Minimizing Power Consumption using Signal Activity Transformations for Very Deep FPGA Pipelines

Jeffrey T. Muehring and John K. Antonio

School of Computer Science
University of Oklahoma
200 Felgar Street
Norman, OK 73019-6151
Phone: 405-325-7859
Fax: 405-325-4044
antonio@ou.edu

Abstract

The density of commercially available FPGAs (field programmable gate arrays) has increased dramatically in the past several years. Because of this trend, it has become possible to move more and more computations associated with high-performance embedded applications from DSPs (digital signal processors) onto FPGA devices. The potential advantages of utilizing FPGAs instead of DSPs, in this context, include reductions in the utilization of size, weight, and power (SWAP) required to implement the computational platform. These types of reductions are of particular interest for military applications such as synthetic aperture radar (SAR) processing, which is often performed on a small unmanned aerial vehicle (UAV) having limited available power for the on-board computational platform.

Two major contributions are presented in this paper. First, it is shown that the core computations from the SAR application, including both the range compression and azimuth processing phases, can be structured as a single deep computational pipeline that can be implemented directly onto an array of FPGAs. Past results for high-throughput SAR processing typically assume the computations are to be mapped onto a distributed memory multiprocessor system in which a subset of the available processing elements (PEs) are assigned to perform range compression and the remaining processors perform azimuth processing. In this type of traditional approach, a number of compressed (processed) range vectors are sent from the range PEs to the azimuth PEs where they are buffered in memory. After a prescribed number of compressed range vectors are present in the memory space of the azimuth processors, azimuth processing commences. Because of the significant intermediate buffer storage required by this approach, and the

associated placing and fetching of data in this memory space by the range and azimuth PEs, respectively, this type of SAR implementation is generally not thought to be a "purely streaming" application. However, as is presented in this paper, these computations (both phases) can in fact be structured as a single computational pipeline, which can be directly mapped onto an array of FPGAs. In the proposed approach, no intermediate memory buffer is required between the two phases of computation. Instead, within the structure of the computational pipeline are long segments of delay elements that effectively provide the intermediate storage associated with the more traditional approach. One potential advantage of the proposed approach is that data need not be continuously stored and then fetched from a separate memory module by PE (which, incidentally, can require significant power consumption). Instead, the data streams continuously through a long computational pipeline. Within this pipeline are the taps of the FIR (finite impulse response) implementations of both the range and azimuth processing, interspersed with long segments of delay elements. Although the resulting pipeline may be thousands of stages long for practical values of SAR parameters, it is a viable approach because end-to-end latencies on the order of 1 millisecond are typically acceptable, provided that the required throughput is achieved.

The second contribution presented in this paper demonstrates how signal activity parameters of incoming data can be transformed, before the data are processed by a computational pipeline, as a means of reducing overall power consumption. The key to understanding this approach is the realization that the activity levels of the input signals to the computational pipeline dictate its level of power consumption. The activity of a given input signal (i.e., bit) is defined as the fraction of times that the signal transitions relative to the system clock. It has been demonstrated that increasing the signal activities of input data to a pipelined circuit implemented on an FPGA also increases the power consumption of the circuit [1]. In the present paper we illustrate how the activities of the input data can be transformed (pre-processed) so that the resulting (transformed) signals that are input into the computational pipeline have activity values that are well-matched with the pipelined circuit in terms of minimizing consumed power. At the end of the computational pipeline, an inverse transformation is applied to the output values to convert them back to their proper (and meaningful) representation. This approach is based on two fundamental assumptions: (1) that the power consumption of the computational pipeline is significantly higher than that of the computational structures implemented to perform the transform and inverse transformation of the data and (2) that the computations performed within the computational pipeline are linear and time invariant.

The final version of this paper will contain further details related to the two contributions outlined here. Details on the structure and depth of the computational pipeline associated with the proposed SAR processing approach will be provided. This approach, in terms of estimated power consumption, will be compared with more traditional approaches that make use of a multicomputer architecture. Also presented will be measurements and estimates of overall power savings possible by using the proposed signal activity transformation approach.

Reference

[1] Timothy Osmulski, Jeffrey T. Muehring, Brian Veale, Jack M. West, Hongping Li, Sirirut Vanichayobon, Seok-Hyun Ko, John K. Antonio, and Sudarshan K. Dhall, "A Probabilistic Power Prediction Tool for the Xilinx 4000-Series FPGA," *Proceedings of the 5th International Workshop on Embedded/Distributed HPC Systems and Applications (EHPC 2000)*, in *Lecture Notes in Computer Science*, sponsor: IEEE Computer Society, Cancun, Mexico, May 2000, pp. 776-783.