lenovo

Optimizing Memory Performance of Intel Xeon E7 v2-based Servers

Introduces the architecture of the X6 servers that use these processors

Explains the testing methodology that is used to measure memory performance

Analyzes the effects the different settings and features have on performance

Provides best practices to maximize memory performance

Charles Stephan

Alicia Boozer

Abstract

This paper examines the architecture and memory performance of Lenovo System x and Flex System X6 platforms that are based on the Intel® Xeon® E7-8800, E7-4800, and E7-2800 v2 processors. These platforms can provide significant performance gains over predecessor products. The performance analysis in this paper covers memory latency, bandwidth, and application performance. In addition, the paper describes performance issues that are related to CPU frequency, memory speed, and population of memory DIMMs. Finally, the paper examines optimal memory configurations and best practices for the Lenovo X6 platforms.

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

For more information about our most recent publications, see this website:

http://lenovopress.com

Contents

Introduction

The Intel Xeon E7 v2 series of EX processors has up to 15 cores and 30 threads per processor socket. Based on 22-nanometer manufacturing technology, these processors are designed to enable systems to scale 1 - 8 processor sockets natively.

The E7 v2 series provides a common building block across the following Lenovo® platforms:

- Lenovo System x3850 X6, a 4U rack server scalable to four processors
- Lenovo System x3950 X6, an 8U rack server scalable to eight processors
- Lenovo Flex System™ X6 Compute Node family that consists of the x280 X6 (scalable to two processors), x480 X6 (scalable to four processors), and the x880 X6 (scalable to eight processors).

Table 1 lists the generational improvements that were introduced in the E7 v2 series over its predecessor.

Table 1 Intel Xeon E7 v2 series generational improvements

	E7 Series	**E7 v2 Series**
Process Technology	32 nm	22 nm
Thermal Design Power (TDP)	130 W, 105 W, 95 W	155 W, 130 W, 105 W
Memory Capacity	► Up to 16 DIMMs per socket ► 32 GB max DIMM capacity ► Up to 2 TB in 4S ► Up to 4 TB in 8S	► Up to 24 DIMMs per socket ► 64 GB max DIMM capacity ► Up to 6 TB in 4S ► Up to 12 TB in 8S
Memory Speed[a]	800, 977, 1066 MHz (DDR3)	1066, 1333, 1600 MHz (DDR3)
Cores/Threads	Up to 10/20 per socket	Up to 15/30 per socket
Last Level Cache Size	Up to 30 MB	Up to 37.5 MB
I/O Bandwidth	Up to 72 lanes PCIe 2.0 (dual IOH)	Up to 32 integrated PCIe 3.0 lanes per socket
Intel QPI Bandwidth	Up to 3x Intel QPI v1.0, 6.4 GT/s max	Up to 3x Intel QPI v1.1, 8.0 GT/s max

a. Memory speed depends on memory configuration and population rules, processor SKU, and memory mode (Independent or Lockstep).

Compared to its predecessors, the E7 v2 series offers higher core counts, more memory capacity, memory controller improvements, and snoop optimizations, which translates into lower memory latencies and higher memory bandwidth.

System architecture

This section describes the architecture of the Lenovo platforms, which support the E7 v2 series processors. Figure 1 shows the block diagram of an E7 v2 CPU.

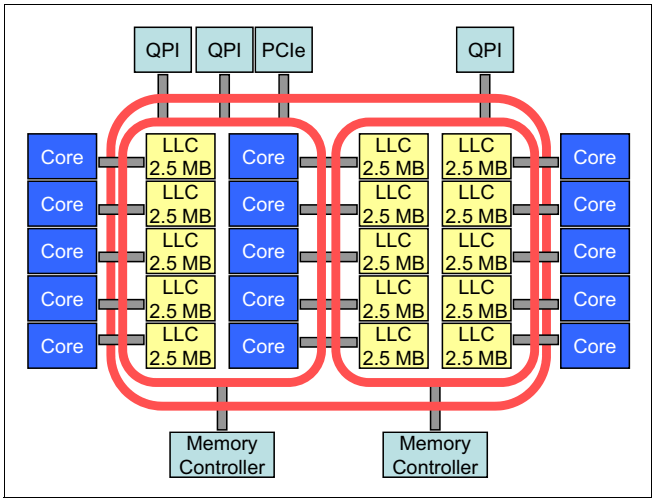

Figure 1 Block diagram of an E7-4890 v2 CPU

The E7 v2 series processors support two integrated memory controllers, four high-speed Scalable Memory Interface 2 (SMI2) links, and eight DDR3 memory channels. Unlike its predecessor (which supported up to 16 DIMMs per socket), the E7 v2 series supports up to 24 DIMMs per socket. The E7 v2 series introduces a high-speed, bidirectional ring that interconnects the processor cores and the uncore components, such as LLC, memory controller, PCI Express, and QPI. The high-speed ring is clocked at the processor core frequency.

Scalable memory buffers provide a bridge between the DDR3 memory channels and the SMI2 channels. The DDR3 memory frequency and the Intel Quick Path Interconnect 1.1 (QPI) rate depend on memory population, memory mode, and processor SKU. Supported QPI rates are 8.0 GT/s, 7.2 GT/s, and 6.4 GT/s.

Lenovo System x3850 X6 and x3950 X6

The Lenovo System x3850 X6 is the follow-on rack product to the x3850 X5™. The x3850 X6 can be configured as a 2- or 4-socket system, and supports up to 48 and 96 memory DIMMs for a maximum capacity of 3 TB and 6 TB. Figure 2 on page 5 shows the block diagram of the x3850 X6 4-socket system architecture.

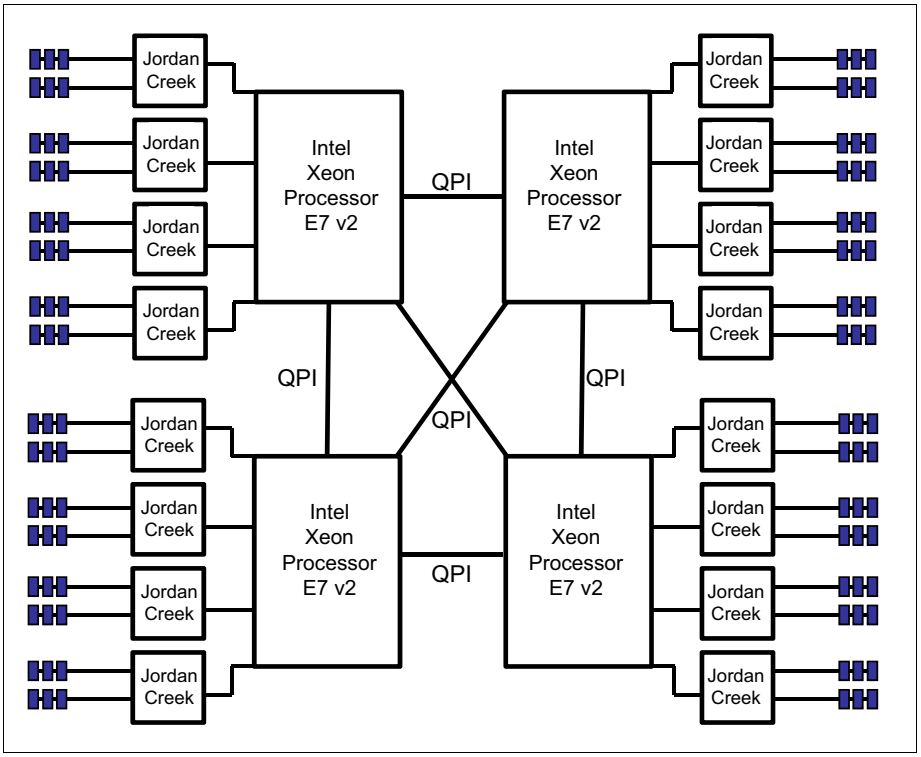

Figure 2 x3850 X6 system architecture

As shown in Figure 2, each processor is directly connected to one another via QPI links. The direct connections provide shorter latencies for processor communication, which can yield higher sustainable memory bandwidths. Not shown are the direct PCIe links that are associated with each socket.

The x3950 X6 consists of two x3850 X6s that are housed in a single chassis. The external scalability cables that are required for an 8-socket X5 were eliminated, and the scalability links are incorporated inside the chassis. The x3950 X6 supports up to 192 memory DIMMs for a maximum capacity of 12 TB. The processor and memory DIMMs for each socket are housed in compute books. Each compute book contains a processor and up to 24 memory DIMMs.

The x3850 X6 and x3950 X6 are shown in Figure 3.

Figure 3 Lenovo x3850 X6 and x3950 X6

Lenovo Flex System x280 X6, x480 X6, and x880 X6

The Flex System X6 Compute Node family includes the three X6 compute nodes: x280 X6, x480 X6, and x880 X6. These servers are based on a two-socket, double-wide compute node and differ only by the following processor types that are installed and the degree to which they can scale up:

- The Flex System x880 X6 Compute Node uses processors of Intel Xeon E7-8800 v2 family, can scale up to 8-socket, and supports 8-socket (consists of four connected compute nodes), 4-socket (consists of two connected compute nodes) and 2-socket (one compute node) configurations.
- The Flex System x480 X6 Compute Node uses processors of Intel Xeon E7-4800 v2 family, can scale up to 4-socket, and supports 4-socket and 2-socket configurations.
- The Flex System x280 X6 Compute Node uses processors of Intel Xeon E7-2800 v2 family and supports 2-socket configurations only.

Figure 4 shows the block diagram of the two-socket, double-wide compute node that forms the basis of the Flex System X6 Compute Nodes.

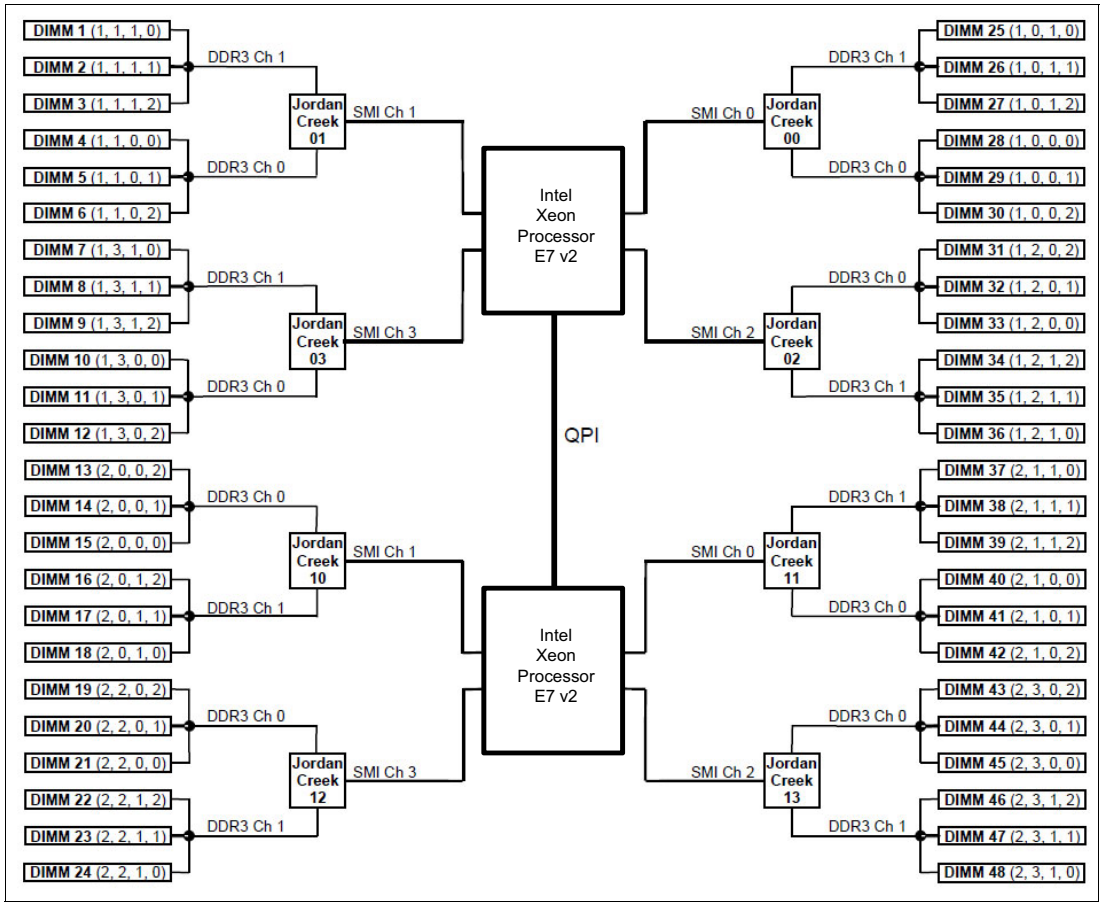

Figure 4 Flex System X6 system architecture

With the x480 X6 and x880 X6, the compute nodes are connected by using a front-mounted interconnect system that joins the QPI links of the processors.

The Flex System X6 models are shown in Figure 5.

Figure 5 Lenovo X6 Flex System family

Memory performance

Memory performance in an Intel Xeon E7 v2 based server depends on different variables, such as memory mode, CPU SKU, memory speed, memory ranks, and memory population.

Measurement configuration

The configuration that is used for the performance data in this paper is listed in Table 2.

Table 2 Performance measurements configuration

Component	Description
System	Lenovo System x3850 X6
Processor	2x E7-4890 v2 (2.8 GHz, 155 W, QPI 8.0 GT/s, 1600 MHz capable)
Memory	► 8 GB (1600 MHz, 1Rx4, 1.35 V) RDIMM ► 16 GB (1600 MHz, 2Rx4, 1.35 V) RDIMM ► 32 GB (1600 MHz, 4Rx4, 1.35 V) LRDIMM

Component	Description
UEFI Settings	► Operating Mode: Maximum Performance ► C-states: Disabled ► C1E: Disabled ► Turbo Mode: Enabled ► Hyperthreading: Enabled
Operating System	Red Hat Enterprise Linux 6 Update 5 x64 Edition

To measure low-level memory performance metrics, an internal Lenovo memory tool was used that accurately measures memory throughput and memory latency. In all of the memory latency figures, lower numbers are better; in all memory throughput figures, higher numbers are better.

The following industry standard applications were also used to measure memory performance:

- SPECint2006_rate_base

  Used as an indicator of performance for commercial applications. This application often is more sensitive to processor frequency and less to memory bandwidth.

- SPECfp2006_rate_base

  Used as an indicator of High Performance Computing (HPC) performance. This application often is more sensitive to memory bandwidth.

- STREAM

  A benchmark that consists of four different memory workloads; however, the data in this paper corresponds to the Triad component. The Triad component of STREAM consists of two read operations and one write operation from the application's perspective.

This paper includes information that is primarily focused on optimal memory performance. For more information about the Lenovo X6 products, see the following publications:

- *Lenovo System x3850 X6 and x3950 X6 Planning and Implementation Guide*, SG24-8208, which is available at this website:

  http://lenovopress.com/sg248208

- *Lenovo Flex System X6 Compute Node Planning and Implementation Guide*, SG24-8227, which, is available at this website:

  http://lenovopress.com/sg248227

Memory modes

The E7 v2 memory controllers support Lockstep, Independent, Mirroring, and Sparing modes. There can be performance ramifications, depending on the memory mode selected. The memory mode settings in the Unified Extensible Firmware Interface (UEFI) are shown in Figure 6.

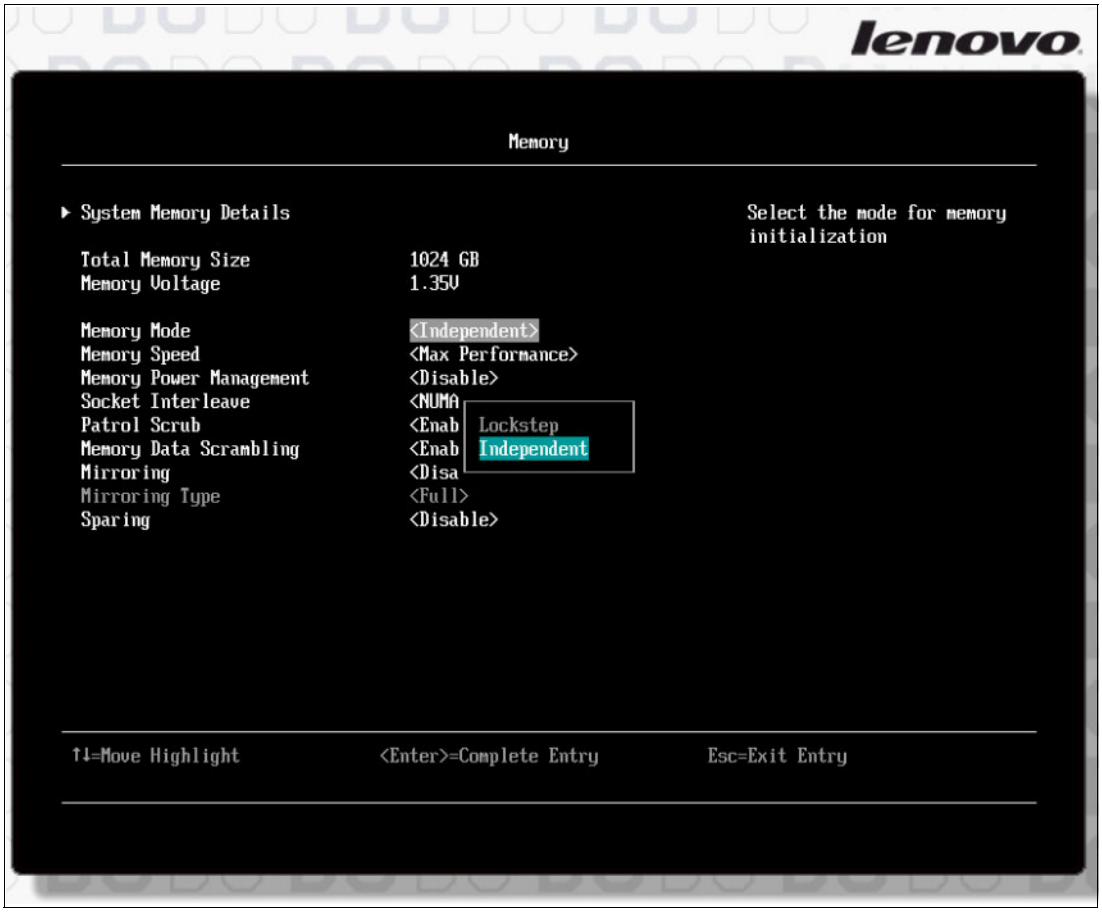

Figure 6 Memory Mode setting in UEFI setup

Independent memory mode

In Independent mode, each DDR3 channel is addressed individually. The SMI2 link interleaves the data from two DDR3 channels and the scalable memory buffer separates the data. In this mode, the SMI2 channel operates at twice the DDR3 data rates. Memory channels can be populated with DIMMs in any order in Independent mode. All eight DDR3 channels per processor can be populated in any order and have no matching requirements. Independent mode should be selected for best memory performance in most production environments.

Best practices for optimal memory performance should follow the DIMM installation order that is listed in Table 3 on page 10 (x3850 X6 and x3950 X6) and Table 4 on page 10 (Flex System X6), regardless of what memory mode is selected. DIMM configurations of 1 DPC, 2 DPC, or 3 DPC should be used for optimal memory performance. Memory configurations that deviate from 1 DPC, 2 DPC, or 3 DPC can result in memory performance degradation.

Table 3 x3850/x3950 X6 Compute Book optimal DIMM installation

Memory Configuration	DIMM Population Order	DIMM Slots to Populate per Compute Book
1 DPC (8x DIMMs)	Pair 1	9 and 15
	Pair 2	6 and 24
	Pair 3	1 and 19
	Pair 4	10 and 16
2 DPC (16x DIMMs)	Pair 5	8 and 14
	Pair 6	5 and 23
	Pair 7	2 and 20
	Pair 8	11 and 17
3 DPC (24x DIMMs)	Pair 9	7 and 13
	Pair 10	4 and 22
	Pair 11	3 and 21
	Pair 12	12 and 18

Table 4 Flex System X6 Compute Node optimal DIMM installation

Memory Configuration	DIMM Population Order	DIMM Slots to Populate per Compute Node
1 DPC (16x DIMMs)	Pair 1	25 and 28
	Pair 2	45 and 48
	Pair 3	7 and 10
	Pair 4	15 and 18
	Pair 5	1 and 4
	Pair 6	21 and 24
	Pair 7	33 and 36
	Pair 8	37 and 40
2 DPC (32x DIMMs)	Pair 9	26 and 29
	Pair 10	44 and 47
	Pair 11	8 and 11
	Pair 12	14 and 17
	Pair 13	2 and 5
	Pair 14	20 and 23
	Pair 15	32 and 35
	Pair 16	38 and 41

Memory Configuration	DIMM Population Order	DIMM Slots to Populate per Compute Node
3 DPC (48x DIMMs)	Pair 17	27 and 30
	Pair 18	43 and 46
	Pair 19	9 and 12
	Pair 20	13 and 16
	Pair 21	3 and 6
	Pair 22	19 and 22
	Pair 23	31 and 34
	Pair 24	39 and 42

Lockstep memory mode

In Lockstep mode, the memory controller uses two DDR3 channels at the same time behind a single memory buffer, which splits a cache line across both channels. In this mode, the SMI2 channel operates at the DDR3 transfer rate. Lockstep mode provides the highest reliability, availability, and serviceability (RAS) features. Paired channels should have the same configuration. Lockstep memory mode does not yield the best memory performance for the system in most cases.

Memory Mirroring mode

Mirroring mode is supported in Independent and Lockstep modes. When Independent mode is selected, DDR3 channel 0 on SMI2 link 0 is mirrored with DDR3 channel 0 on SMI2 link 1, and DDR3 channel 1 on SMI2 link 0 is mirrored with DDR3 channel 1 on SMI2 link 1. The same pattern is true for SMI2 links 2 and 3 on the second memory controller. In Independent mode, the DDR3 channels are operating independently from one another.

When Lockstep mode is selected, DDR3 channels 0 and 1 on SMI2 link 0 are mirrored with DDR3 channels 0 and 1 on SMI2 link 1. This pattern is also true for the DDR3 channels on SMI2 links 2 and 3. In Lockstep mode, the DDR3 channels are operating together in lockstep. Regardless of Independent or Lockstep mode, the total available memory to the system is half of the physical memory that is installed.

When Mirrored Memory mode is selected, memory on CPU socket 0 and CPU socket 1 must be populated with memory that all have the same feature set, such as size, organization, and ranks. The memory channels can have memory with different feature sets, but the same memory DIMM slots across CPU sockets must be populated with memory DIMMs that have the same feature set.

Rank Sparing mode

As with Mirrored mode, Rank Sparing mode can be implemented with Independent and Lockstep modes. Used with Independent mode, each DDR3 bus has its own spare rank. However, in Lockstep mode, a rank-pair is selected across the DDR3 busses off an SMI2 link.

The spare rank or rank-pair is held in reserve and is not used by the system as part of its system memory. The memory sparing algorithm allocates a spare rank of memory from another of the installed memory DIMMs or DIMM-pair to use as active memory if a certain threshold of correctable errors occurs on a rank or rank-pair of memory. The spare rank must have identical or larger memory capacity than all the other ranks on the same channel. After sparing, the sparing source rank is lost.

For more information about memory modes, see the following publications:

- *Lenovo System x3850 X6 and x3950 X6 Planning and Implementation Guide*, SG24-8208, which is available at this website:

  http://lenovopress.com/sg248208

- *Lenovo Flex System X6 Compute Node Planning and Implementation Guide*, SG24-8227, which, is available at this website:

  http://lenovopress.com/sg248227

Memory Mode effect on memory latency and STREAM Triad bandwidth

The loaded latency performance as a function of Lockstep and Independent modes is shown in Figure 7.

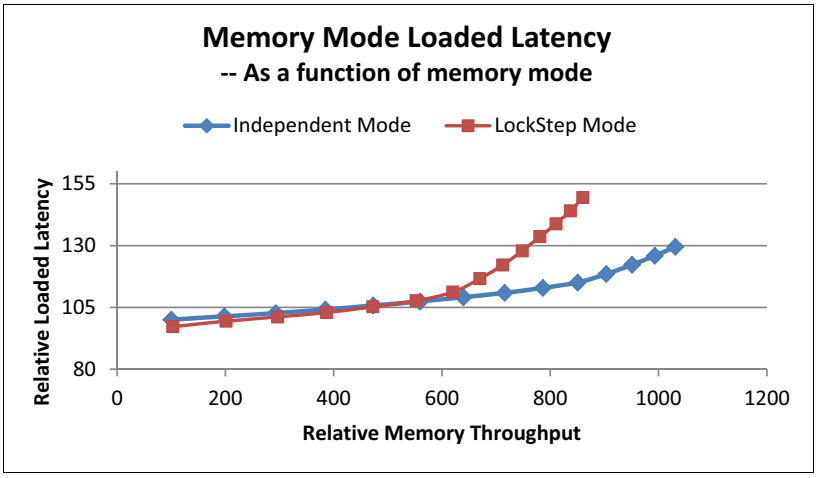

Figure 7 Loaded latency as a function of memory modes

At light loads, Lockstep mode can yield slightly lower latencies than Independent mode. However, as the load is increased from left to right (as shown Figure 7), Independent mode delivers lower latencies and greater memory throughput than Lockstep mode.

Figure 8 shows unloaded latency performance as a function of memory modes. In this case, the Mirrored and Sparing modes were implemented in Lockstep mode.

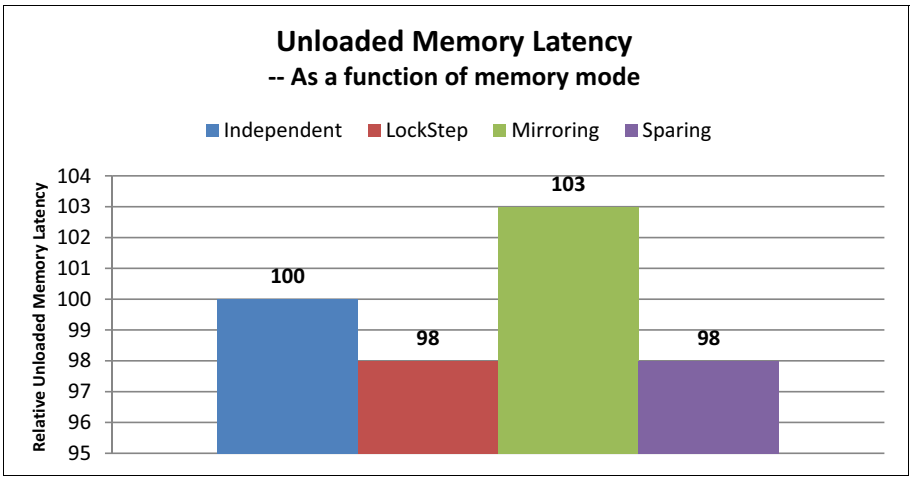

Figure 8 Unloaded latency as a function of memory mode

Memory sparing does not affect system memory latency; however, memory mirroring increases system latency by over 5% when compared to Lockstep mode without any redundancy features implemented.

Figure 9 shows the relative STREAM Triad memory bandwidth for various memory modes. There is a 34% decrease in STREAM Triad throughput when moving from normal Lockstep mode to Lockstep-Mirrored mode. There is a 32% decrease when moving from Lockstep to Lockstep-Sparing mode.

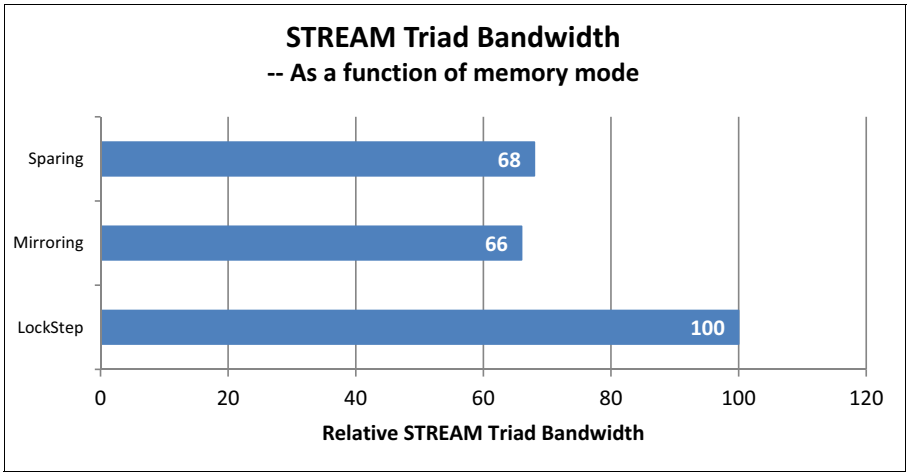

Figure 9 STREAM Triad Bandwidth as a function of memory mode

Memory speed

Memory speed is one of the most critical factors that affects memory performance. It is important to understand the performance characteristics when memory frequency is changed and the factors that control memory speed in a particular architecture.

Factors controlling memory frequency

Memory frequency is limited by the maximum common frequency that is supported by the processors and the DIMMs. In addition, the memory frequency can be set lower by the user. These factors are described next.

Processor type

The E7 v2 series of processors includes 2-, 4-, and 8-socket capable processors that are designated as Advanced, Standard, Basic, and Segment Optimized. Table 5 lists the categories and attributes of each processor.

Table 5 Intel Xeon E7 v2 processor categories and attributes

Category	Processor	TDP (Watts)	Core Freq. (GHz)	Core Count	Cache Size (MB)	QPI Speed (GT/s)	Maximum Memory Frequency (MHz)
Advanced	E7-8890	155	2.8	15	37.5	8	1600
	E7-8880	130	2.5	15	37.5	8	1600
	E7-8870	130	2.3	15	30	8	1600
	E7-4890	155	2.8	15	37.5	8	1600
	E7-4880	130	2.5	15	37.5	8	1600
	E7-4870	130	2.3	15	30	8	1600
	E7-4860	130	2.6	12	30	8	1600
	E7-2890	155	2.8	15	37.5	8	1600
	E7-2880	130	2.5	15	37.5	8	1600
	E7-2870	130	2.3	15	30	8	1600
Standard	E7-8850	105	2.3	12	24	7.2	1600
	E7-4850	105	2.3	12	24	7.2	1600
	E7-4830	105	2.2	10	20	7.2	1600
	E7-4820	105	2.0	8	16	7.2	1600
	E7-2850	105	2.3	12	24	7.2	1600
Basic	E7-4809	105	1.9	6	12	6.4	1333
Segment Optimized	E7-8891	155	3.2	10	37.5	8	1600
	E7-8893	155	3.4	6	37.5	8	1600
	E7-8880L	105	2.2	15	37.5	8	1600
	E7-8857	130	3.0	12	30	8	1600

DIMM Frequency

The DIMM type is another factor that controls the maximum memory frequency. The DDR3 DIMMs that are available for the Lenovo X6 platforms support the following frequencies:

- 1600 MHz
- 1333 MHz
- 1066 MHz

The maximum memory frequency is the lower of the DIMM frequency and the maximum frequency that is supported by the processor.

Table 6 lists the maximum supported memory frequencies for Independent mode.

Table 6 Supported maximum memory frequencies for Independent mode

		Independent mode speed/Voltage that is supported by DIMMs per channel					
DIMM by Rank, Type, Technology	**DIMM Capacity**	**1DPC**		**2DPC**		**3DPC**	
		1.35 V	1.50 V	1.35 V	1.50 V	1.35 V	1.50 V
1Rx4, RDIMM, 2 Gb	4 GB	1333	1333	1333	1333	1066	1333
1Rx4, RDIMM, 4 Gb	8 GB	1333	1333	1333	1333	1066	1333
2Rx4, RDIMM, 4 Gb	16 GB	1333	1333	1333	1333	1066	1333
4Rx4, LRDIMM, 4 Gb	32 GB	1333	1333	1333	1333	1333	1333
8Rx4, LRDIMM, 4 Gb	64 GB	1333	1333	1333	1333	1333	1333

Table 7 lists the maximum supported memory frequencies for Lockstep mode.

Table 7 Supported maximum memory frequencies for Lockstep mode

		Independent mode speed/voltage that is supported by DIMMs per channel					
DIMM by Rank, Type, Technology	**DIMM Capacity**	**1DPC**		**2DPC**		**3DPC**	
		1.35 V	1.50 V	1.35 V	1.50 V	1.35 V	1.50 V
1Rx4, RDIMM, 2 Gb	4 GB	1333	1600	1333	1600	1066	1333
1Rx4, RDIMM, 4 Gb	8 GB	1333	1600	1333	1600	1066	1333
2Rx4, RDIMM, 4 Gb	16 GB	1333	1600	1333	1600	1066	1333
4Rx4, LRDIMM, 4 Gb	32 GB	1333	1600	1333	1600	1333	1600
8Rx4, LRDIMM, 4 Gb	64 GB	1333	1333	1333	1333	1333	1333

System Settings

Memory can be set to a frequency lower than the platform maximum by clicking **System Settings** → **Memory** → **Memory Speed** in the system's UEFI shell. Memory frequency often is set lower to save energy in environments with little memory performance sensitivity.

Figure 10 shows the UEFI shell window with the Memory Speed setting, which includes the options Minimal Power, Balanced, or Max Performance.

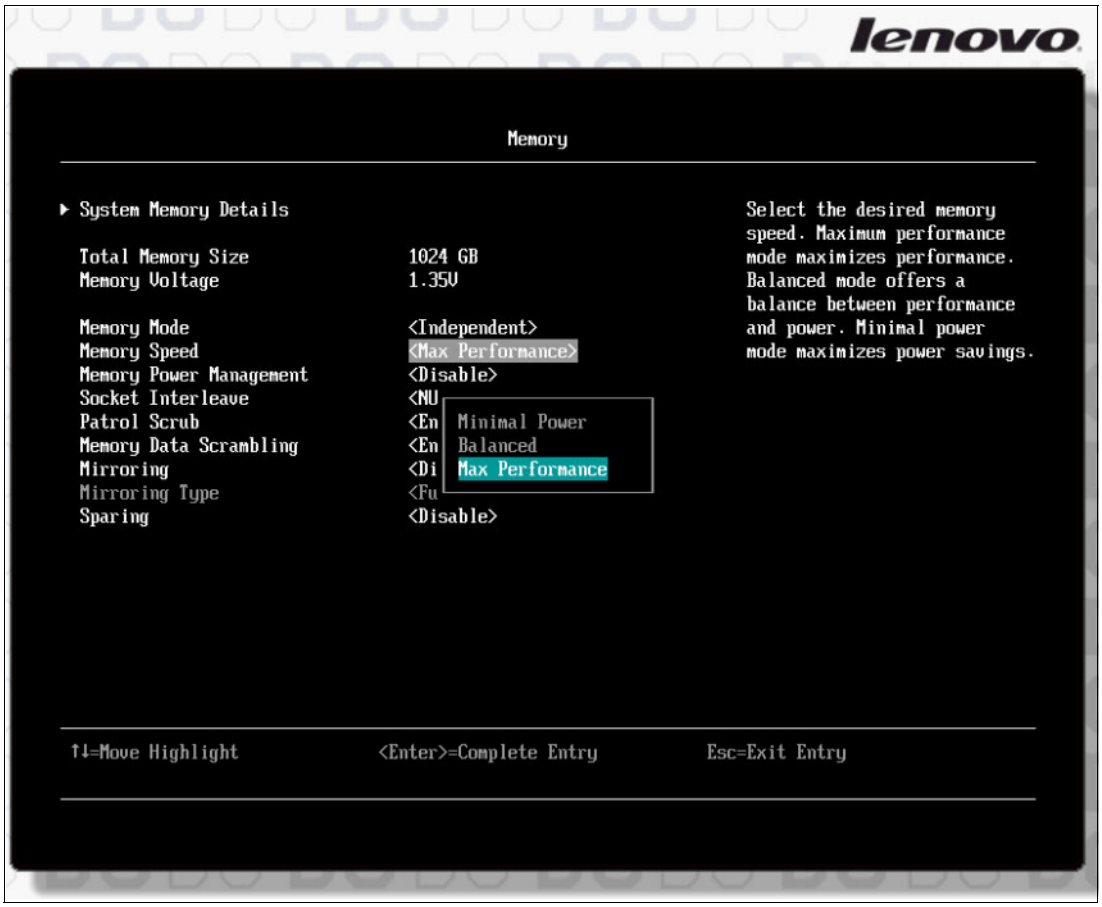

Figure 10 Memory speed setting in UEFI setup

The following Memory Speed options are available:

- Maximum Performance

  Memory runs at the maximum speed as determined by processor SKU and the memory subsystem. Memory voltage is forced to the minimum voltage that is needed to run at the maximum speed.

- Balanced

  Memory runs at one step below the maximum speed. Memory voltage is always set to lowest supported value.

- Minimal Power

  Memory runs at the lowest speed that is allowed by the architecture. Memory voltage is always set to lowest supported value.

Processor and memory frequency effects on memory performance

This section describes the effects on memory performance because of processor frequency and memory frequency. All data that is shown corresponds to Independent Mode, unless otherwise noted.

Figure 11 shows unloaded memory latency as a function of processor and memory frequency.

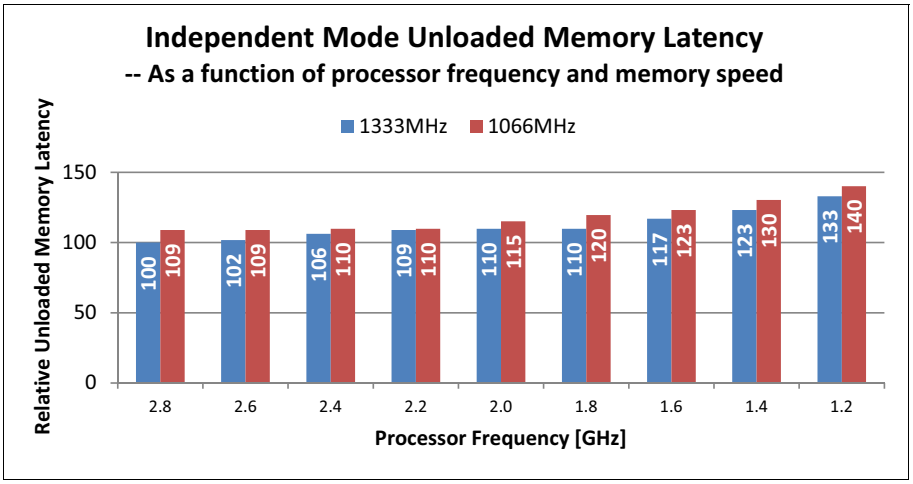

Figure 11 Unloaded memory latency as a function of processor and memory frequency

The rings that are shown in the E7 v2 block diagram in Figure 1 on page 4 operate at the processor core frequency, and the latency is dominated by this factor. As the processor core frequency decreases from left to right (as shown in Figure 11), the memory latency increases significantly. The same effects on memory latency for Lockstep mode are shown in Figure 12.

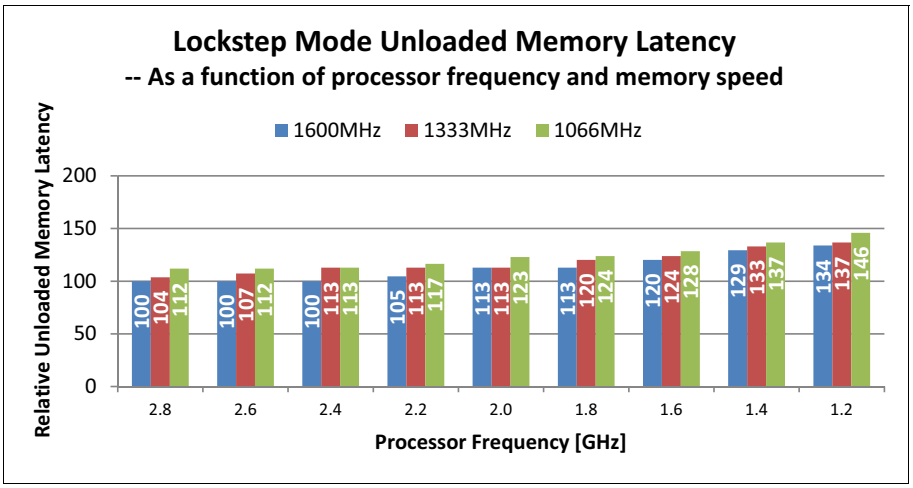

Figure 12 Lockstep mode unloaded memory latency

Figure 13 shows STREAM Triad memory bandwidth as a function of processor and memory frequency.

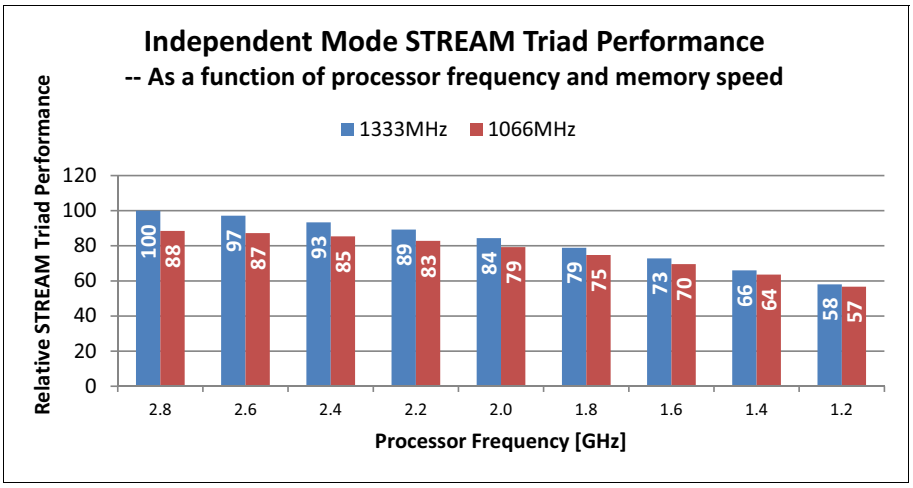

Figure 13 STREAM Triad performance

As the processor frequency is decreased, the memory throughput decreases. The extent to which memory throughput decreases depends on the memory frequency. At 1066 MHz, there is less sensitivity to core frequency because the performance is limited by the throughput that the memory channels can sustain. However, the drop in memory throughput is greater at 1333 MHz. In this case, the performance is limited more by the processor ring and less by the memory channels. The same effects on STREAM Triad bandwidth in Lockstep mode are shown in Figure 14.

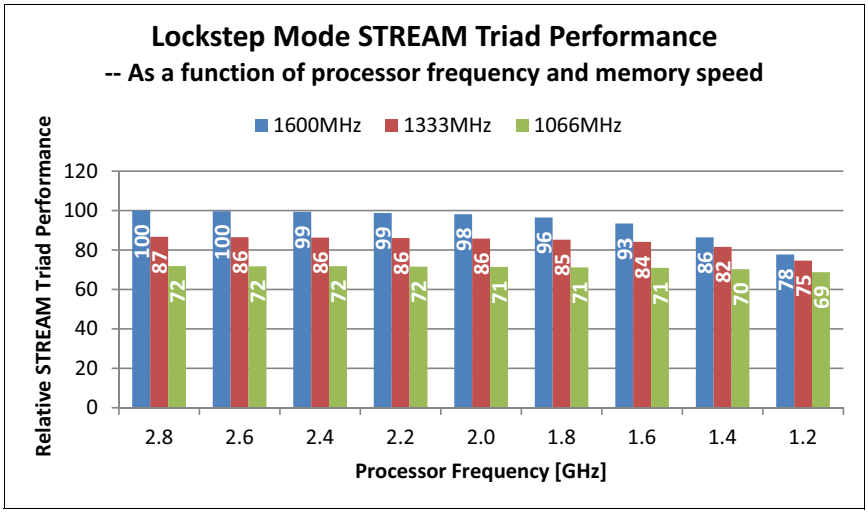

Figure 14 Lockstep mode STREAM Triad performance

Memory speed effects on applications

Figure 15 on page 19 shows the affect on application performance when the processor frequency is kept constant. As expected, the lower memory speed of 1066 MHz does not affect SPECint2006_rate_base as much as SPECfp2006_rate_base because SPECint2006_rate_base is more sensitive to processor frequency rather than memory speed.

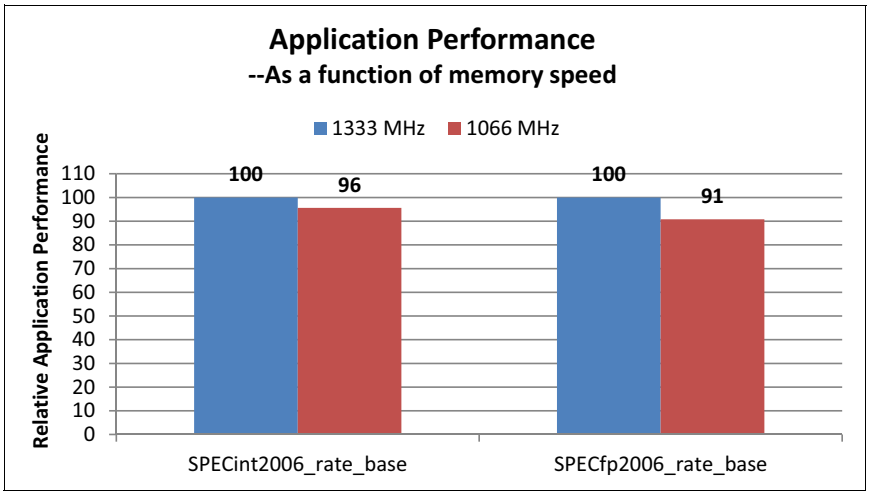

Figure 15 Application performance as a function of memory frequency

DIMM types

The E7 v2 based System x® and Flex System servers support RDIMMs and LRDIMMs. The following sections describe each type of DIMM and show the performance implications of using each type of DIMM.

Registered DIMMs

Registered DIMMs (RDIMMs) are the most prevalent DIMMs that are used in servers. RDIMMs use a register between the memory controller and dynamic random-access memory (DRAM) devices to buffer the address and control signals, which enables the reduction of electrical loading on the memory bus. This configuration allows the memory controller to support more DIMMs and a higher memory frequency, which provides scalability and greater performance.

Load reduced DIMMs

Load reduced DIMMs (LRDIMMs) reduce the electrical loading on the memory bus while maintaining larger capacities than RDIMMs. The register that is used by RDIMMs is replaced with a buffer on LRDIMMS, which isolates address, command, and data signals from the memory controller.

LRDIMMs use a technique that is called *rank multiplication* to work around the chip select limitation per DDR3 channel. Rank multiplication presents many ranks on a DIMM as a smaller number of ranks to the memory controller. For example, a quad-rank LRDIMM appears as a dual-rank memory module to the memory controller. This appearance allows LRDIMMs in the system to achieve a larger memory capacity while maintaining high performance, although with a slightly higher latency. LRDIMMs are targeted at memory capacities that cannot be achieved by using RDIMMs.

Effects of DIMM type on memory latency and bandwidth

Figure 16 shows the relative loaded memory latency and bandwidth of RDIMMs and LRDIMMs. The data that is shown Figure 16 in was achieved by applying a load incrementally across cores as the latency is measured. In the chart, the load applied increases from left to right, and the performance of both types of DIMMs is similar. At light loads, the LRDIMMs loaded latency is approximately 2 - 4% longer than RDIMMs. At heavier loads, the LRDIMMs loaded latency is only approximately 2% longer than RDIMMs.

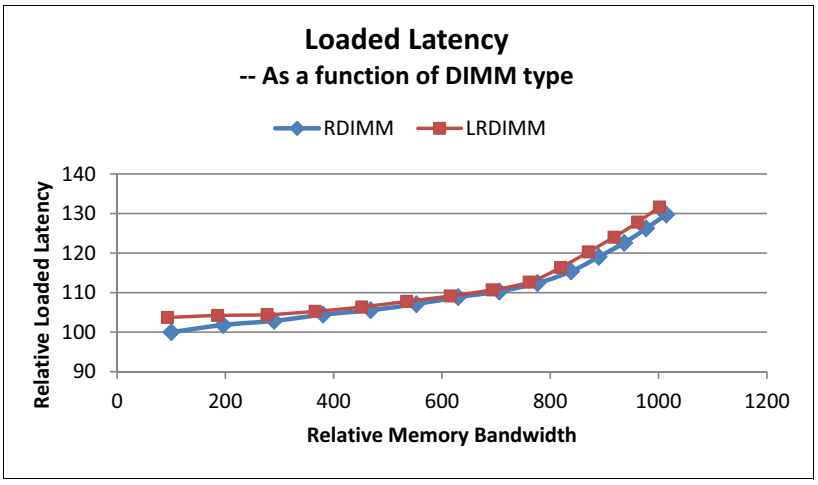

Figure 16 Loaded latency as a function of DIMM type

Figure 17 shows STREAM Triad memory bandwidth for both types of DIMMs for various memory configurations. As with the loaded latency chart, the STREAM Triad bandwidth of the LRDIMMs is at most only 4% lower than RDIMMs.

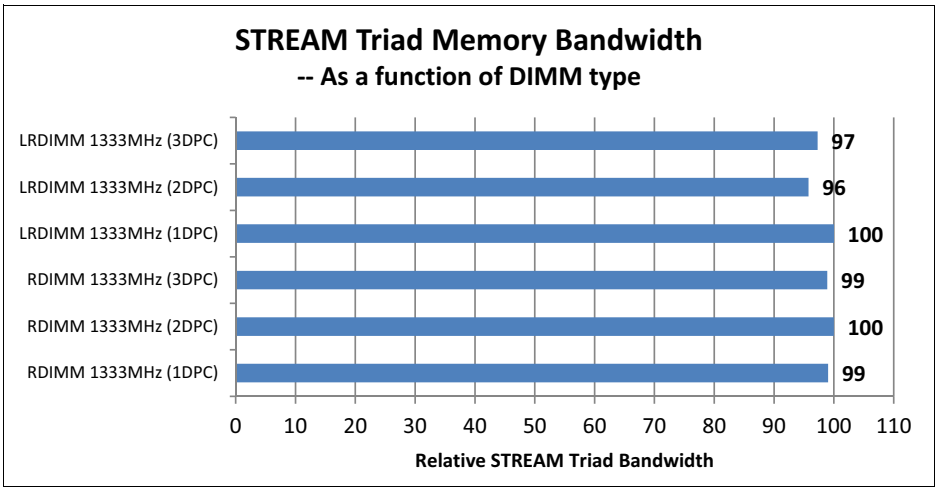

Figure 17 STREAM Triad memory bandwidth as a function of DIMM type

Ranks per channel

The number of ranks per channel (RPC) affects memory performance to a certain extent. In this section, we describe how this process works. The E7 v2 series and Jordan Creek memory buffers support 8RPC.

The following DIMM configurations were used in this section:

- 1 rank per channel: 1x 8 GB 1Rx4 at 1333 MHz
- 2 ranks per channel: 1x 16 GB 2Rx4 at 1333 MHz
- 3 ranks per channel: 1x 16 GB 2Rx4 + 1x 8 GB 1Rx4 at 1333 MHz
- 4 ranks per channel: 2x 16 GB 2Rx4 at 1333 MHz
- 6 ranks per channel: 3x 16 GB 2Rx4 at 1333 MHz

For the measurement results that are included in the following two sections, enough memory was provided to the processors so that the memory capacity differences among the configurations did not affect the results.

DIMM ranks and STREAM Triad

Figure 18 shows the STREAM Triad memory bandwidth performance as a function of RPC. Several observations are apparent. First, transitioning from 1RPC to 2RPC can gain over 23% more memory bandwidth performance. Second, a memory configuration with an odd number of RPC is not optimal. Finally, there is essentially no drop in performance when every memory channel is populated with three RDIMMs.

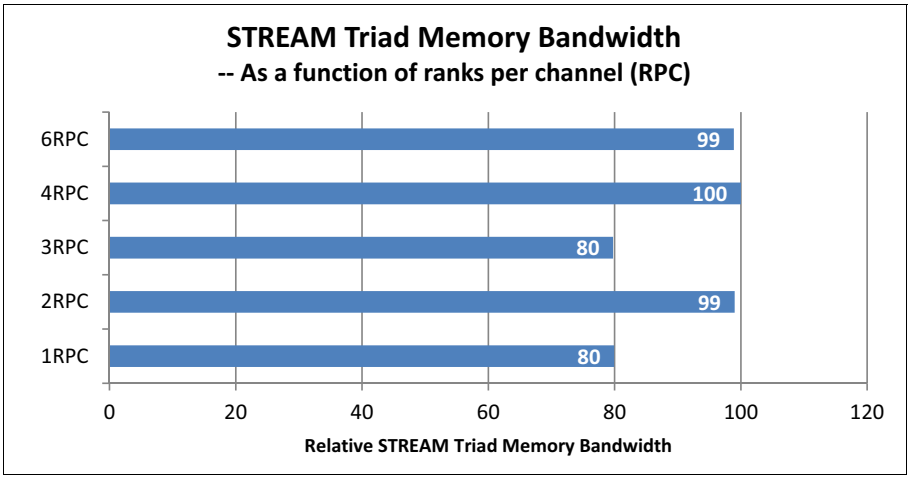

Figure 18 STREAM Triad memory bandwidth as a function of RPC

DIMM Ranks and Applications Performance

Figure 19 on page 22 shows application performance as a function of RPC, and indicates SPECint2006_rate_base is less sensitive to memory configuration than SPECfp2006_rate_base. However, populating DDR3 channels with an even number of ranks provides optimal performance in integer- and floating point-sensitive environments.

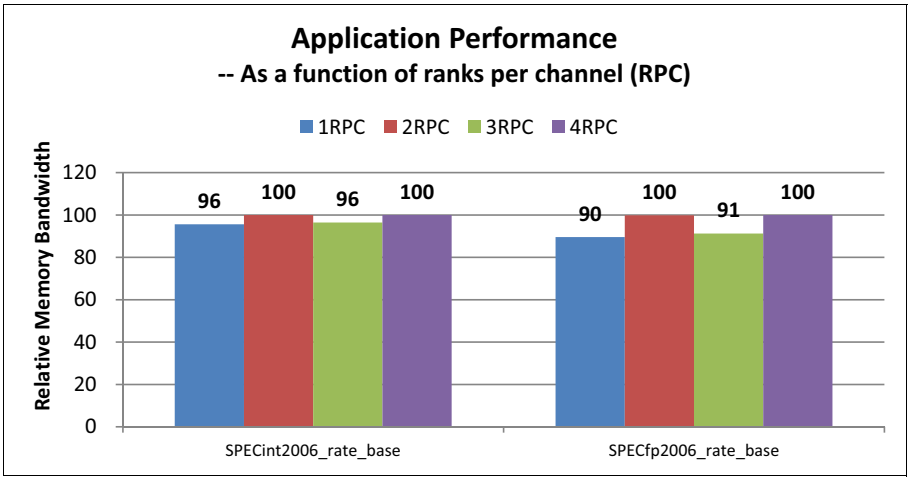

Figure 19 Application performance as a function of RPC

Memory population and balance

Memory interleaving refers to how physical memory is interleaved across the physical DIMMs. A balanced system provides the best interleaving and the best performance. The following rules must be observed for balancing a system for optimized memory performance:

- If all available DIMMs are of the same capacity, distribute the DIMMs such that all memory channels have the same number of DIMMs. Populate memory in groups of eight per compute book for the x3850 and x3950 X6 platforms, and in groups of 16 for the Flex System X6 compute node.
- If all available DIMMs are not of equal capacity, balance all eight channels in each compute book with the same amount of memory for the x3850 and x3950 X6 platforms. For the Flex System X6 compute node, ensure that individual DDR3 channels are populated with the same amount of memory capacity.

It is not uncommon for a system to contain a memory configuration that is poorly interleaved, which can occur for the following reasons:

- By using available DIMMs to reduce parts on the floor.
- By configuring a system that is based solely on memory capacity requirements. At a minimum, physical memory should be over-provisioned to the closest memory capacity, which yields a balanced memory configuration.

Memory channel population effect on STREAM Triad bandwidth

Leaving DDR3 channels unpopulated affects memory performance. Figure 20 shows how STREAM Triad memory bandwidth is affected, depending on how many DDR3 channels are populated per socket.

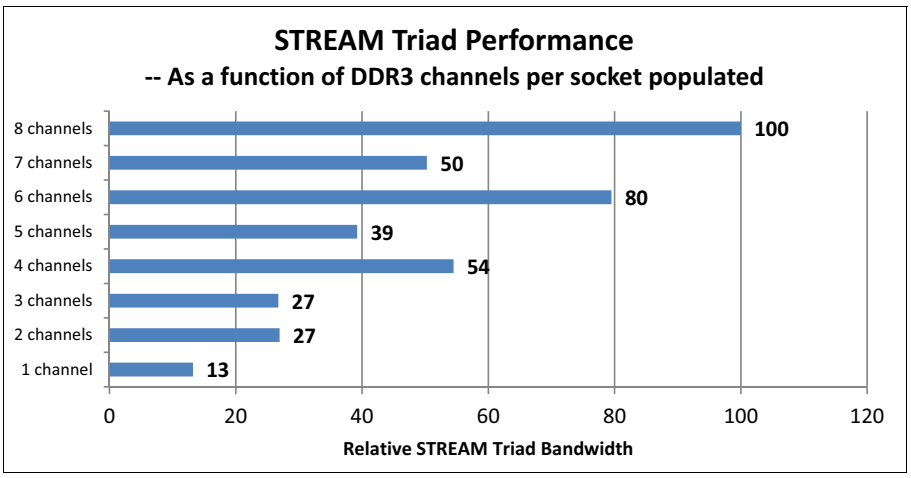

Figure 20 STREAM Triad performance as a function of DDR3 channels populated per socket

As shown in Figure 20, populating all memory channels per socket yields optimal memory performance. If all of the memory channels cannot be populated, it is best to populate even numbers of memory channels. Populating six memory channels yields 80% of maximum STREAM Triad memory bandwidth. However, populating seven memory channels yields only 50% of maximum expected bandwidth. In fact, populating only four memory channels per socket yields slightly more STREAM Triad bandwidth than populating seven channels.

Memory balance effect on memory throughput

In this section, several different DIMM configurations were analyzed to show the effect of an unbalanced memory configuration. All configurations were run at 1333 MHz.

Figure 21 on page 24 shows the possible memory performance degradation because of an unbalanced memory configuration. The capacities of the DIMMs that are used in this exercise were the same. Configurations 1 and 6 represent balanced memory configurations; therefore, they achieve optimal memory performance. The remaining configurations break the rules for a balanced memory configuration and show the performance relative to the most optimal configuration.

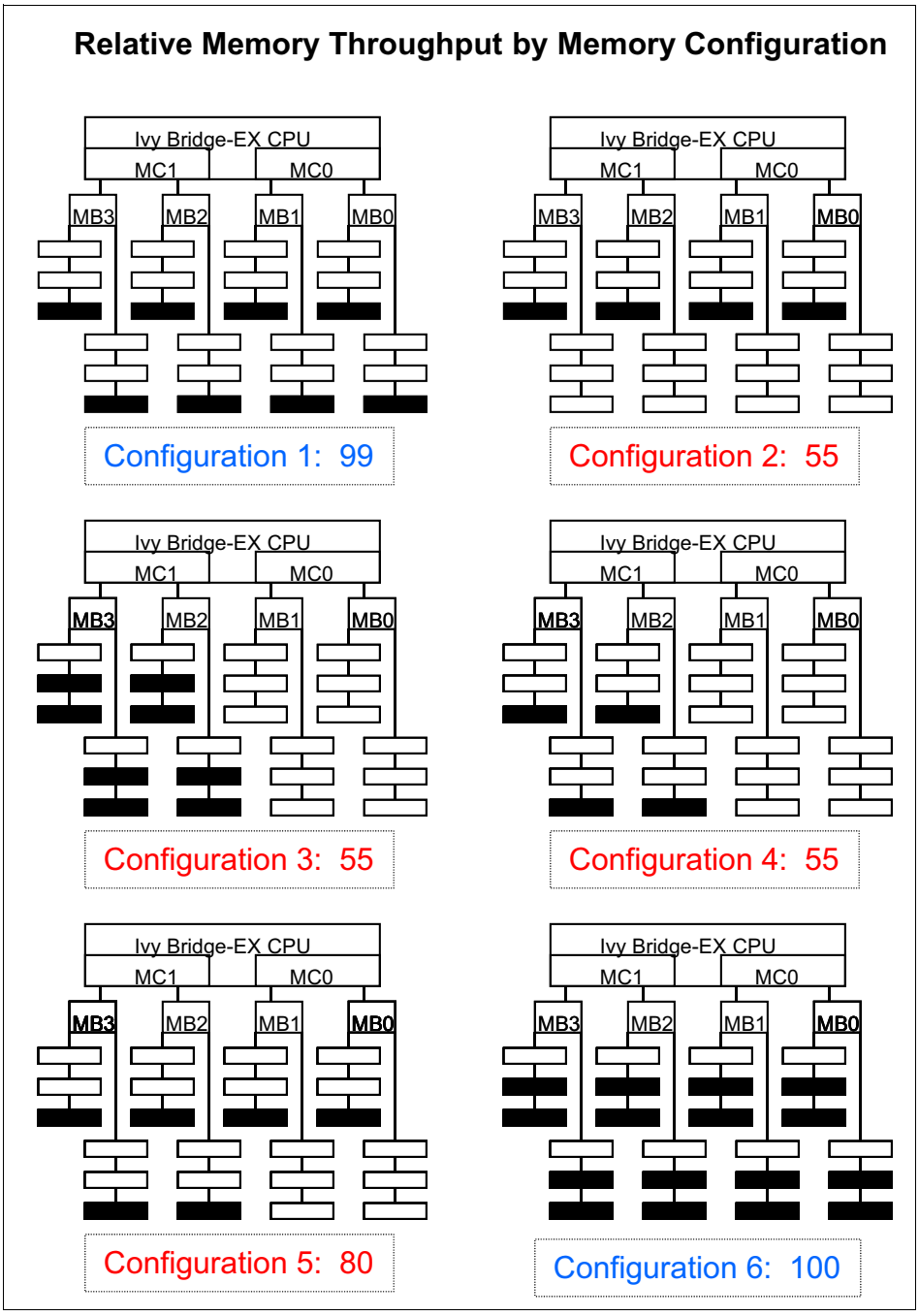

Figure 21 Memory performance of unbalanced memory configurations

Configuration 2 balances memory across both memory controllers, but populates only four of eight of the DDR3 channels per compute book. In this case, the maximum achievable memory throughput is only 55% of optimal.

Configurations 3 and 4 also populate only four out of the eight DDR3 channels, but all of the populated channels are on only one memory controller. Again, the result is 55% of optimal memory throughput.

Configuration 5 populates 6 of 8 DDR3 channels and can achieve 80% of optimal memory performance.

Different DIMM capacities effect on memory throughput

In this section, six different DIMM configurations were analyzed to show the effect of mixing DIMM capacities, as shown in Figure 22.

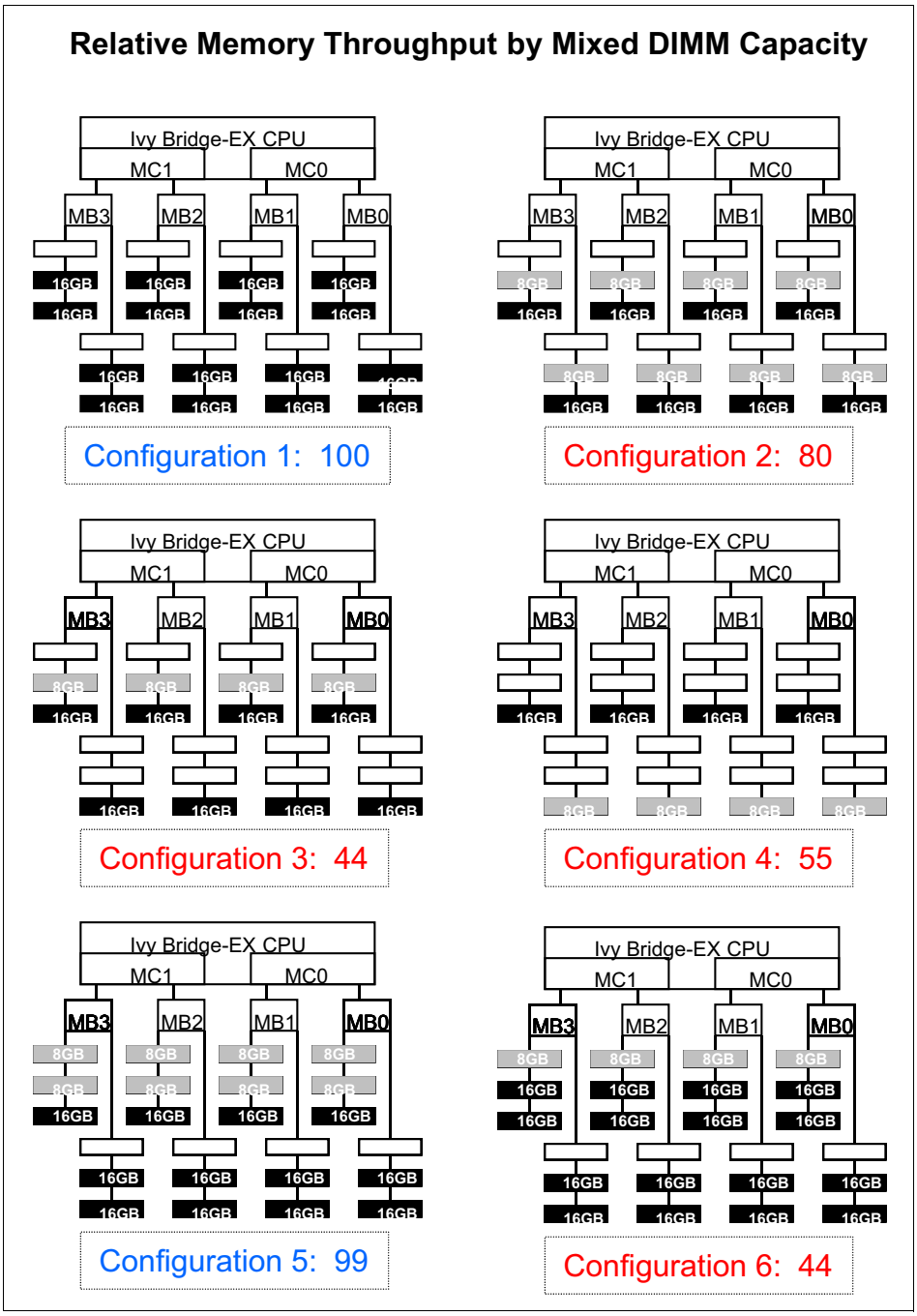

Figure 22 Memory performance of mixed DIMM capacity configurations

Configuration 1 represents a balanced memory configuration and achieves optimal memory performance. Configuration 2 also is a balanced memory configuration, but mixes 8 GB and 16 GB DIMMs. Each DDR3 channel contains the same capacity. However, the performance is reduced by 20%.

Configurations 3, 4, and 6 contain memory configurations with the DDR3 channels off each Jordan Creek memory buffer that contains different capacities. In these cases, the performance degradation is 45% - 56%.

Configuration 5 includes DDR3 channels with the same capacities. However, one DDR3 channel per memory buffer contains 2 DPC and the other 3 DPC. In this case, the performance degradation is nonexistent because the RPC is optimal per DDR3 channel and the capacities are balanced across memory channels.

Balancing memory population on X6 platforms

The Lenovo x3850 X6 and x3950 X6 systems use compute books that contain a processor socket and memory DIMM slots. The memory subsystem for these platforms can be balanced by installing DIMMs in groups of eight per compute book. The front of the computer book is shown in Figure 23.

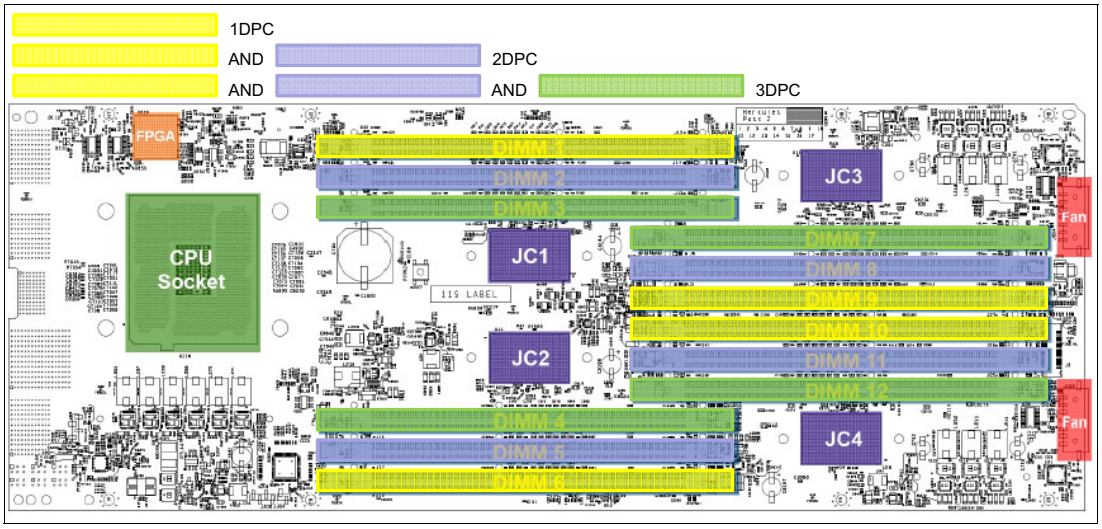

Figure 23 Front of x3850/x3950 X6 compute book

The back of the compute book is shown in Figure 24.

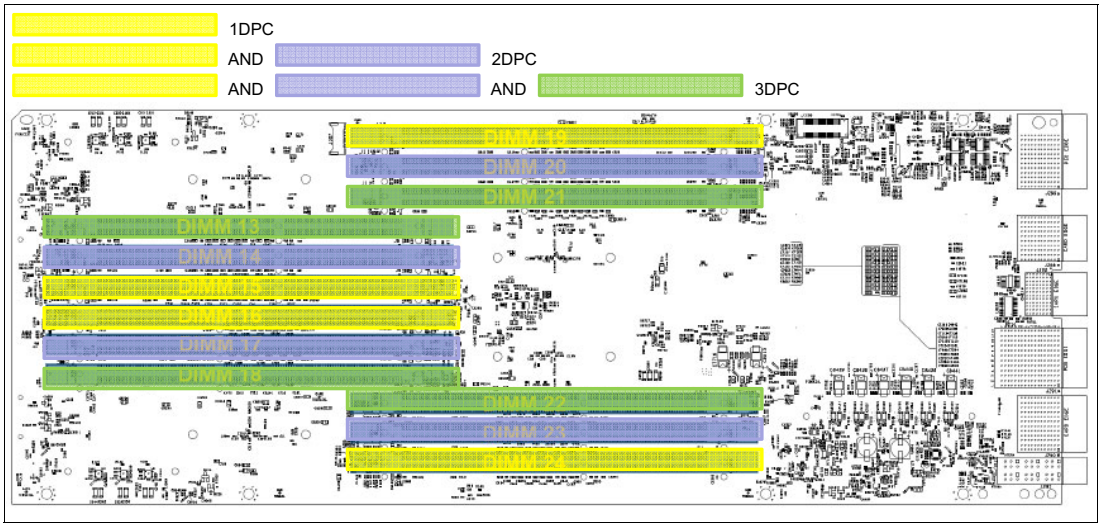

Figure 24 Back of x3850/x3950 X6 compute book

Use the colored legend that is shown at the top of Figure 23 on page 26 and Figure 24 on page 26 to populate the compute books correctly for a balanced memory configuration.

Figure 25 shows the DIMM layout of the Flex System X6 compute node. Use the colored legend that is shown above Figure 25 to populate the Flex System X6 node correctly with a balanced memory configuration.

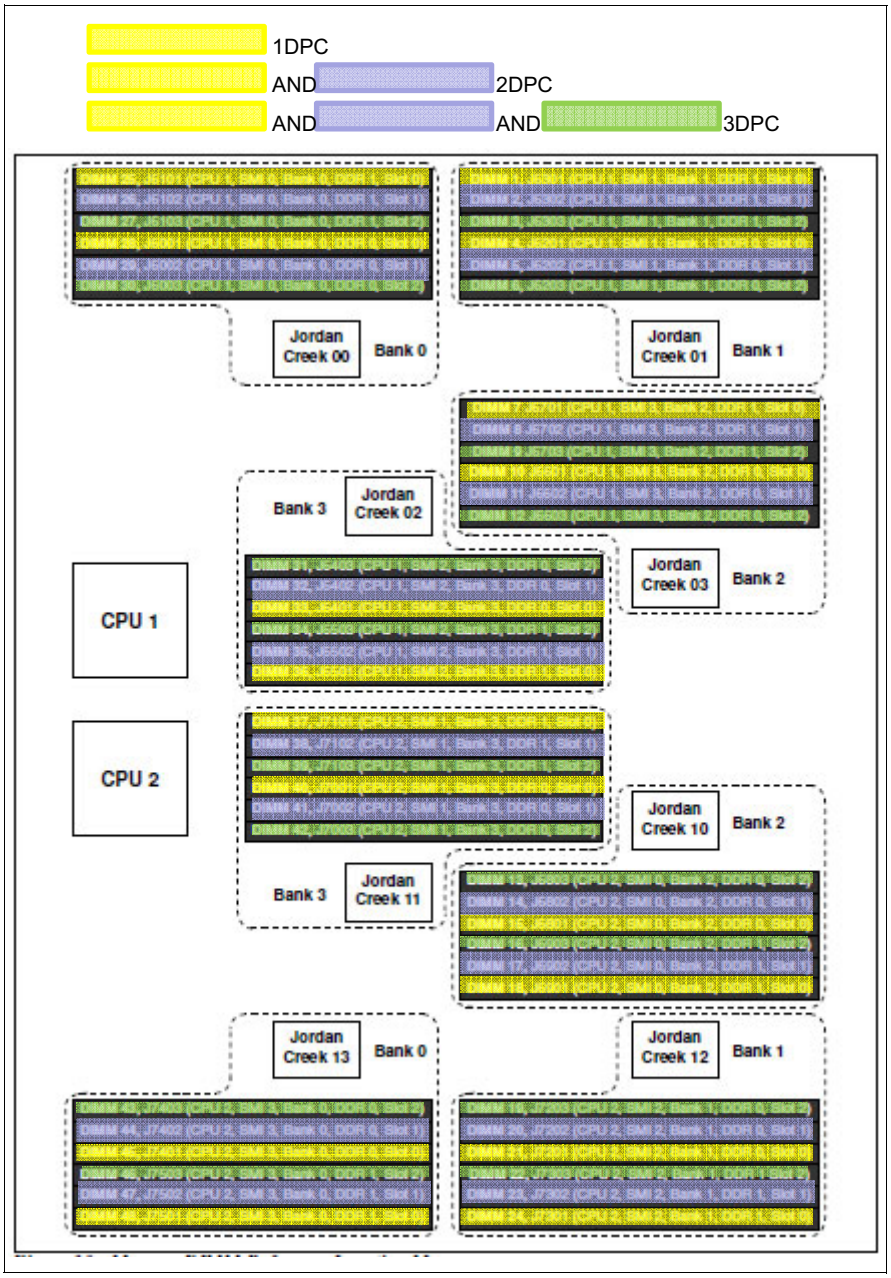

Figure 25 Flex System X6 DIMM layout

Best practices

This section recommends memory population best practices for the E7 v2 series-based platforms. Adhere to the following rules for optimal memory performance:

- Always populate all processors with equal memory capacity to ensure a balanced NUMA system.
- Always populate all eight memory channels on each socket by using identical DIMMs. If this configuration is not possible, the next best option is to populate all channels with equal memory capacity.
- Always populate all eight channels on each socket. If this configuration is not possible, populating an even number of channels is preferable to an odd number of channels.
- Use dual-rank RDIMMs whenever possible. Use LRDIMMs if memory capacity requirement cannot be achieved with RDIMMs.
- Populate memory channels with an even number of ranks when possible.

Conclusion

Lenovo systems that use the E7 v2 series processors offer significant performance gains over their predecessors. The x3850 and x3950 X6, and the Flex System X6 platforms support RDIMM and LRDIMM memory options, which can be optimized for performance and large memory capacity requirements.

Although every application has unique characteristics that might not be affected by the scenarios that are described in this paper, adhering to the best practices that are presented here produces a system that is configured for optimal memory performance.

Authors

Charles Stephan is the Technical Lead for the System Performance Verification team in the Lenovo System x and Flex System Performance Laboratory at the Lenovo EBG campus in Morrisville, NC. His team is responsible for analyzing the performance of storage adapters, network adapters, various flash technologies, and complete x86 platforms. Before transitioning to Lenovo, Charles spent 16 years at IBM as a Performance Engineer analyzing storage subsystem performance of RAID adapters, Fibre Channel HBAs, and storage servers for all x86 platforms. He also analyzed performance of x86 rack systems, blades, and compute nodes. Charles holds a Master of Science degree in Computer Information Systems from the Florida Institute of Technology.

Alicia Boozer is a hardware engineer in the Lenovo System x and Flex System Performance Laboratory in Morrisville, NC. Before starting at Lenovo, Alicia spent 6 years at IBM working in the benchmark area initially, then transitioning to system performance verification. Her current role includes subsystem analysis for all x86 products and performance validation against functional specifications and vendor targets. Alicia holds Bachelor of Science degrees in Mathematics from Spelman College and Electrical Engineering from North Carolina A&T State University and a Master of Science degree from the Massachusetts Institute of Technology.

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

- Randolph Kolvick, Senior Technical Staff Member, Lenovo
- Joseph Jakubowski, Senior Technical Staff Member, Lenovo
- Timothy Wiwel, System x High Performance Server Developer, Lenovo
- Mark Tirpack, System x High Performance Server Developer, Lenovo
- Mike French, Technical Project Manager, Lenovo
- 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 REDP-5223-00 was created or updated on June 4, 2015.

Send us your comments in one of the following ways:
- Use the online **Contact us** review Redbooks form found at:
  ibm.com/redbooks
- Send your comments in an email to:
  redbooks@us.ibm.com

Trademarks