Performance Evaluation of the Intel Sandy Bridge Based NASA Pleiades Using Scientific and Engineering Applications

Subhash Saini, Johnny Chang, Haoqiang Jin

NASA Advanced Supercomputing Division

NASA Ames Research Center

Moffett Field, California 94035-1000, USA

{subhash.saini, johnny.chang, haoqiang.jin}@nasa.gov

Abstract — **We present a performance evaluation of Pleiades based on the Intel Xeon E5-2670 processor, a fourth-generation eight-core Sandy Bridge architecture, and compare it with the previous third generation Nehalem architecture. Several architectural features have been incorporated in Sandy Bridge: (a) four memory channels as opposed to three in Nehalem; (b) memory speed increased from 1333 MHz to 1600 MHz; (c) ring to connect on-chip L3 cache with cores, system agent, memory controller, and QPI agent and I/O controller to increase the scalability; (d) new AVX unit with wider vector registers of 256 bit; (e) integration of PCI-Express 3.0 controllers into the I/O subsystem on chip; (f) new Turbo Boost version 2.0 where base frequency of processor increased from 2.6 to 3.2 GHz; and (g) QPI link rate from 6.4 to 8 GT/s and two QPI links to second socket. We critically evaluate these new features using several low-level benchmarks, and four full-scale scientific and engineering applications.**

Keywords: Intel Sandy Bridge processor, performance evaluation, benchmarking.

I. INTRODUCTION

The Intel Nehalem, a third generation architecture (Xeon 5600 series) introduced in 2009, offers some important initial steps toward ameliorating the memory bandwidth problem [1, 2]. The Intel X5600 launched in 2010 is the Westmere series and it is a 32 nm die shrink of Nehalem. The Nehalem architecture has overcome problems associated with the sharing of the front-side bus (FSB) in previous processor generations by integrating an on-chip memory controller and by connecting the two processors through the Intel QuickPath Interconnect (QPI) and to the input/output (I/O) hub. The result is more than three times greater sustained-memory bandwidth per core than the previous-generation dual-socket architecture. It also introduced hyper-threading (HT) technology (or simultaneous multi-threading, "SMT") and Intel Turbo Boost technology 1.0 ("Turbo mode") that automatically allow processor cores to run faster than the base operating frequency if the processor is operating below rated power, temperature, and current specification limits [3].

However, third generation Nehalem architecture still has performance and scalability bottlenecks due to scalability of L3 cache bandwidth, I/O, limited memory bandwidth, low performance of Turbo Boost, and low HT performance due to inadequate memory bandwidth per thread, low bandwidth between two processors on a node, etc. In 2012, Intel introduced a fourth-generation eight-core architecture Intel Xeon processor E5-2670 ("Sandy Bridge") that introduced new architectural features and extensions and mechanisms, which has significantly improved overall performance [4]. This processor is also used in large-scale heterogeneous systems such as Stampede with co-processor Intel Xeon Phi

based on the Many Integrated Core (code-named Knight's Corner) architecture and Yellowstone [1, 5, 6]. New and extended features of Sandy Bridge architecture are:

a) A ring to connect on-chip L3 cache with cores, system agent, memory controller, and QPI agent and I/O controller to increase the scalability. L3 cache per core has been increased from 2 MB to 2.5 MB.

b) New micro-ops (L0) cache that caches instructions as they are decoded. The cache is direct mapped and can store 1.5 K micro-ops.

c) New Intel Advanced Vector Extensions (AVX) unit with wider vector registers of 256 bit in Sandy Bridge instead of 128 bit in Westmere, thereby doubling the floating-point performance.

d) Integration of PCI-Express 3.0 controllers into the I/O subsystem on chip. PCIe lanes have been increased from 36 to 40. Earlier QPI was used to connect to I/O hub.

e) New Turbo Boost version 2.0 where frequency boost of processor is up to 600 MHz instead of up to 400 MHz.

f) Two QPI links connecting first processor to second processor instead of one link. QPI link rate increases from 6.4 to 8 GT/s.

g) Two loads plus one store per cycle instead of one load plus one store, thereby doubling load bandwidth.

h) Four memory DDR3 channels as opposed to three in Westmere.

i) Memory speed increased from 1333 MHz in Westmere to 1600 MHz in Sandy Bridge.

The potential performance improvement of Sandy Bridge architecture over Nehalem architecture (Nehalem and Westmere processors) is attributed due to increasing three memory channels to four, increasing memory speed from 1333 MHz to 1600 MHz, and new technology/architecture such as ring connecting cores, L3 cache (2.5 MB vs. 2 MB per core), QPI agent, memory controller and I/O controller, and system agent.

In the past, several researchers have evaluated the performance of high performance computing systems [14-20]. To the best of our knowledge, this is the first paper to conduct a:

a) Critical and extensive performance evaluation and characterization of an SGI ICE X cluster based on the

.

Intel Xeon E5-2670, hereafter called "Sandy Bridge", using High Performance Computing Challenge (HPCC) suite, memory latency and bandwidth benchmarks, NAS Parallel Benchmarks (NPB), and four real-world production-quality scientific and engineering applications (Overflow, MITgcm, USM3D, and CART3D) taken from the existing workload of NASA and U.S. aerospace industry [7-13]

b) Detailed comparison of SGI ICE X cluster based on the Intel Xeon E5-2670 connected by 4x FDR IB with an SGI ICE 8400EX based on the Intel Xeon 5670, connected by 4x QDR IB-connected hypercube topology (hereafter called "Westmere") using network latency and bandwidth benchmarks of HPCC suite [7].

c) Detailed performance comparison of AVX and SSE4.2 instructions for Sandy Bridge using NPB and four full-scale applications.

d) Performance evaluation of Turbo Boost 2.0 for Sandy Bridge and its comparison with Turbo Boost 1.0 for Westmere using NPB and four full-scale applications.

e) Performance evaluation of hyper-threading (HT) (or simultaneous multi-threading, "SMT") for Sandy Bridge and Westmere using NPB and four full-scale applications.

f) Measurement of the latency and memory load bandwidth of L1 cache, L2 cache, L3 cache and main memory for Sandy Bridge and Westmere.

The remainder of the paper is organized as follows: Section II provides details of the Pleiades-Sandy Bridge and Pleiades-Westmere systems; in Section III we briefly describe the benchmarks and applications used in the current study; in Section IV we present our results comparing the performance of the two systems; and in Section V we present our conclusions.

II. COMPUTING PLATFORMS

We used NASA's Pleiades supercomputer, an SGI Altix ICE system located at NASA Ames Research Center. Pleiades comprises 11,776 nodes interconnected with an InfiniBand (IB) network in a hypercube topology [1]. The nodes are based on four different Xeon processors from Intel: Harpertown, Nehalem-EP, Westmere-EP and Sandy Bridge. In this study, we used only the Westmere-EP and Sandy Bridge based nodes.

A. Pleiades Sandy Bridge

As shown in Figure 1, the Sandy Bridge-based node has two Xeon E5-2670 processors, each with eight cores. Each processor is clocked at 2.6 GHz, with a peak performance of 166.4 Gflop/s. The total peak performance of the node is therefore 332.8 Gflop/s. Each core has 1.5K μ ops, 64 KB of L1 cache (32 KB data and 32 KB instruction) and 256 KB of L2 cache. All eight cores share 20 MB of last level cache (LLC), also called L3 cache. The on-chip memory controller supports four DDR3 channels running at 1600 MHz, with a peak-memory bandwidth per processor of 51.2 GB/s (and twice that per node). Each processor has two QPI links to

connect with the second processor of a node to form a non-uniform-memory access (NUMA) architecture. The QPI link runs at 8 GT/s ("T" for transfer), at which rate 2 bytes can be transferred in each direction, for an aggregate of 32 GB/s. Each link runs at 16 GB/s in each direction simultaneously [1].
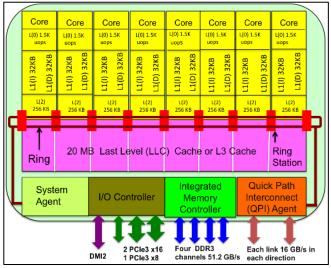

Figure 1.   Schematic diagram of a Sandy Bridge processor.

Following are the new and extended architectural features of Sandy Bridge.

1)   New Features

L0 (μ-ops) Cache: In Sandy Bridge, there is a μ-ops cache that caches instructions as they are decoded. The cache is direct mapped and can store 1.5 K μ-ops. The μ-ops cache is included in the L1(I) cache. The size of the actual L1(I) and L1(D) caches has not changed, remaining at 32 KB each (for total of 64 KB).

Last Level Cache (LLC) / L3 Cache: In Westmere, all cores have their own private path to the L3 cache. Sandy Bridge has a bi-directional 32-byte ring interconnect that connects the 8 cores, the L3-cache, the QPI agent and the integrated memory controller. The ring replaces the individual wires from each core to the L3-cache. The bus is made up of four independent rings: a data ring, request ring, acknowledge ring, and snoop ring. The QPI link agent, cores, L3 cache segments, DDR3 memory controller, and an I/O controller all have stops on this ring bus. The L3 cache is divided into eight slices/blocks, which are connected to the eight cores, and the system agent through a ring interconnect. The red boxes in Fig. 1 are ring stations. Each core can address the entire cache. Each slice gets its own stop station and each slice/block has a full cache pipeline. In Westmere, there is a single cache pipeline and queue that all cores forward requests to, whereas in Sandy Bridge, cache pipeline is distributed per cache slice.

AVX: Intel Advanced Vector Extensions (AVX) is a new set of x86 instruction-set extensions of SSE4.2 [22]. It increases the width of the registers from 128 bits to 256 bits. Each

.

2

register can hold eight single-precision floating-point values or four double-precision floating-point values that can be operated on in parallel using SIMD (single-instruction, multiple-data) instructions. AVX also adds three-register instructions (e.g., z=x+y), whereas previous instructions could only use two registers (x=x+y). Square root and reciprocals vectorize with 128 bit-wide (SSE4.2) but do not vectorize with AVX. In AVX, alignment of data is to 32 bytes boundary, whereas in SSE4.2, it is 16 bytes boundary.

QPI 2.0: In Nehalem/Westmere, one QPI 1.0 link connects the two processors/sockets of the node to form a non-uniform-memory access (NUMA) architecture to do point-to-point communication; the other connects to the IO hub [4]. The QPI link runs at 6.4GT/s, at which rate 2 bytes can be transferred in each direction, for a rate of 12.8 GB/s in each direction per QPI link and a total 25.6 GB/s bidirectional rate per link. In Sandy Bridge, two QPIs at 8.0 GT/s connect the two processors/sockets of the node and deliver 16 GB/s in each direction with a total of 32 GB/s bidirectional. In Westmere, the total inter-processor bandwidth is 51.6 GB/s, whereas in Sandy Bridge, it is 128 GB/s, an increase of 148%.

Memory Subsystem: The improvements to Sandy Bridge's floating-point performance by AVX instruction increase the demands on the load/store units. In Nehalem/Westmere, there are three load and store ports: load, store address, and store data for L1(D) cache. The memory unit can service two memory requests per cycle, i.e., 16 bytes load and 16 bytes store, for a total of 32 bytes per cycle. In Sandy Bridge, the load and store address ports are now symmetric so each port can service a load or store address to L1(D) cache. By adding a second load/store port, Sandy Bridge can handle two loads plus one store per cycle automatically. The memory unit can service three memory requests per cycle, two 16 bytes load and a 16-byte store, for a total of 48 bytes per cycle.

2) Extended Features

Several existing features such as Turbo Boost, HT, the number of memory channels, and the speed of the memory bus of Nehalem architectures (Nehalem-EP, Westmere-EP, etc.) have been significantly enhanced and extended in Sandy Bridge architecture, as described below.

Turbo-Boost 2.0: In Westmere, TB 1.0 provides a frequency-stepping mode that enables the processor frequency to be increased in increments of 133 MHz. The amount of Turbo boost available varies with processor bin. The processor can turbo up to three frequency increments in less than half-subscribed mode—that is, for two or fewer cores per chip busy, the frequency can go up by 3 x 133 MHz and by two bin splits in half-subscribed to fully-subscribed mode (2 x 133 MHz). The frequency is stepped up within the power, current, and thermal constraints of the processor.

In Sandy Bridge TB 2.0, the amount of time the processor spends in the TB state depends on the workload and operating environment, such as the number of active cores, current power consumption and processor temperature.

When the processor is operating below these limits and the workload demands additional performance, the processor frequency dynamically increases until the upper limit of frequency is reached. There are algorithms to manage current, power, and temperature to maximize performance and energy efficiency. The Sandy Bridge processor with a 2.6 GHz clock frequency can boost its frequency up to 3.2 GHz, i.e., an increase of up to 23%.

Hyper-Threading 2.0: Intel provided HT 1.0 in Nehalem. In Sandy Bridge E5-2670, it is enhanced to HT 2.0. HT enables two threads to execute on each core in order to hide latencies related to data access. These two threads can execute simultaneously, filling unused stages in the functional unit pipelines. When one thread stalls, a second thread is allowed to proceed. The advantage of HT is its ability to better utilize processor resources and to hide memory latency. It supports two threads per core, presenting the abstraction of two independent logical cores. The physical core contains a mixture of resources, some of which are shared between threads:

(a) *replicated resources (*register state, return stack buffer, and the instruction queue);

(b) *partitioned resources* (load buffer, store buffer, and reorder buffer);

(c) *shared resources* (L1, L2, and L3 cache); and

(d) *shared resources unaware of the presence of threads* (execution units).

Memory Speed: Memory speed increased from 1333 MHz in Westmere to 1600 MHz in Sandy Bridge, an increase of bandwidth by 20%.

Memory Channels: The number of memory channels increased from 3 in Westmere to 4 in Sandy Bridge, an increase in bandwidth by 33%.

B. Networks Interconnects (FDR and QDR)

The Sandy Bridge nodes are connected to the two fabrics (ib0 and ib1) of the Pleiades InfiniBand (IB) network via the dual-port, four-link fourteen data rate (4x FDR) IB Mezzanine card on each node, as well as via the Mellanox FDR IB switches in the SGI ICE X IB Premium Blade. The FDR runs at 14 Gbits/s per lane. With four links, the total bandwidth is 56 Gbits/s (7 GB/s). On each node, the IB Mezzanine card sits on a sister board next to the motherboard, which contains the two-processor sockets.

There are 18 nodes per Individual Rack Unit (IRU). These 18 nodes are connected to two Mellanox FDR IB switches in an SGI ICE X IB Premium Blade to join the ib0 fabric. Another set of connections between the 18 nodes and a second Premium Blade is established for ib1. However, Westmere nodes are connected via four link quad data rate (4x QDR) IB running at 40 Gbits/s or 5 GB/s. Peak bandwidth of 4x FDR IB is 1.75 times that of 4x QDR (56 Gbits/s vs. 32 Gbits/s).

Table I presents the characteristics of Sandy Bridge and Westmere.

.

TABLE I.   CHARACTERISTICS OF SANDY BRIDGE AND WESTMERE.

Characteristic	Pleiades-Sandy Bridge	Pleiades-Westmere
Processor:		
Processor architecture	Sandy Bridge	Nehalem
Processor type	Intel Sandy Bridge-EP (Xeon E5-2670)	Intel Westmere-EP (Xeon X5670)
Base frequency (GHz)	2.60	2.93
Turbo Boost Version	V2.0, up to 600 MHz	V1.0, up to 400 MHz
Turbo frequency (GHz)	3.2	3.33
Floating/clock/core	8	4
Perf. per core (Gflop/s)	20.8	11.7
Number of cores	8	6
Peak performance	166.4	70.3
L0 (micro-op) Cache	1.5K micro-ops	None
L1 cache size	32 KB (I)+32 KB(D)	32 KB (I)+32 KB(D)
L2 cache size	256 KB/core	256 KB/core
L3 cache size (MB)	20 shared	12 shared
L3 cache network	Ring	Individual links
Memory type	4 channels DDR3 - 2 DIMMS per channel	3 channels DDR3 - 2 DIMMS per channel
Memory speed (MHz)	1600	1333
HyperThreads / core	2	2
I/O controller	On chip	Off chip
PCI Lanes	40 Integrated PCIe 3.0	36 Integrated PCIe 2.0
PCIe 3.0 Speed	8 GT/s	none
Node:		
Number of processors	2	2
Main memory (GB)	32	24
No. of Hype Threads	32	24
Inter socket QPI links	2	1
QPI frequency (GT/s)	8.0	6.4
New instruction	AVX	AES-NI
Number of QPIs	2	1
Perf./node (Gflop/s)	332.8	140.6
Interconnects		
Interconnect type	4x FDR IB	4x QDR IB
Peak network perf. Gbits/s	56	32
Network topology	Hypercube	Hypercube
Compiler, Libraries, operating system and File System:		
Compiler	Intel 12.1	Intel 12.1
MPI library	MPT 2.06	MPT 2.06
Math library	Intel MKL 10.1	Intel MKL 10.1
Type of file system	Lustre	Lustre
Operating system	SLES11SP1	SLES11SP1
System Name	SGI ICE X	SGI ICE 8400EX

III.   BENCHMARKS AND APPLICATIONS

In this section we present a brief description of the benchmarks and applications used in this study.

A. HPC Challenge Benchmarks (HPCC)

The HPCC benchmarks are intended to test a variety of attributes that can provide insight into the performance of high-end computing systems [7]. These benchmarks examine not only processor characteristics but also the memory subsystem and system interconnects.

B. Memory Subsystem Latency and Bandwidth

A deep understanding of the performance of the hierarchical memory system of Sandy Bridge is crucial to understanding application performance. We measured the latency and bandwidth for L1, L2, L3 caches and main memory for both Sandy Bridge and Westmere [8].

C. NAS Parallel Benchmarks (NPB)

The NPB suite contains eight benchmarks comprising five kernels (CG, FT, EP, MG, and IS) and three compact applications (BT, LU, and SP) [9]. We used NPB MPI version 3.3, Class C in our study. BT, LU, and SP are typical of full production-quality science and engineering applications.

D. Science and Engineering Applications

For this study, we used four production-quality full applications representative of NASA's workload.

1) OVERFLOW-2 is a general-purpose Navier-Stokes solver for CFD problems [10]. The code uses finite differences in space with implicit time stepping. It uses overset-structured grids to accommodate arbitrarily complex moving geometries. The dataset used is a wing-body-nacelle-pylon geometry (DLRF6) with 23 zones and 36 million grid points. The input dataset is 1.6 GB in size, and the solution file is 2 GB.

2) CART3D is a high-fidelity, inviscid CFD application that solves the Euler equations of fluid dynamics [11]. It includes a solver called Flowcart, which uses a second-order, cell-centered, finite volume upwind spatial discretization scheme, in conjunction with a multi-grid accelerated Runge-Kutta method for steady-state cases. In this study, we used the geometry of the Space Shuttle Launch Vehicle (SSLV) for the simulations. The SSLV uses 24 million cells for computation, and the input dataset is 1.8 GB. The application requires 16 GB of memory to run.

3) USM3D is a 3-D unstructured tetrahedral, cell-centered, finite volume Euler and Navier-Stokes flow solver [12]. Spatial discretization is accomplished using an analytical reconstruction process for computing solution gradients within tetrahedral cells. The solution is advanced in time to a steady-state condition by an implicit Euler time-stepping scheme. The test case used 10 million tetrahedral meshes, requiring about 16 GB of memory and 10 GB of disk space.

4) MITgcm (MIT General Circulation Model) is a global ocean simulation model for solving the equations of fluid motion using the hydrostatic approximation [13]. The test case uses 50 million grid points and requires 32 GB of system memory and 20 GB of disk to run. It writes 8 GB of data using Fortran I/O. The test case is a ¼ degree global ocean simulation with a simulated elapsed time of two days.

IV. RESULTS

In this section we present our results for low-level benchmarks, HPCC suite, memory subsystem latency and bandwidth benchmarks, NPB, and four full applications (Overflow, Cart3D, USM3D, and MITgcm).

A. Memory Latency and Bandwidth

In this section we present the memory latency and memory load bandwidth of Sandy Bridge and Westmere. Figure 2 shows the memory latency of two systems. It exhibits step function pattern with four steps; each step corresponds to L1 cache, L2 cache, L3 cache and main memory. L1 cache latency is 1.2 ns for both Sandy Bridge and Westmere. L2 cache latency is 3.5 ns and 3 ns for Sandy Bridge and Westmere respectively. L3 cache latency is 6.5 ns for both Sandy Bridge and Westmere. Main memory latency is 28 ns and 24 ns for Sandy Bridge and Westmere respectively. L2 cache latency and main memory latency is higher on Sandy Bridge than that on Westmere.
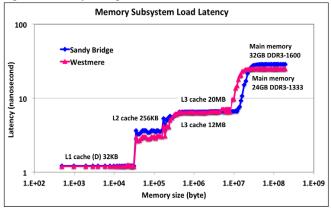

Figure 2. Memory latency of Westmere and Sandy Bridge.

Figure 3 shows the memory load bandwidth of L1 cache, L2 cache, L3 cache and main memory for the two systems. Read and write bandwidth is higher on Sandy Bridge than on Westmere except for L3 cache, where it higher on Westmere. The reason for higher read bandwidth is due to the fact that Sandy Bridge has two memory loads compared to one memory load in Westmere.
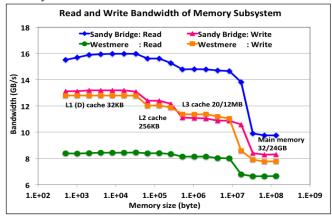

Figure 3. Memory load bandwidth of Westmere and Sandy Bridge.

B. HPC Challenge Benchmarks (HPCC)

In this section we present results for HPCC Version 1.4.1 benchmarks for two systems [7]. We discuss the intra-node and inter-node performance separately.

1) Intra-Node HPCC Performance

In this section we present the intra-node HPCC results for Westmere and Sandy Bridge. In Figure 4 we show the performance of a subset of HPCC suite benchmarks. The performance gains by Sandy Bridge are 66%, 64%, 65%, 66%, 80%, and 141% for G-FFTE, EP-STREAM, G-Random Access, G-PTRANS, EP-DGEMM, and G-HPL, respectively, over Westmere. The performance of Sandy Bridge is superior to that of Westemere due to faster memory speed, extra memory controller, larger L3 cache, higher Gflop/s per core, etc.
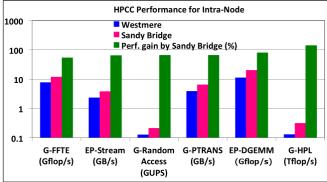

Figure 4. Performance of HPCC on Westmere and Sandy Bridge nodes.

Figures 5 and 6 show the network latency and bandwidth for the intra-node Westmere and Sandy Bridge. The minimum latency corresponds to communication within the socket and the maximum latency across two sockets. Both intra-socket and inter-socket latency is higher for Sandy Bridge than for Westmere. The reason for this is that Sandy Bridge has a ring connecting all the cores to L3 cache, whereas for Westmere, the cores are individually connected with wires. However, the ring bus makes Sandy Bridge more scalable than Westmere and is the method of choice in the new Intel Xeon Phi (MIC), which uses the same ring bus for 60 cores on a die. The higher bandwidth of Sandy Bridge is due to two QPIs connecting the two sockets as opposed to one QPI in Westmere.
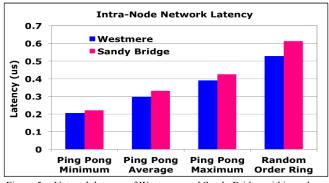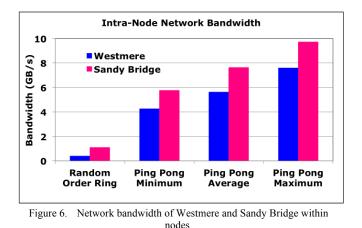

Figure 5. Network latency of Westmere and Sandy Bridge within nodes.

Figure 6. Network bandwidth of Westmere and Sandy Bridge within nodes.

2) *Inter-Node HPCC Performance*

In this section we present inter-node HPCC results for the two systems [7]. In Figure 7, we plot performance of the compute-intensive embarrassingly parallel (EP) DGEMM (matrix-matrix multiplication) for the two systems. The theoretical one-core peak for Sandy Bridge is 20.8 Gflop/s, and for Westmere it is 11.7 Gflop/s. When using Turbo mode on Westmere, the processor core frequency can be increased by up to three 133 MHz increments, raising its peak to 13.32 Gflop/s. The performance gain by Sandy Bridge is 20% to 30% for numbers of cores ranging from 16 to 512 due to the fact that it has higher compute performance per core and has a 20% faster memory speed (1600 MHz vs. 1333 MHz).
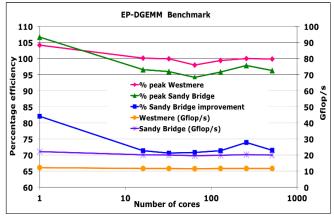

Figure 7. Performance of EP-DGEMM on Westmere and Sandy Bridge.

In Figure 8, we plot performance of the compute-intensive global high-performance LINPACK (G-HPL) benchmark. For both Sandy Bridge and Westmere we give the efficiency for their base frequencies of 2.6 GHz and 2.93 GHz, respectively. The efficiency of Westmere is higher than that of Sandy Bridge and decreases gradually from 16 to 512 cores. In addition, the efficiency of Westmere is higher than that of Sandy Bridge in the entire range of cores except for 16 and 512 cores. The performance gain by Sandy Bridge in terms of floating-point operations is 68% to 87% better than that on Westmere due to better memory bandwidth per core and better Gflop/s performance per core.
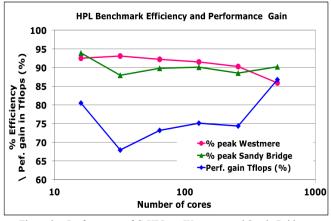

Figure 8. Performance of G-HPL on Westmere and Sandy Bridge.

In Figure 9, we show memory bandwidth for each system using the EP-Stream Triad benchmark. For a single core, the measured bandwidths were 14 GB/s and 9.6 GB/s for Sandy Bridge and Westmere, respectively, i.e., 45.8% higher for Sandy Bridge due to faster memory speed (1600 vs. 1333 MHz; 20% faster on Sandy Bridge) and larger L3 cache (2.5 MB vs. 2 MB per core; 25% larger cache on Sandy Bridge). For 16 cores, these values decreased to 3.8 GB/s and 2.6 GB/s due to memory contention. The aggregate node level bandwidth for Sandy Bridge in fully subscribed mode was then 3.8 x 16 = 60.8 GB/s, which translates into 59 percent of peak-memory bandwidth (102.4 GB/s per node = 2 processors x 4 channels x 8 bytes x 1600 MHz per processor). The faster memory bus enables Sandy Bridge to deliver both higher peak-memory bandwidth and efficiency, producing significant advantages for memory-intensive codes.
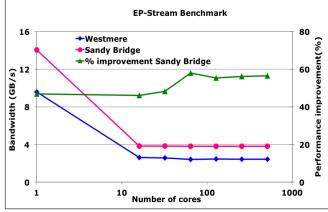

Figure 9. Performance of EP-STREAM on Westmere and Sandy Bridge.

In Figure 10, we show the minimum, average and maximum Ping-Pong latency for Westmere and Sandy Bridge. The minimum latency on both systems is around 0.25 μs, this corresponding to latency within a processor/socket. The maximum latency on both systems is around 2 μs, except for 16 cores where latency for Sandy Bridge is 74% lower than that on Westmere. This is because for Westmere, one needs two nodes (12 cores each), whereas

one needs only one Sandy Bridge node (16 cores). The average latency of Sandy Bridge is lower than Westmere by 12% to 24%, except for 16 cores where it is better by 60%.
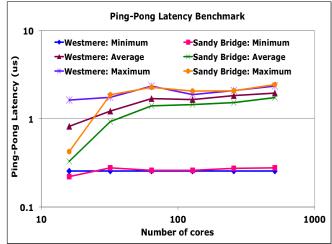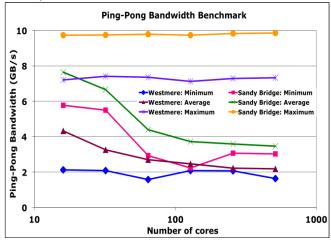

Figure 10. Ping-Pong Latency on Westmere and Sandy Bridge.

Figure 11 shows the minimum, average and maximum ping-pong bandwidth for Westmere and Sandy Bridge. The maximum bandwidth is within 16 cores on one Sandy Bridge node and two Westmere nodes. The maximum bandwidths are 9.8 GB/s and 7.2 GB/s for Sandy Bridge and Westmere, respectively. The reason for this is that for 16 cores, Sandy Bridge has two sockets with 8 cores each connected via 2 QPI with 8 GT/s, whereas Westmere has 2 sockets with 6 cores each connected via one QPI of 5 GT/s and it is via QDR to another node. For a large number of cores, bandwidth is again higher in Sandy Bridge nodes than Westmere nodes, as the former is connected by FDR and latter by QDR.

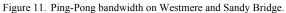Figure 11. Ping-Pong bandwidth on Westmere and Sandy Bridge.

Figure 12 shows the Random Order Ring (ROR) latency for Sandy Bridge and Westmere. For 16 cores, latency for Sandy Bridge is lower than that of Westmere because in the former it is intra-node latency, whereas in the latter it is inter-node latency. For numbers of cores ranging from 32 to 512, latency for Sandy Bridge is higher than that of Westmere by 11% to 70%.
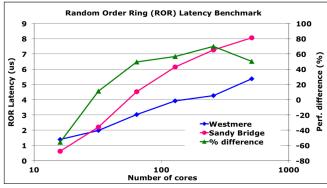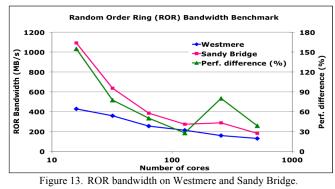

Figure 12. ROR latency on Westmere and Sandy Bridge.

Figure 13 shows the ROR bandwidth of Sandy Bridge and Westmere for numbers of cores ranging from 16 to 512. In the range of 32 to 512 cores, the bandwidth on Sandy Bridge is always higher than that of Westmere by 38% to 80%. At 16 cores, bandwidth is higher on Sandy Bridge than that on Westmere by 155% because in the former it is intra node and in the latter it is inter node via IB QDR.

Figure 13. ROR bandwidth on Westmere and Sandy Bridge.

In Figure 14, we plot performance of the Random Access benchmark as Giga Updates per second (GUP/s) for 16 to 512 cores for the two systems. In the entire range of cores we studied, the performance was much better on Sandy Bridge than on Westmere. Up to 32 cores, the performance on Sandy Bridge is higher than that on Westmere by 17%. The superior performance on Sandy Bridge is due to the FDR IB and higher memory bandwidth. Scaling is very good on Sandy Bridge and Westmere because of the almost constant bisection bandwidth for 512 cores of the hypercube topology used in these two systems.
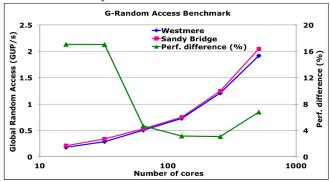

Figure 14. GUP benchmark on Westmere and Sandy Bridge.

In Figure 15, we plot the performance of the PTRANS benchmark for the two systems. The benchmark performance primarily depends on the network and to a lesser extent on memory bandwidth. At 512 cores, it was 74 GB/s for Sandy Bridge and 51.3 GB/s on Westmere. The performance was better by 44% on Sandy Bridge than on Westmere due to the use of FDR IB. Scaling of the benchmark was very good on both systems because of the constant bisection bandwidth for the relatively small number of cores (up to 512) on these two systems.
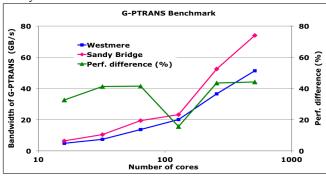

Figure 15. Performance of PTRANS on Westmere and Sandy Bridge.

In Figure 16, we plot the performance of the G-FFT benchmark on Sandy Bridge and Westmere. The benchmark's performance depends on a combination of flops, memory, and network bandwidth. The FDR IB and higher sustained-memory bandwidth enable Sandy Bridge to outperform Westmere. Scaling was better on Sandy Bridge than on Westmere. At 512 cores, performance was 166.2 and 123.4 Gflop/s on Sandy Bridge and Westmere, respectively. We note that the performance on Sandy Bridge is especially high at 16, 64, and 256 cores. The reason for this is that for Sandy Bridge, these numbers correspond to whole number of 1, 4 and 16 nodes, whereas for Westmere, they correspond to 2, 6 and 22 nodes. FFT involves all-to-all communication, which takes much longer in the case of Westmere due to poor network (QDR IB vs. FDR IB).
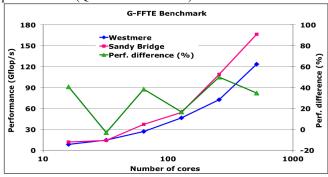

Figure 16. Performance of G-FFTE on Westmere and Sandy Bridge.

C. Science and Engineering Applications

In this subsection, we focus on the comparative performance of four full applications (Overflow, Cart3D, USM3D, and MITgcm) on the two systems [10-13]. The time for all the four applications is for the main loop, i.e., compute and communication time, and does not include I/O time.

Figures 17-20 provide the performance and scalability of the four full-scale applications used in this study. Each figure shows the scaling performance on the Sandy Bridge and Westmere systems along with the percentage performance gain on Sandy Bridge.

Overflow: Figure 17 shows time per step for 16 cores to 512 cores for Overflow. The performance of Overflow on Sandy Bridge is much better than on the Westmere system across the entire range of cores. The Overflow performance on Sandy Bridge is higher by 29% to 46% for cores ranging from 16 to 512 cores. Overflow is a memory-intensive application, and therefore performance was better on Sandy Bridge than on Westmere because memory bandwidth per core of the former is better (3.8 vs. 2.6 GB/s), an advantage of 46%. About 20% of the performance gain of Overflow on Sandy Bridge is from faster memory speed (1600 MHz vs. 1333 MHz). In addition, Sandy Bridge has an advantage, especially for large numbers of cores, as its L3 cache is 2.5 MB per core compared with 2 MB per core of L3 for Westmere, which translates into a gain of 25%.
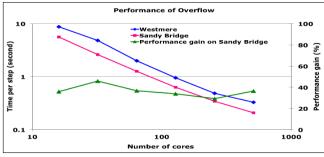

Figure 17. Time per step for Overflow on Westmere and Sandy Bridge.

Cart3D: Figure 18 shows the time to run Cart3D for 16 cores to 512 cores on the two systems. The performance of Cart3D was higher on Sandy Bridge than on Westmere by about 20% due to faster memory speed (1600 MHz vs. 1333 MHz). Using Intel Performance Monitor Unit (PMU) we found that Cart3D is only 1% vectorized so it does not benefit from 256-bit long vector pipeline of Sandy Bridge [21].
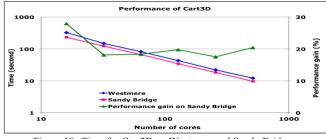

Figure 18. Time for Cart3D on Westmere and Sandy Bridge.

USM3D: Figure 19 shows the USM3D cycle time per step for a range of cores. USM3D is an unstructured mesh-based application that solves a sparse matrix by the Gauss-Seidel method and uses indirect addressing. The L2/L3 caches are poorly utilized, and almost the entire data has to come from main memory. Using PMU, we found that 72% of the data comes from the main memory [21]. Being memory-

intensive, its performance depends exclusively on the memory bandwidth, which is highest for Sandy Bridge (3.8 GB/s) and lowest for Westmere (2.6 GB/s). Because of indirect addressing, USM3D cannot make use of the 256-bit long vector pipe for Sandy Bridge, as it cannot be vectorized. The performance of USM3D was better on Sandy Bridge than on Westmere by 20% to 25%, consistent with the faster memory speed of Sandy Bridge (1600 vs. 1333 MHz), a gain of 20%.
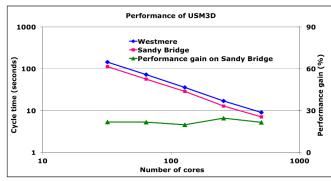

Figure 19. Time for USM3D on Westmere and Sandy Bridge.

MITgcm: Figure 20 shows the time to run the climate modeling application MITgcm [13]. This code is memory-bound and network bound. Since Sandy Bridge provides the higher memory bandwidth (3.8 GB/s vs 2.6 GB/s) and better network (FDR IB vs QDR IB), MITgcm performs much better on this system than on Westmere by about 40%.
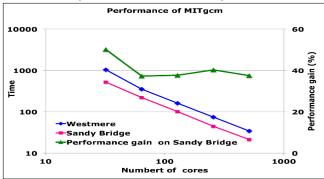

Figure 20. Time for MITgcm on Westmere and Sandy Bridge.

Figure 21 shows a summary of the performance gain by Sandy Bridge over Westmere for four applications: Overflow, Cart3D, USM3D and MITgcm. Using PMU, we found that USM3D and Cart3D have a low vectorization of 20% and 1% respectively and thus cannot make use of the 256-bit long vector pipe [21]. However, both the applications are memory bound; therefore, they benefit from faster memory speed (1600 MHz vs. 1333 MHz; 20% faster on Sandy Bridge) and exhibit performance gains of 17% to 20%. On the other hand, the other two applications, Overflow and MITgcm, have 64% and 50% vectorization, respectively, and are also memory-bound, so their performance gain is much higher (20% to 50%).
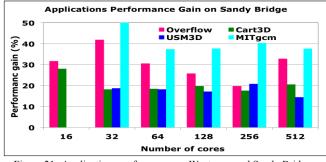

Figure 21. Applications performance on Westmere and Sandy Bridge.

D. Performance Impact of Turbo Boost

In this subsection, we compare results for the MPI version of the NPB with Turbo Boost on and off. Figure 22 shows the measured performance gain of Turbo Boost on Sandy Bridge over Westmere. We ran six NPBs (MG, SP, CG, FT, LU, and BT) for numbers of cores ranging from 16 to 512. We tabulated performance in Gflop/s in both modes on Sandy Bridge and Westmere and calculated the performance gain by Sandy Bridge. The performance gain was in the range of 1% to 10%. In general, Sandy Bridge enjoys a much higher performance gain using Turbo Boost than Westmere except for MG and FT at 512 cores, where Turbo Boost degrades the performance by 1.7% to 3.2%.
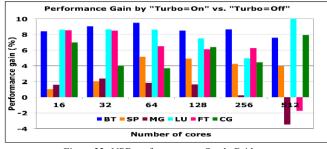

Figure 22. NPB performance on Sandy Bridge.

Figure 23 shows the performance gain in Turbo mode for Sandy Bridge for the applications Overflow, Cart3D, USM3D and MITgcm. The performance gain due to Turbo mode by Overflow and Cart3D is 8% to 10%. For MITgcm and USM3D, the performance gain is about 3% at lower numbers of cores and 6% to 7.5% for higher number of cores.
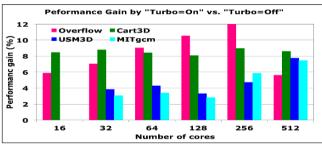

Figure 23. Applications performance on Sandy Bridge.

E. Performance Impact of AVX

Figure 24 shows the performance gain of AVX in Sandy Bridge for the Class C size of six NPB benchmarks [22]. The

largest difference between the AVX and SSE4.2 is for the compute intensive benchmarks (e.g.,BT and LU) and the least gain is by memory-bound benchmarks (e.g., MG and SP). We see for BT the benchmark AVX version versus SSE 4.2 version gives 6-10% improvement, whereas it is 6% for EP, 2-4% for FT, 7-12% for LU, 2-4% for MG, and 1-5% for SP. CG is the only benchmark whose performance degrades in AVX mode. CG uses a sparse BLAS-2 (sparse matrix times vector) and involves indirect addressing, and as such, it cannot be vectorized so unable to use the vector pipeline.
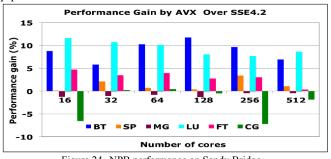

Figure 24. NPB performance on Sandy Bridge.

Figure 25 shows the performance gain in AVX in Sandy Bridge for the four applications. The performance gain for these applications is almost insignificant and ranges from +2% to -3%. Cart3D shows higher performance in AVX mode.However, memory bound applications such as Overflow, MITgcm, and USM3D don't benefit from AVX; in fact, their performance degrades.
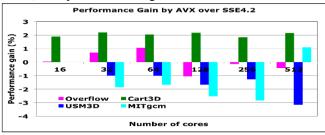

Figure 25. Applications performance on Sandy Bridge.

F. Impact of Hyper-Threading

In Figures 26 and 27, we show the performance gain from HT by Overflow, Cart3D, USM3D and MITgcm on Sandy Bridge and Westmere. With HT, the node can handle twice as many processes (32/24) as without HT (16/12). With more processes per node, there is greater communication overhead. In other words, more processes compete for the same host channel adapter (HCA) on the node. On the other hand, additional processes (or threads) can steal cycles in cases of communications imbalance or memory access stalls. The result is better overall performance for main memory bound applications. For example, USM3D, where 70% of the data comes from main memory because of indirect addressing, can't reuse the L2/L3 cache and thus gets an opportunity to hide the memory latency. Cart3D also benefits from HT as the code is 99% scalar and has more opportunities to schedule the instructions in the pipeline. Overflow and MITgcm are 64%

and 51% vectorized, respectively, so they do not benefit from HT as there is saturation of floating point units. The reason why Overflow does not benefit from HT is because it is very cache-sensitive. Running in HT mode reduces the amount of L3 cache available to each process, so data has to be fetched from main memory instead of from L3 cache, causing HT performance to suffer [3]. On Sandy Bridge, the performance gain by HT for USM3D and Cart3D is almost two times that on Westmere.
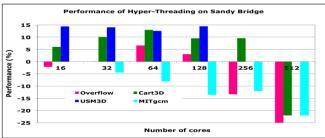

Figure 26. Applications performance gain from HT on Sandy Bridge.
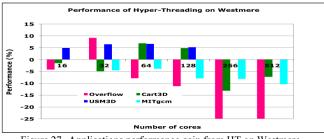

Figure 27. Applications performance gain from HT on Westmere.

V. CONCLUSIONS

In this paper, we conducted a comprehensive performance evaluation and analysis of the Pleiades-Sandy Bridge computing platform, using low-level benchmarks, the NPB, and four NASA applications. Our key findings are as follows:

- The revamped Turbo Boost 2.0 overclocking mechanism on Sandy Bridge is more efficient than the prior implementation TB 1.0 on Westmere. The impact of Turbo Boost in Sandy Bridge is almost doubled relative to Westmere (9% vs. 4%).
- The advantage of AVX over SSE4.2 instructions is insignificant, ranging from +2% to -3%.
- The performance of Hyper-Threading technology on Sandy Bridge is much better than that on Westmere and is helpful in some cases, but for HPC applications this is not universal. The impact of Hyper-Threading on Sandy Bridge is almost doubled that on Westmere for USM3D and Cart3D (10% vs. 4%).
- The memory bandwidth of Sandy Bridge is about 40% higher than that of Westmere.
- The performance of 4x FDR IB is 40% better than that of 4x QDR IB.
- The overall performance of Sandy Bridge is about 20% to 40% better than that of Westmere for the NASA workload.

.

REFERENCES

[1] http://www.nas.nasa.gov/hecc/resources/pleiades.html

[2] S. Saini, A. Naraikin, R. Biswas, D. Barkai and T. Sandstrom, "Early performance evaluation of a "Nehalem" cluster using scientific and engineering applications", Proceedings of the ACM/IEEE Conference on High Performance Computing, SC 2009, November 14-20, 2009, Portland, Oregon, USA.

[3] S. Saini, H. Jin, R. Hood, D. Barker, P. Mehrotra and R. Biswas, "The impact of hyper-threading on processor resource utilization in production applications", **Best Paper**, 8th International Conference on High Performance Computing, HiPC 2011, Bengaluru, India, December 18-21, 2011.

[4] Intel Xeon Benchmark - Intel.com, www.**intel**.com/**Xeon**

[5] Texas Advanced Computing Center - *Stampede*, www.tacc.utexas.edu/**stampede**

[6] NCAR-Wyoming Supercomputing Center (NWSC): https://www2.cisl.ucar.edu/resources/yellowstone/hardware

[7] HPC Challenge Benchmarks, http://icl.cs.utk.edu/hpcc/

[8] R. Schöne, D. Hackenberg, D. Molka, "Memory performance at reduced CPU clock speeds: an analysis of current x86_64 processors", In Proceedings of the 2012 USENIX conference on Power-Aware Computing and Systems (HotPower'12), October 7, 2012, Hollywood, USA http://dl.acm.org/citation.cfm?id=2387869.2387878

[9] NAS Parallel Benchmarks (NPB). http://www.nas.nasa.gov/publications/npb.html

[10] OVERFLOW, http://aaac.larc.nasa.gov/~buning/

[11] D. J. Mavriplis, M. J. Aftosmis, and M. Berger. High Resolution Aerospace Applications using the NASA Columbia Supercomputer, in: Proc. ACM/IEEE SC05, Seattle, WA, 2005.

[12] USM3D: http://tetruss.larc.nasa.gov/usm3d/

[13] M.I.T General Circulation Model (MITgcm), http://mitgcm.org/

[14] S. Saini, D. Talcott, D. Jespersen, J. Djomehri, H. Jin, and R. Biswas. "Scientific application-based performance comparison of SGI Altix 4700, IBM POWER5+, and SGI ICE 8200 supercomputers, High Performance Computing, Networking, Storage and Analysis, SC 2008, Austin, Texas, Nov. 15-21, 2008.

[15] Vitali Morozov, Kalyan Kumaran, Venkatram Vishwanath, Jiayuan Meng,and Michael E. Papka, Early Experience on the Blue Gene/Q Supercomputing System, IEEE IPDPS, Boston, May 20-23, 2013.

[16] Kevin Barker, Kei Davis, Adolfy Hoisie, Darren J. Kerbyson, Mike Lang, Scott Pakin, Jose C. Sancho. Entering the Petaflop Era: The Architecture and Performance of Roadrunner. In *Proceedings of IEEE/ACM Supercomputing (SC08)*, Austin, TX, November 2008.

[17] Kevin Barker Adolfy Hoisie Darren Kerbyson, An Early Performance Analysis of POWER7-IH HPC Systems, SC11, Seattle, Nov.12-18, 2011

[18] Darren J. Kerbyson, Kevin J. Barker, Abhinav Vishnu, Adolfy Hoisie: Comparing the Performance of Blue Gene/Q with Leading Cray XE6 and InfiniBand Systems. ICPADS 2012: 556-563

[19] S. Alam, R. Barrett, M. Bast, M. Fahey, J. Kuehn, C. McCurdy, J. Rogers, P. Roth, R. Sankaran, J. Vetter, P. Worley, and W. Yu, *Early Evaluation of IBM BlueGene/P,* in Proceedings of the ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis (SC08), Austin, TX, November 15-21, 2008

[20] S. R. Alam, R. F. Barrett, M. R. Fahey, J. A. Kuehn, O. E. Messer, R. T. Mills, P. C. Roth, J. S. Vetter, and P. H. Worley, *An Evaluation of the ORNL Cray XT3,* International Journal for High Performance Computer Applications, **22**, February 2008, pp. 52-80.

[21] *PMU Performance Monitoring PerfMon | Intel® Developer Zone*, *software.**intel**.com/en-us/tags/18842*

[22] Intel® ArchitectureInstruction Set Extensions Programming Reference, 319433-014 August 2012,

http://software.intel.com/en-us/avx.

.