smooth concrete with very low backscatter. The grid of bright dots is an array of corner reflectors used for calibrating SAR images. The image in Figure 2 is a reference image created from the same data by a CPU version of CBP using an Intel Pentium processor and 64-bit IEEE floating point.

Qualitatively the image produced by the S1 (left image) has reasonable appearance. To the human eye it looks identical to the reference image (right) produced using double precision floating point.

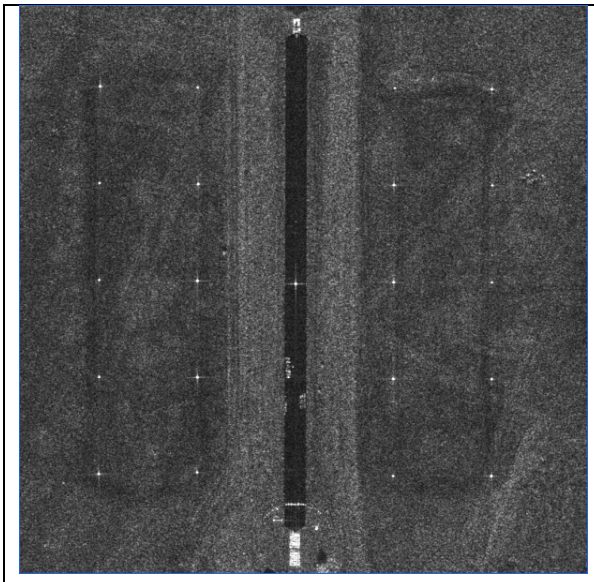

Figure 1 SAR Image Produced by S1

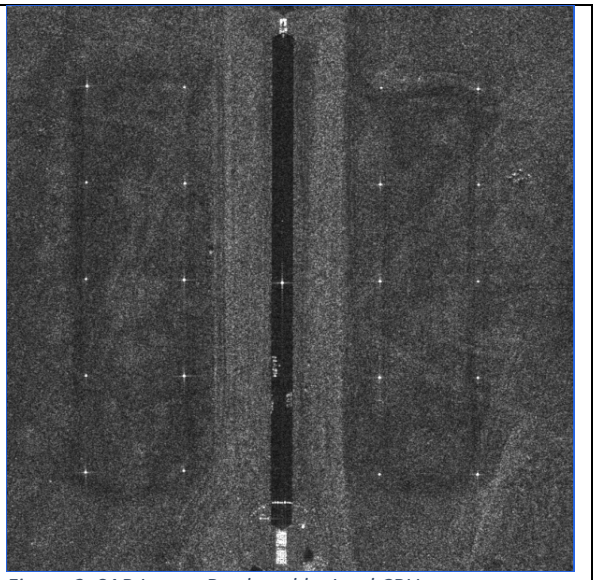

Figure 2 SAR Image Produced by Intel CPU

Impulse Response

To better quantify the quality of the images, Figures 3 and 4 show the analysis of the IPR (impulse response) of the center reflectors in the S1 and CPU images respectively. Overall the IPRs of the two images are similar. The peak in the S1 image is a fraction of a dB weaker than in the CPU image. The differences in the range IPRs are very minor. The differences in the azimuth IPRs are more pronounced.

The most salient difference in the azimuth IPRs is the distinct sidelobe on the right side of the azimuth IPR for the S1 image that is not present in the CPU azimuth IPR. Beyond this sidelobe, the S1 image's azimuth IPR does not have as deep nulls as the CPU image and the main lobe is very slightly broader.

These effects are not enough to substantially change the quantitative metrics appearing at the bottom of the figures. Other than the sidelobe at +20 pixels in the S1's azimuth IPR, the difference between the S1 and CPU IPRs is comparable to the variability between IPRs from repeated collects of the same scene using the same radar and acquisition parameters.

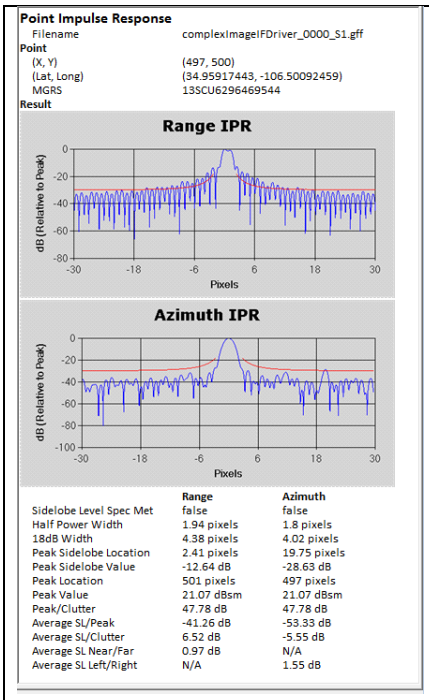

Figure 3 IPR of Center Reflector of S1 Image

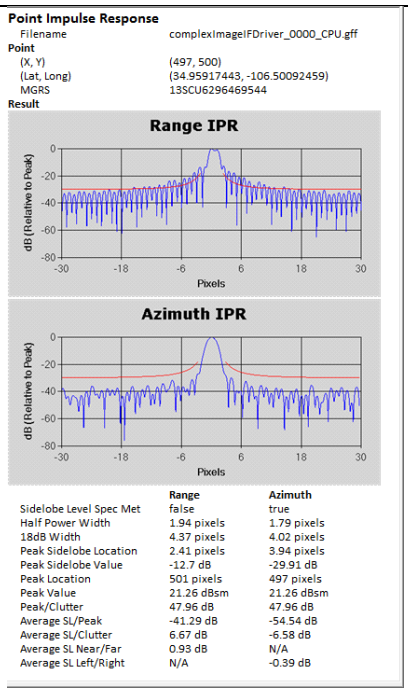

Figure 4 IPR of Center Reflector of CPU Image

Conclusions

While this work demonstrated the ability to form a SAR images on the S1 within the constraints associated with the S1 architecture (approx data type, absence of integer multiply and divide, etc.), the S1 has resource limitations that make it challenging in its current instantiation for forming SAR images.

(1) The amount of memory available is a primary limitation. The largest image we can form on a system with 16 S1 chips is approximately 1024 x 1024 pixels.

(2) I/O bandwidth between the Control Unit and the APEs is another primary limitation. For a 1024 x 1024 image size, data transfers to/from the APEs took 25 times as long as the computation.

The S1 system used in this work is a prototype. Singular Computing expects to be able to scale memory and I/O bandwidth by a factor of hundreds in subsequent generations.

Writing software for the S1 requires managing low-level details: (1) data transfers, (2) memory layout and management, (3) expressing computation at the level of individual operations. Explicitly specifying all the low-level details bloats the software. The CPU version of the CBP is less than 200 lines long. Largely due to having to code at such a low-level, the S1 implementation uses over 2000 lines of code to do the same thing as the 200 lines of CPU code.

Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

References

[1] M. D. Desai, W. K. Jenkins, “Convolution Backprojection Image Reconstruction for Spotlight Mode Synthetic Aperture Radar,” IEEE Transactions on Image Processing, vol 1, no.4, pp. 505-517, October 1992.

[2] C. V. Jakowatz, Jr., D. E. Wahl, D. A. Yocky, “Beamforming as a Foundation for Spotlight-Mode SAR Image Formation by Backprojection,” SAND2008-1033C, Sandia National Laboratories, 2008.

[3] J. Bates, “Practical Approximate Computing,” presented at Neuro-Inspired Computational Elements Workshop, Berkley, California, March 7-9, 2016.

[4] Singular Computing LLC, “Singular S1 Overview,” Revision 2, August 2015.

[5] Singular Computing LLC, “Software Examples for Singular S1,” Revision 1, August 2015.

Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy’s National Nuclear Security Administration under contract DE-NA0003525.