

Configuring Memory Speed for Optimal Memory Latency with AMD EPYC Gen 2 and Gen 3 Processors

Planning / Implementation

Configuring AMD EPYC Gen 2 and Gen 3 based servers for minimal memory latency is essential for getting the highest possible performance in applications that favor lower memory latency over peak memory bandwidth – such as OLTP Database workloads running on SQL Server. Unlike most servers, on AMD EPYC Gen 2 based systems, the highest configurable memory speed will not always result in the lowest memory latency and the highest memory bandwidth. As such, understanding the behavior of memory latency and how it scales with memory speed on these platforms is necessary to ensure the highest levels of performance.

AMD EPYC Gen 2 and Gen 3 processors support a maximum memory speed of 3200 MHz, however, it is important to note that there are implementation differences between the two generations. As a result, memory latency behavior with respect to the configured memory speed also changes. This performance brief will explain the differences in memory speed implementation on AMD EPYC Gen 2 and Gen 3 processors and provide guidance on how to set memory speed to achieve the lowest memory latency possible.

Understanding Memory Bus Speeds

AMD EPYC Gen 2 and Gen 3 processors support DDR4 Memory. Double Data Rate (DDR) technology allows for the memory signal to be sampled twice per clock cycle: once on the rising edge and once on the falling edge of the clock signal. Because of this, the reported memory speed rate is twice the true memory clock frequency.

AMD EPYC processors also feature an Infinity Fabric that serves as an interconnect between the CPU cores and main memory. When optimally configured, the clock speed of this Infinity Fabric is typically equal to the true memory clock frequency, or half of the reported memory speed. For example, at 2933 MHz memory speed (1467 MHz memory clock frequency), the Infinity Fabric frequency is configured to 1467 MHz. Maintaining this 1:1 ratio between the memory clock frequency and the Infinity Fabric frequency yields the best memory latency.

On Lenovo ThinkSystem SR645 and SR665 servers with AMD EPYC Gen 2 or Gen 3 processors, the memory bus speed can run at 2666 MHz, 2933 MHz or 3200 MHz. The speed depends on the RDIMM selection and the number of memory DIMMs installed in each memory channel. When configured with Performance+ RDIMMs, the ThinkSystem SR645 and SR665 servers support a memory bus speed up to 3200 MHz when configured with two memory DIMMs installed in each memory channel.

For more information, see the ThinkSystem SR645 and SR665 product guides:

- SR645 Product Guide: <https://lenovopress.com/lp1280>
- SR665 Product Guide: <https://lenovopress.com/lp1269>

AMD EPYC Gen 2

On AMD EPYC Gen 2 processors, the maximum Infinity Fabric frequency is 1467 MHz. Thus, at a configured memory speed of 3200 MHz, the 1:1 ratio between the memory clock frequency and Infinity Fabric frequency is no longer upheld. This concept is called “Decoupling” and roughly results in a 12ns memory latency penalty to synchronize the two different frequency domains.

The following table shows the relationship between the different clock frequencies and the resulting memory latency and bandwidth on AMD EPYC Gen 2 processors.

Table 1. Memory Latency and Bandwidth vs. Memory Speed on AMD EPYC Gen 2 CPUs

Memory Speed	Infinity Fabric Frequency	Memory Latency	Memory Bandwidth
3200 (1600) MHz	1467 MHz	103 ns	180 GB/s
2933 (1467) MHz	1467 MHz	90 ns	169 GB/s
2666 (1333) MHz	1333 MHz	92 ns	153 GB/s

About the measurements: Multichase was used to measure memory latency, average local node latency reported; AMD stream-dynamic was used to measure bandwidth, stream triad bandwidth reported. All measurements were made in a controlled lab environment. Actual customer measurements may vary depending on configuration and application workload

Due to the decoupling latency penalty at 3200 MHz, 2933 MHz results in the lowest possible memory latency, while maximum memory bandwidth is achieved at 3200 MHz. It is important to understand the memory latency and bandwidth performance tradeoff between 2933 MHz and 3200 MHz to determine which memory speed setting is optimal for your workload. Some workloads may benefit from higher peak memory throughput, while others are more latency sensitive and will perform better with lower memory latency – even at the cost of peak memory bandwidth.

AMD EPYC Gen 3

AMD EPYC Gen 3 processors support a maximum Infinity Fabric frequency of 1600 MHz, meaning there is no instance where the memory clock frequency and the Infinity Fabric frequency are decoupled. As a result, memory latency now decreases more linearly as the memory speed is increased.

The following table shows the relationship between the different clock frequencies and the resulting memory latency and bandwidth on AMD EPYC Gen 3 processors.

Table 2. Memory Latency and Bandwidth vs. Memory Speed on AMD EPYC Gen 3 CPUs

Memory Speed	Infinity Fabric Frequency	Memory Latency	Memory Bandwidth
3200 (1600) MHz	1600 MHz	76 ns	187 GB/s
2933 (1467) MHz	1467 MHz	79 ns	172 GB/s
2666 (1333) MHz	1333 MHz	82 ns	157 GB/s

At a memory speed of 3200 MHz, the Infinity Fabric frequency is set to 1600 MHz and the 1:1 ratio with the memory clock frequency is maintained. This results in the best possible memory latency as well as the highest memory bandwidth. There is no latency penalty at 3200 MHz, thus, to achieve the best memory performance, simply set the memory speed to the highest speed that can be supported based on the DIMM selection and memory population.

Conclusion

While both AMD EPYC Gen 2 and Gen 3 processors can support a maximum memory speed of 3200 MHz, there is a difference in maximum frequency the Infinity Fabric clock can support. As a result, memory latency behavior changes from generation to generation. It is important to understand these differences and the demands of different workloads to ensure memory speed is set optimally:

- For AMD EPYC Gen 2 processors, 2933 MHz results in the lowest memory latency while 3200 MHz provides the highest memory bandwidth.
- For AMD EPYC Gen 3 processors, 3200 MHz provides the lowest memory latency and the highest memory bandwidth.

About the author

Jamal Ayoubi is a Systems Performance Verification Engineer in the Lenovo Infrastructure Solutions Group Performance Laboratory in Morrisville, NC. His current role includes CPU, Memory, and PCIe subsystem analysis and performance validation against functional specifications and vendor targets. Jamal specializes in AMD EPYC architecture, UEFI specification, and performance tuning. Jamal holds Bachelor of Science degrees in Electrical Engineering and Computer Engineering from North Carolina State University.

Related product families

Product families related to this document are the following:

- [Processors](#)

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.
8001 Development Drive
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.

© Copyright Lenovo 2024. All rights reserved.

This document, LP1492, was created or updated on July 15, 2021.

Send us your comments in one of the following ways:

- Use the online Contact us review form found at:
<https://lenovopress.lenovo.com/LP1492>
- Send your comments in an e-mail to:
comments@lenovopress.com

This document is available online at <https://lenovopress.lenovo.com/LP1492>.

Trademarks

Lenovo and the Lenovo logo are trademarks or registered trademarks of Lenovo in the United States, other countries, or both. A current list of Lenovo trademarks is available on the Web at <https://www.lenovo.com/us/en/legal/copytrade/>.

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

Lenovo®

ThinkSystem®

The following terms are trademarks of other companies:

SQL Server® is a trademark of Microsoft Corporation in the United States, other countries, or both.

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