DELL Technologies

Whitepaper

Balanced Memory with 2nd Generation AMD EPYC™ Processors for PowerEdge Servers

Optimizing Memory Performance

Revision: **1.4**
Issue Date: **4/21/2020**

Abstract

Properly configuring a server with balanced memory is critical to ensure memory bandwidth is maximized and latency is minimized. When server memory is configured incorrectly, unwanted variables are introduced into the memory controllers' algorithm, which inadvertently slows down overall system performance. To mitigate this risk of reducing or even bottlenecking system performance, it is important to understand what constitutes balanced, near balanced and unbalanced memory configurations.

Dell EMC has published this brief to educate PowerEdge customers on what balanced memory means, why it is important and how to properly populate memory to 2nd Generation AMD EPYC™ server processors for a balanced configuration.

Revisions

Date	Description
12 September 2019	Initial release for 1st wave of AMD CPUs
21 April 2020	Includes all AMD CPU SKUs

Acknowledgements

This paper was produced by the following people:

Name	Role
Matt Ogle	Technical Product Marketing, Dell EMC
Trent Bates	Product Management, Dell EMC
Jose Grande	Software Senior Principal Engineer, Dell EMC
Andres Fadul	Software Senior Principal Engineer, Dell EMC

Table of Contents

1. Introduction

Understanding the relationship between a server processor (CPU) and its memory subsystem is critical when optimizing overall server performance. Every processor generation has a unique architecture, with volatile controllers, channels and slot population guidelines, that must be satisfied to attain high memory bandwidth and low memory access latency.

2nd Generation AMD EPYC™ server processors, which will be referred to by their code name throughout this white paper, Rome processors, offer a total of eight memory channels with up to two memory slots per channel.[1] This presents numerous possible permutations for configuring the memory subsystem with traditional Dual In-Line Memory Modules (DIMMs), yet there are only a couple of balanced configurations that will achieve the peak memory performance for Dell EMC PowerEdge servers.

Memory that has been incorrectly populated is referred to as an unbalanced configuration. From a functionality standpoint, an unbalanced configuration will operate adequately, but introduces significant additional overhead that will slow down data transfer speeds. Similarly, a near balanced configuration does not yield fully optimized data transfer speeds but it is only suboptimal to that of a balanced configuration. Conversely, memory that has been correctly populated is referred to as a balanced configuration and will secure optimal functionality and data transfer speeds.

This white paper explains how to balance memory configured for Rome processors within Dell EMC PowerEdge servers.

2. Memory Topography and Terminology

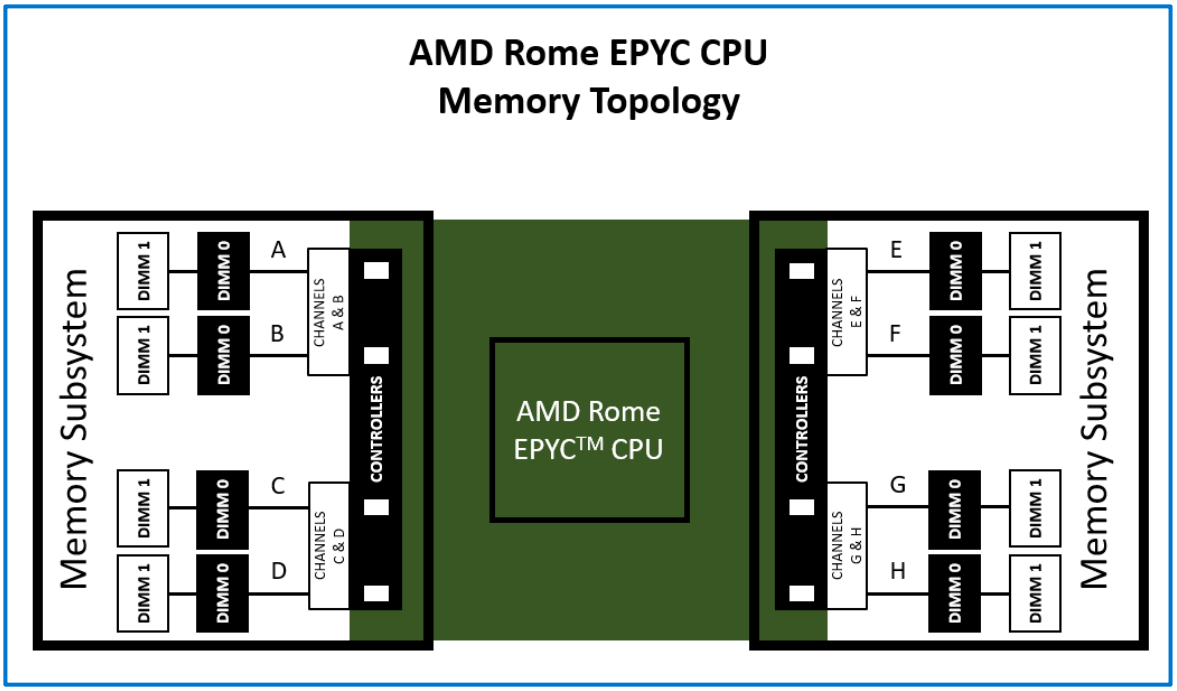

Figure 1: CPU-to-memory subsystem connectivity for Rome processors

To understand the relationship between the CPU and memory, terminology illustrated in Figure 1 must first be defined:

- Memory controllers are digital circuits that manage the flow of data going from the computer's main memory to the corresponding memory channels.[2] Rome processors have eight memory controllers in the processor I/O die, with one controller assigned to each channel.
- Memory channels are the physical layer on which the data travels between the CPU and memory modules.[3] As seen in Figure 1, Rome processors have eight memory channels designated A, B, C, D, E, F, G and H. These channels were intended to be organized into pairs such as two-way (AB, CD, EF, GH), four-way (ABCD, EFGH) or eight-way (ABCDEFGH).
- The memory slots are internal ports that connect the individual DIMMs to their respective channels.[4] Rome processors have two slots per channel, so there are a total of sixteen slots per CPU for memory module population. DIMM 1 slots are the first eight memory modules to be populated while DIMM 0 slots are the last eight. In the illustrations ahead, DIMM 1 slots will be represented with black text marked A1-A8 and DIMM 0 slots will be represented with white text marked A9-A16.
- The memory subsystem is the combination of all the independent memory functions listed above.

3. Memory Interleaving

Memory interleaving allows a CPU to efficiently spread memory accesses across multiple DIMMs. When memory is put in the same interleave set, contiguous memory accesses go to different memory banks. Memory accesses no longer must wait until the prior access is completed before initiating the next memory operation. For most workloads, performance is maximized when all DIMMs are in one interleave set creating a single uniform memory region that is spread across as many DIMMs as possible.[5] Multiple interleave sets create disjointed memory regions.

3.1 NPS and Quadrant Pairing

Rome processors achieve memory interleaving by using Non-Uniform Memory Access (NUMA) in Nodes Per Socket (NPS).[6] There are four NPS options available in the Dell EMC BIOS:

1. **NPS 0** – One NUMA node per system (on two processors systems only). This means all channels in the system are using one interleave set.
2. **NPS 1** – One NUMA node per socket (on one processor systems). This means all channels in the socket are using one interleave set.
3. **NPS 2** – Two NUMA nodes per socket (one per left/right half). This means each half containing four channels is using one interleave set; a total of two sets.
4. **NPS 4** – Up to four NUMA nodes per socket (one per quadrant). This means each quadrant containing two channels is using one interleave set; a total of four sets.

The simplest visual aid for understanding the NPS system is to divide the CPU into four quadrants. We see below in Figure 2 that each quadrant contains two paired DIMM channels that can host up to two DIMMs. The paired DIMM channels in each quadrant were designed to group and minimize the travel distance for interleaved sets. NPS 1 would correlate to all four quadrants being fully populated. NPS 2 would correlate to having either the left or right half quadrant being fully populated. NPS 4 would correlate to having any one quadrant being fully populated.

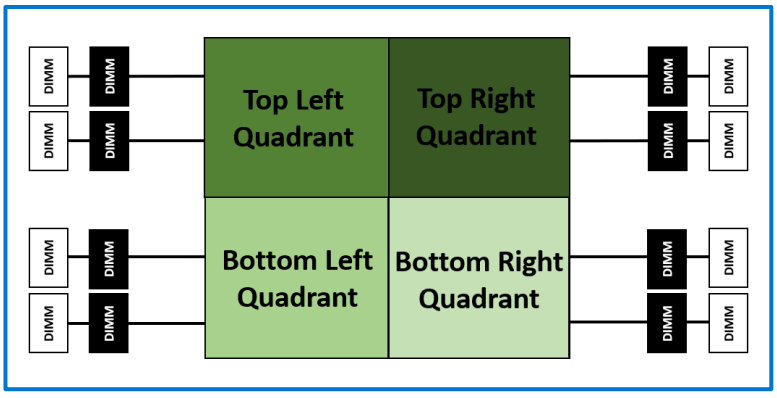

Figure 2: Quadrant layout of Rome processors

3.2 NPS and Quadrant Pairing

NPS 0 and NPS 1 will typically yield the best memory performance, followed by NPS 2 and then NPS 4. The Dell EMC default setting for BIOS NUMA NPS is NPS 1 and may need to be manually adjusted to match the NPS option that supports the CPU model. As seen below in Figure 3 there are various CPUs that will not support NPS 2 or 4 that require awareness of which memory configurations are optimized for each CPU.

CPU Name	DPN	NPS Support
2nd Gen AMD EPYC™ 7742	YVKJ6	1,2,4
2nd Gen AMD EPYC™ 7642	C59HD	1,2,4
2nd Gen AMD EPYC™ 7542	8JWMD	1,2,4
2nd Gen AMD EPYC™ 7702	5PG5C	1,2,4
2nd Gen AMD EPYC™ 7702P	835TD	1,2,4
2nd Gen AMD EPYC™ 7552	3J0XY	**1,2**
2nd Gen AMD EPYC™ 7502	FG4GY	1,2,4
2nd Gen AMD EPYC™ 7502P	3NFJT	1,2,4
2nd Gen AMD EPYC™ 7402	542T2	1,2,4
2nd Gen AMD EPYC™ 7402P	YK5KC	1,2,4
2nd Gen AMD EPYC™ 7452	Y96PT	1,2,4
2nd Gen AMD EPYC™ 7352	F9NJ5	1,2,4
2nd Gen AMD EPYC™ 7302	V99P3	1,2,4
2nd Gen AMD EPYC™ 7302P	3425F	1,2,4
2nd Gen AMD EPYC™ 7262	XPY7D	1,2,4
2nd Gen AMD EPYC™ 7282	J1X8V	**1***
2nd Gen AMD EPYC™ 7252	XJG06	**1***
2nd Gen AMD EPYC™ 7232P	DH26K	**1***
2nd Gen AMD EPYC™ 7272	V0K1X	**1***
2nd Gen AMD EPYC™ 7662	GX27F	1,2,4
2nd Gen AMD EPYC™ 7532	CPHXD	1,2,4
2nd Gen AMD EPYC™ 7H12	MTHGK	1,2,4

Figure 3: A full list of 2nd Gen AMD EPYC™ CPUs and their respective supported NPS models. The CPUs with an asterisk have been optimized to reduce the performance impact of only filling four DIMM channels.

Figure 4 below shows our recommended NPS setting for each # of DIMMs per CPU:

# of DIMMs Per CPU	NPS		
	0/1	2	4
1			*
2			*
3			*
4	*		
5			*
6			*
7			*
8	*		
9			*
10			*
11			*
12		*	
13			*
14			*
15			*
16	*		

Figure 4: Recommended NPS setting for each # of DIMMs per CPU

If the NPS setting for a memory configuration will limit performance (as seen in Figure 5), Dell EMC BIOS will return the following informative prompts to the user:

> UEFI0391: Memory configuration supported but not optimal for the enabled NUMA node Per Socket (NPS) setting. Please consider the following actions:
>
> 1) Changing NPS setting under System Setup>System BIOS>Processor Settings>NUMA Nodes Per Socket, if supported.
>
> 2) For optimized memory configurations please refer to the General Memory Module Installation Guidelines section in the Installation and Service Manual, of the respective server model available on the support site.

In layman's terms, a different NPS setting or memory configuration will result in better memory performance. The system is fully functional when this message appears, but it is not fully optimized for best performance.

# of DIMMs Per CPU	NPS		
	0/1	2	4
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			

Figure 5: Color-coded table illustrating when an informative message will occur (yellow) or no message (green)

4. Memory Population Guidelines

4.1 Overview

DIMMs must be populated into a balanced configuration to yield the highest memory bandwidth and lowest memory access latency. Various factors will dictate whether a configuration is balanced or not. Please follow the guidelines below for best results [7]:

- Memory Channel Population
  - Balanced Configuration
    - All memory channels must be fully populated with one or two DIMMs for best performance; a total of eight or sixteen DIMMs per CPU
  - Near Balanced Configuration
    - Populate four or twelve DIMMs per socket
    - Populate DIMMs in sequential order (A1-A8)
- CPU and DIMM parts must be identical
- Each CPU must be identically configured with memory

4.2 Memory Channel Population

To achieve a balanced configuration, populate either eight or sixteen DIMMs per CPU. By loading each channel with one or two DIMMs, the configuration is balanced and has data traveling across channels most efficiently on one interleave set. Following this guideline will yield the highest memory bandwidth and the lowest memory latency.

If a balanced configuration of sixteen or eight DIMMs per CPU cannot be implemented, then the next best option is a near balanced configuration. To obtain a near balanced population, populate four or twelve DIMMs per CPU in sequential order. When any number of DIMMs other than 4, 8, 12 or 16 is populated, disjointed memory regions are created making NPS 4 the only supported BIOS option to select.

The last guideline is that DIMMs must be populated in an assembly order because Rome processors have an organized architecture for each type of CPU core count. To simplify this concept, the lowest core count was used as a common denominator, so the assembly order below will apply across all Rome processor types. Populating in this order ensures that for every unique Rome processor, any DIMM configuration is guaranteed the lowest NPS option, therefore driving the most efficient interleave sets and data transfer speeds. Figure 6 illustrates the assembly order in which individual DIMMs should be populated, starting with A1 and ending with A16:

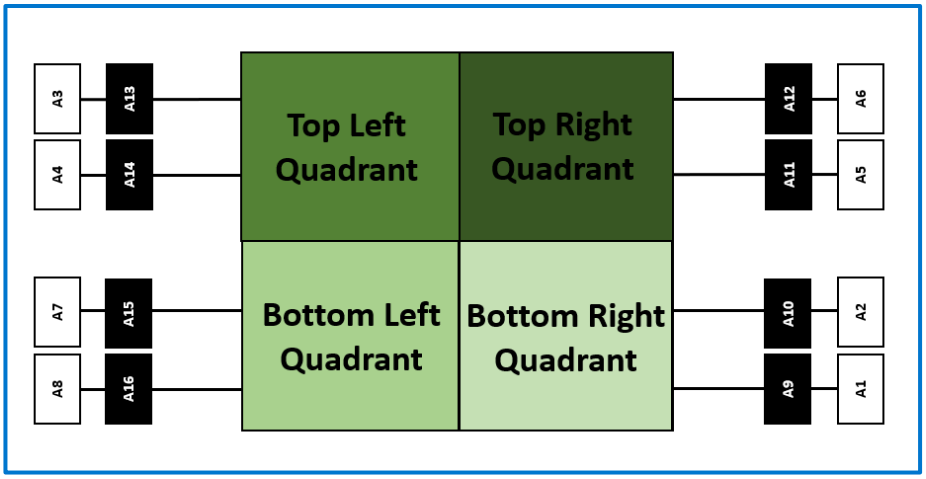

Figure 6: DIMM population order, starting with A1 and ending with A16

4.3 Identical CPU and DIMM Parts

Identical DIMMs must be used across all DIMM slots (i.e. same Dell part number). Dell EMC does not support DIMM mixing in Rome systems. This means that only one rank, speed, capacity and DIMM type shall exist within the system. This principle applies to the processors as well; multi-socket Rome systems shall be populated with identical CPUs.

4.4 Identical Memory Configurations for Each CPU

Every CPU socket within a server must have identical memory configurations. When only one unique memory configuration exists across both CPU sockets within a server, memory access is further optimized. Figure 7 below illustrates the expected memory bandwidth curve when these rules are followed:

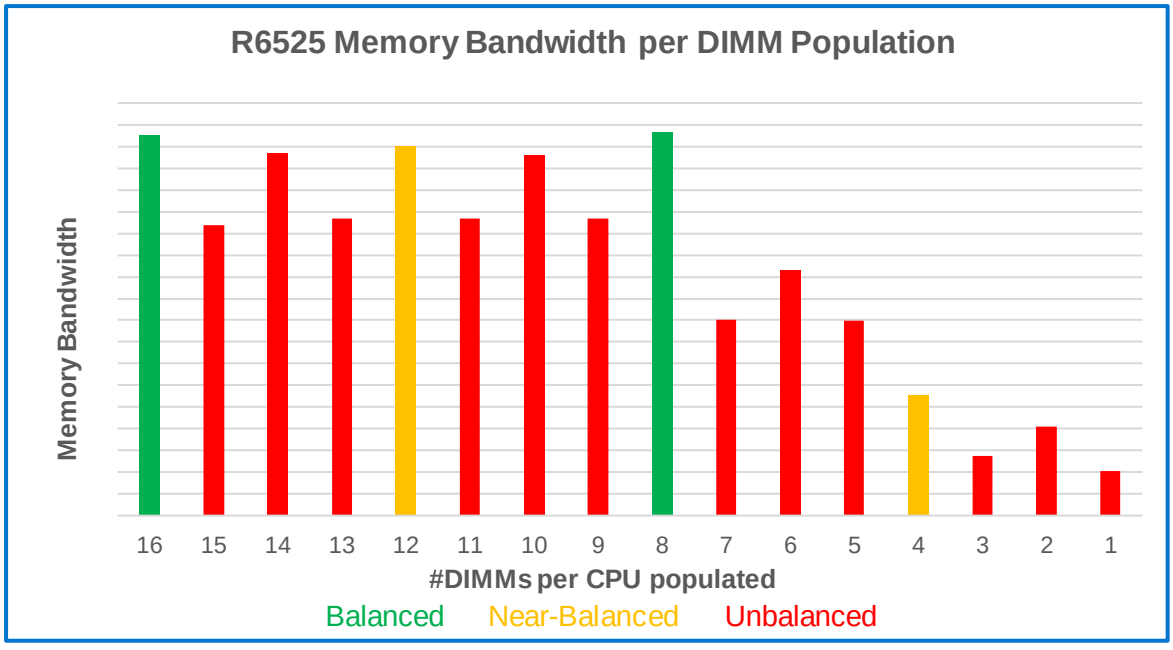

Figure 7: Bar graph illustrating expected performance variation as # of dimms increases

5. Balanced Configurations (Recommended)

Balanced configurations satisfy NPS 0/1 conditions by requiring each memory channel to be populated with one or two identical DIMMs. By doing this, one interleave set can optimally distribute memory access requests across all the available DIMM slots; therefore, maximizing performance. Memory controller logic was designed around fully populated memory channels, so it should come as no surprise that **eight or sixteen populated DIMMs are recommended**. Having eight DIMMs will reap the highest memory bandwidth while having sixteen DIMMs will yield the highest memory capacity.

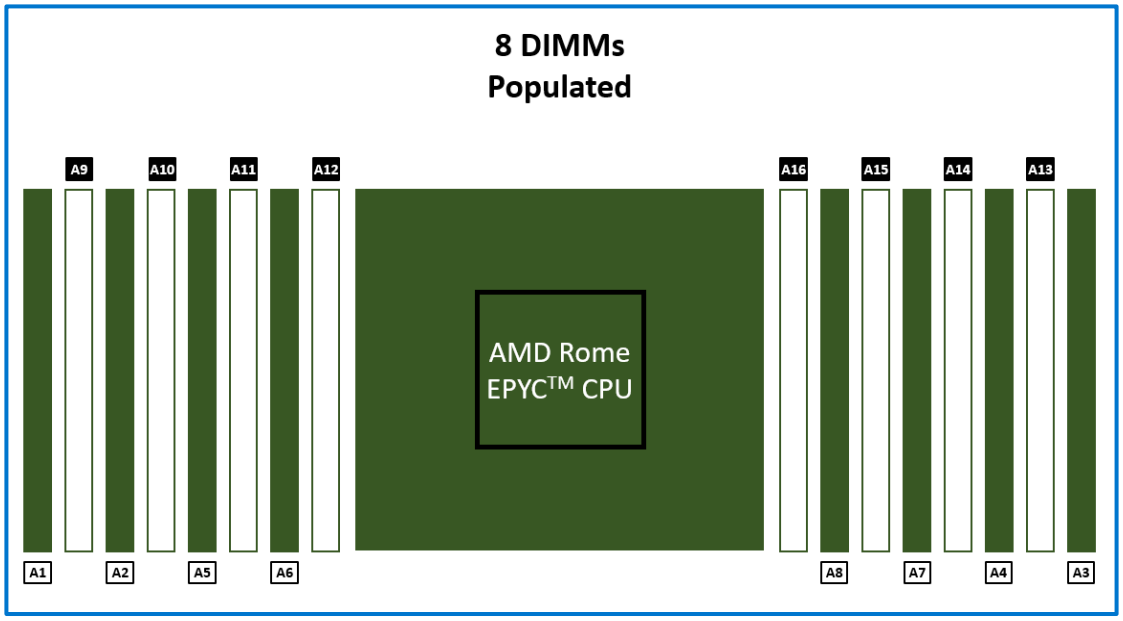

Figure 8: Eight DIMMs are populated in a balanced configuration, producing the highest memory bandwidth while at a lower capacity than sixteen

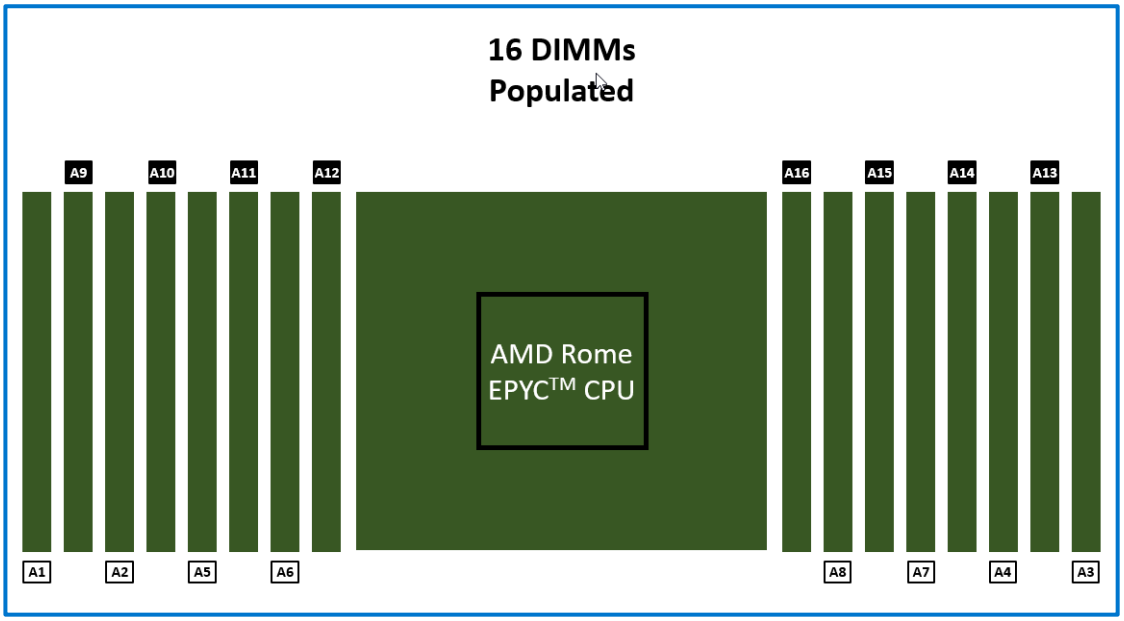

Figure 9: Sixteen DIMMs are populated in a balanced configuration, producing the highest memory capacity while at a lower bandwidth than eight

6. Near Balanced Configurations

Near balanced configurations satisfy NPS 1 or 2 conditions by populating either four or twelve identical DIMMs per CPU. These configurations are not optimized because the channels are partially populated, which creates disjointed memory regions that reduce performance (making it *near* balanced). Performance for near balanced configurations will undergo degradation when compared to balanced configurations. Although the below configurations are adequate for implementation, they are not highly recommended. ***Note that CPUs 7282, 7252, 7232P and 7272 were designed to reduce the performance impact of populating four DIMM channels.**

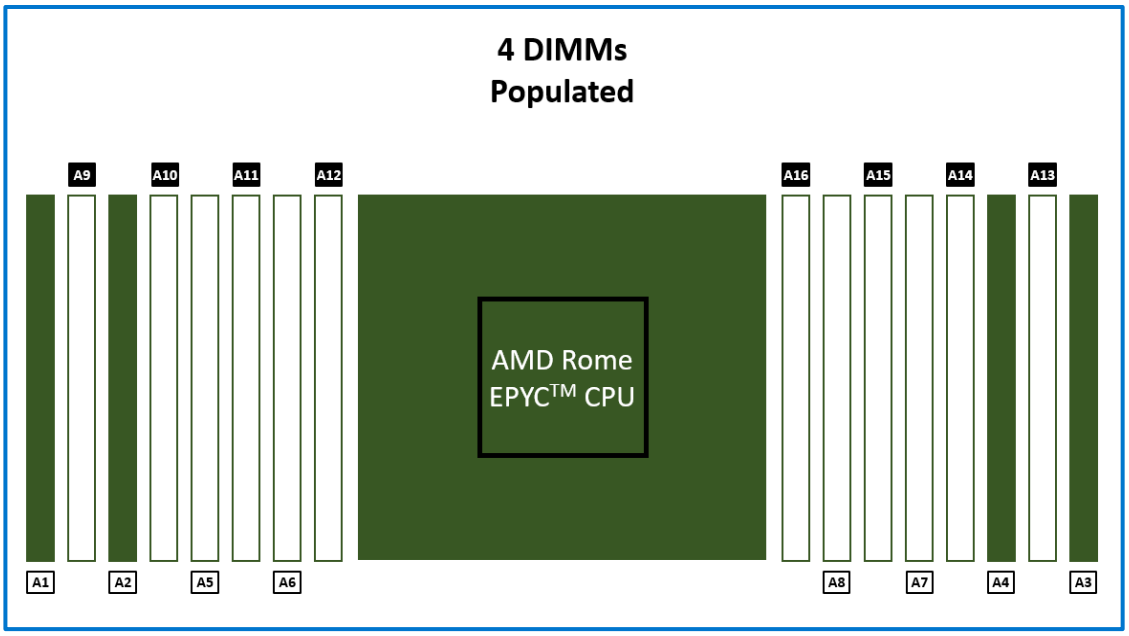

Figure 10: Four DIMMs are populated in a near balanced configuration

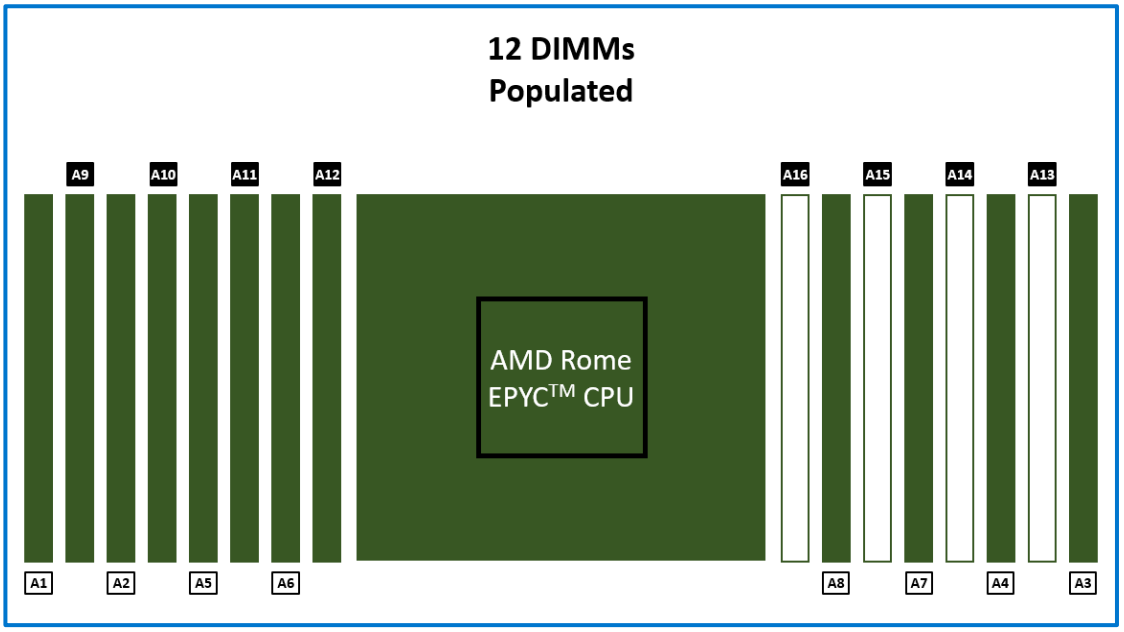

Figure 11: Twelve DIMMs are populated in a near balanced configuration

7. Unbalanced Configurations

Unbalanced configurations can only satisfy NPS 4 conditions. More than two interleave sets can now be introduced to the memory controller algorithm which causes very disjointed regions. Memory performance for the unbalanced configurations below are *significantly* less than balanced or near balanced and are not recommended.

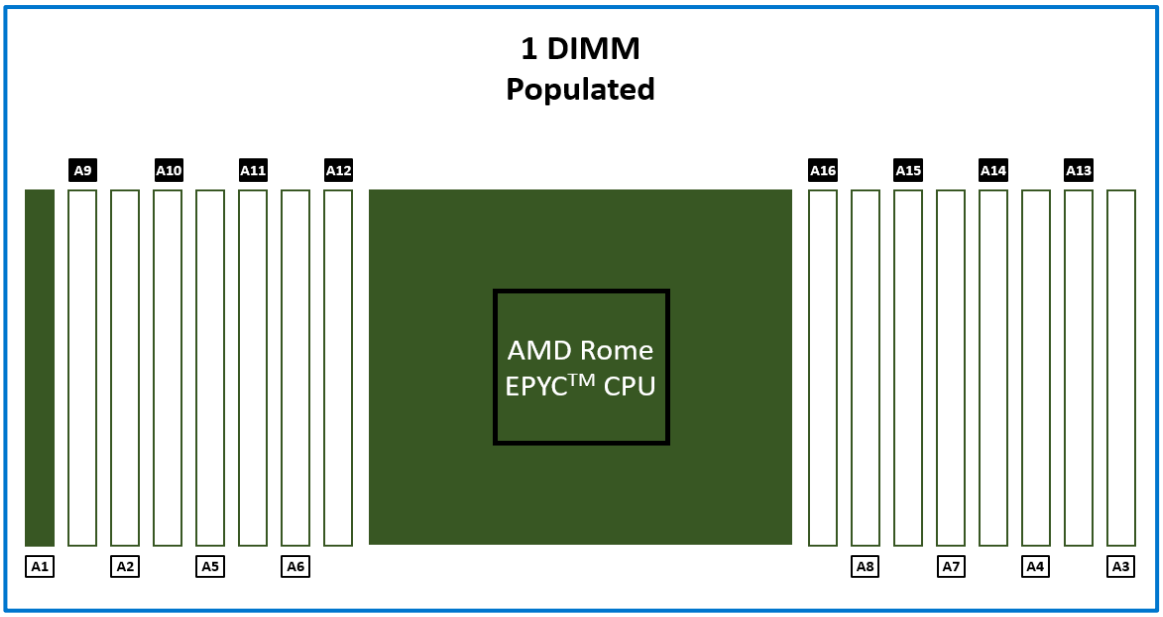

Figure 12: One DIMM is populated in an unbalanced configuration

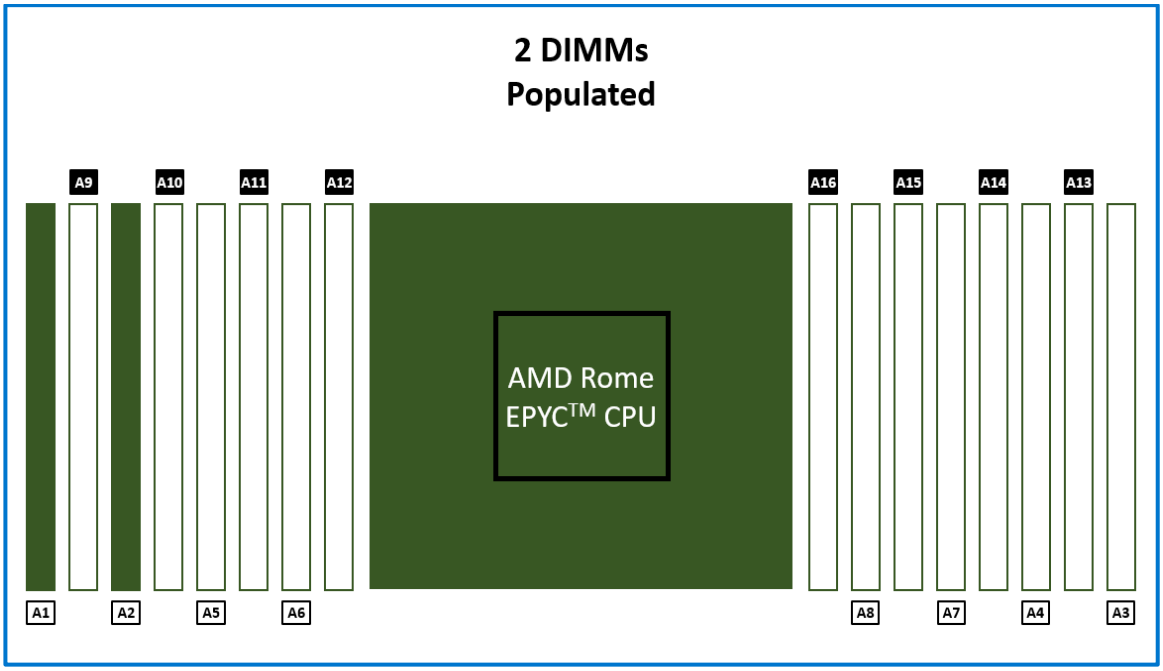

Figure 13: Two DIMMs are populated in an unbalanced configuration

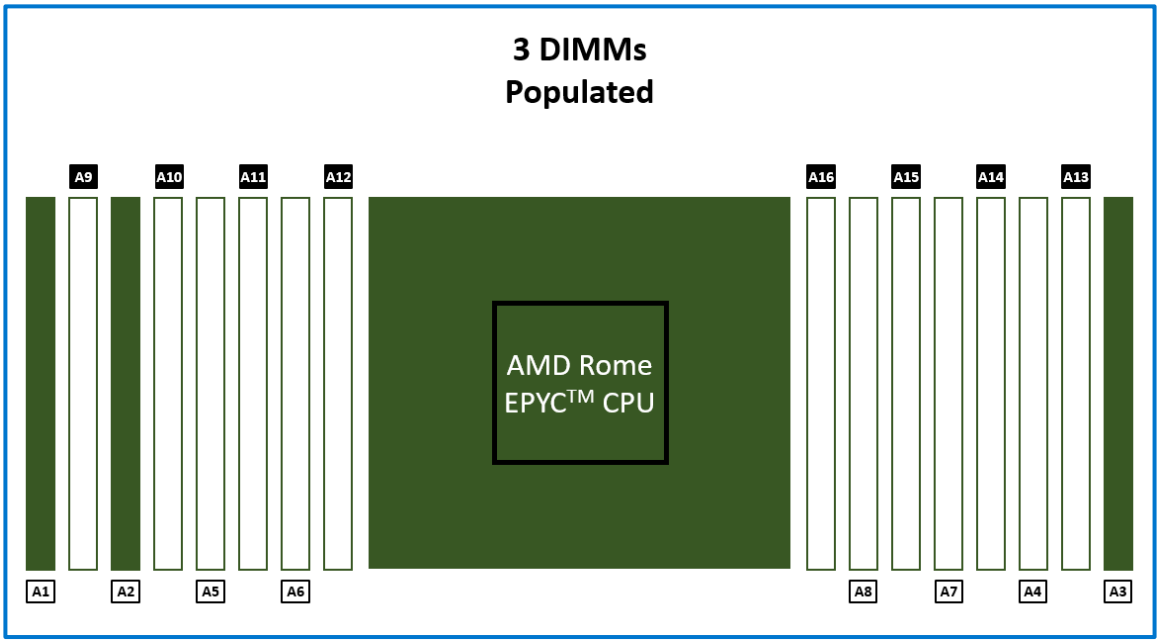

Figure 14: Three DIMMs are populated in an unbalanced configuration

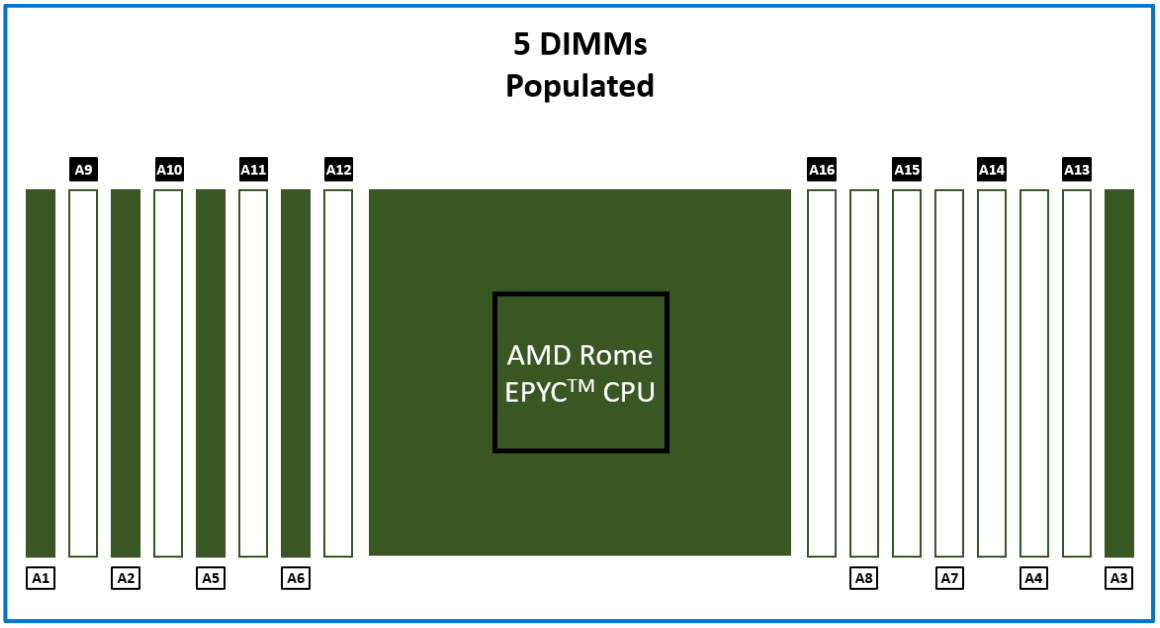

Figure 15: Five DIMMs are populated in an unbalanced configuration

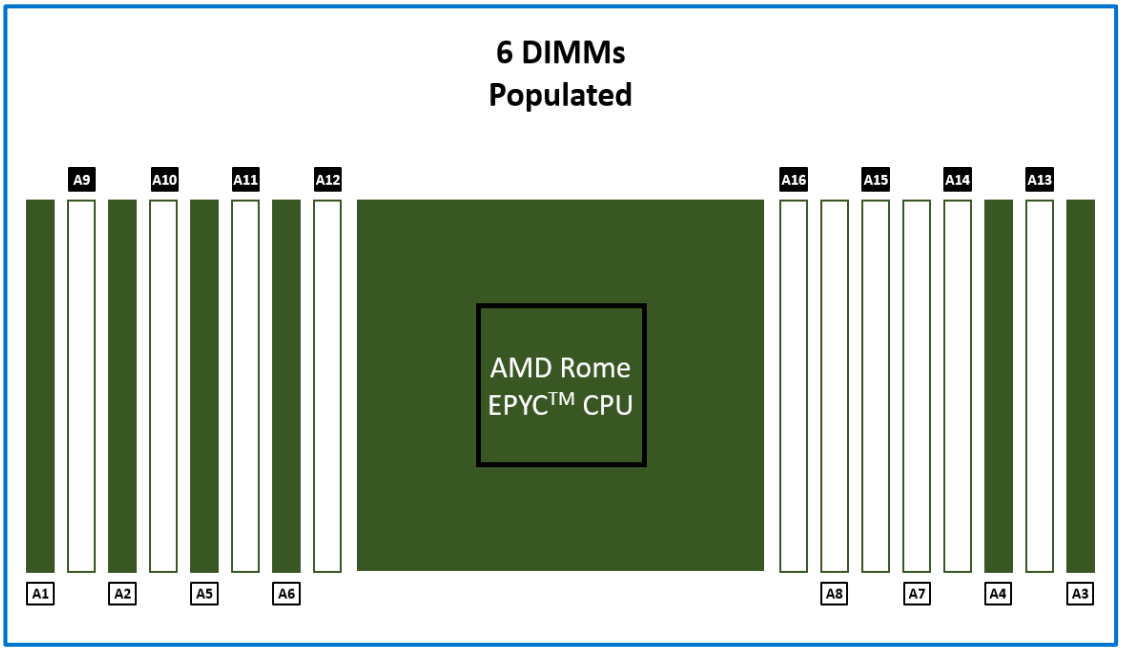

Figure 16: Six DIMMs are populated in a near balanced configuration

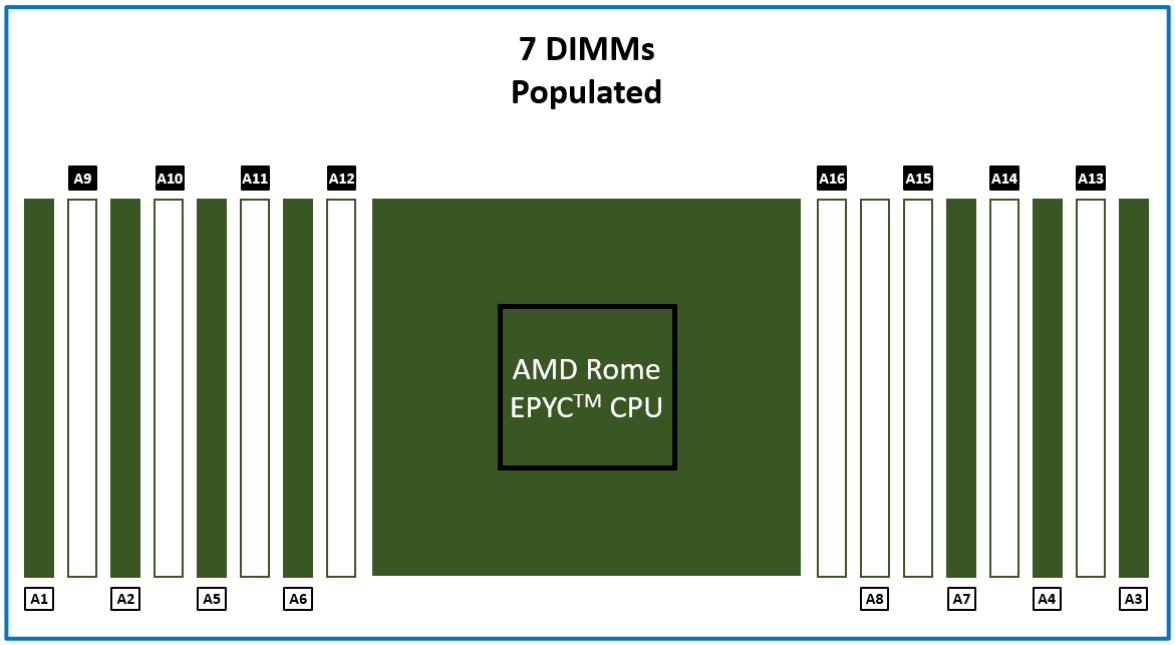

Figure 17: Seven DIMMs are populated in an unbalanced configuration

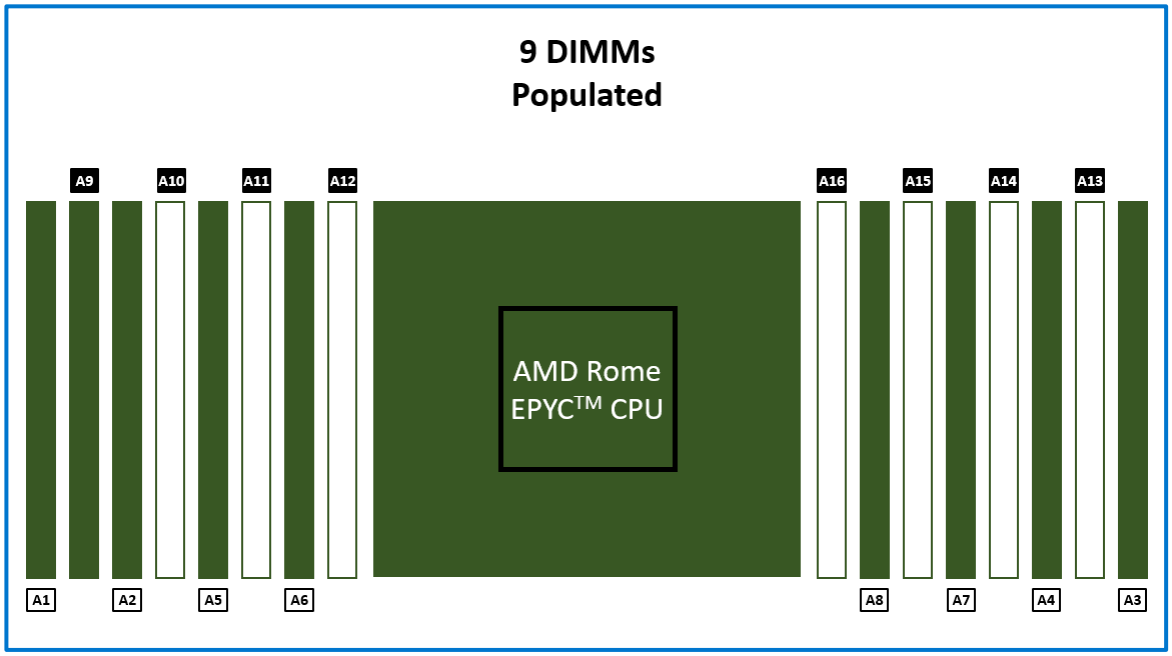

Figure 18: Nine DIMMs are populated in an unbalanced configuration

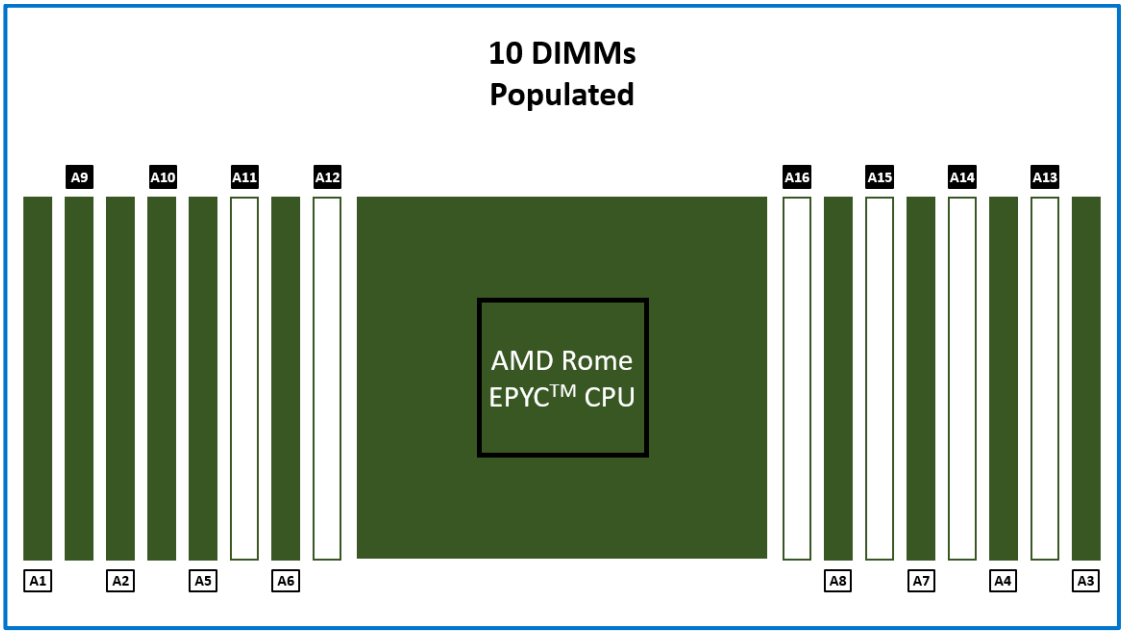

Figure 19: Ten DIMMs are populated in a near balanced configuration

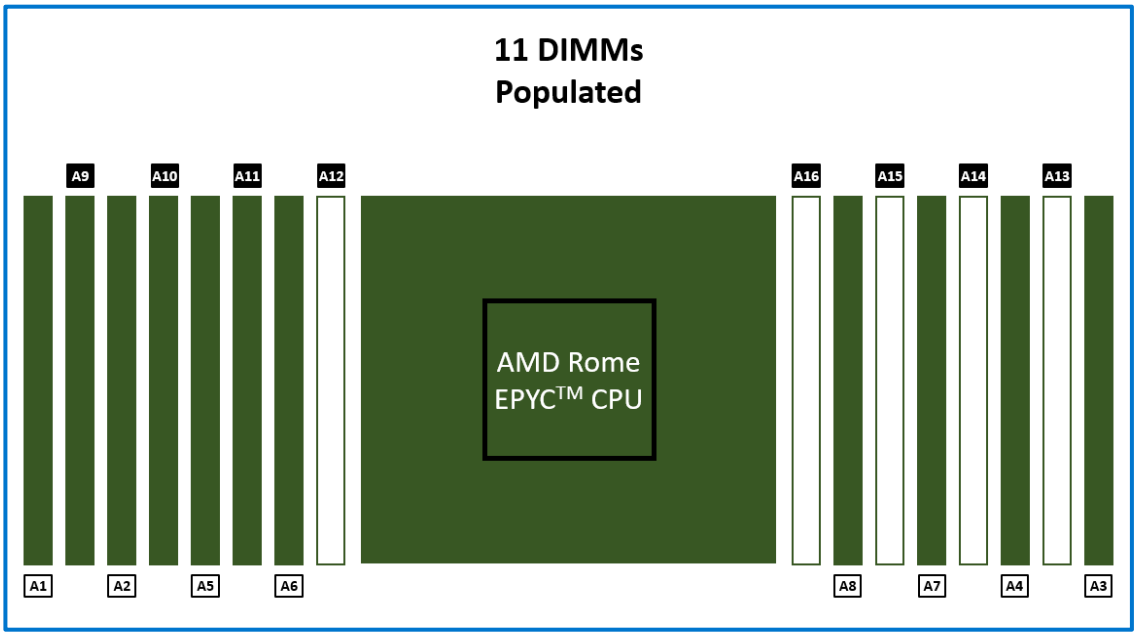

Figure 20: Eleven DIMMs are populated in an unbalanced configuration

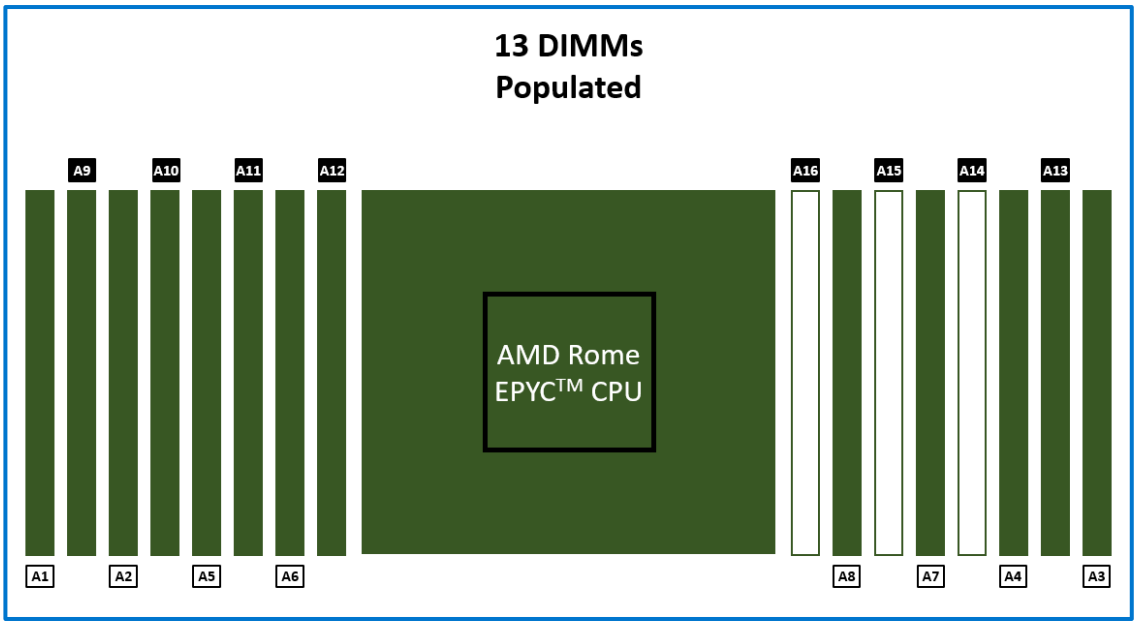

Figure 21: Thirteen DIMMs are populated in an unbalanced configuration

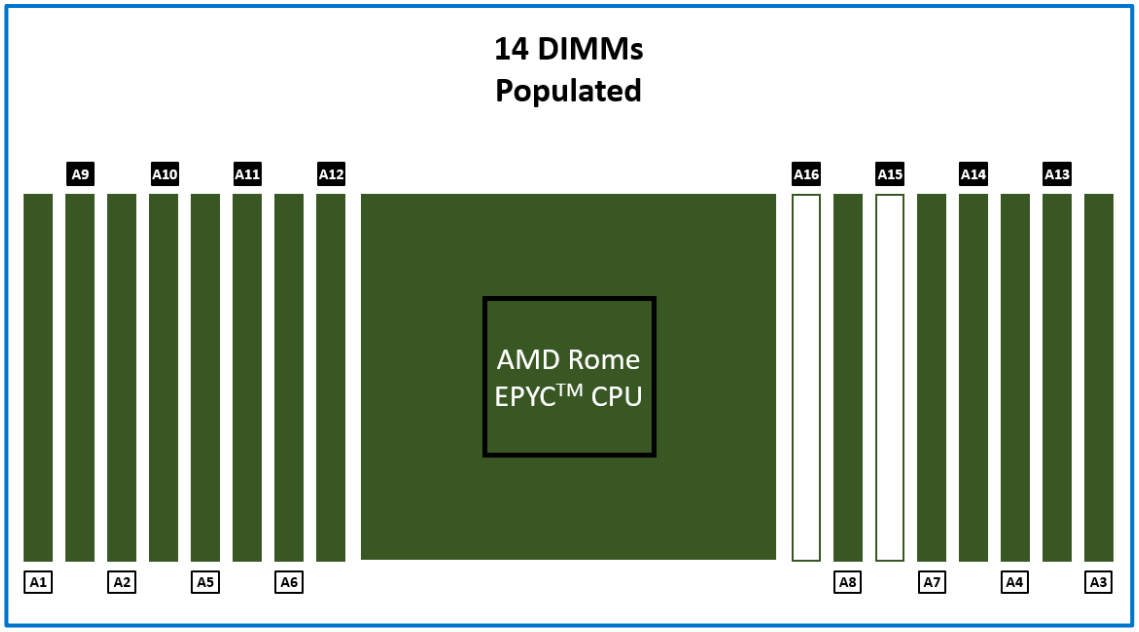

Figure 22: Fourteen DIMMs are populated in a near balanced configuration

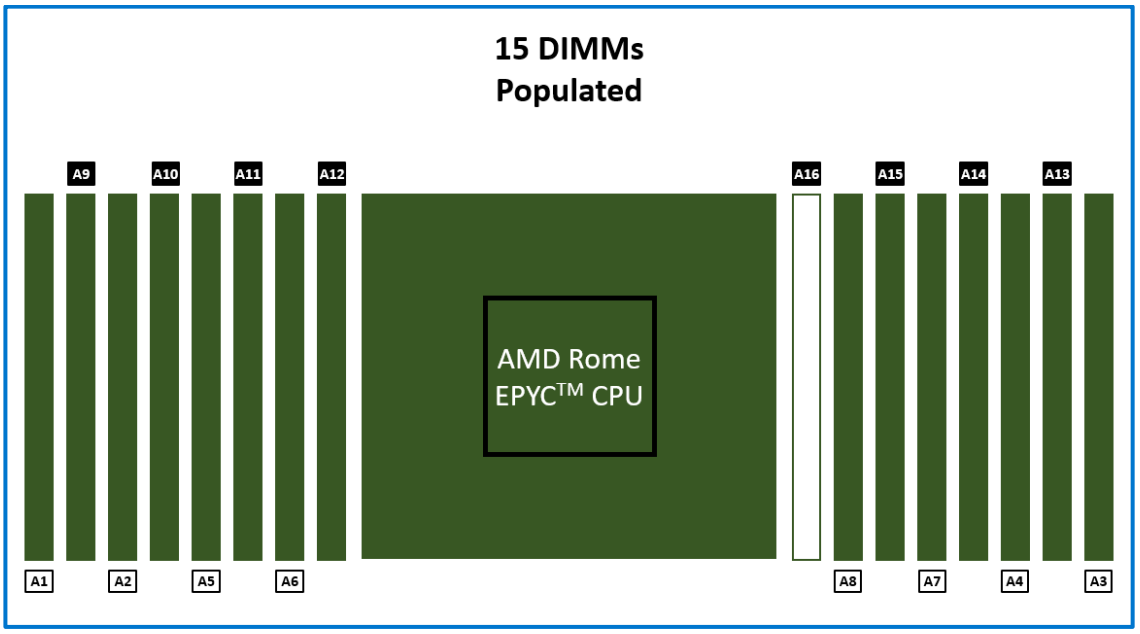

Figure 23: Fifteen DIMMs are populated in an unbalanced configuration

8. Conclusion

Balancing memory with 2nd Generation EPYC™ server processors increases memory bandwidth and reduces memory access latency. When memory modules are configured in such a way that the memory subsystems are identical, and channels are fully populated with one or two DIMMs, one interleave set will create a single uniform memory region that is spread across as many DIMMs as possible. This allows the distribution of data to perform most efficiently on Dell EMC PowerEdge servers.

Applying the balanced memory guidelines demonstrated in this brief will ensure that both memory bandwidth and memory access latency are optimized, therefore ensuring peak memory performance within Dell EMC PowerEdge servers.

9. References

1 https://developer.amd.com/wp-content/resources/56301_1.0.pdf

2 https://www.streetdirectory.com/travel_guide/124468/hardware/computer_memory_controllers_how_they_work.html

3 https://www.computerhope.com/jargon/d/dual-channel-memory.htm

4 https://www.computerhope.com/jargon/m/memoslot.htm

5 https://www.geeksforgeeks.org/memory-interleaving/

6 https://www.amd.com/system/files/2018-03/AMD-Optimizes-EPYC-Memory-With-NUMA.pdf

7 https://developer.amd.com/wp-content/resources/56301_1.0.pdf