Cutting-edge Test Bed Cluster Architecture Based on Intel® Xeon Phi™ Coprocessor

James A. Ang, Manager, Scalable Computer Architectures, Sandia National Laboratories
Simon D. Hammond, Senior Member of Technical Staff, Scalable Computer Architectures, Sandia National Laboratories
Maria McLaughlin, Product Marketing Director, Cray Inc.
Introduction

As part of the National Nuclear Security Administration’s (NNSA) Advanced Simulation and Computing program, Sandia National Laboratories is addressing a critical need for experimental architecture test beds to support path-finding explorations of alternative programming models, architecture-aware algorithms, low-energy runtimes and system software, and advanced memory subsystem development.

The transition from single-core to multicore processor technology and the advent of heterogeneous compute node architectures and accelerators — coupled with the continually increasing demand for more computing cycles — has made Sandia explore cutting-edge technology changes to their high performance computing (HPC) foundation. Sandia is researching how much further an MPI everywhere programming model will be able to scale. Power usage curves of the largest systems are reaching practical limits and are unsustainable when projected to exascale requirements. Current algorithms and solvers must be reconsidered in light of many-core and accelerator technology. Even applications themselves may need to be reworked to take advantage of the anticipated huge node and platform level parallelism. None of these issues will suddenly appear on an exascale-class system. Instead, they will become more apparent as we continue on the path to exascale. Now is the time to explore what and by how much things need to change.

A logical first step is to study these challenges, as well as others, starting from the key building block of an HPC system — the node. Sandia obtained several cutting-edge test beds, ranging in size from one node to a few racks of nodes. These test beds offered various methods designed to increase node level computational ability, all presented as increased node-level parallelism of some type.

The major focus areas were expected to be:

- Alternative programming models
- Architecture-aware algorithms
- Low-energy runtime and system software
- Advanced memory sub-system development

The first-of-its-kind experimental cluster test bed built by Sandia for this project was known as “Arthur” — a cluster supercomputer test bed based on Intel® Xeon® processor E5 family and Intel’s Knights Ferry (KNF) software development platform for the Intel® Many Integrated Core architecture (MIC).¹ The project intended to establish an experimental foundation for Sandia’s exploration of advanced programming models with proxy applications and R&D of advanced system software.

This cluster project was planned in three phases:

1. **Phase one:** Build the Arthur test bed platform based on the Intel Xeon processor 5600 and the KNF coprocessor as a baseline system for the laboratories’ test work.
2. **Phase two:** Develop a new test bed called “Compton,” (launched in May 2012) also based on the Intel Xeon processor E5 family. The plan was to test if the system would improve performance while reducing I/O network latency.
3. **Phase three:** Add pre-production Intel® Xeon Phi™ coprocessors to the Compton system (implemented in September 2012). The plan was to test the system based on the latest Intel Xeon processor combined with the Intel Xeon Phi coprocessor to understand breakthrough performance gains as we move towards the development of the world’s first exascale supercomputer.

¹ Announced in November 2011, the project was a collaboration between Intel and Appro International Inc. The latter was acquired by Cray Inc. in November 2012.
Arthur and Compton – Test Bed System Configurations

Arthur is a first-of-its-kind experimental cluster based on Intel’s Knights Ferry (KNF) software development platform. The system consists of 42 nodes based on a Mellanox Infiniscale IV QDR InfiniBand interconnection network. Each node has two 30-core KNF software development cards at 1.05GHz; each card has 2GB GDDR5 at 1800MHz and one 80GB SSD SATA 3Gb/s MLC NAND Flash drive. Each Arthur node also has two 6-core Intel Xeon processors X5690 running at 3.46GHz.

Compton is a second, first-of-its-kind experimental cluster. Compton also has 42 nodes with the same interconnect and SSD storage infrastructure as Arthur, but it is upgraded with two Intel Xeon E5 processors and two pre-production Intel Xeon Phi coprocessors per node. [1]

The Cray CS300-AC™ cluster supercomputer was used for both test bed configurations. Based on industry-standard, optimized and flexible server platforms, the Cray CS300-AC system offers cutting-edge technologies designed to increase performance while reducing power consumption. The computing nodes and coprocessor used in the system run applications independently to support the demands for more computing cycles, taking advantage of the huge node and platform level parallelism needed today. The system is integrated with Cray’s HPC cluster software stack including Advanced Cluster Engine™ (ACE) management software which offers server, cluster, storage and network management features.

The Intel® Xeon Phi™ Coprocessor

The following features and benefits description comes from Intel’s product brief on the Intel Xeon Phi coprocessor:

The Intel Xeon Phi coprocessor is a complementary solution to the Intel Xeon processor for highly parallel workloads. Based on the Intel® Many Integrated Core (MIC) architecture, the Intel Xeon Phi coprocessor is designed to complement and provide the highest level of flexibility in conjunction with Intel Xeon processor-based systems and clusters.

Now, a wide assortment of programming languages, models and tools support Intel architecture and all of them can be used with both Intel Xeon processors and Intel Xeon Phi coprocessors. Applications that run on one processor family will run on the other. This uniformity can greatly reduce the complexity of developing, optimizing, and maintaining your software code while increasing developer productivity via programming models and tools that are common with the Intel® Xeon® processor.

While the Intel Xeon processor E5 family remains the preferred choice for the majority of applications, Intel Xeon Phi coprocessors provide more efficient performance for highly parallel applications. They include many more and smaller cores, many more threads, and wider vector units. The high degree of parallelism compensates for the lower speed of each individual core to deliver higher aggregate performance for workloads that can be subdivided into a sufficiently large number of simultaneous tasks.

A single Intel Xeon Phi coprocessor provides up to 61 cores and 244 threads and can deliver up to a teraflop of double-precision performance2 for targeted applications. These coprocessors features a standard PCIe x16 form factor that can be added to a supported Intel Xeon processor-based server providing exceptional compute density and energy efficiency.

The Intel Xeon Phi coprocessor can operate under the host server’s operating system (OS), in which case the OS and application run on the Intel Xeon processors and highly-parallel code segments are off-loaded to the Intel Xeon Phi coprocessor to accelerate performance. However, unlike a basic accelerator, the Intel Xeon Phi coprocessor can also function as an independent server node with its

---

2 Claim based on calculated theoretical peak double-precision performance capability for a single coprocessor, 16DP flops/clock/core * 60 cores * 1.053GHz = 1.01088 Tflops
own Linux OS and IP address. In this scenario, it can run applications independently and it can off-load serial code segments (or moderately parallel code segments) to the host system so they perform more quickly and efficiently. [2]

Intel® Many Integrated Core Architecture – Brief Summary

- Wider vector processing units for greater floating point performance/watt
- Highly parallel and highly programmable; standards based with support for data, thread and process parallelism while delivering full support by Intel® Cluster Studio XE
- Higher aggregate performance and higher memory bandwidth compared to the current multicore Intel Xeon processor. [2]

Early Results and Looking Forward

Compton is currently being used to collect application performance data on open-source proxy applications which are intended to show how real applications will use hardware performance characteristics provided by current and new hardware architectures. Researchers at Sandia are also porting several numerical libraries including Trilinos to the Intel MIC architecture enabling larger, complex applications to use performance offered by the Intel Xeon Phi coprocessor.

Alongside much of the porting and optimization efforts, Sandia is also developing expertise around optimizing codes for the Intel MIC architecture for both native and offload execution models and have experimented with scaling to a significant number of threads using several programming models, including OpenMP and Intel Cilk™ Plus and fully hybrid modes of execution combining MPI and threading between many nodes and multiple cards.

Figure 1: MiniFE solver performance on Compton
Note: Lower number is better; Intel Xeon Phi coprocessor data is benchmarked with a pre-production 57-core card
One area of early investigation has focused on finite element assembly and subsequent solution using specialized solvers — a problem which represents complex, sparse, computationally demanding algorithms employed in Sandia’s portfolio of production codes. Using a single code base which has been developed to be portable across Intel’s HPC products, researchers at Sandia have been able to demonstrate not only improved performance on traditional Intel Xeon processors but also a single pre-production Intel Xeon Phi coprocessor out-performing dual-socket Intel Xeon E5 nodes. Figure 1 shows performance results from a single code base of the MiniFE mini-application running on single- and dual-socket Intel Xeon E5 2600 processors and a single pre-production Intel Xeon Phi coprocessor. At small problem sizes, the dual-socket Intel Xeon E5 2600 processor server provides the best performance, but as the problem size increases and more parallelism becomes available, the Intel Xeon Phi coprocessor provides the fastest runtime.

The HPC industry is moving forward rapidly. After reaching petascale computing only a few years ago, industry experts now believe the first exascale supercomputer will be available by 2018. Intel anticipates that its MIC architecture will be an instrumental part of this mission along with the Intel Xeon processor, and systems like Arthur and Compton are helping pave the way.

**References**


**Acknowledgement**

This material is based upon work supported by Sandia National Laboratories for the U.S. Department of Energy’s National Nuclear Security Administration.