OCP SUMMIT

March 20-21
2018
San Jose, CA

OPEN
Compute Project

Intel Architecture 2S Server Tioga Pass – Performance and Power Optimization

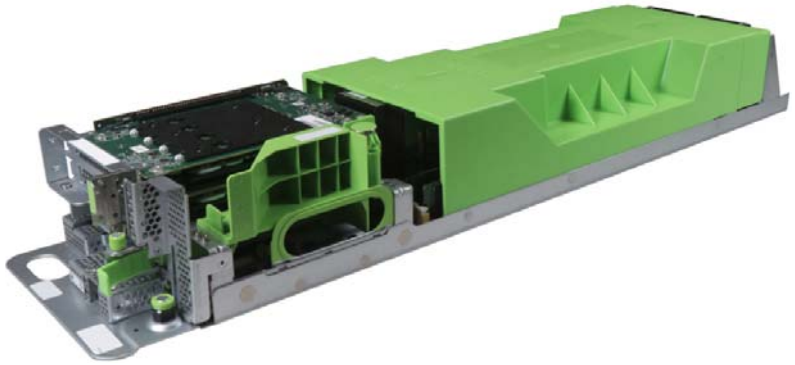
Terry Trausch/Platform Architect/Intel Inc.
Whitney Zhao/HW Engineer/Facebook Inc.

Agenda

- Tioga Pass Feature Overview
- Intel Xeon Scalable Processor Family (Purley/Skylake) Performance and Power
- Optimize Power & Performance for FB workload
- Call to Action

Overview

- Half width 2S system in ORv2
- Intel Xeon Scalable Processor (Purley/Skylake) CPUs up to 165W TDP
- 12 memory channels @ 2666Mhz
- Single sided/Double sided memory SKUs
- 2x PCIe Gen 3 x16 slots via riser
- OCP NIC 2.0 support: 25G to 100G
- Support Intel KR Mezz Card
- Support for Intel At Scale Debug
- Support for OCP LCD debug card



Xeon Scalable Processor Performance & Power

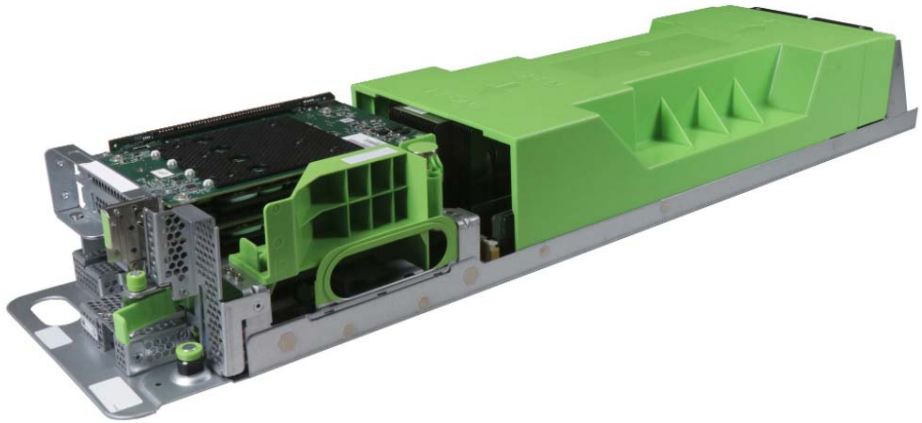
- **Performance and Power optimizations possible Xeon Scalable Processor (Purley/Skylake)**
- **Within the 165W TDP maximum possible CPU SKU:**
- **SPECint_rate: SKX 39% improvement over BDW with just 14% higher TDP**
  SpecPower: SKX 67% performance improvement over BDW with just 14% higher TDP
  SKX CPU SpecPower performance per Watt: 46.5% improvement over BDW

NOTE

1. Performance estimates 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.

2. Results based on Intel measurements and are provided for informational purposes only. Any difference in system hardware or software design or configuration may affect actual performance.
   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 http://www.intel.com/performance/datacenter.

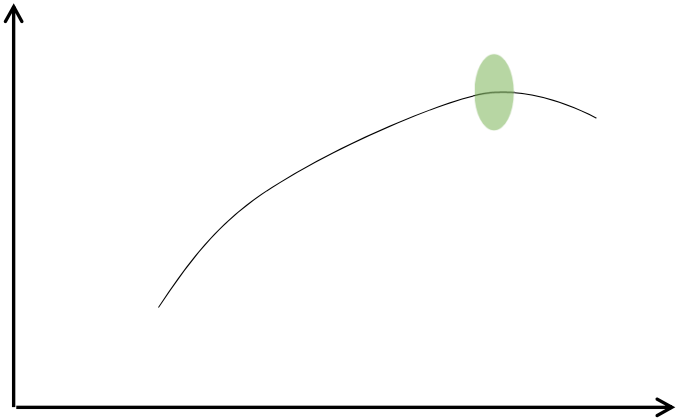
3. Configurations: Shown in backup slides



Optimize Power & Performance for FB workload

In search of best efficiency for scaling out

F(X)



F(X)= System performance /w X

X={x_1,x_2,…,x_n}

	Parameters to explore across different SKUs
X_1	TDP
X_2	Core Count
X_3	Frequency
X_4	Tcase
X_5	Others

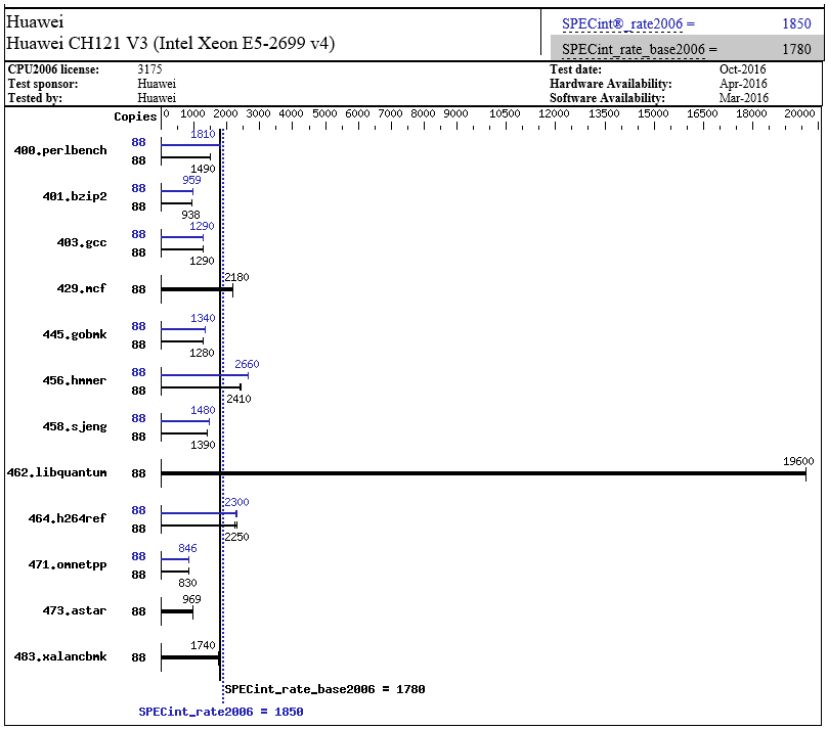
Summary

- Check out the Tioga Pass hardware
  - Facebook
  - Intel
  - Inspur
  - Quanta
  - Wiwynn
- Learn more about the Tioga Pass specification
- Learn more about Intel Xeon Scalable Processor
- Come visit the Intel booth!

QUESTIONS

Intel E5-2699 v4 SPEC Detail

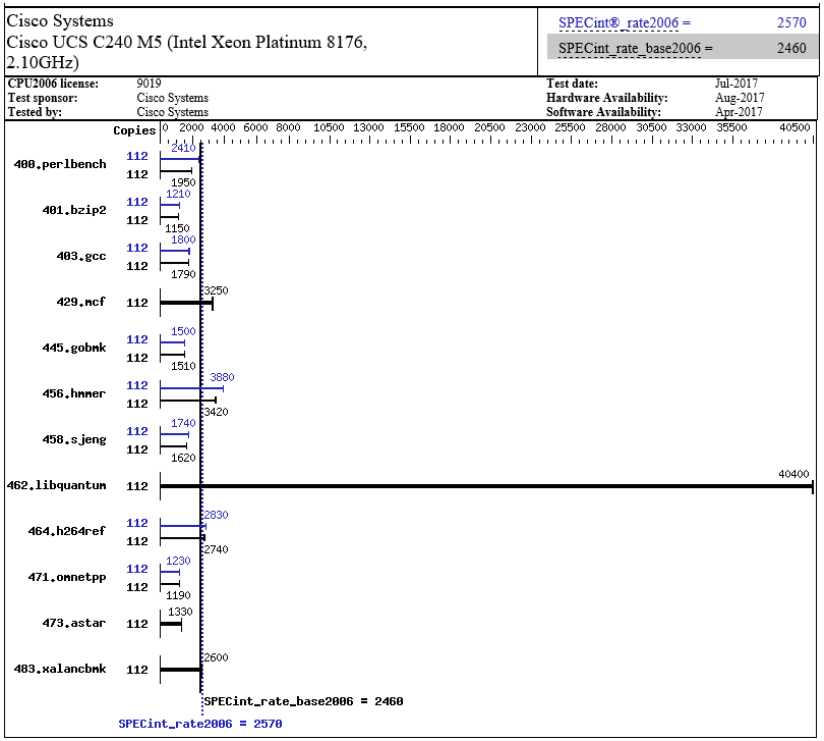
Hardware		Software	
CPU Name:	Intel Xeon E5-2699 v4	Operating System:	Red Hat Enterprise Linux Server release 7.2 (Maipo) 3.10.0-327.el7.x86_64
CPU Characteristics:	Intel Turbo Boost Technology up to 3.60 GHz	Compiler:	C/C++: Version 16.0.0.101 of Intel C++ Studio XE for Linux
CPU MHz:	2200	Auto Parallel:	No
FPU:	Integrated	File System:	xfs
CPU(s) enabled:	44 cores, 2 chips, 22 cores/chip, 2 threads/core	System State:	Run level 3 (multi-user)
CPU(s) orderable:	1,2 chip	Base Pointers:	32-bit
Primary Cache:	32 KB I + 32 KB D on chip per core	Peak Pointers:	32/64-bit
Secondary Cache:	256 KB I+D on chip per core	Other Software:	Microquill SmartHeap V10.2
L3 Cache:	55 MB I+D on chip per chip		
Other Cache:	None		
Memory:	256 GB (16 x 16 GB 2Rx8 PC4-2400T-R)		
Disk Subsystem:	1 x 1000 GB SATA, 7200 RPM		
Other Hardware:	None		



Source: http://spec.org/cpu2006/results/res2016q4/cpu2006-20161020-44646.html

Intel 8176 SPEC Detail

Hardware		Software	
CPU Name:	Intel Xeon Platinum 8176	Operating System:	SUSE Linux Enterprise Server 12 SP2 4.4.21-69-default
CPU Characteristics:	Intel Turbo Boost Technology up to 3.80 GHz	Compiler:	C/C++: Version 17.0.3.191 of Intel C/C++ Compiler for Linux
CPU MHz:	2100	Auto Parallel:	Yes
FPU:	Integrated	File System:	xfs
CPU(s) enabled:	56 cores, 2 chips, 28 cores/chip, 2 threads/core	System State:	Run level 3 (multi-user)
CPU(s) orderable:	1,2 chips	Base Pointers:	32-bit
Primary Cache:	32 KB I + 32 KB D on chip per core	Peak Pointers:	32/64-bit
Secondary Cache:	1 MB I+D on chip per core	Other Software:	Microquill SmartHeap V10.2
L3 Cache:	38.5 MB I+D on chip per chip		
Other Cache:	None		
Memory:	384 GB (24 x 16 GB 2Rx4 PC4-2666V-R)		
Disk Subsystem:	1 x 600 GB SAS 10K RPM		
Other Hardware:	None		



Source: http://spec.org/cpu2006/results/res2017q3/cpu2006-20170725-47928.html

Intel E5-2699 v4 SPEC Detail

Hardware	
Hardware Vendor:	Inspur Corporation
Model:	NF5280 M4
Form Factor:	2U
CPU Name:	Intel Xeon E5-2699 v4
CPU Characteristics:	22-Core, 2.20GHz, 55MB L3 Cache
CPU Frequency (MHz):	2200
CPU(s) Enabled:	44 cores, 2 chips, 22 cores/chip
Hardware Threads:	88 (2 / core)
CPU(s) Orderable:	1,2 chips
Primary Cache:	32 KB I + 32 KB D on chip per core
Secondary Cache:	256 KB I+D on chip per core
Tertiary Cache:	55 MB I+D on chip per chip
Other Cache:	None
Memory Amount (GB):	128
# and size of DIMM:	8 x 16 GB
Memory Details:	16 GB 2Rx8 PC4-2400N ECC; slots CHA0, CHB0, CHC0, CHD0, CHE0, CHF0, CHG0, CHH0 populated
Power Supply Quantity and Rating (W):	1 x 800
Power Supply Details:	Delta DPS-800AB-9X P/N:V07LP0400000000F
Disk Drive:	Intel SSD DC S3500 Series 6Gb/s 80G PN:ZMHD1000018
Disk Controller:	Integrated SATA Controller
# and type of Network Interface Cards (NICs) Installed:	1 x Intel I350 OnBoard LAN
NICs Enabled in Firmware / OS / Connected:	2/2/1
Network Speed (Mbit):	1000
Keyboard:	None
Mouse:	None
Monitor:	None
Optical Drives:	No
Other Hardware:	None

Software	
Power Management:	Enabled ("Balanced Performance" power plan)
Operating System (OS):	Microsoft Windows Servers 2012 R2 Standard
OS Version:	Version 6.3.9600 Build 9600
Filesystem:	NTFS
JVM Vendor:	Oracle Corporation
JVM Version:	Oracle Java HotSpot(TM)64-Bit Server VM(build 24.80-b11,mixed mode),version 1.7.0_80
JVM Command-line Options:	-server -Xmn11g -Xms13g -Xmx13g -XX:SurvivorRatio=60 -XX:TargetSurvivorRatio=90 -XX:AllocatePrefetchDistance=256 -XX:AllocatePrefetchLines=4 -XX:LoopUnrollLimit=45 -XX:InitialTenuringThreshold=12 -XX:MaxTenuringThreshold=15 -XX:ParallelGCThreads=22 -XX:InlineSmallCode=3900 -XX:MaxInlineSize=270 -XX:FreqInlineSize=2500 -XX:+AggressiveOpts -XX:+UseLargePages -XX:+UseParallelOldGC -XX:-UseAdaptiveSizePolicy
JVM Affinity:	start /NODE [0,1,2,3] /affinity 0x3FFFFF
JVM Instances:	4
JVM Initial Heap (MB):	13000
JVM Maximum Heap (MB):	13000
JVM Address Bits:	64
Boot Firmware Version:	4.1.8
Management Firmware Version:	4.15
Workload Version:	SSJ 1.2.10
Director Location:	Controller
Other Software:	None

Performance			Power	Performance to Power Ratio
Target Load	Actual Load	ssj_ops	Average Active Power (W)	
100%	99.7%	3,561,599	245	14,567
90%	89.9%	3,210,954	221	14,551
80%	80.1%	2,859,402	191	14,951
70%	70.1%	2,504,652	171	14,637
60%	59.9%	2,138,538	156	13,676
50%	49.9%	1,783,222	144	12,362
40%	40.0%	1,427,593	130	10,988
30%	29.9%	1,068,982	113	9,483
20%	20.0%	713,846	97.5	7,325
10%	10.0%	357,111	82.1	4,349
Active Idle		0	45.7	0
			∑ssj_ops / ∑power =	**12,296**

Source: http://spec.org/power_ssj2008/results/res2017q3/power_ssj2008-20170807-00775.html

Intel 8176 SPEC Detail

Hardware	
Hardware Vendor:	Huawei Technologies Co., Ltd
Model:	Fusion Server 2288H V5
Form Factor:	2U
CPU Name:	Intel Xeon Platinum 8176
CPU Characteristics:	28-Core, 2.1GHz, 38.5MB L3 Cache (Intel Turbo Boost Technology up to 2.8GHz)
CPU Frequency (MHz):	2100
CPU(s) Enabled:	56 cores, 2 chips, 28 cores/chip
Hardware Threads:	112 (2 / core)
CPU(s) Orderable:	1,2 chips
Primary Cache:	32 KB I + 32 KB D on chip per core
Secondary Cache:	1 MB I+D on chip per core
Tertiary Cache:	39424 KB I+D on chip per chip
Other Cache:	None
Memory Amount (GB):	192
# and size of DIMM:	12 x 16 GB
Memory Details:	12 x 16GB 2Rx8 PC4-2666V ECC RDIMM;slots DIMM000, DIMM010, DIMM020, DIMM030, DIMM040, DIMM050,DIMM100, DIMM110, DIMM120, DIMM130, DIMM140, and DIMM150 populated
Power Supply Quantity and Rating (W):	1 x 550
Power Supply Details:	Huawei P/N 02131255
Disk Drive:	1 x 240GB SSD 2.5" SATA Huawei P/N 02311HGX
Disk Controller:	Integrated SATA controller
# and type of Network Interface Cards (NICs) installed:	1 x Dual-port Intel X722 Gigabit Ethernet controller
NICs Enabled in Firmware / OS / Connected:	2/2/1
Network Speed (Mbit):	1000
Keyboard:	None
Mouse:	None
Monitor:	None
Optical Drives:	No
Other Hardware:	None

Software	
Power Management:	Balanced (recommended) Mode enabled in OS
Operating System (OS):	Microsoft Windows Server 2012 R2 Datacenter
OS Version:	Version 6.3 (Build 9600)
Filesystem:	NTFS
JVM Vendor:	Oracle Corporation
JVM Version:	Oracle Java HotSpot(TM) 64-Bit Server VM (build 24.80-b11, mixed mode), version 1.7.0_80
JVM Command-line Options:	-server -Xmn1300m -Xms1550m -Xmx1550m -XX:SurvivorRatio=1 -XX:TargetSurvivorRatio=99 -XX:ParallelGCThreads=2 -XX:AllocatePrefetchDistance=256 -XX:AllocatePrefetchLines=4 -XX:LoopUnrollLimit=45 -XX:InitialTenuringThreshold=12 -XX:MaxTenuringThreshold=15 -XX:InlineSmallCode=3900 -XX:MaxInlineSize=270 -XX:FreqInlineSize=2500 -XX:+AggressiveOpts -XX:+UseLargePages -XX:+UseParallelOldGC
JVM Affinity:	start /NODE [0,2,] /AFFINITY [3,C,30,C0,C000,30000,C0000,30000000,C0000000,300000000,C00000000,C0000000000,300000000000,C00000000000];start /NODE [1,3,] /AFFINITY [3,C,30,3000,C000,30000,C0000,30000000,C0000000,300000000,30000000000,C0000000000,300000000000,C00000000000];
JVM Instances:	56
JVM Initial Heap (MB):	1550
JVM Maximum Heap (MB):	1550
JVM Address Bits:	64
Boot Firmware Version:	0.20
Management Firmware Version:	2.53
Workload Version:	SSJ 1.2.10
Director Location:	Controller
Other Software:	None

Performance			Power	Performance to Power Ratio
Target Load	Actual Load	ssj_ops	Average Active Power (W)	
100%	99.9%	5,941,377	461	12,882
90%	90.2%	5,362,581	381	14,083
80%	79.9%	4,753,465	313	15,181
70%	70.0%	4,160,796	277	15,029
60%	60.1%	3,573,477	233	15,362
50%	49.9%	2,969,633	201	14,739
40%	39.9%	2,374,390	182	13,065
30%	29.9%	1,778,877	162	10,985
20%	19.9%	1,186,091	142	8,375
10%	10.0%	596,165	120	4,957
	Active Idle	0	49.6	0
			∑ssj_ops / ∑power =	**12,968**

Source: http://spec.org/power_ssj2008/results/res2017q3/power_ssj2008-20170621-00759.html

Notes and Disclaimers

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Performance varies depending on system configuration.

No computer system can be absolutely secure.

Tests document performance of components on a particular test, in specific systems. Differences in hardware, software, or configuration will affect actual performance. For more complete information about performance and benchmark results, visit http://www.intel.com/benchmarks .

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 complete information visit http://www.intel.com/benchmarks .

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.

Intel® Advanced Vector Extensions (Intel® AVX)* provides higher throughput to certain processor operations. Due to varying processor power characteristics, utilizing AVX instructions may cause a) some parts to operate at less than the rated frequency and b) some parts with Intel® Turbo Boost Technology 2.0 to not achieve any or maximum turbo frequencies. Performance varies depending on hardware, software, and system configuration and you can learn more at http://www.intel.com/go/turbo.

Intel's compilers may or may not optimize to the same degree for non-Intel microprocessors for optimizations that are not unique to Intel microprocessors. These optimizations include SSE2, SSE3, and SSSE3 instruction sets and other optimizations. Intel does not guarantee the availability, functionality, or effectiveness of any optimization on microprocessors not manufactured by Intel. Microprocessor-dependent optimizations in this product are intended for use with Intel microprocessors. Certain optimizations not specific to Intel microarchitecture are reserved for Intel microprocessors. Please refer to the applicable product User and Reference Guides for more information regarding the specific instruction sets covered by this notice.

Cost reduction scenarios described are intended as examples of how a given Intel-based product, in the specified circumstances and configurations, may affect future costs and provide cost savings. Circumstances will vary. Intel does not guarantee any costs or cost reduction.

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.

OCP SUMMIT