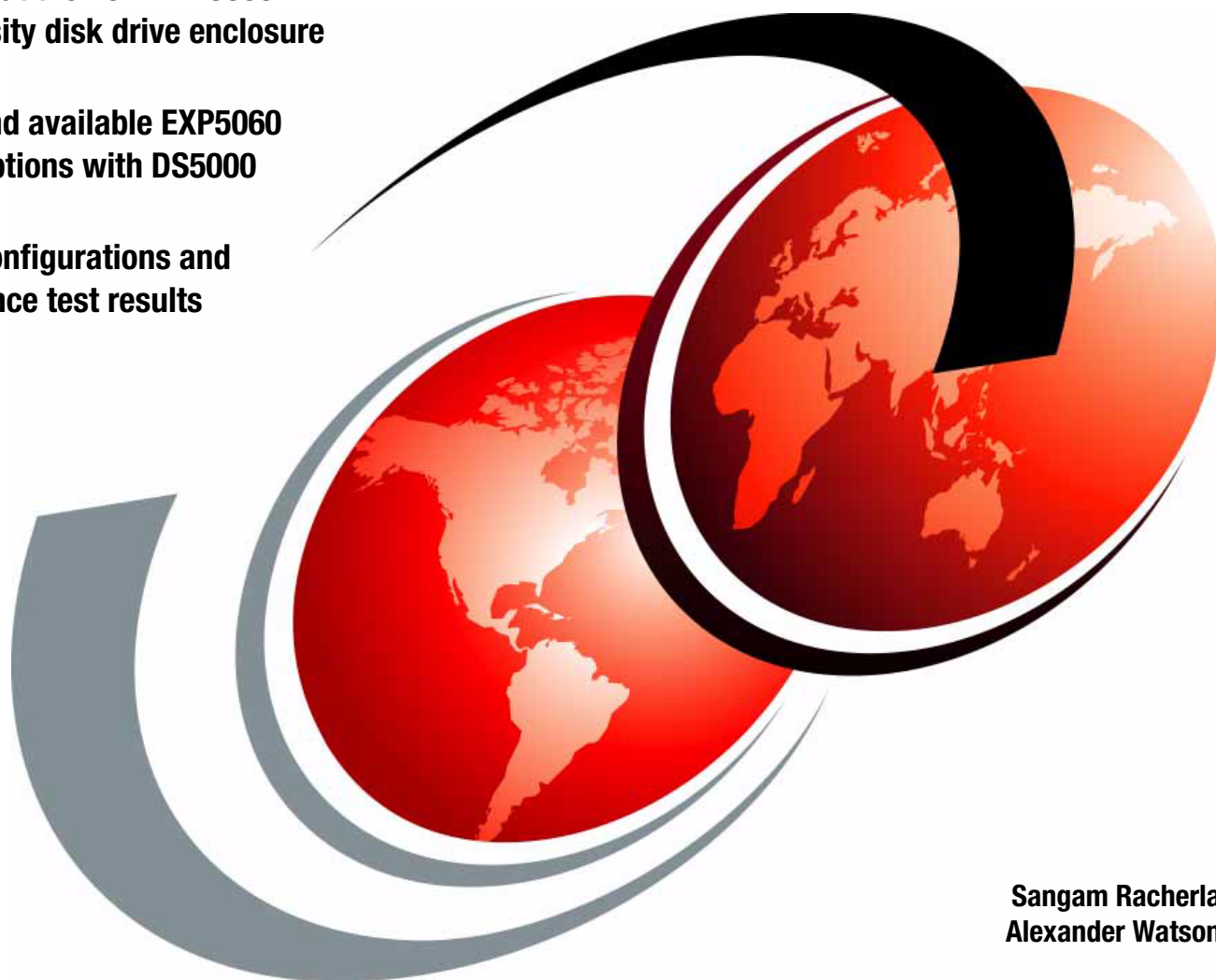


# IBM System Storage EXP5060 Storage Expansion Enclosure Planning Guide

Learn about the new EXP5060  
high-density disk drive enclosure

Understand available EXP5060  
cabling options with DS5000

Review configurations and  
performance test results



Sangam Racherla  
Alexander Watson





International Technical Support Organization

**IBM System Storage EXP5060 Storage Expansion  
Enclosure Planing Guide**

December 2010

**Note:** Before using this information and the product it supports, read the information in “Notices” on page v.

**First Edition (December 2010)**

This edition applies to IBM System Storage EXP5060 Storage Expansion Enclosure connected to IBM System Storage DS5100 and DS5300 controllers operating at firmware level 07.60, or later.

**© Copyright International Business Machines Corporation 2010. All rights reserved.**

Note to U.S. Government Users Restricted Rights -- Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

# Contents

<b>Notices</b> .....	v
Trademarks .....	vi
<b>Preface</b> .....	vii
The team who wrote this paper .....	vii
Now you can become a published author, too! .....	viii
Comments welcome .....	viii
Stay connected to IBM Redbooks .....	viii
<b>Chapter 1. IBM System Storage EXP5060</b> .....	1
1.1 The EXP5060 design and component description .....	2
1.2 EXP5060 design and component description .....	2
1.2.1 EXP5060 chassis design .....	3
1.2.2 EXP5060 disk drive modules .....	4
1.2.3 Environmental Service Modules .....	4
1.3 EXP5060 usage models .....	5
1.3.1 EXP5000 and EXP5060 mixed environment .....	5
1.3.2 Non-trunked EXP5060 only .....	6
1.3.3 Trunked configuration of the EXP5060 .....	6
1.3.4 High availability recommendations .....	6
<b>Chapter 2. Implementing the IBM System Storage EXP5060</b> .....	7
2.1 EXP5060 cabling design .....	8
2.1.1 EXP5000 and EXP5060 mixed environment .....	8
2.1.2 Non-trunked EXP5060 only .....	9
2.1.3 Trunking the EXP5060 .....	11
<b>Chapter 3. Configuring the EXP5060</b> .....	15
3.1 Planning for the needs of your environment .....	16
3.1.1 RAID array types .....	16
3.2 Array groups and logical drives .....	18
3.2.1 Number of disks per array group .....	18
3.2.2 Trunking .....	19
<b>Chapter 4. EXP5060 performance</b> .....	27
4.1 Performance test runs .....	28
4.1.1 Optimal performance layouts .....	28
4.1.2 Non-optimal performance layouts .....	31
4.2 Mixed configuration test runs .....	36
<b>Related publications</b> .....	43
IBM Redbooks publications .....	43
Other publications .....	43
Online resources .....	43
How to get IBM Redbooks publications .....	44
Help from IBM .....	44



# Notices

This information was developed for products and services offered in the U.S.A.

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

IBM 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:

*IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785 U.S.A.*

**The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law:** INTERNATIONAL BUSINESS MACHINES CORPORATION 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 states 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. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM 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 IBM product and use of those Web sites is at your own risk.

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

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

## COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs.

# Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. These and other IBM 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 IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at <http://www.ibm.com/legal/copytrade.shtml>


The following terms are trademarks of the International Business Machines Corporation in the United States, other countries, or both:

AIX®

IBM®

Redbooks®

Redpaper™

Redbooks (logo) ®

System p®

System Storage®

System x®

The following terms are trademarks of other companies:

Microsoft, Windows, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Linux is a trademark of Linus Torvalds in the United States, other countries, or both.

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



# Preface

The IBM® System Storage® EXP5060 High Density Storage Enclosure (Machine Type 1818, Model G1A) provides high-capacity Serial Advanced Technology Attachment (SATA) disk storage for the DS5100 and DS5300 storage subsystems. The storage expansion enclosure delivers fast, high-volume data transfer, retrieval, and storage functions for multiple drives to multiple hosts. The storage expansion enclosure provides continuous, reliable service, using hot-swap technology for easy replacement without shutting down the system and supports redundant, dual-loop configurations. External cables and Small Form-Factor Pluggable (SFP) modules connect the DS5100 or DS5300 storage subsystem to the EXP5060 storage expansion enclosure.

This IBM Redpaper™ publication gives a brief understanding of EXP5060 and serves as a planning guide for attaching the EXP5060 to DS5100 or DS5300.

Links to the additional resources and documentation are provided as needed.

## The team who wrote this paper

This paper was produced by a team of specialists from around the world working at the International Technical Support Organization, Raleigh Center.

**Sangam Racherla** is an IT Specialist and Project Leader working at the International Technical Support Organization (ITSO), San Jose Center. He holds a degree in electronics and communication engineering and has ten years of experience in the IT field. He has been with the ITSO for the past seven years and has extensive experience installing and supporting the ITSO lab equipment for various IBM Redbooks® publication projects. His areas of expertise include Microsoft® Windows®, Linux®, AIX®, System x® and System p® servers, and various storage area network (SAN) and storage products.

**Alexander Watson** is an ATS Specialist for Storage Advanced Technical Skills (ATS) Americas in the United States. He is a Subject Matter Expert on SAN switches and the IBM Midrange system storage products. He has over fifteen years of experience in planning, managing, designing, implementing, and analyzing problems for and tuning SAN environments and storage systems. He has worked at IBM for eleven years. His areas of expertise include SAN fabric networking, Open System Storage I/O, and the IBM Midrange Storage solutions.

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

Danh T. Le  
Michael Roll  
Thomas Phelan  
IBM

David Worley  
Timothy Chau  
LSI Corporation

## Now you can become a published author, too!

Here's an opportunity to spotlight your skills, grow your career, and become a published author - all at the same time! Join an ITSO residency project and help write a book in your area of expertise, while honing your experience using leading-edge technologies. Your efforts will help to increase product acceptance and customer satisfaction, as you expand your network of technical contacts and relationships. Residencies run from two to six weeks in length, and you can participate either in person or as a remote resident working from your home base.

Find out more about the residency program, browse the residency index, and apply online at:

[ibm.com/redbooks/residencies.html](http://ibm.com/redbooks/residencies.html)

## Comments welcome

Your comments are important to us!

We want our papers to be as helpful as possible. Send us your comments about this paper or other IBM Redbooks publications in one of the following ways:

- Use the online **Contact us** review Redbooks form found at:

[ibm.com/redbooks](http://ibm.com/redbooks)

- Send your comments in an e-mail to:

[redbooks@us.ibm.com](mailto:redbooks@us.ibm.com)

- Mail your comments to:

IBM Corporation, International Technical Support Organization  
Dept. HYTD Mail Station P099  
2455 South Road  
Poughkeepsie, NY 12601-5400

## Stay connected to IBM Redbooks

- Find us on Facebook:

<http://www.facebook.com/IBMRedbooks>

- Follow us on Twitter:

<http://twitter.com/ibmredbooks>

- Look for us on LinkedIn:

<http://www.linkedin.com/groups?home=&gid=2130806>

- Explore new Redbooks publications, residencies, and workshops with the IBM Redbooks weekly newsletter:

<https://www.redbooks.ibm.com/Redbooks.nsf/subscribe?OpenForm>

- Stay current on recent Redbooks publications with RSS Feeds:

<http://www.redbooks.ibm.com/rss.html>



# IBM System Storage EXP5060

This IBM Redpaper publication is primarily being provided as a supplemental document to the published standard support installation and user documentation. The major focus of this paper is to provide best practices guidance and performance details to users who want to know how best to configure and use this new expansion to enhance their solution environments. This paper is meant to complement the standard support documentation with its detailed installation usage and maintenance information. For full details of those areas, see the additional documentation listed in “Related publications” on page 43.

In this chapter, we introduce the new IBM System Storage EXP5060 High Density Storage Enclosure. We provide a brief description of this new enclosure, its features and capabilities, usage models, and where it best fits in terms of various storage solutions. We also summarize the functions of the DS Storage Manager software as it pertains to this new enclosure.

If you are unfamiliar with this new expansion, review this chapter to ensure that you are aware of all the new capabilities, design, and component layout that the new EXP5060 can provide, and how they will affect your environment.

## 1.1 The EXP5060 design and component description

The new IBM EXP5060 enclosure is a high capacity expansion, capable of holding up to 60 Serial Advanced Technology Attachment (SATA) disk modules. You can have a maximum of eight EXP5060s installed on either the DS5100 or the DS5300 with the proper addition of the necessary feature keys. The maximum configuration is 480 disk modules.

**Best practice:** For the maximum configuration on the DS5100, include the performance enhancement key as well with the configuration to increase the throughput rate to the maximum.

The design of this expansion is five drawers with 12 disks in each drawer. At the time of writing, the supported disks offered are the 1 TB and 2 TB models. The maximum raw capacity of a single EXP5060 is a maximum of 120 TB, or a maximum system capacity of 960 TB. Figure 1-1 is a front view of the EXP5060 enclosure and an illustration of its front bezel.

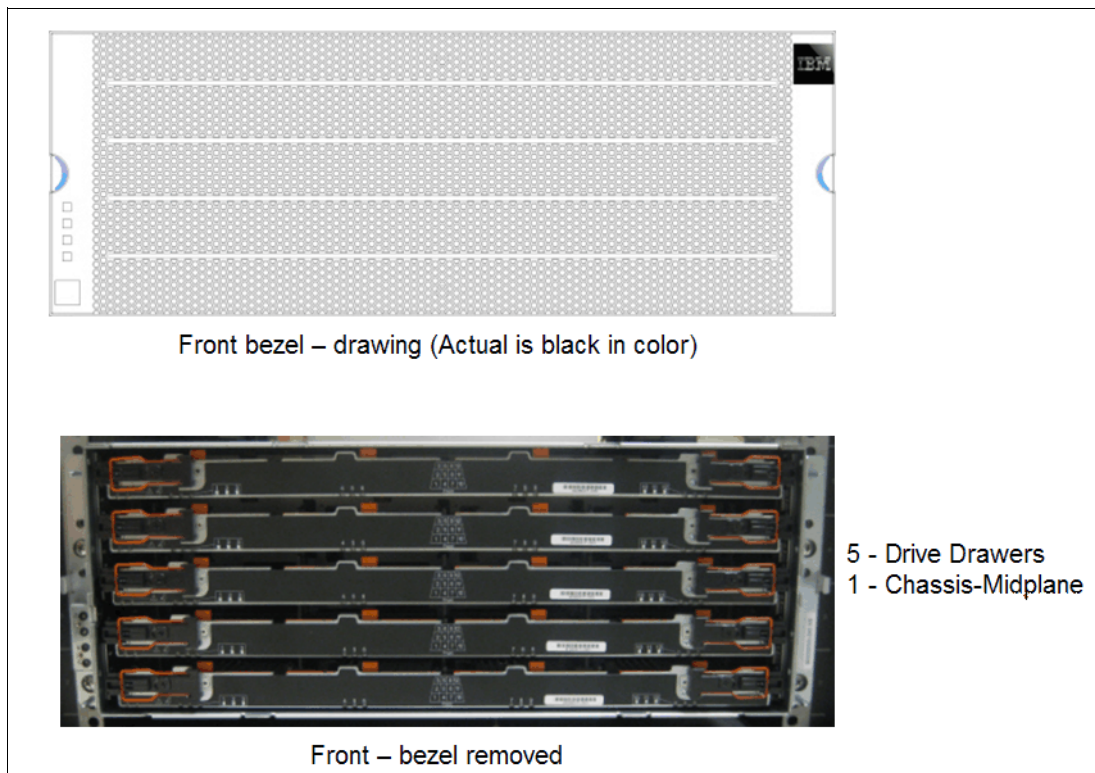


Figure 1-1 EXP5060 front bezel and drive drawers

## 1.2 EXP5060 design and component description

The EXP5060 is similar to the EXP5000 expansion. It has two redundant power supplies, two Environmental Service Modules (ESMs), and it supports SATA 1 TB and 2 TB disk modules, as the EXP5000 can. It connects by a pair of Fibre Channel loop connections to the DS5100 or DS5300 controllers for redundant paths and loops. However, this point is where the comparison ends.

### 1.2.1 EXP5060 chassis design

The EXP5060 is a 5U tall chassis that is built with five drawers in which 12 SATA disk drives can be installed into each, for a total of 60 disks. There is an interconnect midplane, which interfaces the five drawers to the Environmental Service Modules (ESMs). There are service and status LEDs on the front panel for the EXP5060, and drive activity and service LEDs on the front of each drawer for each drive location in the drawer.

For required airflow, a minimum of four drives must be installed in each drawer in the front four drive bays (1, 4, 7, and 10). This requirement places the minimum configuration of this expansion at 20 disks. Figure 1-2 shows the drive slot locations in the drawer and the location of the four required drives.



Figure 1-2 EXP5060 drawer with required drives shown

The EXP5060 is also a heavy chassis with an empty weight of about 56.7 kg (125 lb) and a full chassis weight of 113 kg (250 lb). It is due to this heavy weight that the EXP5060 is not to be installed in any rack over the 32U point. When installing the chassis into the rack, it is a requirement that you use a portable lift tool to install or remove the chassis from the cabinet. Make sure that the lift tool is available on location at the time of these procedures.

**Note:** The ordering procedures for the lift tool vary depending on your location. Direct questions about these procedures to your regional representative.

The EXP5060s power supplies are a higher wattage to handle the increased drive count and therefore require a 208 VAC circuit. Ensure that you order sufficient and proper power distribution units (PDUs) for the rack in which the EXP5060s will be installed.

**Important:** The EXP5060 does not support 90-136 VAC sources. It supports 180-240VAC sources only.

## 1.2.2 EXP5060 disk drive modules

The SATA drive modules that are supported by the EXP5060 differ from the SATA drives that are supported by the EXP5000. The EXP5060 has a fiber to Advanced Technology Attachment (ATA) translator built into the ESMs and does not require the use of the interposer card on the disk modules. Additionally, with the horizontal mounting of the drive in the drawer, IBM has introduced a new carrier with these modules. Figure 1-3 shows the drive module and carrier that support the EXP5060.

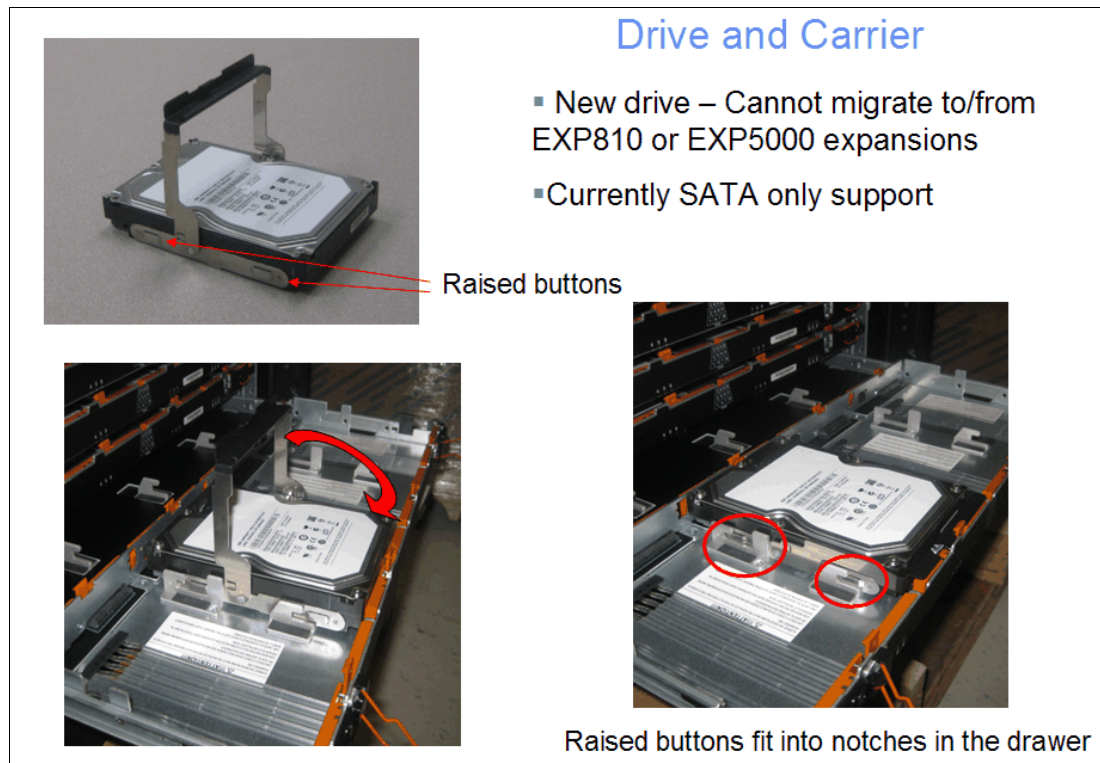


Figure 1-3 EXP5060 drive and carrier assembly

## 1.2.3 Environmental Service Modules

The Environmental Service Modules (ESMs) that are used in the EXP5060 are uniquely designed compared to other DS5000 expansions, because they add both hardware and firmware changes to support the new trunking capability. Each of these modules has a second level of built-in fiber switching and a device control manager (DCM) that manages the device I/O operations. These features play a critical part in the handling, managing, and throughput capabilities of these expansions. We describe their roles and effects in detail in Chapter 2, “Implementing the IBM System Storage EXP5060” on page 7 and Chapter 3, “Configuring the EXP5060” on page 15 of this document.



Figure 1-4 shows the back of the EXP5060 with its fan assemblies, power supplies, and ESMs. Figure 1-4 also shows the ports, ID indicator, LED status lights, and serial maintenance connections.

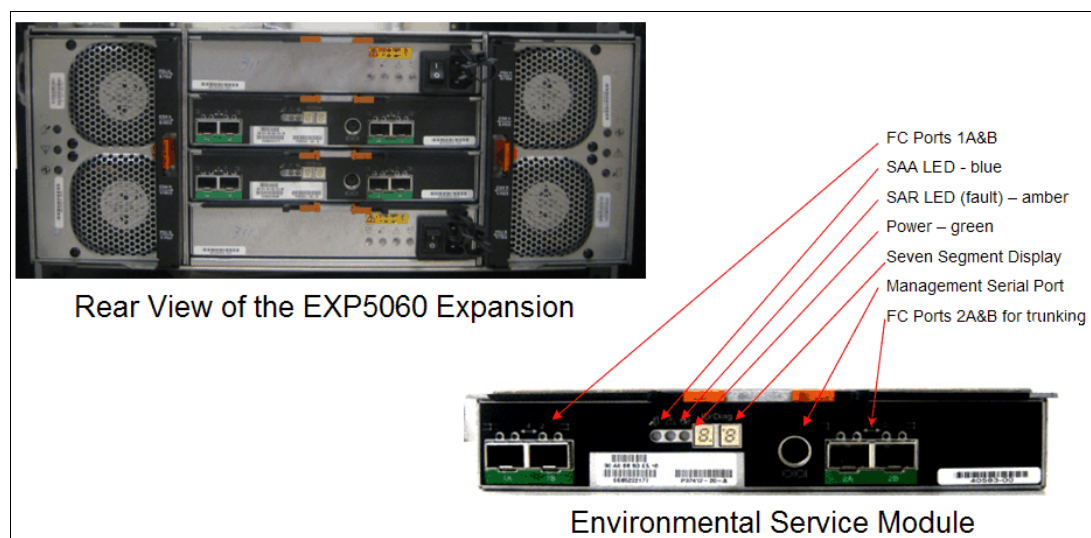


Figure 1-4 EXP5060 rear view and ESM card

## 1.3 EXP5060 usage models

You can configure the EXP5060 enclosure in a number of configurations. Its SATA drive support model is best suited for large block sequential workloads. Therefore, the best application areas for the EXP5060 are imaging, video storage, storage pools, and sequential, high-performance, computing applications. Small random I/O intense workloads, such as accessing online transaction processing (OLTP) databases and certain mail servers, are best served by the Fibre Channel 15K RPM disk that is supported with the EXP5000 expansions. In many VMWare environments, it is common practice to use the low-cost SATA drive solutions; here, it is important to know the expectations of the guest application prior to making a choice. The *IBM Midrange System Storage Implementation and Best Practices Guide*, SG24-6363, discusses these areas of performance differences in detail.

### 1.3.1 EXP5000 and EXP5060 mixed environment

When needed, you can combine the EXP5060 with EXP5000 expansions to create an environment for shared workloads with Fibre Channel drives for the random workloads and large SATA capacity for the sequential and streaming needs. Be careful when you configure the system in this manner, because heavy throughput workloads can greatly affect simultaneous, small, random I/O workloads running on the same channels. Timing conflicts and channel contention can be a major performance concern.

**Best practice:** When mixing EXP5000 and EXP5060 expansions on the same DS5300, dedicate separate channels for the two separate workload types.

We do not recommend that you cascade multiple EXP5000s with the EXP5060, although this configuration is supported. Using many disks on the same loop pair for varying workloads risks higher channel contention with the members. In this configuration, you cannot use trunking to enhance the throughput capabilities.

### 1.3.2 Non-trunked EXP5060 only

You can install the EXP5060 on one loop of the DS5300 in the same manner as the EXP5000s. In this configuration, the system can reach its maximum configuration of eight EXP5060s by placing one EXP5060 on each channel loop behind the DS5300. The maximum configuration cannot have any EXP5000s included with the EXP5060s. To support this maximum configuration, you must order a special maximum disk configuration feature key for the DS5300 midrange system.

**Important:** The special feature key for the DS5100 to support 448 disks is required as a prerequisite to installing the 480 maximum drive configuration feature key.

When configuring the EXP5060 in this non-trunked environment, it is a best practice to follow the recommended cabling, array, and logical unit number (LUN) layout practices that are used with the EXP5000s. For the best throughput and a balanced workload, assign the array and LUN layout to preferred controllers in the same manner as discussed in *IBM Midrange System Storage Implementation and Best Practices Guide*, SG24-6363. See this document for more details in this area.

### 1.3.3 Trunked configuration of the EXP5060

With the new design of the EXP5060 expansion, we now have the ability for trunked expansion across four loops to get the full throughput of a pair of channels (1.6 GBps) using one EXP5060 worth of disks. So, when a solution requires high throughput with fewer EXP5060s and a lower drive count, the trunked design is a great answer to fill this need. The trunked capability of the EXP5060 is aimed at providing high throughput. It does not improve the I/O handling capability of the subsystem. Input/output operations per second (IOPS) is drive spindle type dependent and quantity dependent.

**Important:** Observe the special array layouts and logical drive configuration requirements to ensure the maximum throughput in a trunked configuration. See Chapter 3, “Configuring the EXP5060” on page 15 and Chapter 4, “EXP5060 performance” on page 27 for more information.

In the trunked configuration, you can design a cascaded pair of EXP5060s on a trunked channel pair. In this configuration, you can build a subsystem up to the maximum configuration of eight EXP5060s. Be aware, as the spindle count and potentially the array and LUN configuration increase, the effective throughput is equal to the non-trunked configuration. There is however an advantage gained from the cascaded trunked configuration, because the number of paths that are available for path redundancy is now four paths as opposed to two paths for a non-trunked configuration.

### 1.3.4 High availability recommendations

With the DS5000 and all expansions, you can build arrays and LUNs across expansion trays to enable enclosure loss protection. With newer code releases, the robustness of the networks and improved drive preemptive diagnostics have made the need for this capability less critical. However, with the design of the EXP5060, certain components reside on the drawer assembly. For repair, you might need to remove the drawer from the chassis. When building arrays and LUNs, we recommend when possible that you select drives from across the drawer assemblies to allow for drawer protection with your configuration. With careful planning and care, you can design high throughput configurations in this manner. We explain this topic in detail in Chapter 3, “Configuring the EXP5060” on page 15.





## Implementing the IBM System Storage EXP5060

As described in Chapter 1, “IBM System Storage EXP5060” on page 1, you can configure the IBM System Storage EXP5060 High Density Storage Enclosure (EXP5060) in many configurations with the DS5300 storage subsystem. You can use the EXP5060 in these configurations for various workload purposes. Because the present configuration only supports the 7500 RPM Serial Advanced Technology Attachment (SATA) disk drive modules (DDMs), we recommend that you do not use them for high transaction, I/O-based workloads. However, throughput-based workloads with large sequential host requests can perform well with this storage environment. In this chapter, we discuss the options and ways to configure the EXP5060 disks into arrays and logical unit numbers (LUNs) to provide the highest levels of throughput and the best load balancing across the entire storage subsystem.

## 2.1 EXP5060 cabling design

You can configure the EXP5060 enclosure in many ways. There are two major styles of configurations: non-trunked and trunked. In the non-trunked environments, you can combine the EXP5060 with EXP5000 expansions to create an environment for shared workloads with Fibre Channel drives and large SATA capacity needs. Or, you can trunk the EXP5060 across two loop pairs for dedicated high throughput environments to support large streaming applications. In the trunked configurations, the EXP5060 can reach maximum throughput numbers driving all the back-end channels at their maximum limits.

### 2.1.1 EXP5000 and EXP5060 mixed environment

When mixed with the EXP5000 expansion, place each expansion type on its own loop pair of cables. However, when necessary, you can cable the EXP5060 to cascade directly behind an EXP5000. In this configuration, the single loop pair connects to both of the expansions (as shown in Figure 2-1), which is an example of the cascaded configuration.

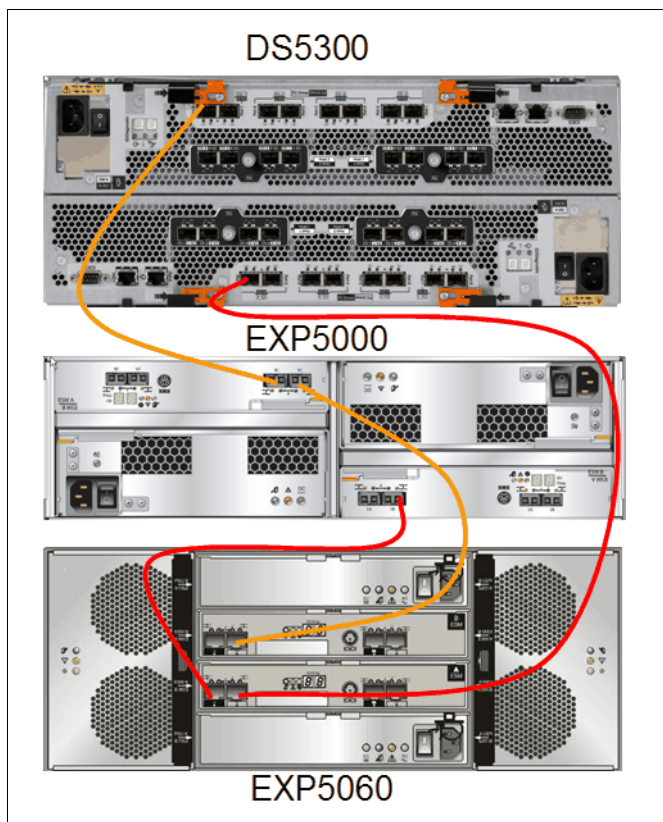


Figure 2-1 EXP5000 and EXP5060 cascaded on the same loop pair

Although it is a supported configuration, we do not recommend that you cascade multiple EXP5000s with the EXP5060. Having a high number of disks on the same loop pair risks higher channel contention from the members. In this configuration, you cannot use trunking to enhance the throughput capabilities.

## 2.1.2 Non-trunked EXP5060 only

You can also install the EXP5060 in the standard configuration with one EXP5060 per loop pair installed on each of the DS5100 or DS5300 loop pairs. In this configuration, the system can reach its maximum configuration of eight EXP5060s. This configuration cannot include any EXP5000s. To support this maximum configuration, you must add a special feature key to the DS5100 or DS5300 storage subsystems.

**Important:** For the maximum configuration special feature key to work on the DS5100, you must order the performance enhancement feature key first.

When configuring the EXP5060 in this non-trunked environment, follow the recommended cabling, array, and LUN layout practices that are used with the EXP5000s. Use the *IBM Midrange System Storage Implementation and Best Practices Guide*, SG24-6363, to determine the preferred way to lay out the arrays and LUNs for the best throughput and balanced workload handling. Figure 2-2 shows an EXP5060 that is connected in a non-trunked configuration.

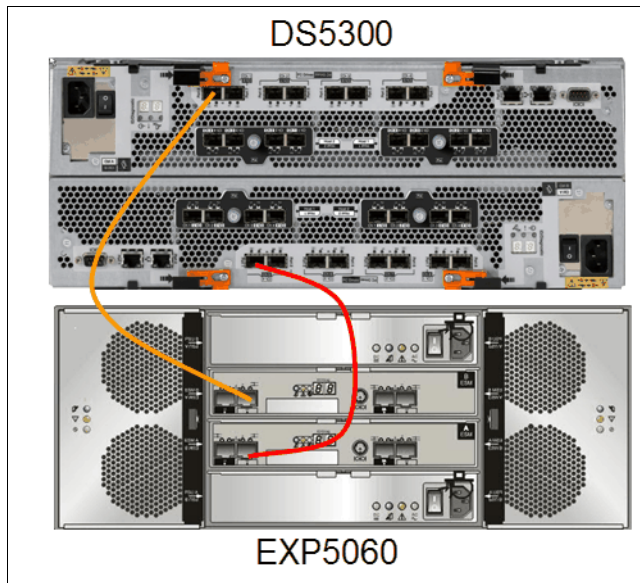


Figure 2-2 EXP5060 cabled non-trunked to the DS5300

You can copy this single unit per port configuration easily for each of the eight expansions on the port pairs for a full configuration. Figure 2-3 shows a simplified drawing of the connections for this full configuration.

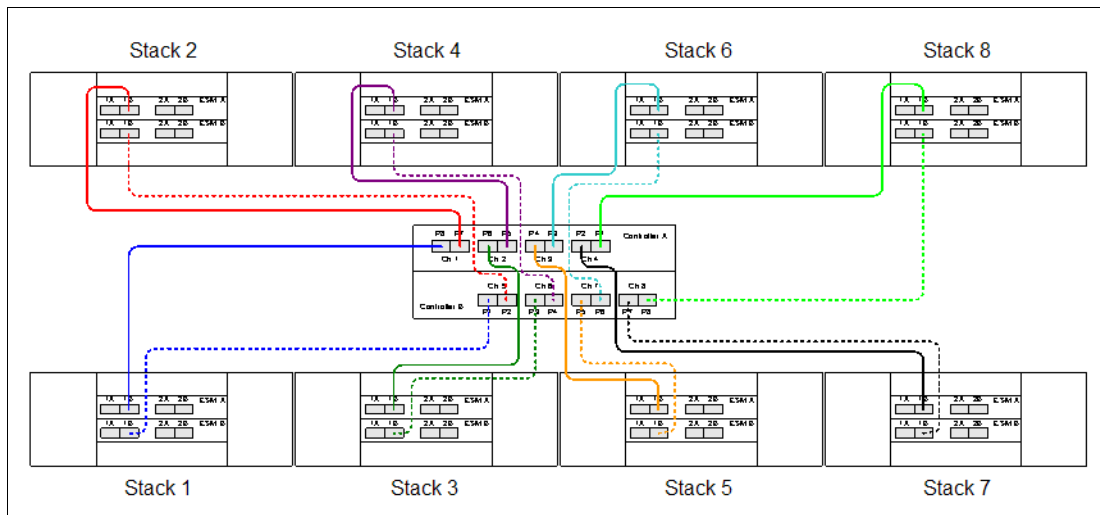


Figure 2-3 Cabling for non-trunked configuration of eight EXP5060s

In certain mixed EXP5000 and EXP5060 environments, it is sometimes necessary to cascade EXP5060s on the same loop pair. Although this configuration is supported, it is not a best practice. If you need large storage capacity that has minimal use, and if disk performance and EXP5000 I/O performance are critical, cascading the EXP5060s by using the non-trunked cabling scheme might meet the need for this environment. However, as shown in Figure 2-4, the disks in the EXP5060s in each drive channel loop pair are limited to an average of half the channel bandwidth for each expansion of 60 disks.

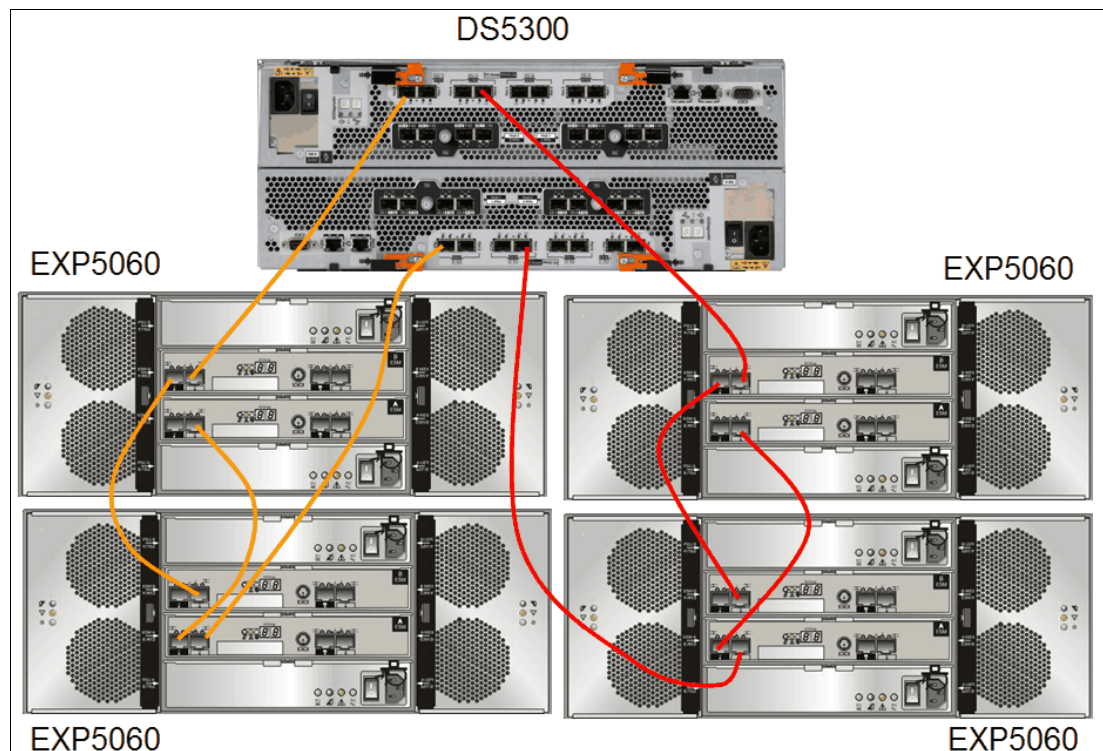


Figure 2-4 EXP5060 non-trunked and cascaded configuration

Normally, in the non-trunked environment, we recommend that you attach a single EXP5060 on a channel loop pair, as shown in Figure 2-2 on page 9. If possible, it is better to cascade the EXP5000 expansion so that the EXP5060s can be spread across more channel pairs or built into a trunked configuration, as discussed in 2.1.3, “Trunking the EXP5060” on page 11. The configurations that are shown in Chapter 4, “EXP5060 performance” on page 27 might offer an option that meets your needs without sacrificing performance. The key point to consider is that throughput has a great need for bandwidth.

### 2.1.3 Trunking the EXP5060

With the new design of the EXP5060 expansion, we now can trunk the expansion across four loops to get the full throughput of a pair of channels using only one EXP5060 worth of disks (60 drives). This design makes it possible to achieve over 6 GBps for the DS5300 subsystem. So, when your requirement is for high throughput with fewer EXP5060s and a lower drive count, the trunked design is a great solution. In the trunked configuration, you can use a total of four EXP5060s to drive the full bandwidth of all of the eight back-end channels. Figure 2-5 shows two EXP5060s in a trunked configuration.

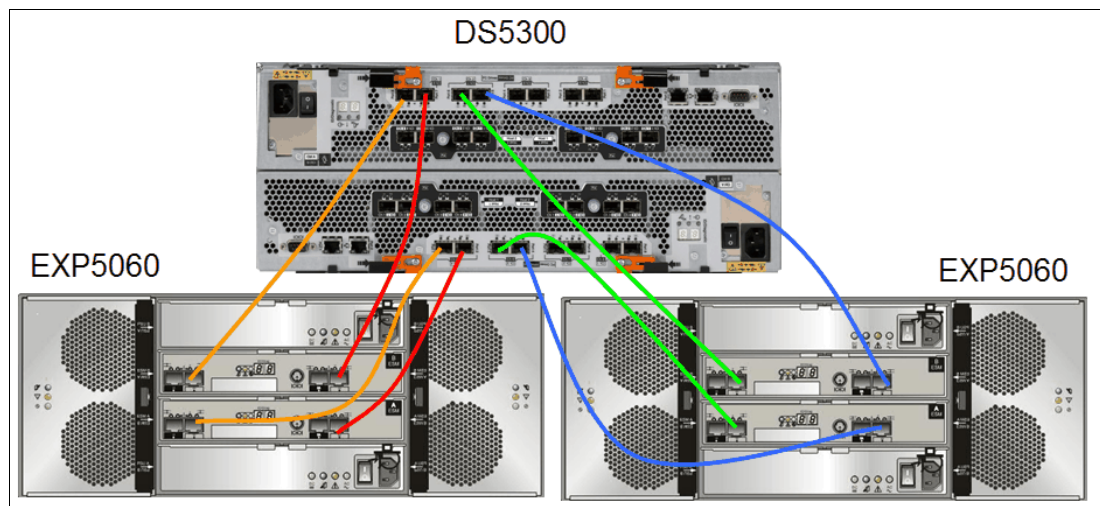


Figure 2-5 EXP5060 trunked configuration

Figure 2-6 shows a drawing of four EXP5060s in a trunked configuration that reaches the full 6 GBps throughput of the DS5300 subsystem.

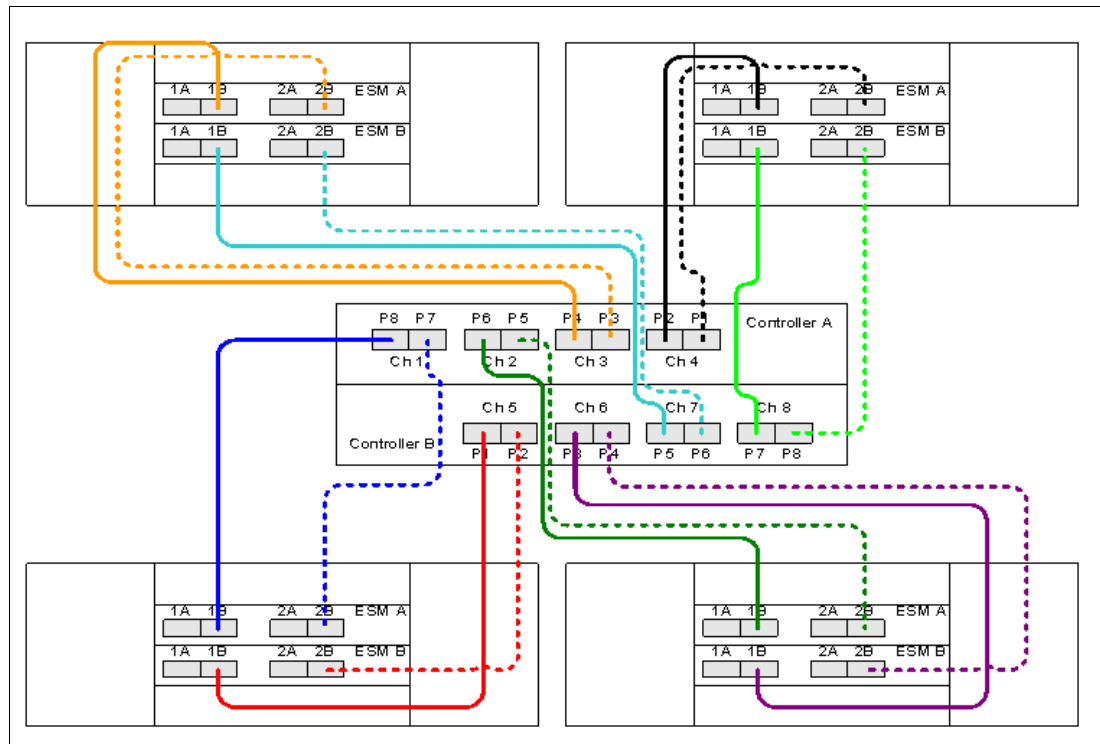


Figure 2-6 Four EXP5060s in trunked configuration for maximum throughput

You also can design a cascaded, trunked configuration with up to the full eight EXP5060s attached. In this configuration, there is no real throughput enhancement over the full non-trunked configuration. The cascaded configuration of the two EXP5060s per trunked channel pair results in the same number of disks being addressed per channel pair as in the non-trunked environment. Therefore, the same number of active disks is possible.



Figure 2-7 shows the four EXP5060s in a cascaded and trunked configuration.

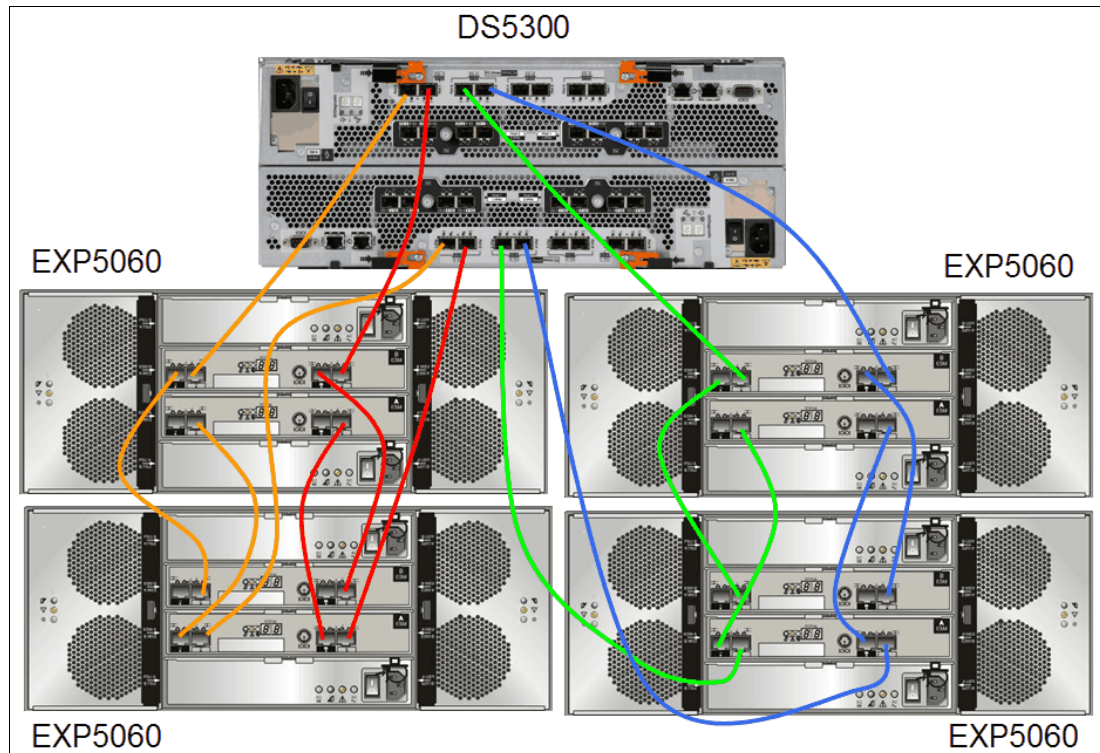


Figure 2-7 EXP5060s in a trunked and cascaded configuration

The major advantage of the cascaded, trunked configuration is the number of paths available for path redundancy. The additional disks can help with I/O per second (IOPS) limits. However, this environment is not designed for I/O-based applications, and the trunking of the channels does not help in this area.

*With the EXP5060, it is critical for throughput performance that you observe the cabling recommendations for best practices.* It is also critical to build the arrays and LUNs so that the resources are properly spread and evenly used. We discuss these recommendations in Chapter 3, “Configuring the EXP5060” on page 15, along with several examples.







## Configuring the EXP5060

In this chapter, we discuss the planning and use of many of the best practice array group configurations for the IBM System Storage EXP5060 High Density Storage Enclosure disks. We look at the design and layout, and how to attain the best performance from the disks housed in this expansion. There are many ways in which to build a configuration, but it is important to understand that in certain cases you might pass on one desired practice to gain a major objective. Only you can decide what is acceptable. In this chapter, we point out each of these practices in a layout and what you can gain or lose with each practice.

As discussed in the Chapter 2, “Implementing the IBM System Storage EXP5060” on page 7, the EXP5060 with its Serial Advanced Technology Attachment (SATA) disks is not the best solution for an I/O per second (IOPS)-intense environment. When you support this workload, design and configure the arrays for the best practices of the EXP5060 and the best design for the workload. In this chapter, our focus is on getting the best throughput out of the expansion, but when appropriate, we also comment on the IOPS capabilities of the configuration as well.

## 3.1 Planning for the needs of your environment

When setting up the DS5000 storage subsystem with the EXP5060, the configuration of the arrays and logical drives is a critical piece in your planning. Knowing the solution goals is a major piece of the planning to be done. Understanding how your configuration affects your workload and the entire solution is crucial to the success of your planning. Several of the design features that you choose to incorporate might have a negative effect on your performance. However, you might still need to use these design features to obtain the availability requirements of the solution. With this consideration in mind, we look at a number of design plans that can be used.

### 3.1.1 RAID array types

When configuring a DS5000 storage subsystem with the EXP5060, you can configure it with array groups of any of the supported Redundant Array of Independent Disks (RAID) types for the DS5000 storage subsystem. Understanding the advantages of each RAID type is critical when selecting the type that you want to use. In certain cases, you might need to create various array groups for separate workload types. The following brief descriptions show the benefits of each RAID type.

RAID0, which is striping without mirroring or parity protection, is generally the best choice for almost all environments for maximum performance; however, there is no protection built into RAID0 at all, and a drive failure requires a complete restore. For protection, it is necessary to look toward either a software mirroring solution or one of the other RAID types.

RAID1 on the DS5000 storage subsystem is disk mirroring for the first pair of disks. Larger arrays of four or more disks use disk mirroring and striping (RAID10 model). This RAID type is a good performer for high random read and write environments. It outperforms RAID5 due to the additional reads that RAID5 requires in performing its parity check when doing the write operations. With RAID1, there are two writes performed per operation. With RAID5, there are two reads and two writes required for the same operation, totaling four I/Os.

Common uses for RAID1 are mail server environments where random writes can frequently outweigh the reads, and database log files where generally high numbers of small sequential writes are needed.

Both RAID1 and RAID5 work well with random reads. RAID1 has an advantage if the layout is designed with greater spindle count given to the set of disks than are designed for the RAID5 layout to handle the same capacity. In this way, RAID1 can outperform RAID5 due to the extra spindles with which it is designed. In many cases, this advantage is too small to justify the added cost of the required extra drives. When the RAID5 array is created with an equal number of spindles to handle the same workload as the RAID1 array, it outperforms the RAID1 array.

**Tip:** There are no absolute choices as to which type of RAID to use. The RAID type depends on the workload read and write activity. A good general guide might be to consider using RAID1 if random writes exceed about 25%.

In the sequential high throughput environment, RAID5 performs well. You can configure it to perform only one additional parity write when using “full stripe writes”, which are also known as “full stride writes”, to perform a large write I/O. Compared to the two writes per data drive (self and its mirror) that are needed by RAID1, RAID5 performs better. As shown in Figure 3-1, this model is a definite advantage with RAID5 array groups.

Optimized RAID-5 (7+P) write activity						
HDDs Written	Read Data	Read Parity	Write Data	Write s Parity	RAID 5	RAID 1
1	1	1	1	1	4	2
2	2	1	2	1	6	4
3	3	1	3	1	8	6
4	3	0	4	1	8	8
5	2	0	5	1	8	10
6	1	0	6	1	8	12
7	0	0	7	1	8	14

Figure 3-1 RAID5 write penalty chart

Another option with the sequential high throughput environment is RAID6, because it performs well in this environment. RAID6 is second to RAID5 due to the extra drive operation required. Tests have shown that a well-configured RAID6 array and logical unit number (LUN) can perform within 15% of a comparable RAID5 configuration. You can see the advantage of both RAID5 and RAID6 when you configure them to perform only one or two additional parity writes when using full stripe writes and handling large write I/Os. RAID5 and RAID6 perform better than the two writes per data drive (self and its mirror) that are needed by RAID1. As shown in Figure 3-1, this model is a definite advantage for RAID5 and RAID6 array groups.

With the EXP5060 and the SATA disk drives that it supports, the best usage environment is a high throughput environment. An array design that uses the full stripe write technique is best suited. Hence, we recommend that you use either RAID5 or RAID6 to give you the throughput boost. Table 3-1 on page 18 shows the expected maximum transaction and throughput performance for the RAID5 and RAID6 array types.

Table 3-1 RAID5 and RAID6 performance comparison

Operation	Non-trunked	Trunked
Sustained I/O rate disk reads: RAID5	34,000 IOPS	34,000 IOPS
Sustained I/O rate disk reads: RAID5	7,000 IOPS	7,000 IOPS
Sustained I/O rate disk reads: RAID6	34,000 IOPS	34,000 IOPS
Sustained I/O rate disk reads: RAID6	4,500 IOPS	4,500 IOPS
Sustained throughput disk reads: RAID5	3,150 MB/sec	6,300 MB/sec
Sustained throughput disk writes: RAID5	2,650 MB/sec	5,300 MB/sec
Sustained throughput disk writes, CME <sup>a</sup> : RAID5	2,650 MB/sec	3,800 MB/sec
Sustained throughput disk reads: RAID6	3,150 MB/sec	6,200 MB/sec
Sustained throughput disk writes: RAID6	2,400 MB/sec	4,700 MB/sec
Sustained throughput disk writes, CME <sup>a</sup> : RAID6	2,400 MB/sec	3,400 MB/sec
Drive count used	220 SATA drives	220 SATA drives

a. CME refers to "Cache Mirroring Enabled" option

As mentioned earlier and illustrated in Table 3-1, this solution is not the most practical for transaction (IOPS)-intense environments.

## 3.2 Array groups and logical drives

With the EXP5060, several design factors influence the configuration of your arrays and LUNs for maximum performance. Consider these major areas:

- ▶ Number of disks per array group
- ▶ Trunking
- ▶ Selection of the members of the array group
- ▶ Number of LUNs per array group
- ▶ Spreading and balancing the workload evenly across all resources

We cover each of these areas in detail.

### 3.2.1 Number of disks per array group

With the EXP5060, the number of drives per array depends on the environment with which the subsystem is used. With the SATA drives, small arrays of only a few drives can result in serious performance bottlenecks. In certain cases, small arrays of only a few drives cause the processes to back up. However, it is also important not to make the array group so large that

the rebuild times are long. For these reasons, we recommend using array groups that consist of five to nine drives for RAID5. And, we recommend using array groups that consist of six to 10 drives for RAID6. In certain cases, you can use larger arrays with the understanding that they have longer rebuild times. But in most cases, smaller arrays encounter conflicts and excessive I/Os per drive and are unable to align block size for matching the “full stripe write” pattern.

In testing, we used four to eight data drives to gain the optimal throughput performance with the SATA drives. These configurations attained the maximum throughputs for these array layouts.

### 3.2.2 Trunking

With the EXP5060, trunking is a new feature that enables you to achieve high throughput with a minimum number of storage expansions and disks. For this capability, we use the four ports of a channel pair to drive the arrays and LUNs in a single EXP5060. Therefore, we effectively give the 60 drives a full 1.6 GB of bandwidth for their use. This capability can enhance the throughput of a single EXP5060 up to twice that of a standard non-trunked configuration, as shown in Table 3-1 on page 18. However, you must follow certain rules and be aware of specific restrictions with this configuration. We discuss both methods of attaching the EXP5060, so that you can determine the best method to implement for your specific environment.

Table 3-1 on page 18 shows that using trunking does not provide a transaction (IOPS) rate advantage. I/O capability is purely a result of the number of spindles. When the configurations are the same, there is no benefit with the increase in bandwidth.

With these facts understood, it is easy to see that the advantage of the EXP5060 lies with its improved throughput handling capabilities or, when needed, its large capacity for use as a low-cost archival system.

#### ***EXP5060 non-trunked configurations***

In the non-trunked configuration, you can attain high capacities with mixed environments of EXP5000 and EXP5060 expansions sharing channels and subsystem resources. In these environments, we recommend that you follow the best practices that are used with the EXP5000 as the basis for the layout of the arrays and LUNs. This way, you can avoid contention with the drive channel switches and controllers. Also, spread the slot selection evenly between odd and even slots to balance the loop port utilization for all arrays and LUNs.

Figure 3-2 shows an example of a layout for a configuration of four EXP5060s that are used for a layout of 8 + P RAID5 or 7 + P + Q RAID6 array groups using two expansions to provide arrays and LUNs to controller A and two expansions to provide arrays and LUNs for controller B. The layout in Figure 3-2 aligns with the dedicated channel model, which is a best practice for both EXP5000 and EXP5060 in the non-trunked configuration.

**Tip:** You can use any RAID5 layout as a RAID6 layout for a configuration with one less user data drive capacity.

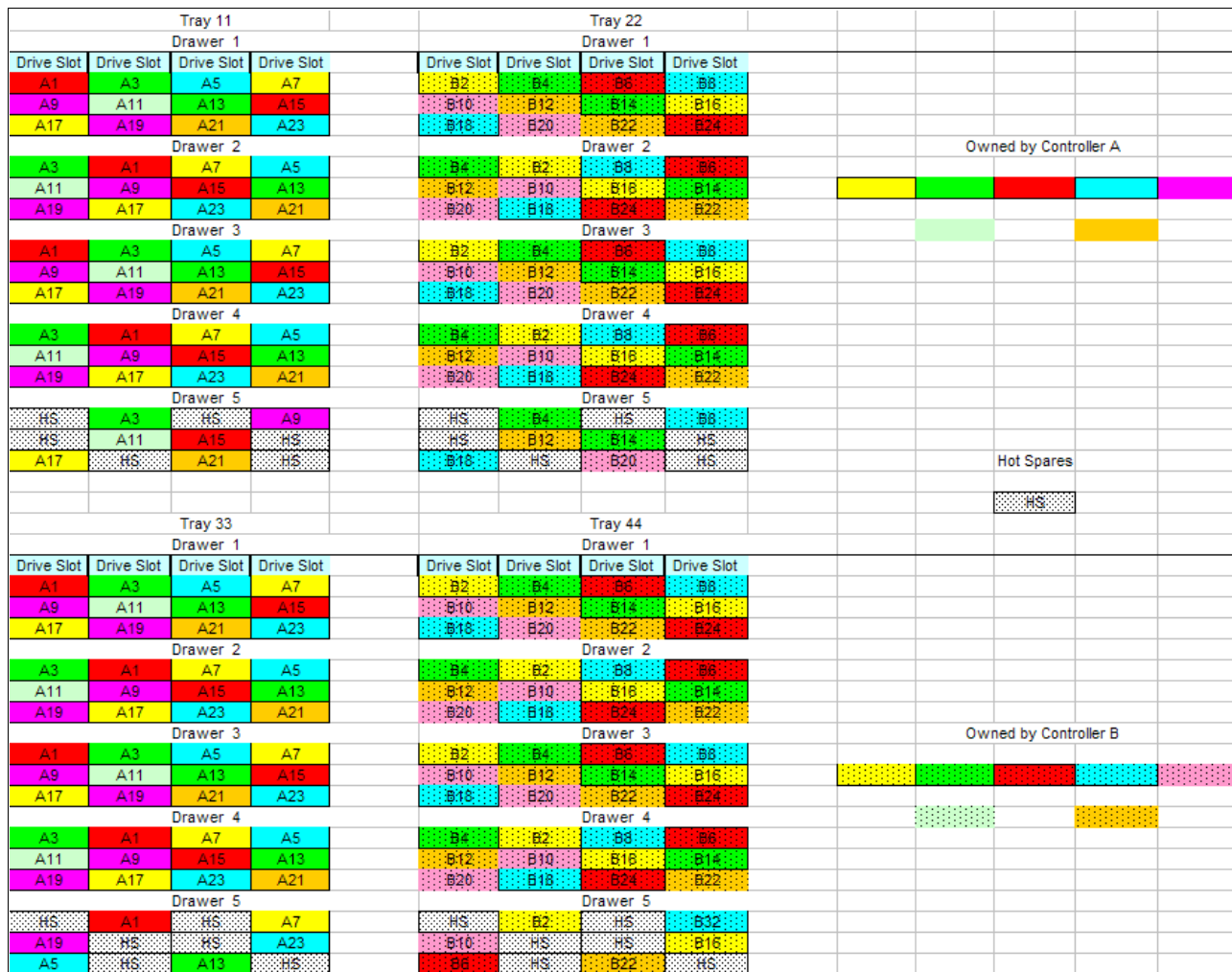
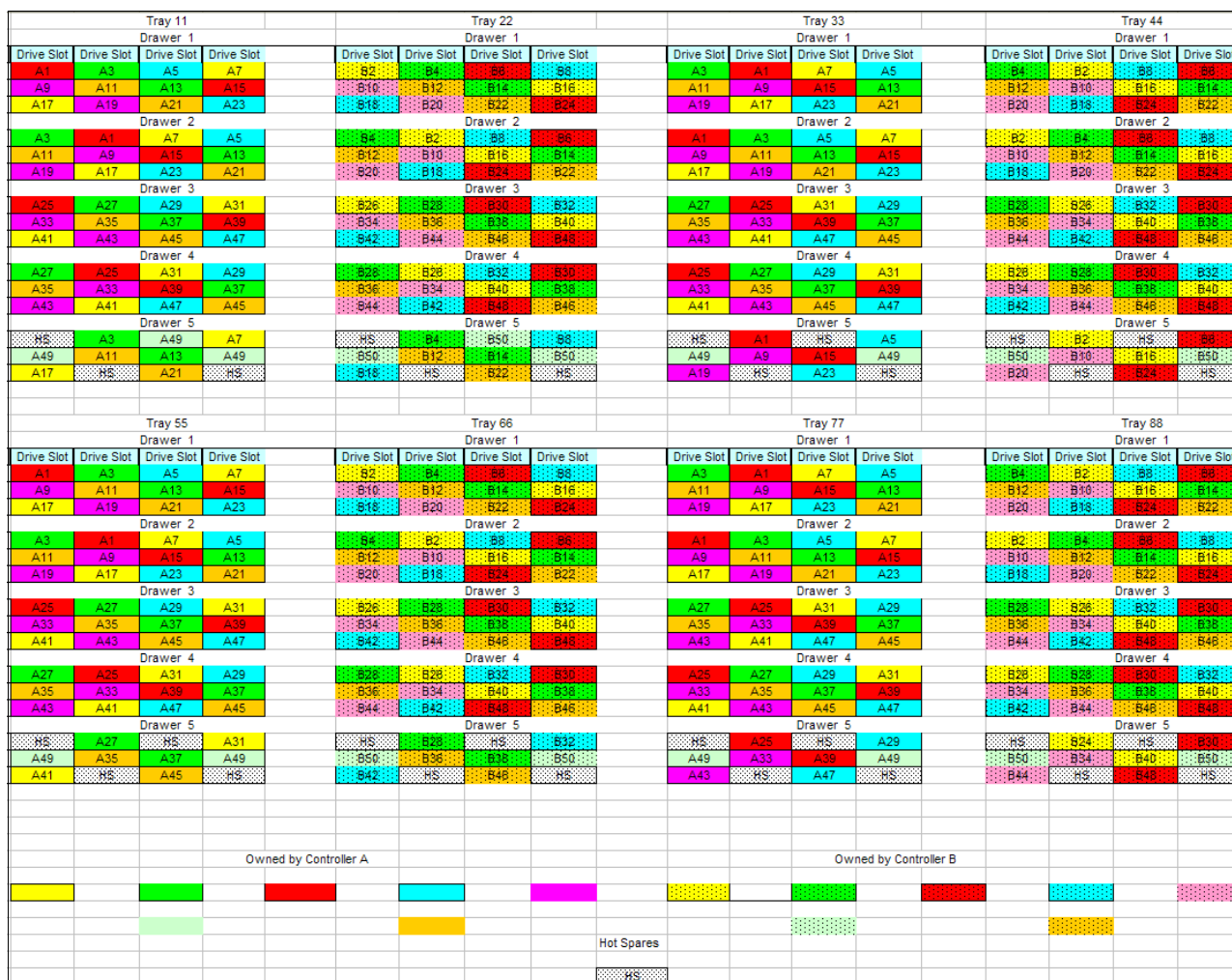


Figure 3-2 Example of four EXP5060s in a non-trunked 8 + P configuration

In dedicated EXP5060 environments, you can use the entire DS5000 storage subsystem to drive up to eight EXP5060 expansions. *When configuring the maximum configuration of eight EXP5060s, there must not be any EXP5000s in the configuration.*

Figure 3-3 shows an example of a layout for a configuration of eight EXP5060s that are used for a maximum configuration with a layout of 8 + P RAID5 or 7 + P + Q RAID6 array groups across all eight expansions. With fewer EXP5060s, you can create shared configurations with EXP5000s on certain channels or cascaded with the EXP5060, creating mixed environments for shared workload patterns.



### Trunked configuration of the EXP5060

With the new trunked design of the EXP5060 expansion, there is more to consider when planning the layout of arrays and LUNs. With the new expansion, we also have two additional switches that are built into the Environmental Service Module (ESM) that are not available in the earlier designs. With the EXP5060, the Fibre Channel drawer consists of a two drawer control monitor (DCM) with one drawer for each of the loop pairs. The DCM uses a microcontroller to manage the environmental data and enclosure control over the I/O loop and System on a Chip (SOC) switch path controls. When configuring the EXP5060 in a trunked configuration for maximum throughput, it is important that you avoid sharing the drive drawers between controllers whenever possible. Create arrays and LUNs by using the disks from two drawers for arrays and LUNs to be owned by controller A, and two other drawers to be owned by controller B. Then, you can use the fifth drawer for hot spares. If all 60 drives are needed, limit the drives using the fifth drawer to only arrays that are built to be used by only one controller to prevent the effect of shared DCMs. Figure 3-4 shows a full four EXP5060 configuration that is built with 8 + P RAID5 or 7 + P + Q RAID6.

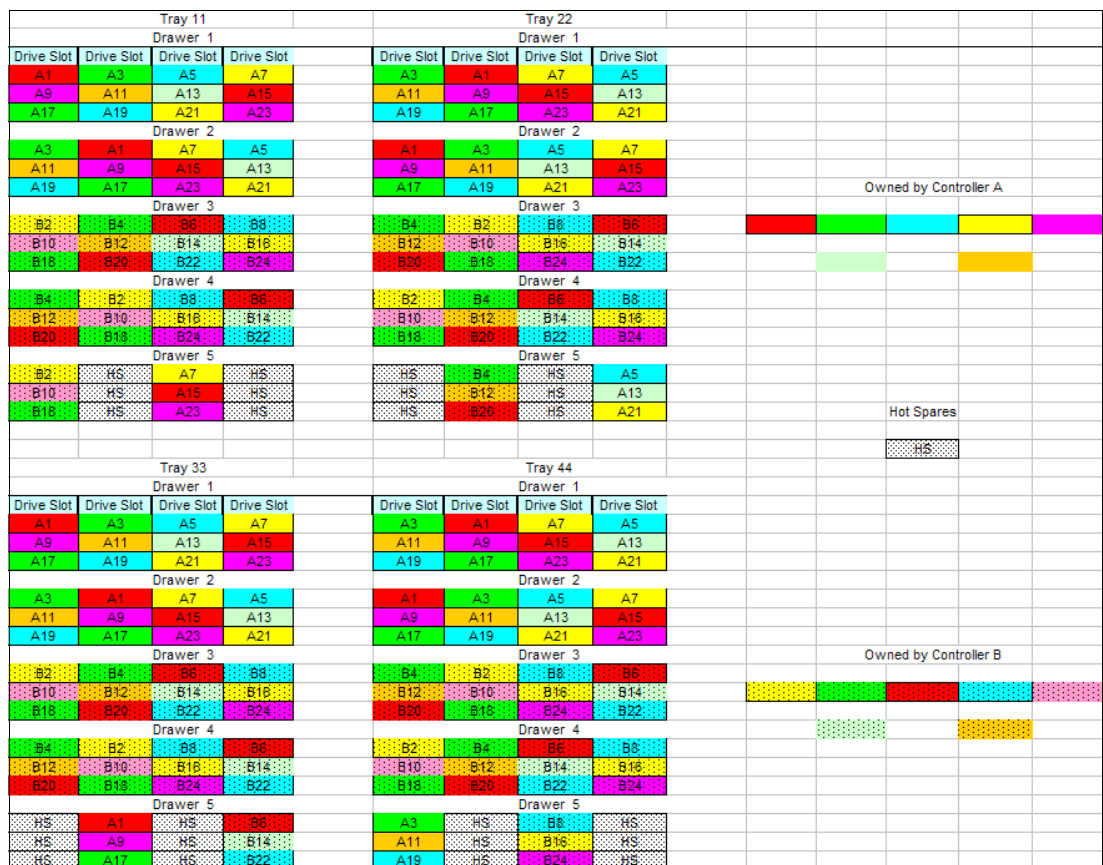


Figure 3-4 Trunked EXP5060 configured with 8 + P arrays with drawer protection

Follow these recommendations to help you plan for smaller configurations or when you want maximum performance:

1. Even when building a small configuration, make good use of all of the available resources for the best performance.
2. Dedicate two drawers, drawers 1 and 2 in the example in Figure 3-4, from all of the expansion trays to build arrays and LUNs that are owned by Controller A.
3. Dedicate two other drawers, drawers 3 and 4 in the example in Figure 3-4 on page 22, from all of the expansion trays to build arrays and LUNs that are owned by Controller B.



4. When using the fifth drawer, build separate array groups (A21, A23, B22, and B24 in the example in Figure 3-4 on page 22) that are split equally between Controller A and B. This design allows the shared DCM to have minimal effect and still spreads the workload as evenly as possible across all of the drive channels.

**Tip:** When less user capacity is needed and layout requirements allow, using only the four disks that are required in drawer 5 for hot spares is the best model for maximum performance. You can create this layout with a 7 + P configuration.

You can design many layouts with the trunked configuration, and if you follow the recommendations, you can achieve good throughput.

### High availability features to use

Clients have always been able to build arrays with enclosure loss protection with the drive layout selection. With earlier expansions and code releases, this capability carried a high degree of importance. However, with the newer switched technology and more robust code design, this requirement has lesser importance. The new EXP5060 expansion design adds a new level of protection at the drawer level. With this protection, you can create an array that spans the expansion trays and is built across multiple drives in separate drawers to incorporate multiple paths for improved throughput performance. This capability helps to strengthen the robustness of your solution. This protection helps to alleviate concerns about enclosure (tray) loss protection when you build an environment that spans four EXP5060s in a trunked environment. To help in this area, you can build the array group by selecting a drive of each of the associated drawers that are preferred to the specific controller, allowing you to use an increased array group size. Additionally, with RAID6, you can extend the number of members in an enclosure to two and still have parity coverage.

You can create many configurations that can provide either optimal performance or maximum protection, but only a few configurations can provide the best levels for both performance and protection. Building these environments requires extensive planning and careful implementation to avoid conflicts and degraded performance. We provide the following sample configurations as additional references to help you with your planning efforts. The following examples are only a subset of all of the possible choices.

Figure 3-5 shows a RAID10 configuration that incorporates the drawer protection best practice.

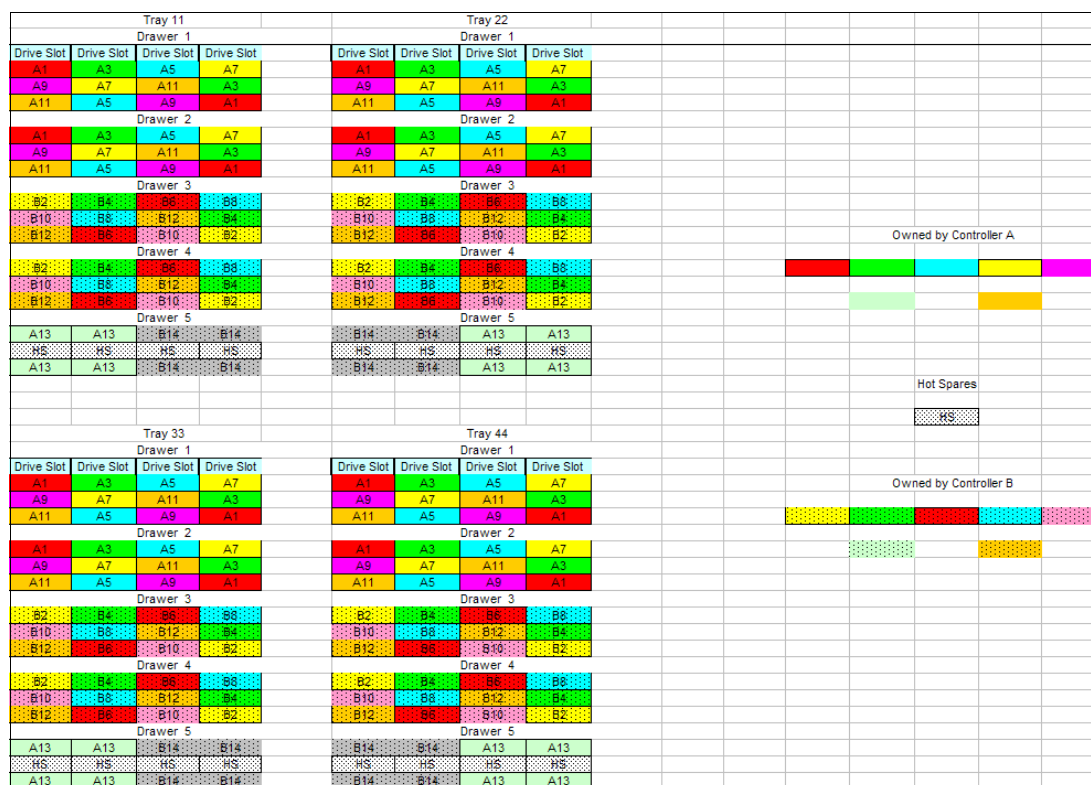


Figure 3-5 Trunked EXP5060 configured with 8 + P RAID10 arrays with drawer protection

For “*High Performance Computing*” environments, we frequently are forced to use small arrays of 4 + P RAID5 layouts. For these environments, we recommend using the layout in Figure 3-6. In this configuration, we build the arrays to include drawer loss protection. Without this requirement, you can improve performance; however with this configuration, you can attain reasonable levels of throughput. See Chapter 4, “EXP5060 performance” on page 27 for details about performance differences.

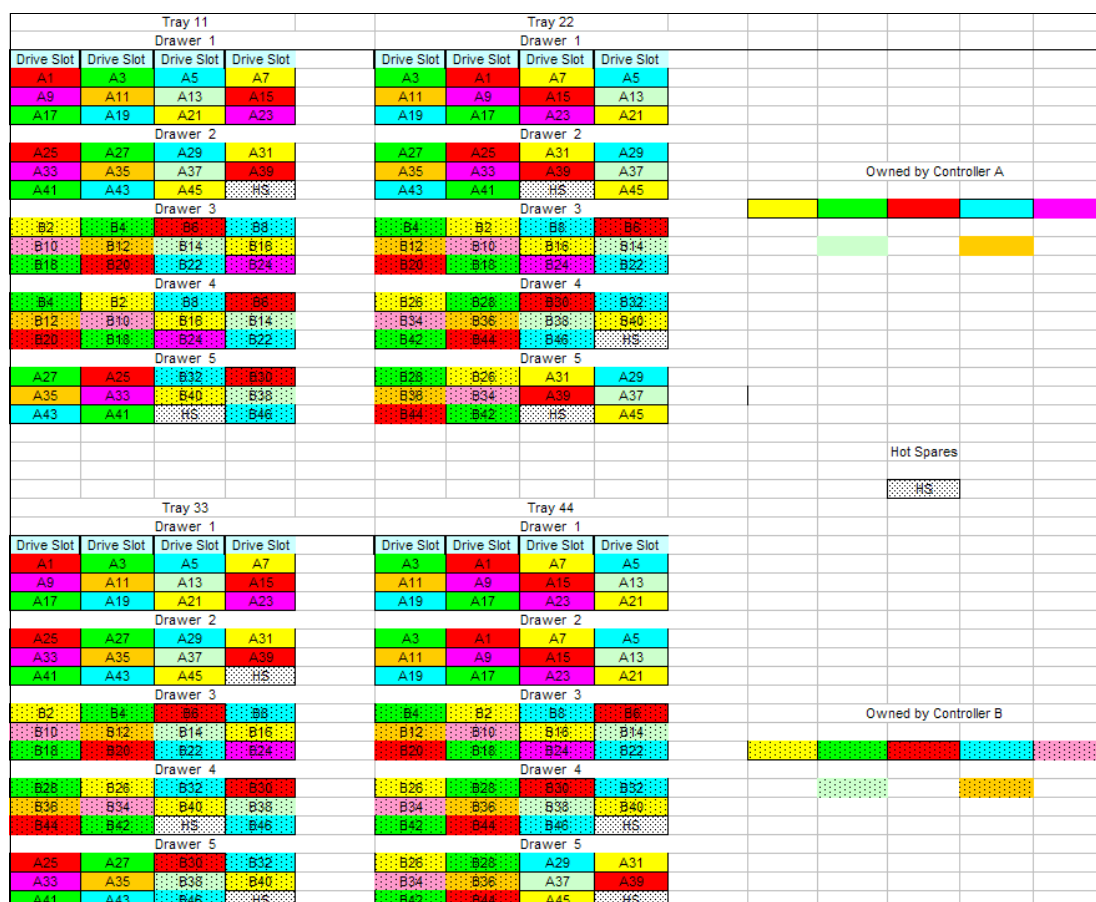


Figure 3-6 Trunked EXP5060 with 4 + P RAID5 arrays

Although you can use the 4 + P layout in Figure 3-6 for a 3 + P + Q RAID6 configuration, we do not recommend it due to too few data drives in the arrays.





## EXP5060 performance

This chapter provides the performance results that we collected during various test scenarios. We used the DS5300 and a configuration of four IBM System Storage EXP5060 High Density Storage Enclosures (EXP5060s) that were installed in a trunked configuration.

**Important:** We reached the performance numbers in this chapter by using test programs in a lab environment. Your results probably will vary from case to case.

## 4.1 Performance test runs

Because the major focus of the EXP5060 expansion is the throughput environment, most of the tests that we performed were to show the maximum throughputs possible with these configurations. In many of the configurations that we used, we did not consider a need for a high availability solution. You need to include planning for high availability when you plan your solution.

We gathered part of these results with best practice configurations for performance, but not all. We wanted to show the differences that you can encounter. In certain cases, the configurations that show the best performance do not fit a high availability model. In these cases, you need to base your choices on how best to meet your business needs, as discussed in Chapter 3, “Configuring the EXP5060” on page 15. We discuss options in this chapter that can help in the decision-making process.

### 4.1.1 Optimal performance layouts

To attain the best possible performance, you must follow these rules to avoid any chance of contention or workload imbalance in the configuration of the array and logical unit number (LUN) layout. We list these rules in the order of their effect on performance:

1. You must build array groups across a balance of odd and even drive slots.
2. You must share each expansion tray evenly between the two controllers by assigning two drawers to controller A and two drawers to controller B from each expansion tray.
3. There is no imbalance in the use of the channels.
4. There are no shared drawers being used.
5. Define hot spares in the front row of the drives in the fifth drawers of the trays to use the required disks that are installed in that drawer.

Sometimes, you might not be able to follow these rules. In certain cases, you might need to make exceptions to these rules to meet the configuration needs of your environment. In these cases, plan to implement the exception so that it has the least effect on the overall configuration. An example of an exception is when the fifth drawer of the EXP5060 is used for arrays and LUNs to be built. See Chapter 3, “Configuring the EXP5060” on page 15 for the best practices in this situation.

**Availability:** Depending on the array group size, it might be difficult to ensure enclosure and drawer loss protection. You need to understand the availability requirements.

We defined the following example test case layouts in 8 + P RAID5 array groups (Figure 4-1). Our testing demonstrated the maximum throughput performance numbers that are shown in Table 4-1.

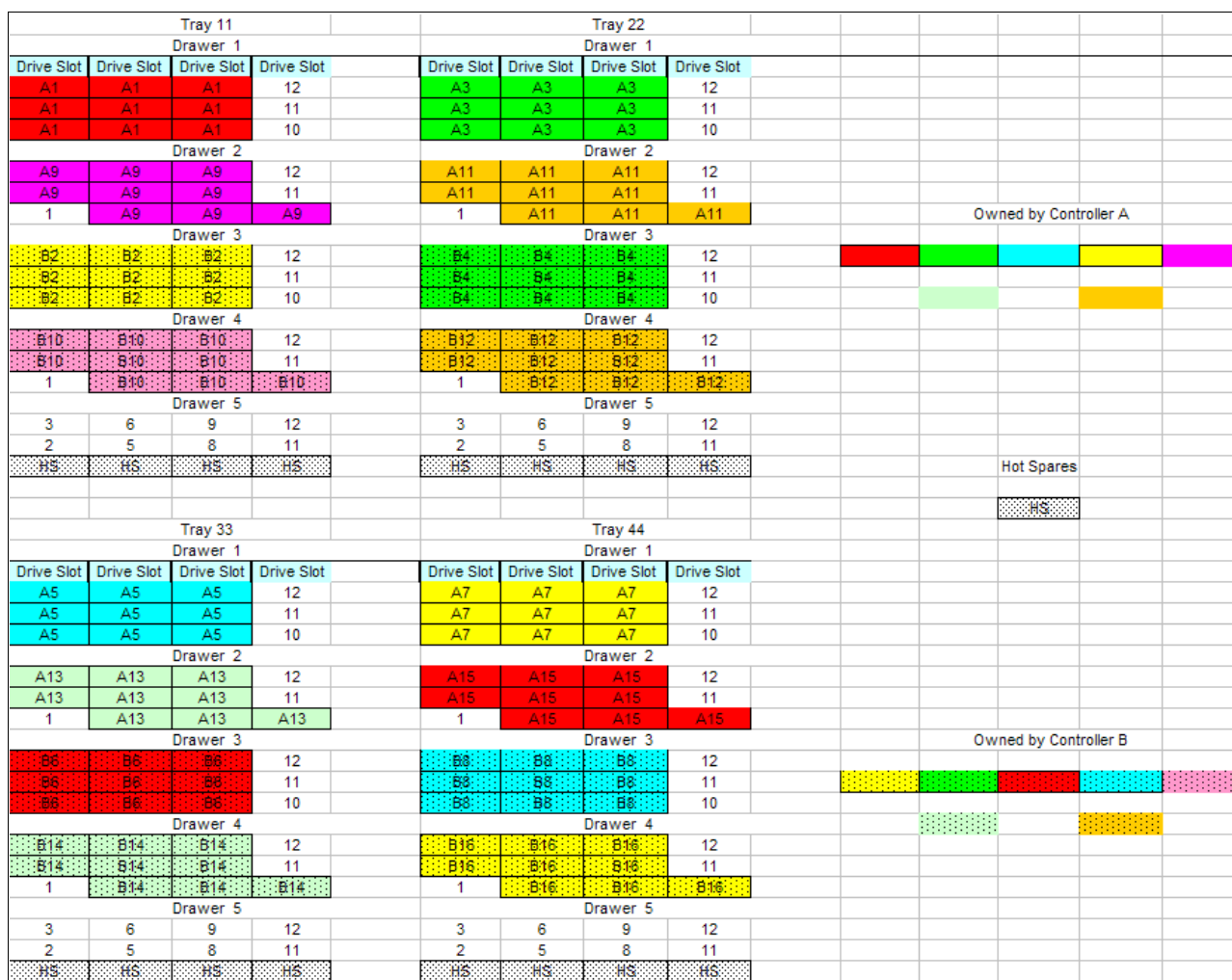


Figure 4-1 An 8 + P single enclosure and a single drawer array group layout

Table 4-1 Maximum performance test results as seen with the Figure 4-1 configuration

Measurement	Result
Throughput read performance	6339.48 MB/sec
Throughput write performance	5369.19 MB/sec
Throughput write with Cache Mirroring Enabled (CME) performance	3803.21 MB/sec

Figure 4-2 on page 30 shows 8 + P shared enclosures with dedicated drawers for the array group layout. Table 4-2 on page 30 shows the maximum performance test results for the Figure 4-2 on page 30 configuration.

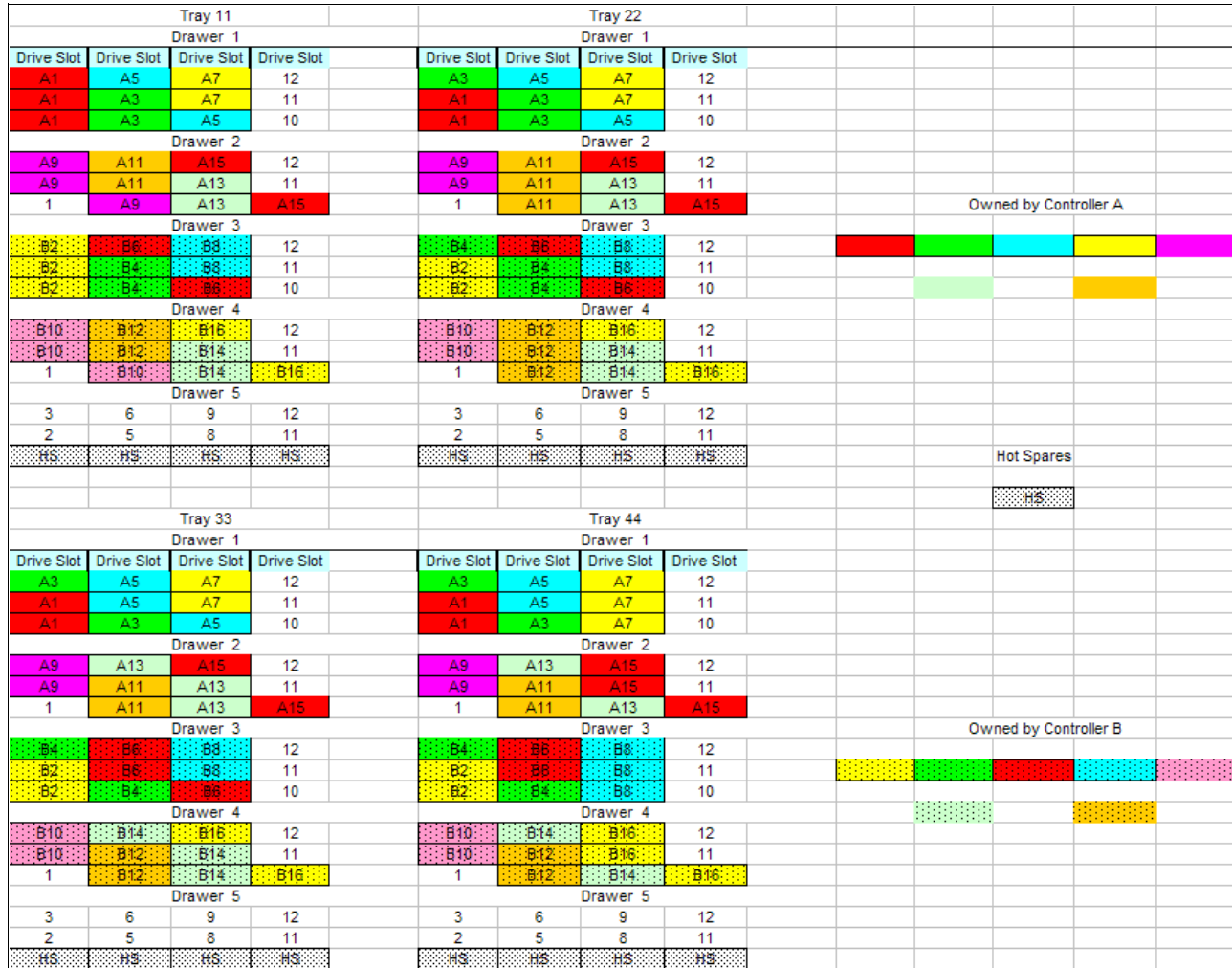


Figure 4-2 Test with an 8 + P shared enclosures and dedicated drawers for array group layout

Table 4-2 Maximum performance test results as seen with the Figure 4-2 configuration

Measurement	Result
Throughput read performance	6320.76 MB/sec
Throughput write performance	5373.92 MB/sec
Throughput write with CME performance	3801.89 MB/sec



As shown in Table 4-1 on page 29 and Table 4-2 on page 30, with Figure 4-1 on page 29 and Figure 4-2 on page 30, these tests adhered to the following best practice recommendations:

- ▶ Dedicate two drawers (in this example, drawers 1 and 2) to build arrays and LUNs that are owned by Controller A. Dedicate two drawers (in this example, drawers 3 and 4) to build arrays and LUNs that are owned by Controller B.
- ▶ Ensure that the array groups were created with as equal a balance of odd and even disks slots as possible for balanced channel or Environmental Service Module (ESM) workloads. Figure 4-2 on page 30 shows this practice. We have an odd number of drives in an array, but we have made sure to build them as equal arrays as much as possible by alternating the ninth drive of each array between the odd and even slots to keep the channels balanced.

Enclosure protection and drawer loss protection are not provided with these configurations. Make sure that you understand your availability requirements. See Chapter 3, “Configuring the EXP5060” on page 15 for guidance in this area. In certain cases, you might only need a slight change in the configuration to meet the high availability requirement (at a small cost to the performance expectations).

### 4.1.2 Non-optimal performance layouts

The following configurations are examples of how poorly the EXP5060 environment can perform when you do not adhere to the best practice rules. We defined these example test case layouts in 8 + P RAID5 array groups so that you can easily compare them to the previous optimal test cases. These examples are worst-case scenarios. While you are in the planning stages, use these examples to help you estimate the effect on performance.

Figure 4-3 on page 32 shows non-shared drawers and expansion trays. Table 4-3 on page 32 shows the maximum performance test results as seen with the Figure 4-3 on page 32 configuration.

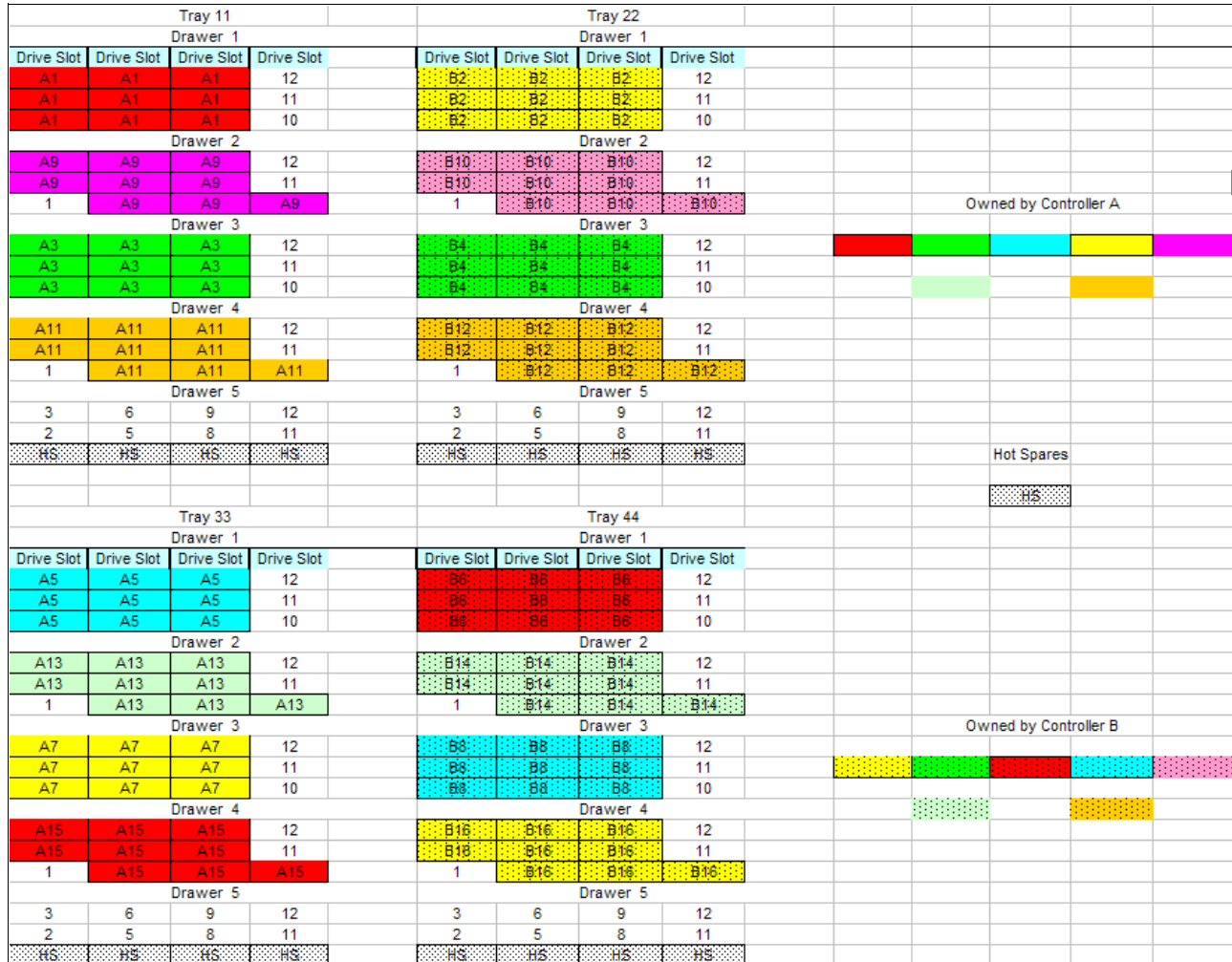


Figure 4-3 A balanced, non-shared drawers or expansions 8 + P layout

Table 4-3 Maximum performance test results as seen with the Figure 4-3 configuration

Measurement	Result
Throughput read performance	3172.75 MB/sec
Throughput write performance	2700.24 MB/sec
Throughput write with CME performance	2692.88 MB/sec

As shown in Table 4-3, the performance of the configuration that is shown in Figure 4-3 is about half of the performance of the two trunked connected channels. The same controller owns all of the array groups and LUNs in the expansion; therefore, it only uses half of the bandwidth.

**Important:** Dividing the drawers in an expansion tray between the controllers is the most critical of all guidelines when using the trunking method to gain increased performance. Failure to adhere to this rule nullifies the performance benefit.

Figure 4-4 shows non-balanced access across the drive channels by controllers in an 8 + P layout. Table 4-4 shows the maximum performance test results that were seen with the Figure 4-4 configuration.

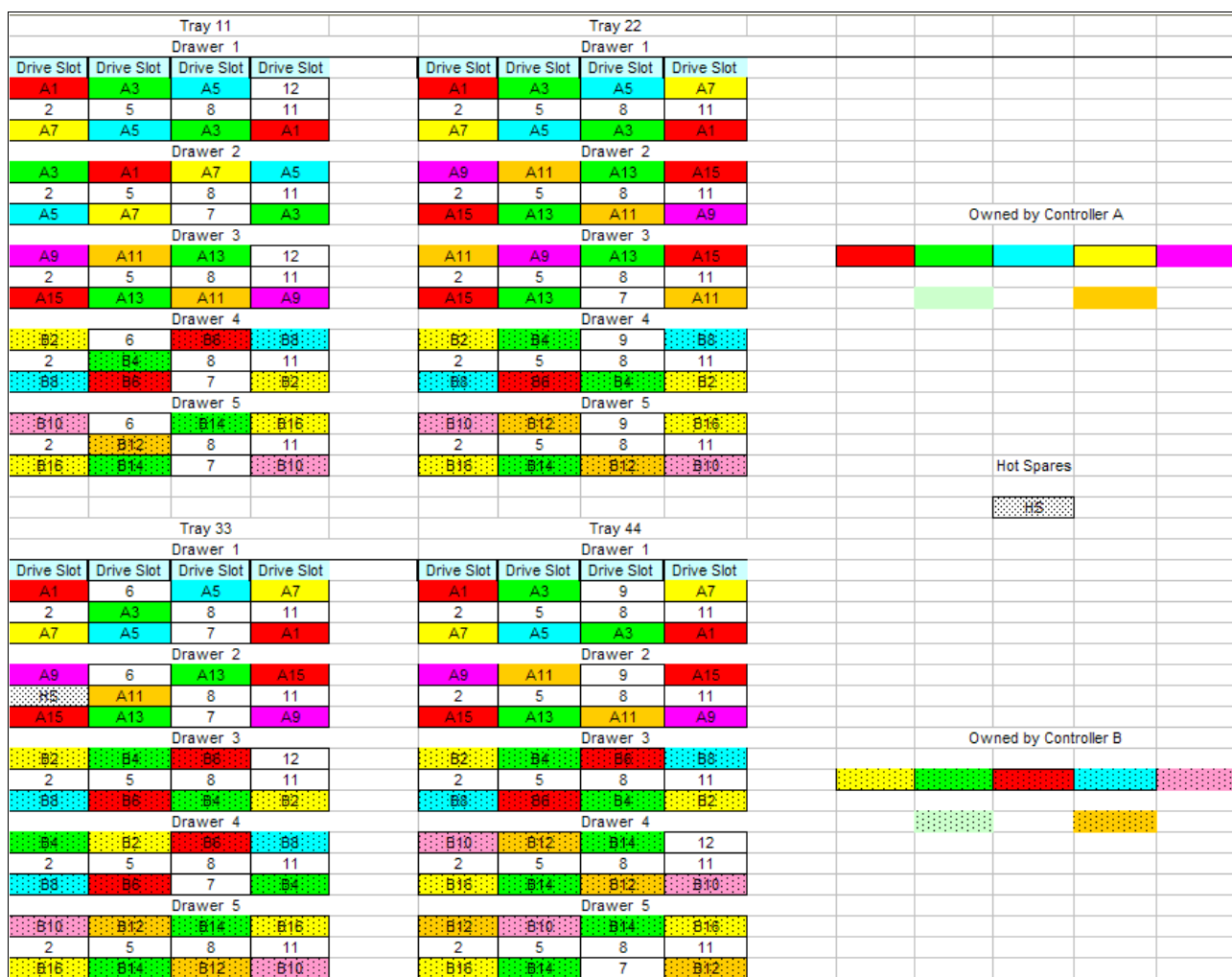


Figure 4-4 Non-balanced access across the drive channels by controllers in an 8 + P layout

Table 4-4 Maximum performance test results that were seen with the Figure 4-4 configuration

Measurement	Result
Throughput read performance	5183.80 MB/sec
Throughput write performance	4412.15 MB/sec
Throughput write with CME performance	3800.59 MB/sec

In Figure 4-4, we share the expansions between the two controllers, which helps to gain access to the full bandwidth of the trunked pair. However, with the addition of the fifth drawers being brought into the mix, we see an imbalance in the channel usage. Controller A has a higher drive count on expansion trays 11 and 12, and controller B has a higher number on expansion trays 33 and 44. As shown in Table 4-4, the performance is fair and might be acceptable, but it is not at an optimal level. When you need to use all five drawers, it is a better layout to split the fifth drawer's disks between the two controllers and to build dedicated arrays

and LUNs with them to minimize the performance effect. Refer to Chapter 3, “Configuring the EXP5060” on page 15 for a configuration example that helps to minimize this negative effect.

Figure 4-5 shows sharing the device control manager (DCM) with controllers in an 8 + P layout. Table 4-5 shows the maximum performance test results that were seen with the Figure 4-5 configuration.

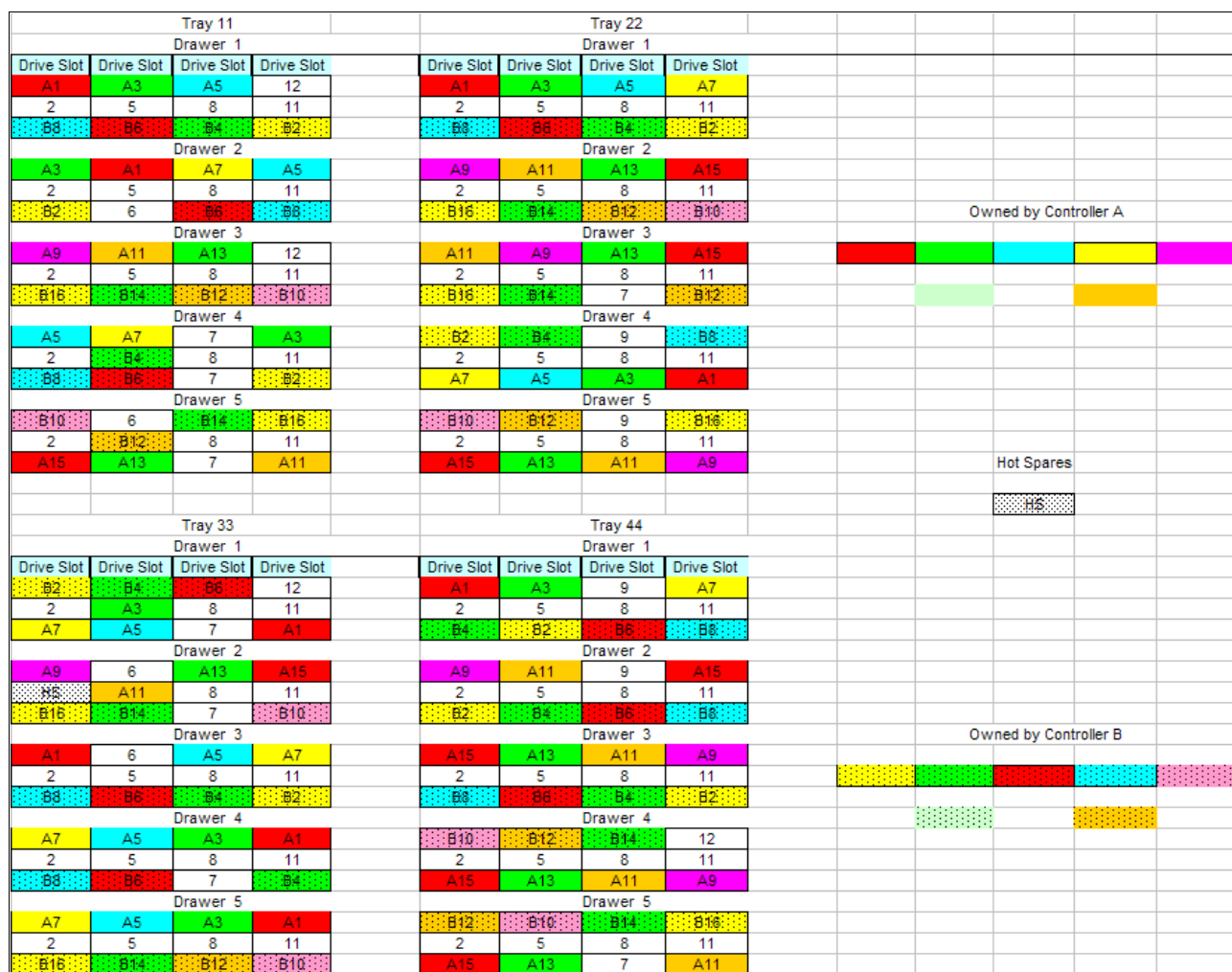


Figure 4-5 Sharing the DCM with controllers in an 8 + P layout

Table 4-5 Maximum performance test results as seen with the Figure 4-5 configuration

Measurement	Result
Throughput read performance	5451.83 MB/sec
Throughput write performance	4516.85 MB/sec
Throughput write with CME performance	3799.36 MB/sec

In Figure 4-5, we share the expansions between the two controllers, which helps to gain access to the full bandwidth of the trunked pair. However, with the addition of the fifth drawers being brought into the mix, we see an imbalance in the DCMs. As shown in Table 4-4 on page 33, the performance is fair and might be acceptable, but it is not at an optimal level. When you need to use all five drawers for storage capacity and drive usage, refer to the

recommendations that are outlined in Chapter 3, “Configuring the EXP5060” on page 15 for ways to minimize negative effects.

Figure 4-6 shows a non-balanced single tray in a single drawer 8 + P configuration. Table 4-6 shows the maximum performance test results that were seen with configuration in Figure 4-6.

Tray 11				Tray 22							
Drawer 1				Drawer 1							
Drive Slot	Drive Slot	Drive Slot	Drive Slot	Drive Slot	Drive Slot	Drive Slot	Drive Slot				
A1	A1	A1	12	A3	A3	A3	12				
A1	A1	A1	11	A3	A3	A3	11				
A1	A1	A1	10	A3	A3	A3	10				
Drawer 2				Drawer 2							
A9	A9	A9	12	A11	A11	A11	12				
A9	A9	A9	11	A11	A11	A11	11				
A9	A9	A9	10	A11	A11	A11	10				
Drawer 3				Drawer 3				Owned by Controller A			
B2	B2	B2	12	B4	B4	B4	12				
B2	B2	B2	11	B4	B4	B4	11				
1	B2	B2	B2	1	B4	B4	B4				
Drawer 4				Drawer 4							
B10	B10	B10	12	B12	B12	B12	12				
B10	B10	B10	11	B12	B12	B12	11				
1	B10	B10	B10	1	B12	B12	B12				
Drawer 5				Drawer 5							
3	6	9	12	3	6	9	12				
2	5	8	11	2	5	8	11				
HS	HS	HS	HS	HS	HS	HS	HS	Hot Spares			
								HS			
Tray 33				Tray 44							
Drawer 1				Drawer 1							
Drive Slot	Drive Slot	Drive Slot	Drive Slot	Drive Slot	Drive Slot	Drive Slot	Drive Slot				
A5	A5	A5	12	A7	A7	A7	12				
A5	A5	A5	11	A7	A7	A7	11				
A5	A5	A5	10	A7	A7	A7	10				
Drawer 2				Drawer 2							
A13	A13	A13	12	A15	A15	A15	12				
A13	A13	A13	11	A15	A15	A15	11				
A13	A13	A13	10	A15	A15	A15	10				
Drawer 3				Drawer 3				Owned by Controller B			
B6	B6	B6	12	B8	B8	B8	12				
B6	B6	B6	11	B8	B8	B8	11				
1	B6	B6	B6	1	B8	B8	B8				
Drawer 4				Drawer 4							
B14	B14	B14	12	B16	B16	B16	12				
B14	B14	B14	11	B16	B16	B16	11				
1	B14	B14	B14	1	B16	B16	B16				
Drawer 5				Drawer 5							
3	6	9	12	3	6	9	12				
2	5	8	11	2	5	8	11				
HS	HS	HS	HS	HS	HS	HS	HS				

Figure 4-6 Non-balanced single tray in a single drawer 8 + P configuration

Table 4-6 Maximum performance test results as seen with the Figure 4-6 configuration

Measurement	Result
Throughput read performance	5714.90 MB/sec
Throughput write performance	4844.83 MB/sec
Throughput write with CME performance	3802.34 MB/sec

In the Figure 4-6 configuration, we see the results of an array layout that meets most of the rules for a best practice configuration, but the layout fails to evenly balance the arrays across the odd and even drives for a balanced use of the connected loops. Table 4-6 shows performance numbers that might be at an acceptable level, but they miss the optimal performance that can be achieved with a slight drive selection layout change (see Figure 4-1 on page 29, for example).

## 4.2 Mixed configuration test runs

The best usage environment for the EXP5060 expansion enclosure and the Serial Advanced Technology Attachment (SATA) drives is to support a high throughput-based transaction volume. However, clients can use this high-capacity expansion enclosure in many nearline environments. Often, the environment supports a mix of workloads. To help you in this planning area, we created a scenario with various arrays and LUNs and tested these arrays and LUNs with multiple host I/O workloads. We focused our testing on the best practices in these environments. We used the *iometer* data generator test tool.

### **Multiple arrays and LUNs in a variety of configurations**

Figure 4-7 shows the initial configuration of various 4 + P arrays and LUNs that was used to test a variety of workload types. With the DS5300 and the EXP5060s configured in this manner, we built an array group that used many creation patterns of trays and drawers. Two LUNs on each tray used 128 KB and 512 KB segment sizes. We used various I/O block sizes to gather data points for a variety of sequential I/O workloads as well as a random pattern test run. We tested these areas:

- ▶ Maximum throughput capabilities with various sequential host block sizes
- ▶ Maximum I/O per second (IOPS) capabilities with various random host block sizes
- ▶ A test case of the effects on a non-optimal configuration

Figure 4-7 shows a test of a trunked EXP5060 with variety of 4 + P test array configurations.

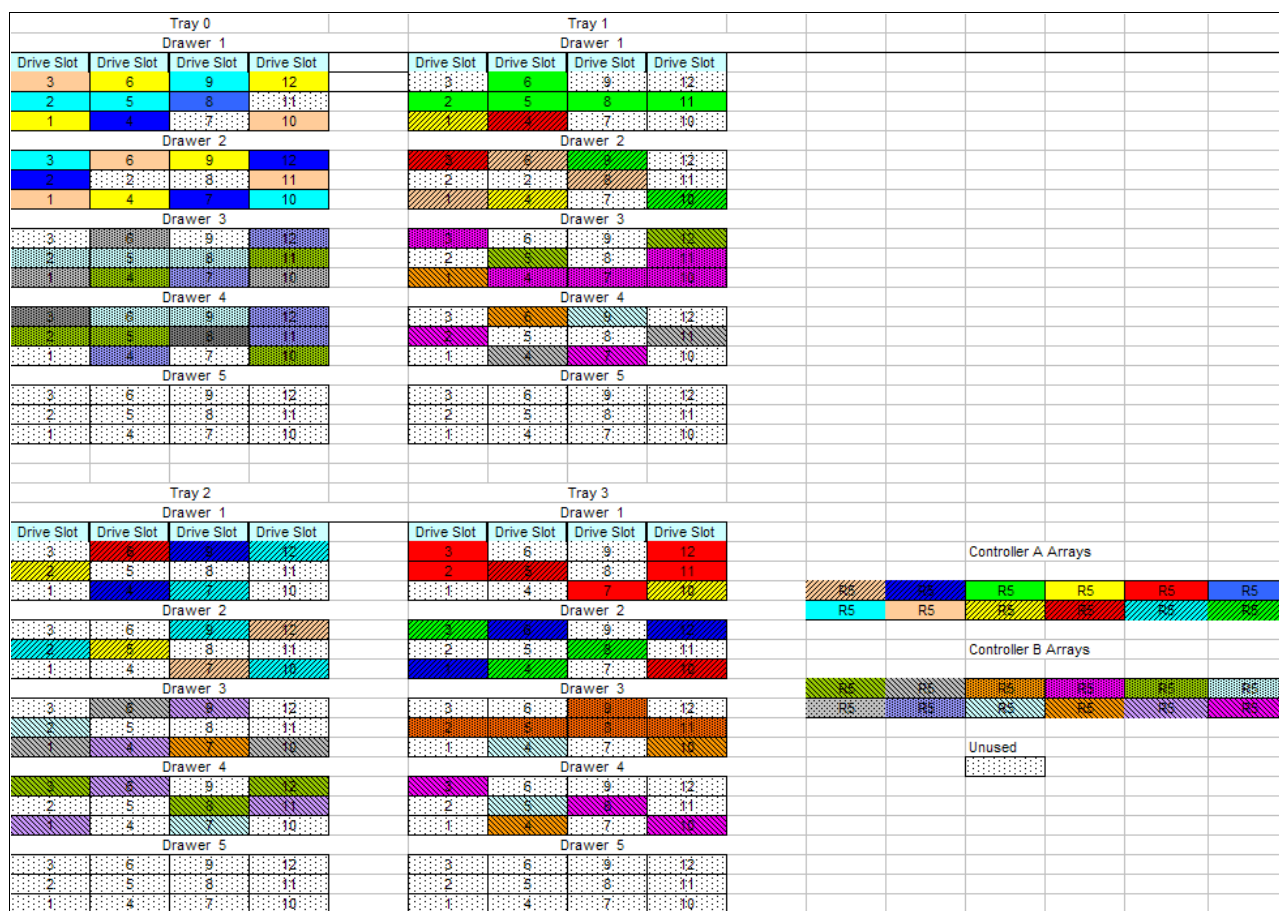


Figure 4-7 Trunked EXP5060 with variety of 4 + P test array configurations

We focused only on the best practices that affected the performance in this test configuration layout. We created the array groups in the following manner:

- ▶ Single EXP5060 with multiple drawers
- ▶ Single EXP5060 with a single drawer
- ▶ Multiple EXP5060s with multiple drawers

To avoid affecting the performance, we did not use Drawer 5.

**Drawer 5:** When using Drawer 5 for arrays and LUNs, use the drives in the drawer to build separate arrays and LUNs that are evenly spread across the two controllers. This layout helps to minimize the effect of the shared drawer on the overall performance. See Chapter 3, “Configuring the EXP5060” on page 15 for details.

We configured each array group with two LUNs of equal size using 128 KB segment size for one LUN and 256 KB segment size for the other LUN. We wanted to test the effects of multiple host I/O block sizes using separate segment sizes. Table 4-7 shows the various results for numerous sequential workload runs with 96 user sessions configured across four hosts. We used host block sizes of 256 KB, 512 KB, 1 MB, and 2 MB to simulate various sequential application types. We collected these results for read and write runs with a queue depth of 1 for reads and a queue depth of 121 for writes. We disabled all cache (both read and write) for the runs that are reported in Table 4-7 to obtain results that were as close to disk throughput results as possible.

*Table 4-7 Maximum throughput with no cache enabled*

Workload operation type	Host block size of I/O	Throughput MB/sec
Sequential read	256 KB	6.889 GB
Sequential read	512 KB	6.915 GB
Sequential read	1 MB	6.914 GB
Sequential read	2 MB	6.914 GB
Sequential write	256 KB	4.008 Gb
Sequential write	512 KB	4.428 GB
Sequential write	1 MB	4.644 GB
Sequential write	2 MB	4.632 GB

The following results show running the same workload test with all cache enabled. In this case, we used varying queue depths to improve the throughput levels. Table 4-8 on page 37 shows the throughput results for queue depths of 1 and 121, as seen in our test runs.

*Table 4-8 Maximum throughput with cache mirroring enabled*

Workload operation type	Host block size of I/O	Throughput MB/sec
Sequential read	256 KB	4.702 GB/6.915 GB
Sequential read	512 KB	5.695 GB/6.917 GB
Sequential read	1 MB	6.509 GB/6.916 GB
Sequential read	2 MB	6.899 GB/5.857 GB
Sequential write	256 KB	3.561 GB/3.239 GB

Workload operation type	Host block size of I/O	Throughput MB/sec
Sequential write	512 KB	3.976 GB/3.265 GB
Sequential write	1 MB	3.330 GB/3.871 GB
Sequential write	2 MB	3.822 GB/4.191 GB

As shown in Table 4-8, the throughput is affected by enabling write cache mirroring. The effect on the throughput is lessened by the increases in the amount of full stripe performance that is reached (with the larger block size). The effect on the throughput is lessened due in part to the number of LUNs that are able to perform full stripe writes at this block size versus the smaller block sizes. In Table 4-9 on page 38, we show a snapshot of the LUN's performance rates.

In Table 4-9 on page 38, we show the results of transaction (IOPS)-based workloads when used with the multiple array/LUN (trunked or "T") configuration that is shown in Figure 4-7 on page 36. We set the queue depth for this test to 1 to minimize the effect on response time results. We ran the tests with cache enabled. We wanted to show that although this layout is not the best layout, it provides performance equal to the capability of the disk type that is being used.

*Table 4-9 Random I/O read and write performance data with a multiple array/LUN configuration T*

Workload type	Host block size of I/O	IOPS	Response times
Random read	4 KB	8940	10.73 ms
Random read	8 KB	8840	10.85 ms
Random read	16 KB	8722	11 ms
Random read	32 KB	8489	11.30 ms
Random read	64 KB	8037	11.94 ms
Random read	128 KB	7040	13.12 ms
Random read	256 KB	5536	17.33 ms
Random read	512 KB	3627	26.45 ms
Random write	4 KB	5416	17.72 ms
Random write	8 KB	5470	17.54 ms
Random write	16 KB	5940	16.15 ms
Random write	32 KB	6526	14.70 ms
Random write	64 KB	4431	21.65 ms
Random write	128 KB	3815	25.15 ms
Random write	256 KB	3054	31.40 ms
Random write	512 KB	2664	35.92 ms

In Table 4-9, understand that we set the queue depth per LUN to 1 to minimize the amount of time that was spent sitting in the queue waiting. As queue depth is increased, the response time value also increases.



### Single array and multiple LUN configuration

In this configuration, we configured the subsystem with a single array, as shown in Figure 4-8 on page 39. We then divided this array into 48 LUNs, which were spread evenly between the two controllers and presented to the hosts, which were running an iometer test suite.

Tray 11				Tray 22			
Drawer 1				Drawer 1			
Drive Slot	Drive Slot	Drive Slot	Drive Slot	Drive Slot	Drive Slot	Drive Slot	Drive Slot
AB1	AB1	AB1		AB1	AB1	AB1	
AB1	AB1	AB1		AB1	AB1	AB1	
AB1	AB1	AB1		AB1	AB1	AB1	
Drawer 2				Drawer 2			
AB1	AB1	AB1		AB1	AB1	AB1	
AB1	AB1	AB1		AB1	AB1	AB1	
AB1	AB1	AB1		AB1	AB1	AB1	
Drawer 3				Drawer 3			
AB1	AB1	AB1		AB1	AB1	AB1	
AB1	AB1	AB1		AB1	AB1	AB1	
AB1	AB1	AB1		AB1	AB1	AB1	
Drawer 4				Drawer 4			
AB1	AB1	AB1		AB1	AB1	AB1	
AB1	AB1	AB1		AB1	AB1	AB1	
AB1	AB1	AB1		AB1	AB1	AB1	
Drawer 5				Drawer 5			
AB1	AB1	AB1		AB1	AB1	AB1	
AB1	AB1	AB1		AB1	AB1	AB1	
AB1	AB1	AB1		AB1	AB1	AB1	
Tray 33				Tray 44			
Drawer 1				Drawer 1			
Drive Slot	Drive Slot	Drive Slot	Drive Slot	Drive Slot	Drive Slot	Drive Slot	Drive Slot
AB1	AB1	AB1		AB1	AB1	AB1	
AB1	AB1	AB1		AB1	AB1	AB1	
AB1	AB1	AB1		AB1	AB1	AB1	
Drawer 2				Drawer 2			
AB1	AB1	AB1		AB1	AB1	AB1	
AB1	AB1	AB1		AB1	AB1	AB1	
AB1	AB1	AB1		AB1	AB1	AB1	
Drawer 3				Drawer 3			
AB1	AB1	AB1		AB1	AB1	AB1	
AB1	AB1	AB1		AB1	AB1	AB1	
AB1	AB1	AB1		AB1	AB1	AB1	
Drawer 4				Drawer 4			
AB1	AB1	AB1		AB1	AB1	AB1	
AB1	AB1	AB1		AB1	AB1	AB1	
AB1	AB1	AB1		AB1	AB1	AB1	
Drawer 5				Drawer 5			
AB1	AB1	AB1		AB1	AB1	AB1	
AB1	AB1	AB1		AB1	AB1	AB1	
AB1	AB1	AB1		AB1	AB1	AB1	

Figure 4-8 Single, large RAID10 array that was built across all expansion trays and drawers

We do *not* recommend this configuration, because it does not honor any of the throughput best practice rules. However, because the DS5300 allows you to build a large array that uses a large number of disks, we tested this configuration to see how the I/O workload compared to the previously tested configuration. Table 4-10 and Table 4-11 on page 40 show the performance that was seen with this configuration. We built this configuration as a RAID10 array with 180 disks; therefore, the test results are not a fair comparison to the test results that were seen in the previous test scenarios. Table 4-10 shows the throughput results for queue depths of 1 and 121 in our test runs.

Table 4-10 Single array throughput with cache but without cache mirroring enabled

Workload operation type	Host block size of I/O	Throughput MB/sec
Sequential read	256 KB	1.345 GB/3.076 GB
Sequential read	512 KB	2.140 GB/6.737 GB
Sequential read	1 MB	4.569 GB/6.762 GB
Sequential read	2 MB	6.454 GB/5.476 GB
Sequential write	256 KB	4.058 GB/1.892 GB

Workload operation type	Host block size of I/O	Throughput MB/sec
Sequential write	512 KB	4.211 GB/1.053 GB
Sequential write	1 MB	4.320 GB/1.422 GB
Sequential write	2 MB	4.3.96 GB/1.353 GB

*Table 4-11 Random I/O read and write performance data with one large array with multiple LUNs T*

Workload type	Host block size of I/O	IOPS	Response times
Random read	4 KB	6817	21.11 ms
Random read	8 KB	7110	20.23 ms
Random read	16 KB	7008	20.54 ms
Random read	32 KB	6908	20.83 ms
Random read	64 KB	6659	21.62 ms
Random read	128 KB	6185	23.27 ms
Random read	256 KB	5383	26.73 ms
Random read	512 KB	4280	33.62 ms
Random write	4 KB	6541	18.58 ms
Random write	8 KB	8632	16.67 ms
Random write	16 KB	7565	19.03 ms
Random write	32 KB	7803	18.44 ms
Random write	64 KB	5785	24.88 ms
Random write	128 KB	5108	28.17 ms
Random write	256 KB	4442	32.39 ms
Random write	512 KB	3361	42.80 ms

You can achieve greater performance in a RAID10 mixed workload configuration where nearline (SATA) storage is an acceptable solution if you build a large array that follows the recommendations that we have outlined. The largest possible array size is one-half of the total disk population, or 120 drives in a 60x60 RAID10 configuration. In this manner, you can build two of these arrays with one array that is owned by controller A and one array that is by controller B. The same controller, to which the array is defined, must own all of the LUNs that are created on these arrays.

Figure 4-9 shows a recommended layout for this type of configuration with two 56x56 RAID10 arrays leaving hot spares for protection.

Tray 11				Tray 22							
Drawer 1				Drawer 1							
Drive Slot	Drive Slot	Drive Slot	Drive Slot	Drive Slot	Drive Slot	Drive Slot	Drive Slot				
A1	A1	A1	A1	A1	A1	A1	A1				
A1	A1	A1	A1	A1	A1	A1	A1				
A1	A1	A1	A1	A1	A1	A1	A1				
Drawer 2				Drawer 2							
A1	A1	A1	A1	A1	A1	A1	A1				
A1	A1	A1	A1	A1	A1	A1	A1				
A1	A1	A1	A1	A1	A1	A1	A1				
Drawer 3				Drawer 3							
B2	B2	B2	B2	B2	B2	B2	B2	Controller A Array/LUNs			
B2	B2	B2	B2	B2	B2	B2	B2				
B2	B2	B2	B2	B2	B2	B2	B2				
Drawer 4				Drawer 4							
B2	B2	B2	B2	B2	B2	B2	B2				
B2	B2	B2	B2	B2	B2	B2	B2				
B2	B2	B2	B2	B2	B2	B2	B2				
Drawer 5				Drawer 5							
A1	A1	B2	B2	A1	A1	B2	B2				
A1	A1	B2	B2	A1	A1	B2	B2				
HS	HS	HS	HS	HS	HS	HS	HS				
Tray 33				Tray 44							
Drawer 1				Drawer 1							
Drive Slot	Drive Slot	Drive Slot	Drive Slot	Drive Slot	Drive Slot	Drive Slot	Drive Slot				
A1	A1	A1	A1	A1	A1	A1	A1				
A1	A1	A1	A1	A1	A1	A1	A1				
A1	A1	A1	A1	A1	A1	A1	A1				
Drawer 2				Drawer 2							
A1	A1	A1	A1	A1	A1	A1	A1				
A1	A1	A1	A1	A1	A1	A1	A1				
A1	A1	A1	A1	A1	A1	A1	A1				
Drawer 3				Drawer 3							
B2	B2	B2	B2	B2	B2	B2	B2	Controller B Array/LUNs			
B2	B2	B2	B2	B2	B2	B2	B2				
B2	B2	B2	B2	B2	B2	B2	B2				
Drawer 4				Drawer 4							
B2	B2	B2	B2	B2	B2	B2	B2				
B2	B2	B2	B2	B2	B2	B2	B2				
B2	B2	B2	B2	B2	B2	B2	B2				
Drawer 5				Drawer 5							
A1	A1	B2	B2	A1	A1	B2	B2				
A1	A1	B2	B2	A1	A1	B2	B2				
HS	HS	HS	HS	HS	HS	HS	HS				

Figure 4-9 Recommended large RAID10 array layout



# Related publications

The publications listed in this section are considered particularly suitable for a more detailed discussion of the topics covered in this paper.

## IBM Redbooks publications

For information about ordering these publications, see “How to get IBM Redbooks publications” on page 44. Note that some of the documents referenced here may be available in softcopy only.

- ▶ *IBM Midrange System Storage Hardware Guide*, SG24-7676
- ▶ *IBM Midrange System Storage Implementation and Best Practices Guide*, SG24-6363
- ▶ *IBM System Storage DS4000 and Storage Manager V10.30*, SG24-7010
- ▶ *IBM System Storage DS3000: Introduction and Implementation Guide*, SG24-7065

## Other publications

These publications are also relevant as further information sources:

- ▶ *IBM System Storage EXP5060 Quick Start Guide*
- ▶ *IBM System Storage EXP5060 Storage Expansion Enclosure Installation, User's and Maintenance Guide*
- ▶ *IBM System Storage DS Storage Manager Installation and Host Support Guide*
- ▶ *IBM System Storage DS Storage Manager Command-line Programming Guide*
- ▶ *IBM System Storage Quick Start Guide for the DS5100/DS5300 Storage Subsystems*
- ▶ *IBM System Storage DS5100/DS5300 Installation, User's, and Maintenance Guide*
- ▶ *IBM System Storage DS Storage Manager Copy Services Guide*

All of these listed publications can be found on the IBM Support website.

## Online resources

These Web sites are also relevant as further information sources:

- ▶ IBM System Storage Disk Storage Systems Support  
<http://www.ibm.com/systems/support/storage/disk>
- ▶ IBM System Storage Interoperation Center (SSIC)  
<http://www.ibm.com/systems/support/storage/config/ssic/index.jsp>
- ▶ IBM DS5100, DS5300, and BladeCenter Premium Feature Activation  
<http://www-912.ibm.com/PremiumFeatures>
- ▶ Storage Area Network (SAN) Support

<http://www.ibm.com/systems/support/storage/san>

## How to get IBM Redbooks publications

You can search for, view, or download IBM Redbooks publications, Redpapers, Technotes or webdocs, draft publications and Additional materials, as well as order hardcopy IBM Redbooks publications, at this Web site:

[ibm.com/redbooks](http://ibm.com/redbooks)

## Help from IBM

IBM Support and downloads

[ibm.com/support](http://ibm.com/support)

IBM Global Services

[ibm.com/services](http://ibm.com/services)





# IBM System Storage EXP5060 Storage Expansion Enclosure Planning Guide



**Learn about the new  
EXP5060 high-density  
disk drive enclosure**

**Understand available  
EXP5060 cabling  
options with DS5000**

**Review  
configurations and  
performance test  
results**

The IBM System Storage EXP5060 Storage Expansion Enclosure (Machine Type 1818, Model G1A) provides high-capacity Serial Advanced Technology Attachment (SATA) disk storage for the DS5100 and DS5300 storage subsystems. The storage expansion enclosure delivers fast, high-volume data transfer, retrieval, and storage functions for multiple drives to multiple hosts. The storage expansion enclosure provides continuous, reliable service, using hot-swap technology for easy replacement without shutting down the system and supports redundant, dual-loop configurations.

This IBM Redpaper publication illustrates and describes the various external cabling options and Small Form-Factor Pluggable (SFP) modules that connect the DS5100 or DS5300 storage subsystem to the EXP5060 storage expansion enclosure.

This paper gives you a brief introduction to the EXP5060 and serves as a supplemental planning guide to attach the EXP5060 to DS5100 or DS5300. This paper describes the best workloads, layouts, and practices for the EXP5060. This paper shows the results of a variety of EXP5060 performance testing.

This paper is intended for anyone who wants to understand the advantages and performance benefits of the EXP5060.

## **INTERNATIONAL TECHNICAL SUPPORT ORGANIZATION**

### **BUILDING TECHNICAL INFORMATION BASED ON PRACTICAL EXPERIENCE**

IBM Redbooks are developed by the IBM International Technical Support Organization. Experts from IBM, Customers and Partners from around the world create timely technical information based on realistic scenarios. Specific recommendations are provided to help you implement IT solutions more effectively in your environment.

**For more information:**  
**[ibm.com/redbooks](http://ibm.com/redbooks)**