Lenovo

Analyzing the Performance of Intel Optane Persistent Memory 200 Series in Memory Mode with Lenovo ThinkSystem Servers

Explains the operation and target usage of the Intel Optane Persistent Memory 200 Series in Memory Mode

Compares performance between supported PMem memory configurations

Discusses performance behavior of the Intel Optane Persistent Memory 200 Series in Memory Mode

Provides tips on how to improve application performance with Persistent Memory in Memory Mode

Nathan Pham

Kevin Huang

LENOVO PRESS

Abstract

The conventional system memory tiering consists of several levels of processor caches, a main memory tier with DDR memory DIMMs, and a system storage tier. In this tiering model, there is a large gap in performance and cost between the main memory tier and system storage tier. Intel Optane Persistent Memory (PMem) was brought to the market to help address this gap.

Intel Optane Persistent Memory 200 Series supports two modes, Memory Mode and App Direct Mode. In Memory Mode, the regular DRAM system memory acts as the next level cache, and the PMem becomes the system main memory. This enables the system to have higher system memory capacity and/or lower TCO. It doesn't require any changes to applications.

In this performance brief, we first explain the operations and the important performance metrics of the PMem 200 Series in Memory Mode. We will then discuss and explain the performance comparison between a few supported memory configurations with PMem. We will end the discussion with a few tips on how to configure system memory with PMem in Memory Mode to maximize system memory and application performance.

This performance brief is for customers, business partners and sellers who wish to have a better understanding on the performance behavior and to learn how to better configure the memory subsystem for performance with Intel Optane Persistent Memory 200 Series in Memory Mode. It is expected that the reader will have a basic understanding on system memory hierarchy, data cache operations, and memory performance.

At Lenovo® Press, we bring together experts to produce technical publications around topics of importance to you, providing information and best practices for using Lenovo products and solutions to solve IT challenges.

See a list of our most recent publications at http://lenovopress.com.

Contents

Introduction

Intel Optane Persistent Memory 200 Series is the latest Persistent Memory (PMem) generation of the Intel Optane Persistent Memory product line. It is supported by Intel 3rd Generation Xeon Scalable platform. It supports up to 3200 MT/s speed and comes in three capacities: 128GB, 256GB, and 512GB.

The following operating modes are supported with the Intel Optane Persistent Memory 200 Series (formerly codenamed Barlow Pass or BPS):

- Memory Mode – In this mode, DRAM acts as the next level cache (L4), and PMem is now the system main memory. This enables the system to have higher system memory capacity and/or lower TCO. This mode requires no changes to applications.
- App Direct Mode – In this mode, both DRAM and PMem are part of main memory. Data stored in PMem enables fast recovery from power loss. Operating System and Virtual Machine are required to be PMem aware. Storage over App Direct feature is also supported in App Direct mode.

This paper describes the use and performance benefits of Memory Mode.

Table 1 provides the comparison between Intel Optane PMem 200 Series and Intel Optane PMem 100 Series.

Table 1 Comparison between Intel Optane persistent memory series

Feature	PMem 100 Series (Apache Pass)	PMem 200 Series (Barlow Pass)
Platform Capacity	Up to 3 TB per socket	Up to 4 TB per socket
Memory Channel	Up to 6 channels per socket	Up to 8 channels per socket
DDR-T Speed	Up to 2666 MT/s	Up to 3200 MT/s
PMem Capacities	128, 256, 512GB	128, 256, 512GB
Media Controller	Elk Valley	Barlow Valley[a]
Sustained TDP	18 W	15 W
Data Persistence	ADR	ADR, eADR[b]
Intel Memory BW Boost	Not available	Smart boost to max TDP[c]

a. The new media controller improves memory bandwidth performance per channel.
b. Improves application performance by avoiding CPU cache flush commands at run time.
c. A new feature that temporarily boost to maximum TDP (18W) to get average 25% additional bandwidth, if thermal headroom is available.

For more details on Intel Optane Persistent Memory 200 series, refer to the Lenovo Press product guide, available from:

https://lenovopress.com/lp1380-intel-optane-persistent-memory-200-series

Memory Population with PMem

There are strict rules for configuring servers with PMem. For Memory Mode, some of the rules are:

- At most 1 PMem DIMM can be populated per DDR channel.
- At least 1 PMem DIMM must be populated per CPU socket.
- All CPU sockets must have identical PMem populations.
- Mixing PMem of different series anywhere in the same server is not supported.

For a more complete set of rules and guidelines for populating memory with PMem in different operating modes, please refer to the product guide for your specific Lenovo ThinkSystem™ server.

DRAM and PMem size ratios

The memory size ratio between DRAM and PMem plays an important role in the overall application performance. As shown in the subsequent sections, a higher DRAM:PMem ratio results in a lower near-memory cache miss rate, and better overall memory and application performance. Intel recommends a ratio between 1:4 and 1:16 (DRAM:PMem system capacity ratio) for Memory Mode.

Table 2 lists DRAM:PMem capacity ratios for all supported DRAM, PMem, and memory populations.

Table 2 DRAM:PMem capacity ratios for all supported memory populations in Memory Mode

Population DRAM + BPS PMem	DRAM Capacity	BPS PMem Capacity		
		128 GB	256 GB	512 GB
8 + 8 / 4 + 4	16 GB	1 : 8	1 : 16	-
	32 GB	1 : 4	1 : 8	1 : 16
	64 GB	-	1 : 4	1 : 8
	128 GB	-	-	1 : 4
8 + 4	16 GB	1 : 4	1 : 8	1 : 16
	32 GB	-	1 : 4	1 : 8
	64 GB	-	-	1 : 4

Performance evaluation test methodology and setup

All performance data discussed in this paper were measured on the ThinkSystem SR650 V2 server with 3rd Gen Intel Xeon Scalable processors. Table 3 shows the system configurations for different memory population options used in the performance evaluation.

Table 3 System configurations for different PMem population options

Configuration	All DRAM	8 + 8	8 + 4	4 + 4
Processor	Intel Xeon Platinum 8380 Processor	Intel Xeon Platinum 8380 Processor	Intel Xeon Platinum 8380 Processor	Intel Xeon Platinum 8380 Processor
DRAM	8 x 32GB RDIMM	8 x 32GB RDIMM	8 x 32GB RDIMM	4 x 32GB RDIMM
BPS PMem	None	8 x 256GB BPS PMem	4 x 256GB BPS PMem	4 x 256GB BPS PMem
Ratio	N/A	1:8	1:4	1:8
OS	RHEL 8.3	RHEL 8.3	RHEL 8.3	RHEL 8.3

In Memory Mode, DRAM acts as the next level cache (L4), and PMem serves as the system main memory. They are also referred to as *near-memory* (DRAM) and *far-memory* (PMem). When there is a miss on the requested data-line in the processor L3 cache, the memory subsystem will look for the data-line in the next level cache L4 (near-memory).

If near-memory has a copy of the requested data-line, the data-line will be copied to the L3 cache, and the memory data request operation is completed. This is a hit to near-memory.

However, if the near-memory does not have a copy of the requested data-line (miss to near-memory), the memory subsystem will get the data-line from the main memory (PMem) for the data request. A miss to near-memory results in longer latency for the requested data-line. For this reason, the near-memory miss rate is an important performance metric for PMem configurations in Memory Mode.

We will discuss the impact of near-memory miss rate throughout the paper.

Intel Memory Latency Checker (MLC) was used as the load generating and measuring tool. All performance tests were done on only one socket (socket 0). Other Intel NDA tool was used to measure latency and near-memory (L4) miss rate. This can also be done using the publicly available Intel VTune Platform Profiler.

The following memory access patterns were used for the test:

- All Read Sequential – all accesses were memory reads, in sequential memory address order.
- All Read Random – all accesses were memory reads, in random memory address order.
- 2R1W Sequential – memory read:write ratio is 2:1, in sequential memory address order.
- 2R1W Random – memory read:write ratio is 2:1, in random memory address order.

Special test setups were utilized to mimic real world usage and to ensure consistency in the test results. These set ups include thread binding, memory buffer allocations, and memory scrambling and/or pre-conditioning. This resulted in relatively higher near-memory miss rates and lower bandwidth measurements even when the memory footprint is small enough to fit in the near-memory, or L4 cache.

The following memory configurations were used for performance evaluation:

- All_DRAM – The server was populated with 8x32GB RDIMMs. All DDR channels were populated with one 32GB RDIMM per channel.
- 8+8 – The server was populated with 8x32GB RDIMMs + 8x256GB BPS PMem. Each DDR channel was populated with one 32GB RDIMM and one 256GB BPS PMem.
- 8+4 – The server was populated with 8x32GB RDIMMs + 4x256GB BPS PMem. Four DDR channels were populated with one 32GB RDIMMs + one 256GB BPS PMem, and four other channels were populated with just one 32GB RDIMMs.
- 4+4 – The server was populated with 4x32GB RDIMMs + 4x256GB BPS PMem. Four DDR channels were populated with one 32GB RDIMM, and the other four channels were populated with one 256GB BPS PMem.

Figure 1 illustrates the memory configurations described above.

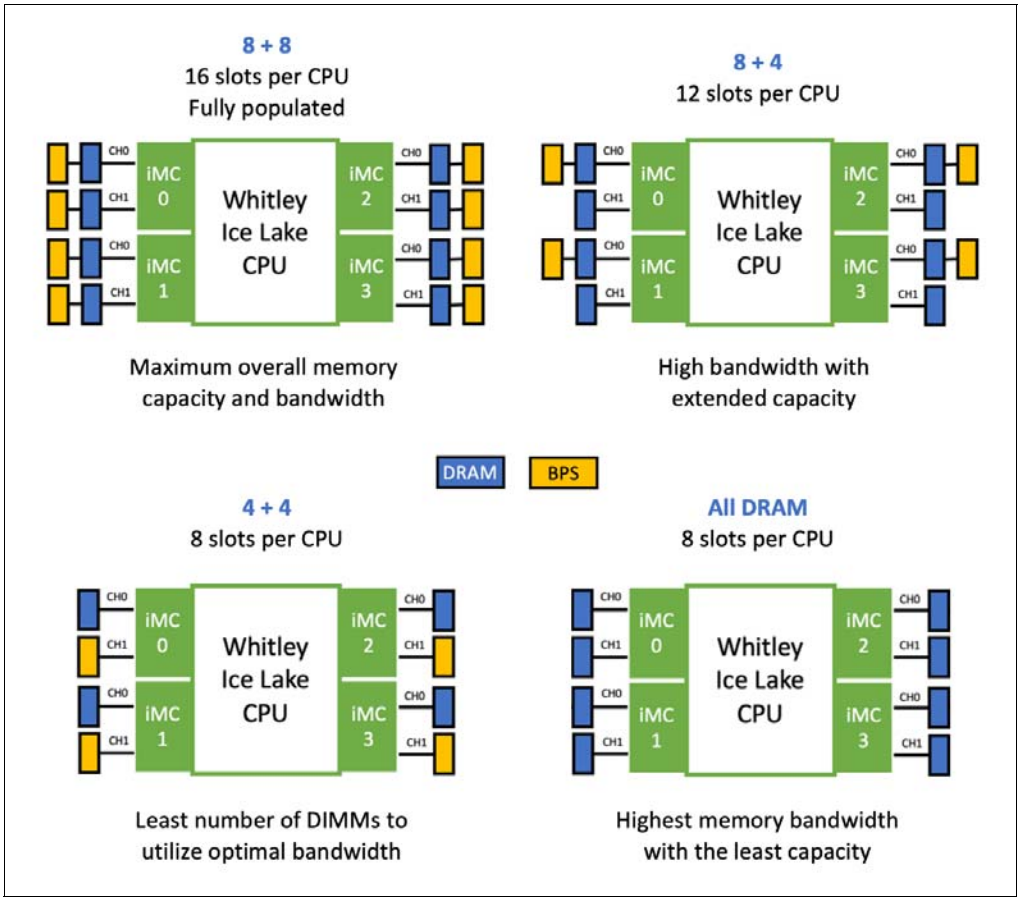

Figure 1 Memory configurations used in performance measurements

Memory bandwidth performance

Memory bandwidth is one of the two key performance metrics for memory sub-system. The other is Memory latency as described in “Memory latency performance” on page 10.

For servers with PMem configurations in Memory Mode, the effective memory bandwidth measures the data transfer rate between near memory and processor L3 cache. When there is a miss to near-memory, there will be data transfer between far-memory (PMem) and

near-memory (DRAM). Both of these data flows share the same DDR buses, but only data transfer between near-memory and processor L3 cache will be considered as effective memory bandwidth.

This section discusses effective memory bandwidth behavior for several supported PMem configurations in Memory Mode.

All-read sequential memory bandwidth performance

Figure 2 shows memory bandwidth performance comparison for all memory configurations with All Read Sequential memory access pattern.

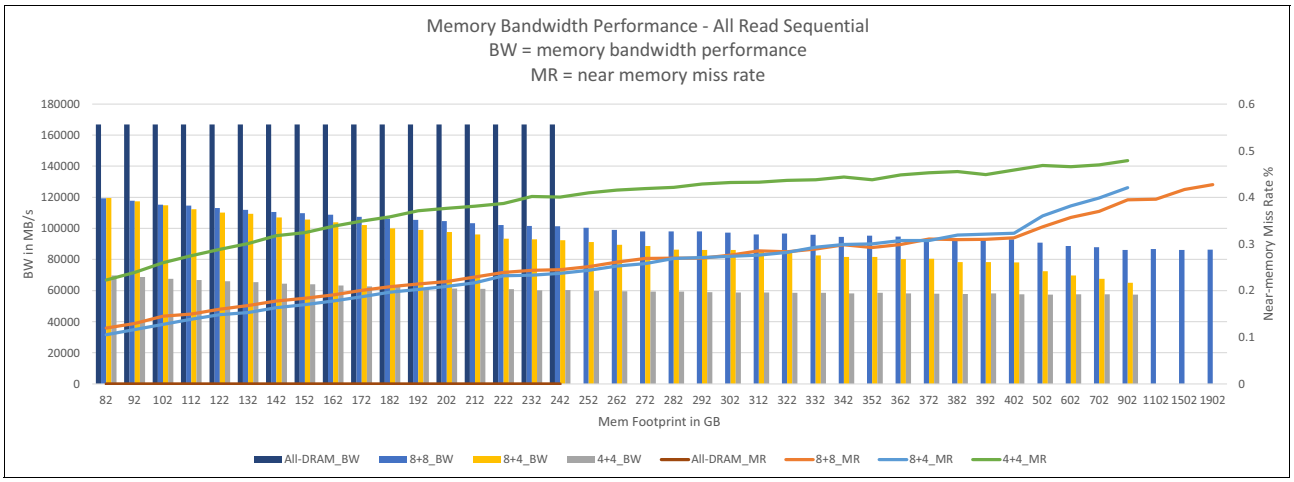

Figure 2 Memory bandwidth performance (higher is better) – all read sequential

The bar chart with the left y-axis has the memory bandwidth performance measured in MB/s. The line chart with the right y-axis has the near-memory miss rate. The x-axis shows the memory footprint in GB used for the test. As the memory footprint increases, the near-memory cache miss rate also increases

As the chart indicates, memory bandwidth performance is best with All_DRAM option. In this configuration, all DDR channels were populated with the same DIMM configuration (1DPC), and the server achieves highest level of memory bandwidth performance.

The second-best configuration for memory bandwidth performance is the 8+8 configuration. Even at a small memory footprint, data request cannot always find the data-line in near-memory cache. When this happens, data-line is copied from far-memory (PMem) to near-memory (DRAM) before moving to the processor caches (L3/L2/L1) and eventually the requested core to complete the transaction.

Both near-memory and far-memory share the same DDR buses, and only data traffic moving from near-memory (DRAM) to the requested core's L3 is counted as effective system memory bandwidth. The more near-memory cache misses there is, the more data movement there is between PMem and DRAM on the DDR buses, and the less DDR bandwidth is available for data movement between DRAM and L3. In other words, the near-memory cache miss rate has a negative impact on effective memory bandwidth performance. As the near-memory cache miss rate increases, effective memory bandwidth decreases.

For all memory configurations with PMem in Memory Mode, as the memory footprint of the application increases, the near-memory cache miss rate also increases, and memory bandwidth performance decreases.

Both the 8+8 and the 8+4 configurations have the same near-memory cache size (8x32GB DIMMs). The 8+4 configuration has only half the number of PMem DIMMs compared to 8+8 configuration, and they are populated on only four out of the eight total DDR channels. This means while the data movement from PMem to DRAM when there is a near-memory miss is mostly on the same DDR channel for the 8+8 configuration, sometimes the data-line is transferred from PMem on one DDR channel to DRAM on another DDR channel for the 8+4 configuration. This results in higher overhead (wasted) bandwidth consumed by the near-memory cache miss with the 8+4 configuration compared to the 8+8 configuration.

This result explains the bandwidth performance behavior comparison between the 8+8 configuration and the 8+4 configuration:

- When the memory footprint is small, the near-memory cache miss rate is low, and the 8+8 configuration only has a small memory bandwidth performance advantage over to the 8+4 configuration.
- As memory footprint increases, the near-memory cache miss rate also increases, and the memory bandwidth performance gap between the 8+8 configuration and the 8+4 configuration also increases.

Both the 8+4 configuration and the 4+4 configuration have the same far-memory size (4x256GB PMem). However, the 8+4 configuration has twice the near-memory cache size compared to the 4+4 configuration (8x32GB DRAM vs 4x32GB DRAM). As a result, for the same memory footprint, the near-memory cache miss rate is always higher with the 4+4 configuration. This results in a lower memory bandwidth performance with the 4+4 configuration compared to the 8+4 configuration.

All-read random memory bandwidth performance

Figure 3 shows memory bandwidth performance comparison for all memory configurations with All Read Random memory access pattern.

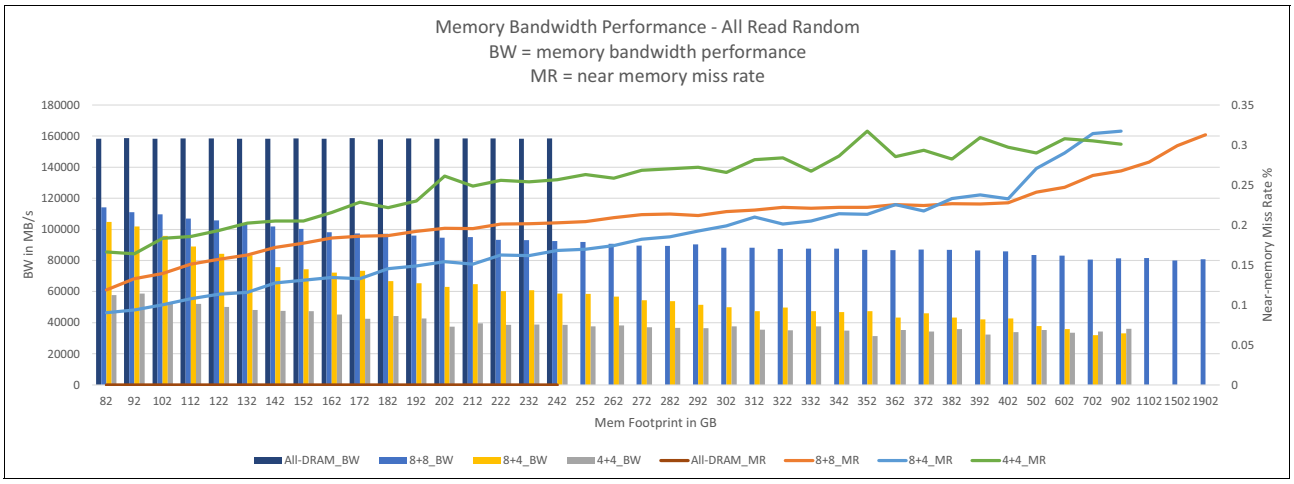

Figure 3 Memory bandwidth performance – all read random

We can observe similar performance behaviors in the All Read Random test, as we did with the All Read Sequential tests:

- The All_DRAM configuration has the best memory bandwidth performance of all memory configurations.

- The 8+8 configuration has the best memory bandwidth performance among all PMem memory configurations. This configuration utilizes all 8 DDR channels, and it has both DRAM and PMem on each memory channel.
- The 8+4 configuration has a small memory bandwidth performance drop compared to the 8+8 configuration when the memory footprint is small. The performance gap increases as the memory footprint increases and the near-memory cache miss rate increases.
- The 4+4 configuration has the lowest memory bandwidth performance. At the same memory footprint, the 4+4 configuration has higher near-memory cache miss rate compared to the 8+4 configuration due to having smaller near-memory (L4) cache. This performance gap decreases as the memory footprint increases and the near-memory cache miss rate increases.

2R:1W Sequential Memory Bandwidth Performance

Figure 4 shows memory bandwidth performance comparison for all memory configurations with 2R:1W Sequential memory access pattern.

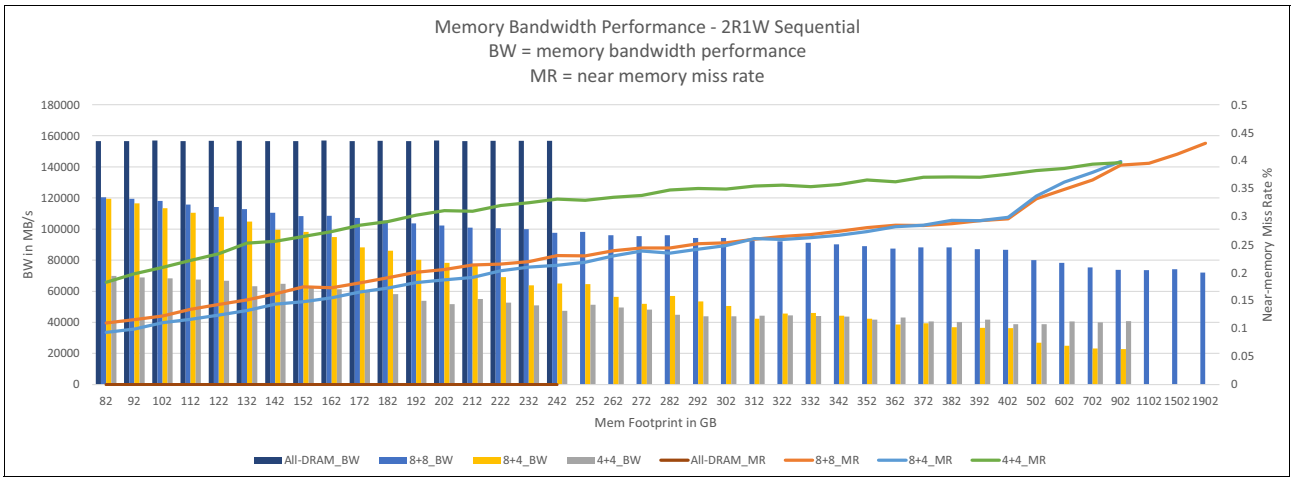

Figure 4 Memory bandwidth performance – 2R:1W sequential

Memory bandwidth performance behavior between the memory configurations with this access pattern is similar to what we discussed above with All Read Sequential/Random access patterns. In general, for a specific memory configuration at the same memory footprint, memory bandwidth measurement is lower with the 2R:1W Sequential data access pattern compared to the All Read Sequential data access pattern.

2R:1W Random Memory Bandwidth Performance

Figure 5 shows memory bandwidth performance comparison for all memory configurations with 2R:1W Random memory access pattern.

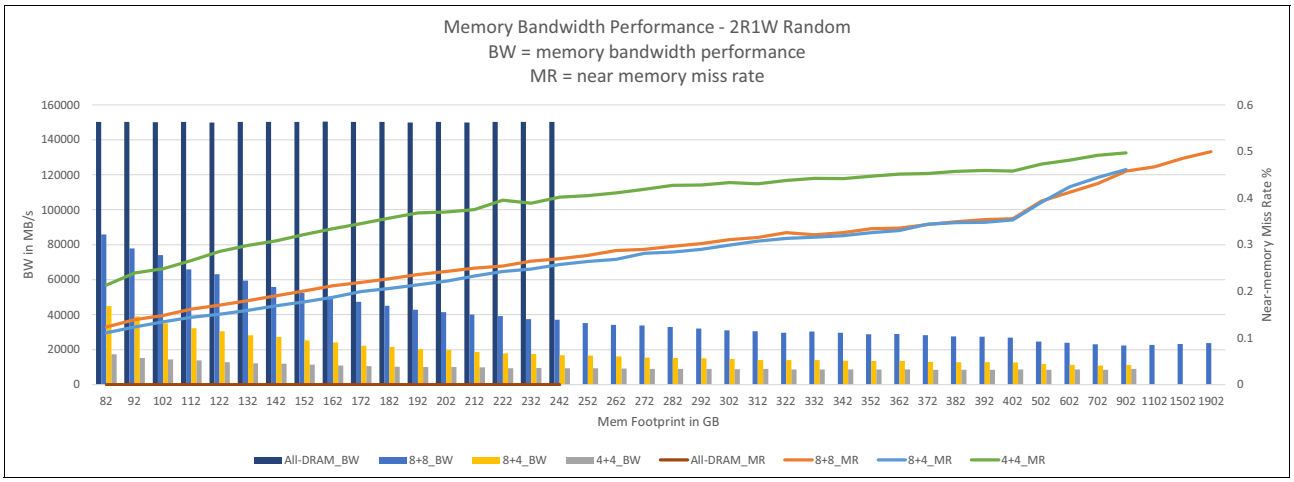

Figure 5 Memory bandwidth performance – 2R:1W random

Memory bandwidth performance behavior between the memory configurations with this access pattern is similar to what we discussed above with All Read Sequential/Random access patterns. In general, for a specific memory configuration at the same memory footprint, memory bandwidth measurement is lower with the 2R:1W Random data access pattern compared to the All Read Random data access pattern.

Memory latency performance

Memory latency is the second important performance metric for memory sub-system, the other being Memory bandwidth as described in “Memory bandwidth performance” on page 6.

When there is a miss to processor L3 cache, memory latency measures the time it takes to get that data-line from memory to the requesting core. In the case when there is a hit to near-memory, the data-line is cached in near-memory, and the latency is close to DRAM latency. However, if there is a miss to near-memory, the data-line is transferred first from far-memory to near-memory. It is then transferred from near-memory to processor L3 cache. As a result, when there is a miss to near-memory, latency will be higher, and it represents PMem latency.

All-read sequential memory latency performance

Figure 6 shows the memory latency performance comparison for all memory configurations with All Read Sequential memory access pattern.

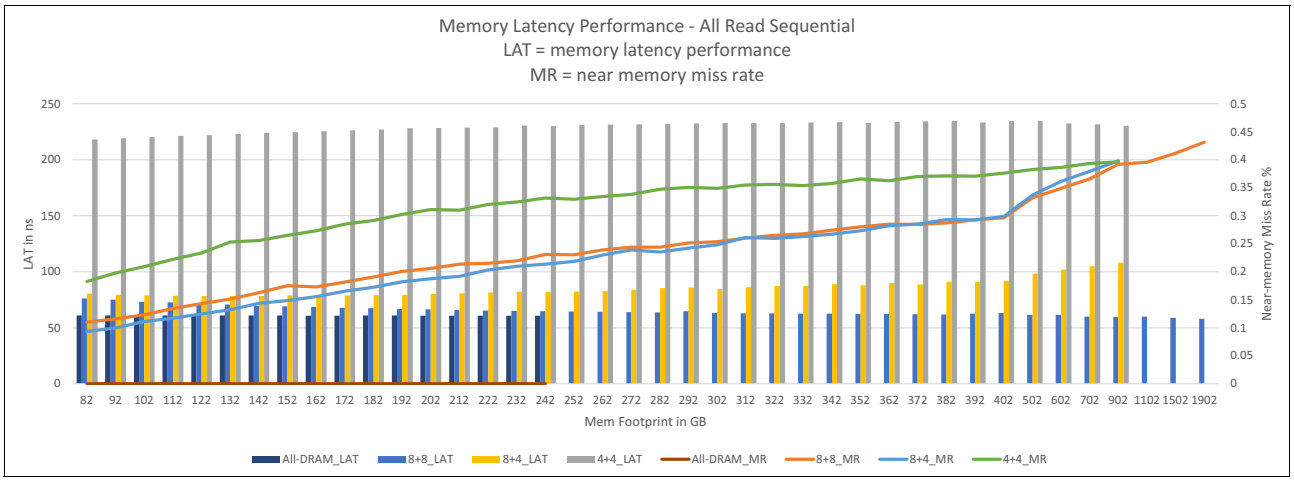

Figure 6 Memory latency performance (lower is better) – All Read Sequential

The bar chart with the left y-axis has the memory latency performance measured in ns. The line chart with the right y-axis has the near-memory miss rate. The x-axis shows the memory footprint in GB used for the test.

As explained in “Memory bandwidth performance” on page 6, as the memory footprint increases, the near-memory cache miss rate also increases.

The latency measured in these tests is “loaded latency”. That means the latency was measured when the system is under heavy load. In these tests, memory accesses were generated at a very high rate to measure the maximum sustainable memory bandwidth and the associated latency.

The two factors that have impact on the loaded latency measured in these tests are:

- Memory bandwidth utilization – As the memory bandwidth performance increases, and the memory bandwidth utilization increases, the loaded latency also increases.
- Near-memory cache miss rate – When there is a miss in the near-memory cache, the data-line is brought into near-memory cache (DRAM) from far-memory (PMem). This results in higher latency. As the near-memory cache miss rate increases, the memory latency also increases.

In Figure 6 on page 11, All_DRAM configuration has the lowest (best) memory latency performance. The latency performance also stays pretty much the same as the memory footprint increases since there is no near-memory cache in this configuration.

At a given memory footprint, all memory configurations with PMem have higher latencies compared to the All_DRAM configuration. It is important to consider the two factors mentioned above when we compare latency performance between All_DRAM configuration to any of the memory configurations with PMem:

- As long as the near-memory cache memory miss rate is not zero, latency performance with any PMem configuration is expected to be worse than latency with All_DRAM configuration because it has to account for the data transfer from PMem to DRAM for the near-memory cache misses.

- All_DRAM configuration has better memory bandwidth performance (higher memory bandwidth utilization) compared to PMem configurations as discussed in the previous memory bandwidth performance section. The higher memory bandwidth utilization will push latency higher on All_DRAM configuration compared to PMem configurations.

The measured latency performance reflects the impact of these two opposite influencing factors.

Both the 8+8 and the 8+4 configurations have the same near-memory cache size (8x32GB DIMMs), but the 8+4 configuration only has half the PMem size compared to the 8+8 configuration. There are data movements from PMem on one DDR channel to DRAM on another DDR channel for some of the near-memory cache misses in the 8+4 configuration. This results in a longer latency with the 8+4 configuration compared to the 8+8 configuration. As the near-memory cache miss rate increases, the latency performance gap between these two configurations also increases.

The 8+4 configuration and the 4+4 configuration have the same far-memory size (4x256GB PMem), but the 4+4 configuration has only half the near-memory cache size compared to the 8+4 configuration. This results in higher near-memory cache miss rate, and thus higher latency, at a given memory footprint in the 4+4 configuration compared to the 8+4 configuration.

All-read random memory latency performance

Figure 7 shows the memory latency performance comparison for all memory configurations with All Read Random memory access pattern.

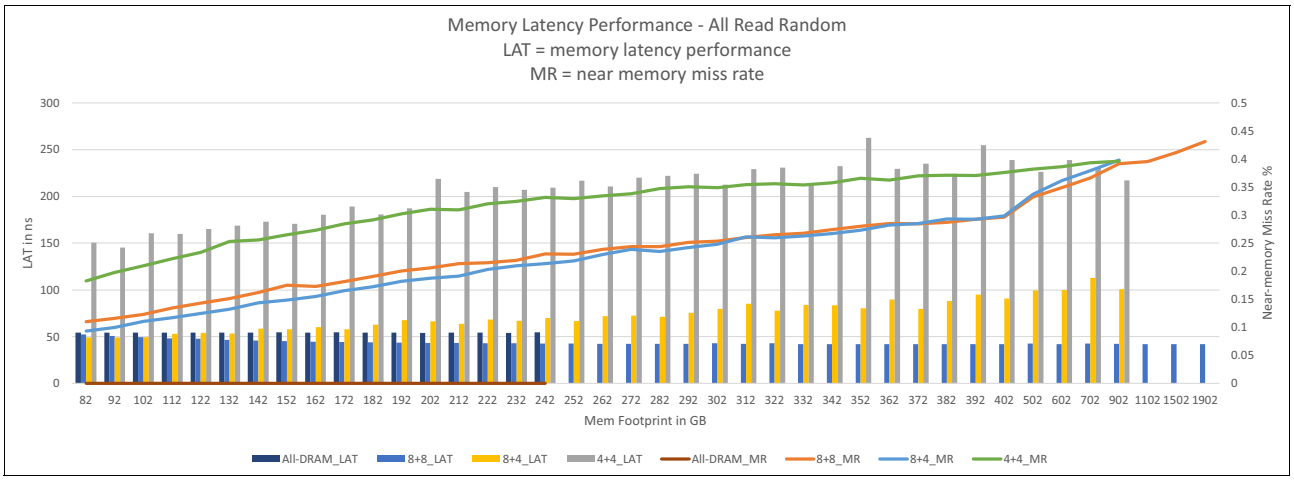

Figure 7 Memory latency performance – All Read Random

The All Read Random tests share the following latency performance behaviors with the All Read Sequential tests:

- The 8+8 configuration has the best memory latency performance among all PMem memory configurations.
- The 8+4 configuration has a small memory latency performance drop (higher latency) compared to the 8+8 configuration when the memory footprint is small. The performance gap increases as the memory footprint increases and the near-memory cache miss rate increases.
- The 4+4 configuration has the lowest memory latency performance. At the same memory footprint, the 4+4 configuration has the highest near-memory cache miss rate compared to

other PMem configurations due to having smaller near-memory cache. The performance gap decreases as the memory footprint increases and the near-memory cache miss rate increases.

The All_DRAM configuration has worse latency performance compared to the 8+8 configuration in the All Read Random tests. This seems odd at first, but we need to take into consideration the impact of the two influencing factors mentioned above to the loaded latency measurements. The gap in memory bandwidth performance between the All_DRAM configuration and the 8+8 configuration is larger in All Read Random memory access pattern.

For this reason, the latency adder component due to memory bandwidth utilization is higher with the All Read Random tests, and that is the reason latency performance comparison between the All_DRAM configuration and the 8+8 configuration is different between the All Read Sequential and the All Read Random tests.

2R:1W Sequential Memory Latency Performance

Figure 8 shows memory latency performance comparison between all memory configurations with 2R:1W Sequential memory access pattern.

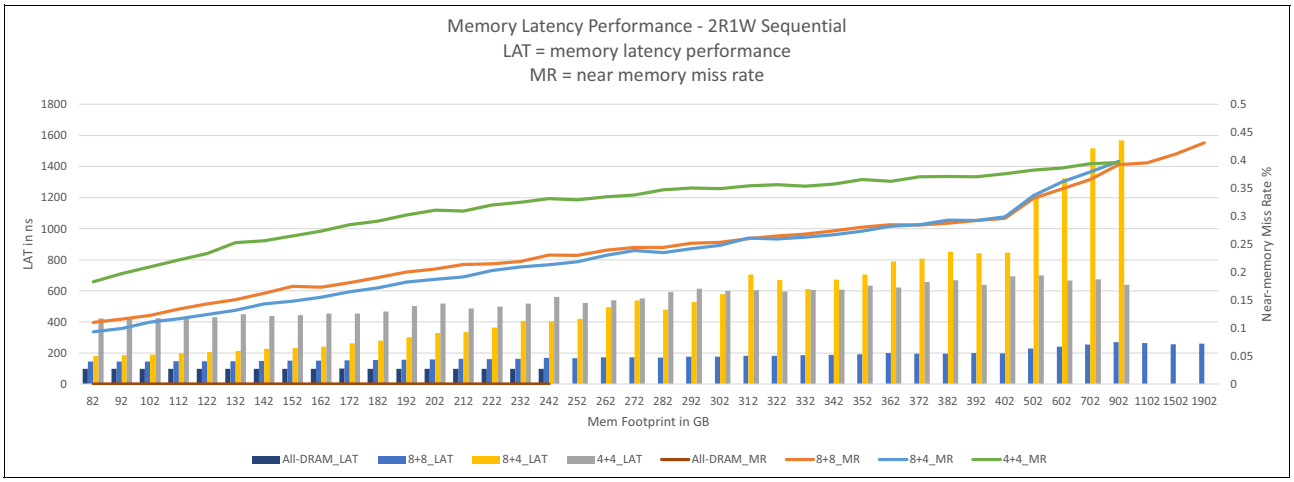

Figure 8 Memory latency performance – 2R:1W Sequential

Memory latency performance behavior between the memory configurations with this access pattern is similar to what we discussed above with All Read Sequential access pattern. In general, for a specific memory configuration at the same memory footprint, memory latency measurement is higher with the 2R:1W Sequential data access pattern compared to the All Read Sequential data access pattern.

2R:1W Random Memory Latency Performance

Figure 9 shows memory latency performance comparison between all memory configurations with 2R:1W Random memory access pattern.

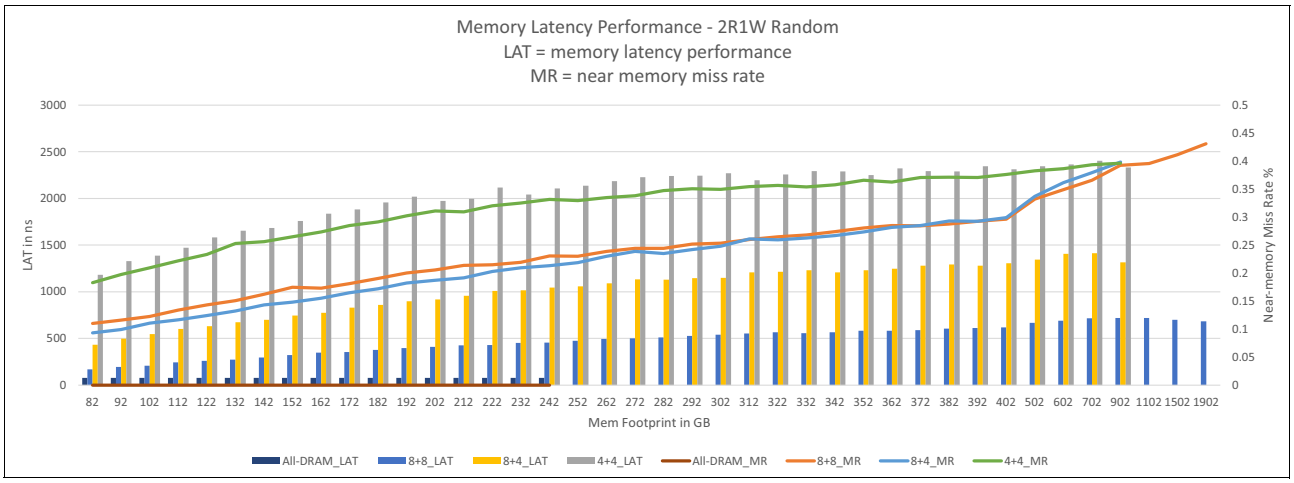

Figure 9 Memory latency performance – 2R:1W Random

Memory latency performance behavior between the memory configurations with this access pattern is similar to what we discussed above with All Read Sequential access pattern. In general, for a specific memory configuration at the same memory footprint, memory latency measurement is higher with the 2R:1W Random data access pattern compared to the All Read Random data access pattern.

Summary

Intel Optane Persistent Memory 200 Series in Memory Mode can be used as an effective way to bridge the gap in cost and capacity between DRAM memory and Storage. Due to higher DIMM capacity support (max 128GB with DRAM DIMMs vs max 512GB with PMem DIMMs), the use of PMem enables a server to support higher total system memory capacity. Also, at the same capacity, PMem DIMMs costs less than DRAM DIMMs, so there can be cost saving using PMem instead of DRAM to support the same total system memory capacity.

The near-memory miss rate has a direct impact to memory bandwidth and latency performance for Persistent Memory configurations in Memory Mode. Near-memory miss rate is dependent on the following factors:

- Application data access pattern
- Near-memory capacity
- DRAM:PMem capacity ratio

Keeping the near-memory miss rate relatively low is important to achieve a good level of performance with Persistent Memory configurations in Memory Mode. For these PMem configurations, to maximize memory sub-system and application performance, consider the following tips:

- Choose a Persistent Memory size that fits most of your application data set. This will minimize I/O accesses and maximize your application performance.

- Higher DRAM:PMem capacity ratio can improve the near-memory cache miss rate and improve application performance. Intel recommends a ratio between 1:4 and 1:16.
- For a given DRAM:PMem capacity ratio, the best memory configuration should have one DRAM DIMM and one PMem DIMM populated per DDR channel. All DDR channels should have identical memory configurations. In other words, we recommend the 8+8 PMem configuration whenever possible. This configuration maximizes the effective memory bandwidth for PMem configuration.

Authors

Nathan Pham is a Senior Engineer in the Lenovo Infrastructure Solution Group Performance Laboratory in Morrisville, NC. He has spent over 19 years between IBM and Lenovo working on system performance. His focus has been in system performance architecture, modeling, analysis, and optimization. In his current role, he is responsible for all aspects of future Lenovo DCG system architecture and design for optimized performance. Nathan holds a Bachelor of Science degree in Computer Engineering and a Master of Science degree in Computer Engineering, both from North Carolina State University.

Kevin Huang is an x86 Enterprise Server Performance Engineer in the Lenovo Infrastructure Solutions Group Performance Lab in the Taipei Design Center. In his current role, he mainly focuses on the CPU and memory subsystem and is responsible for ensuring the subsystem performance to be reliable and competitive before product announcements, for every new generation of x86 server products developed by Lenovo. Prior to working on server performance, Kevin has also worked on automation, system back-end development, and software architecture.

Thanks to the following people for their contributions to this project:

- David Watts, Lenovo Press

Notices

Lenovo may not offer the products, services, or features discussed in this document in all countries. Consult your local Lenovo representative for information on the products and services currently available in your area. Any reference to a Lenovo product, program, or service is not intended to state or imply that only that Lenovo product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any Lenovo intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any other product, program, or service.

Lenovo may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:

Lenovo (United States), Inc.
1009 Think Place - Building One
Morrisville, NC 27560
U.S.A.
Attention: Lenovo Director of Licensing

LENOVO PROVIDES THIS PUBLICATION “AS IS” WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some jurisdictions do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. Lenovo may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

The products described in this document are not intended for use in implantation or other life support applications where malfunction may result in injury or death to persons. The information contained in this document does not affect or change Lenovo product specifications or warranties. Nothing in this document shall operate as an express or implied license or indemnity under the intellectual property rights of Lenovo or third parties. All information contained in this document was obtained in specific environments and is presented as an illustration. The result obtained in other operating environments may vary.

Lenovo may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Any references in this publication to non-Lenovo Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this Lenovo product, and use of those Web sites is at your own risk.

Any performance data contained herein was determined in a controlled environment. Therefore, the result obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

This document was created or updated on September 23, 2021.

Send us your comments via the **Rate & Provide Feedback** form found at
http://lenovopress.com/lp1528

Trademarks

Lenovo and the Lenovo logo are trademarks or registered trademarks of Lenovo in the United States, other countries, or both. These and other Lenovo trademarked terms are marked on their first occurrence in this information with the appropriate symbol (® or ™), indicating US registered or common law trademarks owned by Lenovo at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of Lenovo trademarks is available from https://www.lenovo.com/us/en/legal/copytrade/.

The following terms are trademarks of Lenovo in the United States, other countries, or both:

Lenovo® Lenovo(logo)® ThinkSystem™

The following terms are trademarks of other companies:

Intel, Intel Optane, VTune, Xeon, and the Intel logo are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Other company, product, or service names may be trademarks or service marks of others.