

Symmetrix DMX: Reliability, Availability, and Serviceability

Abstract

The Symmetrix DMX series delivers an advanced fault tolerant storage array design that includes full redundancy, proactive monitoring, and error detection and correction. This paper provides an overview of these Symmetrix DMX capabilities.

Date 07/30/2003

Copyright © 2003 EMC Corporation. All rights reserved.

EMC believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

THE INFORMATION IN THIS PUBLICATION IS PROVIDED “AS IS.” EMC CORPORATION MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND WITH RESPECT TO THE INFORMATION IN THIS PUBLICATION, AND SPECIFICALLY DISCLAIMS IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Use, copying, and distribution of any EMC software described in this publication requires an applicable software license.

EMC², EMC, Symmetrix, AViiON, CLARiiON, CLARAlert, DG, DG/UX, Navisphere, PowerPath, ResourcePak, VisualSAN, and The EMC Effect are registered trademarks and EMC Automated Networked Storage, EMC ControlCenter, EMC Developers Program, EMC Enterprise Storage, EMC Enterprise Storage Network, EMC Enterprise Storage Specialist, EMCLink, EMC OnCourse, EMC Orbit, EMC Proven, The EMC Effect Alliance, the EMC Information Orb, Access Logix, Application Transparent Failover, AutoIS, Automated Resource Manager, AVALONidm, C-Clip, CacheStorm, Celerra, Celerra Replicator, Centera, CentraStar, CLARevent, Connectrix, CopyCross, CopyPoint, CrosStor, Direct Matrix, Direct Matrix Architecture, EDM, E-Infostructure, E-Lab, Enginuity, FarPoint, FLARE, GeoSpan, HighRoad, InfoMover, MirrorView, OnAlert, PowerVolume, RepliCare, SafeLine, SAN Manager, SDMS, SnapSure, SnapView, SnapView/IP, SRDF, StorageScope, SymmAPI, SymmEnabler, TimeFinder, Universal Data Tone, WideSky, and where information lives are trademarks of EMC Corporation. All other trademarks used herein are the property of their respective owners.

C1016.2

Table of Contents

Table of Contents	2
Introduction	3
Reliability, Availability, and Serviceability.....	4
Symmetrix DMX Error Detection and Remote Support.....	5
Symmetrix Component-Level Redundancy	6
Global Memory Director Technology Overview.....	6
Global Memory Director Data Integrity	7
Triple Modular Redundancy with Majority Voting	8
Symmetrix Global Memory Technology Compared to Mirrored Cache	10
Symmetrix DMX Channel Directors	11
Channel Director Redundancy	11
Internal Control Data Path Redundancy.....	11
Power Supply Redundancy	12
Battery Backup	12
Fibre Channel Back-End Redundancy	13
Data Integrity for Fibre Channel Arbitrated Loops	14
Symmetrix DMX Data Protection Options	15
Symmetrix Component-Level Serviceability.....	15
Symmetrix DMX Control and Environmental Matrix.....	16
Communication Control Module	16
Environmental Control Module	17
Non-disruptive Operations	18
Enginuity 5670 Operating Environment Overview	19
Enginuity 5670 Non-Disruptive Operations	19
Inter-Family Non-Disruptive Upgrades (NDU) (5669 to 5670).....	19
Hardware Quality Assurance	19
Conclusion.....	20

Introduction

As increasingly more IT executives declare themselves to be in the service business, their organizational missions can be simply stated as delivering the right level of information services at the lowest overall cost.

Storage itself continues to play a vital role in delivering IT services. It directly impacts the performance, availability, and functionality of most applications – while improving time-to-market and ensuring business continuity. The fundamental need for a robust solution that not only supports uncompromised SLA demands, but also alleviates economic concerns, has challenged storage vendors to rethink their high-end architectures.

As the demands from internal and external customers force IT organizations to deliver the very highest service levels, performance must be augmented with the ability to support sudden “bursts” of unexpected activity. What's more, availability expectations have increased to where any interruption to the access of information is unacceptable.

Important functionality such as local and remote replication of data used to deliver business continuity must cope with more data than ever before, without impacting production activities. Furthermore, at the end of the day, all of these challenges must be met while continually improving economics.

To meet these challenges, EMC developed an innovative, new storage platform architecture called the Direct Matrix™ (DMX), which supports a revolutionary new product family – the Symmetrix DMX™ series.

Highlights of the Symmetrix® DMX include:

- **Architecture** – The Direct Matrix represents a significant advance beyond yesterday's bus and switch architectures. It delivers more performance, better scalability, and lowers costs, and provides a clear roadmap for future enhancements.
- **Performance** – The Symmetrix DMX dramatically resets performance expectations in a broad range of demanding transactional, decision support, and consolidation environments. More importantly, when coupled with the Enginuity™ storage operating environment, Symmetrix DMX has a unique ability to effectively react to bursts of unexpected activity while continuing to deliver high service levels.
- **Availability** – The Symmetrix DMX goes beyond yesterday's designs to set a new standard in availability, including the elimination of buses and switches, and the incorporation of triple-module-voting for key components. Power systems and the ability to do online upgrades have also been dramatically improved.
- **Functionality** – The rich suite of functionality found on previous Symmetrix models has been brought forward to Symmetrix DMX, doing vastly more work. Key local and remote replication functions used for business continuity are now orders of magnitude faster, while minimizing application performance impact. Additionally, as part of EMC's AutoIS® strategy, the majority of Symmetrix software management capabilities can be used to create an open, multi-vendor storage management environment.
- **Economics** – The power of the Direct Matrix Architecture™ allows better use of hardware resources. New modular packaging allows lower cost-of-entry into high-end capabilities. High-performance parity RAID makes better use of disk resources for more environments than ever before. And, new high-performance ports allow more servers per port, reducing the cost of consolidation.

With the introduction of Symmetrix DMX, what customers demand from high-end storage in terms of availability, performance, replication, scalability, and management, has changed. The new high-end availability is more than just redundancy; it means non-disruptive operations and upgrades, and “always online.” Delivering high-end performance means handling all workloads, predictable or not, under all conditions. High-end replication means copying any amount of data anytime and sending it any distance. High-end scalability means more than capacity; it means having the flexibility to handle any service level or application, cost-effectively, no matter how your business changes. And high-end storage management goes beyond monitoring storage arrays; it means managing the service

levels the business expects, from provisioning to business continuity. Symmetrix DMX has redefined the standard in high-end storage.

Reliability, Availability, and Serviceability

Binding service-level agreements (SLAs) commit IT organizations to deliver agreed-to measurable support metrics such as application performance, end-user response time, and system availability. Even in the face of a component failure, such as a faulty disk drive or power supply, these organizations are required to provide the highest levels of performance, availability and serviceability (RAS) – without compromise.

IT executives around the globe have recognized that downtime is not only measured in minutes or hours, but is also calculated as potential revenue loss, missed opportunities, or dissatisfied customers. Any operational impacts resulting from a component failure must be completely transparent to the individual applications and users relying on information availability to drive the business.

Today's mission-critical environments require a high-end storage solution that guarantees uncompromised levels of service, backed by a vendor who will provide quality support and a smooth migration path to new technologies. This solution must include a robust architectural design to withstand any and all potential failures without impacting data availability.

Over the past decade, EMC's Symmetrix product line has set the bar for all other storage arrays to follow. With the introduction of the world's most sophisticated high-end solution, the Symmetrix DMX series has raised the bar even higher. The Symmetrix DMX series offers advanced fault-tolerance capabilities and data integrity features via an innovative, new storage architecture called the Direct Matrix (DMX).

This new generation of Symmetrix is based on a revolutionary design and includes key enhancements that improve the reliability, availability, and serviceability of the new systems – making Symmetrix DMX the ideal choice for critical applications and 24x7 environments demanding uninterrupted access to information.

Highlights of the new Symmetrix DMX series include the following:

- Proactive error detection and remote support
- Nondisruptive Enginuity™ operating environment upgrades and updates
- Fully fault-tolerant design with redundant critical components and concurrent maintenance support
- Triple Modular Redundancy with Majority Voting (TMR-MV) protected Global Memory Directors with redundant region controllers
- Channel Director redundancy with end-to-end automatic channel failover and load balancing
- Internal control data path redundancy
- Two (N+1) power supply redundancy
- Fibre Channel back-end featuring redundant Disk Directors, disk channels, disk ports, and Port Bypass Cards
- Parity RAID and RAID 1 mirroring data protection options
- Advanced Communication and Environmental Control Modules

The objective of this white paper is to provide an overview of the reliability, availability, and serviceability (RAS) features within the Symmetrix DMX architecture.

Symmetrix DMX Error Detection and Remote Support

Symmetrix DMX provides the most reliable hardware design in the industry. However, all hardware is subject to the effects of aging and occasional failures. The unique methods used by Symmetrix DMX for detecting and preventing these failures in a proactive way, set it apart from all other storage solutions. The Symmetrix DMX system is designed with these data integrity features:

- Error checking, correction, and data integrity protection
- Disk error correction and error verification
- Memory error correction and error verification
- Periodic system checks

Remote support is an important and integral part of EMC customer service and support. Every Symmetrix DMX system has an integrated service processor that continuously monitors the Symmetrix environment. The service processor can communicate with the EMC Customer Support Center through a customer-supplied, direct phone line.

Through the service processor, Symmetrix DMX actively monitors all I/O operations by scanning for errors and faults. While tracking these errors during a normal operation, the Symmetrix DMX system can recognize patterns of error activity and predict a potential hard failure before it occurs. This proactive error-tracking capability can often prevent component failures by fencing off, or removing from service, a suspect component before a failure occurs. Additionally, the “call home” capabilities allow Symmetrix to automatically notify EMC Customer Support of potential issues so an EMC Product Support Engineer can handle the call and dispatch a local Customer Service Technician to install a new component without disrupting data access.

In order to provide remote support capabilities, Symmetrix DMX is configured to phone home and alert EMC Customer Support of a failure or potential failure. The appropriate, authorized EMC Product Support Engineer is able to run system diagnostics remotely for further troubleshooting and resolution. EMC products are also configured to allow in-bound dial that enables EMC Customer Support to proactively dial-in to the Symmetrix DMX system and gather necessary diagnostic data or attend to previously identified issues. Dial-in support uses the latest digital key exchange technology for strong authentication, layered application security, and a centralized support infrastructure that places calls through an encrypted tunnel between EMC Customer Support and the service processor located inside the Symmetrix DMX system.

Before any individual can initiate a call to a customer site, that person must be individually authenticated and determined to be an appropriate member of the EMC support team. And, field-based personnel who might be known to the customer must be properly associated with the specific customer’s account.

An essential part of the SymmIP architectural design for remote support is that the call to a customer’s Symmetrix service processor must originate from one of several specifically designed Remote Support Networks within EMC. Inside each of these EMC support centers, the necessary networking and security infrastructure has been built to enable both the call-EMC and call-device functions.

Symmetrix Component-Level Redundancy

All critical components are fully redundant, including Global Memory Directors, internal data paths, power supplies, battery backup, and all Fibre Channel back-end components. The following is an overview of the redundancy of each of these components.

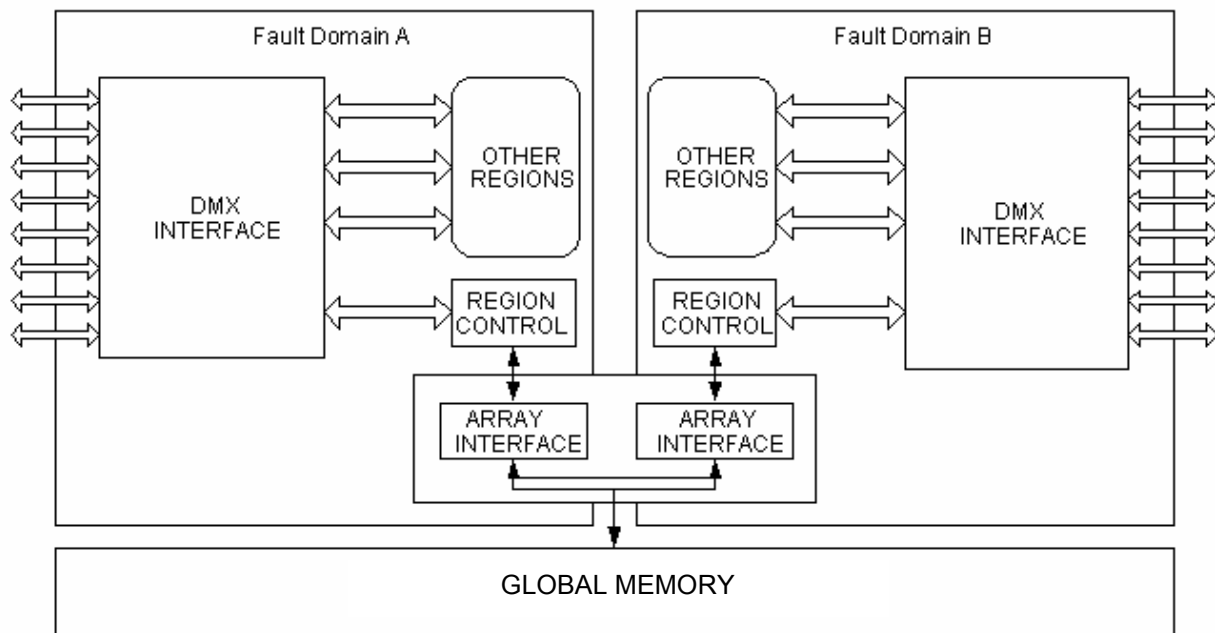
Global Memory Director Technology Overview

The Symmetrix DMX Global Memory Director technology is one of the most crucial components of the system. It delivers a high-performance, fault-tolerant, highly reliable memory design that is capable of providing unsurpassed throughput, bandwidth, and independent memory regions to support multiple concurrent operations.

All read and write operations transfer data to or from the Symmetrix DMX Global Memory Directors. And, any transfers between the host processor, Channel Directors, and Global Memory Directors are achieved at electronic speeds, which are significantly faster than those transfers (involving read/write data) directly to the physical disks. Each Symmetrix Global Memory Director accommodates four separately addressable, simultaneously accessible memory regions, which greatly reduces the probability of contention for memory access. Each memory board is directly connected to each Channel Director and Disk Director through the DMX architecture to ensure high performance and fault tolerance.

Symmetrix DMX Global Memory Directors provide component-level redundancy where each component in the system is protected against failure. The redundancy is done via both mirroring and parity schemes that can be compared to RAID parity-based protection.

The connections from a region controller to the global memory are monitored by the redundant logic within the region controller. Two region controllers are used to transfer data in the course of machine operation and guarantee that a fault will be detected. In other words, the operations are guaranteed “fault secure.” The array interface is likewise constructed using these fault-secure techniques, and provides an independent check of the region control operations. The connections from this array interface to the global memory are likewise protected from any single points of failure. If such a fault is detected (and background diagnostics periodically check to make sure such a fault is detectable) the operating environment will take action to ensure that the data within that Global Memory Director is removed via the correct operating region controller, and is “fenced off” and ready for replacement.



Symmetrix DMX Global Memory Director

The memory components themselves are also protected. Each component is just one in a RAID-like set of components. The Enginuity operating environment is used to detect and correct any possible component failures. Even if an individual component attempts to write data to the wrong address, or does not perform the write at all, it alone is at fault. This error is both detectable and correctable mathematically by the data contained in the other components that make up the region. When such errors are identified, Enginuity performs diagnostic tests to determine if the region has experienced a hard or soft failure. In the event of a hard failure, the memory component is “fenced off” as before. In the event of a soft error, such as an alpha-particle-induced single-bit error, the operating environment logs the error, corrects the data, and continues to operate with that memory. If the soft error rate exceeds a preset threshold, the operating environment will “fence” the memory for replacement.

Enginuity offers superior data protection compared to typical cache protection methods seen in other systems. Its sophisticated fault-detection and correction capabilities ensure data integrity and carry into other areas of machine design while providing fault-recovery capabilities for each path in the system.

Global Memory Director Data Integrity

Digital electronic circuits are made up of analog electronic components, and the voltage thresholds decide the difference between a digital ZERO and a digital ONE. In the absence of a fault-secure design, spurious signals can lead to a variation in voltage levels at the signaling gate that can inadvertently convert a digital ZERO into a digital ONE resulting in error conditions and a malfunctioning system.

The challenge for an enterprise storage design engineer is to ensure that a fault-tolerant design operates per specification in spite of faults, such as a defect occurring in a component. Faults can be permanent, transient, or intermittent. They can also be structured or unstructured. Faults can be detectable, correctable, or recoverable via a retry, or a retry after a failover to redundant components. Faults result in error conditions (incorrect behavior caused by the presence of fault) that the system needs to detect and correct/recover without compromising data integrity. This will ensure that the underlying system is reliable and can perform its specified task/service in spite of faults.

Highly advanced fault-tolerant techniques are mandated in systems deployed in aerospace applications, missile guidance systems, and fly-by-wire applications. These same fault-tolerant techniques are also implemented in the Symmetrix DMX architecture to provide the highest levels of availability and data integrity available in a high-end storage system. The Symmetrix DMX deploys the following techniques to ensure fault security and data integrity within the Global Memory Directors:

- Dual independent fault domains – Symmetrix Global Memory Directors are designed with two independent fault domains on the board. Each one has a separate system clock, reset, and other related logic/functions, in addition to redundant power supplies.
- Signal comparison – This technique is used at every critical stage in the system. The Symmetrix DMX system deploys identical modules across all critical paths and signal comparison techniques deployed from end-to-end (within the system) ensure that the signals at like points are the same.
- Watchdog timers – Symmetrix DMX deploys watchdog timers to detect errors. For every component, a watchdog timer is loaded (with a time value) which the operation must complete. After each operation is completed, the watchdog timer is reset for the next operation. If the operation does not complete within the designated time value for that operation, a fault is detected and the appropriate recovery measures are taken. The timers ensure that the related logic does not hang or become non-operational.
- Background diagnostics – During periods of inactivity, faults are intentionally inserted in a way that does not impact the operation of the storage array. This ensures that the fault-detection components and logic are operating correctly. The background diagnostics also provide a proactive “fire drill” feature that guarantees the fault-tolerance capabilities are functioning properly.

- Voltage-level monitoring – Symmetrix DMX deploys elements that check to see if the voltage is within margin range. If they go out of the margin range, error flags are set to indicate erroneous conditions. Associated logic also ensures error correction/recovery based on feasibility.
- Information redundancy – Symmetrix DMX deploys end-to-end information redundancy from the host (through global memory) to the disks at the back-end. At each and every stage, the information flowing through the system is protected via multiple techniques, namely:
 - Parity Check – Multiple parity bits are associated with every I/O flowing through the Symmetrix DMX. The parity bits are checked for validity at every stage to ensure the information is interchanged.
 - Checksums and CRC – A checksum and CRC (Cyclical Redundancy Check) is generated for every block of information and is appended at the end. The checksum and CRC are then recalculated and verified for accuracy at each stage the information block is moved.
 - Single Nibble Correction Double Nibble Detection (SNCDND) techniques – Symmetrix Global Memory deploys a state-of-the art technique for information protection that results in making every chip on the Global Memory Director redundant. Traditional cache-based systems usually provide for eight bits of parity information to support single-bit error correction and detection in a 64-bit long word. SNCDND is achieved by internally generating 16 bits of error correction coding parity information. This enables the system to correct up to four-bit errors associated with a 64-bit long word. Up to eight-bit errors can be detected using Symmetrix Global Memory Directors.

Symmetrix global memory can interleave 64 bits of information plus 16 bits of parity information (total 80 bits) across 20 memory chips on the memory board. This results in each memory chip storing only a nibble (the equivalent of four consecutive bits) of information corresponding to a word. Thus, a chip-level error will disable access only to the nibble stored on that faulty chip, whereas the SNCDND technique enables regeneration of data from the faulty chip. This leads to chip-level redundancy making sure every chip on the memory board is redundant.

- Instruction monitoring – At every stage, regardless of where the Enginuity operating environment is executing an instruction, errors are detected and corrected in the instruction – whether the error is in either the operation code or in the operand. An example of an Enginuity-executed instruction is an RMW (read-modify-write) instruction from the Channel Director to the Global Memory Director. This ensures error-free instruction execution.
- Address containment – The Symmetrix DMX system was designed with a philosophy to check each and every activity. Following this philosophy, the address component of every command/instruction is checked for validity in order to ensure that the operation undertaken is at the correct address location within the Symmetrix Global Memory Director, and that erroneous addresses are not read from an incorrect location or written to an unexpected location.
- Loop-back testing – This function verifies that signals leaving one point of a circuit arrive at their destination unchanged. It is achieved by providing an independent return path for the signal back to its source, and by comparing the outgoing and return signals to ensure equivalence.

Triple Modular Redundancy with Majority Voting

The Symmetrix DMX Global Memory Director deploys a state-of-the-art data integrity technique called Triple Modular Redundancy with Majority Voting (TMR-MV) across its two independent fault domains.

Three identical modules are designed to allow a comparison of multiple signals. If modules produce identical signals it is assumed that they are fault-free. The underlying idea behind triple modular redundancy with majority voting across two independent fault domains is fairly simple. It deploys a series of techniques to ensure fault avoidance (avoid introduction of faults by deploying redundant design), fault removal (use formal methods to find and eradicate faults), and fault tolerance (operational in spite of faults) by *detecting* (determining that a fault has occurred), *diagnosing* (determining what caused the fault or exactly which subsystem or component is faulty),

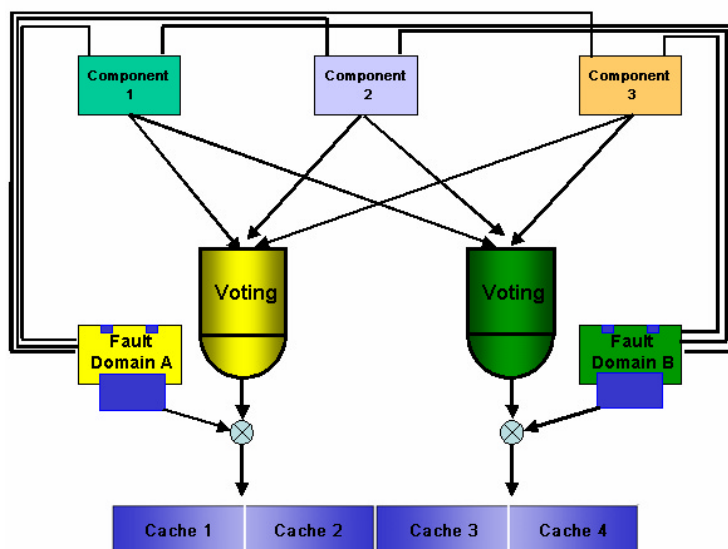
confining (preventing the propagation of faults from their origin in a subsystem to other subsystems), *masking* (ensuring that only correct values get passed to the system boundary in spite of a failed component), *compensating* (providing a response to compensate for output of the faulty subsystem), and *majority voting* and *recovering* (the process in which faults are removed from a system).

In a well-designed system, faults are contained before they propagate to the extent that the system is free from faults to ensure the highest levels of fault security and system dependability. Redundant paths to the Global Memory Directors are on two independent fault domains. The Symmetrix DMX deploys TMR-MV using a set of three identical modules, created from components from three different vendors, to eliminate any unstructured faults and determine the set's output to match the majority of modules. The three modules as depicted in the illustration below receive identical inputs and thus should produce identical outputs. A pair of redundant voters placed across two independent fault domains compares the three outputs.

Dual voters evaluate the output from the three redundant modules, and in the absence of any module faults, the voters will simply present the unanimous view of the modules as its output. On a module failure, the voters will mask the fault by adopting the majority view as the correct output. Dual voters across two independent fault domains eliminate the possibility of voter failure.

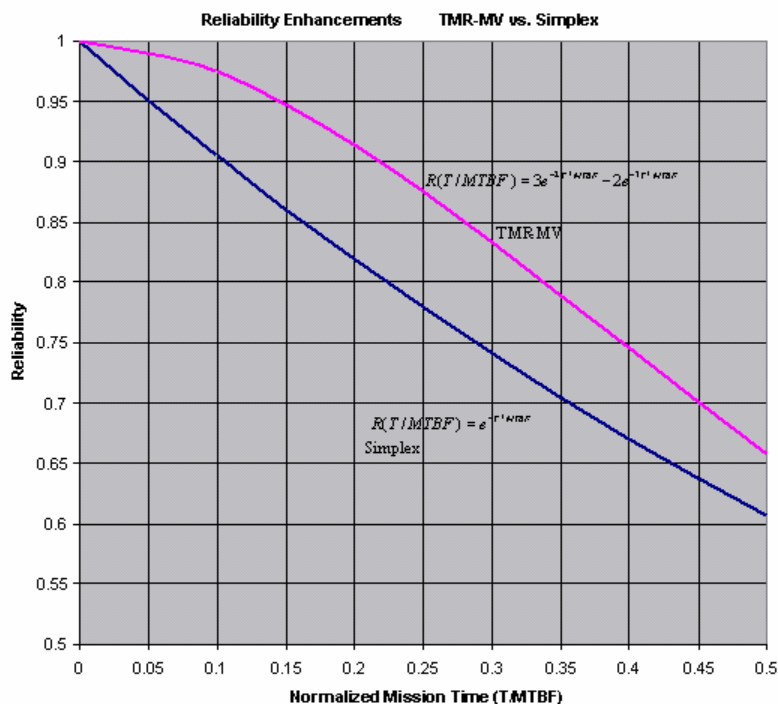
The disagreement detector identifies the failed module and removes it from the voting arrangement. Its place will then be taken by one of the spares.

As depicted in the graph below, TMR-MV techniques increase reliability during the earlier (read useful) part of the system life.



In addition to monitoring recoverable conditions during normal access, all locations in global memory are periodically read and rewritten to detect any increase in single-bit errors. This memory-scrubbing technique maintains a record of errors for each memory segment. If the predetermined error threshold is reached in a certain segment, the service processor generates a call-home for immediate attention. Constant memory scrubbing to detect

and correct single-nibble and non-consecutive double-nibble errors dramatically reduces the potential for multiple or hard errors. Should multiple errors be detected during the scrubbing process, it is considered a permanent error. This error is corrected and the segment is immediately "fenced" (removed from service) – while the segment's contents are moved to another area in global memory. The service processor's call-home capability alerts EMC Global Service Call Centers whenever an unacceptable level of errors has been detected, and a non-disruptive Global Memory Director replacement is ordered. Customer service is immediately notified of all call-home alerts, and a Customer Engineer (CE) can be dispatched with the appropriate parts for speedy repair. Even in cases where errors are occurring and are easily corrected, a call home is executed if the errors exceed a set level.



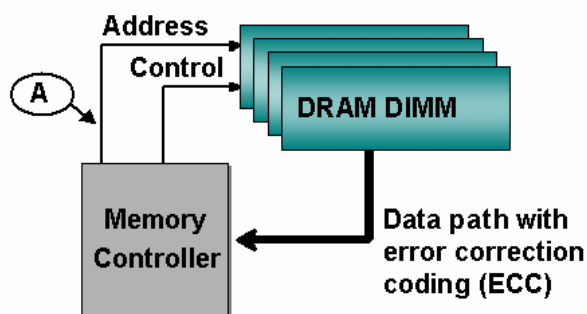
Symmetrix Global Memory Technology Compared to Mirrored Cache

In early design testing, EMC discovered that cache mirroring is an inefficient way of creating redundancy for failsafe operations. Cache mirroring results in two cache operations to complete a read I/O and five cache operations to complete a write I/O. In addition to the added overhead, mirroring requires up to 50 percent of useable cache to provide a single mirror. EMC analysis reveals that memory boards themselves do not fail, unlike the chips on the memory boards. This led to the development of a new architecture that ensures every chip on the memory board is redundant – and eliminates any single point of failure on the Global Memory Directors. This also allows for higher utilization of available global memory resources resulting in higher system throughput.

In a typical mirrored cache, the cache system replicates the data to provide another copy in case the primary components fail. The primary cache is used for all “real” work, and the secondary cache is used as a backup. The problem, however, is that this implementation does not offer comprehensive protection by simply providing a mirrored copy of cached data. There are several factors that could potentially impact data integrity.

First, consider that when the number of components doubles, the chance of having a failure in the total set of components doubles. In addition, if a fault occurs in the secondary cache, it is not detected because read operations only take place from primary cache. In the event of a primary cache failure, there is no guarantee that the data in the secondary cache has not also experienced a fault or been corrupted and was undetected. Thus, if someone removes the primary memory, the “good” copy of the data, then only the potentially “bad” copies of the data remain in the system. When and if these are found, the system no longer has any failover. At this point, there is no protection at all for the data until the defective memory is replaced.

Another potential point of exposure is that there is no protection on the address given to the cache array. For example, an open pin at the point labeled “A” in the diagram will lead to data being written to the wrong address. This not only makes the data write location incorrect, but it also contaminates a completely separate address.



Since reads of the cache only occur from the primary cache, the fault goes undetected by the system. Similar faults can occur in control signals to the memories, where the cache may never actually write the data and stale data is returned to the host system on subsequent reads.

The granularity of a replaceable unit should not be the concern, but rather that all reasonable cases of fault and serviceability have been addressed. EMC continues to use field-proven techniques, such as component-level redundancy, to ensure the best value for the customer.

Symmetrix DMX Channel Directors

Channel Director Redundancy

Channel Director redundancy configures multiple connections from the host servers (direct connect) or Fibre Channel switch (SAN Connect) to the Symmetrix DMX system. With SAN connectivity via Fibre Channel switches, each Symmetrix DMX port can support multiple host attachments – enabling storage consolidation across a large number of host platforms. The multiple connections are distributed across separate Channel Directors to ensure uninterrupted access in the event of a channel failure. A minimum of two connections per server or SAN is required to provide full redundancy. The following is a summary of the number of host/SAN connections available across the Symmetrix DMX series:

Model	Number Channel Directors	Total Fibre Channel Ports
Symmetrix DMX800	2	8-16
Symmetrix DMX1000	2-6	16-48
Symmetrix DMX2000	2-12	16-96
Symmetrix DMX3000	2-8	16-64

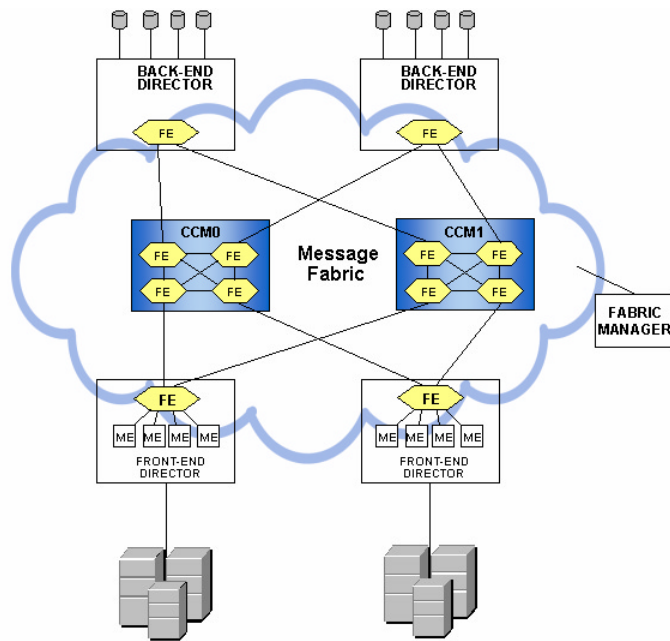
Channel-failover functionality is required to automate the failover and fail-back process, in order to avoid an interruption to data access. Without this functionality, a path failure due to a problem with the host bus adapter, Fibre Channel switch, Fibre cable, or Channel Director would create the potential for the application to go down. EMC's PowerPath[®] software supports end-to-end (or switch-to-storage) channel-failover functionality for open systems and Windows environments. PowerPath provides automatic, non-disruptive failover capabilities by redirecting I/Os from a failed channel to all remaining channels. PowerPath can automatically detect when a path has failed, and sends an error log to the host notifying it that there is an inactive path. PowerPath then fails over the existing I/O request to another active path to maintain data access and application availability – offering intelligent load-balancing capabilities to optimize performance and minimize bottlenecks. All of these functions occur transparently to the host, so the application is not stopped and data is continuously available.

Once the failed path is fixed or repaired, PowerPath automatically detects activity, brings the path back into operation, and automatically sends I/O requests to the now active path. The major benefits of the path-failover capability is the ability to automatically detect a channel failure, transparently redirect an I/O to other available channels, and easily service and restore the failed channel without interrupting applications.

Internal Control Data Path Redundancy

Each Director has two paths to connect the Symmetrix DMX message matrix between each Director. Independent and redundant fault domains provide a plug-in Communication Control Module (CCM). These fault domains within the message matrix fabric are intelligent at the hardware level and load-balance traffic between each domain. In the event that a fault occurs within one of the domains, the traffic is automatically routed by the hardware to the available domain, and the system is alerted of the faulty CCM domain for replacement. The connection is made via the plug-in CCM.

There are two CCMs per system (see diagram below). If we look at a fully configured Symmetrix DMX3000, for example, 16 Directors are configured with 32 point-to-point serial connections. This means 16 connections are routed to the first CCM and the other 16 redundant connections are routed to the second CCM. The message engine (ME) component of the message matrix provides protection against rogue or runaway nodes. This is done at each message engine through error correction coding configurations that allows the nodes to receive messages. Packet CRC and sequence numbers protect each message packet transmitted. Internal data paths with the message engine and fabric element are fully parity-protected.



Power Supply Redundancy

Power supply redundancy in a Symmetrix DMX system is achieved through the combination of redundant techniques. In a typical application, an N+1 power supply system will be used. An N+1 system is comprised of multiple replications of small power supplies integrated together to satisfy the total power requirement of the system. If any single power supply fails, the remaining units can still maintain power to the system. N+1 also allows the failed power supply to be replaced without disruption to the system. *Note: Symmetrix DMX800 and DMX1000 systems implement an N+1 power supply.*

While an N+1 power supply design is highly reliable, there are still possible conditions that make N+1 redundancy susceptible to failure modes and contribute to a complete power system failure. For example, if a transient fault induces a disturbance on the output and propagates to the rest of the system, or a power supply output shorts and drives down the output for the rest of the power system, N+1 and dual-system solutions offer no protection.

In the Symmetrix DMX2000 and DMX3000, 2(N+1) redundancy is provided to eliminate these potential points of exposure. 2(N+1) power redundancy is based on dual N+1 power supplies for a higher level of fault tolerance and availability than standard N+1 power supply implementations. In addition to providing higher levels of availability, 2(N+1) redundancy is a cost-effective approach by providing minimal excess power capacity above what is required to maintain system power in the event of a single power supply failure.

Single-phase or three-phase main and auxiliary AC power is provided to each of the two Power Line Input Modules (PLIMs). These modules contain line circuit breakers and lower power for the EPI (Emergency Power Interrupt) and EPO (Emergency Power Off) functions. The PLIMs then feed main and auxiliary AC power to each of the DC power converters. Two independent battery backup units support the DC power supplies – one for each zone. Critical boards receive power from each zone. The Symmetrix DMX has been designed to continue to operate in the case of a Power Line Input Module (PLIM) failure or replacement.

Battery Backup

The Battery Backup Unit (BBU) ensures data integrity during power outages. In the event of a power failure, the BBU allows the system to continue operation for a several minute interval. If power is not restored within that interval, the system is brought offline and powered down. The BBU delivers sufficient power to safely de-stage all cached data onto the disk drives, where it is therefore preserved indefinitely. The BBU battery capacity is sufficient to ride out two consecutive outages and allow the system to safely withstand multiple power interruptions or failures. To ensure data integrity and proper system operation in the event of more than two consecutive power failures, the BBUs must be sufficiently charged in order to restart the system. Approximately 12 hours of recharge

time is required to fully charge the BBU. The BBU also contains a self-diagnostic controller to ensure availability and correct operation by monitoring the battery condition, charger states, and by providing status updates to the environmental control module.

Fibre Channel Back-End Redundancy

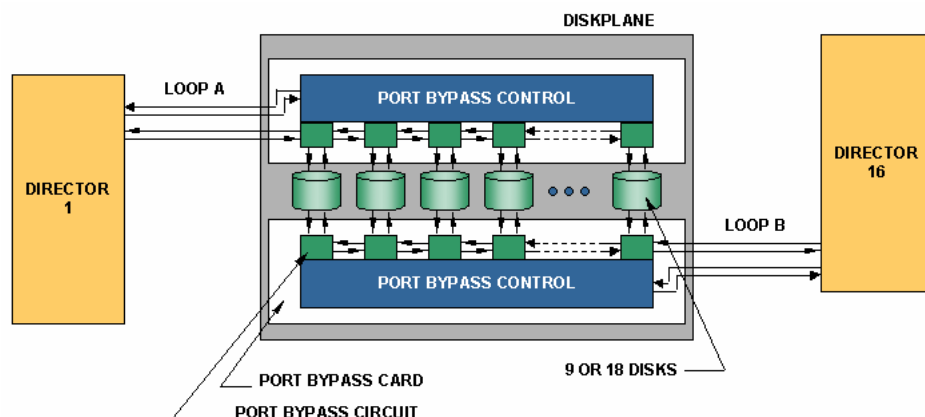
The Symmetrix DMX architecture incorporates a two gigabit-per-second (2Gb/s) Fibre Channel back-end design to ensure high performance and full redundancy. The Fibre Channel back-end provides redundant paths to the data stored on the disk drives to provide non-stop access to information, even in the event of a component failure or replacement. The following is an overview of the redundant components of the Symmetrix DMX back-end:

- **Redundant Disk Directors** – A pair of Disk Directors is used to access each disk drive. Each Disk Director pair is connected to a separate physical path to the drive and is responsible for moving data between the global memory and the disks. Since each pair is connected to the global memory via redundant internal paths, this eliminates any possible single points of failure.
- **Redundant cable paths** – A separate, independent cable assembly – carrying the Fibre Channel loop, control bus, and other protocol signals – attaches each disk director to its associated set of Fibre Channel disks.
- **Redundant disk ports** – Each Fibre Channel disk drive has two fully independent Fibre Channel ports that are designed to connect to two separate loops. In the Symmetrix DMX architecture, these two ports are connected to a different Port Bypass card, which in turn is connected via separate cables to two different Disk Directors. The two pathways are completely independent of one another.
- **Redundant Port Bypass Cards** – Each Disk Director in a redundant pair connects to a separate Port Bypass Card.

Symmetrix DMX systems employ an arbitrated loop design that contains monitoring and control hardware and software to maximize the performance and availability of each loop. The loop is connected in a star-hub topology, with gated hub ports and a bypass switch that allows individual loop segments (connecting to a Fibre Channel disk drive) to be dynamically inserted or removed. The loop initiator is a Symmetrix Fibre Channel Disk Director, which feeds data and controls the hub. All of the monitoring and control logic is contained on the Port Bypass card, which also contains the bypass/hub logic for the loop. To ensure the highest level of availability, the communications link to the monitoring and control functions is not carried by the Fibre Channel loop, but rather is implemented through a separate path. If the Fibre Channel loop is not operating for some reason, the Disk Director can access the Port Bypass Card through the control bus and reconfigure the loop into a working state.

The Port Bypass Card is the heart of the Symmetrix disk Fibre Channel loop and constantly monitors the loop for Fibre Channel layer 0/1 signal connection and protocol errors. If errors are detected, the controlling software program is notified, and if necessary the bypass circuits may be used to switch out the failing loop segment(s), thereby allowing the remaining devices on the loop to operate unaffected. Each disk enclosure houses up to 36 individual disk drives, and contains a pair of Port Bypass Cards. Each Port Bypass Card connects one of the ports from the 36 drives to a Disk Director. For each disk connection, a bypass circuit is provided and allows that disk connection to be shunted. This only occurs when the drive is not installed, the drive was just powered on, the drive was reset, or the drive is malfunctioning. In addition to the bypass circuits themselves, the card contains realtime monitors embedded within the loop that can determine if bad data is being generated on the loop. These monitors report status to the loop control software and can be useful in determining which parts of the loop are being affected by potential hardware failures. The Port Bypass Card also has the ability to control power to the drives, and it maintains basic configuration information so that if the loop configuration is changed (e.g., a new disk assembly is inserted), it reports this to the control software program to take the appropriate action. The Port Bypass Card is located within the disk enclosure, allowing replacement in the case that an isolation device fails. Each port of a disk is connected to a different Port Bypass Card, so access to the data is maintained even during service operations.

The following is an illustration of the back-end redundancy of the Symmetrix DMX system.



Since the disk media is a single, shared resource, this level of redundancy is not enough to protect the system from a disk media failure. In order to defend against these types of failures, some form of data redundancy in the form of RAID is necessary. RAID protection adds information to the other disks in the system to allow the recreation of any other disk's data in the event that a disk becomes unavailable. Combining the two techniques, as is done in Symmetrix DMX, yields a very robust result that ensures the highest level of access and reliability to the customer's data.

Data Integrity for Fibre Channel Arbitrated Loops

In order to ensure that the data transmitted by the Disk Director is the same as the data received by the disk device (and vice-versa), a multi-layer protection strategy is employed. At the highest level, data is transmitted between the global memory and the disk, and a connection referred to as an exchange is established. Logical parameters that identify this connection are initialized, and all of the information transferred as part of this exchange is identified with these parameters.

In addition, as the data in global memory is encapsulated into Fibre Channel frames, it is marked with a sequence number that identifies the order of the frame within the exchange. Thus, a cyclic redundancy checksum is computed over the data and appended to the data frame. After validating the exchange parameters, the receiving port will recompute the checksum for the data received, and compare it to the transmitted checksum determining if there are any bytes in error. It will also check the sequence number to insure that the received frame is the next one expected for that exchange.

The next level of protection is designed at the transmission layer of the Fibre Channel protocol. The data frame, with the additional information appended, is translated into special 10-bit data transmission symbols. These symbols are chosen for their ability to increase transmitted data reliability, through either a fiber optic or copper cable. The transmission codes also help to assist in detecting transmission errors.

Finally, once the data has been received at the disk drive (for writes), and decoded from the transmission codes, it will then be further recoded using a very powerful linear block coding algorithm before being stored on the disk's magnetic surface. This code has the ability to not only detect errors, but to also correct them on the fly – ensuring that the data read from the drive at a later time will be read perfectly.

Symmetrix DMX Data Protection Options

The Symmetrix data protection options ensure uninterrupted access to data in the event of a disk failure. While disk drive technology has improved significantly over the last several years to ensure continuous data availability, the following disk-protection options are available to provide a higher level of data protection, recoverability, and availability. The options listed below can be purchased separately and implemented into the Symmetrix operation.

Data Protection Option	Description
Mirroring (RAID Level 1)	Provides the highest level of performance and availability for all mission-critical and business-critical applications by maintaining a duplicate copy of each volume on two disk devices.
Parity RAID	Provides higher performance and availability than RAID Level 5 offerings for online business applications. A Parity RAID (3+1) group consists of three data volumes to one parity volume. A Parity RAID (7+1) group consists of seven data volumes to one parity volume.
Symmetrix Remote Data Facility (SRDF®)	Provides an information protection/business continuity solution by maintaining a mirror image of data between multiple Symmetrix systems that can be in physically separate locations.
Dynamic Sparing	Increases data availability by copying the data on a failing disk device to a spare disk device until the original disk device is replaced. Dynamic Sparing is used as additional protection for mirrored, Parity RAID, and SRDF volumes.

Additional technical detail on Symmetrix DMX data protection options is available in a separate Engineering White Paper, *Symmetrix Data Protection Options*.

Symmetrix Component-Level Serviceability

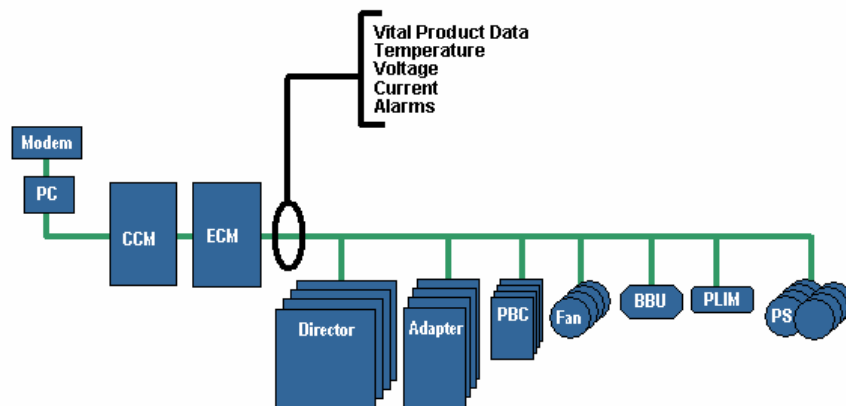
The Symmetrix DMX system implements a modular design (with a low parts count used to minimize the number of failure points) to improve serviceability by allowing non-disruptive component replacement should a failure occur. The Symmetrix DMX system features concurrent maintenance of all major components, including:

- Global Memory Director boards
- Channel Director boards
- Disk Director boards
- Disk devices
- Cooling fan modules
- Environmental Control Module
- Communications Control Module
- Power supplies
- Batteries
- Services Processor

The Symmetrix DMX system provides full component-level redundancy to protect against a component failure and ensure continuous and uninterrupted access to information. The Symmetrix DMX Director boards are mechanically keyed to prevent insertion to invalid slots. And, a “disable interface” feature ensures that during a power-up, the link interfaces between any new Channel or Disk Director and its corresponding Global Memory Directors are disabled. Only after safeguards have been satisfied, and the Director has been properly replaced, will a new board be activated and brought online.

Symmetrix DMX Control and Environmental Matrix

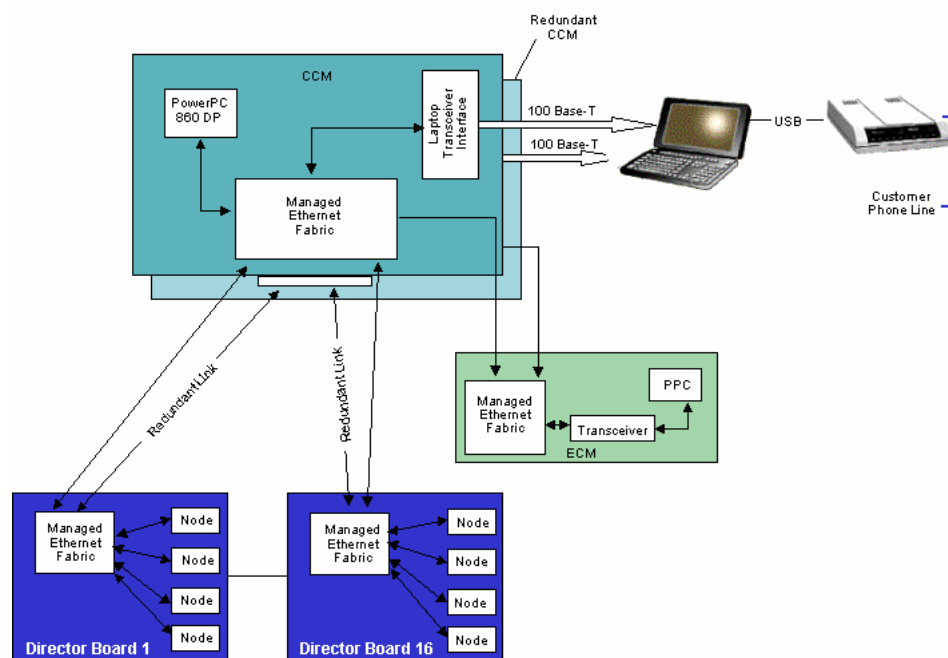
Remote Support is an important and integral part of EMC customer service and support. Every Symmetrix DMX unit has an integrated service processor that continuously monitors the Symmetrix environment. The service processor can communicate with the EMC Customer Support Center through a customer-supplied direct phone line. The service processor automatically dials the Customer Support Center when the Symmetrix DMX system detects a component failure or environmental violation. An EMC Product Support Engineer (PSE) at the Customer Support Center can also run diagnostics remotely via the service processor to determine the source of a problem and potentially resolve it before the problem becomes critical. Within the DMX control e-net matrix are the Environmental Control Module and Communication Control Module. These two modules provide the low-level system-wide communications for running application software, monitoring, and system diagnostics from the service processor.



The Communication Control Module and Environmental Control Module feature non-disruptive replacement by utilizing redundant module functions with the DMX architecture in the event of a module failure.

Communication Control Module

The Communication Control Module's (CCM) primary function is to act as a communications agent between the laptop service processor and the embedded processing nodes within the system. The figure below illustrates the Ethernet fabric interconnecting the processing nodes to the service processor. External connections to the service processor provide dial-home capabilities for remote monitoring and diagnostics.

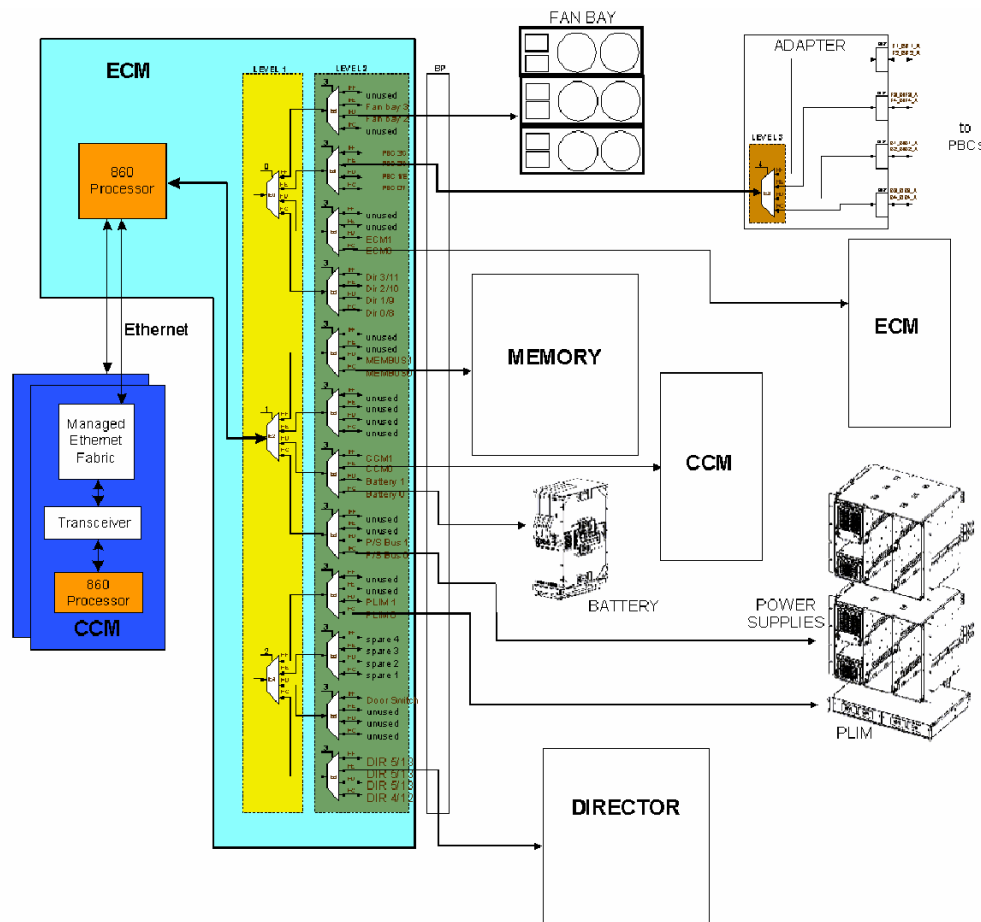


Software-driven fail-over mechanism exists such that the Ethernet fabric remains intact should a CCM become temporarily unavailable. As a backup to the Ethernet fabric, the CCM also contains RS232 multiplexing logic to allow for an alternate means of serial communication to embedded processing nodes within the DMX. The primary service processor can use its RS232 serial port to communicate to individual nodes should the use of the Ethernet fabric become unavailable. The CCM also has the ability to issue remote commands to the Channel Directors, Global Memory Directors, Disk Directors, and Environmental Control Module (ECM). These commands can be issued from the laptop service processor, or remotely by the EMC Customer Support Center, providing a rich set of intelligent serviceability functions. Also within the CCM is the top-level fabric for the message matrix communications system. This is a high-speed communications fabric within the DMX that allows for fast, reliable messaging between compute nodes.

Environmental Control Module

The Environmental Control Module (ECM) monitors and logs environmental events across all critical components and reports any operational problems. The ECM also allows revision numbers and date codes on individual components to be searched for by the EMC Customer Product Lab. Critical components include Channel and Disk Directors, Global Memory Directors, Port Bypass Cards, power supplies, fans, and switches. The ECM is capable of monitoring each component's local voltages, ensuring optimum power delivery. Temperature of all Channel, Global Memory, and Disk Directors and is also continuously monitored. Fan speed is also varied as required to maintain optimum operating temperature. The AC Power Main is checked for AC failures, transfer to auxiliary, DC failures, current-sharing between DC supplies DC output voltage, specific notification of over-voltage condition, the current from each DC supply, and voltage drops across major connectors. Ailing components can be detected and replaced before a failure occurs.

Additionally, the ECM centralizes system FRU (Field Replaceable Unit) management by acquiring all VPD (Vital Product Data), status, and control of the FRUs via the I²C bus or an Ethernet interface. It also monitors temperature, voltage, and fan speed remotely via Programmable Interface Controllers (PICs).



Non-disruptive Operations

Symmetrix Enginuity is the intelligent operating environment for the Symmetrix DMX systems. Enginuity manages and ensures the optimal flow and integrity of information through the different hardware components of the Symmetrix DMX system. Enginuity manages all Symmetrix operations, from monitoring and optimizing internal data flow, to ensuring the fastest response to the users' requests for information, to protecting and replicating data.

Enginuity provides the following services for the Symmetrix DMX systems:

- Manages system resources to intelligently optimize performance across a wide range of I/O requirements
- Ensures system availability through advanced fault-monitoring, detection, and correction capabilities while providing concurrent maintenance and serviceability features
- Interrupts and prioritizes tasks from microprocessors to ensure that fencing off failed areas takes precedence over other operations
- Offers the foundation for specific software features available through EMC's disaster recovery, business continuity, and storage management software
- Provides functional services for both Symmetrix-based functionality and for a large suite of EMC storage application software
- Defines priority of each task including basic system maintenance, I/O processing, and application processing including: EMC ControlCenter™, SRDF, TimeFinder™, and EMC ControlCenter Symmetrix Optimizer
- Provides uniform access through industry-standard APIs for internal calls, and provides an external interface to allow integration with other software providers and ISVs

Enginuity upgrades and interim updates are performed at the customer site by the EMC Customer Engineer (CE), and provide enhancements to performance algorithms, error recovery, reporting techniques, diagnostics, and code fixes.

Non-disruptive Enginuity upgrades from one version to the next, as well as interim updates (loads), are available for Symmetrix DMX and take advantage of its multiprocessing and redundant architecture. Release levels can be also be loaded online without interruption to data availability.

During an online upgrade, the EMC Customer Engineer (CE) downloads the newest version of Enginuity to the Symmetrix DMX service processor. The new operating environment code loads into the EEPROM areas within the Channel and Disk Directors, but remains idle until it is made operational. The Symmetrix system does not require manual intervention to perform this function. All Channel and Disk Directors remain online to the host, thus maintaining uninterrupted application access. The Symmetrix DMX system will load the new version of Enginuity at selected windows of opportunity within each Director hardware resource until all Directors have been loaded.

Once the new version of Enginuity is loaded, internal processing is synchronized and the new code becomes operational. During an online load within a code family (or interim upgrade), the full version is loaded and consists of the same base code plus any additional patches. The ability to perform non-disruptive Enginuity upgrades is a critical feature to providing uncompromising levels of system availability and data access.

In addition to non-disruptive microcode upgrades, Enginuity also supports a wide range of non-disruptive operations allowing Symmetrix DMX to support mission-critical environments for applications that require uninterrupted access to information and uncompromising service levels. The following are examples non-disruptive operations supported by Enginuity:

- Hardware upgrades, such as adding additional disk units, Channel Directors, and Global Memory Directors
- Configuration updates and modifications, such as reconfiguring or moving existing storage resources to support new host platforms
- Modifying data protection options, such as converting RAID 1-protected devices to Parity RAID-protected devices, and vice versa
- Configuration updates and modification for TimeFinder and SRDF volumes, such as the ability to dynamically add or remove SRDF groups, or convert BCV capacity into standard devices

Enginuity 5670 Operating Environment Overview

Symmetrix Enginuity 5670 is a landmark Enginuity release and provides the industry's leading non-disruptive upgrade capabilities and features. With this enhancement, Symmetrix DMX series systems are the only high-end storage arrays to provide online, non-disruptive upgrades without requiring additional hardware and/or software on the hosts. Symmetrix DMX systems can be upgraded without interruption to most hosts or applications. The only exception is when supporting Windows boot devices. No vendor currently supports non-disruptive upgrades for Windows hosts that are in the process of booting from the array. This is due to the limitations of Windows, Windows NT, Windows 2000, and Windows XP. EMC is working with Microsoft to address these limitations in future products.

In addition, the Symmetrix DMX does not require multiple connections to the host to support non-disruptive upgrades, unlike other competitive products, such as the Hitachi 9980V, where according to the published documentation, non-disruptive code loads are supported only if hosts have dual connections to the array and are running path fail-over software on those hosts.

Enginuity 5670 Non-Disruptive Operations

Enginuity 5670 delivers non-disruptive operations. Customers can non-disruptively reconfigure and re-deploy their storage resources. Redeployment of storage resources involves freeing up the disk space currently under use and then re-using that free space to create new devices for re-deployment elsewhere. Enginuity 5670 supports the ability to non-disruptively delete devices no longer required to free up disk space used by devices that are candidates for redeployment.

Enginuity 5670 also allows customers to reconfigure and reassign allocated capacity non-disruptively on a Symmetrix DMX system, and make that capacity (physical or logical disk space) accessible to any combination of mainframe and open systems applications. Enginuity 5670 also enables customers to convert a mainframe (CKD) device into an open systems (FBA) device, and vice-versa, online without business disruption. It requires that the EMC Enterprise Storage Platform (ESP) software has been pre-installed on the array so that the Symmetrix DMX has been configured to accept mainframe (CKD) devices.

Inter-Family Non-Disruptive Upgrades (NDU) (5669 to 5670)

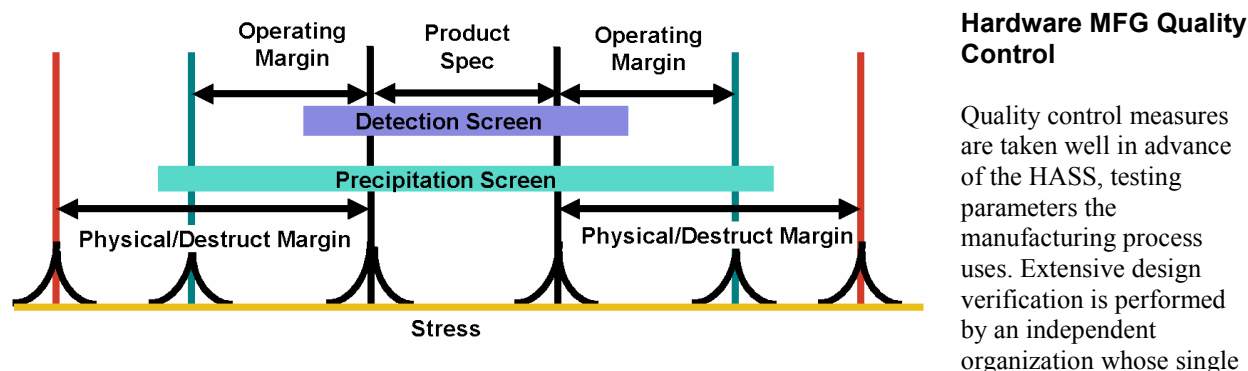
Enginuity 5670 provides non disruptive operation and delivers the highest levels of availability in the industry. Enginuity 5670 enables customers to upgrade existing Enginuity 5669 versions across code families without business disruption. All new features to be incorporated into future Enginuity versions will assure and enable non-disruptive upgrades to the future Enginuity versions. Inter-family NDU occurs quickly and seamlessly with no business disruption and is comparable to current 5669 intra-family upgrades which do not cause disruption to system operations.

Hardware Quality Assurance

World-class quality assurance (QA) practices and techniques are used to prepare, test, and validate all Symmetrix DMX systems. Accelerated life and stress testing are performed on *all* electro-mechanical assemblies ensuring customers the highest level of reliability and successful plug-and-play. The testing processes that all Symmetrix DMX systems must complete and pass before they are shipped to customers are outlined in this section.

Highly Accelerated Life Testing (HALT) is a design-margins verification process and is performed early in product development. The HALT test is used to screen all components, detecting design flaws, manufacturer defects, and identifies components with unacceptably low or narrow margins of operation. The HALT process is used to simulate and accelerate major failure conditions. Each failure is tracked and plotted to determine if there is an increasing failure rate. Repeated cycles are then further analyzed to determine the root cause. This can categorize the failure to determine if overstress or a screen-precipitated latent defect is present. Data gathered from the HALT process provides a wealth of component-level data and statistics to improve product reliability. EMC Manufacturing, EMC Design Engineering, and EMC Supply Base Management divisions utilize the data extensively. The data is also shared with the component vendors to help improve their designs and overall reliability.

An immediate use for HALT data is in prescribing stress-screening values in the manufacturing process. This process, referred to as Highly Accelerated Stress Screening (HASS), allows EMC Manufacturing to detect and precipitate latent manufacturing defects quickly. The parameters chosen for HASS are derived from the HALT. A “proof of screen” qualifies HASS testing where several numerical methods are used: **Weibull Analysis (Infancy, Wearout)**, **Six Sigma Methods For Pass/Fail Criteria**, **Simple A/B Population (Normal Statistics)**, **Pareto** and **Arrhenius Modeling**.



intention is finding potential problems early in the design phase. Primarily the HWQA group scrubs and audits electrical designs to establish and verify that operation of the design falls completely within each constituent component’s specified margins.

The design-phase verification tests by HWQA are comprehensive, including *timing verification* (that different signals are within the timing specifications of the different components), *signal integrity* (verifying that the different signals follow the components specifications), and *negative testing* (fault injection) used to ensure system availability and data integrity remain despite any component failure.

Furthermore, software and hardware co-verification tests changes in a live runtime environment through a separate software regression cycle. Finally, System Level Testing is run on a large sample of systems, with different configurations through heat room, cool room, power cycling, voltage margining, and vibration testing. The goal is to expose system-type issues, compatibility, and functionality issues.

Conclusion

Over the past decade, EMC’s Symmetrix product line has set a high standard for all other storage arrays to follow. With the introduction of the world’s most sophisticated high-end solution, the Symmetrix DMX series has raised that standard even higher.

The Symmetrix DMX series offers advanced fault-tolerant capabilities and data integrity features via an innovative, storage architecture. The Symmetrix DMX architecture was developed to help customers address their high-end storage challenges for the next decade. By delivering this architecture in a revolutionary product family, the Symmetrix DMX series, EMC has elevated customer expectations for high-end storage performance, availability, functionality, and economics to a new level. With Symmetrix DMX, EMC has significantly improved the economics of high-end storage, enabling customers to:

- Deploy high-end functionality across more of their infrastructures
- Keep pace with growing high-end storage demands
- *Drive greater business value through consolidation, performance, improved service levels, and faster time-to-market*

This generation of Symmetrix is based on a revolutionary design and includes key enhancements that improve the reliability, availability and serviceability of the systems – making Symmetrix DMX the ideal choice for critical applications and 24x7 environments demanding uninterrupted access to information.