Song Sun
Ph.D.
Computer Engineering

Abstract:
With the continued development of computation and communication technologies, we are overwhelmed with electronic data. Ubiquitous data in governments, commercial enterprises, universities and various organizations records our decisions, transactions and thoughts. The data collection rate is undergoing tremendous increase. And there is no end in sight. On one hand, as the volume of data explodes, the gap between the human being's understanding of the data and the knowledge hidden in the data will be enlarged. The algorithms and techniques, collectively known as data mining, are emerged to bridge the gap. The data mining algorithms are usually data-compute intensive. On the other hand, the overall computing system performance is not increasing at an equal rate. Consequently, there is strong requirement to design special computing systems to accelerate data mining applications. FPGAs based High Performance Reconfigurable Computing (HPRC) system is to design optimized hardware architecture for a given problem. The increased gate count, arithmetic capability, and other features of modern FPGAs now allow researcher to implement highly complicated reconfigurable computational architecture. In contrast with ASICs, FPGAs have the advantages of low power, low nonrecurring engineering costs, high design flexibility and the ability to update functionality after shipping. In this thesis, we first design the architectures for data intensive and data-compute intensive applications respectively. Then we present a general HPRC framework for data mining applications.

Frequent Pattern Mining (FPM) is a data-compute intensive application which is to find commonly occurring item sets in databases. We use systolic tree architecture in FPGA hardware to mimic the internal memory layout of FP-growth algorithm while achieving higher throughput. The experimental results demonstrate that the proposed hardware architecture is faster than the software approach.

Sparse Matrix-Vector Multiplication (SMVM) is a data-intensive application which is an important computing core in many applications. We present a scalable and efficient FPGA-based SMVM architecture which can handle arbitrary matrix sizes without preprocessing or zero padding and can be dynamically expanded based on the available I/O bandwidth. The experimental results using a commercial FPGA-based acceleration system demonstrate that our reconfigurable SMVM engine is more efficient than existing state-of-the-art, with speedups over a highly optimized software implementation of 2.5X to 6.5X, depending on the sparsity of the input benchmark.

Accelerating Text Classification Using SMVM is performed in Convey HC-1 HPRC platform. The SMVM engines are deployed into multiple FPGA chips. Text documents are represented as large sparse matrices using Vector Space Model (VSM). The k-nearest neighbor algorithm uses SMVM to perform classification simultaneously on multiple FPGAs. Our experiment shows that the classification in Convey HC-1 is several times faster compared with the traditional computing architecture.

Map Reduce Reconfigurable Framework for Data Mining Applications is a pipelined and high performance framework for FPGA design based on the Map Reduce model. Our goal is to lessen the FPGA programmer burden while minimizing performance degradation. The designer only need focus on the mapper and reducer modules design. We redesigned the SMVM architecture using the Map Reduce Framework. The manual VHDL code is only 15 percent of that used in the customized architecture.