Performance Modeling and Simulation of Communications-Intensive ProcessingHPC Systems

Andrew Huang
(bunnie@ai.mit.edu)

The Problem:

Project Aries is architecting a communications-intensive processing system with a focus on addressing the issues of reliability, efficiency, and accessibility. Reliability includes both fault tolerant hardware and hardware-supported software constructs that enable the execution of code in multiple domains. Efficiency is measured in terms of computation per silicon area, power dissipation, and cost per result to the end user. Programmability–referred to as accessibility–implies the existence of easy to use tools which help a third party programmer utilize the specialized hardware features and quickly visualize where their applications are slowing down. A high performance computer platform designed without regard to accessibility issues is about as useful as a short from power to ground. Trying to design a system that addresses all of these issues requires a simulation tool that is up to the task of emulating multiple processors and their network with reasonable performance.

Motivation:

This diverse set of foci reflects Aries' system-oriented approach. The difficulty with trying to design such a system lies in creating an adequate performance model that allows one to optimize all of the complex interactions of the individual elements of a high performance computer (HPC) system. Factors as diverse as special hardware features, the accessibility of these features, compiler power, the complexity of interesting problems, network topology and fault-tolerance all play an important role in the final performance equation. Creating a accurate simulation model that accounts for all these factors is difficult, but crucial for making informed decisions about the architecture. A simulation is accurate if it can be used to correctly predict the performance bottlenecks of existing HPC systems. In addition to being accurate, the simulation must also be fast and flexible, so as to allow the user to quickly explore the design space with adequate feedback from well-chosen performance benchmarks.

Previous Work:

Previous efforts at creating performance models of computers often use a software simulation approach. Stanford's SimOS [1] is a good example of an all-software solution which provides various performance models and profiling features. The major problem with using a system like SimOS to model a large, multiprocessor system is the lack of performance. Stanford's DASH project found that software simulations of their architecture would incur a real-time slow down in the range of 1000x-10000x. [2] Their solution was to build a hardware prototype with a plethora of performance monitoring features, so they could profile very large datasets within a reasonable amount of time. However, the DASH prototype is specific to the DASH architecture. The prototype being developed by project Aries has a similar performance motivation behind its design, but with the added capability of being able to target a broader range of architectures.

Approach:

The project Aries prototyping platform will attempt to provide an accurate simulation of the processor-in-memory architecture that is proposed for this machine, as well as its high performance interconnect network based on the MIT Transit Project's METRO Network [3]. The initial prototype will include numerous hooks for monitoring network congestion and throughput; power consumption; and specialized innovations that include hardware support for multiple domains of execution, tags, network coprocessors, tag manipulation in DRAM, and memory hierarchy slicing. These specialized computation structures will be implemented using reconfigurable hardware for good performance and short design cycles. The prototype will also feature a physical design targeted at enabling experimentation with several network topologies, using commercially available high speed point-to-point cabled data link chipsets. Development efforts for the initial prototype will have a strong emphasis on establishing a well documented, reusable design database, enabling rapid design iterations. Parallel efforts in developing software toolchains will thus be required to have enough generality to target the rapidly evolving hardware platform, which will hopefully yield a robust, platform-independent toolchain.

Once the experimental hardware platform has been built, existing high performance system architectures, such as the SGI Origin 2000's ccNUMA hypercube architecture, will be implemented with simplifications, so the performance predictions of the simulation can checked against reality. This reality check is required to lend credibility to the performance predictions made by the simulation on hypothetical architectures.

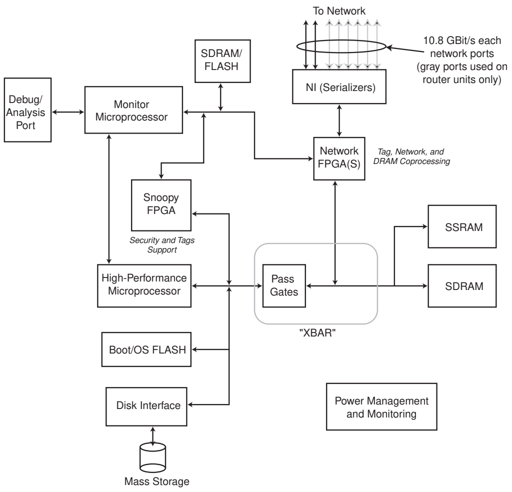

Figure: Detailed architecture of a processor node

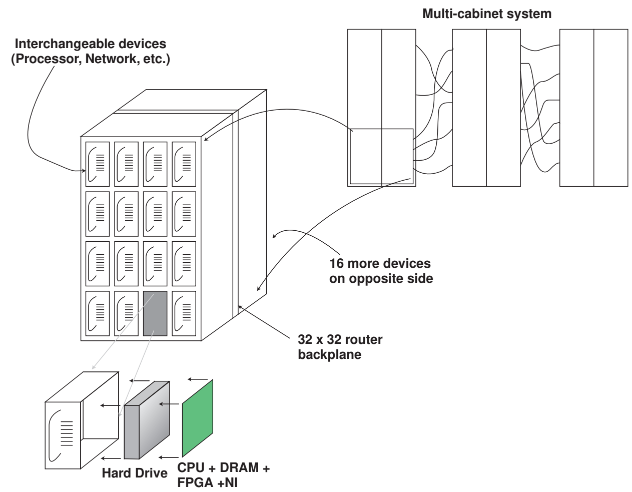

Figure: System prototype physical design overview

Impact:

The primary result of this prototype will be a performance model, based on empirical data garnered through simulations, of the issues involved in creating high performance systems capable of running real applications. The analysis of the data yielded by this performance model will lead to a finalized implementation of the architecture. Also, the validation of this performance model will lead to the development of reference designs based on existing architectures. The availability of reference designs of multiple architectures on a common platform will hopefully find use in other research projects. Finally, a reusable, well-documented design database will be created which will enable future progress in this project, as well as assist other research projects within the lab.

Research Support: Support for this research was provided by the Air Force Research Laboratory, agreement number F30602-98-1-0172, "Active Database Technology".

References:

[1] M. Rosenblum, S. Herrod, E. Witchel, A. Gupta "Complete Computer System Simulation: The SimOS Approach" Available at http://simos.stanford.edu/papers.html

[2] D. Lenoski, J. Laudon, T. Joe, D. Nakahira, L. Stevens, A. Gupta, J. Hennessy "The DASH Prototype: Logic Overhead and Performance". Available at http://www-flash.stanford.edu/architecture/papers/paperlinks.html

[3] A. Dehon "Robust, High-Speed Network Design for Large-Scale Multiprocessing". MIT AI Lab Tech Report 1445, AI-TR 1445, Massachusetts Institute of Technology, 1993.