

ComSys2019

INTRODUCING OPTANE DC PERSISTENT MEMORY

Techinical Solution Specialist Fumiyasu Ishibashi

MEMORY AND STORAGE HIERARCHY GAPS THE CAPACITY GAP





BIG AND AFFORDABLE MEMORY 128, 256, 512GB MODULES DDR4 PIN COMPATIBLE

BYTE ADDRESSABLE DIRECT LOAD/STORE ACCESS

HIGH PERFORMANCE STORAGE NATIVE PERSISTENCE

HIGH RELIABILITY AND SECURITY TWO OPERATIONAL MODES



INTEL[®] OPTANE[™] MEDIA TECHNOLOGY

High Resistivity – 'O' Low Resistivity – '1'

Attributes

- + Non-volatile
- + Potentially fast write
- + High density
- + Non-destructive fast read
- + Low voltage
- + Integrate-able w/ logic
- + Bit alterable

First Generation Capacities: 128 GB 256 GB 512 GB

Selectors allow dense packing And individual access to bits

Cross-Point Structure



Breakthrough Material Advances

Compatible switch and memory cell materials

High Performance

Cell and array architecture that can switch fast



Scalable

Memory layers can be stacked in a 3D manner

MORE TO BE GAINED BY BEING ON MEMORY BUS

IDLE AVERAGE RANDOM READ LATENCY¹





See Aopendix J

Performance results are based on testing as of July 24, 2018 set forth in the configurations and may not reflect all publicly available security updates. See configuration disclosure for details. No product can be absolutely secure. For more complete information about performance and benchmark results, visit www.intel.com/benchmarks.

PERFORMANCE Latency vs. Load

(70Read/30Write Random, 4kB for SSD and 256B for Memory)



See Appendix K

Performance results are based on testing as of February 22, 2019 set forth in the Configurations and may not reflect all publicly available security updates. See configuration disclosure for details. No product can be absolutely secure. For more complete information about performance and benchmark results, visit <u>www.intel.com/benchmarks</u>.

PERFORMANCE DETAILS

- Intel[®] Optane[™] DC persistent memory is programmable for different power limits for power/performance optimization
 - 12W 18W, in 0.25 watt granularity for example: 12.25W, 14.75W, 18W
 - Higher power settings give best performance
- Performance varies based on traffic pattern
 - Contiguous 4 cacheline (256B) granularity vs. single random cacheline (64B) granularity
 - Read vs. writes

Granularity	Traffic	Module	Bandwidth
256B (4x64B)	Read		8.3 GB/s
256B (4x64B)	Write		3.0 GB/s
256B (4x64B)	2 Read/1 Write	256GB, 18W	5.4 GB/s
64B	Read		2.13 GB/s
64B	Write		0.73 GB/s
64B	2 Read/1 Write		1.35 GB/s



COMPLETE SYSTEM ON A MODULE

PMIC Generates all the rails for

media and controller

SPI Flash

Where FW is saved

Intel[®] Optane[™] media

11 parallel devices for data + ECC+ spare



DQ buffers

Need for high bit rate signal integrity

AIT DRAM

Where address indirection table lies

Energy store caps

Ensures flushing of all module queues at power fail



INTEL[®] OPTANE[™] DC Persistent memory controller



DESIGNED TO PROVIDE DATA AT REST SECURITY



HARDWARE-ENCRYPTED MEMORY

- Full module protection using 256b AES-XTP encryption engine on board
- Security keys are stored on module in hardware
- Secure cryptographic erase and module over-write for secure repurposing or discard
- Firmware authentication and integrity



MEMORY LEVEL ENDURANCE

FRANKLARD FRANK WEI FRANKLART WEI FRANKLART

Endurance targets are spec'd in Petabytes Written (PBW)

Let's do the math: PBW for 100%wr

- = BW*%write*3600s/hr*24hr*356days/yr*5years/1e6
- = 2.3*3600*24*356*5
- = 353.72PBW

ENDURANCE FOR THE ENTERPRISE

Usage reporting through SMART registers 5-year product warranty

Example Intel[®] Optane[™] DC Persistent Memory Endurance in PBW @15W (4 CL) 256GB

400



Write 15W Write 15W

BANDWIDTH 100% READ 15W 256B	6.8 GB/s	Fro
BANDWIDTH 100% WRITES 15W 256B	2.3 GB/s	m Da
BANDWIDTH 100% READ 15W 64B	1.75 GB/s	Itash
BANDWIDTH 100% WRITES 15W 64B	0.58 GB/s	eet



15W

TWO OPERATIONAL MODES





PERSISTENCY: EARLY ENGAGEMENT WITH INDUSTRY

PERSISTENT MEMORY PROGRAMING MODEL developed through SNIA

FILE MEMORY MANAGEMENT UI **APPLICATION** APPLICATION APPLICATION USER SPAC Standard Standard Load Raw Device File API Store Access MANAGEMENT LIBRARY "DAX" **FILE SYSTEM** PMEM-AWARE MMU MAPPINGS **FILE SYSTEM** PA **GENERIC NVDIMM DRIVER** Ê PERSISTENT MEMORY

PERSISTENT MEMORY DESIGN KIT (PMDK) available on <u>http://pmem.io</u>





INTRODUCING SECOND GENERATION INTEL® XEON® SCALABLE PROCESSORS

PROCESSOR SKU STRUCTURE

PROCESSOR GENERATION

1 FIRST GENERATION

2 SECOND GENERATION

INTEL[®] XEON[®] PLATINUM # 2 # # $\alpha \alpha$ PROCESSOR

PROCESSOR SKU

E.G. 20, 34, ...

PROCESSOR LEVEL

9 PLATINUM 8 PLATINUM 6 GOLD 5 GOLD 4 SILVER 3 BRONZE

NFV: NETWORK FUNCTION VIRTUALIZATION SST-BF: INTEL® SPEED SELECT TECHNOLOGY-BASE FREQUENCY) SST-PP: INTEL® SPEED SELECT TECHNOLOGY-PERFORMANCE PROFILE ALL INFORMATION PROVIDED 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.

PROCESSOR OPTIONS

L LARGE DDR MEMORY TIER SUPPORT (UP TO 4.5TB)

- M MEDIUM DDR MEMORY TIER SUPPORT (UP TO 2.0TB)
- 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 (SST-PP, "3 CPUS IN 1")



S	ECOND GENERATION			
	NIEL AEUN	ADVANCED	PERFORM	ANC
S	CALABLE PROCESSORS	9242	48 CORES	3. TUP
	inte	9222	32 CORES	3. TUF
	PLATIKUM Pratic	9221	32 CORES	3. TUP
	CUSTOMED WORKLOAD	OPTIMIZED F	OR HIGH	EST
	OBSESSED OPTIMIZED	8280	28 CORES	4 . TUR
	INTEL [®] XEON [®] Platinum 9200 processors	8270	26 CORES	4 . TUF
	INTEL® XEON® PLATINUM 8200 processors	8268	24 CORES	3. TUF
1	INTEL [®] XEON [®] GOLD 6200 & 5200 PROCESSORS	8256	4 cores	3. TUF
1		6254	18 CORES	4 . TUR
[[4200 PROLESSORS	6246	12 CORES	4. TUR
	3200 PROCESSORS Available processor options	6244	8 CORES	4 . TUR
L LARGI M MEDIL	E DDR MEMORY TIER SUPPORT	6242	16 CORES	З. тоғ
N NETW S Searc	ORKING & NFV SPECIALIZED (INCL. SST-BF) (H value specialized	6234	8 CORES	4. TUR
V VM DF	MAL & LUNG-LIFE CYCLE SUPPURI Insity val IIF speciai I7FD	6226	12	3.
Y INTEL	° SPEED SELECT TECHNOLOGY-PP ("3 IN 1")	ULLU	CORES	TUF
TURBO M	AXIMUM INTEL® TURBO BOOST TECHNOLOGY 2.0 Requency (in GHz)	5222	4 CORES	3. TUF
BASE B	ASE FREQUENCY (IN GHZ) Rocessor Cache (in Mr)	5217	8	3.
TDP T	HERMAL DESIGN POWER (IN WATTS)		CORES	TUF
SST-PP	ITEL® SPEED SELECT TECH-PERFORMANCE PROFILE	5215	10	3.
RCP R	TEL SPEED SELECT TECH-BASE FREQUENCY ECOMMENDED CUSTOMER PRICING (\$ US DOLLARS)	ALC: NO.	COMED	
NFV N	ETWORK FUNCTION VIRTUALIZATION	4215	8	3.

VM

VIRTUAL MACHINE

NEBS NETWORK EQUIPMENT-BUILDING SYSTEM

DVANCED P	ERFORM	ANCE			_	_	
9242	48 CORES	3.8 TURBO	2.3 BASE	71.5 cache	350 TDP	-	
9222	32 cores	3.7 TURBO	2.3 base	71.5 CACHE	250 TDP		
9221	32 CORES	3.7 TURBO	2.3 BASE	71.5 cache	250 TDP		
PTIMIZED F()R HIGH	EST PER	-CORE	SCALABLE	PERFOR	MANCE	
3280	28 cores	4.0 TURBO	2.7 base	38.5 cache	205 TDP		2.0TB & 4. DDR4 MEMO CAPACITY SUF SKUS AVAILA
8270	26 CORES	4.0 TURBO	2.7 base	35.75 CACHE	205 TDP		
8268	24 CORES	3.9 Turbo	2.9 base	35.75 CACHE	205 TDP		
8256	4 CORES	3.9 Turbo	3.8 base	16.5 cache	105 TDP		
6254	18 cores	4.0 TURBO	3.1 BASE	24.75 CACHE	200 TDP	OPTANE DC 00	
6246	12 CORES	4.2 TURBO	3.3 base	24.75 CACHE	165 TDP		
6244	8 CORES	4.4 TURBO	3.6 base	24.75 CACHE	150 TDP		
6242	16 CORES	3.9 TURBO	2.8 base	22 cache	150 TDP		
6234	8 CORES	4.0 TURBO	3.3 BASE	24.75 CACHE	130 TDP		
6226	12 CORES	3.7 TURBO	2.7 base	19.25 CACHE	125 TDP		
5222	4 CORES	3.9 TURBO	3.8 BASE	16.5 cache	105 TDP		
5217	8 cores	3.7 TURBO	3.0 base	16.5 cache	115 TDP	OPTANE DO	
5215	10 cores	3.4	2.5 BASE	16.5 CACHE	85 TDP		2.0TB & 4. DDR4 MEMC CAPACITY SUP SKUS AVAILA
4215	8 CORES	3.5 TURBO	2.5 BASE	16.5 cache	85 TDP		

5TB

SCALABLE PERFORMANCE								
8276	28 cores	4.0 TURBO	2.2 base	38.5 cache	165 TDP		2.0TB & 4.5 TB DDR4 MEMORY CAPACITY SUPPORT SKUS AVAILABLE	
8260	24 CORES	3.9 TURBO	2.4 BASE	35.7 cache	165 TDP		2.0TB & 4.5TB DDR4 MEMORY CAPACITY SUPPORT SKUS AVAILABLE	
8253	16 CORES	3.0 TURBO	2.2 BASE	35.7 cache	165 TDP			
6252	24 CORES	3.7 TURBO	2.1 base	35.75 CACHE	150 TDP			
6248	20 cores	3.9 TURBO	2.5 base	27.5 CACHE	150 TDP			
6240	18 cores	3.9 TURBO	2.6 BASE	24.75 CACHE	150 TDP		2.0TB & 4.5TB DDR4 MEMORY CAPACITY SUPPORT SKUS AVAILABLE	
6238	22 CORES	3.7 TURBO	2.1 BASE	30.25 сасне	140 TDP		2.0TB & 4.5TB DDR4 MEMORY CAPACITY SUPPORT SKUS AVAILABLE	
6230	20 CORES	3.9 TURBO	2.1 BASE	27.5 CACHE	125 TDP			
5220	18 cores	3.9 TURBO	2.2 base	24.75 CACHE	125 TDP			
5218	16 cores	3.9 TURBO	2.3 base	22 cache	125 TDP			
4216	16 cores	3.2 TURBO	2.1 BASE	16.5 cache	100 TDP			
4214	12 CORES	3.2 TURBO	2.2 BASE	16.5 cache	85 TDP			
4210	10 cores	3.2 TURBO	2.2 BASE	13.75 CACHE	85 TDP			
4208	8 CORES	3.2 TURBO	2.1 BASE	11 cache	85 TDP			
3204	6 CORES	1.9 TURBO	1.9 BASE	8.25 CACHE	85 TDP			

FEATURING INT	EL° SPEE	D SELEC	T TECH	PERFORMA	NCE PROI	FILE (SST-PP, "3 IN 1")
8260Y	24 CORES	3.9 TURBO	2.4 BASE	35.75 CACHE	165 TDP	OPTANE DC CO
6240Y	18 CORES	3.9 TURBO	2.6 base	24.75 CACHE	150 TDP	
4214Y	12 CORES	3.2 TURBO	2.2 BASE	16.5 cache	85 TDP	
NETWORKING	/NFV SI	PECIALIZ	ZED (IN	CL. INTEL® S	PEED SEL	.ECT TECH -BF)
6252N	24 CORES	3.6 TURBO	2.3 BASE	35.75 CACHE	150 TDP	
6230N	20 CORES	3.5 TURBO	2.3 base	27.5 cache	125 TDP	
5218N	16 CORES	3.9 TURBO	2.3 base	22 cache	105 TDP	
VM DENSITY \	/ALUE S	PECIALI	ZED			
6262V	24 CORES	3.6 TURBO	1.9 base	33 cache	135 TDP	GPTANE DC
6222V	20 CORES	3.6 тигво	1.8 base	27.5 cache	115 TDP	
LONG-LIFE CY	/CLE AN	D NEBS-	THERM	IAL FRIEND	LY	
6238T	22 CORES	3.7 тигво	1.9 base	30.25 CACHE	125 TDP	
6230T	20 CORES	3.9 тигво	2.1 base	27.5 cache	125 TDP	
5220T	18 CORES	3.9 Turbo	1.9 base	24.75 CACHE	105 TDP	
5218T	16 CORES	3.8 TURBO	2.1 BASE	22 CACHE	105 TDP	
4209T	8 CORES	3.2 TURBO	2.2 BASE	11 сасне	70 TDP	
SEARCH APP	LICATIO	IN VALU	E SPEC	IALIZED		
5220S	18 CORES	3.9 TURBO	2.7 base	24.75 CACHE	125 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 OR CONTACT YOUR INTEL REPRESENTATIVE TO OBTAIN THE LATEST INTEL PRODUCT SPECIFICATIONS. © COPYRIGHT 2019, INTEL CORPORATION.

INTEL PERSISTENT MEMORY RECOMMENDATIONS

	Populatio	Max Capacity			
Ratio	Population	DCPMM	DRAM	AD	MM
		512GB x 6	128GB x 6	3.77 TB	3 TB
4:1	2-2-2	256GB x 6	64GB x 6	1.84 TB	1.5 TB
		128GB x 6	32GB x 6	960 GB	768 GB
		512GB x 4	64GB x 6	2.43 TB	2 TB
5.3:1	2-2-1	256GB x 4	32GB x 6	1.22 TB	1 TB
		128GB x 4	16GB x 6	608 GB	512 GB
		512GB x 6	64GB x 6	3.38 TB	3 TB
8:1	2-2-2	256GB x 6	32GB x 6	1.69 TB	1.5 TB
		128GB x 6	16GB x 6	864 GB	768 GB
10 7.1	2 2 1	512GB x 4	32GB x 6	2.24 TB	2 TB
10.7.1	2-2-1	256GB x 4	16GB x 6	1.12 TB	1 TB
16.1	2_2_2	512GB x 6	32GB x 6	3.26 TB	3 TB
10:1	2-2-2	256GB x 6	16GB x 6	1.63 TB	1.5 TB



Sample Cascade Lake (CLX) Platinum/Gold Socket with Apache Pass (DCPMM)

- All Modes:
 - Max 1 DCPMM per channel
- Memory Mode (MM):
 - Min 1 DRAM DIMM + 1 DCPMM per IMC
 - 8:1; 5.3:1; 4:1 DCPMM to DRAM populated
 - DCPMM total capacity counted ONLY. DRAM used as cache.

- App Direct Mode (AD):
 - Min 1 DCPMM anywhere on platform
 - DRAM and DCPMM additive for total capacity.



A STRONG MEMORY & STORAGE FUTURE

	TODAY	FUTURE		
Cintel XEON PLATINUM inside	2 ND GEN INTEL XEON SCALABLE (CASCADE LAKE)	COOPER LAKE / ICE LAKE	SAPPHIRE RAPIDS	FUTURE INTEL XEON PROCESSOR
Intel OPTANE DC S OPTANE DC S	APACHE PASS	BARLOW PASS	3 RD GEN DC PERSISTENT MEMORY	4 TH GEN DC PERSISTENT MEMORY
intel OPTANE DC Solid STATE DRIVE	Intel® SSD DC P4800X (COLDSTREAM)	ALDER STREAM	NEXT GENERATION	NEXT GENERATION
INTEL® 3D NAND SSD	INTEL [®] SSD DC P46XX/P45XX	CLIFFDALE-R/ARBORDALE + (96-L, 144-L)	NEXT GENERATION	NEXT GENERATION
	1 1 1			





USECASE-DATABASE



SAP HANA SAP



- Volatile data structures remain in DRAM
- Column Store Main moves to
 Persistent Memory
 - DIMM form-factor, replacing DRAM
 - Could be configured for each table, partition, or column
 - Loading of tables into memory at startup becomes obsolete
 - Lower TCO, larger capacity
- No changes to the persistence

SAP HANA CONTROLS WHAT IS PLACED IN PERSISTENT MEMORY AND WHAT REMAINS IN DRAM



ORACLE EXADATA : PERSISTENT MEMORY ACCELERATOR FOR OLTP



- Exadata Storage Servers will add Persistent Memory Accelerator in front of Flash memory
- **RDMA** bypasses the software stack, giving 10X faster access latency to remove Persistent Memory
- Persistent Memory mirrored across storage servers for faulttolerance
- Persistent memory used as a shared cache effectively increases its capacity 10X vs using it directly as expensive storage
- Log Writes will use RDMA to achieve super fast commits

10X LOWER LATENCY



Slide courtesy of Oracle

USECASE-STORAGE



NETAPP[®] MAX DATA

Application

NetApp MAX Data* with NetApp Max FS*



Storage Tier NetApp AFF A-Series* and NetApp AFF8000* Series All Flash Arrays

- NetApp MAX Data runs on servers equipped with 2nd Generation Intel[®] Xeon[®] Scalable processors and Intel Optane DC persistent memory
- NetApp MAX File System* (FS) for Optane and auto tiering
- Your applications don't require any changes (App Direct mode)



MAX RECOVERY HIGH-LEVEL ARCHITECTURE



- ONTAP® Ha PAIR LUN
- MAX Recovery enables memory-to-memory replication between the MAX Data Server and the MAX Recovery server.
- MAX Recovery enables recovery in minutes instead of hours.
- Four MAX Data servers can replicate to a single MAX Recovery server.



ORACLE*, INTEL® OPTANE™ DC PERSISTENT MEMORY, AND NETAPP MAX DATA* PERFORMANCE SUMMARY

IOPS HIGHER

Oracle 18c—75% Select



Oracle 18c—75% Select



DAOS: DISTRIBUTED ASYNCHRONOUS OBJECT STORAGE

A new open-source, highperformance storage software solution architected for DCPMM

- Small I/Os are stored in Intel Optane DC persistent memory
- Bulk I/Os go straight to the NVMe SSDs
- Built entirely in userspace





CUSTOMER SUCCESS STORY: EXASCALE HPC



"What excites me most about exascale systems like Aurora is the fact that we now have, in one platform and one environment, the ability to mix simulation and artificial intelligence. This idea of mixing simulation and data-intensive science will give us an unprecedented capability, and open doors in research which were inaccessible before, like cancer research, materials science, climate science, and cosmology."

Rick Stevens, associate laboratory director for computing, environment and life sciences at Argonne National Laboratory and professor of computer science at the University of Chicago



Customer: Argonne National Laboratory supports about 3,500 researchers with a billion-dollar budget each year and spearheads scientific research in disciplines like physics, chemistry, genomics and more. **Challenge**: Deploying Aurora, the first exascale supercomputer in the United States, represents an enormous undertaking. The complexity of next-generation research and engineering requires a system offering the prowess to tackle workloads involving massive data sets like advanced simulation and modeling, artificial intelligence and data science. Solution: Working closely with Intel and Cray, the Argonne team collaborated to design and implement Aurora's exascale architecture. Scheduled for deployment in 2021, Aurora will feature future Generation Intel® Xeon® Scalable processors, Intel's X^e compute architecture, future Intel® Optane[™] DC persistent memory and Intel® One API.



USECASE-VIRTUALIZATION



WHEN IS VM MEMORY EXPANSION A GOOD FIT FOR YOUR CUSTOMER?





VMWARE ESXI VMMARK FOR INCREMENTAL MEMORY

WORKLOAD – INCREASE VMs PER NODE



Performance results are based on testing as of 4/02/2019 and may not reflect all publicly available security updates. See configuration disclosure for details. No product can be absolutely secure. For more complete information about performance and benchmark results, visit www.intel.com/benchmarks. Configuration: See <u>VMWare VMMark config slide</u>



A COMPLETE HIERARCHY

DRAM PERSISTENT MEMORY (intel) OPTANE DC ()) PERSISTENT MEMORY

COMPUTE Cache

IN PACKAGE MEMORY

Intel OPTANE DC STORAGE

NAND SSD

Intel[®] 3D NAND

HDD-TAPE

INTEL® OPTANETM PERSISTENT MEMORY Brings more data into memory

INTEL[®] OPTANE[™] SSDs Bring storage closer to the processor

> INTEL[®] QLC 3D NAND SSDS brings more data into solid state storage



LEGAL DISCLAIMER

All information provided here is subject to change without notice.

The products described in this document 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.

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 intel.com.

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 <u>www.intel.com/benchmarks</u>.

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 data. You should review this content, consult other sources, and confirm whether referenced data are accurate.

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.

© Intel Corporation. Intel, the Intel logo, and other Intel marks are trademarks of Intel Corporation or its subsidiaries. Other names and brands may be claimed as the property of others.



