Maximum Performance Computing with Dataflow Engines

MAXELER Technologies

This is a summary of a paper copublished by the IEEE CS and AIS July/August 2012.

About the Authors

Oliver Pell is VP of HPC Solutions at Maxeler Technologies and Vitali Averbukh is a Professor at Imperial College of London.

Summary

The computational power and memory resources of a computer directly determine the complexity of the scientific problems that can be attempted, in addition to the achievable accuracy and reliability of the particular program. Despite great advances in the past years, performance gains from processor frequency scaling have nearly ended, and is now one of the main limiting factors in high performance computing (HPC). Computer architectures in modern computing can be divided up into general purpose and program-specific. Maxeler Technologies is leading the development of Dataflow Computing, a class of special purpose computers that are programmable and can be specialized to run different applications orders of magnitude faster than conventional CPUs.

On a dataflow engine (DFE), data is streamed from memory onto a chip where the data moves directly from one functional unit to another, without being written to off chip memory until the entire process is complete, as shown in Figure 1. Control flow core performs operations at differents points in time on the same functional unit ("computing in time"), whereas on a dataflow core, computation is laid out spatially on a chip ("computing in space"). In a DFE there is no need for instructions because the DFE itself represents the computation. As a result, there is also no need for memory decode logic, branch prediction or out of order scheduling, allowing the chip to dedicate all its resources to performing computation.

DFEs must be integrated with conventional CPUs, networking, and storage to build a balanced compute system at cluster level. Two types of Maxeler Maximum Performance Computing (MPC) dataflow compute nodes used are the MPC-C and MPC-X. The MPC-C permits standalone deployment of dataflow technology with a fixed combination of coupled CPUs and DFEs. The MPC-X allows a heterogeneous system designed to balance

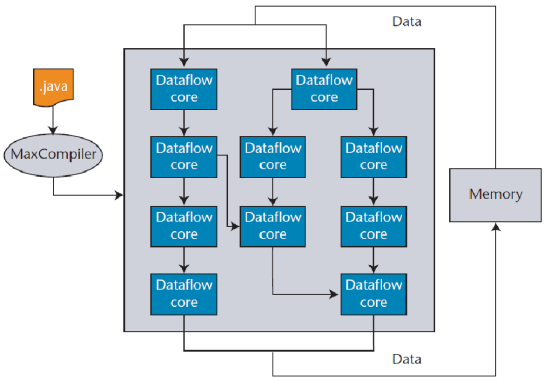

Figure 1: Model of a dataflow compute engine.

different compute technologies, since the balance of CPUs and DFEs is flexible at runtime. The system software manages the DFEs, allocating them to different applications as they are released by previous applications.

Dataflow computing accelerates a wide variety of applications such as Monte Carlo simulations, financial tree-based PDE solvers and finite difference solvers. Finite difference is a numerical method often used to model waves. One of the primary challenges in wave modeling is that the amount of computation required increases with the fourth power of the wave frequency. Modeling high frequencies like 70 Hz can easily require hundreds of gigabytes of memory. In the accelerated wave model, CPUs control the application and instruct the DFEs to compute each step, as shown in Figure 2. Multiple DFEs can work together on a single problem dividing the domain into subdomains to be handled by each DFE.

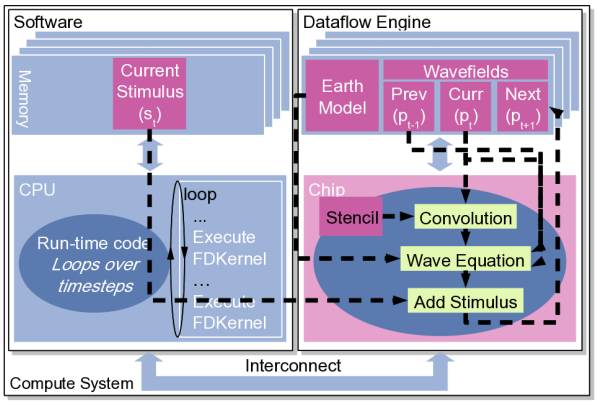

Figure 2: Structure of an accelerated wave-modeling application.