Xeon E5-2600 v4 series overview

details of the new Intel Xeon "Broadwell" systems

About Us

- Advanced Clustering provides customized turn-key High-Performance Computing solutions
- Our solutions include: compute nodes, networking, storage, infrastructure, software and support services

advanced clustering
technologies, inc.

Company Background

- Company founded in 2001
  - Same ownership, and management
  - Focused on providing HPC solutions since inception
  - Completely self-funded and profitable for years
  - Small business located in Kansas City, KS
- Intel Platinum Partner (highest level)

advanced clustering technologies, inc.

Intel Tick / Tock Model

Platform: Tylersburg		Platform: Romley		Platform: Grantley	
Nehalem X5500 Q1 2009	Westmere X5600 Q1 2010	SandyBridge E5-2600 Q1 2012	Ivy Bridge E5-2600v2 Q3 2013	Haswell E5-2600v3 Q3 2014	Broadwell E5-2600v4 Q1 2016
45 NM	32 NM	32 NM	22 NM	22 NM	14 NM
TOCK New Micro-Architecture	TICK Die-shrink	TOCK New Micro-Architecture	TICK Die-shrink	TOCK New Micro-Architecture	TICK Die-shrink

advanced clustering technologies, inc.

Haswell vs. Broadwell

	Xeon E5-2600v3 - Haswell	Xeon E5-2600v4 - Broadwell
Number of CPUs per system	2	
Number of Cores per CPU	Up to 18	Up to 22
AVX support	AVX 2.0 (16 DP FLOPS/cycle)	
QPI speed	Max 9.6GT/s	
Integrated memory controller	Yes	
Memory support	DDR4 2133MHz ECC Registered	DDR4 2400MHz ECC Registered
Memory channels CPU/System	4/8	
Maximum DIMMs	16	
PCI-e Gen 3.0 lanes	80 (per 2 socket system)	
Power draw	85W - 135W	

E5-2600v4 New Features

CPU micro-architecture

- Improved branch prediction
- Lower latency on AVX multiply instructions
- Performance improvements focused on encryption
- Ability to monitor cache utilization and memory bandwidth

Power Management

- Hardware controlled power management (HWPM)
- Per-core AVX frequency scaling

More cores / faster Memory

- Up to 22 cores available
- Faster DDR4 memory

advanced clustering technologies, inc.

Processor SKUs

Standard

- E5-2620v4 — 8c 2.1 GHz (85W)
- E5-2630v4 — 10c 2.2 GHz (85W)
- E5-2640v4 — 10c 2.4 GHz (90W)

- 20-25MB (2.5MB/core) L3 Cache
- 8.0 GT/s QPI
- DDR4 2133MHz

Advanced

- E5-2650v4 — 12c 2.2 GHz (105W)
- E5-2660v4 — 14c 2.0 GHz (105W)
- E5-2680v4 — 14c 2.4 GHz (120W)
- E5-2690v4 — 14c 2.6 GHz (135W)

- 30-35MB (2.5MB/core) L3 Cache
- 9.6 GT/s QPI
- DDR4 2433MHz

Other specialty SKUs from min of 4 cores (3.5GHz) to max of 22 cores (2.2GHz) available

Memory Population Guidelines

Type	DIMM Capacity	1 DIMM per channel	2 DIMM per channel
RDIMM SR or DR	4GB	2400 MHz	2133 MHz
RDIMM SR or DR	8GB	2400 MHz	2133 MHz
RDIMM SR or DR	16GB	2400 MHz	2133 MHz
LRDIMM QRx4	32GB	2400 MHz	2400 MHz
LRDIMM QRx4	64GB	2400 MHz	2400 MHz

- Only E5-2650v4 and above support 2400 MHz memory. Slower CPUs max at 2133MHz.
- Quad channel memory controller per socket, must use all channels for best performance:
  - 1 DIMM per channel = 4 DIMMs per CPU, or 8 DIMMs per system
  - 2 DIMM per channel = 8 DIMMs per CPU, or 16 DIMMs per system

TDP vs AVX frequency

- Thermal Design Power (TDP) = the total power consumption of the CPU (i.e. 85, 120, 135W)
- In the past, the CPU could always run at the marked frequency and consume at or below the TDP
- E5-2670v2 "Ivy Bridge" example: Processor is 2.5GHz and 115 watts. In all cases the cores could run at 2.5GHz and would never exceed 115 watts of power
- ***This changes started with Haswell and the introduction of AVX 2.0 instructions***
- AVX instructions now use more power to execute, therefore when using them the processor will run slower than the marked frequency
  - There is a 300-500MHz frequency reduction when using AVX instructions

What is AVX?

AVX 1.0

- Launched with Sandy Bridge processors (E5-2600)
- 256-bit wide vector instruction set extension to Intel SSE and is designed for applications that are Floating Point (FP) intensive.

AVX 2.0

- Launched with Haswell processors (E5-2600v3) and included in Broadwell (E5-2600v4)
- Extends 256-bit vector support for integer vector operations
- 256-bit Fused Multiply-Add (FMA)
  - Up to ~2x the FLOPs/core vs. prior generation
- Adds support for Gather, Permutes/Blend, Vector Shifts
- 2x L1/L2 cache bandwidth to support higher FLOPS

AVX frequency reduction

new in AVX 2.0

1. Certain high-power instructions get run on the CPU (mostly Integer and FP AVX instructions)
2. These instructions cause the power control unit (PCU) to increase the voltage
3. With the increased voltage the CPU could now be drawing more power than TDP, so frequency is reduced to stay within TDP limits
4. 1ms after the last high power instruction has executed, the voltage will be reduced back to non-AVX levels

Highly optimized AVX workloads (like Linpack) will be more likely to operate at reduced frequency

Broadwell Theoretical Peak Floating Point

Clock Frequency [TDP|AVX] * Cores * 16 DP FLOPs per cycle

- Example *TDP* Frequency: Broadwell E5-2690v4:
  - 2.6 (GHz) * 14 (Cores) * 16 (FLOPs) = 582.4 GFLOPs per socket
  - 1,164.8 GFLOPs per dual socket system
- Example *AVX* Frequency: Broadwell E5-2690v4:
  - 2.1 (GHz) * 14 (Cores) * 16 (FLOPs) = 479.36 GFLOPs per socket
  - 958.72 GFLOPs per dual socket system

What does this mean for Linpack?

- Linpack requires AVX for good performance
- Broadwell CPU may run up to 500MHz lower than the marked frequency on the processor
- Efficiency of measured performance vs. AVX peak varies based on processor SKU
- The more thermal headroom the better the efficiency, and the closer measured performance will be to TDP frequency
- Lower core count at higher wattage seems to show the best efficiencies

What about other applications?

It depends on the application. CPU frequency varies based on:

- Number of active compute cores
- Temperature of CPU
- Power consumption
- Use of high-power instructions

Heavy AVX usage = lower frequency to stay within TDP

Intel Turbo Boost

When there is thermal headroom, this runs the processor at faster than marked frequency. Built-in CPU overclocking.

What determines the actual Turbo Boost frequency?

- Processor SKU
- Number of active cores
- Power consumption
- Processor temperature
- Type of workload

Intel Turbo Boost 2.0

When there is thermal headroom, this runs the processor at faster than marked frequency. Built-in CPU overclocking.

What determines the actual Turbo Boost frequency?

- Processor SKU
- Number of active cores
- Power consumption
- Processor temperature
- Type of workload
- Variability of Part

Now "opportunistic" and not guaranteed. Turbo performance can vary from part to part even in the same SKU.

Intel Turbo Boost 2.0

Variability of Part

Identically marked parts running the same workload all are valid possibilities:

No turbo

+100MHz turbo

+300MHz turbo

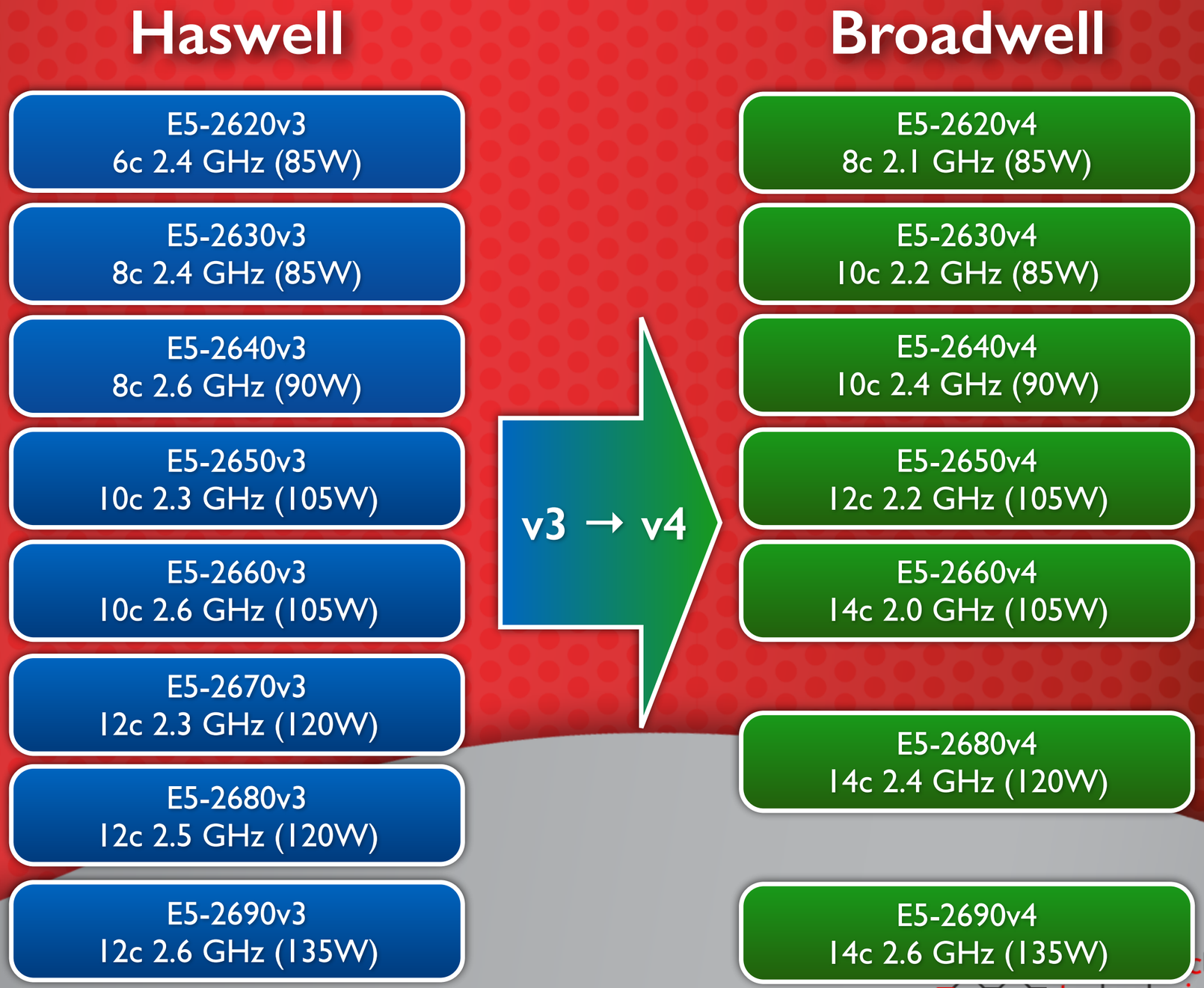
never to exceed max turbo table

Want to avoid variability?

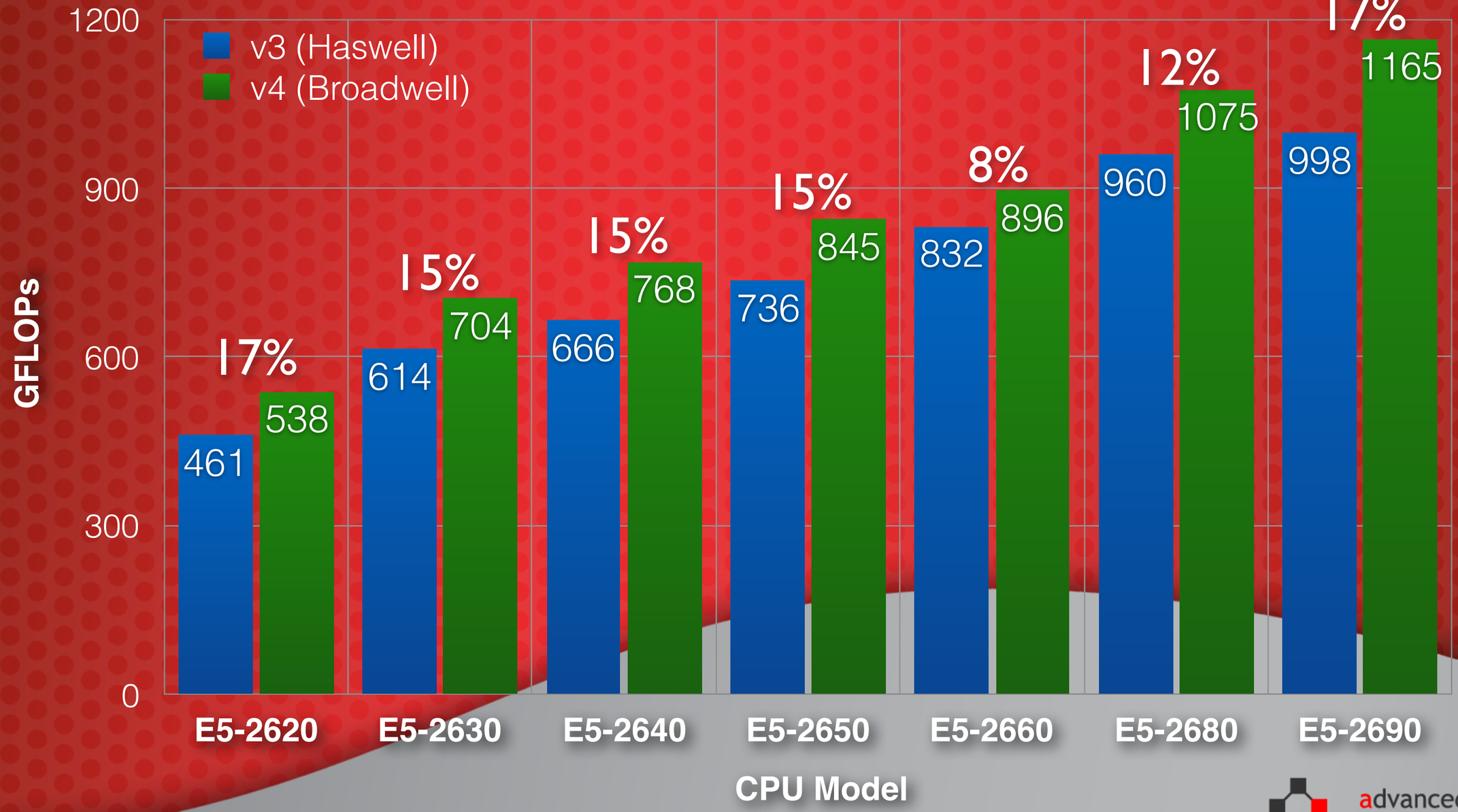
Limit AVX workloads or turn off Turbo mode in the BIOS

Detailed Comparison of Haswell vs. Broadwell

Haswell to Broadwell Processor Transition



Rpeak Theoretical GLOPs Haswell vs Broadwell, 2 socket system



advanced clustering technologies, inc.