



INTRODUCTION OF CASCADELAKE ---- THE SECOND GENERATION INTEL® XEON-SP PLATFORM

Legal Disclaimer

You may not use or facilitate the use of this document in connection with any infringement or other legal analysis concerning Intel products described herein. You agree to grant Intel a non-exclusive, royalty-free license to any patent claim thereafter drafted which includes subject matter disclosed herein.

No license (express or implied, by estoppel or otherwise) to any intellectual property rights is granted by this document.

All information provided here is subject to change without notice. Contact your Intel representative to obtain the latest Intel product specifications and roadmaps.

The products described may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request.

Results have been estimated or simulated using internal Intel analysis or architecture simulation or modeling, and provided to you for informational purposes. Any differences in your system hardware, software or configuration may affect your actual performance.

Copies of documents which have an order number and are referenced in this document may be obtained by calling 1-800-548-4725 or by visiting:
<http://www.intel.com/design/literature.htm>.

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. Check with your system manufacturer or retailer or learn more at <http://www.intel.com>.

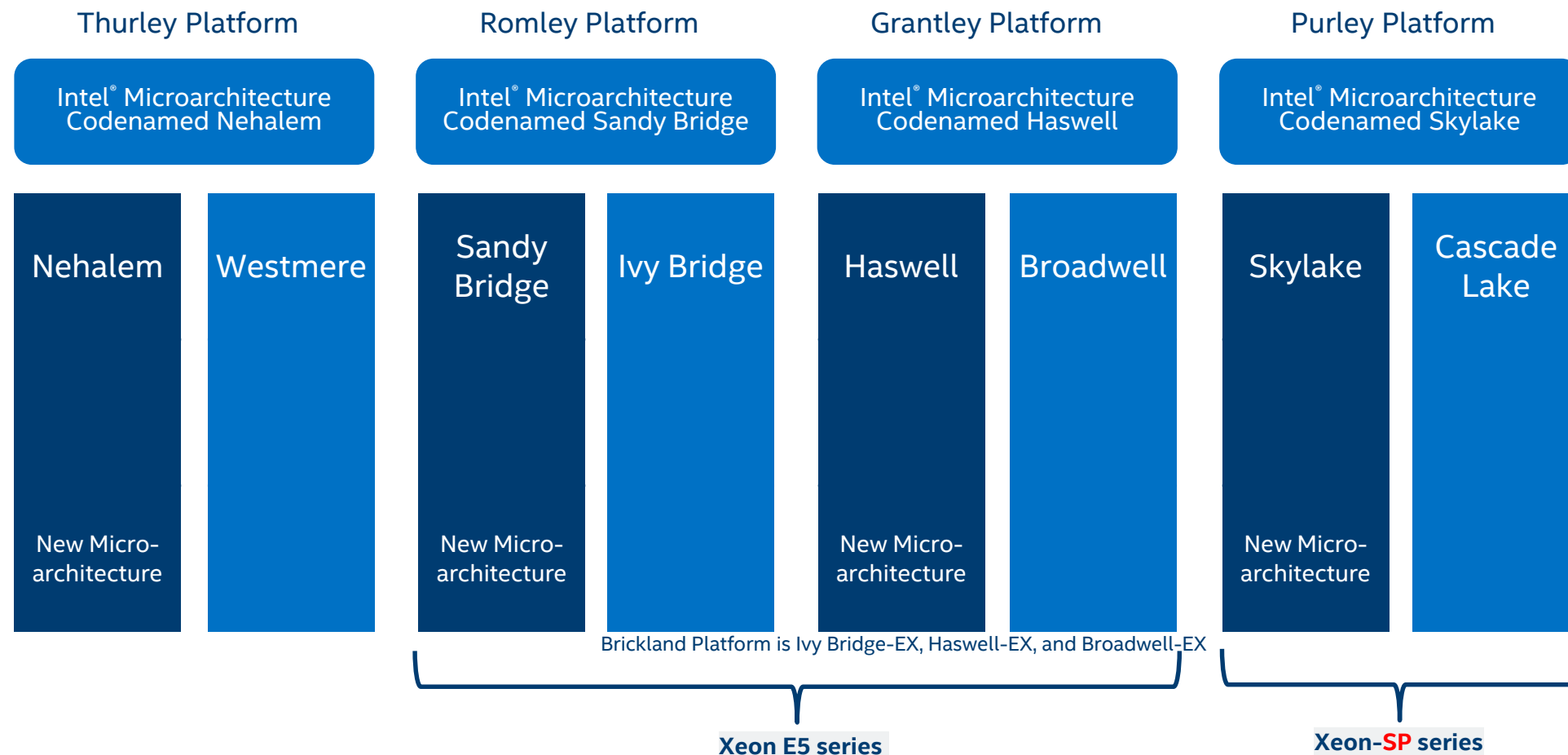
Intel and the Intel logo are trademarks of Intel Corporation in the U.S. and/or other countries.

*Other names and brands may be claimed as the property of others.

Copyright © 2018, Intel Corporation. All rights reserved.

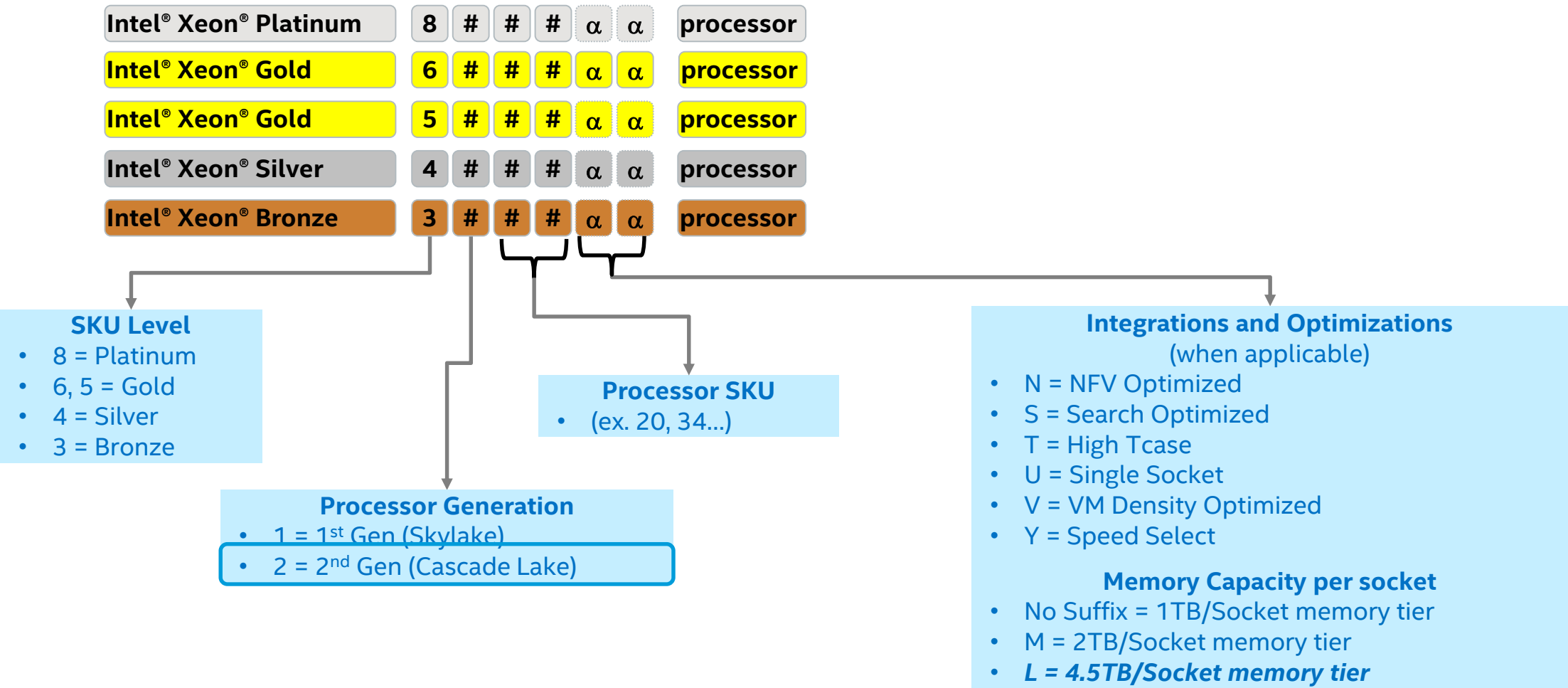
Overview of CascadeLake

2-socket+ Intel® Xeon® Roadmap



Skylake microarchitecture delivers ~10% (geomean) IPC improvement v. Broadwell

Cascade Lake Product Numbering Convention for Intel® Xeon® Scalable Processors



Note: All information provided here is subject to change without notice. Intel may make changes to specifications and product descriptions at any time, without notice. Contact your Intel representative to obtain the latest Intel product specifications and roadmaps

SECOND GENERATION INTEL® XEON® SCALABLE PROCESSORS



CUSTOMER
OBSESSED & WORKLOAD
OPTIMIZED

- INTEL® XEON® PLATINUM 9200 PROCESSORS
- INTEL® XEON® PLATINUM 8200 PROCESSORS
- INTEL® XEON® GOLD 6200 & 5200 PROCESSORS
- INTEL® XEON® SILVER 4200 PROCESSORS
- INTEL® XEON® BRONZE 3200 PROCESSORS

AVAILABLE PROCESSOR OPTIONS

- L LARGE DDR MEMORY TIER SUPPORT UP TO 48 TB
- M MEDIUM DDR MEMORY TIER SUPPORT UP TO 24 TB
- N NETWORKING & NFV SPECIALIZED (INCL. SST-BF)
- S SEARCH VALUE SPECIALIZED
- T THERMAL & LONG-LIFE CYCLE SUPPORT
- V VM DENSITY VALUE SPECIALIZED
- Y INTEL® SPEED SELECT TECHNOLOGY-PP (3 IN 1*)
- TURBO MAXIMUM INTEL® TURBO BOOST TECHNOLOGY 2.0 FREQUENCY (IN GHZ)
- BASE BASE FREQUENCY (IN GHZ)
- CACHE PROCESSOR CACHE (IN MB)
- TDP THERMAL DESIGN POWER (IN WATTS)
- SST-PP INTEL® SPEED SELECT TECH-PERFORMANCE PROFILE
- SST-BF INTEL® SPEED SELECT TECH-BASE FREQUENCY
- RCP RECOMMENDED CUSTOMER PRICING (\$ US DOLLARS)
- NFV NETWORK FUNCTION VIRTUALIZATION
- VM VIRTUAL MACHINE
- NEBS NETWORK EQUIPMENT-BUILDING SYSTEM

ADVANCED PERFORMANCE

9282	56	3.8	2.6	77	400	
CORES	TURBO	BASE	CACHE	TDP		
9242	48	3.8	2.3	71.5	350	
CORES	TURBO	BASE	CACHE	TDP		
9222	32	3.7	2.3	71.5	250	
CORES	TURBO	BASE	CACHE	TDP		
9221	32	3.7	2.1	71.5	250	
CORES	TURBO	BASE	CACHE	TDP		

OPTIMIZED FOR HIGHEST PER-CORE SCALABLE PERFORMANCE

8280	28	4.0	2.7	38.5	205	SUPPORT FOR OPTANE DC CPU RESILIENT MEMORY
CORES	TURBO	BASE	CACHE	TDP		
8270	26	4.0	2.7	35.75	205	SUPPORT FOR OPTANE DC CPU RESILIENT MEMORY
CORES	TURBO	BASE	CACHE	TDP		
8268	24	3.9	2.9	35.75	205	SUPPORT FOR OPTANE DC CPU RESILIENT MEMORY
CORES	TURBO	BASE	CACHE	TDP		
8256	4	3.9	3.8	16.5	105	SUPPORT FOR OPTANE DC CPU RESILIENT MEMORY
CORES	TURBO	BASE	CACHE	TDP		
6254	18	4.0	3.1	24.75	200	SUPPORT FOR OPTANE DC CPU RESILIENT MEMORY
CORES	TURBO	BASE	CACHE	TDP		
6246	12	4.2	3.3	24.75	165	SUPPORT FOR OPTANE DC CPU RESILIENT MEMORY
CORES	TURBO	BASE	CACHE	TDP		
6244	8	4.4	3.6	24.75	150	SUPPORT FOR OPTANE DC CPU RESILIENT MEMORY
CORES	TURBO	BASE	CACHE	TDP		
6242	16	3.9	2.8	22	150	SUPPORT FOR OPTANE DC CPU RESILIENT MEMORY
CORES	TURBO	BASE	CACHE	TDP		
6234	8	4.0	3.3	24.75	130	SUPPORT FOR OPTANE DC CPU RESILIENT MEMORY
CORES	TURBO	BASE	CACHE	TDP		
6226	12	3.7	2.7	19.25	125	SUPPORT FOR OPTANE DC CPU RESILIENT MEMORY
CORES	TURBO	BASE	CACHE	TDP		
5222	4	3.9	3.8	16.5	105	SUPPORT FOR OPTANE DC CPU RESILIENT MEMORY
CORES	TURBO	BASE	CACHE	TDP		
5217	8	3.7	3.0	16.5	115	SUPPORT FOR OPTANE DC CPU RESILIENT MEMORY
CORES	TURBO	BASE	CACHE	TDP		
5215	10	3.4	2.5	16.5	85	SUPPORT FOR OPTANE DC CPU RESILIENT MEMORY
CORES	TURBO	BASE	CACHE	TDP		
4215	8	3.5	2.5	16.5	85	SUPPORT FOR OPTANE DC CPU RESILIENT MEMORY
CORES	TURBO	BASE	CACHE	TDP		

2.0TB & 4.5TB
DDR5 MEMORY
CAPACITY SUPPORT
AVAILABLE

2.0TB & 4.5TB
DDR5 MEMORY
CAPACITY SUPPORT
AVAILABLE

SCALABLE PERFORMANCE

8276	28	4.0	2.2	38.5	165	SUPPORT FOR OPTANE DC CPU RESILIENT MEMORY
CORES	TURBO	BASE	CACHE	TDP		
8260	24	3.9	2.4	35.7	165	SUPPORT FOR OPTANE DC CPU RESILIENT MEMORY
CORES	TURBO	BASE	CACHE	TDP		
8253	16	3.0	2.2	35.7	165	SUPPORT FOR OPTANE DC CPU RESILIENT MEMORY
CORES	TURBO	BASE	CACHE	TDP		
6252	24	3.7	2.1	35.75	150	SUPPORT FOR OPTANE DC CPU RESILIENT MEMORY
CORES	TURBO	BASE	CACHE	TDP		
6248	20	3.9	2.5	27.5	150	SUPPORT FOR OPTANE DC CPU RESILIENT MEMORY
CORES	TURBO	BASE	CACHE	TDP		
6240	18	3.9	2.6	24.75	150	SUPPORT FOR OPTANE DC CPU RESILIENT MEMORY
CORES	TURBO	BASE	CACHE	TDP		
6238	22	3.7	2.1	30.25	140	SUPPORT FOR OPTANE DC CPU RESILIENT MEMORY
CORES	TURBO	BASE	CACHE	TDP		
6230	20	3.9	2.1	27.5	125	SUPPORT FOR OPTANE DC CPU RESILIENT MEMORY
CORES	TURBO	BASE	CACHE	TDP		
5220	18	3.9	2.2	24.75	125	SUPPORT FOR OPTANE DC CPU RESILIENT MEMORY
CORES	TURBO	BASE	CACHE	TDP		
5218	16	3.9	2.3	22	125	SUPPORT FOR OPTANE DC CPU RESILIENT MEMORY
CORES	TURBO	BASE	CACHE	TDP		
4216	16	3.2	2.1	16.5	100	
CORES	TURBO	BASE	CACHE	TDP		
4214	12	3.2	2.2	16.5	85	
CORES	TURBO	BASE	CACHE	TDP		
4210	10	3.2	2.2	13.75	85	
CORES	TURBO	BASE	CACHE	TDP		
4208	8	3.2	2.1	11	85	
CORES	TURBO	BASE	CACHE	TDP		
3204	6	1.9	1.9	8.25	85	
CORES	TURBO	BASE	CACHE	TDP		

2.0TB & 4.5TB
DDR5 MEMORY
CAPACITY SUPPORT
AVAILABLE

2.0TB & 4.5TB
DDR5 MEMORY
CAPACITY SUPPORT
AVAILABLE

2.0TB & 4.5TB
DDR5 MEMORY
CAPACITY SUPPORT
AVAILABLE

2.0TB & 4.5TB
DDR5 MEMORY
CAPACITY SUPPORT
AVAILABLE

FEATURING INTEL® SPEED SELECT TECH-PERFORMANCE PROFILE (SST-PP, 3 IN 1*)

8260Y	24	3.9	2.4	35.75	165	SUPPORT FOR OPTANE DC CPU RESILIENT MEMORY
CORES	TURBO	BASE	CACHE	TDP		
6240Y	18	3.9	2.6	24.75	150	SUPPORT FOR OPTANE DC CPU RESILIENT MEMORY
CORES	TURBO	BASE	CACHE	TDP		
4214Y	12	3.2	2.2	16.5	85	
CORES	TURBO	BASE	CACHE	TDP		

NETWORKING/NFV SPECIALIZED (INCL. INTEL® SPEED SELECT TECH-BF)

6252N	24	3.6	2.3	35.75	150	SUPPORT FOR OPTANE DC CPU RESILIENT MEMORY
CORES	TURBO	BASE	CACHE	TDP		
6230N	20	3.5	2.3	27.5	125	SUPPORT FOR OPTANE DC CPU RESILIENT MEMORY
CORES	TURBO	BASE	CACHE	TDP		
5218N	16	3.9	2.3	22	105	SUPPORT FOR OPTANE DC CPU RESILIENT MEMORY
CORES	TURBO	BASE	CACHE	TDP		

VM DENSITY VALUE SPECIALIZED

6262V	24	3.6	1.9	33	135	SUPPORT FOR OPTANE DC CPU RESILIENT MEMORY
CORES	TURBO	BASE	CACHE	TDP		
6222V	20	3.6	1.8	27.5	115	SUPPORT FOR OPTANE DC CPU RESILIENT MEMORY
CORES	TURBO	BASE	CACHE	TDP		

LONG-LIFE CYCLE AND NEBS-THERMAL FRIENDLY

6238T	22	3.7	1.9	30.25	125	SUPPORT FOR OPTANE DC CPU RESILIENT MEMORY
CORES	TURBO	BASE	CACHE	TDP		
6230T	20	3.9	2.1	27.5	125	SUPPORT FOR OPTANE DC CPU RESILIENT MEMORY
CORES	TURBO	BASE	CACHE	TDP		
5220T	18	3.9	1.9	24.75	105	SUPPORT FOR OPTANE DC CPU RESILIENT MEMORY
CORES	TURBO	BASE	CACHE	TDP		
5218T	16	3.8	2.1	22	105	SUPPORT FOR OPTANE DC CPU RESILIENT MEMORY
CORES	TURBO	BASE	CACHE	TDP		
4209T	8	3.2	2.2	11	70	
CORES	TURBO	BASE	CACHE	TDP		

SEARCH APPLICATION VALUE SPECIALIZED

5220S	18	3.9	2.7	24.75	125	SUPPORT FOR OPTANE DC CPU RESILIENT MEMORY
CORES	TURBO	BASE	CACHE	TDP		

ALL INFORMATION PROVIDED IS SUBJECT TO CHANGE WITHOUT NOTICE. INTEL MAY MAKE CHANGES TO SPECIFICATIONS AND PRODUCT DESCRIPTIONS AT ANY TIME, WITHOUT NOTICE. PLEASE VISIT [INTEL.COM/XEON](https://www.intel.com/xeon) OR CONTACT YOUR INTEL REPRESENTATIVE TO OBTAIN THE LATEST INTEL PRODUCT SPECIFICATIONS. © COPYRIGHT 2019. INTEL CORPORATION.

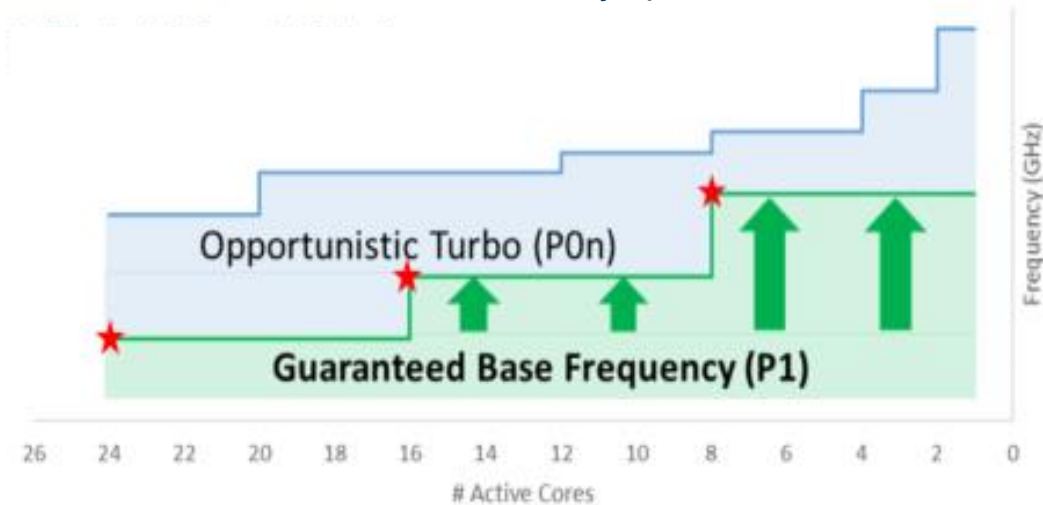
Speed Select Technology

Intel® Speed Select Technology on Cascade Lake

Intel® Speed Select Technology is an umbrella term for a collection of features that provide more granular control over CPU performance

Intel® Speed Select Technology–Performance Profile

(Intel® SST-PP, formerly *Speed Select*)



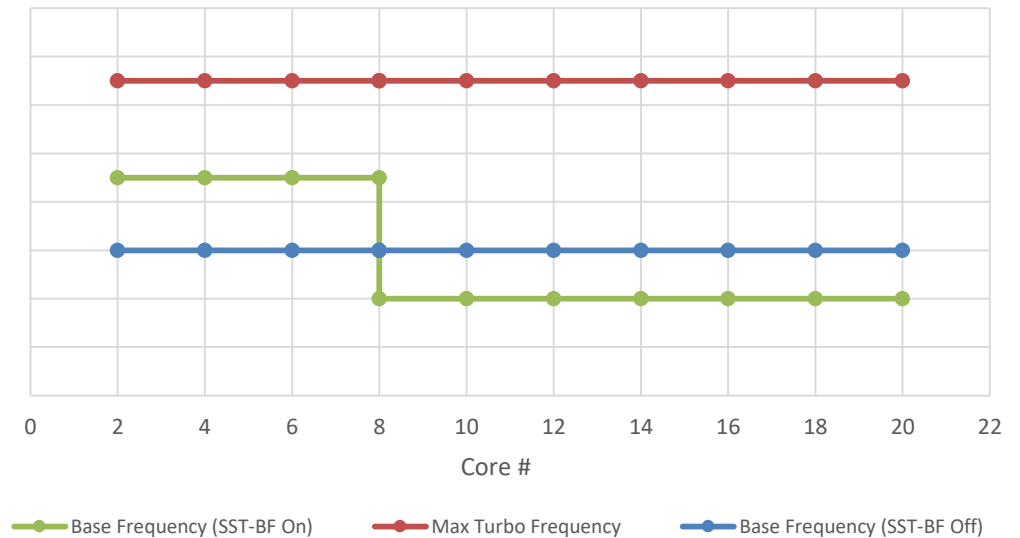
Description: Operate CPU at 3 distinct operating points (ex. with different # of active cores and base frequencies)

Value Prop: TCO optimization by capacity management/SKU consolidation

SKUs: Intel® SST-PP will appear on three CLX Roadmap SKUs: 8260Y, 6240Y, and 4214Y

Intel® Speed Select Technology–Base Frequency

(Intel® SST-BF, formerly *Prioritized Base Frequency*)



Description: Increase base frequency of certain cores and lower base frequency on others

Value Prop: Improve performance by directing base frequency to high priority/bottleneck cores

SKUs: Intel® SST-BF will appear on three CLX Roadmap SKUs: 6252N, 6230N, and 5218N ("N" SKUs are optimized for Network Function Virtualization workloads)

Intel® Speed Select Technology-Performance Profile on Cascade Lake

Capability to configure the CPU to run at 3 distinct operating points

- Each operating point defined by core count with a base frequency associated to that core count
 - Higher core count with lower base frequency
 - Lower core count with higher base frequency
- SKU Stack will include Speed Select specific SKUs

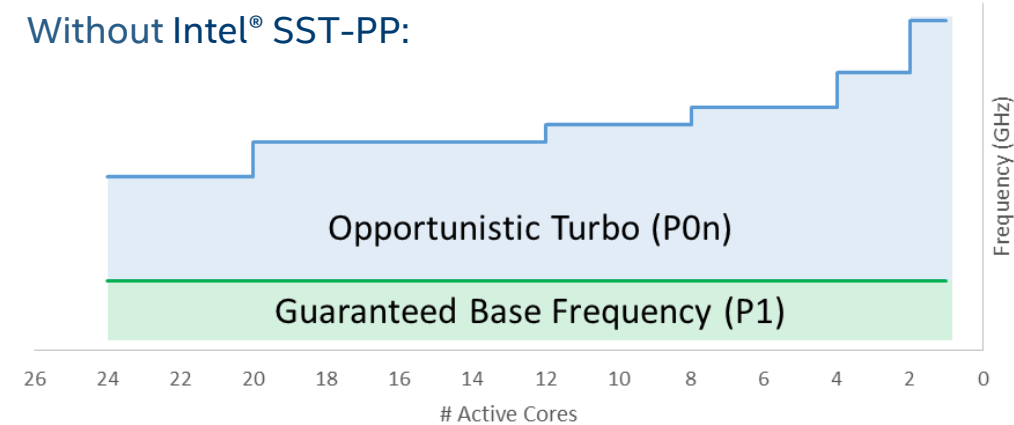
Static Boot Time Configuration

- BIOS discovers capability and prompts user to select from core count / base frequency configurations at boot

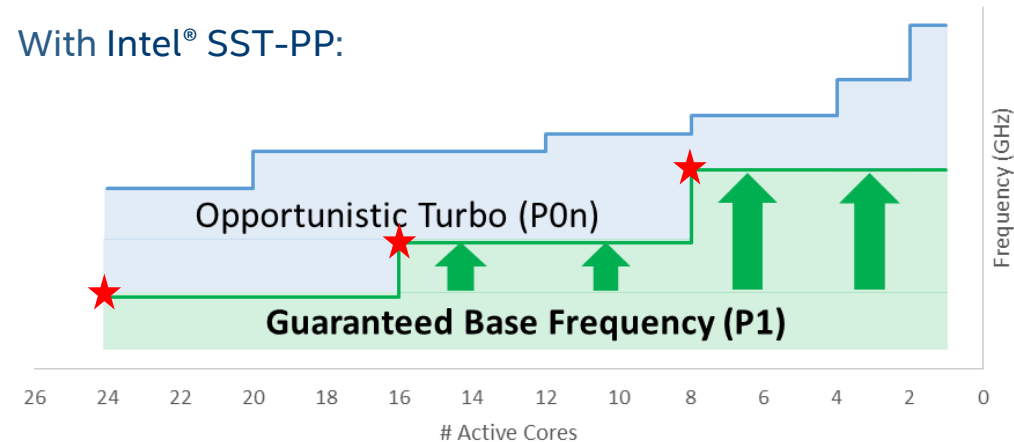
Key Value Prop

- Improved server utilization through SKU consolidation
- Multiple CPU personalities based on workload/VM Needs
- Improved guaranteed per-core performance SLAs

Without Intel® SST-PP:



With Intel® SST-PP:



* Frequency and Core Count for Illustration Only

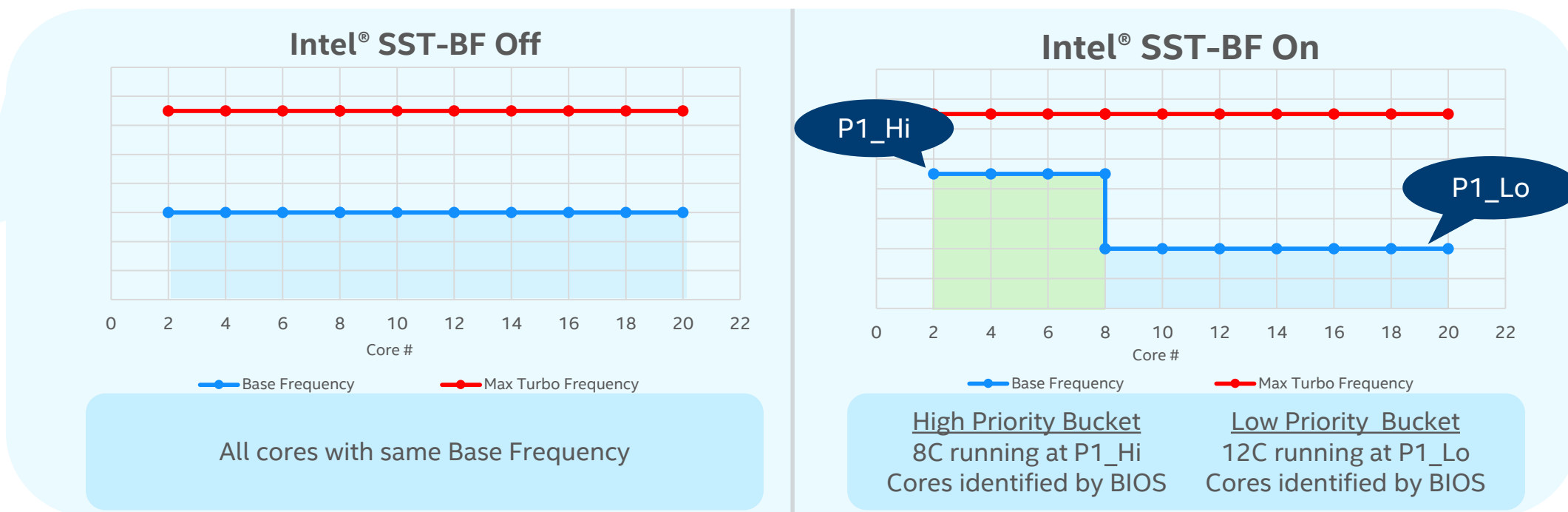
Intel® Speed Select Technology-Base Frequency (Intel® SST-BF)

Description:

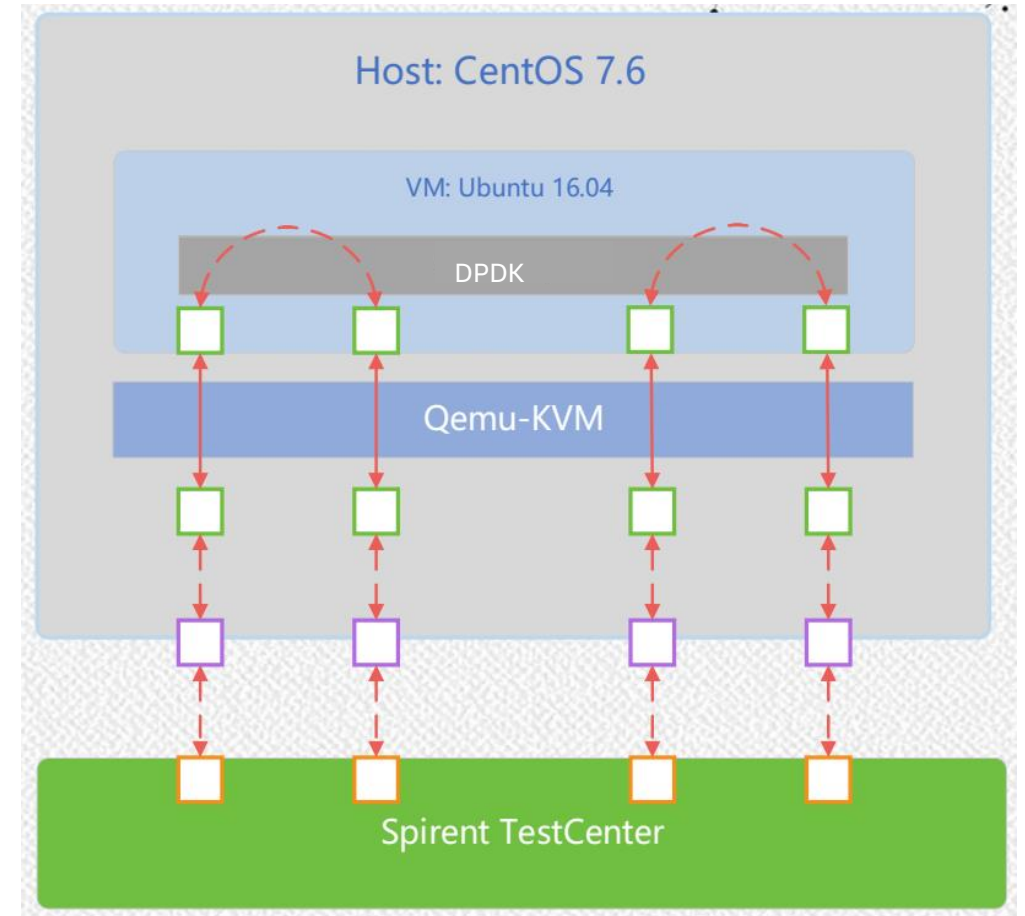
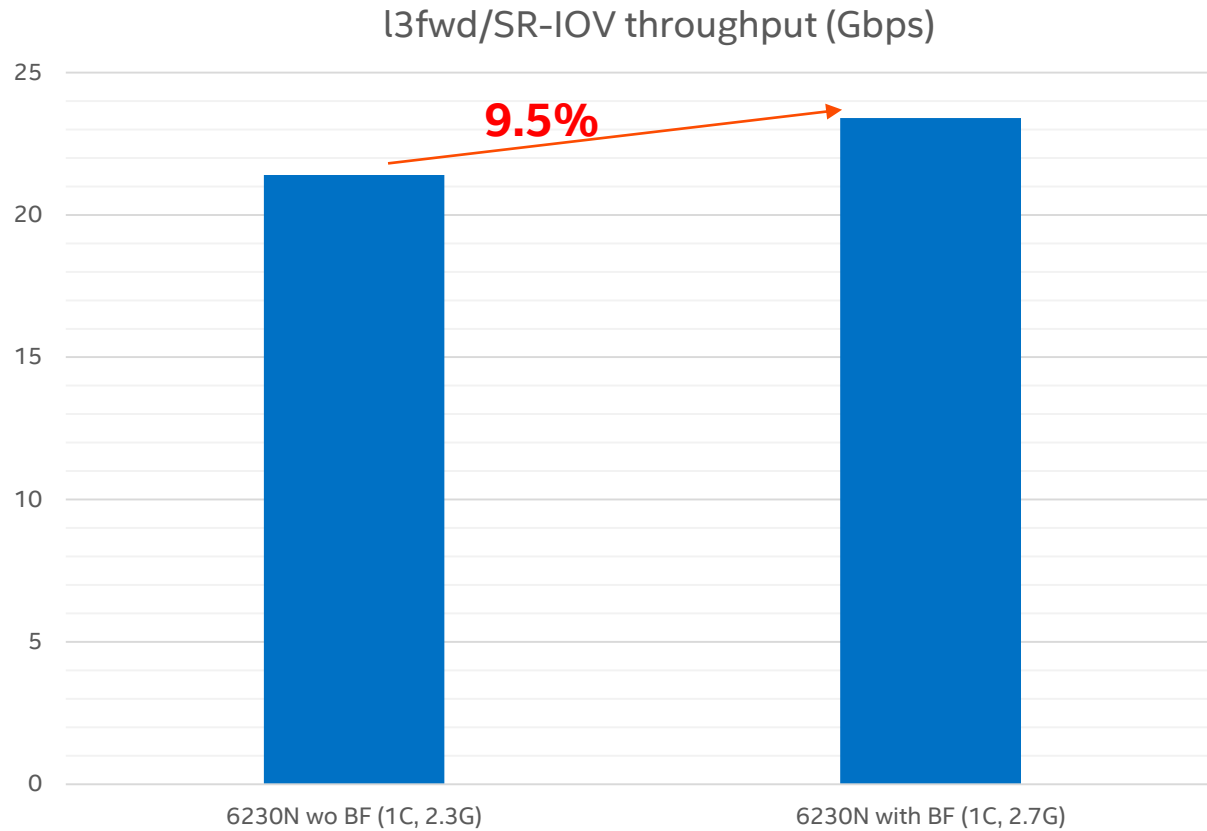
- Ability to increase base frequency on certain cores (*High Priority* cores) in exchange for lowering the base frequency on rest of a CPU's cores (*Low Priority* cores).
- On Intel® SST-BF enabled SKUs, Intel designates whether each core is either a High or Low Priority core. A core's priority designation can not change.

Value: Improve overall performance by boosting base frequency on critical cores

Dynamic: Intel® SST-BF can be activated/deactivated at boot or runtime (after initial BIOS enablement)



SR-IOV Performance with ISS-BF



1 core in Xeon 6230N, 4 * 10G, SR-IOV passthrough, DPDK l3fwd in VM, packet size 64B

OVS-DPDK Performance with ISS-BF

2*Intel Xeon 6230N + 6*10G in a system

Only used 16 cores in CPU 1

6 cores for OVS-DPDK data plane

Low priority Core ID:

ovs-vswitchd Core ID: 20

High priority Core ID:

ovs-pmd: 21,26,27,33,34,36

3 cores for every VPP vRouter VM

Low priority Core ID:

VM 1: 22,23,24

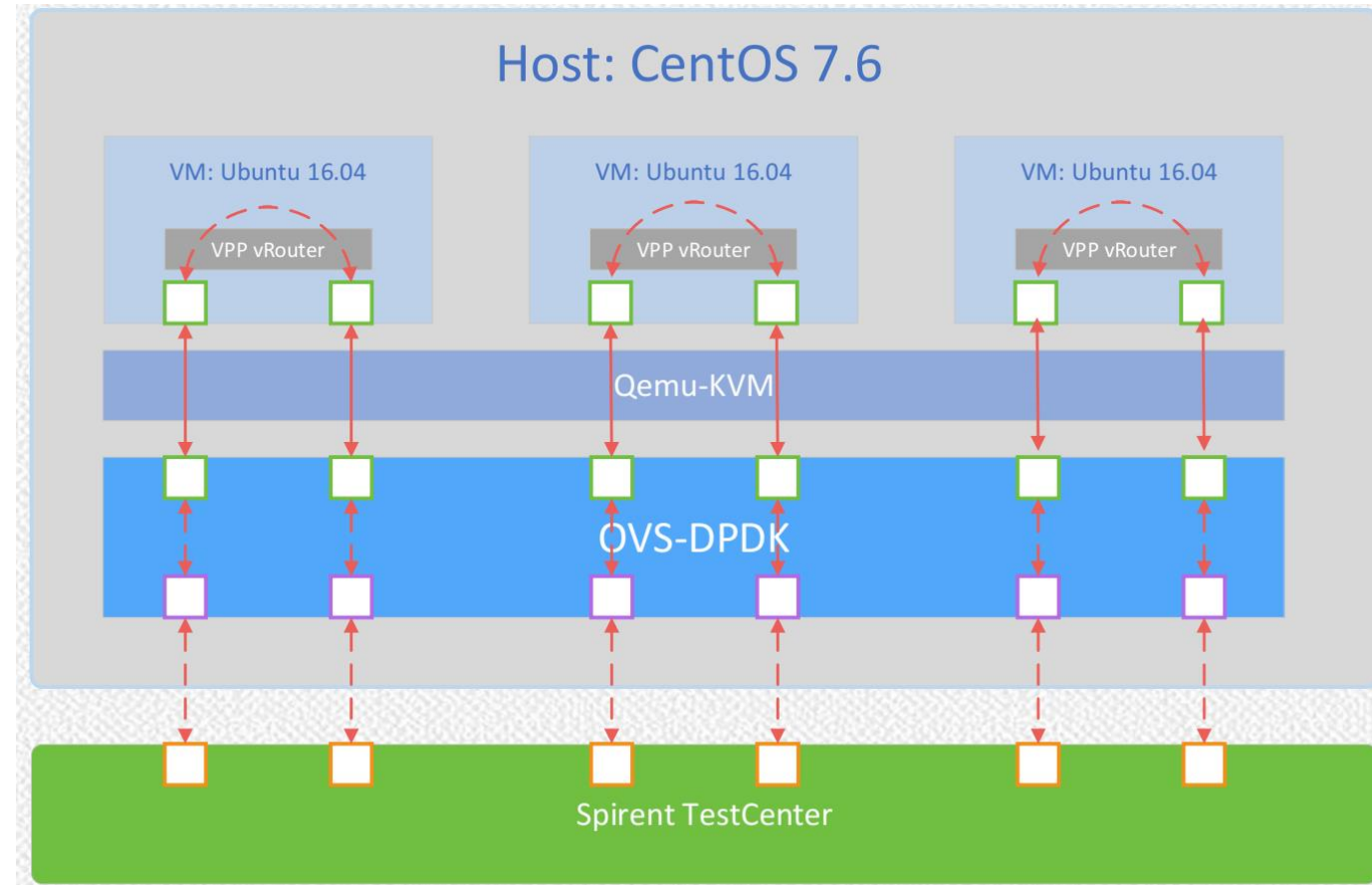
VM 2: 28,29,30

VM 3: 32,35,37

VPP VM core configuration:

VM Core 0: for control plane

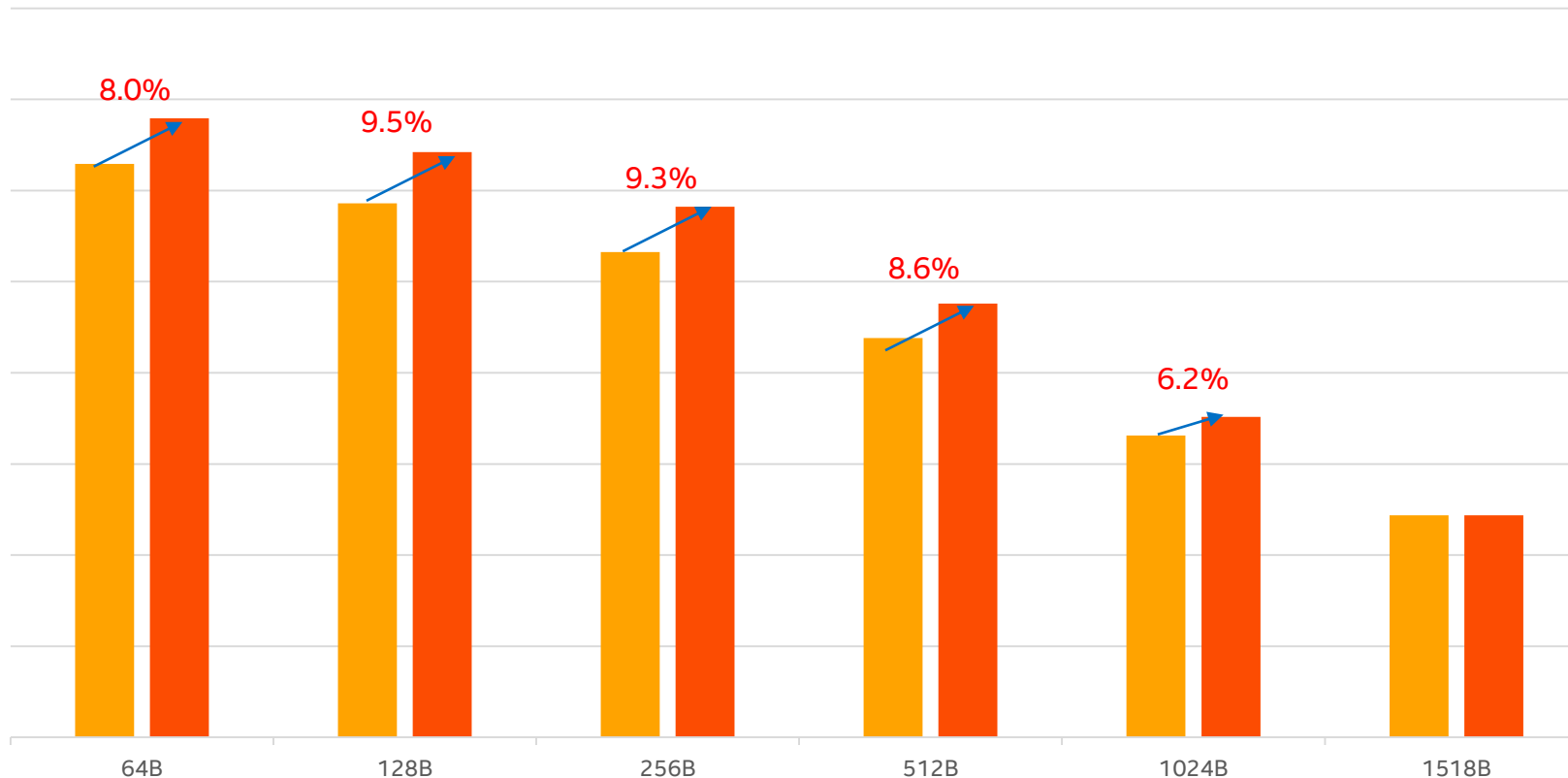
VM Core 1,2: VPP data plane



OVS-DPDK Performance with ISS-BF

Enable ISS-BF(VPP vRouter 2.1Ghz/OVS-DPDK 2.7Ghz) vs Disable ISS-BF(All Core 2.3Ghz)

OVS-DPDK/VPP vRouter Throughput (Mpps)

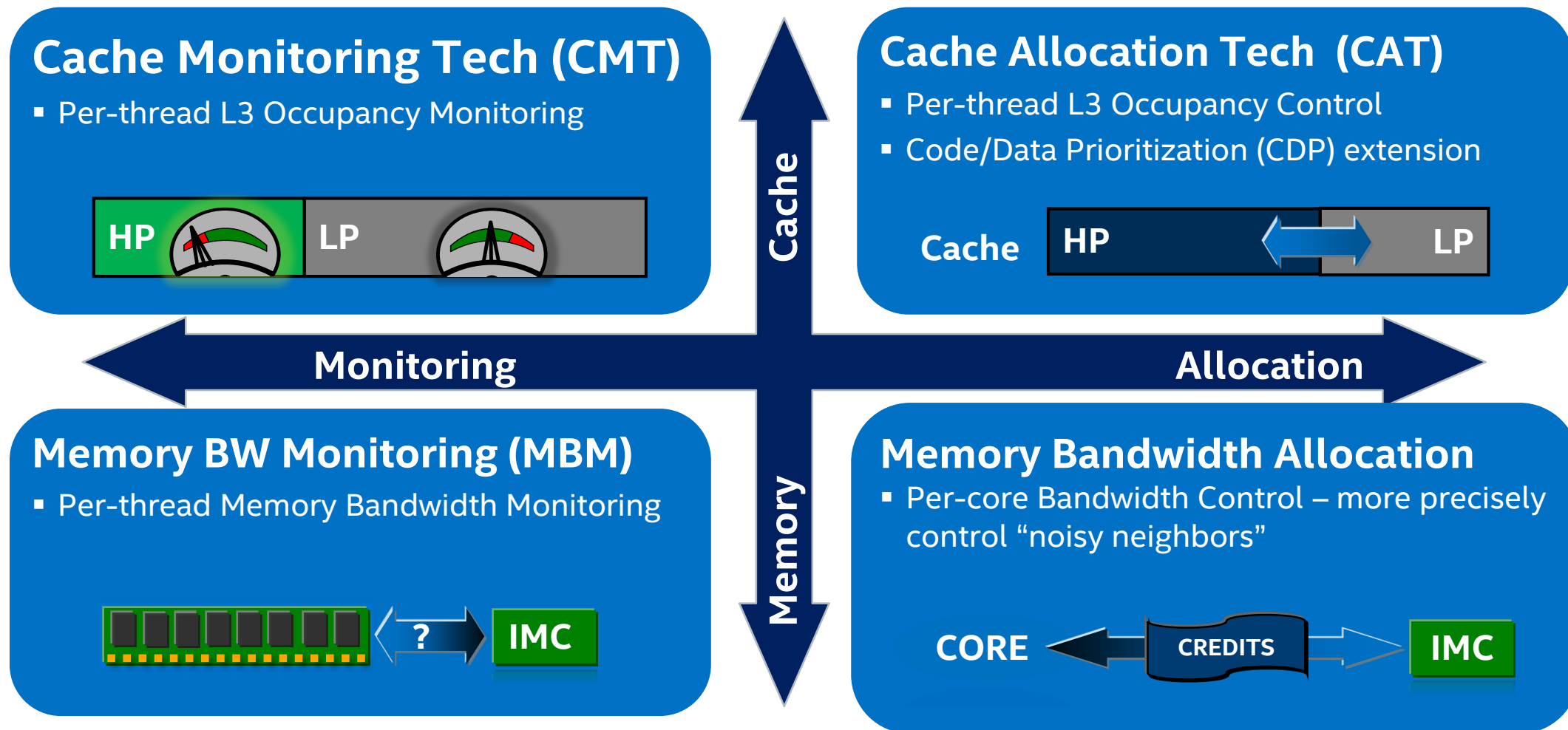


■ 6230N w BF(20C, 2.3G) w 3VM Total 16C

■ 6230N w BF(20C, 2.7/2.1G) w 3VM Total 16C

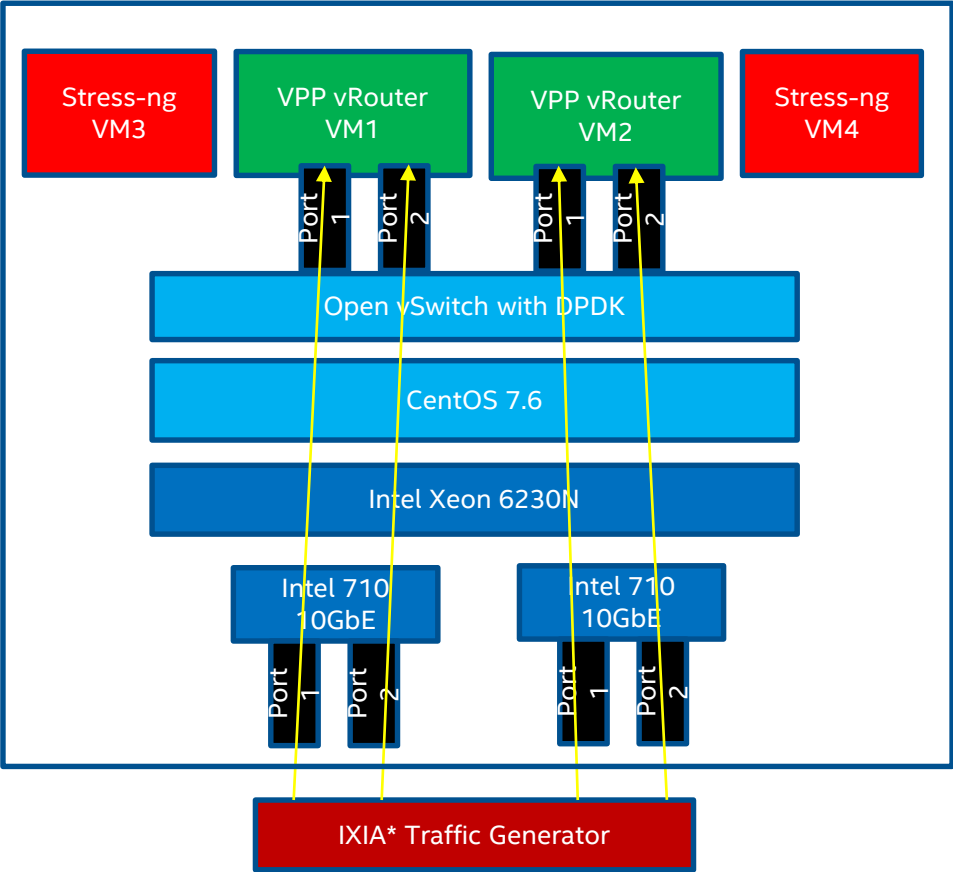
Resource Director Technology

Intel® Resource Director Technology (RDT)



Key Technologies for Improved Visibility and Performance Determinism

RDT Test configuration



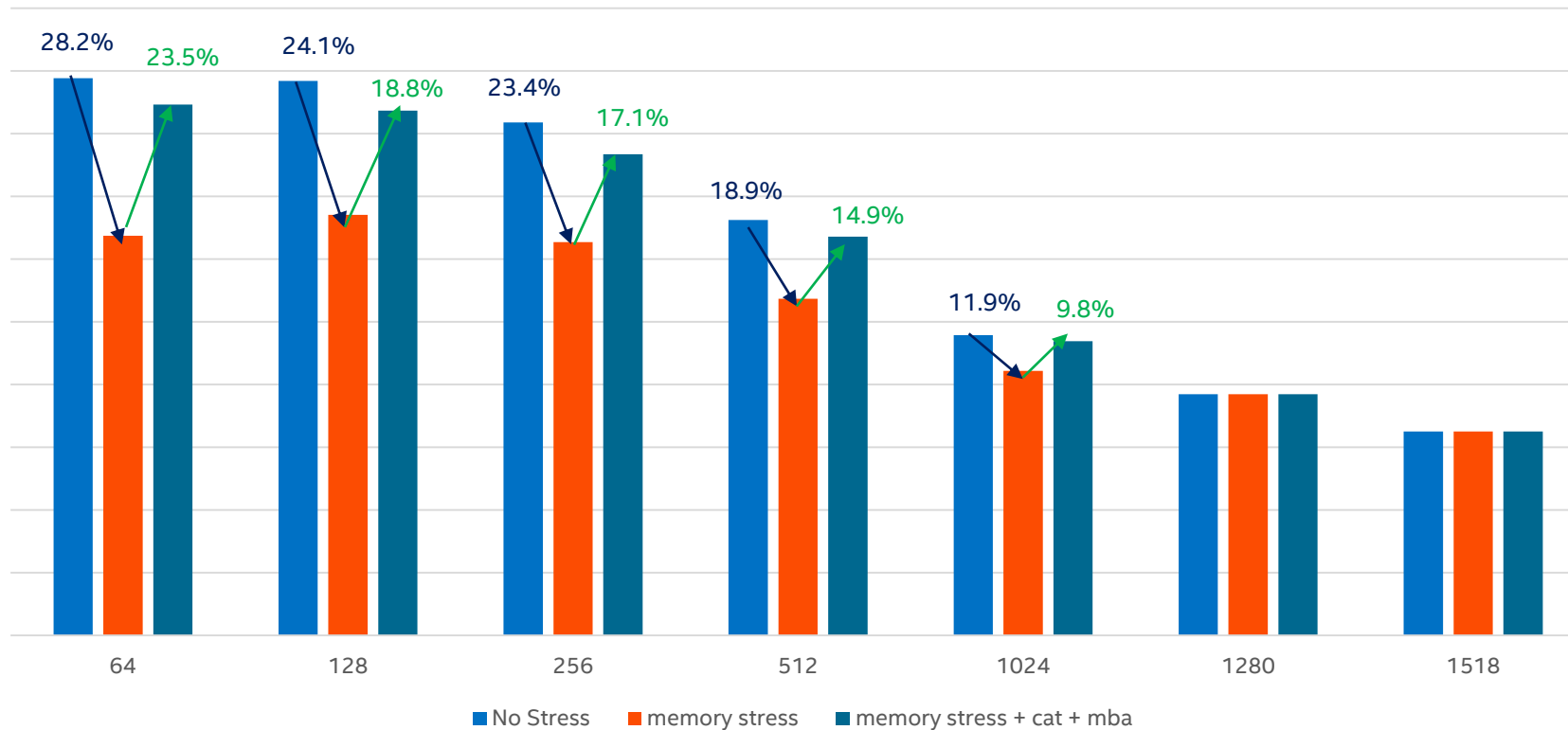
```
pqos -e "llc:0=0x7ff;llc:1=0x7f0;llc:2=0x3;llc:3=0xc;"
pqos -e "mba:0=100;mba:1=100;mba:2=10;mba:3=10;"
pqos -a "llc:0=0-19,40-59"
```

Physical Core	Process/VM	“CAT with Aggressors” Case		Cache Allocation Scheme												Memory Bandwidth Allocation Scheme
		CoS	Capacity Bit Mask (CBM)	11 bit CBM representation												
33,35,37 60,73,75,77	Other App	3	0xC	10	9	8	7	6	5	4	3	2	1	0	10%	
20	ovs-vswitchd	3	0xC	10	9	8	7	6	5	4	3	2	1	0	10%	
21,26,27, 28,29,36	OVS-DPDK PMD	1	0x7F0	10	9	8	7	6	5	4	3	2	1	0	100%	
22,23,24	VM1 - SUT	1	0x7F0	10	9	8	7	6	5	4	3	2	1	0	100%	
30,31,32	VM2 - SUT	1	0x7F0	10	9	8	7	6	5	4	3	2	1	0	100%	
25,34,65,74	VM3 - Noisy Neighbor	2	0x3	10	9	8	7	6	5	4	3	2	1	0	10%	
38,39,78,79	VM4 - Noisy Neighbor	2	0x3	10	9	8	7	6	5	4	3	2	1	0	10%	
0-19,40-59	OS on CPU 0	0	0x7FF												100%	

```
pqos -a "llc:1=21,26,27,28,29,36,22,23,24,30,31,32;llc:2=25,34,38,39;llc:3=20,33,35,37"
pqos -a "llc:1=61,66,67,68,69,76,62,63,64,70,71,72;llc:2=65,74,78,79;llc:3=60,73,75,77"
```

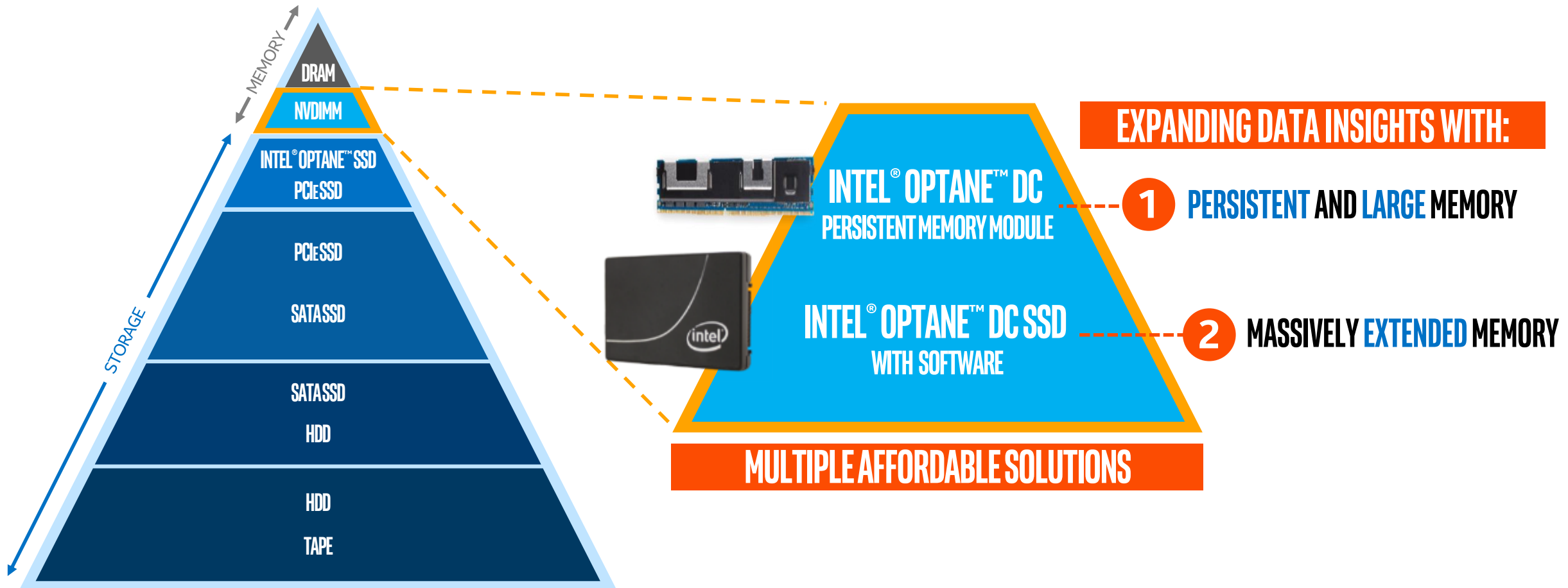
Performance data with RDT

OVS-DPDK/VPP vRouter performance throughput Mpps



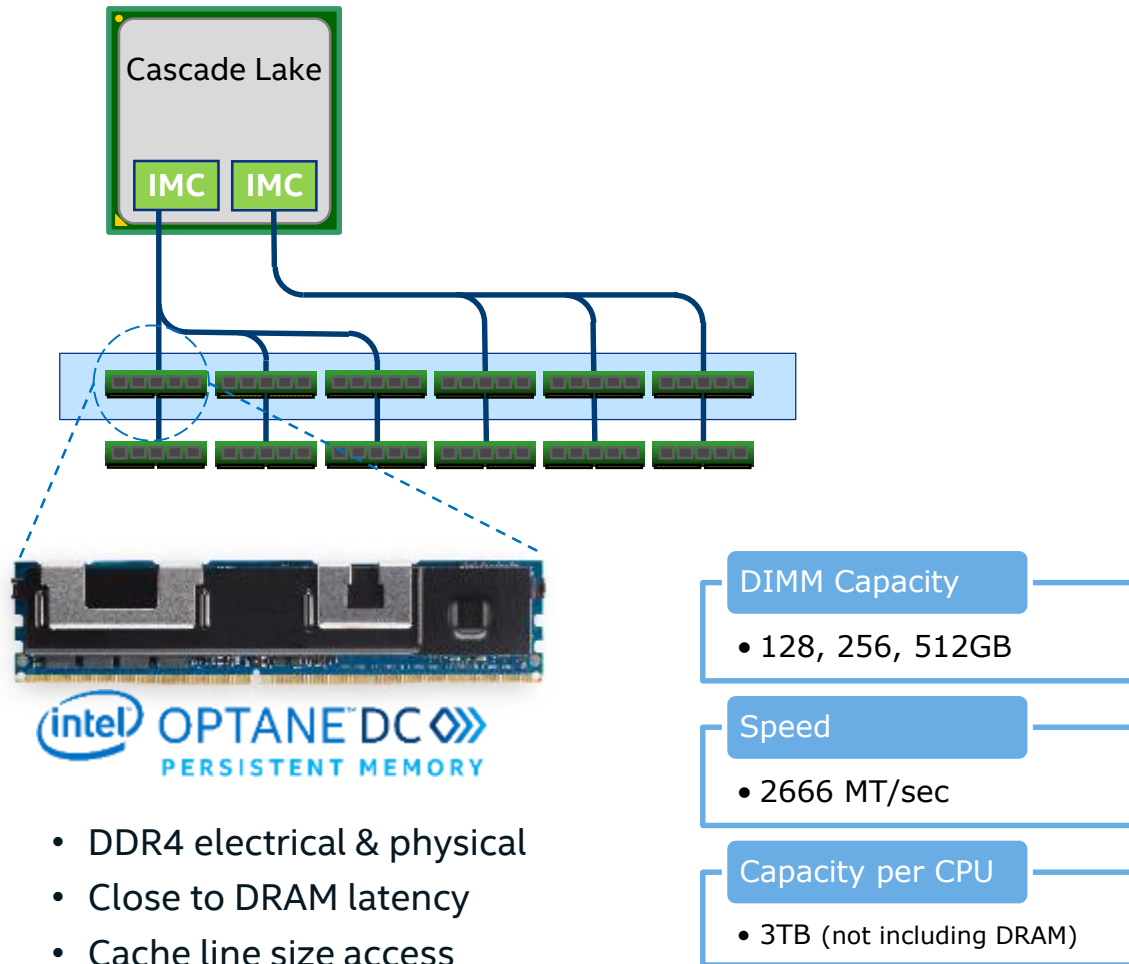
Data Center Persist Memory Module

REDEFINING THE MEMORY / STORAGE HIERARCHY

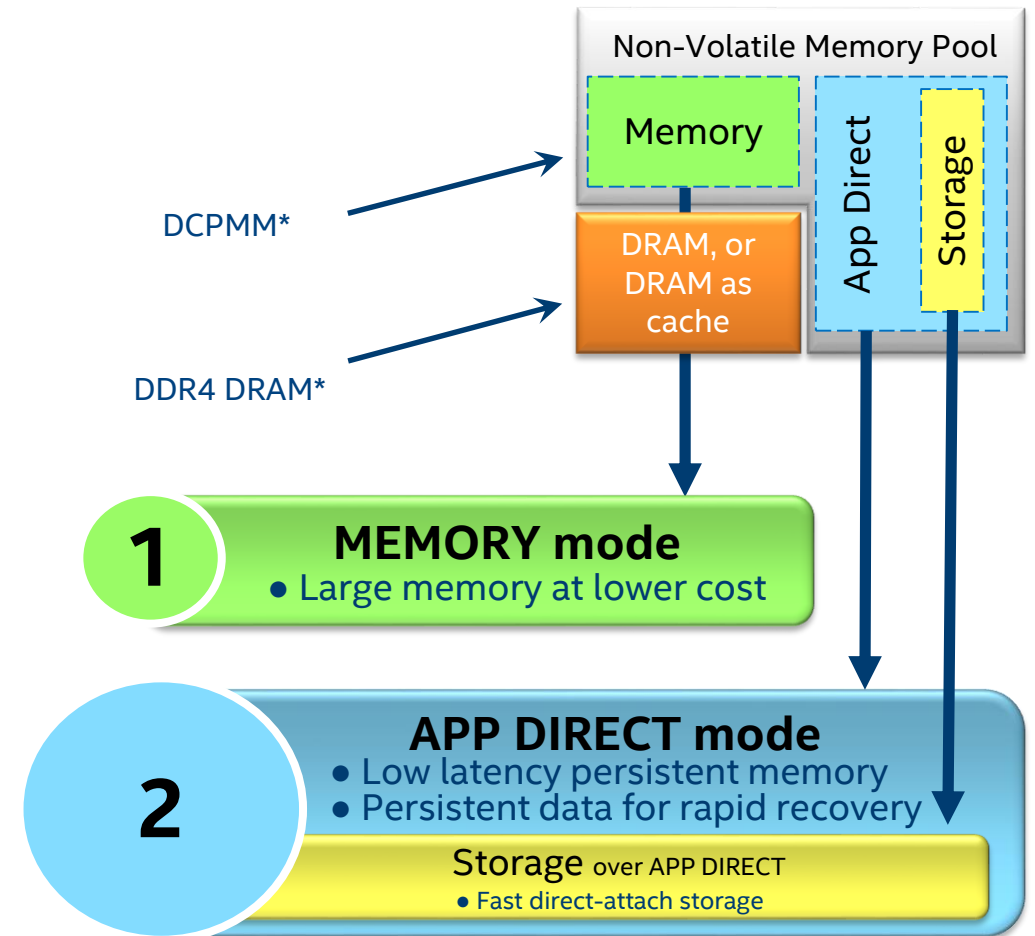


INTEL® OPTANE™ DC PERSISTENT MEMORY - PRODUCT OVERVIEW

(Optane™ based Memory Module for the Data Center)

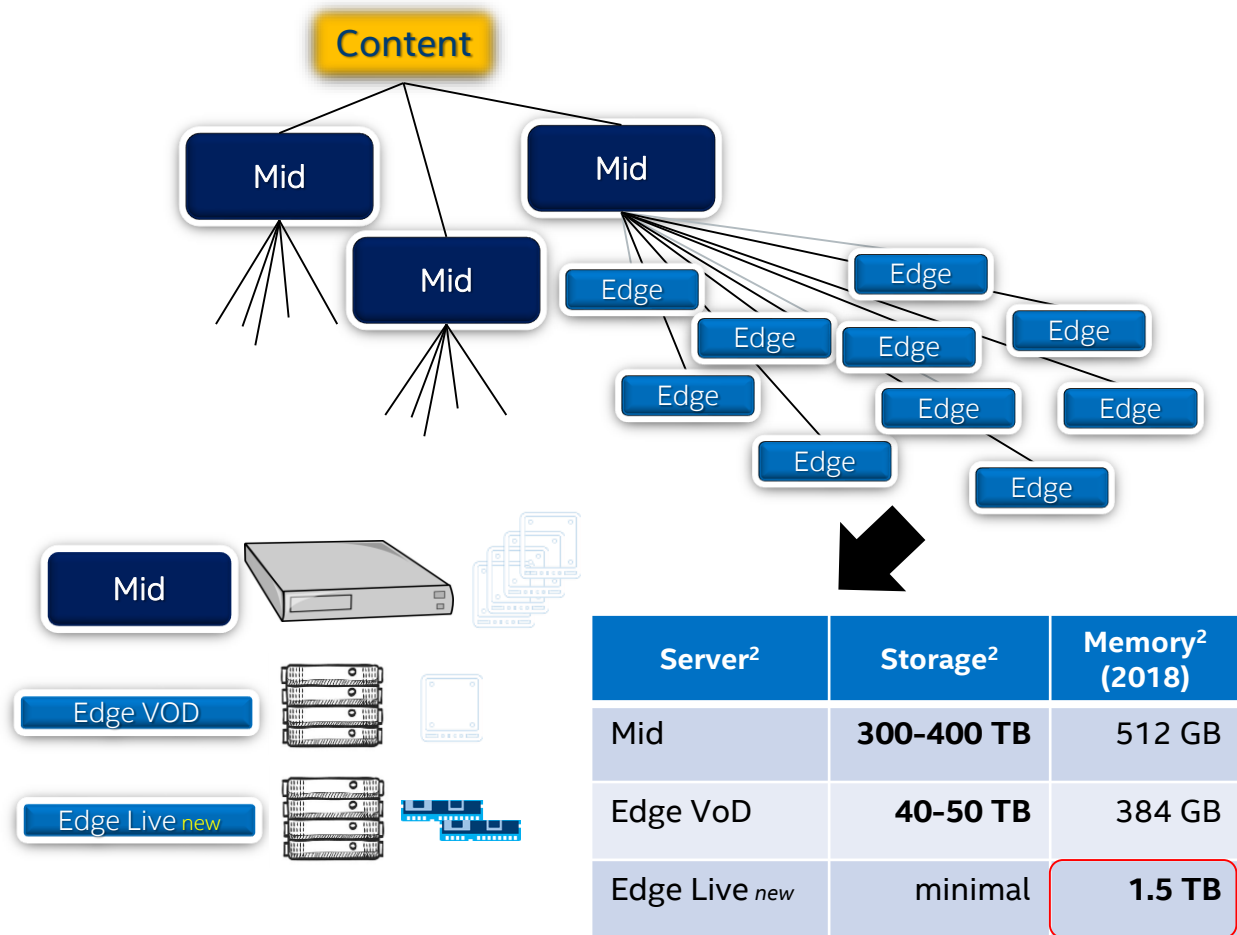


Flexible, Usage Specific Partitions



* DIMM population shown as an example only.

CDN (CONTENT DELIVERY NETWORK): OVERVIEW



CUSTOMER PAIN POINTS

- IP traffic grows **3x** from 2017 to 2022; **82%** is video¹
- Live/linear video traffic grows **15x** in same time frame¹
- For highest quality², hottest VoD content is best cached in **memory (fast)** vs. storage (slow)
- Increasing # of streams per server requires **more memory**
- Memory is costly

SOLUTION

- Intel® Optane™ DC persistent memory delivers huge

VALUE PROPOSITION

- **Scale memory capacity at lower cost**
- Additional opportunity to merge live/linear and VoD nodes

¹Source: Cisco Visual Networking Index: Forecast and Trends, 2017–2022, Nov. 28, 2018, <https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-741490.html>

²Source: Tier 1 CDN vendor(s) serving as lead market definition partner(s)

New Instruction to Boost AI/DL

Cascade Lake: Inference Enhancements with Intel® Deep Learning Boost

Intel® Deep Learning Boost (VNNI) on future Intel® Xeon® Scalable processor (codename “Cascade Lake”) is designed to deliver significant, more efficient Deep Learning (Inference) acceleration.

- Intel® DL Boost (VNNI): A new Intel® Advanced Vector Extension (Intel® AVX-512) instruction
 - 8-bit (VPDPBUSD) new instruction, to accelerate Inference performance.
- No hardware changes are required to support Intel® DL Boost on Purley Platform
 - Minimal OS/VMM enabling if Intel® AVX-512F (foundation) state pre-exists
 - SW development support will be enabled through optimizations on popular AI/Deep Learning frameworks (eg: TensorFlow, Caffe & MXNet) and libraries (Intel® Math Kernel Library – Deep Neural Networks)
- Intel® DL Boost instruction is available on all CLX-SP XCC B-step, HCC and LCC SKUs

