LA-UR-17-20129
Approved for public release; distribution is unlimited.

Title: (U) Status of Trinity and Crossroads Systems

Author(s): Archer, Billy Joe
Lujan, James Westley
Hemmert, K.S.
Laros, J.H. III

Intended for: NECDC 2016 meeting Proceedings

Issued: 2017-01-10

(U) Status of Trinity and Crossroads Systems

[1]B.J. Archer, [1]J.W. Lujan, [2]K.S. Hemmert, [2]J.H. Laros III

[1] Los Alamos National Laboratory, Los Alamos, NM

[2]Sandia National Laboratories, Albuquerque, NM

Abstract

(U) This paper provides a general overview of current and future plans for the Advanced Simulation and Computing (ASC) Advanced Technology (AT) systems fielded by the New Mexico Alliance for Computing at Extreme Scale (ACES), a collaboration between Los Alamos Laboratory and Sandia National Laboratories. Additionally, this paper touches on research of technology beyond traditional CMOS. The status of Trinity, ASCs first AT system, and Crossroads, anticipated to succeed Trinity as the third AT system in 2020 will be presented, along with initial performance studies of the Intel Knights Landing Xeon Phi processors, introduced on Trinity. The challenges and opportunities for our production simulation codes on AT systems will also be discussed. Trinity and Crossroads are a joint procurement by ACES and Lawrence Berkeley Laboratory as part of the Alliance for application Performance at EXtreme scale (APEX) http://apex.lanl.gov.

Introduction

The ASC program has set a national platform strategy that consists of two types of platforms. The Commodity Technology Systems (CTS) are fairly large off-the-self systems that provide the computational resources for the majority of the day-to-day Directed Stockpile Work (DSW) workload. The Advanced Technology Systems (ATS) have a dual purpose, provide the computational resources for the largest DSW simulations, and to advance the computing technology provided by the vendors so that the weapon codes can prepare for the next generation of computing. ASC has established a cadence for the AT systems as shown in Figure 1. The AT systems alternate between ACES and LLNL on a 2.5 year cadence, with each site delivering a new system every five years. This paper describes the status of the first two ACES AT systems, Trinity (ATS-1) and Crossroads (ATS-3). The last section briefly describes the status of the D-Wave system installed at LANL.

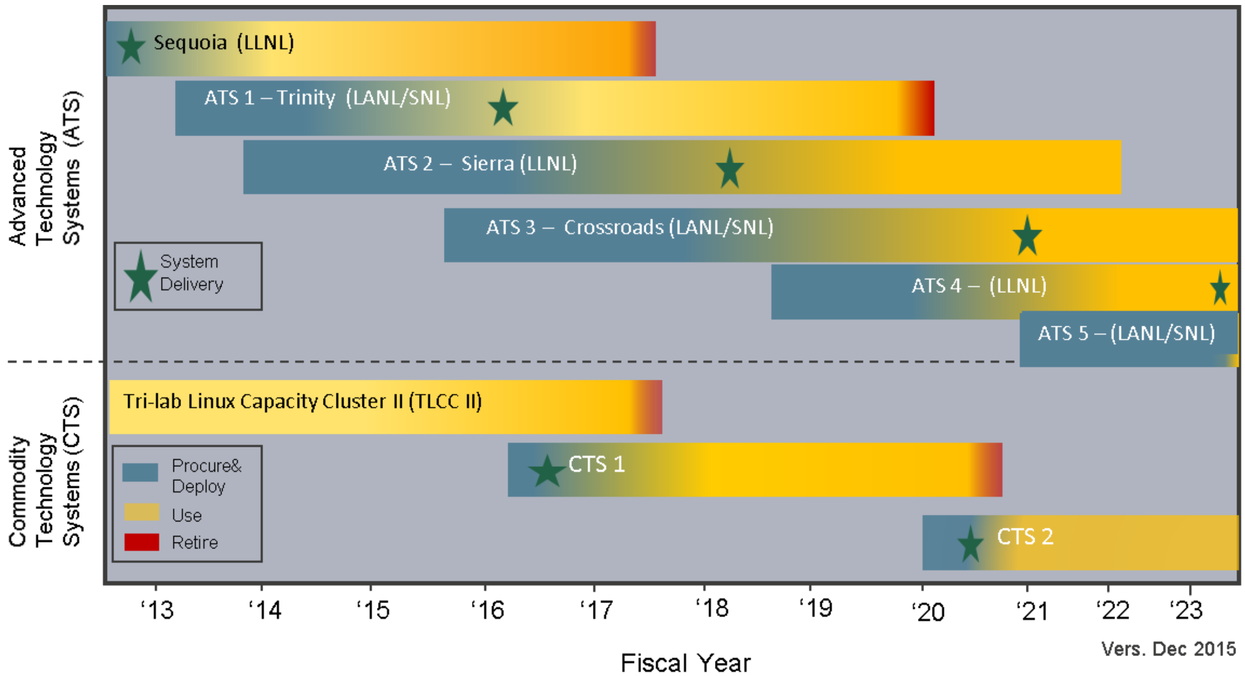

Figure 1. ASC platform timeline showing the five year ATS cadence for each site.

Trinity

Trinity is a Cray XC40™ system that provides over 2 petabytes of memory for DSW simulations. It is a single system that has two partitions with different types of processors. The first partition contains Intel "Haswell" Xeon E5-2698v3 traditional heavyweight processors. The second partition contains Intel Xeon Phi "Knights Landing" multi-core lightweight processors. Details of the system are shown in Table 1.

Trinity has two advanced technology features. The multicore Knights Landing (KNL) processors, with Trinity receiving the first production processors produced by Intel. The KNL processors provide three times the computing compared to the Haswell processors, but at the same electrical power. Trinity is also the first example of Cray's implementation of burst buffers (BB), known as DataWarp™, which uses high speed solid state disks for intermediate storage.

The XC40™ architecture uses a Cray Aries 'Dragonfly' high-speed interconnect that provides advanced adaptive routing. The parallel file system (PFS) is a Cray Sonexion™ storage system that uses Lustre™ to provide 78 petabytes of usable disk space and a sustained bandwidth of about 1.5 terabytes/s. The Cray DataWarp™ software layer supports 576 burst buffer nodes that provide 3.7 petabytes of storage at a sustained bandwidth of about 3.3 terabytes/s.

Table 1. Trinity details.

Trinity			
Node Architecture	KNL + Haswell	Intel "Haswell" Xeon E5-2698v3	Intel Xeon Phi "Knights Landing"
Processor Type		Dual socket, 16 cores/socket, 2.3 GHz	1 socket, 68 cores/socket, > 3 Tflops/KNL
Memory Capacity	2.07 PiB	1.15 PiB	0.91 PiB
Memory Types		128 GiB DDR4	96 GiB DDR4 + 16GiB HBM
Memory BW	>6PB/sec	>1 PB/s	>1PB/s +>4PB/s
Peak FLOPS	41.5 PF	11.1 PF	30.4 PF
Number of Nodes	19,420	9,436	9,984
Number of Cores	980,864	301,952	678,912
Number of Cabs (incl I/O & BB)	110	54	56
PFS Capacity (usable)	78 PB usable		
PFS Bandwidth (sustained)	1.45 TB/s		
BB Capacity (usable)	3.7 PB		
BB Bandwidth (sustained)	3.3 TB/s		

The Advanced Technology Computing Campaign (ATCC) process is used to allocate resources on Trinity. This is the same resource allocation process that was used for Cielo and is used for Sequoia. The Advanced Technology Planning Advisory Committee (ATPAC) balances the use of the system based on priorities of Programs (DSW, ASC, Sci, ICF, GS). The goal is to effectively allocate and schedule AT computing resources among all three NNSA laboratories for weapons deliverables that merit priority on this class of resource. The

Scheduling Governance Model can be found at http://trinity.lanl.gov/ under the Resources tab.

The Haswell partition debuted as #6 on the November 2015 Top500 list at 8.1 petaflop/s. An open science period was conducted on the Haswell partition from January to March 2016 to shake down the system. The Haswell partition was transitioned to the secure and entered general availability in July 2016. The KNL partition was delivered to Los Alamos in September 2016 and was accepted in December 2016. There will be an open science period on the KNL partition in early 2017, after which it will be transitioned to the secure, merged with the Haswell partition, and enter general availability.

Preparing the applications to use the advanced technology is a key goal for all the AT systems, including Trinity. The Trinity Center of Excellence (CoE) brings together experts from Intel, Cray, and ACES to assist the code teams from all three laboratories with application readiness. The Deep Dives and Discovery Sessions occur at a laboratory site with 3-4 Intel and Cray experts focusing on 2-3 codes. The Dungeon Sessions and Bootcamps occur at the vendor sites with about 10 experts focusing on 3 or 4 codes.

The Trinity Haswell partition is currently providing compute resources to the tri-lab DSW community, and the Trinity KNL partition will enter production in 2017.

Crossroads

Crossroads will be the third Advanced Technology system (ATS-3) scheduled for delivery in 2020. It and NERSC-9 are being procured by the Alliance for application Performance at EXtreme scale (APEX), which is a collaboration between ACES and LBL NERSC. APEX has a history of successful collaborations that started with Cielo and Hopper and continued with the Trinity and Cori procurement. By combining our efforts we have a deeper pool of expertise in system procurement, deployment, and integration activities, as well as broader and different perspectives from partnering labs for the technology evaluation process. Another benefit is that the single larger procurement gets the attention of vendors, and allows them to take more risk, resulting in a deeper and more fruitful consideration of all technology alternatives.

Crossroads will enable an increase in predictive capability while extending the state of the art in both system technology and applications. As shown in Table 2, the main goal is to provide a significant application capability improvement over current platforms (>> Sequoia, > Trinity). The desired increase in application performance drives system improvements. Future higher fidelity models drive an increase in the system aggregate memory capacity. Sustaining time to solution of simulations drives increases in system computational capabilities, memory bandwidth, and scaling characteristics. Improvements in performance of current and future programming models and overall science workflow efficiency drive system improvements in reliability, resilience, energy efficiency, and more. As with Trinity the main KPPs in Table 2 are memory capacity and application performance improvements (not floating point operations per second). APEX is extending our procurement methodology yet again beyond that of Trinity by specifying 90% efficiency of applications instead of I/O speeds.

Table 2. Crossroads draft key performance parameters (KPP); final KPP will be set in CD-2/3b.

Description of Scope	Threshold (Minimum) KPPS	Objective KPPs
Capability Performance Improvement over Trinity KNL	6x	12x
Memory Capacity	3 PB	6 PB
Platform Storage	30x Main memory	>30x Main memory
Peak Power	<18MW	<15MW

The vendor bids were evaluated by APEX in December 2016 and ACES is currently in the approval process with NNSA. A conceptual timeline for the Crossroads procurement is shown in Figure 2.The initial market surveys of the vendors occurred in February 2015, it takes over two years just to get to a contract for such major procurements. CD-0 was approved on July 22, 2015, allowing the Request for Proposals (RFP) to be drafted. Vendor feedback on draft technical specifications and vendor briefings on their technology roadmaps occurred in December 2015. A joint ASC/ASCR Design Review was held in January 19-20, 2016. Final vendor roadmap briefings were in June 2016, followed by CD-1/3a approval on August 5, 2016. The RFP was released to vendors in October 2016, with the vendor bids received in November 2016. ACES expects to award two contracts in 2017, one for the system delivery in 2020, and another for Non-Recurring Engineering (NRE). The NRE contract is a vehicle for ACES to work with the selected vendor to mature technology that will improve the performance of the system. NERSC will separately negotiate two similar contracts for NERSC-9.

Figure 2. Crossroads conceptual procurement timeline.

As with Trinity a Center of Excellence will be part of the Crossroads acquisition. The CoE will allow direct support by the selected vendor subject matter experts to the tri-lab application code teams. Application improvements are one of the main drivers of the Advanced Technology Systems and the long-term exascale strategies.

Ising – D-Wave

Nationally research is underway for when Moore's Law (doubling of transistors per area every two years) no longer holds true in the late 2020's. ASC labels this research as Beyond Moore's Law (BML) and it is focused on what to do when computing performance plateaus?

As part of that effort, ASC has procured a D-Wave 2X ™ adiabatic quantum system that is at LANL. The D-Wave system, named Ising, is intended to begin exploring and learning about new ways to compute. It is open to tri-lab users, and to DOE Office of Science, and to Universities. Information on how to get access to Ising is at http://dwave.lanl.gov/.

Figure 3. A D-Wave 2X adiabatic quantum system.

Acknowledgements

The authors want to acknowledge the many staff at LANL, SNL, LLNL, and NERSC who have been involved with the selection and installation of Trinity, and developing the Crossroads/NERSC-9 RFP and with evaluating the vendor bids. This effort could not have succeeded without the large team. This was apparent at the vendor bids review that had over 120 staff from the four laboratories!