New 13G servers – what's new and how much better are they for HPC?

Garima Kochhar, September 2014

It's been an exciting week –Intel Haswell processors for two-socket servers, DDR4 memory and new Dell servers were just released. We've had a busy few months leading up to this announcement – our team had access to early server units for the HPC lab and we spent time kicking the tires, running benchmarks, and measuring performance. This blog describes our study and initial results and is part one of a three part series. The next blog will discuss the performance implications of some BIOS tuning options available on the new servers, and a third blog will compare performance and energy efficiency across different Haswell processor models.

Focusing on HPC applications, we ran two benchmarks and four applications on our server. Our interest was in seeing how the server performed and specifically how it compared to the previous generations.

The server in question is part of Dell's PowerEdge 13th generation (13G) server line-up. These servers support DDR4 memory at up to 2133 MT/s and Intel's latest Xeon® E5-2600 v3 Product Family processors (based on the architecture code-named Haswell). Haswell (HSW) is a net new micro-architecture when compared to the previous generation - Sandy Bridge/Ivy Bridge. HSW processors use a 22nm process technology, so there's no process-shrink this time around. Note the "v3" in the Intel product name – that is what distinguishes a processor as one based on Haswell micro-architecture. You'll recall that "E5-2600 v2" processors are based on the Ivy Bridge micro-architecture and plain E5-2600 series with no explicit version are Sandy Bridge based processors. Haswell based processors require a new server/new motherboard and DDR4 memory. The platform we used is a standard dual-socket rack server with two Haswell-EP based processors. Each socket has four memory channels and can support up to 3 DIMMs per channel (DPC). For our study we used 1 DPC for a total of eight DDR4 DIMMs in the server.

From an HPC point of view, one of the most interesting aspects is the Intel® AVX2 technology that allows the processor to execute ***16 FLOP per cycle***. The processor supports 256 bit registers, allows three-operand non-destructive operations (i.e. A = B+C vs. A = A+B), and a Fuse-Multiply-Add (FMA) instruction (A = A*B+C). The processor has two FMA units each of which can execute 4 double precision calculations per cycle. With two floating point operations per FMA instructions, HSW can execute 16 FLOP/cycle. This value is ***double*** of what was possible with Sandy Bridge/Ivy Bridge (SB/IVB)! There are many more instructions introduced with HSW and Intel® AVX2 and these are described in detail in this Intel programming reference or on other blogs.

Double the FLOP/cycle - does this mean that HSW will have 2x the theoretical performance of an equivalent IVB processor? Close but not quite - read on. In past generations, we've looked at the ***rated base frequency*** of a processor and the available Turbo bins/max Turbo frequency. For example, the Intel® Xeon® E5-2680 v2 has a base frequency of 2.8 GHz and a maximum of 300 MHz of turbo available when all cores are active. HSW processors will consume more power when running the new Intel® AVX2 instructions than when running non-AVX instructions. And so, starting with Haswell product family there will be two rated base frequencies provided. The first is the traditional base frequency which is the frequency one could expect to run non-AVX workloads. The second frequency is the base frequency for workloads that are running AVX code, the "***AVX base frequency***". For example, the HSW Xeon® E5-2697

Available at http://dell.to/1w8xlKI

v3 has a base frequency of 2.6 GHz and an AVX base of 2.2 GHz. Compare that with the Xeon® E5-2680 v2 IVB processor running at 2.6 GHz. For the 2.6 GHz IVB processor, we would calculate HPL theoretical maximum performance as (2.6 GHz * 8 FLOP/cycle * total number of cores). But for a HSW processor with the same rated base frequency of 2.6 GHz and an AVX base of 2.2 GHz, we now calculate HPL theoretical maximum using the AVX base as (***2.2 GHz*** * 16 FLOP/cycle * total number of cores) since HPL is an AVX-enabled workload. In terms of FLOPs a HSW processor will perform much better than an IVB, close to 2x but not exactly 2x due to the lower AVX base frequency. Of course, enabling Turbo mode can allow higher core frequencies when there is power/thermal headroom. But due to the extra power consumption of AVX instructions, non-AVX codes/portions of the code may run at higher Turbo bins than the AVX portions.

The goal of communicating a separate "AVX base frequency" is two-fold – AVX codes will run "hotter" (consume more power) than non-AVX, so lower frequencies on those codes are expected and are by design. Secondly, by providing this secondary "AVX base frequency" Intel is providing a baseline expected frequency for highly optimized AVX workloads.

There are many other significant new aspects in this release – DDR4 memory technology, improvements in Dell server design and energy efficiency, and new features in systems management to name a few. All of these aspects are also factors in improvements in server performance in this generation. These factors are not discussed in this blog.

Now, getting down to the fun part - the results. Table 1 below details the applications we used.

Table 1 - Applications and benchmarks

Application	Domain	Version	Benchmark
Stream	Synthetic benchmark to measure memory bandwidth	v5.9	Triad
HPL	Synthetic processor bound benchmark to measure floating point computation capability	From Intel MKL	Problem size 90% of total memory
Ansys Fluent	Computational fluid dynamics	v15.0	Six benchmark data sets as noted in the graph
LS-DYNA	Finite element analysis	v7_0_0_79069	car2car with endtime=0.02
WRF	Weather Research and Forecasting	v3.5.1	Conus 12km and Conus 2.5km
MILC	Quantum chromo dynamics	v7.7.3	Input data file from Intel

For reference, Table 2 describes the test configuration on the new 13G server. Data for some of the tests on the previous generation systems was gathered with the most current versions available at that time. The performance improvements noted here are mainly due to architectural improvements generation-over-generation, the software versions are not a significant factor.

Table 2 - Server configuration

Components	Details
Server	PowerEdge R730xd – engineering sample unit/prototype
Processor	2 x Intel® Xeon® E5-2697 v3 – 2.6 GHz, 14c, 145W
Memory	128GB - 8 x 16GB 2133 MHz DDR4 RDIMMs
Hard drive	1 x 300GB SAS 6Gbps 10K rpm
RAID controller	PERC H330 mini
Operating System	Red Hat Enterprise Linux 6.5 x86_64
Kernel	2.6.32-431.el6.x86_64
BIOS settings	Performance mode, Turbo enabled, Cstates disabled, Cluster on die, Node interleaving disabled, Logical processor disabled
MPI	Intel® MPI 4.1.3.049
Math Library	Intel® MKL 11.1.3.174
Compilers	Intel® 2013_sp1.3.174 - v14.0.3.174 Build 20140422

All the results shown here are based on single-server performance. The following metrics were used to compare performance:

- Stream - Triad score as reported by the stream benchmark.
- HPL - GFLOP/second.
- Fluent - Solver rating as reported by Fluent.
- LS DYNA – Elapsed Time as reported by the application.
- WRF – Average time step computed over the last 719 intervals for Conus 2.5km
- MILC – Time as reported by the application.

Figure 1 shows the measured memory bandwidth as reported by Stream Triad on the 13G server when compared to previous generations. In the graphs below, "11G – WSM" denotes the Dell 11th generation (11G) servers that support Intel Westmere (WSM) i.e. Intel® Xeon® X5600 series processors. "12G – SB" and "12G – IVB" are the Dell 12th generation servers (12G) that support Intel Sandy Bridge and Ivy Bridge processors. Full system ***memory bandwidth is ~ 116 GB/s*** on the 2133 MT/s HSW system. This is an 18% improvement over the 12G-IVB system that could support memory speeds of up to 1866 MT/s. Even with the increased number of cores on HSW, memory bandwidth per core has remained mostly constant from the previous generation and is ~4.2 GB/s per core.

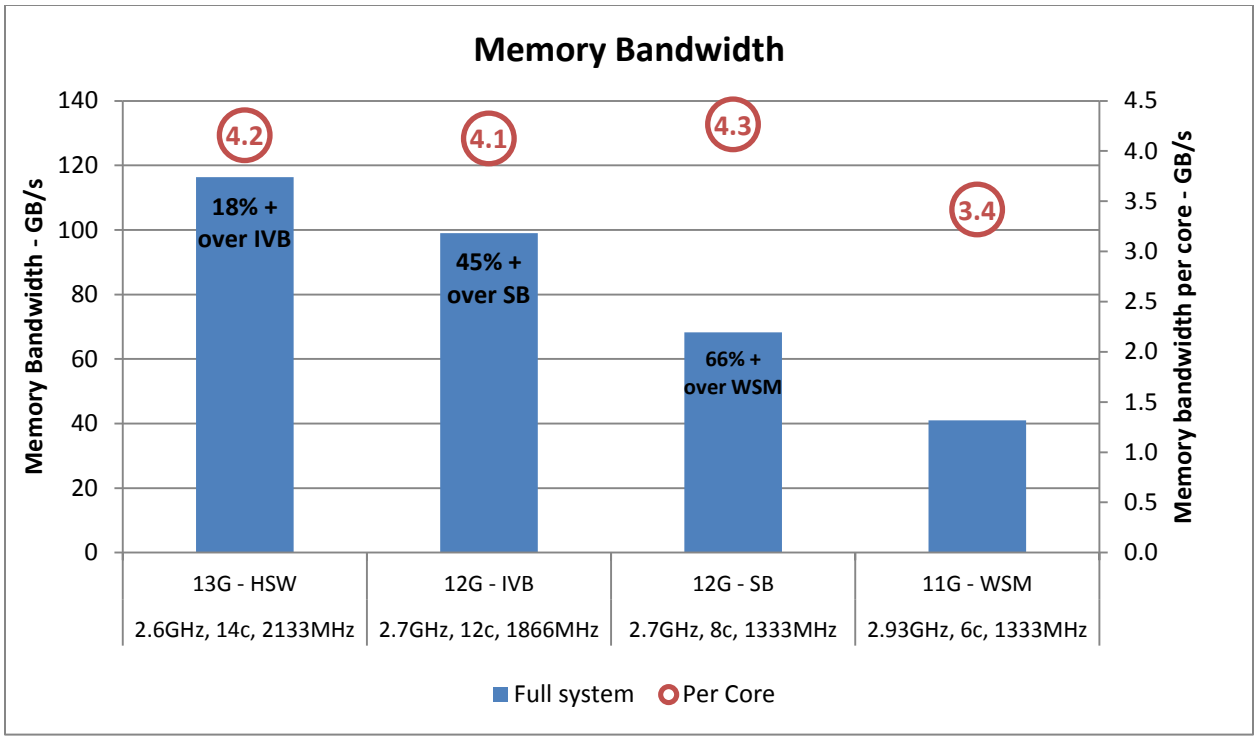

Figure 1 - Memory bandwidth

Figure 2 shows the HPL performance generation-over-generation. With the HSW system we measured close to 1 TFLOPS on a single server! This improvement in HPL performance is due to the increase in floating point capability. Note that the HPL efficiency of 93% for the HSW system is computed using the AVX base frequency and not the rated frequency as discussed above. For the processor used, the AVX base is 2.2 GHz.

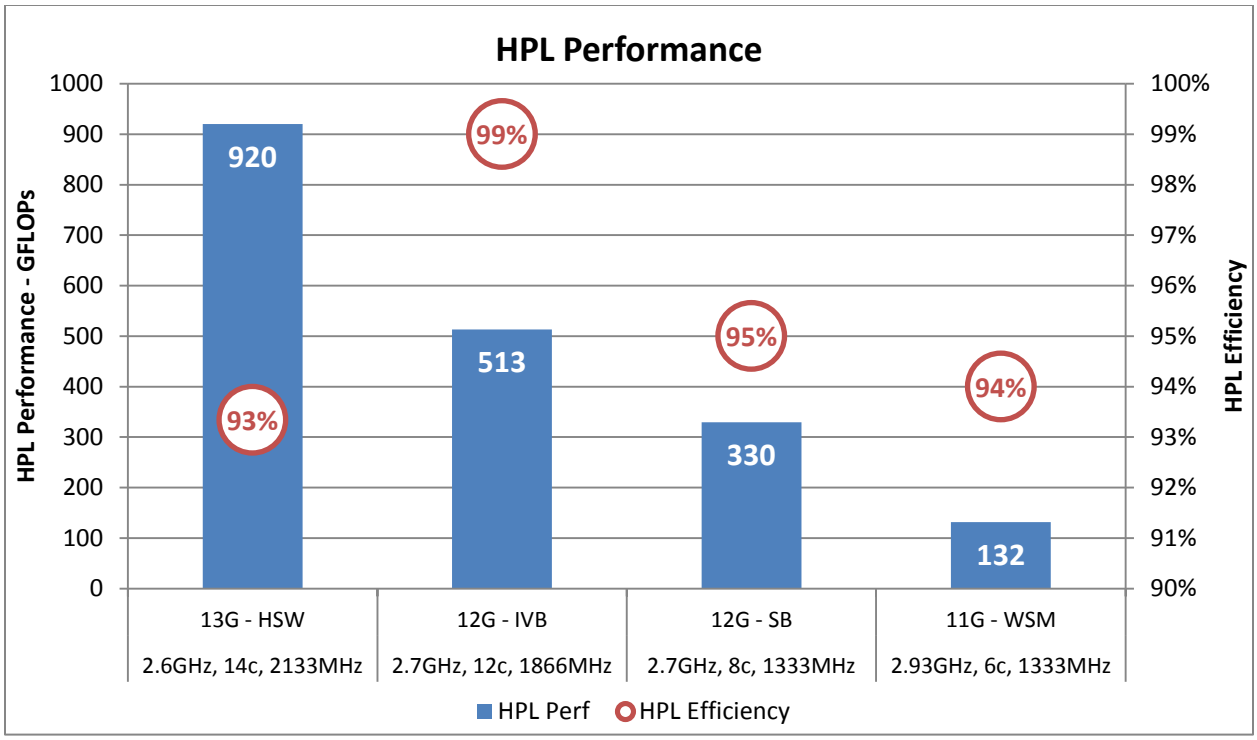

Figure 2 - HPL performance

Figure 3 shows Ansys Fluent performance on 13G HSW when compared to 12G IVB. The Solver Rating as reported by Fluent is the metric used for performance. The 12G system used was a PowerEdge C6220 II with dual Intel® Xeon® E5-2680 v2 @ 2.8 GHz (10 cores each), 128 GB (8 x 16GB 1866MHz) memory. The 13G system was configured as described in Table 2. Note that the all-core Turbo on both the 12G IVB and the 13G HSW system is 3.1 GHz so this is a good comparison across the two generations. There is a significant performance improvement with HSW – 33% to 48% depending on the dataset used. Since Ansys Fluent has a per-core license, we wanted to weight this performance with the increase in number of cores from 12G IVB to 13G HSW (20 cores to 28 cores, a 40% increase in cores.). The per core performance improvement depends on the data set used, with truck_poly_14m demonstrating a 6% performance improvement with 13G even when accounting for the increased number of cores.

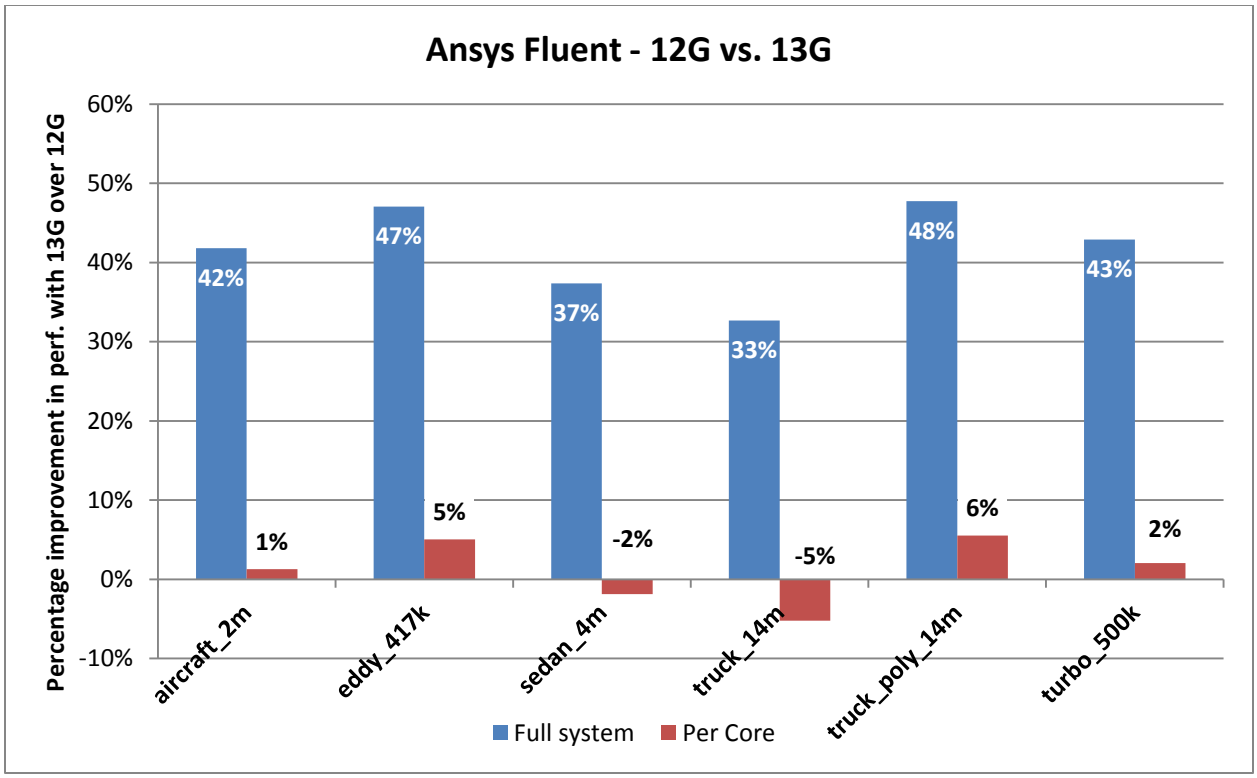

Figure 3 - Ansys Fluent performance

Figure 4 shows WRF performance generation-over-generation for the Conus 12km data set. The average time step computed over the last 149 intervals for Conus 12km is the metric used for performance. There is a 40% improvement from 12G-IVB and a 3.2x improvement over the Westmere platform. The improvement with HSW is likely due to the better memory throughput and micro-architecture enhancements.

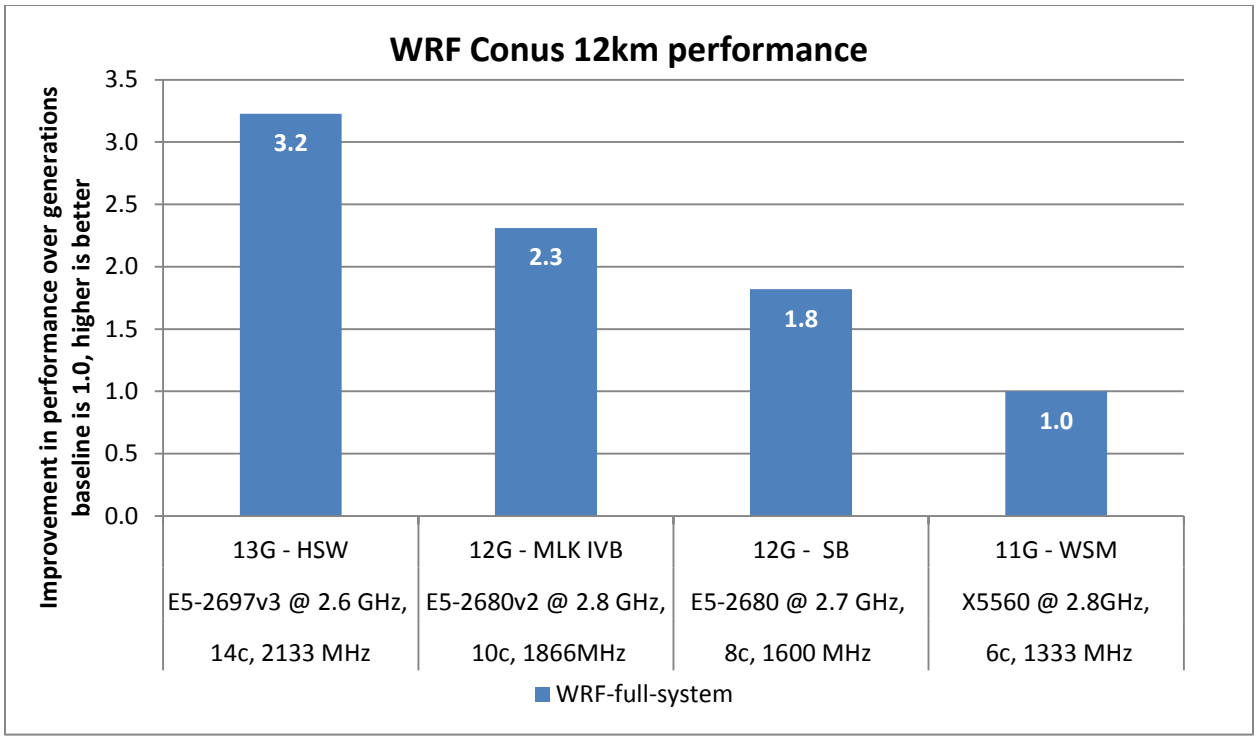

Figure 4 - WRF Conus 12km performance

MILC performance is shown in Figure 5. Total time as reported by the MILC application is the metric used for performance. We measured a 10% improvement in performance with HSW when compared to a 12G IVB system and a 30% improvement when compared to a 12G SB platform.

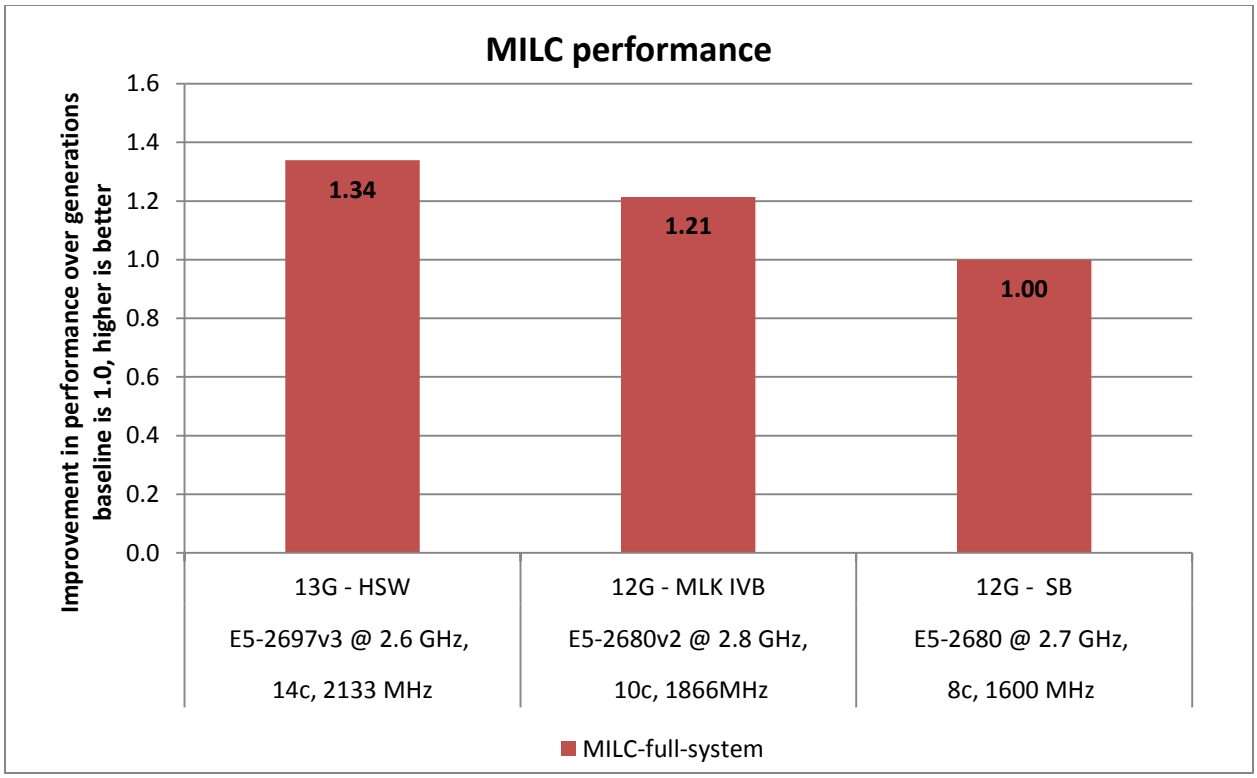

Figure 5 - MILC performance

LS DYNA with car2car and WRF with Conus 2.5km performance comparisons are shown in Figure 6. Elapsed time as reported by LS DYNA is used to compare performance, and average time step computed over the last 719 intervals for Conus 2.5km is the WRF metric used for performance. Both applications perform 38-37% better with 13G-HSW when compared to 12G-IVB.

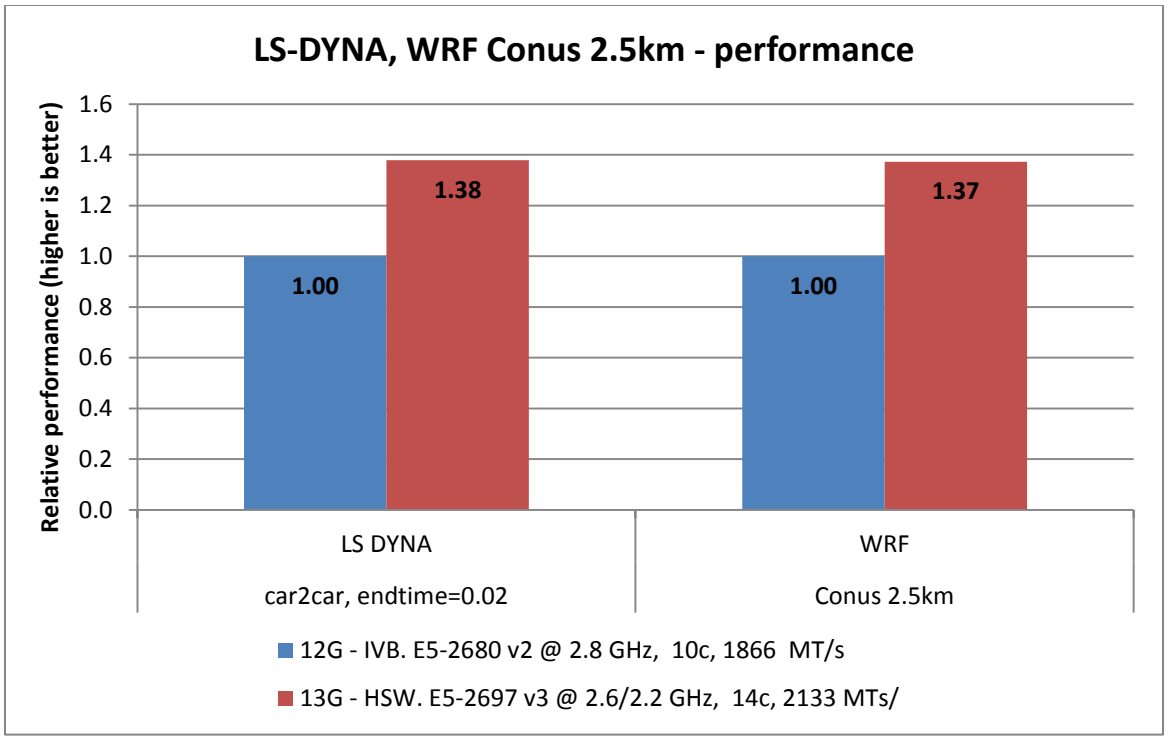

Figure 6 - LS DYNA, WRF 2.5km performance

In conclusion, the new 13G servers show performance improvements across all the applications studied here; this study provides some early comparisons and quantifies these performance improvements.

Look out for the next two blogs in this series. Blog 2 will discuss the performance and energy efficiency implications of some BIOS tuning options available on the 13G servers, and the third blog will compare different Haswell processor models.