**Topic Area:** Heterogeneous Systems, Multi core Systems, Large Scale Simulations

**A Novel Many Core Simulator for very Large Clusters running Multiple Applications**

Venkateswaran Nagarajan¹, Ram Srivatsa Kannan², Prashanth Thinakaran², Rajagopal Hariharan², Vinesh Srinivasan², Bharanidharan Vasudevan⁷, Prashanth Manohar², AswinKumar Sridharan³, Karthikeyan Palavedu Saravanan³, Vignesh Adhinarayanan³, Vignesh Veppur Sankaranarayanan³.

¹Director Waran Research Foundation, ²Research Trainee Waran Research Foundation, ³Former Research Trainee Waran Research Foundation.

Advent of multi-core designs has paved way towards breaking the Exa-scale performance barrier and hence it is been a major focus for modern computer architecture research. It is evident that future generation super computers would have hundreds of multi-core processors at the node level and proportionately a higher number at the cluster level too. Simulation has become a primary technique for evaluating the performance of new design proposals in computer architecture. These node architectures which are generally simulated for its correctness are scaled up to the cluster level. So, there arises a necessity to scale node level simulators to evaluate the metrics of cluster with the help of large scale simulation. Hence, a proactive paradigm shift from the conventional simulation techniques towards effective simulation of cluster level design and addressing its needs becomes inevitable. This research trend accentuates the need for developing a cycle accurate multi-core processor simulator that is scalable to cluster level and therefore we present a scalable simulator that couples a cycle-accurate node simulator with a generic supercomputer network model. At this context, there occurs a necessity towards simulating the node architecture consisting of functional units, network topology, memory hierarchy etc. with cycle accuracy since the Architectural simulators without cycle accuracy which examines the impact of low level system changes on application performance have not historically scaled well [1] [2]. For example these coarse-grained simulators skew dramatically of the order 100x when these changes are scaled up over hundreds of systems at the cluster level.

Recent trends in supercomputing for instance, deployment of Kei computer by Japan and development of blue waters by IBM shows that Massive efforts and huge monetary investments are made in development of such supercomputing clusters in order to achieve Exa flop performance. This articulates the need for developing an effective cluster level simulator that quantifies all the parameters that influences the cluster design. Proposed changes in [3] include techniques for providing accurate estimates of performance and various other machine statistics. Unfortunately, the impact of these models on application performance becomes very difficult to predict analytically when it comes to cluster level owing to complex interaction with in and across nodes, operating system and the applications. These simulators become uncertain when scaled up to cluster level simulations owing to the fact that not being accurate at the node level.
Moreover the core perceived as a tile as proposed by the simulator is claimed to model the computational pipeline which again can be proved to be ambiguous since one cannot treat an entire core as a black box for simulation purposes. On the other hand, in our proposed model for simulation, we try to bring out the interaction between each and every functional units contained in the underlying architecture with its associated network topology in a cycle accurate manner.

The DARSIM [4] is a cycle accurate network simulator that proposes techniques for simulating network traffic using packet traffic generator as trace driven injector. This again does not depict the actual data traffic across nodes in a cluster environment and therefore, these methods are though closely accurate enough to capture the essential details of node level simulations but when scaled up fails to provide a model which has wider design space that encompasses a large set of parameters that are required to design various levels of cluster.

As a milestone step towards the design and automation of supercomputing clusters we propose a processor simulator for a very large cluster environment which is a novel framework that encompasses wide range of parameters pertaining to design of the different levels of cluster with cycle level accuracy at node level. The simulator presented here is to the best of our knowledge the first ever cycle accurate processor simulator integrated with an optimizer that prunes the design space exploration of the various architectural components which are heterogeneous in nature. The heterogeneous multi-core processor simulator developed is also shown to be scalable up to the Cluster Level. Such an initiative for cluster design is a first-of-its-kind that was proposed here, consolidates a large set of parameters, both stochastic and deterministic which is accurate enough to capture the dynamics of the entire cluster tending towards cent percentage accuracy.

CUBEmach SIMulator (CUBESIM) is a cycle-accurate, optimizer integrated framework used to enable detailed architectural evaluation of Custom built heterogeneous multicore architectures [9]. Despite of the fact that cycle accurate simulators are slow, they provide extremely accurate results in order achieve detailed modeling and realistic evaluations of node level architecture. Thus providing detailed and accurate statistics on switching events within and across the cores by simulating the integrated working of node architecture that includes on-core interconnect networks, dynamic hierarchical hardware based compiler, heuristic driven memory subsystems including cache hierarchies with full cache coherence and integrated optimization engine based on the combination of simulated annealing and game theory to prune the design space exploration with regard to performance and power. The design of CUBESIM is modular and object oriented thus adaptable to different models in order to simulate different architectures which can trade off performance for power thus optimizing wide range of design parameters to suit user needs. CUBESIM runs on commodity based machines thus extending its compatibility to support the execution of boost threads. CUBESIM will be made open-source to foster research and development for future multi-core processors.
SYSTEM ARCHITECTURE:

As a case the CUBESIM processor level simulator [9] is designed for more generic architecture CUBEMACH (CUstom Built hEterogeneous Multi-core ArCHitecture) [Fig.1][5], which support the simultaneous execution of multiple applications without space time sharing addressing the needs of future generation supercomputers.

![CUBEMACH System Architecture](image)

**Figure 1. CUBEMACH System Architecture**

Functional Unit of core:

The basic unit of the core is composed of higher level functional units as well as conventional ALUs. It is evident that the future generation computers would be composed of higher level functional units [5] [6] to tackle the application complexity. Hence the Node level simulator is adaptable to higher level units called ALFU which is capable of executing an algorithm of reasonable size.
Compiler cum scheduler:

To overcome the system software overhead posed by the conventional compiler to generate the required instructions which drives the hundreds of underlying functional units and its corresponding mapping of applications across the cores a Hardware based compiler cum scheduler owing to its architectural significance referred as Compiler-On-Silicon is simulated.

Network architecture:

Cost effective high performance computing being the order of the day, conventional grid based Network-on-Chip (NoC) cannot suffice to the highly critical bandwidth requirements for future generation super computers, thereby necessitating the need for a high bandwidth cost effective NoC for large scale heterogeneous multi-core architectures like the CUBEMACH. Hence we use a novel, cost effective On Chip Network called the On Node Network Architecture (ONNET) which is more generic and in turn supports several topologies taking into account the application characteristics of the inputs given to super computers and the communication complexity of very large scale heterogeneous multi-cores.

Memory system:

The memory system of the CUBESIM composed of a three tier hierarchical cache model. The memory system of the simulator is built using generic modules such as cache, heuristics based mapping and replacement, memory controllers, coherence protocols. Currently the architecture has provision to accommodate data packets belonging to multiple applications that are run by different users in the same level of cache with customizable associativity. The heuristics driven cache mapping strategy maps the data of multiple applications to different levels of cache driven by the cache controller. The interaction between the scheduler and the cache controller is a first of its kind, while mapping the data of multiple applications. Each event related to hit, fetch or miss is associated with a clock cycle timestamp. Thus the Cache performance is modeled by tracking the average latency in memory access time using these timestamps.

Optimizer:

Heterogeneous multi-core architectures used in super computers require a powerful optimization methodology for their design due to the number of varying parameters involved. The Optimizer Engine works by the combination of game theory and simulated annealing for the processor level simulator which prunes the search of the design space in order to find architectures that satisfy power, performance or power to performance criteria of the multiple applications being executed. A subset of the parameters from the architectural space is selected using Game Theory based on certain heuristics as shown in figure which helps in selecting the parameters which affects the solution and how close the current architecture state about to satisfying the requirements
of the multiple applications. The values of the parameters governing the system temperatures are varied suitably using simulated annealing technique [7]. Thus the architecture is simulated to get the communication pattern, on the basis of which core formation [Fig 1] takes place, to group highly communicating Functional Units [8] being grouped together to minimize communication latency. The application of optimization theory for the extensive design space exploration of architectural components in heterogeneous multi cores is a first of its kind initiative.

Conclusion:

we have described the need for cycle accurate simulation of supercomputing clusters and how a processor level simulator is critical in this regard. We have elaborated the CUBEMACH design paradigm and also proposed a clock driven, cycle accurate simulator (CUBESIM) that encompasses all node level parameters. The working of an integrated Optimizer Engine that finds the most suitable architecture with respect to power and performance. To substantiate the efficiency of CUBESIM processor level simulator, thus can be easily scaled up for cluster level simulation. A framework as strong as the CUBESIM cluster level simulator will serve as a platform for in the quest towards design automation of high productive supercomputing clusters.

References:


9. This is the link to the CUBEMACH design paradigm simulator: http://www.warftindia.org/joomla/index.php?option=com_content&view=article&id=122&Itemid=102

Corresponding Author 1’s information:

Name Prashanth Thinakaran
Affiliation Research Trainee, WAran Research FoundaTion (WARFT)
Email prashanth5192@yahoo.in
Address Fno:65, Majestic Park, Off 48 Arcot Rd, Saligramam, Chennai-93, India.

Other contact:

Name Venkateswaran Nagarajan
Affiliation Director, WAran Research FoundaTion (WARFT)
Email waran@warftindia.org
Address 46B, Mahadevan street, West mamabalam, Chennai, India.