Part Number: MP4001C

9500 Shared Virtual Array

Introduction

StorageTek
June, 2000

Information contained in this publication is subject to change. In the event of changes, the publication will be revised. Comments concerning its contents should be directed to the address shown below.

StorageTek
One StorageTek Drive
Louisville, Colorado
80028-2201

Contents

Preface

Notices

United States FCC Compliance Statement

The following is the compliance statement from the Federal Communications Commission:

Note: This equipment has been tested and found to comply to the limits for Class A digital devices pursuant to Part 15 of the FCC Rules. These limits are designed to provide reasonable protection against harmful interference when the equipment is operated in a commercial environment. This equipment generates, uses, and can radiate radio frequency energy and, if not installed in accordance with the instruction manual, may cause harmful interference to radio communications. Operation of this equipment in a residential area is likely to cause harmful interference, in which case the user will be required to correct the interference at his or her own expense.

Some of the cables used to connect peripherals must be shielded and grounded as described in the installation manual. Operation of this equipment with the required cables that are not shielded and correctly grounded may result in interference to radio and TV reception.

Changes or modifications not expressly approved by StorageTek could void the user's authority to operate the equipment.

Industry Canada Compliance Statement

This digital apparatus does not exceed the Class B limits for radio noise emissions from digital apparatus set out in the radio interference regulations of the Canadian Department of Communications.

Le présent appareil numerique német pas de bruits radioélectriques dépassant les limites applicables aux appareils numériques de la classe B prescrites dans le Reglément sur le brouillage radioélectrique édicté par le ministére des Communications du Canada.

Japanese Compliance Statement

The following is the compliance statement from Japan:

　この装置は、クラスＡ情報技術装置です。この装置を家庭環境で使用すると電波妨害を引き起こすことがあります。この場合には使用者が適切な対策を講ずるよう要求されることがあります。　VCCI-A

Note: This equipment is in the Class A category information technology equipment based on the rules of Voluntary Control Council For Interference by Information Technology Equipment (VCCI). When used in a residential area, radio interference may be caused. In this case, user may be required to take appropriate corrective actions.

Consequently, when used in residential area or in an adjacent area thereto, radio interference may be caused to radios and TV receivers, etc. Read the instructions for correct handling.

Taiwan Warning Label Statement

The following is the warning label statement from Taiwan, R.O.C.:

警告使用者：這是甲類的資訊產品，在居住的
環境中使用時，可能會造成射頻干擾，在這種
情況下，使用者會被要求採取某些適當的對策

About This Book

This book provides detailed physical specifications, cable information, environmental specifications, and power requirements for the 9500 Shared Virtual Array. Two checklists are also included within this book: the wiring comments checklist and the physical planning checklist.

Who Should Read This Book

The information in this book is for the use of the customer, StorageTek marketing representatives, CSEs, and independent consultants involved with installation planning.

Trademarks

Related Documents

The documents listed below comprise the complete SVA 9500 set. See "Viewing and Ordering Documents" (below) to obtain documents through available distribution channels.

Shared Virtual Array (SVA) Subsystem

Documents below are available online, on CD-ROM, and as bound books.

- *9500 Shared Virtual Array Introduction (MP4001B).*
- *9500 Shared Virtual Array Operation and Recovery (MP4002B)*
- *9500 Shared Virtual Array Planning, Implementation and Usage (MP4003B)*
- *9500 Shared Virtual Array Physical Planning (MP4004B)*

- *9500 Shared Virtual Array Reference (MP4005B)*
- *9500 Shared Virtual Array System Assurance (MP4006B)*
- *Peer to Peer Remote Copy Configuration Guide (MP4007A)*

Shared Virtual Array Administrator (SVAA) for OS/390

Documents below are available online and on CD-ROM.

- *SVAA for OS/390 Configuration and Administration (PN 3112905xx)*
- *SVAA for OS/390 Reporting (PN 3112906xx)*
- *SVAA for OS/390 Messages and Codes (PN 3112907xx)*
- *SVAA for OS/390 Installation, Customization, and Maintenance (PN 3112908xx)*

SnapShot for OS/390

Documents below are available online and on CD-ROM.

- *SnapShot for OS/390 User's Guide (PN 3112912xx)*
- *SnapShot for OS/390 Installation, Customization, and Maintenance (PN 3112913xx)*

Shared Virtual Array Administrator (SVAA) for Solaris

Documents below are available online and on CD-ROM.

- *SVAA for Solaris User's Guide (PN 3112909xx)*
- *SVAA for Solaris Messages (PN 3112910xx)*
- *SVAA for Solaris Installation (PN 3112911xx)*
- *SVAA for Solaris Quick Start Guide (PN 3112971xx)*

SnapShot for Solaris

Documents below are available online and on CD-ROM.

- *SnapShot for Solaris User's Guide (PN 3112914xx)*
- *SnapShot for Solaris Quick Start Guide (PN 3112915xx)*

Shared Virtual Array Console (SVAC) for Windows NT

The document below is available online and on CD-ROM.

- *SVAC for Windows NT Quick Start Guide (PN 3112993xx)*

Viewing and Ordering Documents

Viewing the Documents Online

SVA 9500 documents can be viewed (and printed on your printer - these are PDF files) at the Customer Resource Center website at:

`http://www.support.storagetek.com/wwcss/SilverStream/Pages/pgCRCHome.html`

Click on 'Disk Subsystems', then 'Docs' under the 9500 section.

Note: A password is required. You may obtain the password from a StorageTek marketing representative.

Ordering Documents

SVA 9500 documents are available on CD-ROM, and bound book. Consult a StorageTek marketing representative to order the various manuals relating to the 9500.

CD-ROM

- Customer hardware documents: a CD-ROM of SVA 9500 customer documents is available by requesting the *9500 Customer Documentation, PN MP-9500x*.
- Software documents: a CD-ROM of SVA 9500 software documents is available by requesting *SVA Software Publications, PN 311295301-xx*.

Bound Books

Individual bound books of SVA 9500 documents are available through Software Manufacturing Distribution (SMD); request by document title and/ or part number.

Other Documents

The following IBM documents may also assist you in using SVA 9500:

- *Planning For IBM Remote Copy SG24-2595-xx*
- *Remote Copy Administrator's Guide and Reference SC35-0169-xx*

Do You Have the Complete Book?

This book consists of page 1 through page 74.

Note: The last two pages are the reader comment form and its mailer. These pages may not be there if someone sent the comment form to StorageTek.

History of Changes

Rev A - initial release. November 1999

Rev B - First revision. December 1999.

- Changed pagination to reflect the page number used in the PDF file.
- Document ordering information updated.
- Effective cache figures updated.

Rev C - Second revision. June 2000.

- Minor changes and updates.

Chapter 1 Introduction

StorageTek's advanced Shared Virtual Array (SVA) delivers a wide range of options to meet today's diverse storage requirements. "One size fits all" storage system solutions of the past are no longer adequate. The breadth of the SVA gives customers the freedom to choose the optimal balance of function, performance, capacity, and affordability to meet their specific needs today with flexibility to grow and change in the future.

Figure 1-1 The 9500 Shared Virtual Array System

The StorageTek Shared Virtual Array offers outstanding operational flexibility with exceptional value. The SVA redefines the availability, operational considerations, performance management, and total cost of ownership of online storage.

The Shared Virtual Array represents a major advancement in direct access storage for the IBM-compatible mainframe marketplace. The Shared Virtual Array combines proven technologies and a state-of-the-art architecture to address many of the limitations inherent in traditional online storage. The Shared Virtual Array system includes:

- the Shared Virtual Array Subsystem (Figure 1-1 on page 11)
- and the Shared Virtual Array Administrator (SVAA), Version 2 Release 1 software product.

This manual identifies the significant benefits of the StorageTek Shared Virtual Array Subsystem, provides an overview of its features and capabilities, and describes how the virtual storage architecture makes these important benefits possible.

The 9500 Shared Virtual Array is a single-frame control unit/storage unit with turbo shared memory, sixteen internal data paths, and ESCON/Fibre channels.

In addition, the SVA includes the Shared Virtual Array Administrator (SVAA), a host software product that expands the functionality of Shared Virtual Array, while providing a convenient host interface for controls, reporting, and management function.

Introducing the Shared Virtual Array

The StorageTek Shared Virtual Array is based on a new virtual storage architecture. This virtual storage architecture provides extended capabilities in the areas of system availability, storage management, capacity, performance, connectivity, nondisruptive service, and total cost. These extended capabilities have been achieved without the compromises inherent in other CKD and Redundant Array of Independent Disks (RAID) architectures.

Virtual storage architecture offers the advantages of industry standard, online storage systems including independent channel and Direct Access Storage Device (DASD - a disk drive) transfers, high-speed processing, DASD fast write, expandable cache memory, rapid access to data in cache, efficient cache management, data protection, nonvolatile storage, ESCON and SCSI attachment.

In addition, two Shared Virtual Array software products--SVAA and SnapShot--provide more functionality of an SVA subsystem. The Shared Virtual Array Administrator (SVAA) is a host software product that unlocks the power of the virtual architecture by placing unprecedented storage management capabilities in the hands of the user. SnapShot is an advanced duplication function made possible by the virtual architecture of the Shared Virtual Array.

SnapShot, a separately-priced software product, creates virtual copies of volumes or data sets within seconds. Together, these innovative and synergistic products provide a total storage solution.

Highlights of Virtual Storage Architecture

Continuous System Operation

The virtual architecture provides a truly fault-tolerant design to assure continuous access to data even in the event of a component or device failure. All data stored in an SVA subsystem is automatically protected by the fault-tolerant design.

The Shared Virtual Array features the following fault-tolerant hardware enhancements:

- Dual redundant arrays
- Concurrent License Internal Code (LIC) Activation, which allows nondisruptive loading of microcode.
- In most cases, even multiple failures do not limit access to data.
- No single failure significantly degrades performance (no more than 25%)[1].
- Shared Virtual Array servicing is 99% nondisruptive.
- System hardware upgrades may be installed without affecting operations.
- Media maintenance is automatic and nondisruptive.
- Peer-to-peer Remote Copy (PPRC), which allows two 9500 systems to contain and exchange the same information, providing superior protection against data loss.

Extended Performance Capabilities

The Shared Virtual Array's advanced storage architecture permits a large number of parallel data transfers within the system by utilizing high internal bandwidths and multiple, powerful control processors. In addition, high throughput for online transaction processing and batch workloads can be sustained.

Extended Compatibility

The Shared Virtual Array features 3990 CU compatibility and is supported in MVS and VM environments. Both 3380 and 3390 track geometries are supported. In addition, the SVA offers unique compatibility features that offer unparalleled configuration flexibility.

1. For fault tollerant configurations with a minimum of four interface cards of a given type.

Host Software Compatibility

The most current levels of host software maintenance are highly recommended. Shared Virtual Array host software — SVAA — is available to facilitate and optimize the capabilities of the SVA and exploit the features available through the virtual architecture design.

Current DASD Coexistence

The Shared Virtual Array, which is compatible with the 3990 CU images and easily coexists with traditional online storage systems.

Complements Systems Managed Storage

The Shared Virtual Array is readily adapted into Systems Managed Storage (SMS), including hierarchical environments, and simplifies SMS implementation for the storage administrator.

Virtual DASD

Virtual DASD, which is enabled by the Shared Virtual Array control unit function, maximizes the capabilities of Shared Virtual Array's unique architecture. With virtual DASD, the SVA provides up to 1024 functional 3380[2] or 3390 volumes, independent of the system physical capacity, number of arrays, or physical devices. (The term "functional" describes the host view of the SVA's capacity.)

Deleted Data Space Release (DDSR)--a facility provided by the Shared Virtual Array Administrator (SVAA)--further capitalizes on virtual DASD by allowing the SVA to reclaim and reuse storage capacity.

Investment Durability

Rapid advancements in DASD technology can make DASD systems obsolete within three to five years of their release. However, because the Shared Virtual Array's architecture features virtual DASD and device independence, the functional device image that the host views is isolated from the physical devices that populate the system. Therefore, if new host technology requires changes to the functional device image, Shared Virtual Array can accommodate those changes without requiring changes to the physical devices.

Likewise, the SVA can accommodate new device technology without requiring changes to the host image.

Extended Management Facilities

The Shared Virtual Array system offers the storage administrator more flexibility and control when managing DASD storage capacity and performance. With virtual DASD, physical space management is elevated from the individual volume level to the system level. Data sets may be spread across more volumes. Storage management is simplified: fewer

---

2. Not supported on SCSI attachments.

"out of space" abends occur, fewer volume defragmentations and reorganizations are required, and productivity increases. Also, because virtual DASD can tolerate highly skewed workloads, tuning requirements are minimized.

These are just some of Shared Virtual Array's storage management capabilities; the Shared Virtual Array Administrator (SVAA) provides additional capabilities that unlock the powerful storage management features inherent in the Shared Virtual Array data storage solution using virtual storage architecture.

Extended Service Capabilities

The Shared Virtual Array employs expert-system techniques to provide predictive service analysis, which identifies and fences off a field replaceable unit (FRU) that requires service. The system may initiate maintenance via ServiceTek, a machine-initiated maintenance facility that automatically initiates system audits, orders replacement parts, and schedules repairs.

Connectivity

The Shared Virtual Array can connect to ESCON channels (128 host paths and 16 ESCON physical links) in native mode. SVA native ESCON support allows SVA connectivity to ESCON channel ports on CPUs and/or ESCON channel ports on ESCON directors. Additionally, support for 20 MB/sec instantaneous channel transfer, as well as EMIF, is provided.

Connections to open system hosts can be made via the Shared Virtual Array Extended SCSI Attach Feature (XSA) or native fibre channel attachment.

Summary

The StorageTek Shared Virtual Array subsystem offers many functional improvements over traditional DASD systems and provides uninterrupted online services in a small footprint. With its innovative virtual storage architecture, comprehensive fault-tolerant design, configuration flexibility, storage and performance management facilities, virtual DASD, extended connectivity and service capabilities, the SVA truly heralds a new era in online storage.

This page intentionally left blank

Chapter 2 Shared Virtual Array System

This chapter provides an overview of the StorageTek 9500 Shared Virtual Array (SVA). This chapter describes the architecture and primary functions of the 9500.

System Overview

The 9500 is compatible with the IBM 3990 Models 3 and 9 and supports MVS and VM environments. In addition, the SVA supports intermixed 3380 and 3390 volume types, and attaches to hosts with IBM 390-equivalent or ESCON links.

The 9500 Shared Virtual Array

The StorageTek 9500 combines the disk storage and control unit into a single frame. The 9500 offers outstanding operational flexibility and exceptional total cost of ownership. Built on the same highly flexible virtual architecture as earlier models, 9500 offers the same high level of functionality and availability, but with improved performance and environmentals.

The earlier models of the Shared Virtual Array (9200) required two, or for the higher effective storage capacities, three frames. With the 9393, the entire storage system, including control unit functions, cache storage, and all the disk drives are provided in one single frame. The 9500 further reduced the size, and results in no only a considerable environmental savings, especially in floor space, but also on power consumption and heat generation (for more detailed information, refer to the *Shared Virtual Array Physical Planning Guide*). The bottom line: considerably lower total cost of ownership.

Customers using IBM System/390, ES/9000, and ES/3090 processors and S/390 Parallel Enterprise and Multiprise Servers will find the new model of the SVA an excellent addition to their storage options from StorageTek. The combination of low cost, operational, flexibility, high capacity, availability, and performance provides significant benefits.

Pre-Requisites

The StorageTek 9500 SVA applies to System/390, ES/9000, and ES/3090 customers running MVS/ESA, VM/ESA, and/or VSE/ESA.

StorageTek recommends that VSE/ESA be run as a guest under VM/ESA to get the benefits that the Shared Virtual Array Extended Facilities Product (IXFP) or Shared Virtual Array Administrator (SVAA) provides for the SVA.

Architecture Highlights

The 9500 employs an advanced virtual storage system architecture where logical volumes are dynamically mapped to physical volumes, emulating both IBM 3380 and 3390 direct access storage devices. Additional highlights of this flexible architecture include:

Comprehensive Fault Tolerant Design

The 9500, with dual parity RAID 6 architecture, is designed to help provide access to data even when two simultaneous disk failures occur in the same array. The SVA has a fault-tolerant design that enables you to maintain access to data and system performance in the event of most hardware failures. Hot-pluggable drives, redundant power supplies, and duplicate system components reduce the possibility of system outages.

Automatic System Tuning

The virtual disk architecture employed by the 9500 automatically spreads data across many different disk drives. Disk “hot spots” are practically eliminated, which helps improve overall system performance.

Support for 3380 and 3390 Device Types

The 9500 supports both 3380[1] and 3390 device types in single, double, and triple capacity volume sizes. These can be intermixed within a single system. This capability greatly simplifies migration from older technology systems and adds flexibility for disaster recovery.

Virtual Storage Architecture

The 9500 employs an advanced storage system architecture where logical volumes are indirectly mapped to physical volumes within the system. By operating in this manner, 9500 is able to deliver unique outboard storage management using Shared Virtual Array Administrator (SVAA) software. Plus, it can more efficiently use storage capacity by storing only data that is actually written--not reserving capacity for un-allocated space as well as allocated space, which is unused.

Built-In Compression and Compaction

The virtual architecture also allows the 9500 to employ its own compression and compaction algorithms. Using a sophisticated hardware compression algorithm, the 9500 compresses data at the channel interface in an operation transparent to the host. The 9500 complements host-based data that are not supported by host-based compression. In addition, compaction improves storage efficiency even for compressed data sets. The 9500 eliminates the inter-record gaps associated with storing disk in CKD or ECKD formats, enabling compaction of data onto the disk storage.

1. Not supported for SCSI attachments.

The 9500 employs its own compression and compaction algorithms, achieving typical compression ratios of 2:1 for cache and 4:1 for disk storage, which are used to calculate effective cache and storage capacities. Greater ratios have typically been observed for both cache and disk[2].

High Granularity of Capacity and Cache

System cache is available in sizes up to 5888 MB of effective cache storage. Sixty four MB of effective nonvolatile cache (NVS) is standard. When the granularity of effective storage capacity is combined with the flexibility of adding effective cache capacity in increments up to 5888 MB, the Shared Virtual Array can be tailored to meet a wide variety of configuration requirements.

Because data is always compressed and compacted when stored in a Shared Virtual Array, the effective or usable capacity represents the total apparent system capacity as reported to the host software. Actually-realized storage will vary, depending on the actual data being compressed. Some data sets will achieve higher ratios and some will achieve lower ratios. Experience has shown that the capacities described, which assume a compression ratio of 2:1 for cache and 4:1 for disk storage, are representative for information stored in a typical SVA. Capacities expressed in this way may be directly compared to the capacities of traditional storage systems. The SVA also compresses data before it is stored in cache, so cache and NVS sizes are quoted in terms of effective capacity, as described above. Again, compare these sizes with uncompressed cache sizes in traditional storage systems.

Concurrent License Internal Code (LIC) Activation

The Shared Virtual Array subsystem supports nondisruptive loading of system microcode. Without Concurrent LIC Activation, code changes require that maintenance downtime be scheduled, and that host-to-system links and system power be disabled, causing significant disruption of system operations. With Concurrent LIC Activation, code changes do not require host-to-system links and system power to be disabled; instead, code changes are enabled by warm-booting the system, significantly reducing the impact of code loads on system resources.

Note: The first time that Concurrent LIC Activation is loaded, it is disruptive; thereafter, it is non-disruptive.

Volatile Cache

The 9500 system supports 1920 MB to 2944 MB of physical customer cache, providing 3840 MB to 5888 MB of effective customer cache at a 2:1 compression. Larger volatile cache improves subsystem performance capabilities in heavy database workload environments[3].

2. Compression numbers are valid for OS390 only.
3. Effective cache may be less than this value and is based on the number of devices configued.

Sixteen Front-End (Internal) Data Paths

The 9500 systems support 16 data paths between the storage clusters and cache, doubling the internal data throughput capability of earlier systems. Each of the 16 data paths between the storage clusters and effective cache transfers data at 19.6 MB/sec[4], providing an aggregate internal data transfer rate of 313.6 MB/sec.

SnapShot V1.2 for MVS/ESA

The 9500 systems support SnapShot Version 1.2 for MVS/ESA. This version expands SnapShot capabilities by supporting more data types, including VSAM, multivolume, and extended format data sets.

Peer-to-Peer Remote Copy (PPRC)

The 9500 systems support the peer-to-peer remote copy (PPRC) feature. PPRC allows two discrete 9500 systems to contain and exchange the same information, providing superior protection against data loss from catastrophic events.

Once enabled, PPRC runs virtually unattended, allowing data from a primary DASD system at one location to be copied in real time to a secondary (or recovery) DASD system at another location. With PPRC, records at the recovery site are continuously updated to match those at the primary site, providing an additional level of data redundancy beyond that available in a single 9500 system. PPRC minimizes the potential for data loss during catastrophic events, and reduces the impact of such disasters by allowing faster recovery of critical data.

As a subset of its disaster recovery function, PPRC can also be used for workload migration, allowing data from one system to be easily moved to another system with minimal impact to system operations.

SCSI Host Attachment

The 9500 systems support attachment to SCSI host systems using the StorageTek Shared Virtual Array Extended SCSI Attach Feature (SVA Attach Feature). The SVA Attach Feature extends the distance of a pair of ultra-wide SCSI buses up to three kilometers while maintaining the SCSI system timing requirements.

Fibre Channel Attach

The 9500 system supports attachment to fibre channel host systems by replacing ICE cards with ICF cards. Unlike SCSI attachments, no additional hardware is necessary.

Serial Storage Architecture (SSA)

The 9500 systems support the IBM serial storage architecture (SSA) standard. SSA gives the 9500 system a data transfer bandwidth of about 80 MB/sec on the SSA loops.

4. With compression.

3390-9 Logical Volumes

The 9500 systems support IBM 3390 Model 9 (3390-9) logical volumes. This feature allows up to 341 logical volumes to be defined in a 9500 system. Each 3390-9 volume provides three times the storage capacity of a corresponding 3390-3 volume.

1024 Volume Virtual Addressing

The 9500 systems support 1024 volume virtual addressing. This feature allows users to define up to 256 logical 3380 or 3390 volumes (with the exception of 3390-9s as noted above) for each of four virtual control units in a 9500 system, or 1024 logical volumes total.

Sophisticated Storage Management

The Shared Virtual Array Administrator (SVAA), Version 2, Release 1 provides sophisticated host software, which enables the configuration and monitoring of a Shared Virtual Array, even while the system is operating.

Serviceability

The 9500 provides a function whereby it will automatically call StorageTek Service under some circumstances in the event of a hardware problem. This greatly shortens the time to repair and/or recover from problems. To enable this, customers must provide a phone line near the SVA. This permits the SVA to pass error logs and other pertinent information to the StorageTek support personnel. Remote service connectivity to the phone line is provided by StorageTek and StorageTek service personnel, who will establish and test the connectivity during the installation of the SVA.

Primary Functions

The 9500 system attaches to one to 32 OS390 hosts and directs the transfer and storage of host data to the arrays. Its primary functions are to manage data transfers between the host(s), cache, and arrays; interpret and execute channel commands; monitor operational status; supply status information to the host(s); and perform system diagnostics. Host data are sent to the controller for temporary storage and sorting, then assigned to arrays (back end) for permanent storage.

Data are compressed and compacted before storage, then de-compacted and decompressed before return to the host(s).

Key components of the 9500 include:

- Storage clusters (two) to transmit and receive data via ESCON links to the host(s). The storage clusters include the following:
  - Multi-path storage directors (MPSDs) to receive data from host channels, transfer it via internal paths to cache for temporary storage and sorting, then send it to the arrays for permanent storage.
  - Shared memory to store data about system physical and functional device status.

  - Support facility to monitor operations, process status information, support the local operator panel, and initiate maintenance procedures by separating (fencing) defective logical paths or physical components from operating areas and parts to prevent interruption of operations.
- Volatile cache to retain recently-used data for fast retrieval by a host or Controller; lost if AC power is disrupted.
- 1920 MB to 2944 MB of physical customer cache.
- Nonvolatile cache to back up volatile cache during write operations and to store control information. Available physical cache is 32 MB with and effective capacity of 64 MB.
- Redundant battery backup power to protect cache nonvolatile memory data from loss for at least 72 hours if AC power is disrupted.

Primary functions of the system array region are to transmit and store host data; format drives; control drive power; monitor and transmit environmental and drive statistics; and support maintenance, diagnostic, RPCI, and media error detection, recovery, and correction functions. An array is a logical entity of:

- 7 or 8 physical drives partitioned in groups of 5+2 or 5+2+1
- 15 or 16 physical drives partitioned in groups of 13+2 or 13+2+1

Note:

5+2+1 array = 5 data drives + 2 redundancy/parity drives + 1 spare;
13+2+1 array = 13 data drives + 2 redundancy/parity drives + 1 spare.

System Volatile Cache Capacity Options

Table 2-1 below shows the amount of physical cache available with the various CVX (cache memory cards) installed.

Table 2-1 9500 Cache Configuration

# of CVX2 cards	Physical Cache size - ordered/installed
4	2048
5	2560
6	3072

Fault Tolerance

The Shared Virtual Array combines the beneficial characteristics of RAID 6 and virtual DASD, plus fault-tolerant design, to provide continuous data access even if a hardware failure occurs. The SVA implements RAID 6 storing data across linked groups of small disks (arrays) instead of across one large disk, greatly minimizing the potential for data loss or corruption from drive failures. If a drive fails, its data is automatically reconstructed onto a “hot spare” using redundant data from surviving drives, while the system continues operating at near-peak efficiency.

Physical Design

The physical design of the Shared Virtual Array provides further safeguards. Each potential point of hardware failure has a backup, including cooling systems, power supplies, emergency batteries, cables, drive modules, electronics, data paths, etc. This redundancy allows components to be removed and replaced without shutting down the system or interrupting user operations. Scheduled outages are also reduced, since most upgrades and maintenance can be performed non-disruptively without noticeable degradation in response time or throughput.

Data Paths

Figure 2-1 on page 24 shows the various data paths for 9500 systems. Each system has different types of data paths, which may include:

- System-to-Host/ESCON Director--Four, eight, or 16 physical ESCON links connected to an ESCON director to allow up to 128 host paths (32 host paths with 4 paths each).
- Front-End (Internal) -- Sixteen data paths between the storage clusters and cache.
- Back-End (Internal) --Sixteen copper paths between the front end (controller) and back end (arrays).

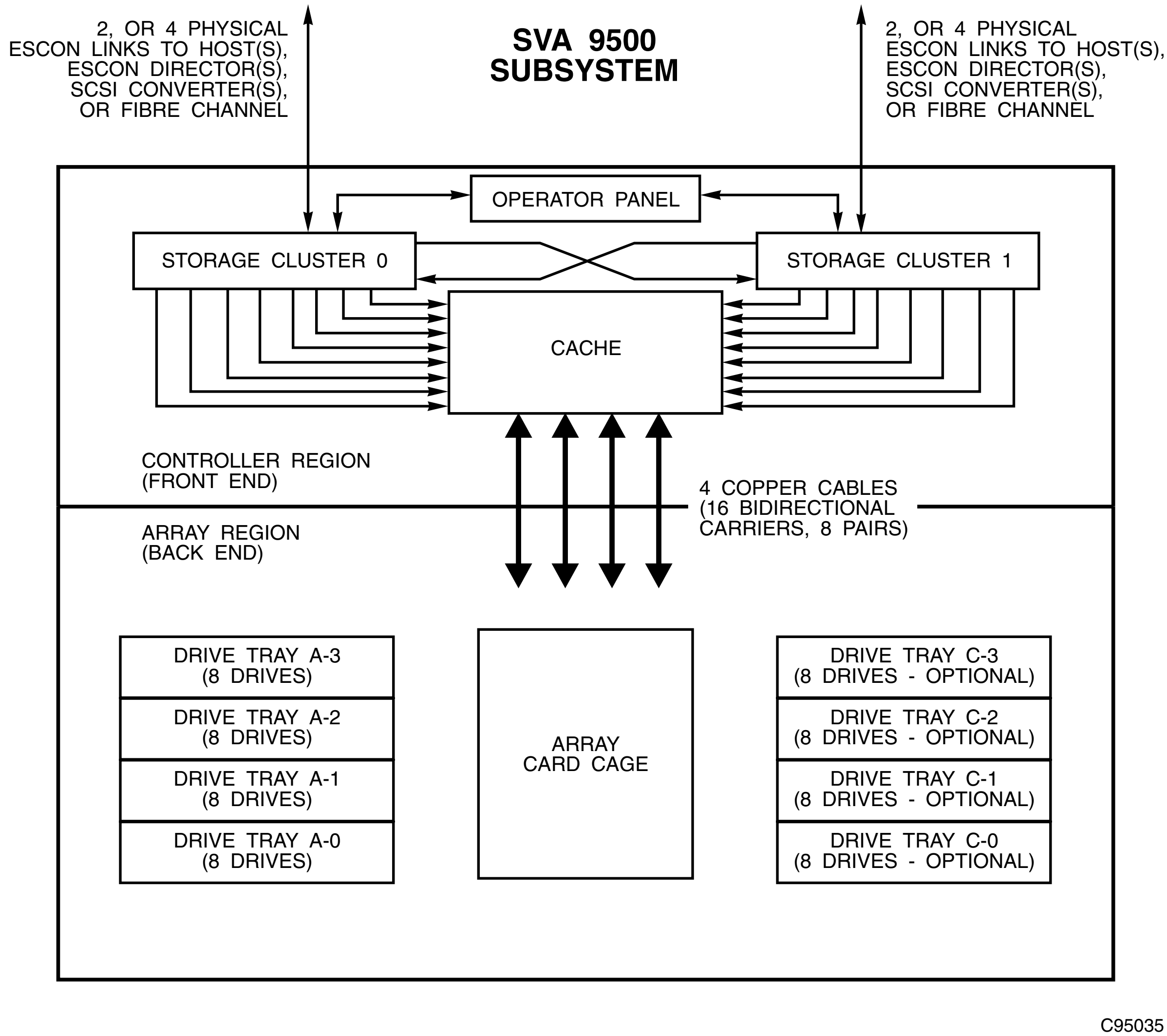

Figure 2-1 9500 System Data Paths

128 Physical Host Connections to Shared Virtual Array

Depending on the number of cards, there are eight or 16 physical connections to the Control Unit. Each connection has 64, or 32 logical paths, respectively.

For Connectivity

With an ESCON director complex (dynamic switch), the 9500 supports 128 physical host path connections. With this feature, one to 32 hosts attach to an ESCON director complex. Each host has four physical ESCON fiber connections to the director complex. Each connection can address any of four virtual control unit (VCU) images defined in the SVA. Each host therefore has a minimum of 4x4 (16) logical paths into the ESCON director complex. 32 hosts x 16 logical paths per host = 512 logical paths contained on 128 physical connections. Each path has a sustained transfer rate of 19.6 MB/sec.

For Performance

The 9500 supports eight host paths each with sixteen ESCON fibre connections (one for each 9500 intenal data path).

SnapShot

SnapShot is a high-speed data duplication solution that operates on the data set and volume levels. SnapShot is the next exploitation of Shared Virtual Array's virtual disk architecture. SnapShot works in conjunction with a new release of Shared Virtual Array Administrator (SVAA) and the Shared Virtual Array.

SnapShot virtually duplicates data content within the SVA by creating multiple pointers to one functional track or group of tracks. SnapShot allows you to "redesign" your data movement processes to reduce recovery windows, reduce the risks associated with current methods of testing application changes using a subset of production data, and increase the time that data is available for product access.

Customers duplicate data for:

- Data recovery--due to an application or hardware failure or a disaster (business resumption)
- Tuning--relocate for performance and volume space management
- Change control--verification of application or hardware changes
- Business and legal reasons--regulations (such as financial information for seven years), market research (data mining), and environment (for example, year-2000 data conversion, testing, and implementation).

This capability will be exploited by the Shared Virtual Array users to dramatically reduce duplication time and resources, as well as eliminate delays in production data access associated with traditional duplication techniques.

Virtual Data Duplication with SnapShot

SnapShot allows customers to virtually duplicate data at cache speeds (seconds or minutes vs. minutes or hours) for production data alone. Now customers can "rethink" how and when they make this duplication. With SnapShot, you can:

- Shorten the time data is unavailable for production access by redesigning your duplication processes

  Replace your traditional data duplication methods with SnapShot. You can also add more snaps to shorten data recovery windows, making data more available to end users.
- Redesign your testing philosophy to do more thorough testing by supplying full test duplicates without requiring significant resources.
- Reduce outages due to system changes and increase development productivity.

- Use fewer duplication resources
- Use less media, CPU, channel, and data communication resources.

Note: SnapShot is a separately-priced product for the Shared Virtual Array.

Chapter 3 Virtual Storage Architecture

This chapter provided an overview of the powerful capabilities made possible through the unique virtual storage architecture used in the Shared Virtual Array (SVA). This chapter describes how this innovative architecture sets a new standard for online storage systems.

Dynamic Mapping

At the heart of the Shared Virtual Array architecture is dynamic mapping, a unique method of mapping data that provides powerful online storage capabilities, while presenting a 3990 interface to the host and user. How does dynamic mapping work? When configuring an SVA subsystem, the user defines the system configuration in terms of traditional 3380 and/or 3390 DASD volumes. To the host, data transferred to an SVA subsystem appears to be stored on these traditional DASD volumes, but these volumes are functional volumes and do not exist as physical devices. The term “functional” in the SVA terminology describes the host view of the SVA -- the virtual DASD system configuration.

The data stored on an SVA subsystem is dynamically mapped across a collection of physical devices organized into a logical group called an array (refer to Figure 3-1 on page 28). Mapping between functional 3380 and 3390 volumes and the SVA is accomplished with tables of pointers located in the SVA Control Unit.

Functional Volumes and Tracks

Just as traditional 3380 and 3390 volumes consist of specific data tracks, Shared Virtual Array functional 3380 and 3390 volumes consist of specific functional tracks containing user data.

Data is stored on a functional track in much the same way that data is stored on a physical 3380 or 3390 track. However, the data on a functional track is compressed. In addition, the physical placement of a functional track onto the array is different from the placement of 3380 or 3390 track onto a traditional device.

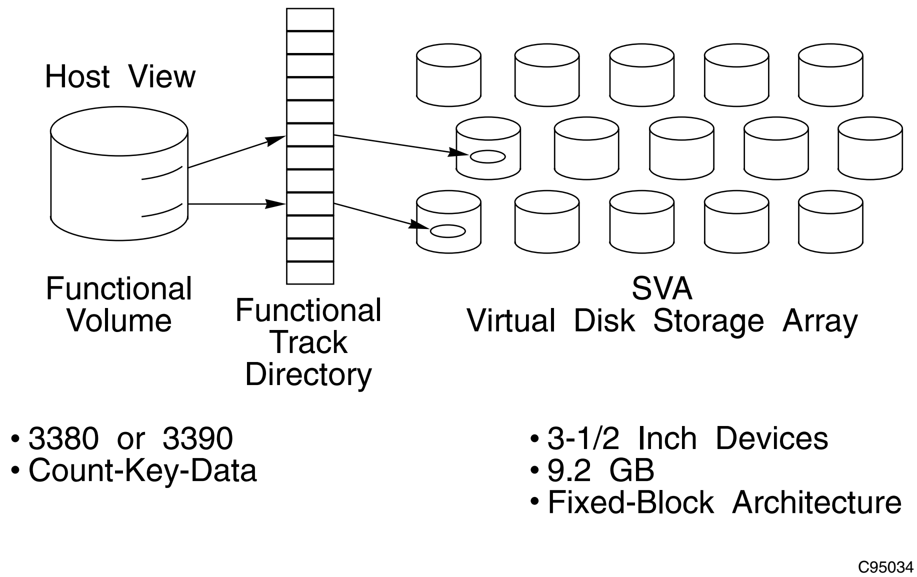

Figure 3-1 Host View of the Shared Virtual Array

The placement of a functional track on the physical devices is not predetermined or "fixed." Instead, the Shared Virtual Array Control Unit dynamically allocates a physical location for each functional track. The track is de-staged to the arrays, and the track location is stored in the functional track directory (FTD), which resides in the control unit base memory. In addition, the SVA generates redundancy information that protects all of the data stored in the system from loss. The redundancy information is also de-staged to the arrays.

The functional tracks of a functional volume are distributed among and across the available physical devices in an array. In this way, a functional volume has many more actuator arms and paths to stored data. This architecture provides a high degree of parallelism, and allows parallel pre-staging and de-staging of functional tracks for high performance.

The SVA's physical devices are formatted with physical tracks that are divided into fixed sectors. When a functional track is read or written, the process must begin on a physical sector boundary; however, it may begin on any sector in a physical track. The number of sectors required to contain a functional track is determined by the functional track's size, and by how well the data in the functional track compacts and compresses. The SVA always transfers entire functional tracks between the cache and the devices. This method of allocating functional tracks allows more efficient use of storage capacity. Traditionally, a shortcoming of CKD systems is that they must reserve, in advance, a fixed location and a fixed capacity for each track. In the SVA, the physical capacity occupied by a functional track depends on the amount of physical capacity the data requires. The SVA Control Unit assigns only as much physical capacity as the functional track needs.

Also, the SVA never updates data in place. When a functional track is updated, it is written to a new physical location. The location of the original functional track is marked as free space, and its storage capacity is released to the SVA free space collection facility for future use.

Finally, the SVA is not subject to pinned tracks from media defects. When a media fault is detected in a volume attached to other cached DASD controllers, the data cannot be written to that same spot; a pinned track results. The SVA automatically rewrites the data on such defective spots to another location without interrupting host access to data. No pinned tracks are experienced, and no user intervention is required.

Arrays, Array Tracks, and Array Cylinders

Functional volumes and functional tracks are the host view of the Shared Virtual Array subsystem. Arrays, array tracks, and array cylinders are the SVA view of the physical storage devices.

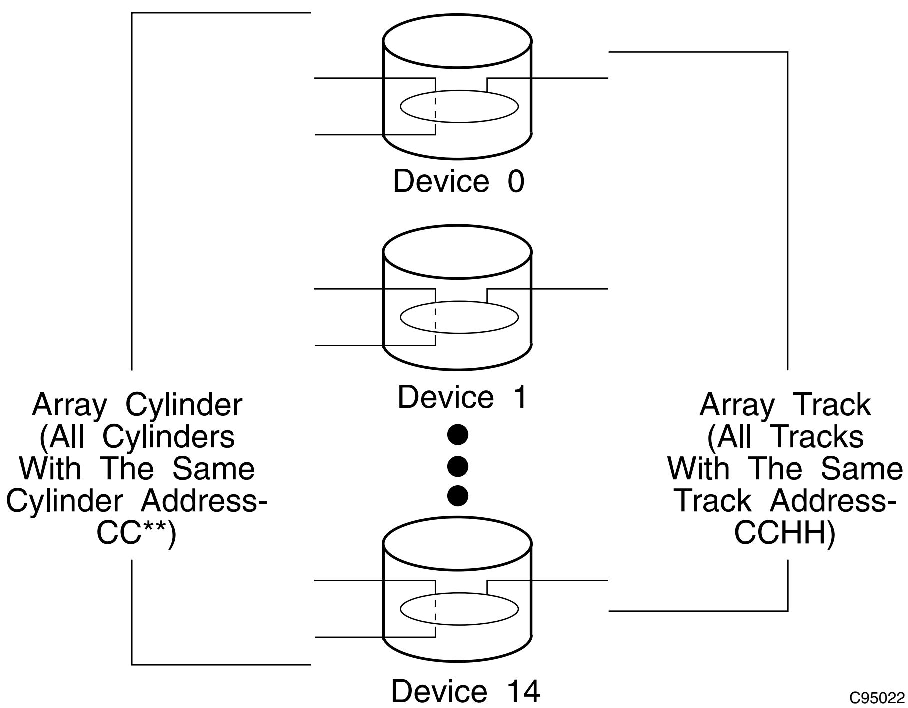

Figure 3-2 A Shared Virtual Array, Array Track, and Array Cylinder

An array is a logical grouping of physical devices. An array track is the set of all physical tracks in an array that have the same physical track address (same cylinder and head). An array cylinder is the collection of all of the array tracks that may be accessed at a common actuator position (same cylinder). In the Shared Virtual Array, a functional track may reside on one or more physical devices. Figure 3-2 on page 29 illustrates an array, an array track, and an array cylinder.

Functional Device Table

Dynamic mapping uses tables of pointers, assigned and located in the Shared Virtual Array Control Unit, to map the location of the data in the functional volumes.

One set of pointers is known as the functional device table (FDT). This table, which is stored in the control unit's shared memory, contains the functional configuration of the system (functional device type, functional volume address, functional volume size, etc.) as defined by the user. A backup copy of the FDT is stored on a mirrored storage disk in the control unit and on the disk arrays. (Refer to Figure 3-3.)

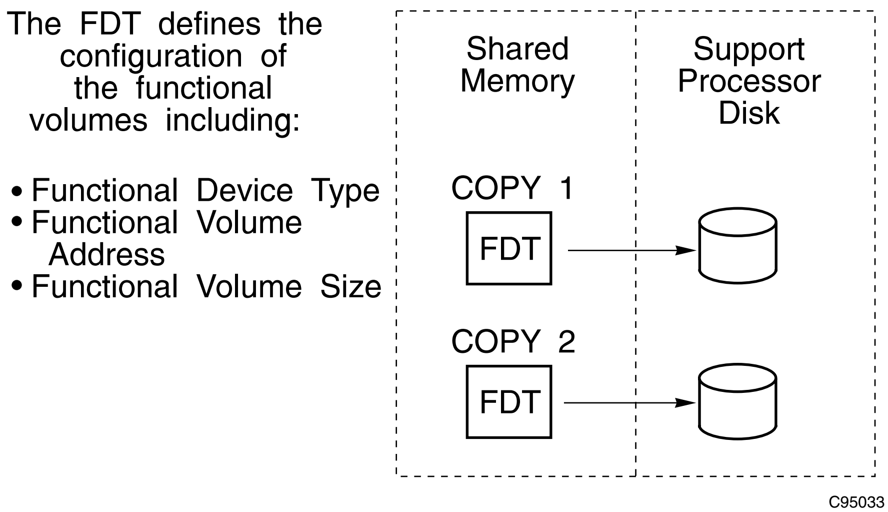

Figure 3-3 Shared Virtual Array Functional Device Table

Functional Track Directory

A second table of pointers, known as the functional track directory (FTD) has an entry for each functional track defined in the system. The entry points to the physical location of the functional track on the disk arrays, and is used to convert functional cylinder and head information from channel programs to physical storage information.

The FTD is kept in the control unit's base memory. (Base memory, which is memory unavailable to host applications, maintains data required by the system for operations.) The system continually makes new copies of the FTD and writes them to predefined locations on the arrays. The copies are "fuzzy image" because the system continues to allow changes to the FTD while the copy is being made.

During the FTD copy process, changes to the FTD are journaled to nonvolatile storage (NVS). These changes are eventually written to the disk arrays. If a failure occurs in base memory, the control unit rapidly reconstructs the FTD by restoring the last completed copy and applying the changes recorded in the journal. Figure 3-4 on page 31 illustrates the functional track directory recording and protection process.

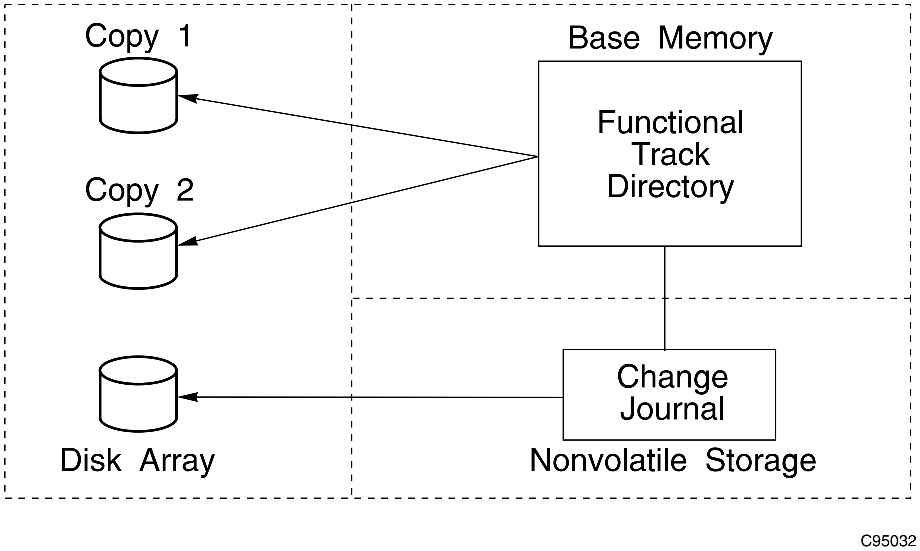

Figure 3-4 Functional Track Directory Recording and Protection

Actuator Level Buffers

The Shared Virtual Array provides additional device buffering for asynchronous data transfers. The SVA has actuator level buffers (ALBs) that allow each device in the array to read, write, or seek concurrent with data transfer operations.

ALBs allow non-synchronous data transfers to occur independently of device transfer rate, seek activity, or rotational position. Rotational latency is minimized because the rotation position of the physical device does not have to be synchronized with the data transfer to and from cache. An SVA physical device performs an arbitrary full track read or write operation in, at most, 105% of a single revolution (one revolution plus one sector latency).

Data Transfer

All data transfers in the Shared Virtual Array subsystem go through the cache in the SVA control unit. Front-end (channel transfer) operations are independent of back-end (device transfer) operations. See Figure 3-5 on page 32.

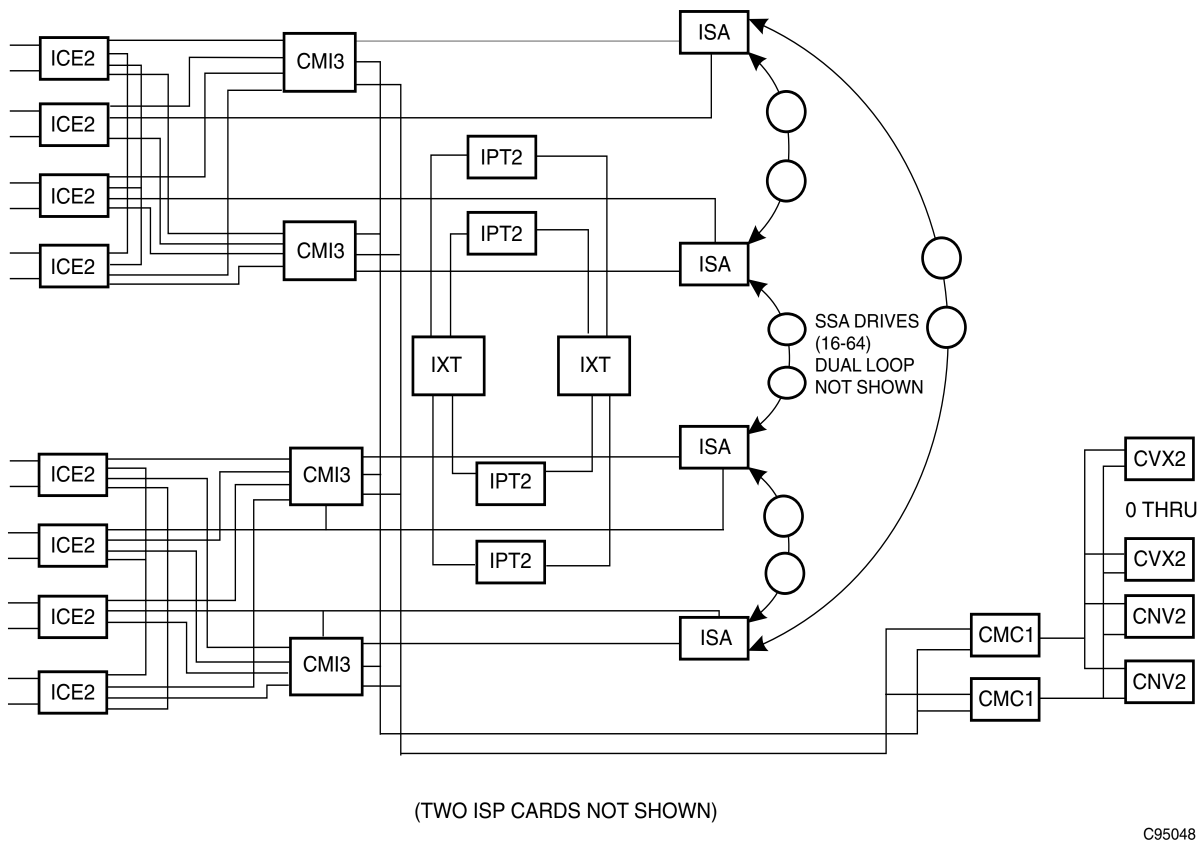

Figure 3-5 Data Transfer Internal Block Diagram

Channel Transfers

Channel transfers involve movement of data between the host channel and cache. This movement is identical to channel-to-cache transfers in other online storage systems, except that the Shared Virtual Array provides more parallel channel attachments, more internal data paths (16), and a proprietary algorithm to compress and decompress data to improve throughput.

All channel transfer operations pass through one of the two16 channel interfaces to the associated storage cluster. The SVA has 16 storage paths and eight microprocessors that allow it to process up to 24 concurrent channel operations. (See note below.)

Note: The functional processor (IUPs) are a pooled resources. That is to say they operate in a multitasking/multiprocessing fashion. A given IUP can in fact be simultaneous processing a backend process as well as initiating a new SIO. A IUP is not required to perform a data transfer and is available to take on new work. A 9500 gives the highest priority within the subsystem to starting new I/Os. During the data transfer of a given I/O the 9500 will start process another SIO.
On a fully populated 9500, there are 16 frontend data paths. It is possible for all eight processors to start addition I/O while these data transfers are occurring. So, for a 100% Read hit case there can be 24 concurrent I/Os in progress (16 that are executing data transfers and 8 that are doing

command processing). Note that the architecture has 16 channels. Each channel can talk to any of eight frontend data paths within its cluster. Again, the data paths are a pool of resources and do not have a fixed allocation to either a channel or an IUP.

Each data path has excess bandwidth capacity. A given channel is not capable of saturating a frontend data path. The fact that a channel has multiple data paths to select from, further extends this capability.

DASD Fast Write Operations

All writes to the Shared Virtual Array are DASD fast writes to cache memory and NVS. When the host system writes a record to cache, the record is transferred simultaneously and at channel speed to cache and NVS. NVS acts as backup to the cache. When the operation is complete, the SVA signals the host, the channel is released, and the track is eventually queued for transfer from cache to the disk arrays. Figure 3-6 illustrates this process.

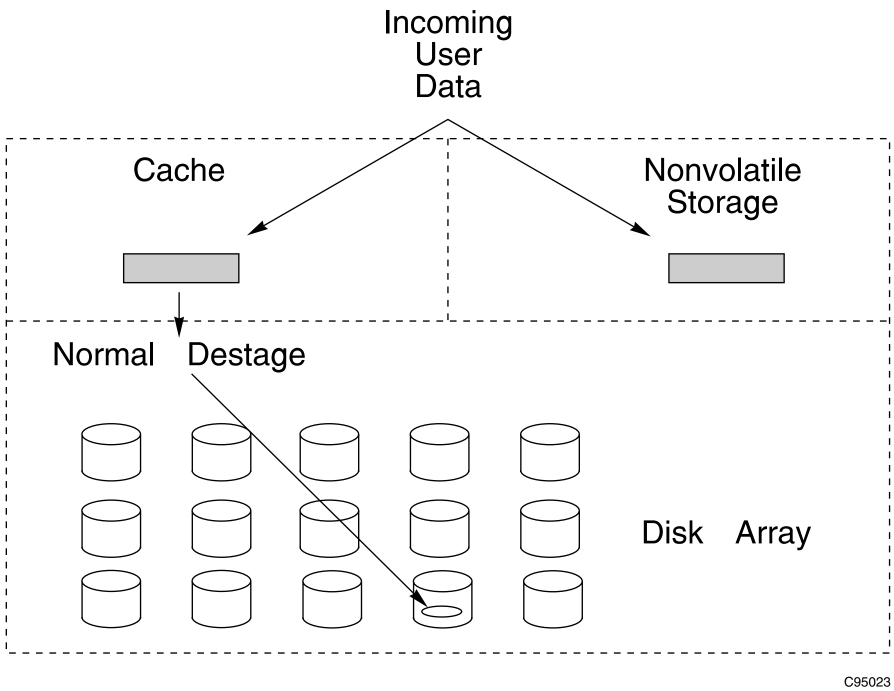

Figure 3-6 Shared Virtual Array Write to Cache (DASD Fast-Write)

If a cache failure occurs before or during the de-staging process, the track is recovered by applying the changes stored in NVS to the old data from the backend. If a power failure occurs, this same process occurs automatically upon resumption of power.

A functional track stays in cache as long as it is active; otherwise it is displaced by other, more frequently used functional tracks. If a functional track that has been chosen for replacement in cache contains modified data, the functional track is de-staged and then displaced. Cache and NVS can be freed as soon as the functional track is de-staged and the selected array track is successfully written.

When the host system transfers data to the SVA subsystem, storage path hardware in the SVA Control Unit automatically compresses the data fields of the functional track before they enter the cache. Because most of the data stored in cache memory is compressed, effective cache capacity is much larger than its physical capacity.

Cache Fast Write Operations

The Shared Virtual Array also treats cache fast write operations as DASD fast write operations. The SVA does so to ensure the integrity of user data. In a traditional cache fast write mode, a cache or power failure would destroy cache fast write data. In the SVA, the cache fast write data is protected by NVS and is not destroyed by a cache or power failure.

Fibre Channel/SCSI Data Transfers

Known as the “Record Format Assist Feature”, for SCSI data transfers. This feature, available in operational microcode levels 1.5.xx and higher provides the following features:

- Improved compression due to longer run lengths.
- Increases the maximum amount of user cache available to the customer because fewer directory extents are required.
- Better track utilization and increased virtual device (LUN) capacity. For example, the use of 4K records doubles the track utilization (49152 bytes/track versus 24576 bytes/track). This allows device capacity to increase from 3.5GB to 7GB for mod 9 devices.
- Supports multiple virtual devices per LUN.
- A two times performance improvement for write and read hit accesses.

Read Data Transfers[1]

For read operations that are aligned to a valid block size, the data transfer proceeds in a normal fashion. The transfer will be handled by the CIP decoding the logical block address sent by the host and mapping it to a configured block size. For unaligned accesses (512 byte accesses that don’t align on a block boundary), a new command is issued by the CIP to the IUP. This command has additional information that includes number of 512 blocks to be transferred and the starting block address. The record that contains the unaligned data is read into the ICE card’s full track buffer. Please note that the compressed data CRC is checked at the ICE card in the decompression logic. The data is then written uncompressed to temporary space in cache. Accumulated CRC values are saved for each block, including the original CRC bytes. The accumulated CRC is used to verify the data has been stored correctly in cache. The host data is then sent to the ICE card with CRC appended. Data integrity is ensured by this

1. While fibre channel attachmet supports RFA, fibre channel operation for read and write data transfers is slightly different.

CRC as data is sent to the host. After the successful transmission of data to the host the temporary cache space is freed up.

Write Data Transfers

For write data transfers that are aligned to a block size, the data transfer proceeds in a normal fashion. For unaligned accesses, the new SCSI command is issued by the CIP to the IUP. As with the reads, the record that contains the unaligned data is read into the ICE card's full track buffer. The data is then written uncompressed to temporary space in cache. Accumulated CRC values are saved, including the original CRC bytes. The accumulated CRC is used to verify the data has been stored correctly in Cache. Host data is transferred to cache with CRC generated from the ICE card. The updated host data along with the unchanged data is transferred from cache to the ICE card buffer. Virtual field CRC is generated and appended to the data by the sector CRC logic on the CMI3 card. The PERC logic on the CMI3 card verifies the integrity of the data that is being transmitted to the ICE2 card. The last step is for all of the data to be sent through the compression logic on the ICE2 card and written to cache as a compressed record. After the successful transmission of data to cache, the temporary write space is reclaimed.

Table 3-1 Data Utilization based on Block Size

Block Length (rec size byt)	rec/trk	byt/trk (used)	byt/cyl (used)	byt/dev (MByte-10^6)	
				Mod 3	Mod 9
512	48	24,576	368,640	1,230	3,691
2048	20	40,960	614,400	2,050	6,153
4096	12	49,152	737,280	2,460	7,384
8192	6	49,152	737,280	2,460	7,384
16384	3	49,152	737,280	2,460	7,384

Positive Side Effects

The new track format with larger block sizes presents a number of positive side effects:

Track utilization/Virtual Capacity

The current implementation of storing data in 512 byte fixed sized records limits the number of records on a track. This limits a track to holding only 24K bytes of customer data. The new format with 4K byte records doubles the amount of customer data per track to 48K bytes. This allows base device sizes to go from 1.2GB to 2.4GB for mod 3 devices and from 3.5GB to 7GB for mod 9 devices.

Intra Track Overhead

Storing twice as much data per track reduces by half the number of track boundary crossings for large transfers. SCSI code implementation is designed to have a separate write or read multiblock command for each track. Current measurements with a channel monitor show that this intra track time can be as high as 900 µsec. The new format allows a 1M byte IO operation to save 19 msec.

Cache Utilization

Increases the maximum amount of user cache available to the customer because fewer directory extents are required.

Cache Bandwidth

The software transfers each record in cache as an individual entity. In addition, the cache hardware breaks up large transfers into 512 byte segments to allow the available bandwidth to be distributed among the active cache ports. Records that are less than 512 bytes compressed reduce the effective cache bandwidth because of the additional arbitration, control and status cycles. 512 byte records with a 2:1 compression ratio reduce the cache bandwidth down to a number between 120MB/sec and 140MB/sec. Tracks formatted with 4K byte records with a 4:1 compression ratio allow the cache bandwidth to maintain its current throughput of 160MB/sec.

Compression Ratio

The current data compression design within Iceberg makes use of a history algorithm. The greater the data set to generate history over, the better the compression ratio. At 512 byte block size it is predicted that the compression ratio will be approximately 2:1. A 4K byte block size should allow the compression ratio to approach the 4:1 value that is observed with OS390 data.

Bypass/Inhibit Cache Operations

Since all Shared Virtual Array data transfers must pass through cache, 3990 bypass cache operations are treated as cache operations. However, as soon as the host request is completed, the tracks in cache associated with that transaction are placed in the least recently used (LRU) queue before tracks associated with normal cache operations. Bypass cache tracks do not accumulate in the SVA cache, and high throughput and performance are maintained. Inhibit cache loading operations are handled in a similar manner.

Device Transfers

Device transfers are the movement of data between the cache and the physical disk storage. The Shared Virtual Array supports 16 simultaneous data transfers, which are independent of channel activity, between the SVA Control Unit and disk storage.

Because of dynamic mapping and actuator level buffering, the device transfer operations in the SVA are radically different from those in traditional CKD DASD systems. While staging operations are performed in response to a host I/O request as in other cached DASD systems, de-staging operations are collected into groups for bulk transfer. Staging and de-staging operations are discussed more fully in the following sections.

Destage Operations (Write Operations)

A normal destage operation (Figure 3-7 on page 37) is a data transfer (or update) from cache memory to a disk array, with redundancy information constructed in the control unit.

Through its storage management algorithms, the control unit software chooses the arrays on which to place the data. Devices are written to in the order in which they appear in the array, and array tracks with the same head number are scheduled before array tracks with a higher head number.

A write request is sent to the physical device and the ALB is filled with data from the cache. The disk seeks to the specified cylinder and selects the specified head. When the head is oriented to the correct sector, data is transferred from the ALB to the disk.

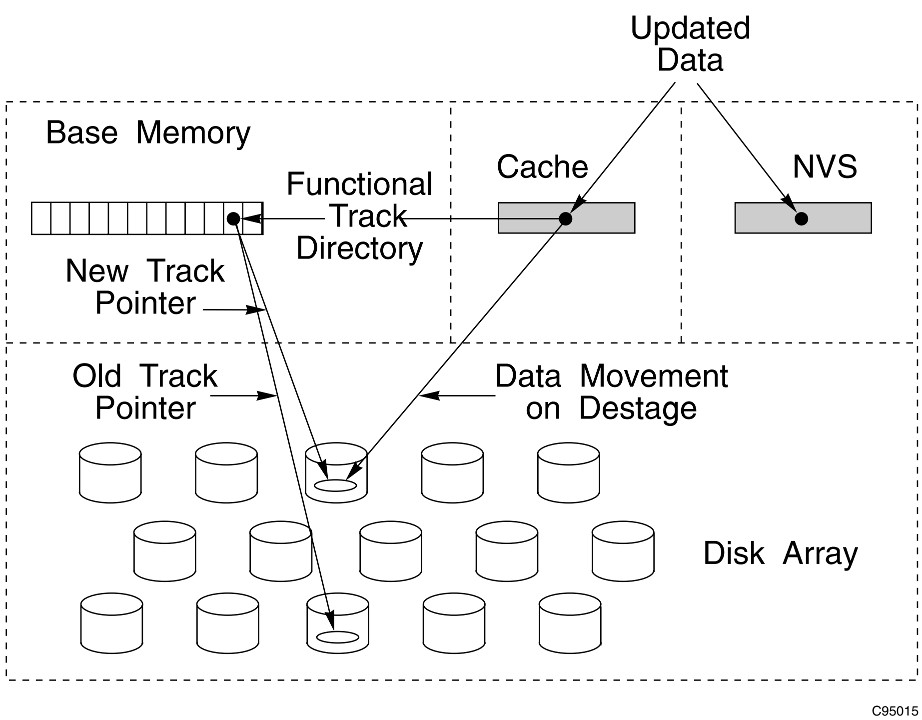

Figure 3-7 Shared Virtual Array De-staging

As data is transferred from the cache to the disk array, redundancy information is created and accumulated in the redundancy data buffers. When a complete array track has been written, the redundancy data is also written to the array. The control unit software is informed that the destage requests have been satisfied, and the space in NVS that held the modified data is released. The tracks in cache are released when the normal cache management algorithm selects them for displacement.

Dynamic mapping never updates in place, which allows for some of the important storage management features discussed later in this chapter.

Stage Operations (Read Operations)

When the host system requests a functional track from the system, the control unit looks for the functional track in cache memory. If the functional track is not found in cache (a cache miss), the control unit initiates a stage operation (Figure 3-8 on page 38). In a stage operation, the control unit allocates space in the cache, determines the location of the functional track from the FTD, and initiates a read request to the physical devices.

When a device that contains all or part of the requested functional track receives a read request, the disk seeks to the physical track and begins to read data into its ALB.

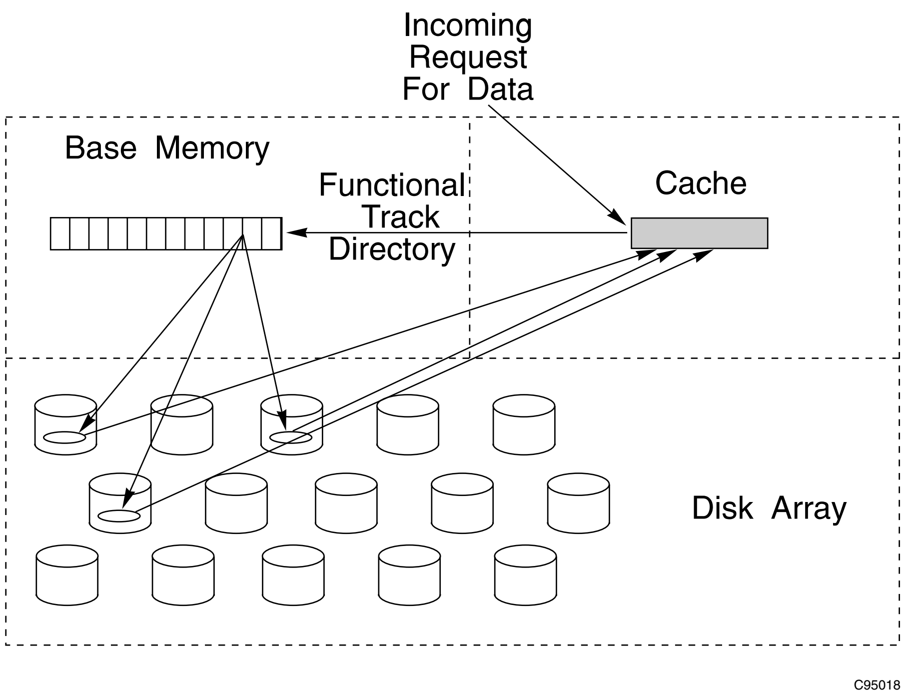

Figure 3-8 Shared Virtual Array Staging

When the control unit determines that the request has been satisfied, the cache releases the data to the host. The data is decompressed as it moves from the control unit to the channel. This cache-to-channel transfer occurs at channel speed.

A normal stage operation (Figure 3-8) is a data transfer that can span multiple devices (from one, two or at the most, three physical devices) to cache. All staging operations occur with full functional tracks. If more than one physical device is involved, the transfers occur in parallel.

Pre-stage Operations

Functional tracks are automatically spread across and among all of the devices in an array. Whenever sequential processing is implied or inferred, the Shared Virtual Array Control Unit can pre-stage sequential tracks into cache. Since sequential tracks are not on the same physical device (as they are on traditional DASD), they can be pre-staged in parallel. This parallelism permits host data transfers to occur at channel speed through cache, rather than at device speed, directly from the DASD.

Dual Redundancy Arrays

Dual-redundancy arrays are the innovative method that the Shared Virtual Array uses to protect user data from loss or destruction due to a device failure. The dual redundancy arrays in the SVA offer the user secure data storage and retrieval without the loss of capacity. Furthermore, while dual copy and simple RAID 5 provide only one level of redundancy, the SVA offers two levels of redundancy to protect user data.

Array Implementation

There are two configuration options for an Shared Virtual Array: a 15-device array or an 8-device array. In a 15-device array, which is referred to as a 13+2+1 array, one physical device in the array is reserved as a spare device and is globally available to the system. The SVA also supports 5+2+1 and 5+2 redundancy groups (five data drives + two redundancy/parity + one spare). This configuration, used with 4.5 GB drives, provides improved storage capacity granularity. All configurations support dual redundancy.

Within the group, user data is recorded on identically addressed tracks on all but two of the drives. The identically addressed tracks on the other two drives are reserved for the two levels of redundancy data generated by the system. One level of redundancy data, provided through the generation of simple “P” parity, is recorded on one of the drives. A second level of redundancy data, provided through a second parity calculation and referred to as “Q” parity, is recorded on the other drive.

The SVA’s implementation of P and Q parity is known as *Reed-Solomon redundancy,* and is generated from the data recorded on the drives in the array. However, specific drives are not dedicated to redundancy data. User data and redundancy data are rotated among all tracks on all of the drives in the array. Figure 3-9 on page 40 diagrams how the SVA generates Reed-Solomon redundancy data in a 15-drive array.

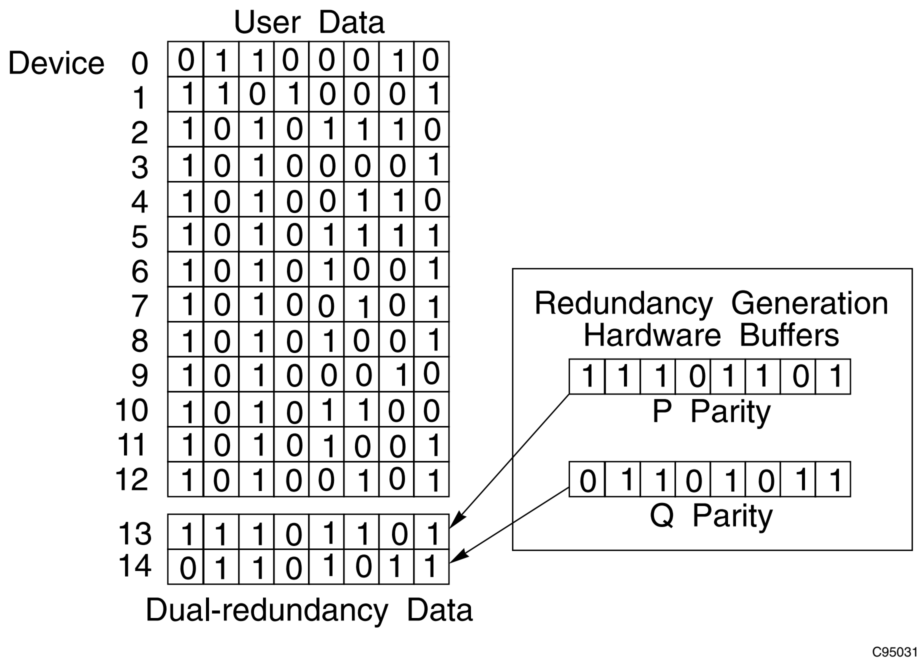

Figure 3-9 Reed-Solomon Redundancy in the Shared Virtual Array Subsystem

Array Track Recovery

As stated, an array track is a set of all physical tracks in an array that have the same physical track address. Any time a physical track is unreadable, the Shared Virtual Array initiates an array track recovery process. Array track recovery takes advantage of the SVA dual redundancy arrays to recreate lost data.

During the array track recovery process, all of the readable physical tracks in the array track are read into a redundancy data construction buffer, along with the P and Q parity tracks. Then the unreadable physical track is recreated in cache and, if possible, written back to the disk array at the same location it previously occupied. If the track cannot be written back to the array at the same location, the recreated track is automatically written to an alternate location.

Functional Track Recovery

Each functional track resides on all or part of one or more physical tracks. When a physical track is unreadable, the functional tracks contained on that physical track cannot be read. The Shared Virtual Array functional track recovery (FTR) process uses array track recovery to recreate a functional track. If the system receives a request for data stored on the lost physical track before the array track recovery is complete, the FTR task becomes a high priority task. The high back-end data bandwidth and parallelism of the architecture allow the immediate reconstruction of host-requested data in real time. Host operations proceed uninterrupted.

Note: All readable tracks are copied to the spare. Unreadable tracks are recovered using the functional track recovery process.

Globally Switchable Spare Drives

Standby redundancy for the physical drives is provided by the globally switchable spare drives. Because they are not logically allied with an array, these spare drives may be logically configured by the Shared Virtual Array Control Unit into any position in any array containing drives of the same type as the spare. When a drive fails, the system automatically switches to a spare drive from the pool of spares, and begins the drive reconstruction process (described in the following section). The array is restored to a fully redundant state. Two (2) spare drives are provided with each system, which is adequate to meet the most stringent availability requirements. Additional spares are optionally available.

Drive Reconstruction

When a drive in an array fails, the data from the remaining drives, including the redundancy information, may be used to recover the lost data. The drive is reconstructed onto a spare drive using the array track recovery process.

Drive reconstruction is an automatic, nondisruptive, background task. Drive reconstruction recreates all of the data contained on the drive in less than two hours, (typically 25 minutes or less), and the user has access to all system data the entire time. The failed device may be removed and replaced with a new drive, which becomes part of the pool of globally switchable spare drives. Figure 3-10 on page 42 illustrates the drive reconstruction process.

Reconstruction of a single failed device is a low-priority background task. Should a second device in the same array fail during the recovery process, reconstruction is escalated to emergency priority, and both devices are completed in 25 minutes or less. Again, access to data, even on the failed drives, is maintained at all times, but a small increase in access time (20%) will be experienced.

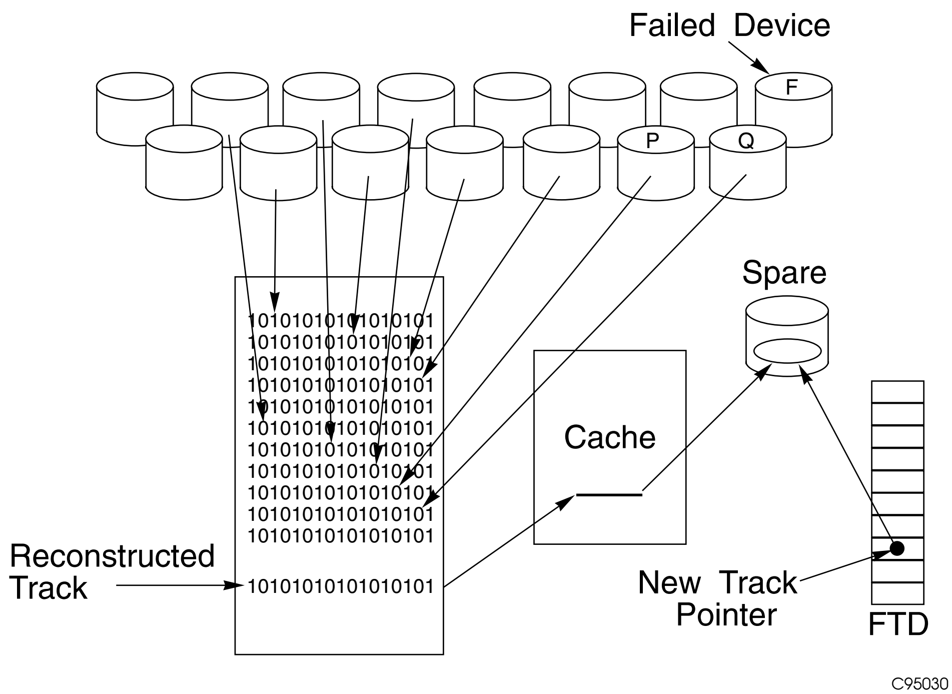

Figure 3-10 Shared Virtual Array Drive Reconstruction Process

Automatic Media Maintenance

The Shared Virtual Array offers two levels of automatic media maintenance: device level and control unit level.

Media defect handling, error recovery, error correction, and error retry are performed at the device level and are transparent to the control unit. If a drive experiences a read error, the drive electronics retry the operation and apply error checking and correction (ECC). If possible, the corrected data is read and written back to the original sector. Otherwise, the corrected data is read and assigned an alternate sector, and the original sector is marked as defective.

In the event of an uncorrectable media error, the control unit takes over. In the control unit, the system recreates the data via the array track recovery process. The recreated data is written to an alternate location on the original device, and the original location is marked as defective.

Automatic media maintenance is a standard SVA feature; no special utility program is necessary. It occurs as a background task and is completely transparent to the host. Should a device exceed a threshold of media errors, the system considers it a failed device and begins the drive reconstruction process.

Dual-Redundancy Arrays

Because the Shared Virtual Array automatically protects all data stored in the system with dual redundancy arrays, it does not require dual copy to protect host data. Dual-redundancy arrays offer two significant advantages over dual copy:

- The Shared Virtual Array provides two levels of redundancy, so data is always protected against single failures and many multiple failures. Dual copy offers only a single level of protection; should one volume fail, the data is no longer protected.
- The Shared Virtual Array provides two levels of data protection without sacrificing user storage capacity. Dual copy requires an additional volume for every volume protected by dual copy, doubling the capacity required to store dual copy data.

Thus, dual redundancy arrays offer greater operational redundancy than dual copy, at a significantly lower price and performance cost. Also, all the rated capacity of the SVA may be used for user data. The capacity that supports the SVA dual redundancy arrays and spare drives is not included in the SVA’s rated capacity. In summary, dual copy capability is unnecessary, because it is an inefficient use of the SVA resources.

Cyclic Redundancy Check (CRC)

In addition to parity checking, the Shared Virtual Array employs cyclic redundancy checking for added error detection capabilities. Hardware at the channel interface generates a cyclic redundancy check (CRC) for each data field. This CRC remains with the data field as it moves throughout the system. The CRC is checked by the channel interface hardware when the data field is read by the host. A beginning-to-end data integrity check is provided that guarantees the field has not changed anywhere within the system.

Shared Virtual Array Data Management

There are several advanced storage management capabilities that result from the unique architecture of the Shared Virtual Array. Data compression, data compaction, self-tuning operations, and free space collection (all discussed in more detail in the following sections) significantly increase the internal bandwidth and performance of the SVA subsystem. These features reduce the amount of physical capacity required to store data as compared to conventional CKD DASD. In addition, they enable a unique the SVA offering, functional capacity, which is also described below.

Advanced Data Compression

All operations in the Shared Virtual Array subsystem are performed on compressed data. Each storage path between the host and the cache includes hardware that compresses data as it arrives from the host, and decompresses data as it returns to the host. Cache utilization is greatly improved, as is the effective data transfer rate between the control unit and disk arrays. The compression algorithm used by the SVA is a proprietary modified Lempel-Ziv algorithm. The data compression process is totally transparent to the host. The compression ratio for the SVA varies with the type of data being stored, but typically about a 3:1 ratio (uncompressed to compressed data lengths) results.

Data Compaction

On conventional DASD, the gaps associated with blocking data on CKD devices occupy physical capacity. The amount of capacity lost to gaps varies with block size, and may be significant. However, dynamic mapping enables automatic data compaction. The gaps are not mapped with the functional tracks, and therefore do not consume disk array capacity. The Shared Virtual Array offers efficient use of storage capacity and improved access to data with a higher data throughput.

Free Space Collection

When data that has been staged is modified in cache, it is not written back to its previous location on disk arrays since that would invalidate the redundancy data for that array track. When a functional track has been updated, the updated track is written to a new location and the data in the previous location is marked as physical free space. Over a period of time, an array cylinder accumulates “holes” of physical free space. To make use of this capacity, the un-released data remaining on the partially filled array cylinder is read, combined with other segments of un-released data, and rewritten to a new array cylinder. This process is known as free space collection.

Free space collection relocates data from an array cylinder based on how much free space it contains. All of the un-released data from that array cylinder is staged into cache and then de-staged to a new array cylinder.

When all of the data from the array cylinder has been released, the array cylinder is added to the free array cylinder pool for use in later de-staging operations.

Free space collection uses the same staging and de-staging processes as described previously except that free space collection is a lower priority task than demand staging and (except in extreme circumstances) a lower priority task than de-staging.

Dynamic Configuration

Through dynamic configuration, the Shared Virtual Array provides significant configuration flexibility. Dynamic mapping of SVA separates the functional configuration (the host view) of the system from the physical devices in the system. Thus, the functional configuration of the system is easily modified to fulfill the users' requirements without being constrained by the attached physical devices.

The SVA architecture allows the user to define the SVA subsystem as up to four functional 3990 systems, each with 256 functional volumes. Therefore, one SVA subsystem may address up to 1024 functional volumes with 3380 or 3390 track geometries, regardless of the size of capacity of the system. These 3380 and 3390 functional volumes may be intermixed within the functional configuration of the system. (The only constraints on the intermixing of functional volumes are those inherent in the host configuration software.)

Dynamic configuration offers distinct advantages. Consider that MVS and VM allow only one I/O operation to a DASD volume at a time. Dynamic configuration increases the number of accessible volumes in an SVA subsystem, which in turn proportionally increases the number of I/Os that can be issued to the system in parallel. This enables reduced I/O queueing, increased path utilization, and increased system throughput.

Functional Capacity

Functional capacity is an innovative capability that is standard with Shared Virtual Array. Functional capacity exploits dynamic configuration, data compaction, and data compression to provide a system with functional storage capacity that substantially exceeds the physical capacity of the disk arrays.

With functional capacity, the SVA user may define up to 1024 functional (virtual) 3380 or 3390[2] volumes with maximum capacity, no matter what the physical capacity of the attached disk arrays. The only restriction is that the physical capacity of the attached disk arrays support the actual net capacity load (NCL) of the system. (The net capacity load is the physical disk array capacity needed to store all of the compressed functional tracks actually occupied by data in a functional configuration.)

So if, for example, the user defined the SVA subsystem as 256 functional 3380-K volumes, functional capacity enables an addressable functional capacity of 483 gigabytes. If the user defined the SVA subsystem as 256 functional 3390-3 volumes, functional capacity enables an addressable functional capacity of 726 gigabytes.

2. 341 functional volumes for 3390-9s.

Because data compression and data compaction decrease the amount of capacity required to store data, they decrease the net capacity load of the system. This decrease makes functional capacity even more powerful. Depending on functional track blocking, functional volume utilization, and how well the stored data compresses, functional capacity is projected to effectively support three to four times the physical disk array capacity.

Functional capacity offers the storage administrator greater flexibility and productivity. The SVA can better tolerate over-allocation, because over-allocation of functional capacity for data sets does not consume physical space. The data sets require less copying and less monitoring as they expand. Functional capacity enables the reduction of “out of space” abends and unit control block (UCB) queueing on the host processors. Functional capacity also elevates storage management from the physical volume level to the system level.

Self-Tuning Architecture

At the physical device level, the Shared Virtual Array provides an inherently self-tuning architecture. Functional tracks associated with a functional volume are typically spread across the physical devices in the system. Read and write operations are effectively distributed uniformly across all of the devices in the system.

This automatic load balancing of physical devices along with automatic free space collection effectively allows the SVA to self-tune the performance of the physical devices.

Simplification of SMS Storage Management

In the Systems Managed Storage (SMS) environment, the Shared Virtual Array offers several important storage management simplifications.

The SVA simplifies storage class and storage group definition. Because the SVA provides caching and DASD fast write, selection of storage class parameters can be reduced to default settings. Thus, the storage administrator does not need to define multiple storage classes to exploit the various device characteristics. Because the SVA provides fault tolerance, caching, DASD fast write, and 3380 and/or 3390 emulation, storage groups can be oriented by business application, not device characteristics. Also, the fault-tolerant capabilities of the SVA make dual copy unnecessary, further simplifying storage class and storage group definitions.

As a result of these advantages, the need to segregate data based on performance and availability requirements is minimized. An SVA subsystem provides enhancements to the SMS environment beyond traditional DASD systems.

Summary

The virtual storage architecture of Shared Virtual Array opens the door to new online storage capabilities--most importantly those of continuous access to data and virtual DASD, providing the lowest total cost of ownership for online storage.

Chapter 4 Shared Virtual Array Operability

The Shared Virtual Array subsystem operates within the constraints of industry standard data streaming channel architectures including IBM or compatible 30XX, 43XX, 9370, ES/9000, and S/390 series processors and servers. This chapter describes the operation of the SVA within those constraints in terms of:

- Configurability
- Attachability
- Operating Environment
- Interface extensions
- Functionality
- Error recovery.

Configurability

The configuration of the SVA can be modified via a local operator panel or at the host with the Shared Virtual Array Administrator (SVAA). Using the local operator panel menus and dialogs, the user can define the hardware that comprises the subsystem, and then initiate the host I/O configuration program (IOCP) or Hardware Configuration Definition (HCD).

Sixteen Path Capability

A 9500 Shared Virtual Array Control Unit has 16 storage paths between the channel and cache, which allows the SVA to perform up to 24 simultaneous transfers between host channels and cache.

Attachability

The Shared Virtual Array subsystem attaches to any host that supports ESCON channel specifications and 3990 hardware attachment. The SVA supports synchronous (ESCON-compatible) channels through ESCON ports.

Operating Environment

The Shared Virtual Array subsystem presents a 3990 image and is supported by the operating systems described in the following table.

Table 4-1 Minimum Levels of Software to Support Shared Virtual Array

Operating System	
OS/390	Version 1, Release 2
MVS/EAS	Version 5, Release 2.2
VM/ESA	Version 2, Release 2.1/3
VSE/ESA	Version 1, Release 4.0

Note: VSE/ESA does not have SVAA support which, although not required, is highly recommended. By running VSE/ESA as a guest operating system under VM/ESA, SVAA can be used to monitor the Shared Virtual Array.

Interface Extensions

All management and monitoring functions that are provided for host interface with the Shared Virtual Array subsystem are controlled through a proprietary StorageTek interface extension. The SVA microcode communicates with the host in a packet data structure. The packet contains individual elements that execute the functions necessary to complete the system actions. The packets support the transfer of information between the host and the system without disrupting host operations.

StorageTek has an Extended Control and Monitoring (ECAM) interface that is available to software vendors under licensed agreement. This interface allows the host software to directly communicate with an SVA subsystem to gather statistics and issue commands.

Functionality

The Shared Virtual Array functionality surpasses that of a 3990, and the way in which the SVA provides that functionality is often different. Because of architectural differences, the SVA subsystem defaults are not always the same as those for the 3990.

Channel Commands

The Shared Virtual Array supports most channel commands and parameters of channel commands that the 3990 supports. Those commands that the SVA does not support are described in the *Shared Virtual Array Reference*.

Because the architecture of the SVA is different, commands that are applicable to the SVA subsystem may be handled in a different manner than in the 3990.

Commands that are not applicable to the SVA subsystem and are not in direct conflict with the SVA functionality are handled in a manner that is transparent to the host. Only in those cases where a command is in direct conflict with the SVA functionality does the SVA inform the host of an incompatibility.

Dual Copy

Dual copy commands are not supported by the Shared Virtual Array, because the dual redundancy arrays in the SVA provide the data protection capability that is dual copy's major function.

Cache Fast Write Operations

The Shared Virtual Array handles a cache fast write command as a DASD fast write command. Therefore, the data is written to both cache and nonvolatile storage. However, cache fast write functionality is emulated within the SVA. In a cache fast write operation, the system:

- Accepts requests to enable and disable cache fast write
- Emulates the cache fast write ID, incrementing it as appropriate
- Maintains and reports the statistical counters that report activity for cache fast write data, reflecting the activity associated with writes that would have been cache fast writes in a native CKD environment.

Cache Management

Because the Shared Virtual Array requires cache for any read or write operation, the system manages cache in a different manner than the 3990. For example, in a bypass cache operation or an inhibit cache loading operation, if a read or write command is issued for a track that is not in cache, the track is staged from the disk arrays into cache. When the operation is complete, the newly staged track is queued for immediate release or de-staging to the arrays.

Sense Information

Sense information presented to the host by the Shared Virtual Array subsystem may not be identical to that presented by the 3990. (Refer to the *Shared Virtual Array Reference* for more information about sense data.) In general, sense information relating to a functional operation is identical to the 3990. However, sense data fields needed to maintain the system are specific to SVA and compatible with StorageTek's Performance Monitoring and Predictive Maintenance (PM2) program.

Error Recovery Procedures

The error recovery routines provided by the host system control programs (e.g. MVS/ESA) support the Shared Virtual Array subsystem. The Shared Virtual Array does not require any additional error recovery routines for successful incorporation into the host system environment.

Wherever possible, error recovery actions are executed within the SVA. In most cases, the SVA avoids involving the host error recovery system until after the component requiring service has been isolated within the system. Only in a few specific cases does the SVA use host error recovery actions to their full extent. (Refer to the *Shared Virtual Array Operation and Recovery* for more information about error recovery.)

Service Information Messages (SIMs)

SIMs are reports sent to alert an operator of system problems requiring service or to convey service-related information. SIM functions are controlled by the support facility. When the Predictive Service Analysis (PSA) function determines that a field replaceable unit (FRU) has failed, is about to fail, or is running with degraded performance, it sends a SIM to the host indicating:

- The general area requiring service (i.e., storage path, channel, FRU, etc.)
- The severity level of the problem
- The most likely FRU needing replacement
- The impact (if any) that servicing will have on normal operations.

Based on information received from SIM messages, the remote support center either services the problem or dispatches a CSE for servicing as needed. SIM information allows the remote support center to dispatch a CSE with the required FRU on the first service call, reducing the mean time between repairs.

Chapter 5 Shared Virtual Array Components

This chapter discusses the capacity options and models of the Shared Virtual Array (SVA) subsystem. Important components of the SVA's fault-tolerant design are also described. As always, StorageTek manufactures the SVA with the highest quality components available chosen for their superiority and reliability.

Shared Virtual Array Control Unit Components

The components of the Shared Virtual Array Control Unit are shown in Figure 5-1. and discussed in detail in the following sections.

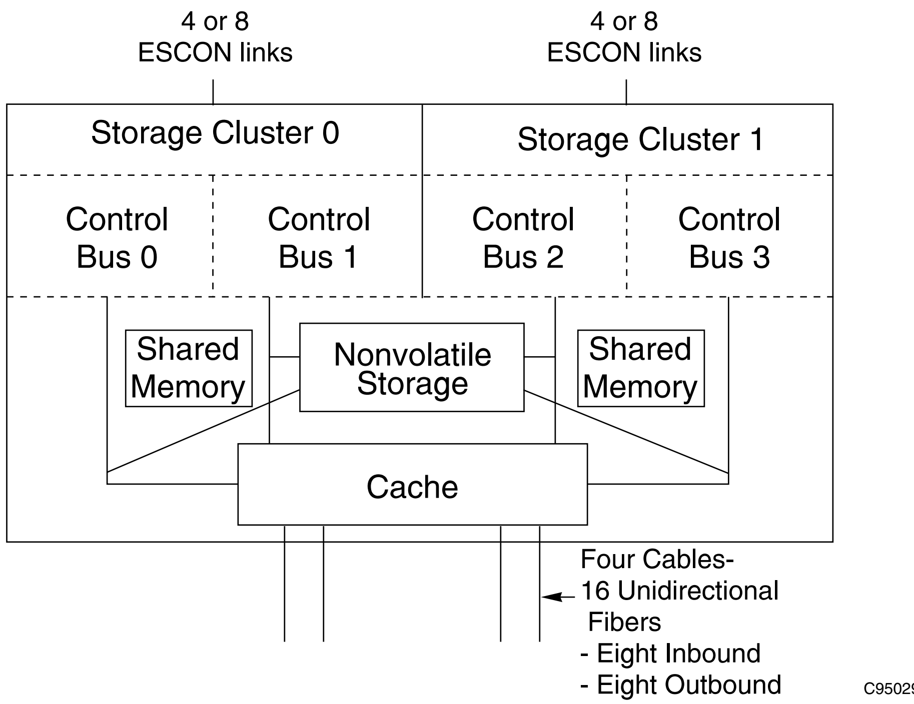

Figure 5-1 Major Components of the Shared Virtual Array Control Unit

Data Paths

The Shared Virtual Array offers the following:

- Up to 16 external channels between the host system and SVA
- Sixteen independent data paths between the channel and the cache
- Eight high-speed microprocessors

External

The Shared Virtual Array offers a choice of external channel interfaces. The system attaches up to 16 ESCON or fibre channel interfaces.

The SVA native ESCON support allows the SVA connectivity to ESCON channel ports on CPUs and/or ESCON channel ports on ESCON directors. Additionally, support for 19.6 MB/sec sustained channel transfer speeds, as well as EMIF, is provided.

Internal

A 9500 Shared Virtual Array Control Unit includes 16 internal data paths between the channel and the cache arranged in two storage clusters; each storage cluster with eight storage paths. The 16 storage paths are linked so that a disconnected operation may be switched to any free storage path in the control unit.

Each storage cluster supports eight simultaneous channel transfers (one per storage path).

A 9500 Shared Virtual Array Control Unit has 16 internal data paths between the channel and the cache, which allows the SVA to perform up to sixteen concurrent I/O transfers between host channels and cache.

Memory

Four kinds of memory are located in the control unit: cache memory, nonvolatile storage memory, base memory, and shared memory. Each type of memory has a specific function within the high performance, fault-tolerant design of the Shared Virtual Array subsystem.

Cache Memory

Cache memory performs high-speed data storage and retrieval operations, as initiated by the host. All data transfers go through cache and all writes to cache are DASD fast writes. Cache operations include, but are not limited to, data transfers between a channel and the cache, a device's ALB and the cache, and a microprocessor and the cache. The data stored in cache is transferred to and from the host at channel speeds.

Up to 32 concurrent operations in the cache result in improved efficiency and a higher data throughput. Data compression improves efficiency and throughput even more. Finally, Shared Virtual Array implements several caching algorithms that efficiently manage cache resources.

Nonvolatile Storage

Nonvolatile storage (NVS) acts as backup to cache during write operations. The Shared Virtual Array NVS is partitioned onto two nonvolatile memory cards.

Base Memory

Base memory is the memory required to operate the Shared Virtual Array system. Base memory is cache memory that is not available to the user. Because it is not included in the stated cache size, base memory does not subtract from user cache.

Shared Memory

The Shared Virtual Array's duplexed shared memory performs two major functions. First, it provides storage for functional and physical device status information. Second, it is the memory used by the microprocessors to communicate with each other. Each shared memory card is attached to the shared memory bus of both storage clusters, and contains the same control information.

Microprocessors

The Shared Virtual Array has eight powerful, high-speed microprocessors which can process channel programs on up to 16 separate channels simultaneously. Thus, the SVA may process up to 16 independent channel data transfers at the same time.

Local Operator Panel

The local operator panel, the hardware module that provides the operator access to the system is recessed into the left front door of the Shared Virtual Array Control Unit cabinet. The local operator panel has a liquid crystal display (LCD) and a key pad with 10 “Soft Keys,”16-Key hex (0 - F) keys, and an “ENTER” and “CLEAR” key. See Figure 5-2 on page 55.

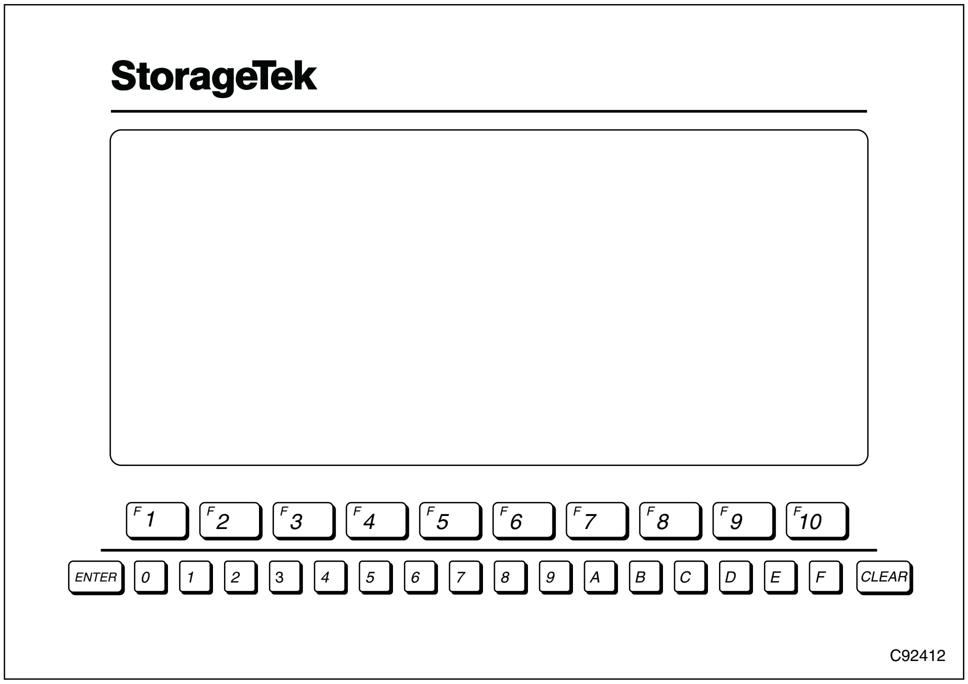

Figure 5-2 Shared Virtual Array Operator Panel Display

Shared Virtual Array Support Processor

The Shared Virtual Array support processor is the microprocessing unit in the Control Unit that monitors SVA operations and initiates repair activity. The ISP generates service information messages (SIMs), provides remote maintenance support, and makes concurrent maintenance possible. In addition, the ISP provides the following functions.

- Loads and validates the operational and diagnostic code
- Audits the field-replaceable units (FRUs) for type and machine-level compatibility

- Monitors voltages, temperatures, and the operations of the microprocessors
- Detects and reports errors
- Provides the communication links for the Customer Service Remote Center (CSRC).

Physical Devices

Drives (9 GB, 3-1/2" form factor) are organized into "redundancy groups," or arrays.

9500 System Array Capacities

Table 5-1 lists the system capacity configurations that are possible with the 9500.

Table 5-1 9500 9.0 GB SCSI/SSA Array Capacities

Equivalent System Capacity[1]	System Array Configuration(s)[2]	No. of Arrays by Configuration	Number of Data Drives	Number of Redundancy Drives	Total Drives in System (with Spares)[3]
422 GB[4]	13+2	1	13	2	16[5]
585 GB	5+2 13+2	1 1	18	4	24[5]
845 GB[4]	13+2	2	26	4	32[5]
1007 GB[4]	5+2 13+2	1 2	31	6	39[5]
1267 GB	13+2	3	39	6	47[5]
1429 GB[4]	5+2 13+2	1 3	44	8	54[5]
1689 GB[4]	13+2	4	52	8	62[5]

Note:

1. Equivalent capacity = number of data drives (n) x drive capacity (8.95 GB) x typical compression ratio (4:1)
2. Number of data drives + redundancy/parity drives per array
3. Two spare drives are required per system
4. Capacity orderable only as an upgrade to an existing system (163 MB/5+2 or 422 MB/13+2, as applicable
5. Total of all data drives + redundancy/parity drives + spare drives in system.

Shared Virtual Array Fault Tolerance

The Shared Virtual Array provides redundancy at every potential point of failure within the system. Also, all major hardware components in the system, with the exception of the motherboards, are hot-pluggable, field-replaceable units (FRUs) and may be replaced in the field by trained CSEs without loss of access to data, shutting down the system, or powering down any components.

The following sections describe more aspects of the Shared Virtual Array's fault-tolerant design.

Control Unit Functions

The Shared Virtual Array Control Unit features a high degree of modularity in addition to component redundancy. The fault-tolerant and modular components unique to the Control Unit include those discussed in the following sections. The fault-tolerant components or redundant components found in both the Control Unit and the Disk Array Units are discussed in the section "System Level Redundancy" on page 58.

Storage Clusters

Each of the two storage clusters in the control unit is comprised of eight independent storage paths. Any single failure or service action in a control bus impacts only 25% of available path resources as opposed to 50% in the comparable DASD controllers.

Cache

Besides including multiple cache cards for each cache configuration, cache in the Shared Virtual Array subsystem is also modular. If an integrated circuit on a card fails, the card is fenced off; the remainder cache continues to operate. Since all FRUs except the motherboards are hot-pluggable, the failed card may be removed and replaced without affecting any other cards in the system, and without disabling cache operations.

Nonvolatile Storage (NVS)

The Shared Virtual Array Control Unit has 32 physical megabytes of NVS. NVS is modular; if a card fails, the card is fenced off and the remainder of NVS remains operative. One card failure has minimal impact on system performance, and complete data availability is maintained. Again, the failed NVS card may be removed and replaced without affecting any other cards in the system and without impacting data availability.

Shared Memory

The shared memory cards in the control unit are shared by all of the components of the system and are duplexed. The primary shared memory bus of one storage cluster is connected to the secondary shared memory bus of the other storage cluster. Redundancy is achieved by writing and reading simultaneously to the primary and secondary busses of each storage cluster. If one shared memory card fails, the other shared memory card supports system operations without interruption.

NVS Battery Backup

In the event of a power outage, a battery backup system protects NVS for at least 72 hours (assuming the batteries are fully charged). A primary battery provides at least 72 hours of DC power in the event of an AC power failure. A redundant battery in the battery backup system offers at least 72 hours of protection should the primary battery fail, or in some cases, 72 hours of additional protection to that supplied by the primary battery.

Shared Virtual Array Support Processor

There are two Shared Virtual Array support processors in an SVA Control Unit. One ISP performs the maintenance and support functions while the other is on standby. In the event of a failure in the active ISP, the standby ISP assumes control of the maintenance and support functions. The two ISPs are interconnected, providing inter-processor communication, reliability, and processor failure recovery.

System Level Redundancy

At the system level, every major component except the motherboard is a hot-pluggable, field-replaceable unit (FRU) with a redundant counterpart. Because most online storage system failures are caused by power supply interruptions, special attention was focused on the AC and DC power supplies. The redundant components are discussed in the following sections.

AC Power Distribution Assemblies

The Shared Virtual Array Control Unit has two AC power distribution assemblies. For true AC redundancy, the two AC power distribution assemblies should receive power from two independent sources that are not likely to fail simultaneously. If both power distribution assemblies are connected to different circuits from the same AC source, reliability is enhanced, but true AC redundancy is not realized.

DC Power Supplies

The controller section of the 9500 has redundant DC power supplies. Each tray of drives in the disk array section also contain redundant DC power supplies. The supplies share the load under normal conditions, and one supply is capable of handling the load if the redundant power supply fails.

Cooling System

The Shared Virtual Array Control Unit has two separate cooling systems. One for the power supplies, consisting of four small fans, and four larger fans over the logic section. Each system has four blowers. The fans all have sensors on them. If one fan fails, the rest will still maintain sufficient cooling. In the event of a fan failure, the 9500 will post sense information indicating that a fan has failed.

This page intentionally left blank

Chapter 6 Shared Virtual Array Serviceability

Continuous system availability requires that servicing and installation of upgrades to the system must be nondisruptive. The design of the Shared Virtual Array (SVA) supports nondisruptive hardware servicing and installation. Nearly all major components are hot-pluggable, field-replaceable units (FRUs). Power or service regions are automatically powered down when replacing a component. Also, SVA offers a shorter mean time for repairs, and often a single service call may be scheduled to service multiple problems. These time-saving service and installation features are discussed in more detail in this chapter.

Predictive Maintenance/Service Management

The Shared Virtual Array includes a powerful predictive maintenance and service management system, and an extensive customer service operation to complement its nondisruptive servicing. The fault-tolerant system design, which compensates for a component that requires service, is discussed in “Shared Virtual Array Fault Tolerance.” The design that automatically detects components that require service, and allows a CSE to replace them without disrupting system performance, as discussed here, is an important advantage of the SVA.

Service Management

The Shared Virtual Array has a comprehensive service management system that includes predictive service analysis (PSA), FRU fencing, service information messages (SIMs), error and event logging, ServiceTek, guided FRU replacement (GFR), general service information (GSI), FRU validation, and diagnostic information. These service management features are described in the following sections.

Predictive Service Analysis

The predictive service analysis (PSA) for the Shared Virtual Array monitors the system’s operational performance. PSA is a knowledge-based system that resides in the SVA support processor, while its detection capability is integrated into SVA’s operational code. PSA monitors the system’s operational performance and, using a data base of possible situations and sets of rules, detects “change-in-status” situations, including:

- If a system component requires service
- If a system component is running with degraded performance or has error or statistical counters that need to be down-loaded
- If information about the status of a system operation should be reported.

If PSA detects a problem, it initiates problem isolation. If necessary, PSA dynamically reroutes data flows by fencing off the FRU or a subset of the FRU. PSA accurately isolates components that require service, and delivers the information to the host through service information messages (SIMs) and to Customer Service Remote Center (CSRC) through ServiceTek.

Note: If PSA detects an abnormality for which it cannot specifically identify the FRU that requires service, the Shared Virtual Array also offers a general service information procedure for non-diagnosable faults. This procedure, which is discussed later (“General Service Information (GSI),” helps the CSE troubleshoot and test for more obscure problems.

Service Information Messages (SIMs)

If PSA identifies a change-in-status condition that should be reported, the ISP sends a message to the host as a service information message (SIM). The SIM includes a description of the condition and its cause and, in the case of a component failure, the impact of the failure and the impact of the repair. Depending on the severity of the SIM, a SIM alert message may be sent to the host console.

All SIMs are passed to the host system, but not all SIMs are relayed to the host’s operator console. The SVA allows the user to select the severity level for SIMs that are to be passed as SIM alerts to the operator console.

Performance Monitoring/Predictive Maintenance (PM2)

The Shared Virtual Array also sends the SIM to the host error recovery program (ERP) where it is logged to the error recording data set (ERDS). StorageTek’s Performance Monitoring and Predictive Maintenance (PM2) program allows you to print reports that include SIM information. PM2 provides performance and error reports for maintenance history, daily reporting, and monthly statistics. PM2 reports also aid the CSE in troubleshooting system problems by correlating SIM and ServiceTek information with the error and event logs.

Program for Online System Testing (POST)

POST offers an additional diagnostic tool that tests and verifies the reading or writing to functional volumes from the host. POST reports all errors as they are reported in the sense data, and includes all SIM information in both online and stand-alone environments.

ServiceTek

The Shared Virtual Array is supported by ServiceTek, StorageTek’s machine-initiated maintenance facility, where the system will automatically report problem isolation information to the Customer Service Remote Center (CSRC).

ServiceTek calls are driven either by the detection of a fault, or by the expiration of the down-load timer (a time interval set by the StorageTek CSRC). The down-load timer ensures the regular collection of the system event log. If a problem is detected, ServiceTek sends an alert to the CSRC.

ServiceTek periodically downloads the event log to the Customer Service Remote Center (CSRC). The CSRC system integrates the event logs from all SVA ServiceTek users into an engineering data base. StorageTek engineers analyze this data base and create and update PSA rules. These rules are integrated into the new code releases, which can be non-disruptively loaded onto the system. This cycle allows us to constantly improve our service to SVA users.

Guided FRU Replacement (GFR)

If PSA identifies that a FRU (or part of a FRU) requires service, a SIM relayed to the host system displays specific information about that FRU. If the system is supported by ServiceTek, this information is also automatically relayed to the CSRC. If the system is not supported by ServiceTek, the user contacts the CSRC with this information. Thus, the CSE can bring the required FRU on the first service call, greatly shortening the mean time for repairs.

Once on-site, the CSE initiates a maintenance session without interrupting the system's normal activities. In this maintenance session, the system directs and coordinates the CSE's actions to ensure that the failed FRU is rapidly and accurately identified. To complete the maintenance procedure, the CSE pulls the suspect FRU and hot-plugs the new one. The system verifies that the correct FRU has been placed in the correct slot, checks the engineering change (EC) level of both the hardware and the microcode, and runs a limited set of diagnostics. Finally, the GFR facility removes the fence around the new FRU and the FRU is placed into service. The entire operation is nondisruptive.

All components in the system (including circuit cards) may be serviced in this manner, with the exception of motherboards. As noted, all maintenance activities are verified by the system.

General Service Information (GSI)

When PSA detects a problem but cannot pinpoint which FRU requires service from a list of suspect FRUs, the CSE may access a menu-driven, general service information facility. This program gathers and displays information about non-isolated failures. With this information, the CSE can run diagnostics on-site to isolate the problem.

Diagnostics

In special cases where PSA cannot suggest any FRUs, another ServiceTek feature, the Engineering alert, notifies the CSRC and engineering. Trained diagnostic engineers can access the system via a remote link that allows them to access specialized diagnostic routines. Through these, the CSRC personnel can obtain sense data and identify errors. When the problem is resolved, the PSA rules are updated and sent to the field.

Note: Although CSEs and CSRC personnel can use the local operator panel menus for diagnostic and maintenance information, they do not have access to any customer data.

Other Service Features

Other features contribute to the nondisruptive serviceability of the Shared Virtual Array. Its unique fault-tolerant design provides redundancy at nearly every point of potential system failure. (Motherboards are not redundant.) For example, should a cache card fail, no more than 25% of user cache is fenced off with only a slight impact on system performance.

The SVA offers several other nondisruptive service features, including:

- Two operationally redundant SVA support processors
- Modular diagnostics
- A time-stamped system event log
- Operational microcode ‘trace’ and ‘state save’
- Microcode support tools and file transfer
- Remotely-readable machine serial number
- Remotely-readable FRU ID data
- FRU identification by bar code labels.

Nondisruptive De-Installation of Disk Arrays

The Shared Virtual Array features a nondisruptive procedure for de-installing disk arrays. This operator-invoked procedure allows the user to drain data out of an array before taking the array off-line. Most often, data from the units is stored elsewhere in the system. In some cases, data may have to be removed from the system if its remaining storage capacity will be inadequate for the current data load.

Other Nondisruptive Procedures

Some of the other nondisruptive installation and de-installation procedures for Shared Virtual Array include:

- Installation of hardware engineering changes, whether to resolve a specific problem or for general level upgrade
- Installation of changes to the SVA support processor read-only databases, including the following tables.

  Fault symptom code to FRU relations

  FRU compatibility rules

  GFR rules
- Installation of additional cache memory
- Installation of additional channel interfaces.

Disruptive Services

Motherboard replacement requires that the system be powered down. However, the Shared Virtual Array system design minimizes the time required to power down the system, replace the motherboard, and power up again. In addition, initial microcode load (IML), diagnostics, and microcode changes may require system interruption.

Service

StorageTek products receive exceptional product support and service from Customer Services personnel. Support and service are available through the Customer Service Remote Center (CSRC), a 24 hours-a-day, 7 days-a-week, multi-level organization.

At the CSRC, diagnostic engineers investigate users' system problems and determine corrective actions. There are two levels of CSRC support. Central Support Level 1 is responsible for end-user customer support. Central Support Level 2 is responsible for CSE telephone support.

- Central Support Level 1

  Level 1 support receives customer calls and diagnoses problems. Using diagnostics, the event log, ServiceTek information, and the incident tracking database, the primary goal of Level 1 support is to work with the user to resolve system malfunctions without sending a CSE to the site. If required, Level 1 support dispatches a CSE to the site with the appropriate FRU and continues to minimize the impact of the system malfunction until the CSE arrives.

- **Central Support Level 2**

  Level 2 support provides telephone assistance to StorageTek technical specialists and CSEs. Level 2 uses all the tools available to assist the CSE in resolving a system problem in a timely manner.

Summary

StorageTek's goal is to offer users an online storage system that provides uninterrupted access to data. Nondisruptive serviceability, nondisruptive installation, exceptional quality, and a professional customer service organization are critical factors in providing continuous system availability. With the advent of the Shared Virtual Array, the quality and performance of online storage has reached a new standard.

This page intentionally left blank

Chapter 7 Shared Virtual Array Administrator (SVAA)

This chapter provides an overview of a host software product designed to complement the 9500 Shared Virtual Array (SVA)--the Shared Virtual Array Administrator (SVAA). An SVA subsystem does not require SVAA to operate, but SVAA facilitates the operation and enhances the capabilities of SVA, so it is highly recommended.

SVAA provides host terminal access and control (under MVS and VM) of the SVA subsystem and offers extended capabilities in the following areas.

- Dynamic Configuration
- Deleted Data Space Release (DDSR)
- Data collection and reporting with the Reporter function

DDSR and the Reporter function are powerful tools, available only through SVAA, that optimize the SVA's capacity utilization and system reporting. Dynamic Configuration is available through the local operator panel, but with SVAA these tools are more flexible, easier to use, and offer more options.

Dynamic Configuration

The SVAA allows the operator to query and modify the configuration of all attached systems from a host-attached terminal. Configurable elements are:

- The functional device configuration, including such selections as the number, type, and addresses of the functional devices
- The system configuration, including such selections as the number of global spares
- The channel configuration, including such selections as the channel status (enabled or disabled), the channel speed, addresses of devices, and the number of devices
- The drive module configuration, including forming arrays and assigning drives to partitions.

Virtual DASD allows the operator to define up to 1024 functional 3380 and/or 3390 devices (with the exception of 3390-9s). Without the SVAA, each functional device must be configured individually from the SVA local operator panel. With the SVAA, the operator may configure a range of functional devices at a time; this greatly improves the accuracy and ease of system configuration.

Configuration Reporting

The SVAA allows the operator to generate configuration and status reports on the configurable elements of one or more Shared Virtual Array subsystems. These reports may be printed or viewed online.

Deleted Data Space Release (DDSR)

The Deleted Data Space Release function allows the Shared Virtual Array to efficiently manage resources and significantly reduce its net capacity load (NCL). A reduction in the net capacity load translates to greater exploitation of the system's extended capacity and higher performance.

MVS systems delete a data set by flagging the file as deleted in the volume table of contents (VTOC). VM systems delete a file or minidisk by removing the file entry from the directory. In either case, the data set or minidisk is no longer accessible to the host, but it continues to use physical storage capacity until it is overwritten by another file in the same location on the same volume (Figure 7-1 on page 68).

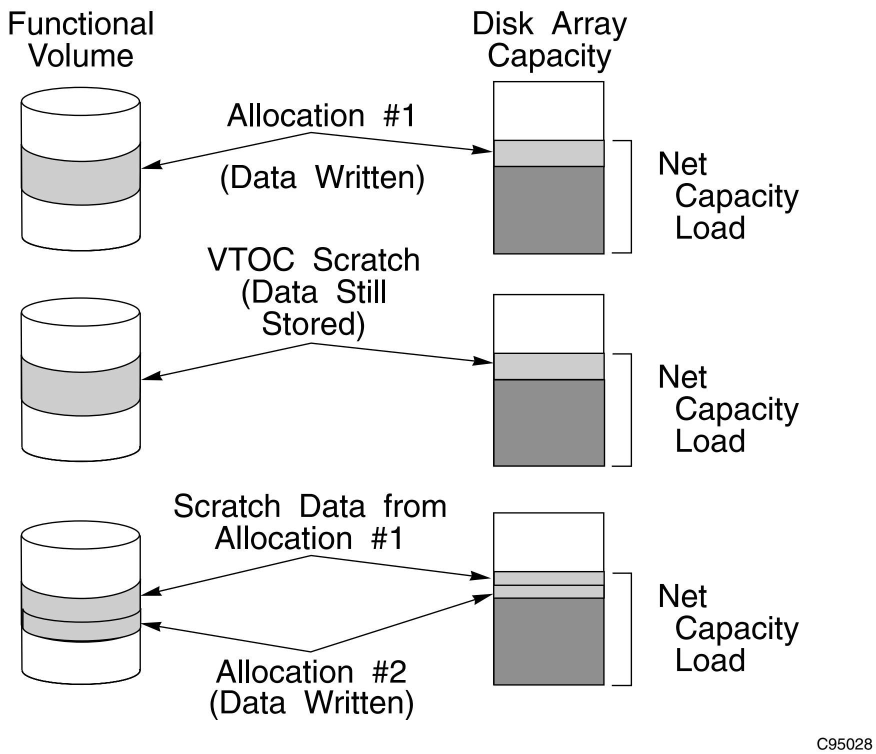

Figure 7-1 Deleted Data Set without DDSR

With the SVAA software, when a data set is no longer valid, the DDSR function directs the SVA to mark the space as free space available for collection by the free space collection facility (Figure 7-2 on page 69). This increases the amount of free space in the system and decreases the net capacity load of the system.

The operator controls the timing and scope of DDSR activity. In MVS, when certain conditions are met (such as data set scratch), the SVAA automatically performs DDSR functions. However, the operator may choose to direct the timing of the release activity or specify the functional volumes that are to be included in DDSR activity.

Refer to the *Shared Virtual Array Planning, Implementation, and Usage Guide* for more information on net capacity load (NCL) and deleted data space release (DDSR).

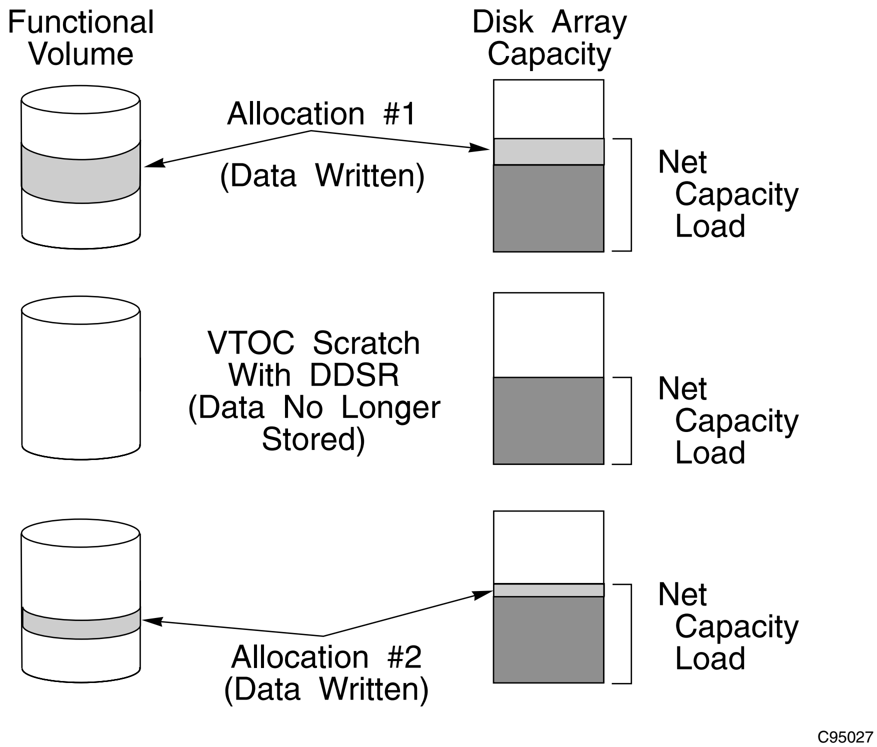

Figure 7-2 Deleted Data Set Release in a Shared Virtual Array Subsystem

Reporter

The SVAA collects and reports data on system utilization and performance with the Reporter facility. Data collection, performance history file maintenance, and reporting (for the Shared Virtual Array and all 3990 compatible systems) may be set up for specific time periods and intervals. Data collection may be synchronized with other host data collection tasks, such as RMF.

The SVAA provides a set of SAS and SAS/Graph programs to generate reports and graphic displays of performance data, or data and reports may be manipulated by other generalized report writer (GRW) and graphic display tools.

Data Collection

Data collection is a continuously executing task that wakes up at specified intervals to collect and log system data similar to SMF and RMF operations. If desired, data collection can be an SMF operation. Five categories of system data are collected.

- Functional volume performance
- Cache performance
- Functional space usage and capacity

- Physical (array) space usage and capacity
- Physical device performance.

Short-term performance data may be collected on demand, and performance and trend data may be collected continuously into a flat file or SMF.

Reporting

The Reporter facility generates performance and utilization reports corresponding to the five categories of data collected. These reports give the storage administrator valuable data; they may be used in optimizing performance and capacity utilization, analyzing performance, or historical trend analysis.

Functional device performance reports and cache performance reports are easily customized. The user can designate the dates and time periods represented in the reports.

The following performance reports are available:

Space-utilization report

> This report provides a view of system conditions at a specified time. This report shows the amount of functional space defined, allocated, and stored, and the amount of physical capacity occupied by each functional volume.

Overall summary report

> This report provides an overall summary of device performance and cache utilization for the defined report period by system, by individual functional device, and by reporting period.

Interval summary report

> This report provides a summary of statistics for each system over a defined time interval.

Time-ordered report

> This report shows (by functional device, partition, or system) the value of specified system parameters for each report interval.

Summary reports

> These reports provide a summary of statistics (daily, weekly, and/or monthly) for each functional device and each partition within the system.

Exception report

> This report lists functional devices and/or systems that fell outside the threshold limits for reported parameters during a specified time interval.

Performance bar chart

> This report plots the values for specified parameters (from one or more of the other types of reports) for a specified time interval.

Performance History File Maintenance

The SVAA extracts data from the data collection files and the Reporter facility summarizes the data into a SAS-format performance history file. The operator can access the performance history file for long-term detail analysis and historical trend reporting.

Shared Virtual Array Administrator (SVAA) Interfaces

The SVAA offers a selection of interface options, including:

- Interactive System Productivity Facility (ISPF)

  This option runs on any 3270-type EBCDIC (or compatible) terminal. The interactive menus have been designed to be as similar as possible to the Shared Virtual Array local operator panel menus. Wherever possible, both menu sets use consistent terminology.

- TSO command line interfaces

  This command line interface uses syntax similar to that of MVS/DFP Access Method Services. It also allows the operator to invoke the SVAA sub-commands in CLISTS or REXX EXECs. Because this interface is familiar, operators require little or no training to use it.

- Batch utility

  Using this option, commands can be executed in batch mode.

- Console operator commands

  This option allows you to enter a subset of the SVAA configuration commands at the system operator console.

Shared Virtual Array Administrator (SVAA) Requirements

Table 4-1 on page 50 describes the minimum host software requirements for an Shared Virtual Array subsystem. Table 7-1 on page 72 and Table 7-2 on page 72 describe the additional minimum software requirements for the SVAA functionality in MVS and VM environments respectively.

Table 7-1 MVS System Software Requirements (With SVAA)

Software Description	Minimum Version/Release
Operating System	
VSE/ESA	Version 1, Release 4.0
Dialog Support	
ISPF	Version 3.5
TSO/E	Version 2.0
Report Software	
SAS	Version 6.07

Note:

1. Report graphics and PC-based reporting are optional SVAA functions that require additional software.

Table 7-2 VM System Software Requirements (With SVAA)

Software Description	Minimum Version/Release
Operating System	
VM/ESA	Version 1, Release 2.1
Dialog Support	
ISPF	Version 3.2
Report Software	
SAS	Version 6.07 for mainframes
SAS/GRAPH[1]	
SAS/PC[1]	

Note:

1. Report graphics and PC-based reporting are optional SVAA functions that require additional software.
2. VSE/ESA does not have SVAA support which, although not required, is highly recommended. By running VSE/ESA as a guest operating system under VM/ESA, SVAA can be used to control and monitor the Shared Virtual Array.

Note: As indicated, if the SVAA is implemented without the required levels of supporting software, some corresponding functions will not be available.

READER'S COMMENT FORM

Manual Name: ______________________________ Manual PN: __________

Please check or fill in the items; adding explanations/comments in the space provided.

Which of the following terms best describes your job?

☐ Field Engineer	☐ Manager	☐ Programmer	☐ Systems Analyst
☐ Engineer	☐ Mathematician	☐ Sales Representative	☐ Systems Engineer
☐ Instructor	☐ Operator	☐ Student/Trainee	☐ Other (explain below)

How did you use this publication?

☐ Introductory text	☐ Reference manual	☐ Student/Trainee	☐ Instructor text

☐ Other (explain) ______________________________

Did you find the material easy to read and understand? ☐ Yes ☐ No (explain below)

Did you find the material organized for convenient use? ☐ Yes ☐ No (explain below)

Specific criticisms (explain below):

Clarifications on pages ______________________________

Additions on pages ______________________________

Deletions of pages ______________________________

Errors on pages ______________________________

Explanations and other comments:

Note: *Staples can cause problems with automated mail sorting equipment. Please use pressure sensitive or other gummed tape to seal this form. If you would like a reply, please supply your name and address on the reverse side of this form.*

Thank you for your cooperation. No postage stamp necessary if mailed in the U.S.A.

TO COMPLY WITH POSTAL REGULATIONS, FOLD EXACTLY ON DOTTED LINES AND TAPE (DO NOT STAPLE)

NO POSTAGE
NECESSARY
IF MAILED
IN THE
UNITED STATES

BUSINESS REPLY MAIL

FIRST CLASS PERMIT NO. 2 LOUISVILLE, CO U.S.A.

POSTAGE WILL BE PAID BY ADDRESSEE

STORAGE TECHNOLOGY CORPORATION
ENTERPRISE DISK INFORMATION DEVELOPMENT
ONE STORAGETEK DRIVE
LOUISVILLE CO 80028-2201

FOLD HERE AND TAPE DO NOT STAPLE FOLD HERE AND TAPE

If you would like a reply, please print:

Your Name: ______________________________

Company Name: ______________ Department: ______________

Street Address: ______________________________

____

Cut or Fold
Along Line