Persistent Memory の最新技術動向紹介

今村 智史 ǂ （富士通株式会社）
柿原 俊男 † （HGST Japan）
SNIA-J 次世代不揮発メモリ分科会
† 会長、ǂ 副会長

自己紹介（今村 智史）

- 所属：富士通株式会社 ICTシステム研究所
  - メモリ／ストレージシステムに関する研究に従事
  - PMEM の性能分析やユースケース検討
- 専門分野：コンピュータアーキテクチャ
  - OS（特に仮想メモリ周り）も得意です
- 出身：九州大学（博士（工学）取得）
- **SNIA-J 次世代不揮発メモリ分科会 副会長**

はじめに

- 本講演では Persistent Memory 製品の性能評価結果をご紹介します

- 今回の性能評価実験はすべて柿原さん（HGST Japan）に実施していただき、その実験結果をご提供いただきました

目次

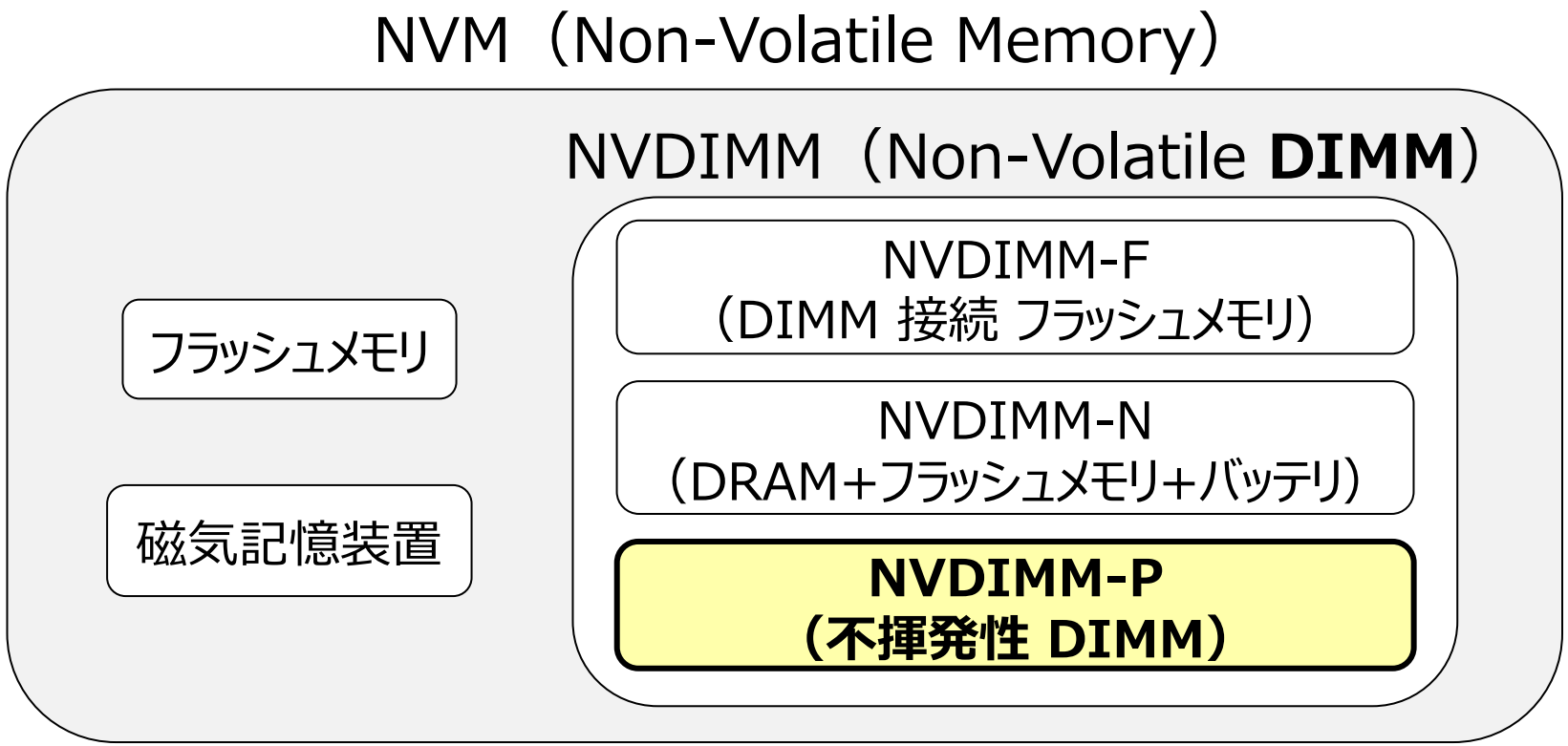
- Persistent Memory（PMEM）とは
- ２世代の PMEM 製品
- 性能評価

Persistent Memory（PMEM）とは

Non-Volatile DIMM（NVDIMM）

“**ロード／ストア**プログラミングモデルに
適した性能特性を持つ**ストレージ**テクノロジ”

出典：SNIA NVM Programming Model Version 1.2



Persistent Memory（PMEM）

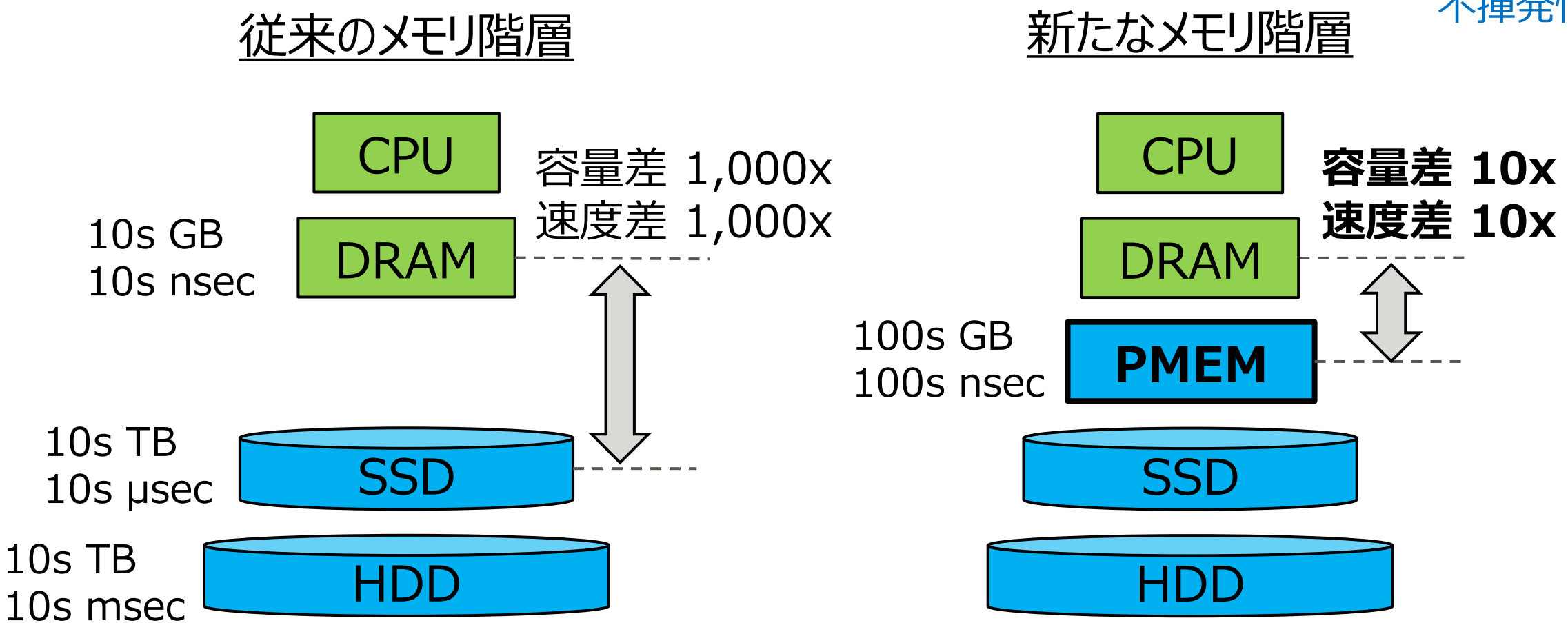
出典：Intel

出典：Lenovo

- 世界初の NVDIMM-P 製品
  - **3D XPoint™ メモリ**技術を採用
- DIMM 当たり最大 **512 GB**
  - CPU あたり最大 **4 TB**
- DIMM スロットに接続
- PMEM 対応の CPU が必要

PMEM の狙い

揮発性
不揮発性



PMEM の特徴

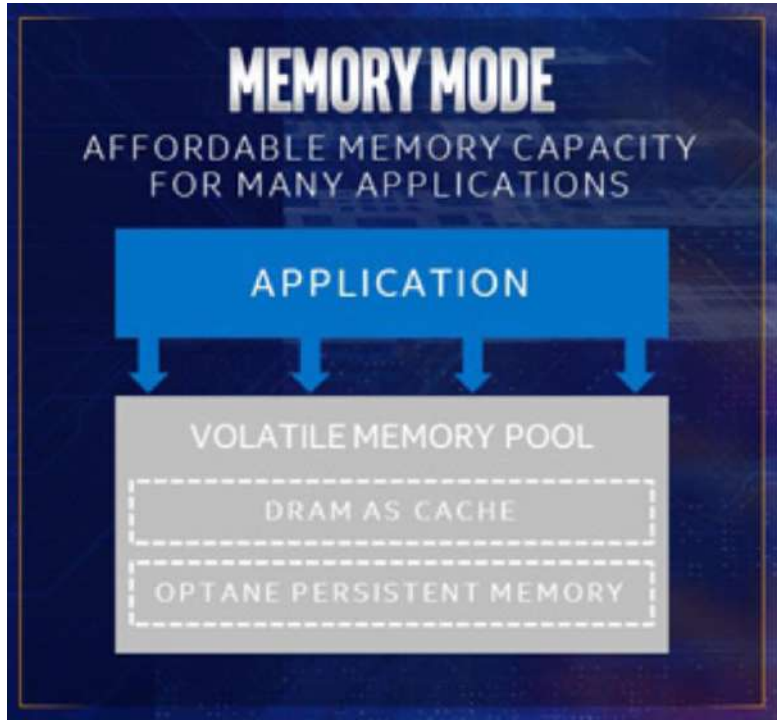
利点

- DRAM に比べ容量単価が安く大容量
- バイト単位のロード／ストアアクセスが可能
- 不揮発性を有する（電源遮断時でもデータを保持可能）
- SSD に比べ高性能かつ高書き込み耐性

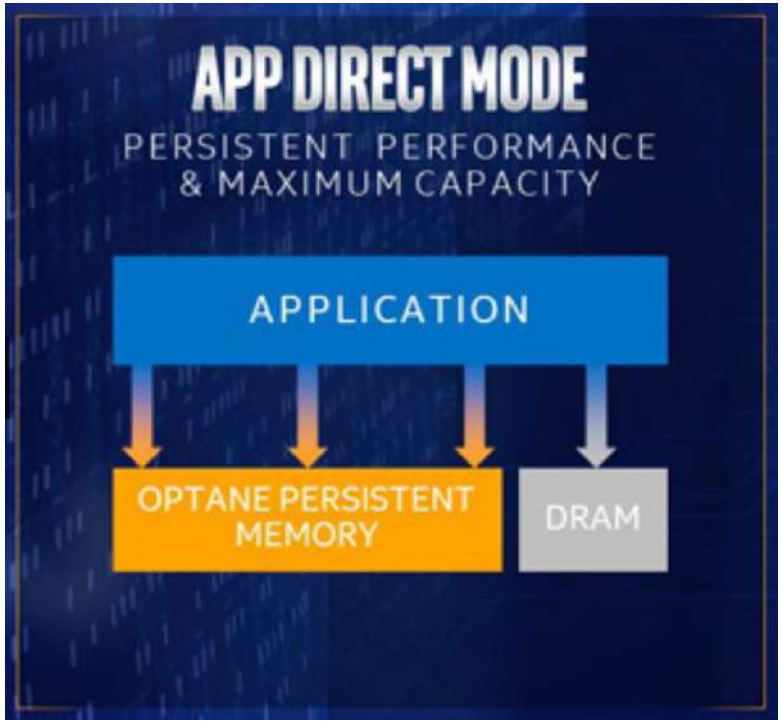
欠点

- DRAM に比べ低性能（特に書き込み性能が低い）
- DRAM より低い書き込み耐性

PMEM の２種類のモード



- PMEM を安価で大容量の**揮発性メモリ**として使用
- DRAM はキャッシュとして機能
- アプリの修正必要なし



出典：インテル

- DRAM と PMEM を使い分け可能
- PMEM の不揮発性を活用可能
  - PMEM プログラミングが必要

PMEM 活用方法の選択肢

選択肢	モード	メリット
DRAM 代替の揮発性メインメモリ	Memory	DRAM と同様に利用可能 （アプリの修正必要無し）
DRAM とは別個の揮発性メインメモリ	App Direct	DRAM/PMEM を容易に使い分け可能
永続ブロックストレージ	App Direct	• 従来のファイル API が使用可能 • データを永続化可能
不揮発性メインメモリ	App Direct	メインメモリ上でデータを永続化可能

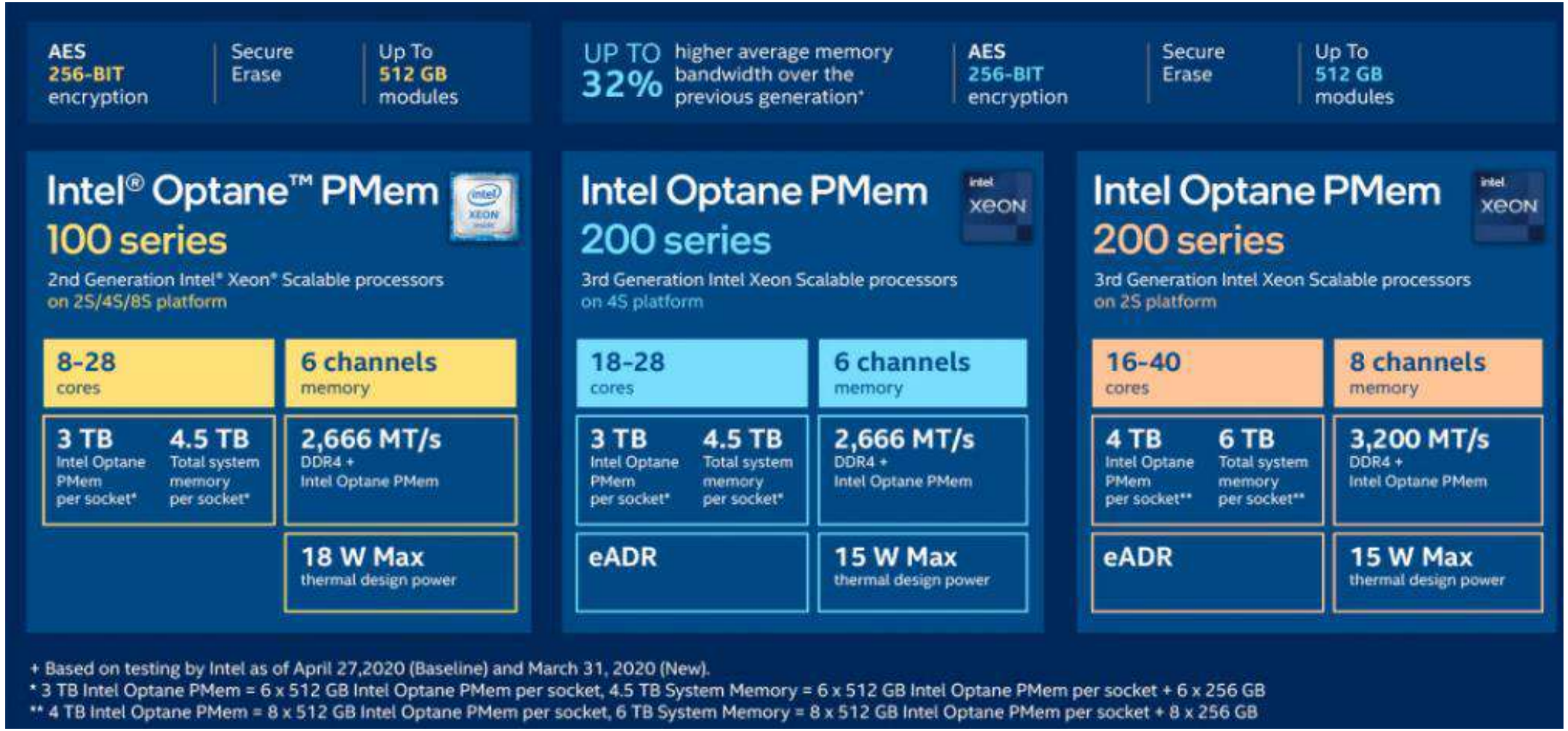
2 世代の PMEM 製品

PMEM 100 Series と PMEM 200 Series

- PMEM 100 Series
  - 第１世代の Optane™ Persistent Memory（旧 Optane™ DCPM）
  - コードネーム：**A**pache Pass
  - リリース時期：Q2’19

- **PMEM 200 Series**
  - 第２世代の Optane™ Persistent Memory
  - コードネーム：**B**arlow Pass
  - リリース時期：Q2’20

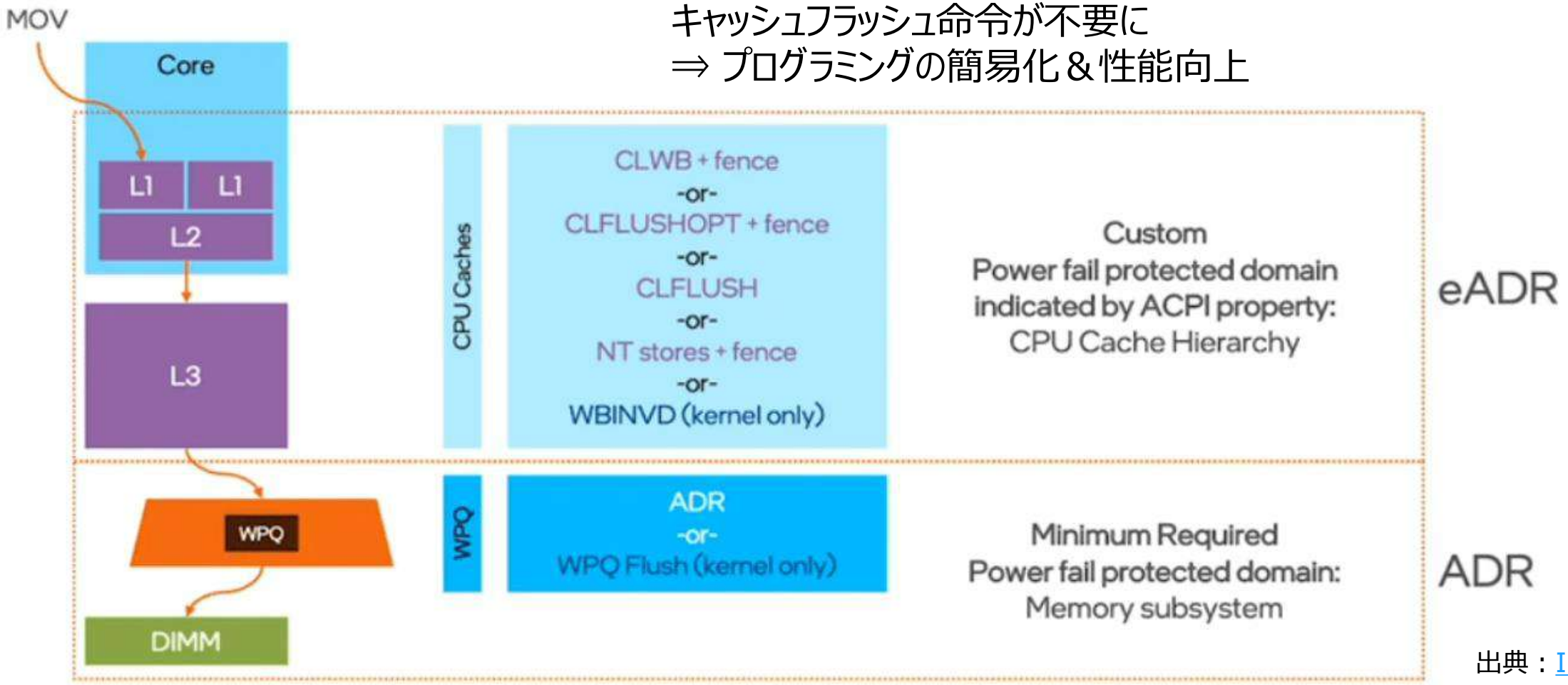
PMEM 100 Series と 200 Series の比較



出典：Intel

eADR（extended Asynchronous DRAM Refresh）

キャッシュフラッシュ命令が不要に
⇒ プログラミングの簡易化＆性能向上



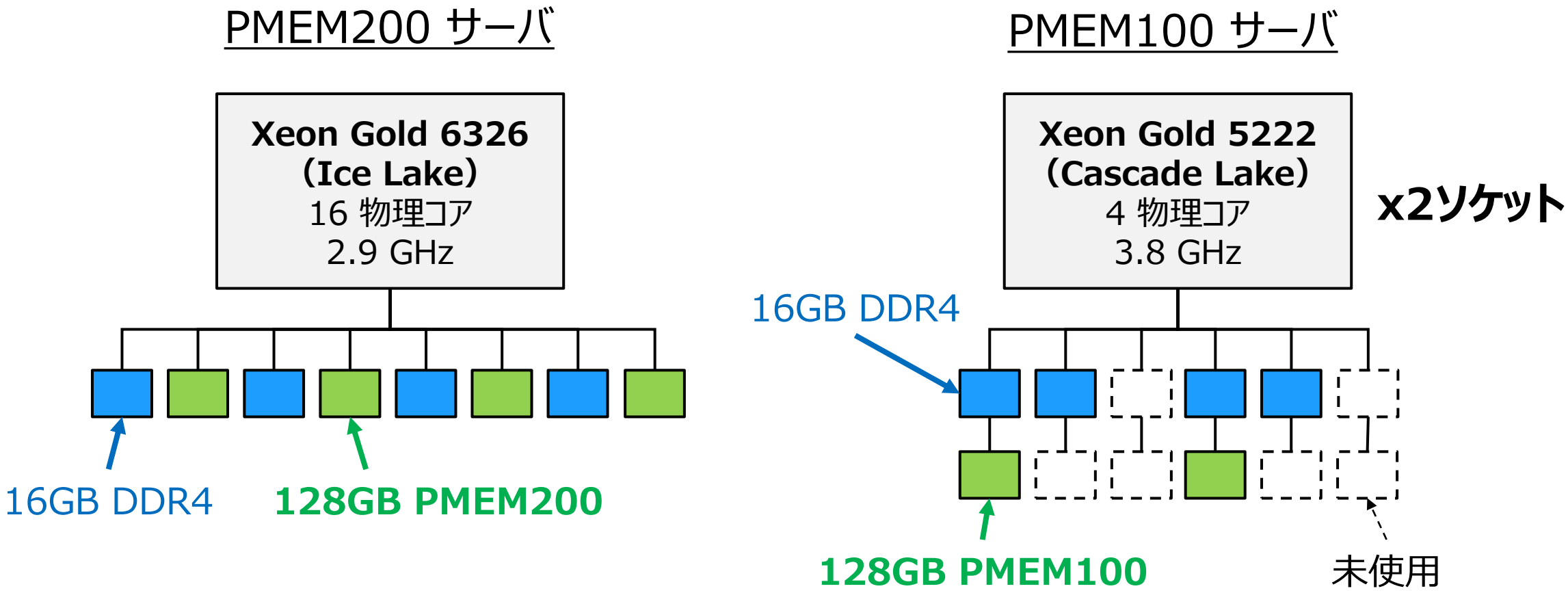
出典：Intel

性能評価

評価項目

- メインメモリとしての基礎性能
- ブロックデバイスとしての基礎性能
- Spec CPU 2017 ベンチマークの性能
- インメモリキーバリューストアの性能

評価環境



KMEM DAX

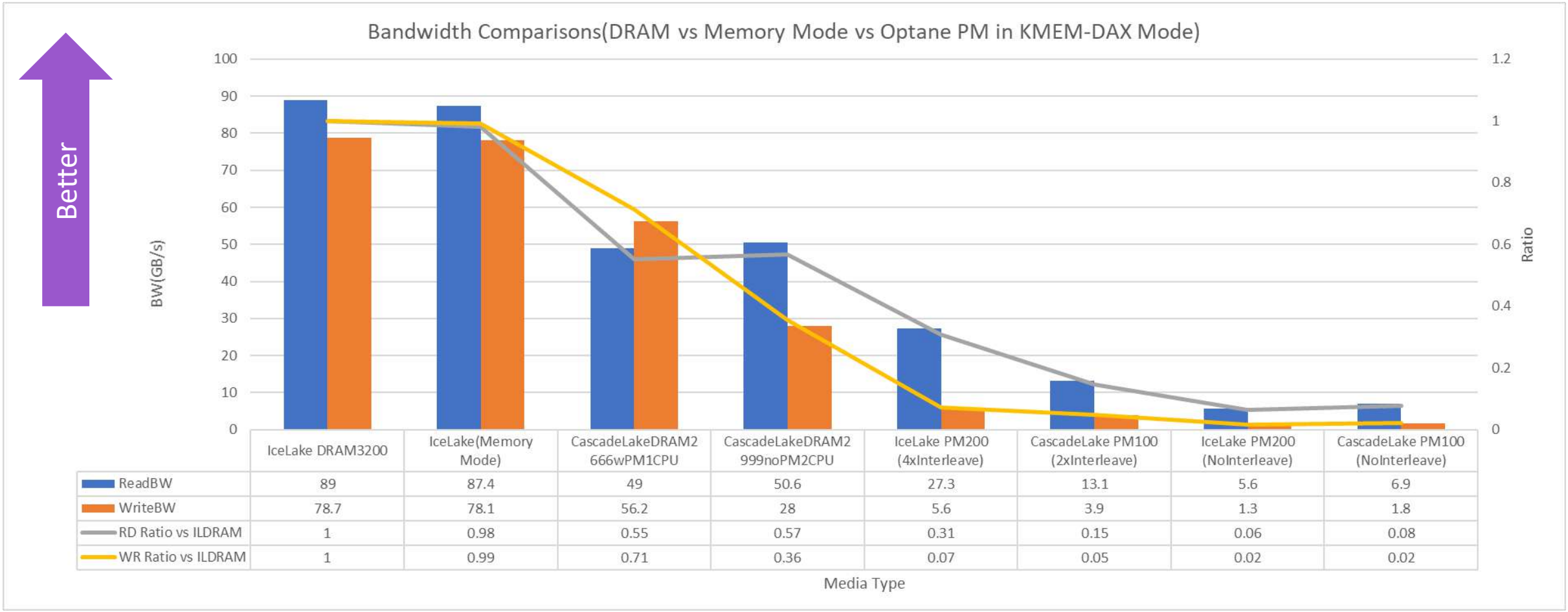
- PMEM を**揮発性メモリ**として活用するための機能
- PMEM のメモリ空間を DRAM とは別の NUMA ノードとして見せる
  - アプリを修正することなく DRAM と PMEM を使い分け可能
- Linux kernel version 5.1 からサポート

```
[Ice Lake]# numactl -H
available: 2 nodes (0-1)
node 0 cpus: 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
19 20 21 22 23 24 25 26 27 28 29 30 31
node 0 size: 63930 MB
node 0 free: 51967 MB
node 1 cpus:
node 1 size: 507904 MB
node 1 free: 507904 MB
node distances:
node   0   1
  0:  10  17
  1:  17  10
```

DRAM

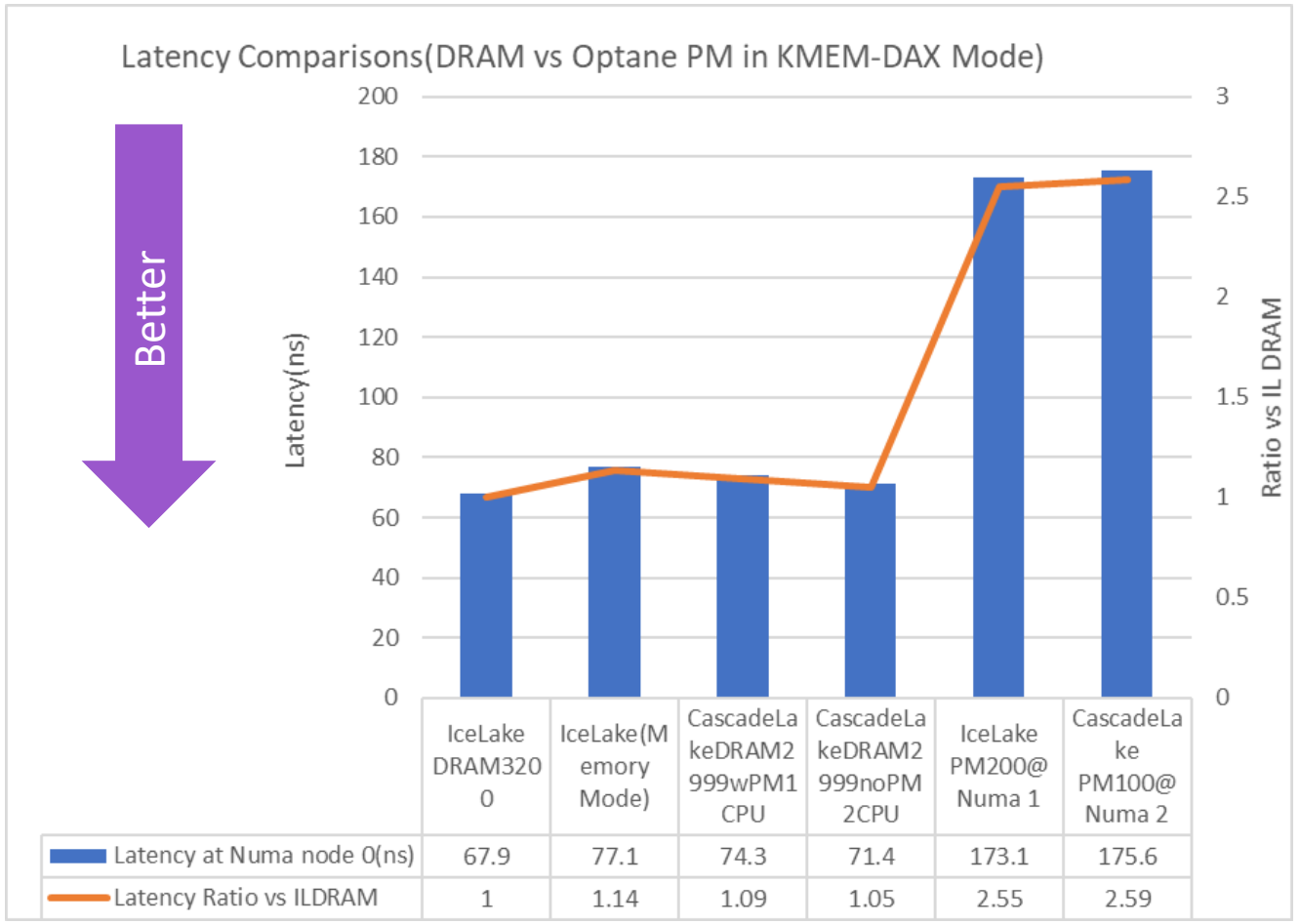
KMEM-DAX (PMEM)

メモリメモリとしてのバンド幅

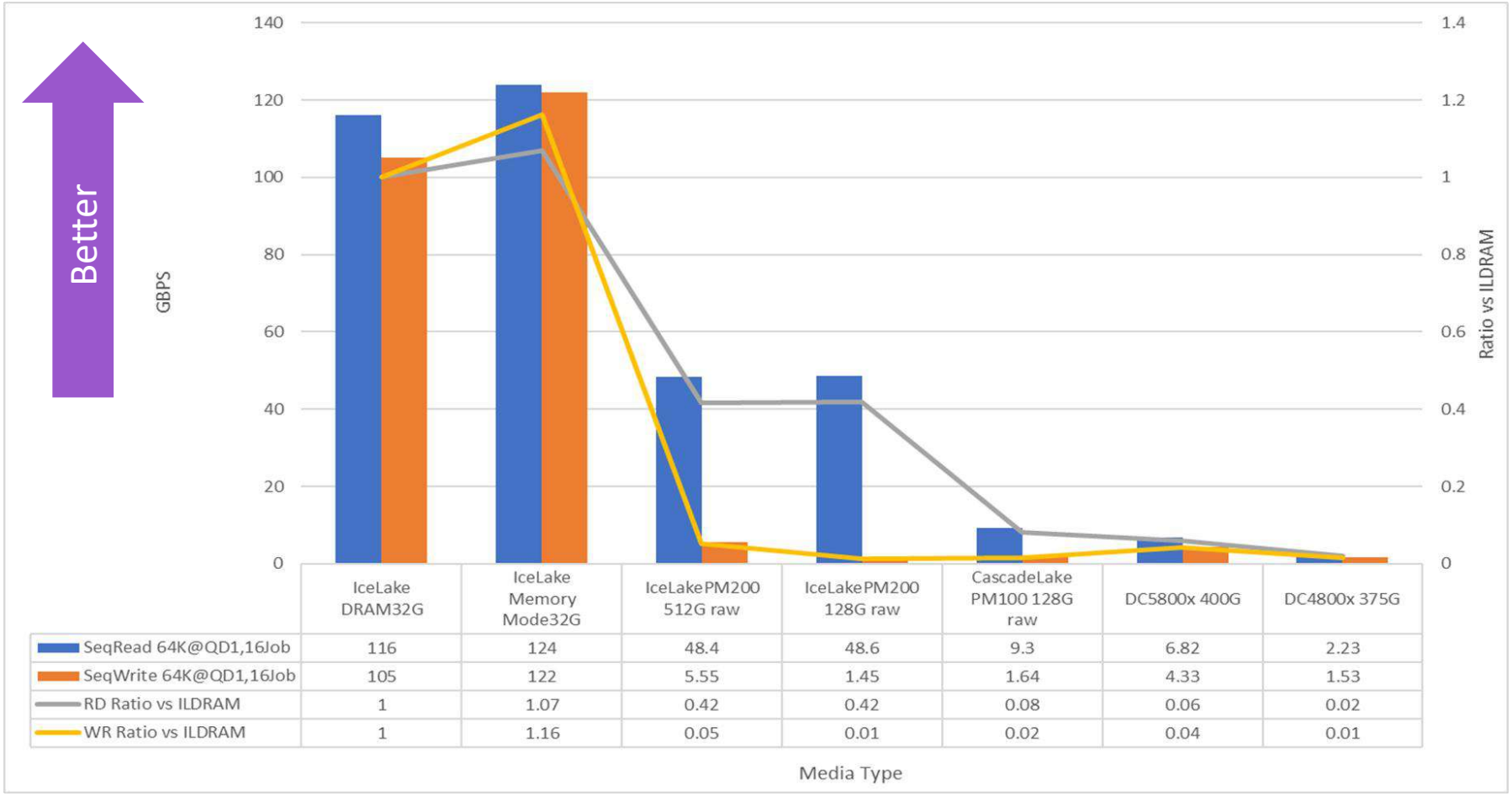


	IceLake DRAM3200	IceLake(Memory Mode)	CascadeLakeDRAM2 666wPM1CPU	CascadeLakeDRAM2 999noPM2CPU	IceLake PM200 (4xInterleave)	CascadeLake PM100 (2xInterleave)	IceLake PM200 (NoInterleave)	CascadeLake PM100 (NoInterleave)
ReadBW	89	87.4	49	50.6	27.3	13.1	5.6	6.9
WriteBW	78.7	78.1	56.2	28	5.6	3.9	1.3	1.8
RD Ratio vs ILDRAM	1	0.98	0.55	0.57	0.31	0.15	0.06	0.08
WR Ratio vs ILDRAM	1	0.99	0.71	0.36	0.07	0.05	0.02	0.02

メインメモリとしてのアクセスレイテンシ

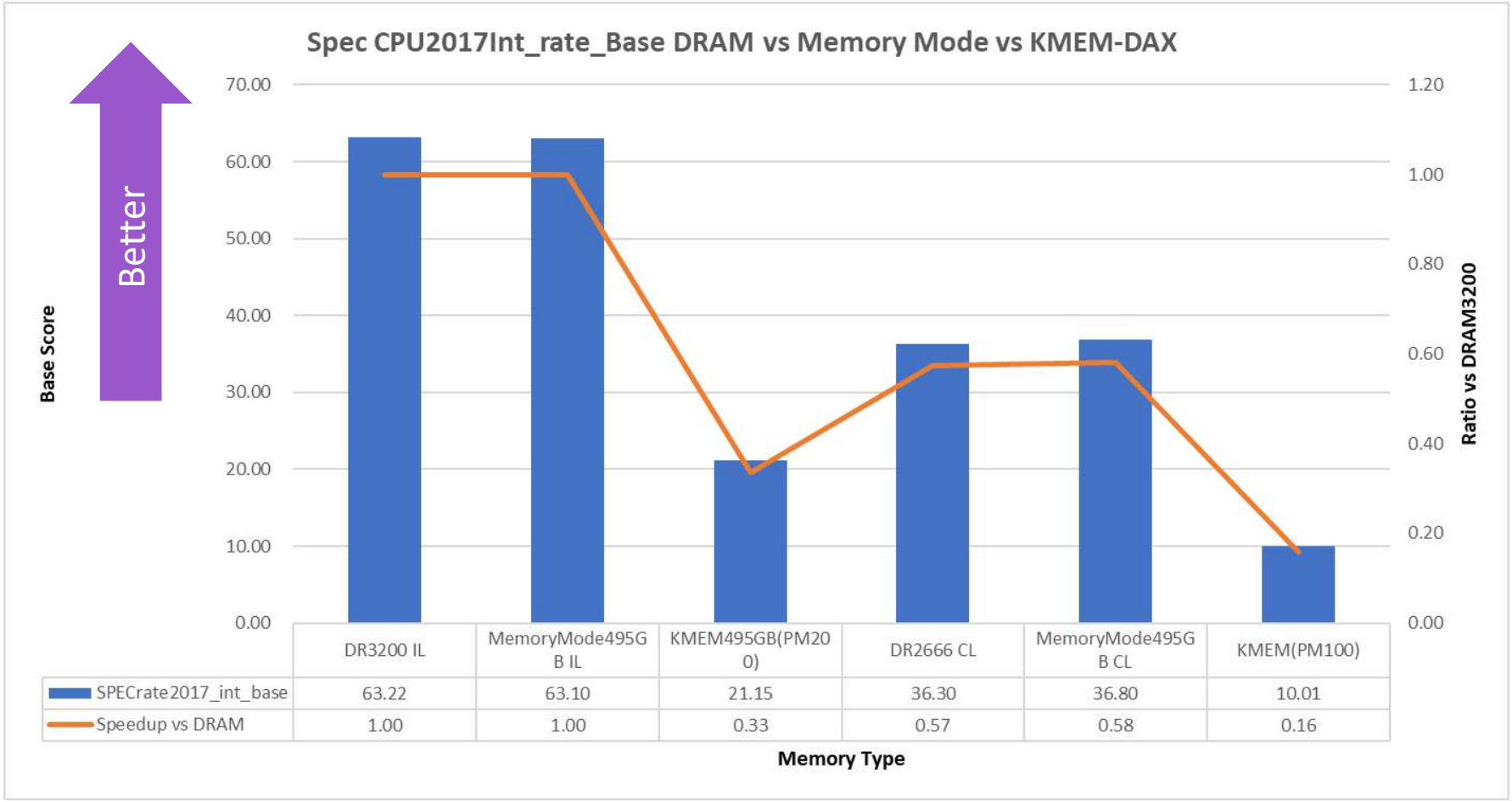


ブロックデバイスとしてのバンド幅

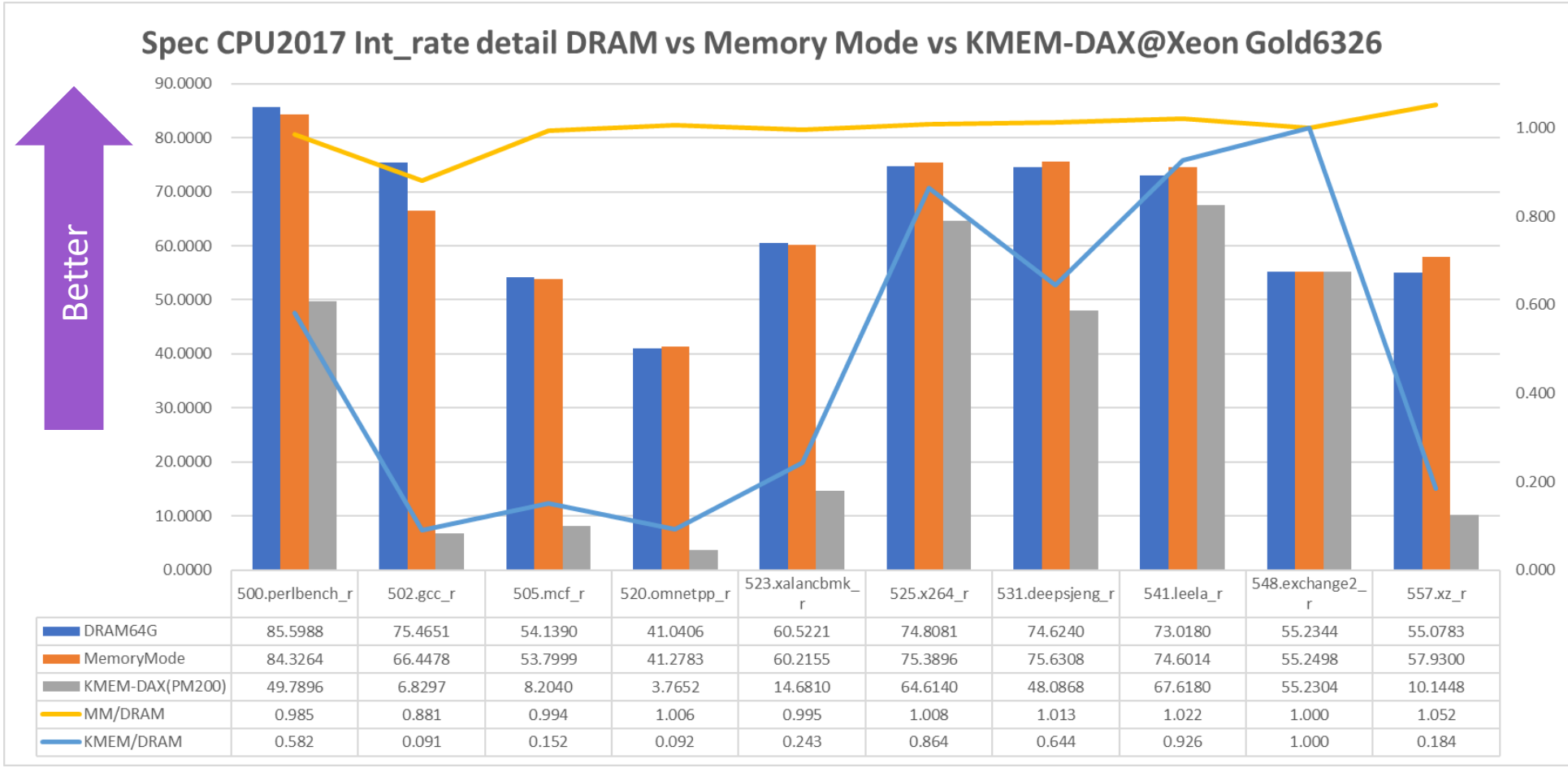


	IceLake DRAM32G	IceLake Memory Mode32G	IceLakePM200 512G raw	IceLakePM200 128G raw	CascadeLake PM100 128G raw	DC5800x 400G	DC4800x 375G
SeqRead 64K@QD1,16Job	116	124	48.4	48.6	9.3	6.82	2.23
SeqWrite 64K@QD1,16Job	105	122	5.55	1.45	1.64	4.33	1.53
RD Ratio vs ILDRAM	1	1.07	0.42	0.42	0.08	0.06	0.02
WR Ratio vs ILDRAM	1	1.16	0.05	0.01	0.02	0.04	0.01

SPEC CPU® 2017 ベンチマーク int_rate

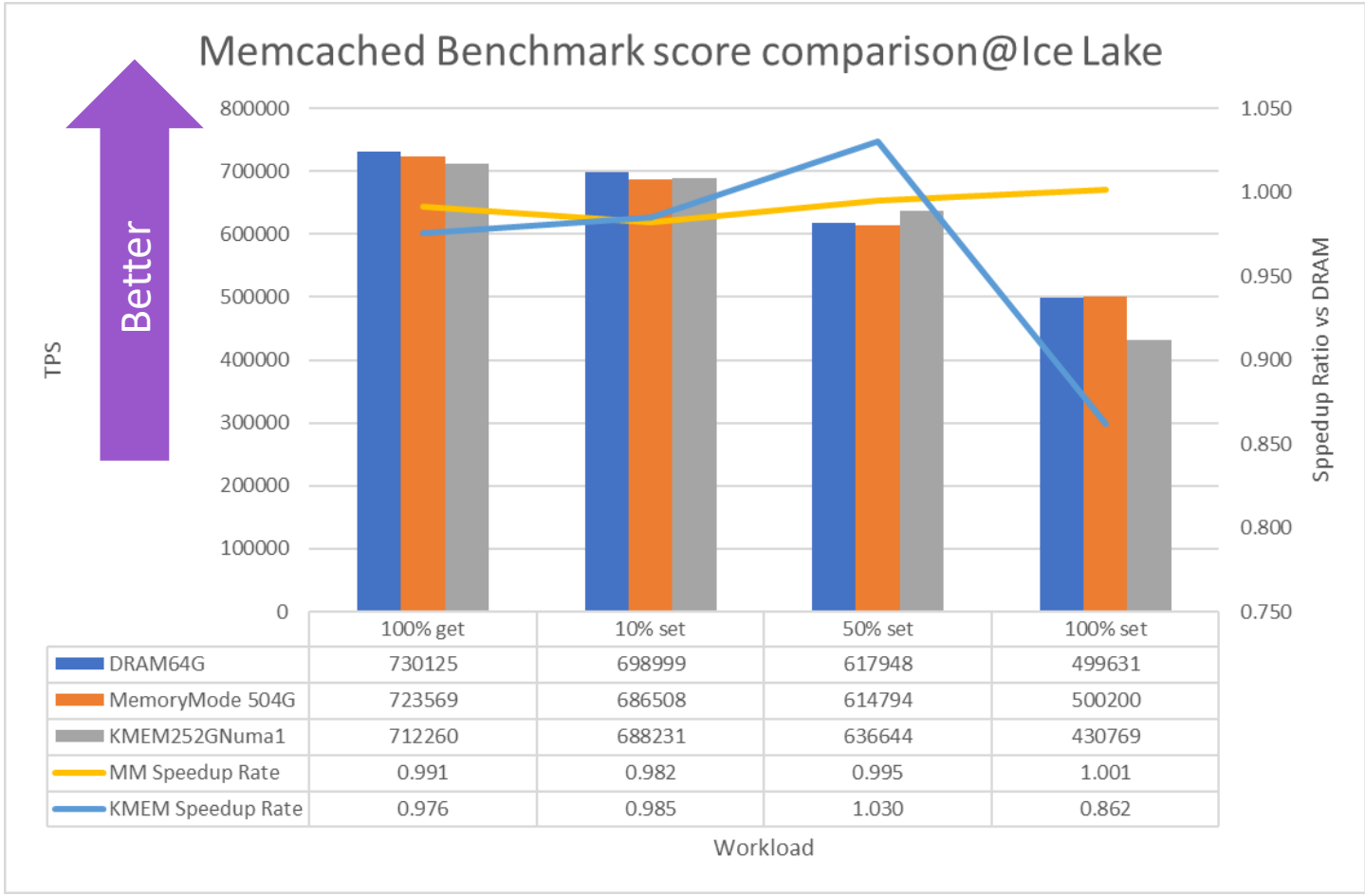


SPEC CPU® 2017 int_rate（ベンチマーク毎の詳細）



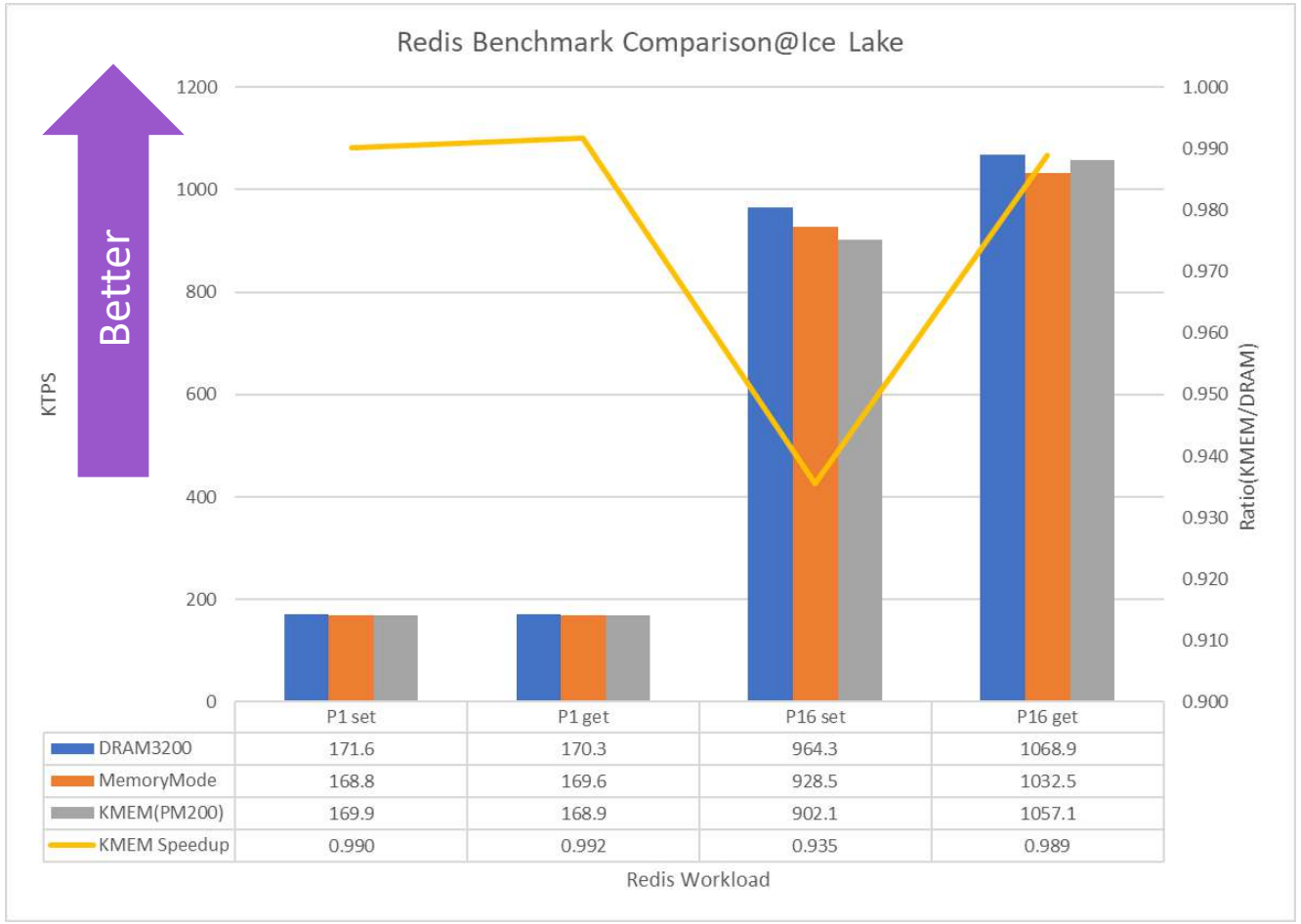
	500.perlbench_r	502.gcc_r	505.mcf_r	520.omnetpp_r	523.xalancbmk_r	525.x264_r	531.deepsjeng_r	541.leela_r	548.exchange2_r	557.xz_r
DRAM64G	85.5988	75.4651	54.1390	41.0406	60.5221	74.8081	74.6240	73.0180	55.2344	55.0783
MemoryMode	84.3264	66.4478	53.7999	41.2783	60.2155	75.3896	75.6308	74.6014	55.2498	57.9300
KMEM-DAX(PM200)	49.7896	6.8297	8.2040	3.7652	14.6810	64.6140	48.0868	67.6180	55.2304	10.1448
MM/DRAM	0.985	0.881	0.994	1.006	0.995	1.008	1.013	1.022	1.000	1.052
KMEM/DRAM	0.582	0.091	0.152	0.092	0.243	0.864	0.644	0.926	1.000	0.184

Memcached（インメモリキーバリューストア）



	100% get	10% set	50% set	100% set
DRAM64G	730125	698999	617948	499631
MemoryMode 504G	723569	686508	614794	500200
KMEM252GNuma1	712260	688231	636644	430769
MM Speedup Rate	0.991	0.982	0.995	1.001
KMEM Speedup Rate	0.976	0.985	1.030	0.862

Redis（インメモリキーバリューストア）



	P1 set	P1 get	P16 set	P16 get
DRAM3200	171.6	170.3	964.3	1068.9
MemoryMode	168.8	169.6	928.5	1032.5
KMEM(PM200)	169.9	168.9	902.1	1057.1
KMEM Speedup	0.990	0.992	0.935	0.989

まとめ

- PMEM とは「バイトアクセス可能な DIMM タイプの不揮発性メモリデバイス」
- 100 series から 200 series へ性能面・機能面で進化している
- メインメモリ、ブロックストレージ、不揮発性メインメモリとして活用可能
- 基礎性能は DRAM に劣る
- ただし、実アプリに対しては DRAM との性能差は意外と見えない
  ⇒ 大容量・低バイト単価という面は DRAM に対してアドバンテージ
  - 実際に仮想化基盤への導入事例なども出てきた（ソフトバンク、NEC）

Q&A

