Towards Efficient Many-Task Computing on Accelerators in High-End Computing Systems

Scott J. Krieder, Benjamin Grimmer, Dustin Shahidehpour, Jeffrey Johnson
Justin M. Wozniak, Michael Wilde**, Ioan Raicu***
Department of Computer Science - Illinois Institute of Technology, Chicago, IL, USA
*Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL, USA
**Computation Institute – University of Chicago, Chicago, IL, USA

Abstract

Current software and hardware limitations prevent Many-Task Computing (MTC) workloads from leveraging hardware accelerators boasting Many-Core Computing architectures. This work aims to address the programmability gap between MTC and accelerators, through the innovative CUDA middleware GeMTC. By working at the warp level, GeMTC enables heterogeneous task scheduling and 10x number of workers compared to CUDA. In order to span multiple accelerators across nodes, we have adopted the Swift parallel programming system, which can both support fine grained millisecond tasks and extreme scale supercomputers at 100K-cores.

GeMTC

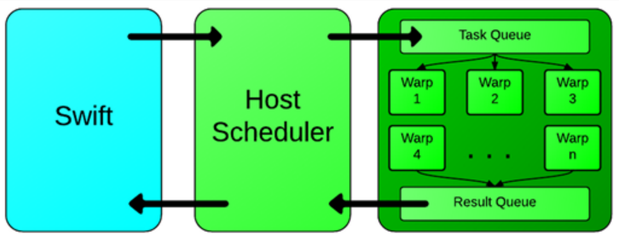

Fig 1: Flow of a task in GeMTC

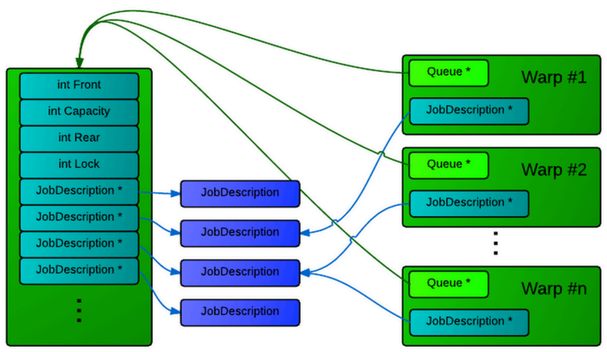

Fig 2: Worker interaction with Queues

Preliminary Results

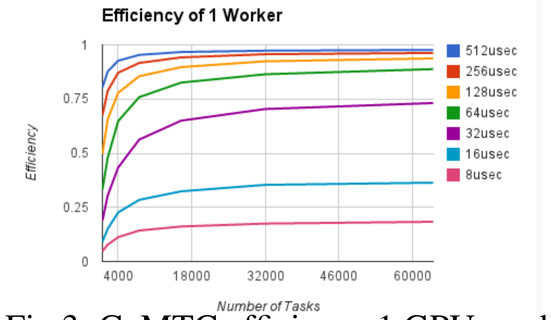

Fig 3: GeMTC efficiency 1 GPU worker

Fig 4: Efficiency 84 Workers

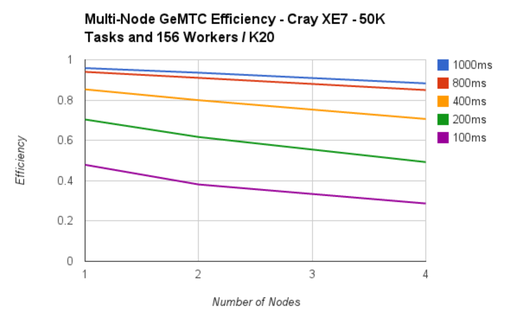

Fig 5: GeMTC + Swift/T efficiency

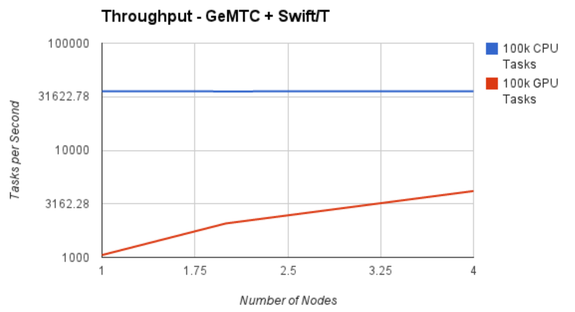

Fig 6: GeMTC + Swift/T 1 to 4 Cray XK7 K20 Nodes

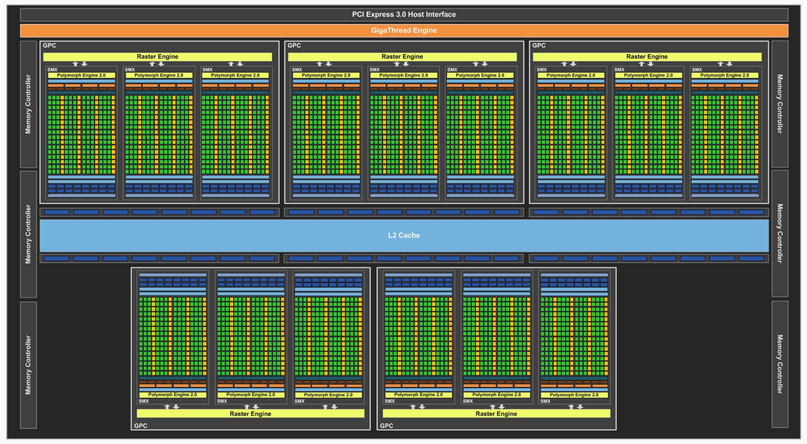

Fig 7: GK110 Block Diagram

- 15 SMX
- O(100) Warps
- O(1000)Cores

Swift/T

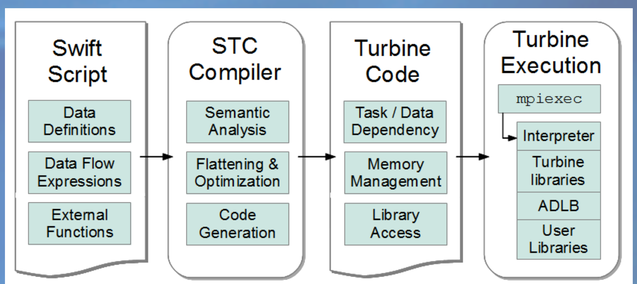

Fig 8: Overview of Swift/T

- Implicitly Parallel Scripting Language
- Data-flow driven scheduling of parallel tasks
- Distributed executor eliminates centralized bottlenecks
- Optimizing compiler detects errors, improves efficiency
- Scales to 100k cores
- Portable to most MPI-based clusters
- Syntax similar to C, Java

Conclusions

- *Designed* GeMTC Framework
- *Improved* Memory Management
- *Integrated* GeMTC + Swift/T
- *Evaluated* Synthetic Benchmarks

Future Work

- *Abstract* for other Accelerators
- *Evaluate* Real Applications
- *Improve* Current Performance

References

GeMTC - http://datasys.cs.iit.edu/projects/GeMTC/
Swift – http://www.mcs.anl.gov/exm/local/guides/swift.html
CUDA - http://www.nvidia.com/object/cuda_home_new.html