

HPC Benchmarks and Applications Performance Study on Broadwell-EP 4S Processor

Author: Neha Kashyap, August 2016 (HPC Innovation Lab)

The intent of this blog is to illustrate and analyze the performance obtained on Intel Broadwell-EP 4S processor with the focus on HPC applications, two synthetic benchmarks along with four applications namely, High Performance Linpack (HPL), STREAM, Weather Research and Forecasting (WRF), NANoscale Molecular Dynamics (NAMD) and Fluent. These runs have been performed on a standalone single server PowerEdge R830. Combinations of System BIOS profiles and Memory Snoop modes are compared for better analysis.

Table 1: Details of Server and Applications used with Intel Broadwell processor

Server	Dell PowerEdge R830
Processor	4 x E5-4669 v4 @2.2GHz, 22 core, 135W, 55M L3 Cache AVX Base Frequency @1.7GHz
Memory	32 x 16GB DDR4 @ 2400MHz (Total=512GB)
Power Supply	2 x 1600W
Operating System	Red Hat Enterprise Linux 7.2 (3.10.0-327.el7.x86_64)
BIOS options	System profile – Performance and Performance Per Watt (DAPC) Snoop modes – Cluster on Die (COD) and Home Snoop (HS) Logical Processor and Node Interleaving – Disabled I/O Non-Posted Prefetch – Disabled
BIOS Firmware	1.0.2
iDRAC Firmware	2.35.35.35
STREAM	v5.10
HPL	From Intel MKL (Problem size is 253440)
Intel Compiler	From Intel Parallel studio 2016 update3
Intel MKL	11.3
MPI	Intel MPI – 5.1.3
NetCDF	4.4.0
NetCDF-Fortran	4.4.2
FFTW	2.1.5
WRF	3.8

NAMD	2.11
Ansys Fluent	v16.0

The server used for obtaining results is Dell’s PowerEdge 13th generation server. It is a high-performance four-socket, 2U rack server. It supports massive memory density (up to 3TB). The Intel® Xeon® Processor E5-4669 v4 (Product Family E5-4600 v4) is a 14nm die supporting two snoop modes in 4 socket systems - Home Snoop and Cluster on Die (COD). It is based on the micro-architecture code named Broadwell-EX.

Default snoop mode being Home Snoop mode. Cluster on Die is available only on processor models having more than 12 cores on Intel’s Broadwell series. In this mode, there is a logical split in the socket into two NUMA domains and are exposed to the operating system. The total number of cores, level 3 Cache is divided equally in the two sliced NUMA domains. Each having one home agent with equal number of cores cache slices in each NUMA domain. The NUMA domain (cores plus home agent) is called a cluster. The COD mode is best suitable for highly NUMA optimized workloads.

The **STREAM** is a synthetic HPC benchmark. It evaluates the sustained memory bandwidth in MB/s by counting only the bytes that the user program requested to be loaded or stored to the memory. The “TRIAD function” score reported by this benchmark is used to analyze the performance of stream bandwidth test. Operation carried out by TRIAD is - TRIAD: $a(i) = b(i) + q*c(i)$

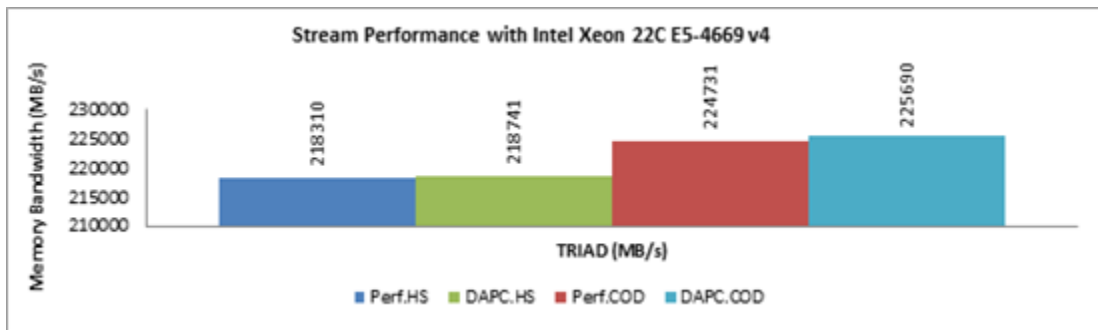


Figure 1: Memory Bandwidth with STREAM

From figure 1 it can be observed that the DAPC.COD performs the best. DAPC and Performance profiles have similar performance, the memory bandwidth varies slightly from 0.2% - 0.4% for HS and COD mode respectively. COD snoop mode performs ~2.9-3.0% better than HS.

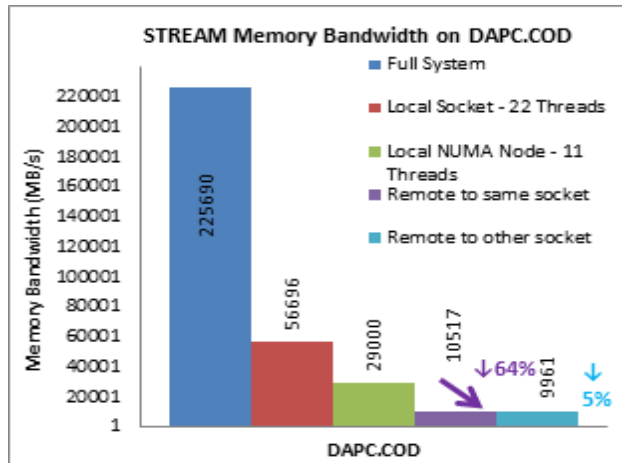


Figure 2: STREAM Memory Bandwidth on DAPC.COD

Here, figure 2 shows the performance obtained with STREAM Triad memory bandwidth in DAPC.COD configuration taking into account local, local NUMA and remote bandwidth. The full system memory bandwidth is ~ 226 GB/s.

To obtain the local memory bandwidth, the processes are bind to a socket and only the memory local to that socket is accessed (same NUMA node with NUMA enabled). The local NUMA node with 11 threads gives almost half the performance when compared to local socket with 22 threads since the number of processors becomes half. For remote memory, the memory bandwidth from remote to same socket drops by 64% while from remote to other socket, the processes are bind to another socket and memory which is remote to that socket is accessed through QPI link (remote NUMA node). Since there is limitation to QPI bandwidth, the remote memory bandwidth of remote to other socket is going down another 5% when compared to remote to same socket.

High Performance LINPACK (HPL) is an industry standard compute intensive benchmark. It is traditionally used to stress the compute and memory subsystem. It measures the speed with which computer solves linear equations and calculates a system's floating-point computing power. Intel's Math Kernel Library (software library to perform numerical linear algebra on digital systems) is required by it.

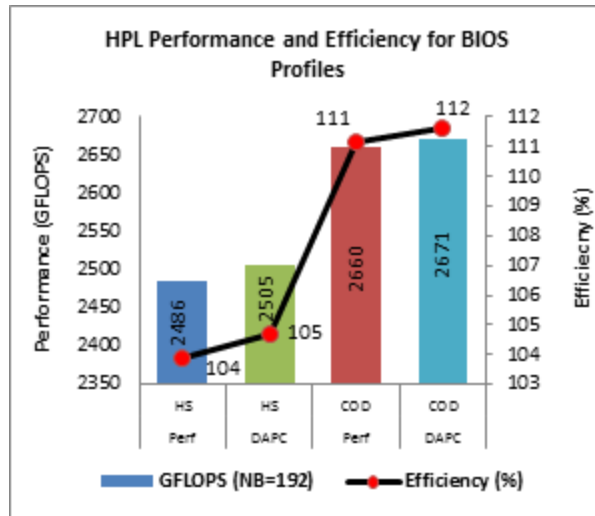


Figure 3: HPL Performance and Efficiency

Figure 3 illustrates HPL performance benchmark. From this graph, it is clear that DAPC and Performance profiles are giving almost similar results whereas there is difference when comparing HS with COD. With DAPC, COD is 6.2% higher than HS whereas with Performance profile, COD yields 6.6% higher performance than HS. The HPL performance is analyzed in “GFLOP/second”. The Efficiency is more than 100% because of the AVX base frequency.

The **Weather Research and Forecasting (WRF)** Model is a next-generation mesoscale numerical weather prediction used for both atmospheric research and weather forecasting needs. It generates atmospheric simulations using real data (observations, analysis) or idealized conditions. It features two dynamical cores, a data assimilation system, and a software architecture which allows for parallel computation and system extensibility. For this study, CONUS12km (small) and CONUS2.5km (large) data set is taken and the “[Average Time Step](#)” computed is used as a metric to analyze the performance.

[CONUS12km is a single domain and medium size \(48hours, 12km resolution case over the Continental U.S. \(CONUS\) domain from October 24, 2001\) benchmark with a time step of 72 seconds. CONUS2.5km is a single domain and large size \(Latter 3hours of a 9hours, 2.5km resolution case over the Continental U.S. \(CONUS\) domain from June 4, 2005 with a time step of 15 second\).](#)

The [number of tiles is chosen based on best performance obtained by experimentation](#) and is defined by setting the environment variable “[WRF_NUM_TILES = x](#)” where x denotes number of tiles. The application was compiled with the “sm + dm” mode. The combinations of MPI and OpenMP processes that were used are as follows:

Table 3: WRF Application Parameters used with Intel Broadwell processor

E5-4669v4	CONUS12km			CONUS2.5km			
	Total no. of cores	MPI Processes	OMP Processes	TILES	MPI Processes	OMP Processes	TILES
	88	44	2	44	44	2	56

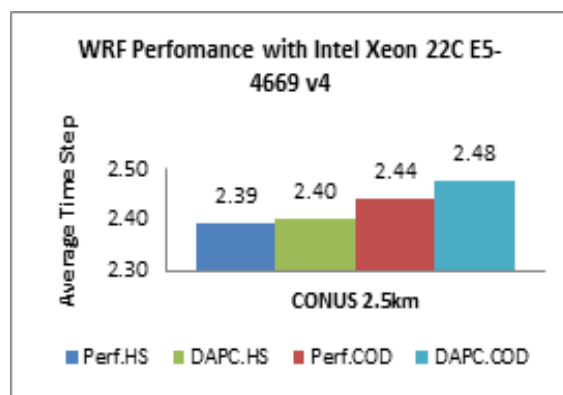
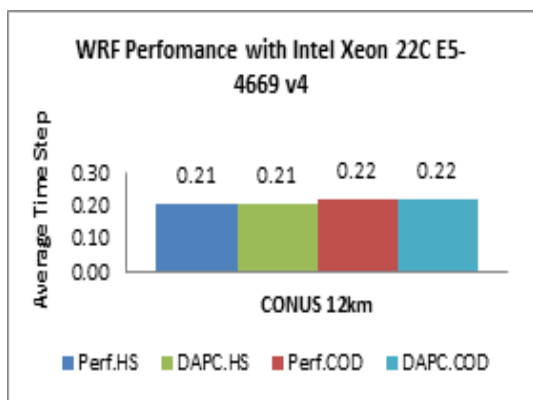


Figure 4: Performance with WRF, CONUS12km **Figure 5: Performance with WRF, CONUS2.5km**

Figure 4 illustrates CONUS 12km. With COD there is 4.5% improvement in the average time step when compared with HS. While for CONUS2.5km, DAPC and Performance profiles are showing a variance from 0.4% to 1.6% for HS and COD respectively. COD performs ~2.1% - 3.2% better than HS. Due to the large dataset size, CONUS2.5km can more efficiently utilize larger number of processors. For both datasets DAPC.COD performs the best.

NAMD is a molecular dynamics research application which is portable, parallel and object oriented designed specifically for high-performance simulations of large bio molecular systems. It is developed using charm++. For this study, the three widely used datasets have been taken, namely ApoA1 (92,224 Atoms) – standard NAMD cross-platform benchmark, F1ATPase (327,506 Atoms) and STMV (virus, 1,066,628 Atoms) – useful for demonstrating scaling to thousands of processors. ApoA1 is the smallest and STMV being the largest dataset.

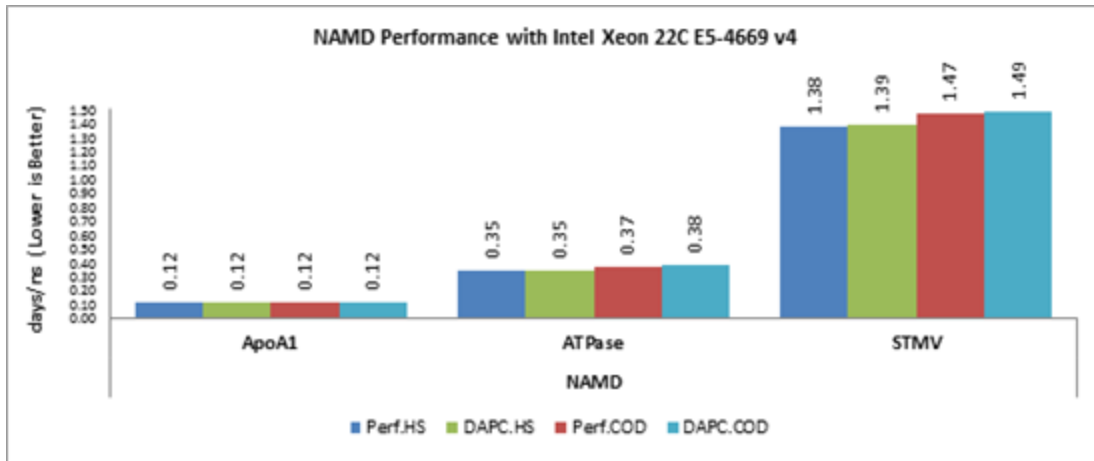


Figure 6: Performance of NAMD on BIOS Profiles (The Lower the Better)

The performance obtained on ApoA1 across BIOS profiles is same while for ATPase it is almost similar. The difference is visible in case of STMV (larger dataset) over other two datasets since the number of atoms increases which in turn helps in proper utilization of processors. In case of STMV, DAPC and Performance profiles vary from 0.7% to 1.3% for HS and COD respectively. Also the COD snoop mode performs ~6.1% - 6.7% better than HS.

Ansys Fluent is a powerful computational fluid dynamics (CFD) software tool. “Solver Rating” (the higher the better) is considered as a metric to analyze the performance on six input data sets for Fluent.

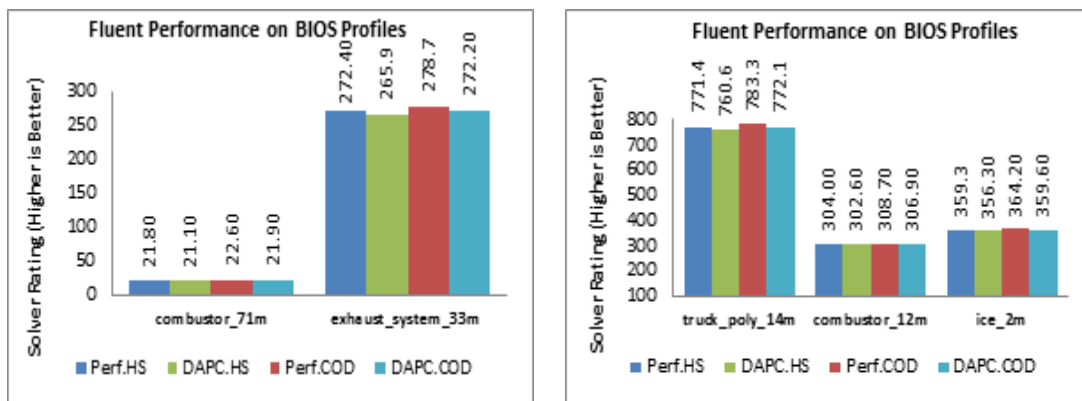


Figure 7: Performance comparison of Fluent on BIOS Profiles (The Higher the Better)

[The Perf.COD is expected to perform best with Fluent.](#) For all datasets the Solver Rating varies from 2-5%.

To conclude, the R830 platform is performing upto the mark with all expected output. It is a good pick for HPC workloads giving best results with DAPC.COD system BIOS profile. It is a great choice in terms of overall system architecture improvements and supporting latest processor.