

Application Performance of Intel Skylake and Intel Knights Landing Processors on Stampede2

June 2018

H17212

White Paper

Abstract

This white paper presents the significant performance advantage of the Intel Skylake processor at a multinode scale in four commonly used applications: NAMD, LAMMPS, Gromacs, and WRF.

Dell EMC Solutions



Copyright

The information in this publication is provided as is. Dell Inc. makes no representations or warranties of any kind with respect to the information in this publication, and specifically disclaims implied warranties of merchantability or fitness for a particular purpose.

Use, copying, and distribution of any software described in this publication requires an applicable software license.

Copyright © 2018 Dell Inc. or its subsidiaries. All Rights Reserved. Dell Technologies, Dell, EMC, Dell EMC and other trademarks are trademarks of Dell Inc. or its subsidiaries. Intel, the Intel logo, the Intel Inside logo and Xeon are trademarks of Intel Corporation in the U.S. and/or other countries. Other trademarks may be trademarks of their respective owners. Published in the USA 06/18 White Paper H17212.

Dell Inc. believes the information in this document is accurate as of its publication date. The information is subject to change without notice.

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more information go to www.intel.com/benchmarks

Intel does not control or audit third-party benchmark data or the web sites referenced in this document. You should visit the referenced web site and confirm whether referenced data are accurate.

Benchmark results were obtained prior to implementation of recent software patches and firmware updates intended to address exploits referred to as "Spectre" and "Meltdown." Implementation of these updates may make these results inapplicable to your device or system.



Contents

- Executive summary 4**
- Stampede2 overview..... 5**
- NAMD benchmark 5**
- LAMMPS benchmark..... 6**
- Gromacs benchmark 7**
- WRF benchmark 8**
- Conclusion..... 10**
- Appendix: Build information by application..... 11**

Executive summary

Introduction

A collaboration between the Texas Advanced Computing Center (TACC) at The University of Texas at Austin, Dell EMC, and Intel studied the performance of the 14 nanometer (nm) Intel Xeon Scalable processor using different high-performance-computing (HPC) application benchmarks. TACC staff performed the tests on its flagship supercomputer, Stampede2.

This white paper compares the performance of the Intel Xeon Scalable Platinum 8160 (Skylake or SKX) and Intel Xeon Phi Processor 7250 (Knights Landing or KNL), and demonstrates the significant advantages of the Skylake processor at a multinode scale in four commonly used applications: NAMD, LAMMPS, Gromacs, and WRF. If you have a VASP license, you can contact your Dell EMC representative for the VASP benchmark results.

Document purpose

This white paper is intended to provide information to help decision makers and end users evaluate the Intel processors based on application benchmarks.

Additional information and support

We encourage you to contact your Dell EMC representative with any questions about your specific business environment and requirements.

Mailing address: Dell EMC, One Dell Way, Round Rock, TX 78682 USA

For information about the Dell EMC HPC Community, see www.dellhpc.org.

We value your feedback

Dell EMC and the authors of this guide welcome your feedback on the solution and the solution documentation. Contact the [Dell EMC Solutions team](#) with your comments.

Authors: Qing Ye, Aighne Kearney, Karen Johnson

Stampede2 overview

Stampede2 is a powerful HPC system that is built on Dell EMC servers. It includes the Intel Xeon Scalable processor and many-core Intel Xeon Phi technologies. Stampede2, which is funded by a National Science Foundation (NSF) grant, is located at TACC at The University of Texas at Austin and supported by Intel and Dell EMC. Stampede2 consists of 5,940 compute nodes, a 28 PB storage subsystem, and 24 additional login and management servers, all interconnected by an Intel Omni-Path Architecture (OPA).

The compute nodes are based on two different processors:

- 1,736 Dell EMC PowerEdge C6420 nodes with dual-socket Intel Skylake processors and 192 GB of DDR4 memory at 2,666 MHz
- 4,204 Dell EMC PowerEdge C6320P nodes with an Intel Knights Landing bootable processor and 96 GB of DDR4 memory at 2,400 MHz

The default Stampede2 software environment is compiled using Intel compiler 17.0.4 and Intel MPI 2017.4.196 running Centos 7.4 with Linux kernel version 3.10.0.

For more information about Stampede2, see [Stampede2](#) on the TACC website.

NAMD benchmark

NAMD is a parallel, object-oriented, molecular dynamics code that is designed for high-performance simulation of large biomolecular systems. NAMD was written using the Charm++ parallel programming model at the University of Illinois at Urbana-Champaign.

For more information about NAMD, see the [NAMD](#) page on the University of Illinois at Urbana-Champaign website.

The analysis in this document is based on the latest stable release of NAMD 2.12, Charm++ 6.7.1, using the Satellite Tobacco Mosaic Virus (STMV)-20 benchmark. STMV is a large-scale, 1-million-atom PME model that is commonly used as a benchmark. The 1-million-atom test case is small enough to fit on one node, yet large enough to scale up to 128 nodes. For optimal performance, NAMD was built with the Intel compiler 16.0.3 instead of Stampede2's default compiler, Intel 17.0.4. Elements that are relevant for the build are the Charm++ version and build options, the fast Fourier transform (FFT) and the tool command language (TCL) versions, and the specific build options for NAMD itself. Contact your Dell EMC representative for more information.

The following figure shows the Skylake performance in nanoseconds (ns) per day, using the STMV benchmark with 4 MPI tasks per node and 11 threads per task for all but the largest 128-node count. The largest node count used 6 MPI tasks per node with 7 threads per task. Knights Landing tests used 4 MPI tasks per node and 32 threads per task for 1 to 4 node counts, and 13 MPI tasks per node with 4 threads per task for the higher node counts. For each node count, the test engineers chose the best-performing task/thread configuration.

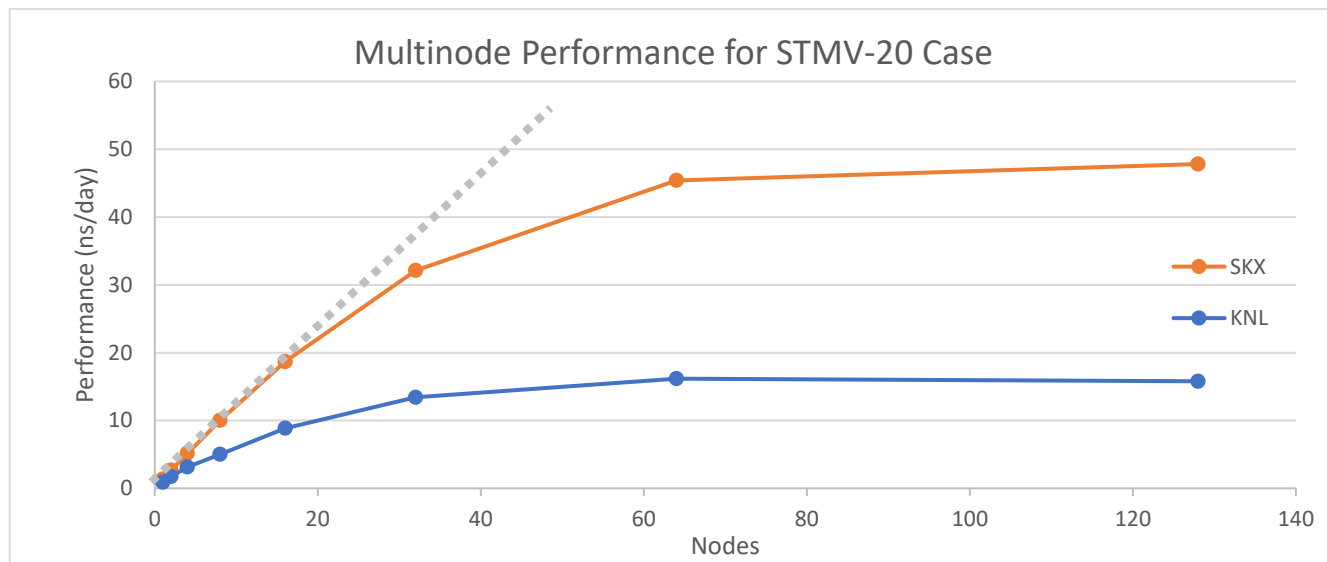


Figure 1. Performance of Skylake and Knights Landing nodes with NAMD

The gray dotted line is a linear trend to show the parallel efficiency of the application at scale. NAMD’s efficiency is reduced after 16 nodes.

Skylake clearly offers a significant advantage at higher node counts, running almost three times faster than Knights Landing. This higher speed is because NAMD does not use the High-Bandwidth Memory (HBM) that is available on the Knights Landing nodes. As the test case scales out and the local memory footprint shrinks, the 16 GB HBM on each Knights Landing node does not help the performance of NAMD. Instead, a greater number of MPI tasks increases the overhead for NAMD on the Knights Landing nodes.

LAMMPS benchmark

Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) is a molecular dynamics code from Sandia National Laboratories. LAMMPS can be used to model atoms or, more generically, as a parallel particle simulator at the atomic, meso, or continuum scale.

For more information about LAMMPS, see the LAMMPS documentation on the [Sandia website](#).

The analysis in this document is based on the latest stable release of LAMMPS, lammps-11Aug17, using the [rhodopsin protein case](#) as the benchmark. This case was built using the default environment. Because this case is too small for multinode benchmarking, the test engineers replicated it six times in X,Y, Z dimensions, resulting in an approximately 7-million-atom test case.

The following figure shows the Skylake and Knights Landing performance of the replicated rhodopsin protein benchmark. The Skylake tests used 48 MPI tasks per node and 1 thread per task. The Knights Landing tests used 64 MPI tasks per node and 1 thread per task.

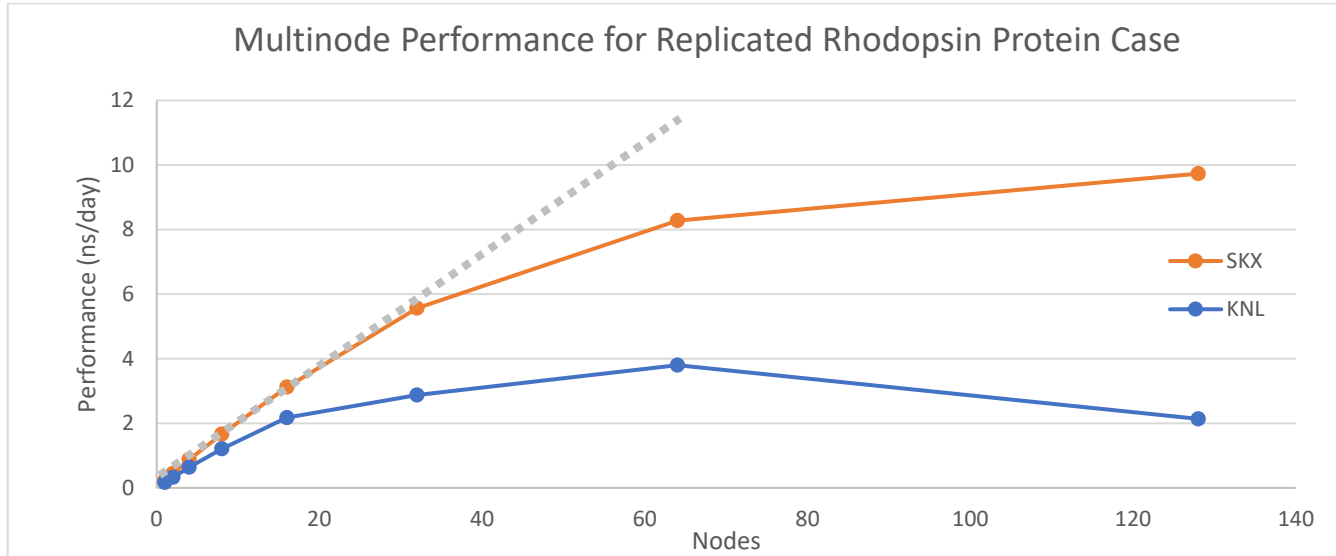


Figure 2. Performance of Skylake and Knights Landing nodes with LAMMPS

Skylake outperforms Knights Landing on a node-to-node basis and scales linearly up to 32 nodes. Analysts suspect that, at 32 nodes, the overhead from the MPI tasks on the Knights Landing begins to impact the performance of LAMMPS. Further improvement of the results might be possible by reducing the number of MPI tasks performing the 3D-FFT, but this requires more time to set up. Contact your Dell EMC representative for more information.

Gromacs benchmark

GROningen Machine for Chemical Simulations (Gromacs) is a molecular dynamics package that is designed mainly for simulations of proteins, lipids, and nucleic acids. Gromacs was originally developed in the University of Groningen Biophysical Chemistry Department and is one of the fastest and most popular software packages that can run on CPU and GPU.

For more information about Gromacs, see the documentation on the [Gromacs website](#).

In the study described in this white paper, pure water solutions are simulated using Gromacs version 2016.3 and built using the default environment. The simulated systems consist of a total of 3 million atoms. The test engineers obtained the initial coordinates and

WRF benchmark

simulation parameters from the Gromacs website. They performed all the simulations in the isothermal isobaric (NpT) ensemble at 300 K and 1 atm.

The following figure shows the Skylake and Knights Landing performance in the pure water benchmark test. The Skylake tests used 48 MPI tasks per node and 1 thread per task. The Knights Landing tests used 64 MPI tasks per node and 1 thread per task.

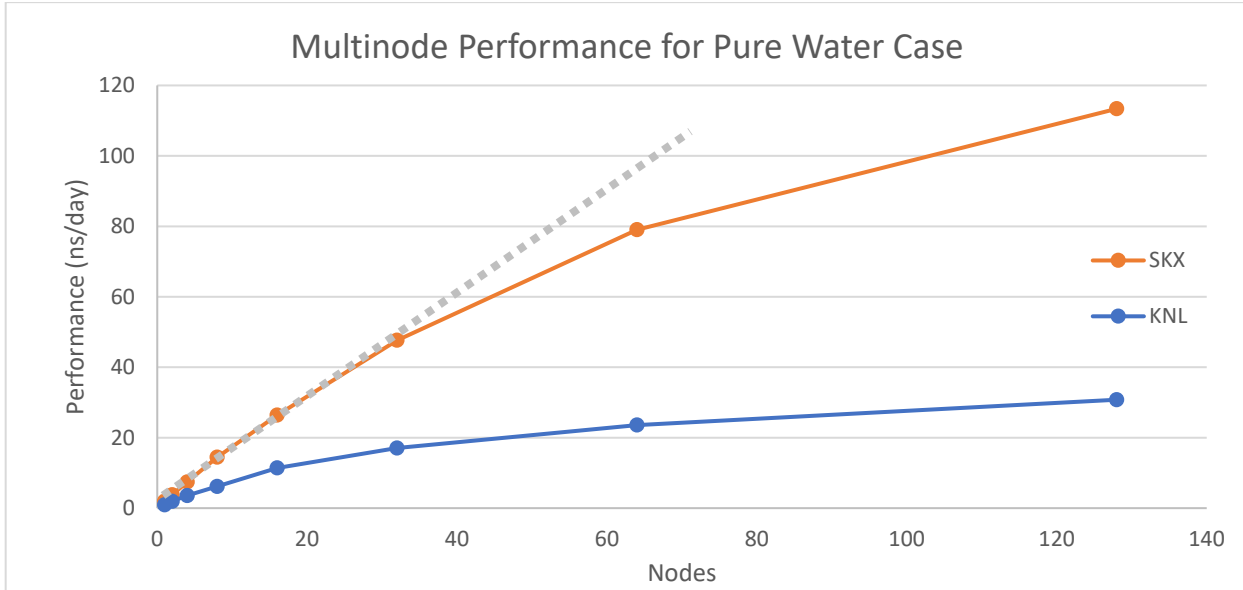


Figure 3. Performance of Skylake and Knights Landing nodes with Gromacs

As with LAMMPS, in the Gromacs benchmark, Skylake outperforms Knights Landing on a node-to-node basis and scales linearly up to 32 nodes. We expect that Knights Landing would perform better with 4 simultaneous multithreads (SMTs) than with 1 thread per task.

WRF benchmark

Weather Research and Forecasting (WRF) is a mesoscale numerical weather-prediction system that is designed to serve both atmospheric research and operational forecasting needs. The model serves a wide range of meteorological applications across scales ranging from a few meters to thousands of kilometers. WRFv3.6.1 was built with netcdf 4.3.2.1 and parallel-netcdf 1.8.1 for I/O. For more information, see the [WRF website](#).

The build configuration varies slightly between Skylake and Knights Landing. For more information about the build, contact your Dell EMC representative.

The following figure shows the Skylake and Knights Landing performance of the 2.5 km continental United States (CONUS) benchmark. The Skylake tests used 4 MPI tasks per node and 12 threads per task. The Knights Landing tests used 4 MPI tasks per node and 17 threads per task. The exception was the 32-node count, which used 2 MPI tasks per node with 34 threads per task.

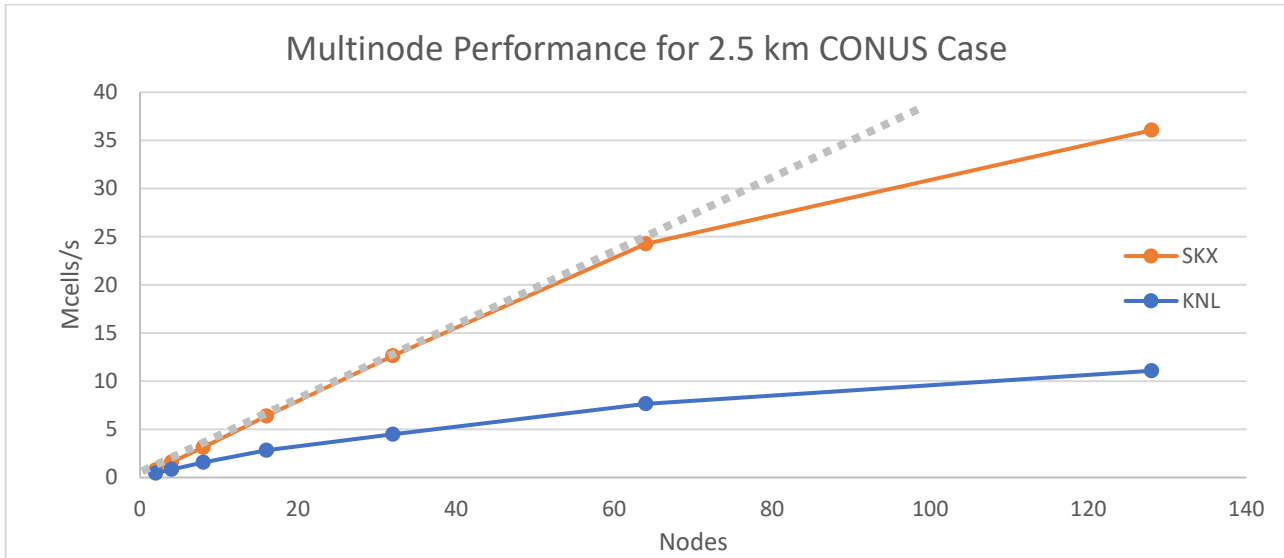


Figure 4. Performance of Skylake and Knights Landing nodes with WRF

As the code scales out, the Skylake nodes outperform the Knights Landing nodes by more than a factor of three. The decrease in performance at high Knights Landing node counts is most likely due to the impact of MPI on the node, even with only 4 MPI tasks per node. Skylake demonstrates good scalability and scales linearly up to 64 nodes. On the Knights Landing nodes, the parallel efficiency steadily drops with the node count. This might be due to higher latency on the Knights Landing nodes.

This study used the 2.5 km CONUS benchmark data set with a time interval from 2005-06-04_00:00:00 to 2005-06-04_01:00:00. This benchmark runs for 240 time steps; so the measured time is the total computing time for processing the domain 1 time steps except for the first step. This timing excludes the startup time, in which the restart file is read in and does not include any I/O for output.

We converted the domain 1 times into a performance measurement of 1 million cells per second. Because WRF has a fixed-size grid, this measure represents the number of horizontal grid cells simulated per second using the following equation:

$$\text{Mcells/s} = (\# \text{ East-West grid cells}) \times (\# \text{ North-South grid cells}) \times (\# \text{ of domain 1 time steps}) / (\text{domain 1 time})$$

Because the total grid size is fixed for every run, this is analogous to $1/(\text{domain 1 time})$ for all node counts. The engineers investigated only the hybrid MPI/OpenMP WRF build. Because differences between cache mode and flat mode were observed to be minimal for single node runs, this study presents only Knights Landing runs in cache mode.

Conclusion

The goal of the study described in this paper is to show the capabilities of the latest Intel Xeon Scalable processor Skylake at multinode scale by comparing its performance with the previous generation of the Intel Xeon Phi processor Knights Landing. Based on the Stampede2 application benchmark measurements, the Skylake processor clearly offers a significant performance advantage and good scalability.

To realize the full potential of HPC systems, all aspects of system design, including hardware, software, application characteristics, data center environment, power, and cooling must be carefully designed and tuned. Dell EMC provides extensive evaluation and best-practice recommendations, including build details such as compilation flags and runtime parameters, for different architectures. Today, HPC is expanding from traditional modeling and simulation to include new, high-growth applications in high-performance data analytics (HPDA) and artificial intelligence (AI). Wherever your focus resides, be it academic research, life sciences, or digital manufacturing, we can help you extract the maximum value from your cluster with validated products, knowledge of tuning at scale, and customized training services.

Appendix: Build information by application

This section provides build information for each of the applications used in the benchmark testing.

NAMD benchmark

NAMD 2.12 with Charm++6.7.1 was built with Intel compiler 16.0.3 and Intel MPI 2017.4.196 using the MKL FFT package and tcl8.5.9. Our experiences indicate little to no performance difference when using the prebuilt FFTW package that the NAMD developers provided.

The build for Skylake was the same as for Knights Landing but used `-CORE-AVX512` in place of `-MIC-AVX512`. This required a change of the optimization flag in the `arch/Linux-KNL-icc.arch` file to use `xCORE-AVX512` in place of `xMIC-AVX512`. This version of NAMD does not yet have a Skylake option.

LAMMPS benchmark

LAMMPS-11Aug17 was built with Intel compiler 17.0.4 and Intel MPI 2017.4.196.

LAMMPS was built with specific support for Knights Landing using `-MIC-AVX512` and specific support for Skylake using `-CORE-AVX512`.

Gromacs benchmark

Gromacs 2016.3 was built with Intel compiler 17.0.4 and Intel MPI 2017.4.196.

The test engineers obtained the initial coordinates and simulation parameters from ftp://ftp.gromacs.org/pub/benchmarks/water_GMX50_bare.tar.gz. To access the tar file (132 MB), go to <ftp://ftp.gromacs.org> and open the `gromacs` folder.

Stampede2 has a single binary that supports Knights Landing using `MIC-AVX512` and Skylake using `CORE-AVX512`.

WRF benchmark

WRFv3.6.1 was built with Intel compiler 17.0.4 and Intel MPI 2017.4.196, using netcdf 4.3.2.1 and parallel-netcdf 1.8.1 for I/O. The configuration varies slightly between the Knights Landing and Skylake builds.

For the Knights Landing build, the Knights Corner option, #17, was used as the basis for `configure.wrf`. In the `configure.wrf` file, the testers removed the `-mmic` flag from all the compile commands and added the `-MIC-AVX512` option to `OPTKNC`.

For the Skylake build, the Sandy Bridge option, #21, was used as the basis for `configure.wrf`. In the `configure.wrf` file, the `OPTAVX` entry becomes: `OPTAVX= -xCORE-AVX512`.

WRF was allowed to partition the domain using its default configuration, with no changes made to the default distribution or tiling.