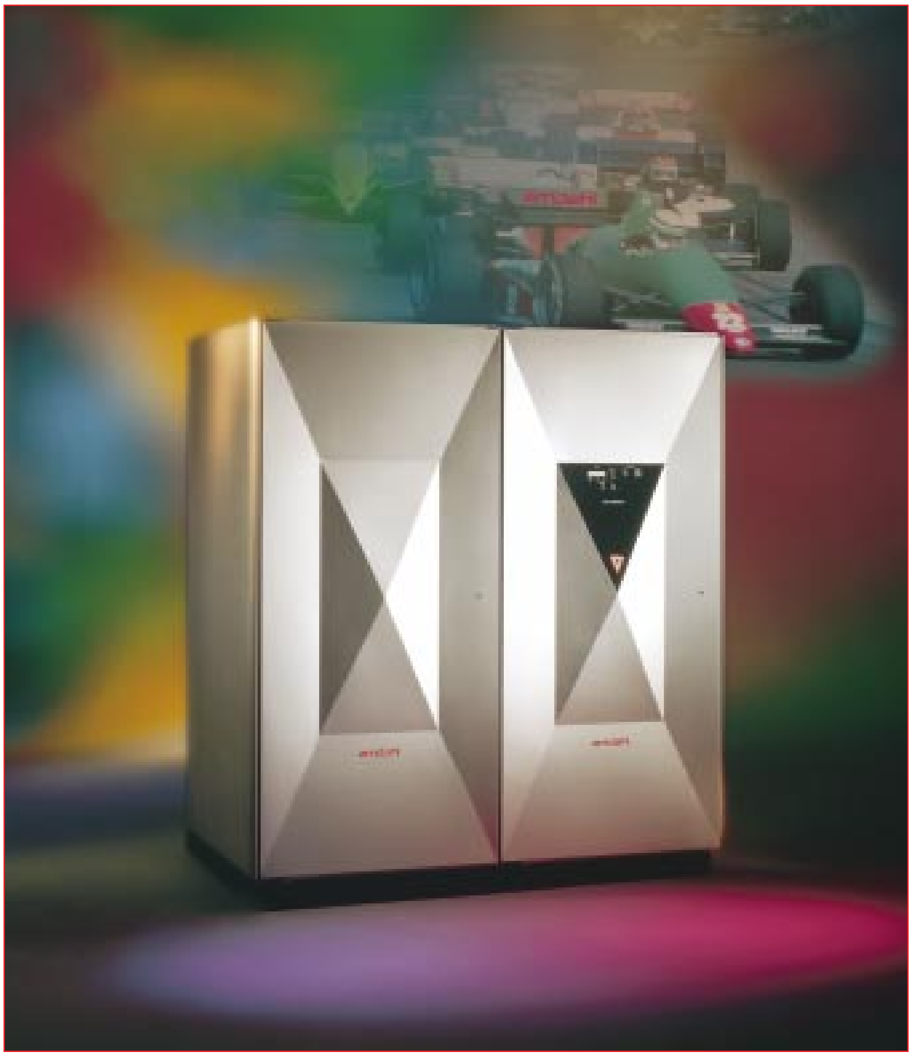
Millennium 400/500 Series Servers

General Information Guide



Abstract

Millennium Servers: General Information Guide is an introduction to the features, characteristics, functions, and components of the Millennium server family, with additional information about both the Millennium Global Server and the Millennium Coupling Server. It is intended for data processing technical management, data center planners, systems programmers, operations management, and senior computer operators. Users of the manual should be familiar with *IBM Enterprise Systems Architecture/390 Principles of Operation.*

Millennium 400/500 Series Servers

General Information Guide

Revision Notice

This is the second edition.

For Further Information

If you wish to obtain further information about the Amdahl Millennium servers discussed in this publication, contact your Amdahl representative, or write:

Amdahl Corporation
S/390 Processor Marketing
P.O. Box 3470
1250 East Arques Avenue
Sunnyvale, CA 94088-3470

Table of Contents

Figures

Tables

1 Introduction

Millennium™ Global Servers are versatile, high-performance computing systems that can be used in traditional System/390, Parallel Sysplex, and Open Systems environments. Using complementary metal oxide semiconductor (CMOS) technology, Millennium servers provide the performance and availability that mission-critical applications need in a compact system that is very economical to operate.

The Millennium family offers a wide range of configurations with up to 8 central processing units (CPUs), 8 gigabytes (GB) of processor storage, and 256 channels in the largest models. Millennium Global Servers have the following characteristics:

Performance. With large 256-kilobyte (KB) Level 1 caches and very large 8-megabyte (MB) Level 2 caches in each central processing unit (CPU), Millennium servers have fast CPUs that deliver stable performance at very high processor utilizations.

Availability. Millennium servers maximize availability by combining fault avoidance from a highly reliable CMOS technology with fault tolerance from mature recovery processes evolved over generations of Amdahl processors. Availability is further enhanced by concurrent maintenance, rapid upgrade features, and supplemented by Parallel Sysplex capability.

Versatility. Millennium Global Servers are versatile systems that are compatible with Enterprise Systems Architecture/390 (ESA/390) and can be used for traditional System/390 processing, as Parallel Sysplex nodes, and in Open Systems environments.

Flexibility. Millennium Global Servers have large configurations that are easy to change, upgrade, or convert to other uses, giving customers wide latitude for both customizing their configurations and adapting them to changing needs. Configuration flexibility includes adding a second Millennium server, such as a Millennium Coupling Server, in the same footprint occupied by an existing server to form a Dual Server.

Lower cost of computing. Power and cooling costs are a fraction of those of a comparable processor using emitter coupled logic (ECL)—just one of several reasons that Millennium Global Servers are economical to operate.

1.1 Performance

Millennium servers combine fast CPUs with a sophisticated storage system that is distinguished by very large caches. The largest Millennium model, GS585 with 8 CPUs, has from 1.33 to 1.47 the performance of an Amdahl 5995-4570M. Each Millennium CPU has a 256-KB Level 1 cache and an 8-MB Level 2 cache, and all CPUs can exchange cache data with each other. The large caches improve individual CPU performance and are particularly beneficial to the performance of online transaction processing (OLTP) and database workloads.

The large caches and cache data exchange facility also greatly reduce the demand on processor storage, which minimizes storage queuing effects. Typically, transaction performance degrades at high processor utilizations because of lengthening storage queues and other multiprocessing overhead, but Millennium servers can be run at very high utilizations with stable transaction performance. The storage system also reduces variability in performance as more CPUs are added. Millennium servers scale well, resulting in predictable performance for applications as CPUs are added to the server.

1.2 Benefits of CMOS Technology

The primary logic technology in Millennium servers is 0.35 micron CMOS. CMOS technology has improved to the point of now being an appropriate technology for most workloads, allowing the many benefits of CMOS technology to be obtained in a mainframe server.

CMOS technology consumes much less power and generates much less heat than ECL technology. Millennium servers do not require the water cooling technology required in ECL mainframes, and the costs of operation can be as much as 97 percent less than for an ECL processor of comparable performance. Because they generate less heat, CMOS components can have significantly higher circuit density than ECL components. A typical Millennium CMOS chip can have hundreds of thousands of circuits per chip, whereas an ECL chip of the same dimensions may have tens of thousands of circuits. Millennium servers require over 90 percent less floor space than comparable ECL processors. Because its density enables a system to be built with markedly fewer components than a comparable system using ECL technology, CMOS technology also possesses dramatically better reliability characteristics than ECL technology.

1.3 Availability

Millennium servers combine a highly reliable CMOS technology, providing fault avoidance, and a mature set of recovery facilities evolved over generations of Amdahl processors, providing fault tolerance. The number of scheduled outages is minimized through concurrent maintenance of the most frequently replaced components. The duration of scheduled outages can also be reduced through the QuickCapacity™ feature, which provides for rapid model upgrading.

ConServ. To reduce the incidence of scheduled outages, Millennium servers have ConServ™, providing concurrent maintenance of the most frequently upgraded or maintained components. Firmware on Millennium servers can be updated concurrently with normal operation. All maintenance to the channels, cooling fans, many parts of the service processor, operator stations, and the operator station local area network (LAN) can also be done concurrently with normal operation.

QuickAvailability. QuickAvailability™ features specially designed for Millennium servers enhance concurrent maintenance capabilities and facilitate rapid upgrades. QuickAvailability features include the following:

- QuickCapacity feature for rapid CPU upgrades
- QuickChannels™ feature for concurrent channel upgrades
- QuickSwitch™ feature for dynamic substitution of CPUs
- QuickMemory™ feature for rapid memory upgrades

The QuickCapacity and QuickSwitch features enhance server availability using the Millennium CPU Fencing™ capability. Millennium servers each have the following maximum CPU capacity for their model class: either 5 or 8 CPUs for 500 Series servers and 2 CPUs for 400 Series servers. CPUs that are not enabled are fenced off into a reserve. The QuickSwitch feature enables dynamic replacement of a non-functioning CPU with one from the fenced reserve, thus avoiding a server outage. The QuickCapacity feature allows the rapid addition of CPUs from a fenced reserve. The QuickChannels feature allows for groups of parallel channels, ESCON channels, coupling links, or OSA features to be added concurrently with normal operation. Spare IOPA capacity is required for this feature which can be made available by pre-installing the Concurrent Channel Group Upgrade feature during the initial system installation. The QuickMemory feature allows for rapid memory config-uration upgrades from a fenced reserve of memory.

Parallel Sysplex. Participation in a Parallel Sysplex environment can augment the recovery and continuous operation characteristics of Millennium servers.

Support and Service. Millennium customers benefit from access to Amdahl's well-developed support network, including the Worldwide Customer Support Center (WWCSC), a network of problem specialists, standard problem escalation procedures, and a comprehensive spares system. Amdahl offers a variety of additional operational services beyond maintenance support, allowing customers to tailor service to their needs.

1.4 Versatile Feature Set

Millennium Global Servers have a rich feature set that makes them suitable for multiple processing environments, including traditional System/390 processing, Parallel Sysplex, and open systems environments.

Compatibility. Millennium Global Servers are compatible with ESA/390 and support the execution of the OS/390, MVS/ESA, VSE/ESA, UTS®, Transaction Processing Facility (TPF), and TPF High Performance Option (HPO) operating systems.

Configuration. The models in the Millennium Global Server family range from the GS415 uniprocessor to the GS585 with eight CPUs.

Millennium Global Server 500 Series models have

- from three to eight CPUs
- from 512 MB to 8 GB of processor storage
- up to 256 Enterprise Systems Connection (ESCON) serial channels, up to 128 parallel channels, up to 32 coupling links, and up to 16 open systems adapter (OSA) features

Millennium Global Server 400 Series models have

- one or two CPUs
- from 256 MB to 4 GB of processor storage
- up to 128 ESCON channels, up to 64 parallel channels, up to 32 coupling links, and up to 16 OSA features

Processor storage can be configured as either main or expanded storage under operator control as long as a minimum of 128 MB of main storage is provided. For all models, the total maximum number of channels possible in any particular configuration is dependent on the number of each type of channel installed. The largest possible channel configuration—256 channels on a 500 Series model and 128 on a 400 Series model—is one composed entirely of ESCON channels. Series 500 models have a minimum of 32 channels, and Series 400 models have a minimum of 8 channels.

The Multiple Domain Feature™ (MDF™), which is standard on all models, allows the resources of a Millennium server to be logically partitioned among up to ten logical processing systems called domains. MDF operator screens are Logical-Partitioning-compatible (LPAR-compatible) and can be used in an IBM Hardware Management Console (HMC) group.

1.4.1 Parallel Sysplex Environment

Millennium servers' performance and availability coupled with Parallel Sysplex features make them excellent choices for Parallel Sysplex nodes.

- With up to 32 coupling links, Millennium servers are suitable for the largest Parallel Sysplex environments.
- Millennium Global Servers can have up to two External Time Reference Attachments (ETRA), or Parallel Sysplex timer ports.
- Customers can construct a single-system or multi-system Parallel Sysplex environment by running Amdahl Coupling Control Code™ (ACCC) in an MDF domain, creating a Coupling Domain™. A Coupling Domain provides coupling facility function for other domains that run MVS/ESA images and communicate with the Coupling Domain via coupling links.
- All Millennium models meet the hardware requirements of the IBM, Parallel Sysplex License Charge (PSLC), which can substantially lower software license fees. By running ACCC in a Coupling Domain with coupling links, customers who satisfy other PSLC prerequisites can construct single- and multiple-system Parallel Sysplex environments to take advantage of PSLC pricing without attachment to a standalone coupling facility.
- Millennium servers can be operated locally from a customer operator station, remotely through a telecommunications link, or automatically through various automated operations software programs. With the optional Hardware Management Console Interface Feature (HMCIF), the server can also be operated from an IBM Hardware Management Console (HMC). HMC compatibility facilitates the integration of Millennium servers into a Parallel Sysplex environment.
- The Millennium Global Server is one element of the Amdahl Practical Solutions for Parallel Sysplex, which includes Parallel Sysplex nodes, coupling facility options such as Coupling Domains, PSMF, and Coupling Servers, disk storage systems, migration facilities, and services.

1.4.2 Open Systems Environment

A Millennium Global Server can function as a server in an open systems environment. To support this role, Amdahl will integrate an open systems adapter feature to provide LAN connectivity to the server. Amdahl also offers the UTS system, support for OS/390, open systems connectivity tools, and direct ESCON channel attachment to EnVista™ servers.

1.5 Configuration Flexibility

Millennium Global Server features not only give customers the versatility of operating in traditional S/390 as well as in Parallel Sysplex and open systems environments, but also provide wide latitude for customizing their configurations.

- With a large range of I/O and storage configurations, customers can select a configuration ideally suited to their needs. For example, a large number of parallel channels allows customers to convert to ESCON channels at a slower pace.
- Millennium availability features make these large configurations easy to change. The QuickCapacity feature allows CPUs to be quickly added to the configuration. The QuickChannels feature allows channels to be added or changed during normal operation. The QuickMemory feature allows memory to be quickly added to the configuration.
- Moreover, the Dual Server feature extends this ease of expansion to adding a second server in the same set of frames and footprint occupied by an existing server. As an example, a Millennium Coupling Server can be installed in the same frames with a Millennium Global Server, providing the principal components of a Parallel Sysplex in a single compact footprint.
- Millennium servers can also be readily converted to other uses. Millennium Global Servers can be converted to Millennium Coupling Servers and Millennium Coupling Servers can be converted to Millennium Global Servers. Thus, a Parallel Sysplex node could be easily converted to a coupling facility, or vice versa.

This flexibility to change configurations easily, to add an additional server in the same system footprint, and to convert servers to other uses enhances an investment in either a Millennium Global Server or a Millennium Coupling Server.

1.6 Lower Computing Costs

Millennium Global Servers can lower computing costs in several ways.

Protection of investment. Millennium servers enhance the value of existing System/390 applications through lower costs of computing, stable performance, and participation in a Parallel Sysplex environment. The value of Millennium servers is extended by their versatility, making them good choices for organizations operating in multiple environments or making transitions from one environment to another, such as from traditional System/390 processing to Parallel Sysplex.

Lower environmental costs. An advantage of a CMOS-based server is that it uses much less power, needs much less cooling than does an ECL processor, and uses much less floor space. Moreover, cooling is simpler. Millennium servers can be operated in raised or non-raised floor environments with simple air cooling.

For example, a Millennium GS585 with 4 GB of processor storage, 160 ESCON channels, and 48 parallel channels occupies only 6.0 m^2 (64.6 ft^2) of floor space (including service clearance), a 92-percent reduction from the floor space needed by an Amdahl 5995-8650M, an ECL processor of nearly identical processing capacity, configured with 2 GB of main storage, 2 GB of expanded storage, and the same channel configuration as the GS585. The GS585 requires 4.2 kVA at 60 Hz and 13.6 kBTU/hour of air conditioning, both 97-percent reductions from the 5995-8650M.

Lower software costs. Software costs can account for half of an information technology budget. Millennium servers can facilitate lowering software costs in two ways. All Millennium models allow customers to take advantage of the IBM multi-system PSLC software pricing. With the Amdahl Parallel Sysplex Migration Facility (PSMF) which packages a Coupling Domain with two internal coupling links, customers can take advantage of PSLC pricing in a single-system Parallel Sysplex environment. And using the Multiple Domain Feature (MDF), workloads from smaller processors can be consolidated onto one Millennium server, providing economies of scale and potentially large software cost savings.

Lower costs through improved availability. Millennium servers can reduce losses due to outages. Fault avoidance from a highly reliable technology and fault tolerance from recovery facilities evolved from earlier Amdahl processors produce a server that will rarely fail. Planned outages are reduced through concurrent maintenance of the most frequently maintained components. QuickAvailability features can reduce the duration of planned outages. Integration into a Parallel Sysplex environment can provide another level of protection for both recovery and maintenance actions.

Lower operation and system maintenance costs. The costs of operation can also be reduced from those of earlier processors. Millennium servers can be operated in various automated operation environments or from an IBM HMC. MDF provides operational and systems management LPAR compatibility, facilitating common operations and system management practices with other logical partitioning systems.

Millennium servers are explained further in the remaining chapters of this manual.

Chapter 2, “Features,” provides an explanation of Millennium Global Server features.

Chapter 3, “Compatibility,” provides information on ESA/390 architectural compatibility and related issues.

Chapter 4, “Multiple Domain Feature,” explains logical partitioning using MDF.

Chapter 5, “System Design and Technology,” provides additional information on performance, technology, and system design.

Chapter 6, “Availability,” explains Millennium server availability features and characteristics.

Appendix, “Millennium Coupling Servers,” provides information on the Millennium Coupling Server, a product related to the Millennium Global Server.

2 Features

Millennium Global Servers are versatile systems that are suitable for traditional System/390 compatible operation, for integration into a Parallel Sysplex environment, and for use as a server in an open systems environment. The Millennium Global Server is one element of the Amdahl Practical Solutions for Parallel Sysplex, a comprehensive suite of Parallel Sysplex products and services. Millennium Global Servers are adaptable to local, remote, and automated operation, as well as operation from an IBM HMC. Millennium servers have a broad range of configuration options, model conversions, and rapid update facilities that gives customers the flexibility to adapt Millennium servers to their needs.

2.1 Millennium Models

A Millennium Global Server can have from one to eight CPUs. Millennium servers have very low multi-processing overhead, resulting in predictable scalability and stable application performance as CPUs are added to a system. Except for processor storage and I/O channels, each model can have the same feature set as any other model. Table 2-1 shows CPU configuration by model number.

Millennium models can be upgraded with additional CPUs to higher models (see section 2.9, “QuickCapacity Feature”). In addition, a Millennium Global Server can be converted to a Millennium Coupling Server, and a Coupling Server can similarly be converted to a Global Server. This conversion flexibility enhances an investment in a Millennium server by extending its potential uses.

Table 2-1. Millennium Global Server Models

Model	Number of CPUs
GS415	1
GS422	2
GS425	2
GS535	3
GS545	4
GS555	5
GS565	6
GS575	7
GS585	8

2.2 Processor Storage

On Millennium Global Servers, processor storage can be configured as either main or expanded storage, and the assignment can easily be changed under operator control. Changing the storage configuration requires a system reset. Each 500 Series model can have from 512 MB to 8 GB of processor storage, and each 400 Series model can have from 256 MB to 4 GB.

Main storage can be configured from a minimum of 128 MB to the maximum of all available processor storage. The maximum amount of expanded storage that can be configured on any model is up to 128 MB less than the whole of processor storage. For example, given 8 GB of processor storage, up to 8,064 MB can be configured as expanded storage.

The hardware system area (HSA) requires a minimum of 16 MB of main storage, but the total amount of HSA is highly dependent upon the Millennium configuration.

2.3 I/O Channels

Millennium Global Servers offer ESCON channels, parallel channels, coupling links, and OSA features. All 500 Series Millennium models have a minimum of 32 channels and can have up to 256 channels, if all of them are ESCON channels. The following are the maximum number of channels by channel type for 500 Series models:

- 256 ESCON channels
- 128 parallel channels
- 32 coupling links
- 16 OSA features

All 400 Series Millennium models have a minimum of 8 channels and can have up to 128 channels, if all of them are ESCON channels. A 400 Series model can have up to the following number of channels by channel type:

- 128 ESCON channels
- 64 parallel channels
- 30 coupling links
- 16 OSA features

The total number of channels that can be installed varies according to the numbers of each type of channel installed. Each server has a maximum channel card capacity and the number of channels per card is fixed by channel type. There are up to 32 channel cards in a 500

Series model and up to 16 channel cards in a 400 Series model. A channel card has eight ESCON channels, four parallel channels, two coupling links, or two OSA features.

2.3.1 ESCON Channels

Millennium ESCON channels can operate at an effective data transfer rate of up to 17 MB/second. When an ESCON channel is supporting the parallel Original Equipment Manufacturer's Interface (OEMI) by means of a serial-to-parallel converter, it can transfer data at up to 4.5 MB/second. In addition, each ESCON channel can operate at less than 4.5 MB for compatibility with older devices.

Any ESCON channel can be specified (in the I/O configuration data set) to be any one of the following:

- CNC (serial channel)
- CBY (parallel OEMI via ESCON converter for byte multiplexer channels)
- CVC (parallel OEMI via ESCON converter for block multiplexer channels)
- CTC (basic or extended modes)

For CBY and CVC channels, Millennium Global Servers support attachment of the IBM 9034 Model 1 or the Amdahl 1034 ESCON converter (serial-to-parallel converters). For CTC channels, channel-initiated retry and auto-disconnection are supported.

2.3.2 Parallel Channels

All standard features of parallel channels* are supported on Millennium Global Servers as well as the following:

- Data streaming (not applicable to byte multiplexer channels)
- High-speed transfer
- Command retry (not applicable to byte multiplexer channels)
- Dynamic reconnection (not applicable to byte multiplexer channels)

Block multiplexer channels can operate at speeds up to 4.5 MB/sec in data-streaming mode. Byte multiplexer channels can operate at up to 200 KB/second in burst mode.

The potential to attach up to 128 parallel channels accommodates customers wishing to migrate to ESCON channels at a slower pace.

2.4 Parallel Sysplex Environment

A Millennium Global Server can be integrated into a multi-system Parallel Sysplex as a node, can run a Coupling Domain to provide coupling facility function in a domain, or use PSMF to create a single-system Parallel Sysplex.

2.4.1 Coupling Links

Coupling links allow a Millennium server to attach to a coupling facility, such as a Millennium Coupling Server or an IBM 9674 Coupling Facility. Millennium supports 100 MB/sec 9 micrometer single-mode fiber coupling links that can operate for a maximum distance of three kilometers and 50 MB/sec 50 micrometer multimode fiber coupling links that can operate at up to 1 kilometer. Millennium coupling links can be defined for the Coupling Facility Send (CFS) function, allowing an MVS/ESA image to participate in a Parallel Sysplex. A coupling link can be defined for the Coupling Facility Receive (CFR) function for a coupling facility domain running ACCC.

The potential for 32 coupling links on Millennium servers allows customers to increase redundancy and improve performance while participating in the largest Parallel Sysplexes.

2.4.2 External Time Reference Attachments

Millennium Global Servers can have up to two External Time Reference Attachments (ETRA), or Parallel Sysplex timer ports, per system. A Millennium server may be operated with no External Time Reference Attachments for traditional operation or a single-system Parallel Sysplex environment, one for attachment to a Parallel Sysplex, or two for redundancy.

2.4.3 Coupling Domains

A Millennium server can be used as a coupling facility. Coupling facility function is provided by ACCC, which runs in a Coupling Domain on either a Millennium Coupling Server or a Millennium Global Server. For an overview of the Millennium Coupling Server, see the Appendix, "Millennium Coupling Server."

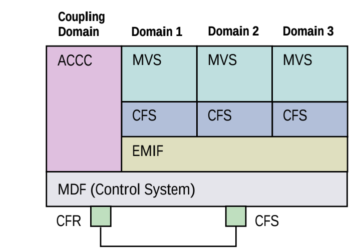
Customers can construct a single-system Parallel Sysplex by using PSMF to satisfy the requirements of the PSLC single-system option, allowing customers who satisfy other PSLC prerequisites to take advantage of PSLC pricing on a single Millennium server without attachment to a external coupling facility. PSMF packages a Coupling Domain with two coupling links,

* Parallel channel features are described in *IBM System/360 and System/370 I/O Interface Channel to Control Unit Original Equipment Manufacturers' Information.*

providing the basis for a single-system Parallel Sysplex for customers who wish to take advantage of single-system PSLC pricing.

ACCC can be the basis of a progressively expanding Parallel Sysplex environment:

1. In the most basic configuration with PSMF. ACCC runs in one domain and MVS/ESA Parallel Sysplex images run in other domains (figure 2-1a). In this single-system Parallel Sysplex implementation, no ETRA is



ACCC = Amdahl Coupling Control Code

CFR = Coupling Facility Receiver Link

CFS = Coupling Facility Sender Link

EMIF = ESCON Multiple Image Facility

MDF = Multiple Domain Feature

necessary. MVS/ESA domains are connected through real coupling links to the Coupling Domain. The MVS/ESA domains can share coupling links defined as CFS via ESCON Multiple Image Facility (EMIF). This configuration provides a realistic Parallel Sysplex testing environment for Parallel Sysplex.

2. When a second system is added to the Parallel Sysplex, ETRAs are necessary to synchronize the two systems' clocks, and coupling links are added from the second system to the Coupling Domain on the first system (figure 2-1b).

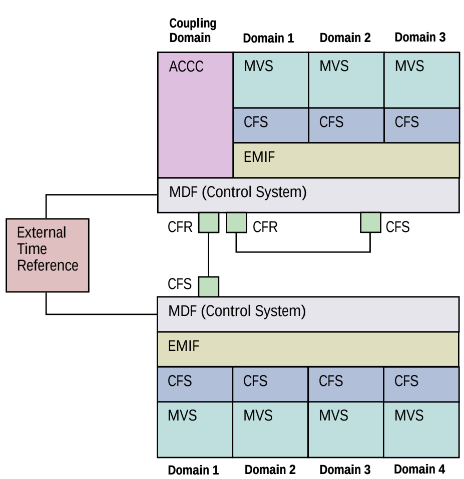


Figure 2-1b. Two-System Parallel Sysplex using a Coupling

3. The final stage of evolution is to replace the Millennium Global Server Coupling Domain with a standalone coupling facility, such as a Millennium Coupling Server, and to interconnect the Coupling Server and the MVS/ESA images with coupling links (figure 2-1c). The Coupling Domain can still be used with other domains as a self-contained test system for Parallel Sysplex or as a backup coupling facility.

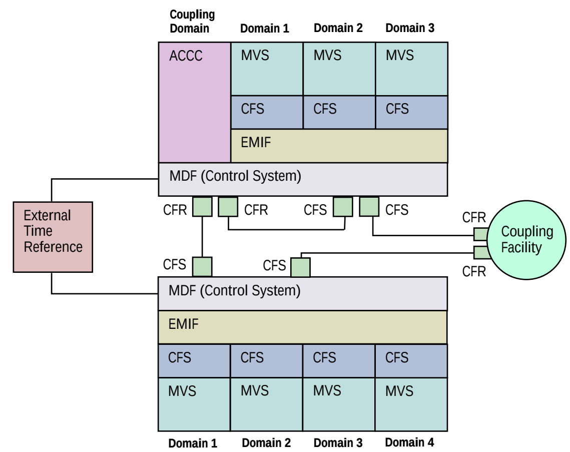


Figure 2-1c. Two-System Parallel Sysplex with Coupling Server and Coupling Domain test system.

2.4.4 Practical Solutions for Parallel Sysplex

The Millennium Global Server is one element of a comprehensive suite of Amdahl products and services that implement and help customers manage their transition to a Parallel Sysplex environment. The Amdahl Practical Solutions for Parallel Sysplex offering includes the following products and services.

▼ Hardware offerings:

• Amdahl provides two processor families that can serve as Parallel Sysplex nodes, the Millennium Global Servers and the 5995M Processors. 5995M processors, available in 20 models, including the 12-way 5995-12670M, are ECL processors that can handle the most demanding workloads.

• With the Millennium Coupling Server, Amdahl provides a coupling facility that can interconnect up to 32 Parallel Sysplex nodes. See the Appendix, “Millennium Coupling Server,” for more information.

• The Hardware Management Console Interface Feature (HMCIF) enables Millennium Coupling Servers, Millennium Global Servers, and 5995M Processors to be operated from an IBM HMC. See section 2.7.3, “Hardware Management Console Interface Feature,” for more information.

• Coupling links provide the interconnection of Parallel Sysplex nodes and coupling facilities, whether the latter are standalone coupling facilities or Coupling Domains.

• Spectris™ RAID Storage Subsystems provide the consistently high performance and availability needed for Parallel Sysplex environments.

• Amdahl 1032 Model 3 Directors provide dynamic fiber optic ESCON channel connections between Parallel Sysplex nodes and attached storage control units.

- Software offerings:

• ACCC is the foundation of Amdahl coupling facility offerings. ACCC provides coupling facility capability on Millennium Coupling Servers and in Coupling Domains on Millennium Global Servers and 5995M Processors, allowing up to 32 MVS/ESA images to concurrently share access to the same databases.

- Migration facilities:

• Coupling Domains provide coupling facility functionality using ACCC without a standalone coupling facility.

• PSMF packages a Coupling Domain with two coupling links to facilitate the creation of a single-system Parallel Sysplex.

- Expertise and services:

• The Amdahl Enterprise Computing Center researches the behavior of the Parallel Sysplex architecture in a real-world environment.

• Amdahl offers a suite of services to help customers make a transition to Parallel Sysplex. The Assessment Service for Parallel Sysplex provides a clear understanding of the benefits and risks associated with deploying Parallel Sysplex. The Planning Service for Parallel Sysplex provides customers with a detailed plan for making the transition to a Parallel Sysplex environment. The Implementation Services for Parallel Sysplex executes this plan and minimizes the time required to successfully complete the migration.

2.5 Open Systems Environment

A Millennium server can function as a server in an open systems client/server environment. Millennium servers offer open systems adapter (OSA) features for LAN connectivity and support for open systems environments.

2.5.1 Open Systems Adapter Feature

The integrated open systems adapter (OSA) feature combines a new type of S/390 channel with standard physical network media connectors. The OSA feature allows direct Millennium Global Server connection to industry-standard network technologies without requiring the use of an external channel-attached control unit. Direct network connectivity allows the Millennium Global Server to be positioned as an open distributed application server in a manner similar to the way non-S/390 platforms are positioned as distributed application servers.

The OSA feature allows attachment to the following networks:

- Fast Ethernet (100 Mb/sec) with full duplex capability
- Fiber Distributed Data Interface (FDDI) with dual ring attachment capability
- Asynchronous Transfer Mode (ATM) 155 Mb/sec

The configuration of OSA features is managed using a combination of standard S/390 management tools and new OSA specific management tools. The channel configuration and associated channel resources are managed using existing S/390 applications such as Hardware Configuration Definition (HCD) and I/O configuration program (IOCP). In addition, the OSA channel type is supported for dynamic reconfiguration management (DRM) actions and is eligible for domain sharing in a multiple domain configuration.

Each OSA channel card contains two OSA features with each OSA feature supporting one of the following physical network connnections:

- Two Fast Ethernet connectors (100BAJE-TX Cat 5 UTP)
- Two FDDI connectors
- One 155 Mb/sec ATM connector and function to support Token-Ring or Ethernet LAN emulation

Millennium Global Servers can be configured with up to 16 OSA features per system, in increments of one.

Two major communication environments are supported when using an OSA feature:

- Transmission Control Protocol/Internet Protocol (TCP/IP) passthrough mode
- Systems Network Architecture (SNA) mode

By supporting both TCP/IP and SNA environments, the OSA feature enables the implementation of applications based on a variety of distributed processing models. These models include:

- Client/server (distributed and point-to-point)

- SNA (hierarchical and peer-to-peer)
- Remote Procedure Call (RPC)

Applications can take advantage of the higher bandwidth and availability of the Millennium Global Server while satisfying the requirements imposed by a distributed processing environment.

2.5.2 Open Systems Support

Millennium Global Servers support the Amdahl UTS operating system, as well as OS/390. The UTS system offers the unique combination of true, native UNIX® with the reliability and scalability of System/390 hardware. With the UTS system, enterprises can gain the benefits of open system interoperability, productivity, and connectivity while maintaining the power of legacy systems and protecting investment in information assets.

Amdahl also offers a suite of connectivity tools and software that enables access to mainframe data from distributed UNIX or Windows NT systems. These include a direct ESCON channel to EnVista servers and database gateways that facilitate OLTP access from user personal computers.

2.6 Multiple Domain Feature

MDF, a standard feature of Millennium Global Servers, enables system resources to be logically partitioned among up to 10 logical processing systems or domains. MDF operator screens and systems management parameters are compatible with the IBM Processor Resource/Systems Manager (PR/SM) product and can be operated as a member of a group from an IBM HMC or can be controlled from a customer operator station.

MDF has three options for allocating CPU resources:

- CPUs can be dedicated to all of a domain's logical processors.
- A domain's logical processors can share the pool of CPUs with other domains, with sharing based on customer-specified weights.
- Dedicated and shared allocations can be combined in the same domain. This Amdahl unique hybrid allocation combines the advantages of low overhead from dedicated CPUs and the dynamic workload allocation of shared CPUs.

When CPUs are shared among domains, MDF normally allows a domain's share to decrease and increase dynamically to better use CPU resources. However, a domain's share can be "capped" so as not to expand when the domain would normally be eligible for more CPU time, using either a relative limit based on a proportion of weights or an absolute limit using MDF's unique Max Capping™ feature. The Max Capping feature enforces an absolute percentage limit to a domain's share of the CPU pool, even if only one domain is active.

MDF supports Dynamic Storage Reconfiguration (DSR) and Automatic Reconfiguration Facility (ARF). MDF also supports EMIF, allowing all active domains to share ESCON channels, coupling links, and OSA features with other domains and DRM, which allows any I/O configuration data set (IOCDS), even the active IOCDS used by active domains, to be dynamically altered.

MDF is described in detail in chapter 4, "Multiple Domain Feature."

2.7 Operational Features

Millennium Servers can function in many operational environments, including direct and automated operation, local and remote operation, and operation from an IBM HMC.

2.7.1 Operator Stations

Millennium operator stations are Amdahl supplied PC compatible computers operated with graphical user interfaces (GUIs). Operator stations are connected to the service processor by a dedicated Ethernet LAN.

Each Millennium server has one service operator station (SOS), which is used by Amdahl personnel to maintain the server. In addition to its LAN connection, the SOS is connected to the service processor by a serial cable and can be located up to 15 m (49 ft) from the Millennium system frame.

There are from one to five customer operator stations (COSs). One COS functions as the primary operator station and is used to control the server, including all domain activities. Other COSs are for display only and may serve as alternates for the primary COS. COSs can be located up to 100 m (328 ft) from the service processor.

Each COS supports ESA/390 operator facilities as well as Amdahl specific functions. Through the primary COS, or an HMC connected through the optional Hardware Management Console Interface Feature (HMCIF), an operator or an automated operation program can control the server. From the primary COS, an operator can:

- Perform power on, power off, system reset, initial microprogram load (IMPL), and initial program load (IPL)

- Configure CPUs, channels, and processor storage
- Configure domains using MDF
- Display and update register and processor storage

COSs are not intended for system control program (SCP) operation. However, Millennium Global Servers support Console Integration, allowing COSs to be used as SCP consoles for backup purposes. In addition, a COS can be configured as an SOS by a service representative, if required. The SOS may also serve as a backup COS.

Through the SOS, or through a communications link to the Amdahl WWCSC, a service representative can diagnose and maintain the server.

2.7.2 Automated and Remote Operation

The primary COS has a Simple Network Management Protocol (SNMP) agent function, allowing the server to be operated from automated operation programs. If the server is being operated from an HMC, the HMC SNMP agent function supersedes the SNMP agent function in the primary COS. A Token-Ring adapter is required for this facility.

The primary COS also has Distributed Console Access Facility (DCAF) for remote console access. DCAF is configured for TCP/IP to connect to the Amdahl Worldwide Customer Support Center (WWCSC) and for customer remote console access. Customers may supply an external modem for connection to a remote PC running DCAF. These two uses of DCAF are mutually exclusive at any one time, because DCAF gives the user exclusive control of the console session, locking out other users.

2.7.3 Hardware Management Console Interface Feature

The optional HMCIF enables a Millennium Global Server to be operated from an IBM HMC. The HMCIF is composed of a PC application and a Token-Ring Adapter to connect to an HMC complex. The HMCIF is installed on the primary COS.

As part of an HMC complex, Millennium servers will:

- Respond to HMC commands targeted to a Millennium server, or a Millennium domain, as either a single system or as a part of an HMC group.
- Participate in remote interactions, such as IOCDS Write or initial program load (IPL) information access, with support elements on other processors in an HMC group. (See also section 3.5, “Remote IOCDS and Remote IPL.”)
- Provide real-time status information to the HMC so that Millennium server status can be displayed.
- Provide for the consolidation of hardware and operating system messages on the HMC.

2.7.4 Alarms

A Millennium server has a local alarm and can attach an optional remote alarm that is provided by the customer. The alarm is similar to the Remote Alarm Interface RPQ for 5995A Processors and the Remote Audible Alarm Standard on IBM ES/9000 9022 Processor Controllers. Conditions that can cause the alarm to sound include:

- Power abnormalities or failures
- System environmental abnormalities or failures
- Critical failure of the service processor
- Activation of the alarm by console command

The alarm remains activated until reset manually or by console command or until service processor power is turned off. The alarm can also be muted while the condition causing the alarm is cleared.

2.7.5 Support Center Connection

The service processor has a wide area network (WAN) router and modem for connection to the Amdahl Worldwide Customer Support Center (WWCSC). Amdahl WWCSC personnel can utilize extensive support tools to provide remote service without compromising customer security provisions.

2.7.6 Operational Environments

Millennium Global Server features allow operation in several types of environments:

Direct local standalone operation. A Millennium server can be operated directly using Millennium COSs. This operational environment requires a minimum of an SOS and a COS. Millennium operation is directed using the COS, which can be up to 100 m (328 ft) from the service processor (figure 2-2a).

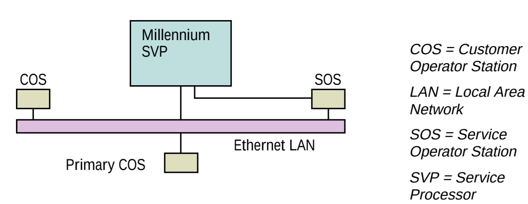
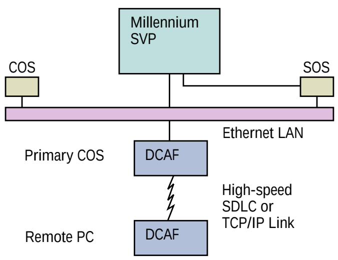


Figure 2-2a. Local direct operation.

Direct remote standalone operation. In this environment, the COS is remote, more than 100 meters (328 ft) from the Millennium service processor, but otherwise operation is the same as for direct local operation. In this method, a remote PC running DCAF connects via TCP/IP or Synchronous Data Link Control (SDLC) to the DCAF interface on the primary COS and control is switched to the remote station (figure 2-2b).



DCAF = Distributed Console Access Facility

PC = Personal Computer

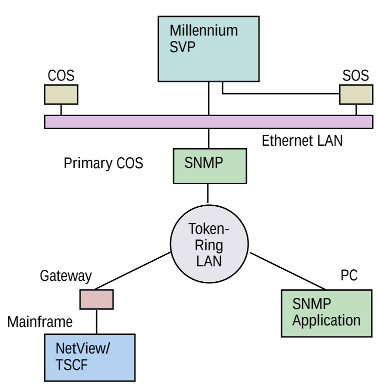
SDLC = Synchronous Data Link Control

SNMP = Simple Network Management Protocol

TCP/IP = Transmission Control Protocol/ Internet Protocol

Figure 2-2b. Remote direct operation.

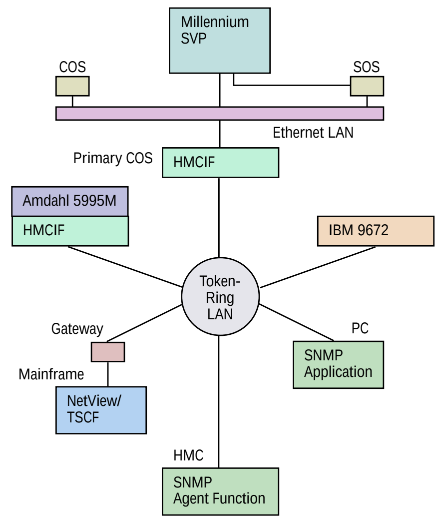
Automated standalone operation. Using the SNMP agent function in the primary COS, Millennium operation can be automated by system management applications from several vendors. Similarly, Millennium operation can be automated from a focal point mainframe running NetView or Target System Control Facility (TSCF) under NetView. In this case, both the gateway to the focal point mainframe and the Millennium primary COS are connected to the same Token-Ring LAN (figure 2-2c).



TSCF = Target System Control Facility

Figure 2-2c. Automated operation.

Direct or automated Parallel Sysplex operation. A Millennium Global Server with HMCIF can be operated in a Parallel Sysplex environment from an HMC, where all Parallel Sysplex nodes are connected with the HMC over a Token-Ring LAN. The systems connected through the HMC can be operated directly or operations can be automated via an SNMP application or a focal point mainframe, as in the previous example. In this case, the HMC has an SNMP agent function that supersedes that of the Millennium primary COS (figure 2-2d).



HMC = Hardware Management Console

HMCIF = Hardware Management Console Interface Facility

Figure 2-2d. Automated Sysplex operation.

2.8 Performance Monitors

Millennium servers have several features that provide performance measurement information and aid capacity planning. The System Activity Monitor (SAM) is a real-time performance monitor for collecting and reporting the activity of Millennium CPUs, channels, and domains. SAM performance data can be displayed in real time on a COS or accumulated by a data collection facility such as the Amdahl Performance Analysis Facility (APAF).

The following categories of performance data are collected:

- CPU busy time is collected as supervisor or problem state time and is segregated by domain, and optionally by CPU number. CPU busy time can also be qualified by key or exclusion of a single key.
- Channel path busy information can be displayed for both shared and non-shared channels. When channel paths are shared among domains under EMIF, SAM can display a domain's share of a channel path or the aggregate usage of that path by all domains.

The source of data for SAM is MDF, which maintains configuration information and collects scheduling data continuously. SAM provides a subset of this data. Other performance, tuning, and capacity planning information can be found in Machine Readable Scheduler Data

(MRSD), which is provided as a feature of MDF. Subsets of this data can be accessed by various performance monitors and reporting programs. MRSD data and programs to access it are described in section 4.8, “Performance and Tuning.” Millennium Global Servers can also run third-party measurement products.

2.9 QuickCapacity Feature

The QuickCapacity feature provides a way to rapidly add CPU capacity. On 500 Series systems there are either five or eight CPUs and on 400 Series systems there are two. Millennium servers have CPU Fencing capability—CPUs not enabled are fenced off into a CPU reserve. For example, in a Millennium GS535 there are three CPUs activated and two held in reserve. In a Millennium GS555 there are five CPUs activated and three held in reserve. Customers may acquire additional CPUs from the fenced reserve as their capacity needs increase. This upgrade capability is applicable to all models except the Millennium GS585 and GS425.

After the upgrade, the results of the STORE CPU ID instruction will reflect the additional CPUs. Customers are responsible for changes to software licensing.

CPUs can also be exchanged dynamically from the reserve in case of a CPU repair action. See section 6.5, “QuickSwitch Feature,” for more information.

2.10 QuickChannels Feature

Millennium servers allow concurrent maintenance and upgrades for channels and coupling links. (See section 6.6, “Concurrent Maintenance.”) The number of channels or coupling links that can be added during normal operation is limited to the unused capacity of existing I/O Processor Adapters (IOPAs). (See section 5.6, “I/O Processing.”) Thus, adding IOPAs increases the potential for adding channels and coupling links concurrently with normal operation. Because each IOPA accommodates up to 4 channel cards, adding an IOPA provides for the future concurrent upgrade of up to 32 ESCON channels, up to 16 parallel channels, up to 8 coupling links, up to 8 OSA features, or some combination of the 4 types. Having more than 4 IOPAs requires installation of the channel expansion frame. (See section 5.9, “Frames.”)

2.11 QuickMemory Feature

The QuickMemory feature provides a way to rapidly upgrade Millennium server memory configurations.

A fenced reserve of memory on Millennium servers can be used to increase memory capacity quickly. QuickMemory upgrades are applicable to all models. The size of the fenced reserve varies by the size of the memory configuration.

2.12 Dual Server

Two Millennium servers in any combination of models, including Millennium Global Servers or Millennium Coupling Servers, can be installed together in the same system frame. Installing a Millennium Coupling Server as the second server allows the major components of a Parallel Sysplex to reside in one compact system footprint.

Except for sharing the same frames and cooling, the second server is a completely separate system with its own processor storage, channels, coupling links, service processor, operations LAN, and operator stations. Each server has its own control panel. Both servers’ system boards have separate power supplies, and one server’s system board can be powered off while the other continues operation. There is one Emergency Power Off (EPO) switch for the whole complex.

With the Dual Server feature, the maximum number of channels for both servers combined is the same as for a single 500 Series Millennium Global Server. Physically, there are 8 possible IOPAs, which are divided into 2 contiguous groups. IOPA assignment starts at IOPA0 for one server and starts at IOPA7 for the other. For example, if a Millennium Coupling Server and a Millennium Global Server are installed as a Dual Server, the Coupling Server could be assigned IOPA0, allowing it up to 8 coupling links, while the Global Server could be assigned the remaining 7 IOPAs. If not all IOPAs are installed, the unused IOPA positions remain unassigned.

2.13 Power and Cooling

Millennium Global Servers have highly reliable power subsystems. Millennium uses 208 VAC, three-phase input voltage. Power may be conditioned or nonconditioned 50/60 Hz. Power converters are n+1 redundant.

The Power Control System turns server power on and off and monitors power supply operating status. The server can be powered on or off manually through a customer-accessible switch located on the system panel in the channel frame. Power can also be turned on or off by command through a COS or through a remote DCAF console. Turning power off by command turns off power to the server except for the service processor, system console interface, system power controller, and fans. With the Remote Power Control Facility, these units, except for minimal SVP functions, can also be powered on or off by command.

Millennium servers are air-cooled. A Millennium server can operate without underfloor cooling in a raised floor environment, but underfloor cooling airflow is the preferred operating environment. A Millennium server is self-monitoring and can protect itself through shutdown from out-of-specification operating conditions.

Maximum heat dissipation for a Millennium GS585 with 4 GB of processor storage and 208 channels (assuming 160 ESCON and 48 parallel channels) is on the order of 14.3 kBTUs per hour or about 3,600 kilocalories/hour. Cooling for this amount of heat dissipation would require 4.0 kW/hour of air conditioning.

2.14 Power Source Options

To protect customer information and facilitate continuous operations, Millennium servers can be configured with a battery backup system and/or a dual power feed.

Battery backup capabilities. S3 UPStation and AP4310 Uninterruptible Power Systems (UPS) models are available for Millennium servers as part of a relationship established with Liebert Corporation.

The battery backup capability of the local UPS provides full power for approximately 20 to 160 minutes depending upon the system configuration to allow continued operation of the entire system throughout that period.

Amdahl provides Power Save state capability for Millennium Coupling Servers running in non-volatility mode using the local UPS to preserve the contents of memory during an extended outage.

Dual power feed. Millennium servers can be configured with an optional dual power feed. This feature extends the redundancy already present in Millennium internal power supplies to the external power cord. Customers can attach Millennium servers to two separate external sources of data center power for very critical environments. Dual power feed can be implemented in one of two ways:

- The Liebert S3 and AP4310 UPSs accept dual power input and then provide a single feed into the Millennium server;
- Millennium servers can have a factory-installed native dual power feed capability direct into the server. This option is not available as a field upgrade and removes the capability of installing a Dual Server.

3 Compatibility

The Millennium Global Server reaffirms Amdahl's commitment to providing System/390 compatible systems. Millennium Global Servers are fully compatible with the ESA/390 architecture. Millennium systems can operate as a single-image system, in which all resources are allocated to a single domain, or can be logically partitioned into multiple domains. In all cases, SCPs operate in domains, and each domain is compatible with ESA/390.

This chapter identifies the architectural features available for Millennium Global Servers, as well as model-dependent instructions and model-dependent software modifications.

3.1 Architectural Compatibility

Millennium Global Servers conform to the following specifications:

- Domains conform to the specifications in the *IBM Enterprise Systems Architecture/390: Principles of Operation.*
- Parallel channels conform to the specifications in *IBM System/360 and System/370 I/O Interface Channel to Control Unit Original Equipment Manufacturer's Information.*
- ESCON channels conform to the specifications in *IBM Enterprise Systems Architecture/390: Enterprise Systems Connection Architecture I/O Interface.*
- Coupling links conform to the specifications for communication within a Parallel Sysplex environment and are based on ANSI Fibre Channel standard X3.230 using FC-0 and FC-1 levels.
- OSA features conform to S/390 channel interface protocols in communicating with SCPs. They are also based on the following standards:
  - FDDI: ANSI X3T9.5/ISO 9314
  - Ethernet: IEEE 802.3u
  - ATM: ATM UNI 3.0, 3.1, and 4.0

3.1.1 Program Compatibility

Any program written for ESA/390 mode can operate on a Millennium Global Server, provided that the program:

- Is not time-dependent.
- Does not depend on the presence of system facilities (such as storage capacity, I/O equipment, or optional features) when these facilities are not included in the configuration.
- Does not depend on the absence of system facilities when the facilities are included in the configuration.
- Does not depend on results or functions that are defined as unpredictable or model-dependent in the *IBM Enterprise Systems Architecture/390: Principles of Operation.*
- Does not depend on results or functions that are defined in this publication as being differences or deviations from the appropriate Principles of Operation publication.

Any problem state program written for System/370 operates in ESA/390 mode, and any problem state or control program written for 370-XA or ESA/370 operates

Table 3-1. Architectural Features and Assists

Asynchronous Data Mover Facility
Asynchronous Pageout Facility
Automatic Reconfiguration Facility
Broadcasted Purging
Channel Subsystem Call
Concurrent Channel Maintenance
Console Integration
Control Unit Queueing Measurement Facility
Data Compression
DB2 Sort Facility
Dynamic Reconfiguration Management
Dynamic Storage Reconfiguration
Enhanced Move Page
ESCON Channels
• CNC • CBY
• CVC • CTC (Basic and Extended)
ESCON Multiple Image Facility
Expanded Storage
Extended Sorting
Fault Tolerant Dynamic Memory Arrays
I/O Device Self-Description
Logging Volume Reduction (CUSE)
Logical String Assist
LPAR Interface*
Move Inverse (MVCIN)
Move Page
Multiple Domain Feature†
Program Event Recording 2 (PER 2)
SCP-Initiated Reset
Service Call Logical Processor
Subspace Group Facility Support
Subsystem Storage Protection
Suppression on Protection
Sysplex Timer
Virtual-Address Enhancement of Suppression on Protection

* LPAR Interface provides support for Diagnose instructions used in an LPAR environment.
† The Multiple Domain Feature provides logical partitioning on Millennium servers.

in ESA/390 mode, provided that in each case the program:

- Observes the limitations in the preceding statements.
- Does not depend on any programming support facilities that are not provided or that have been modified.
- Takes into account other changes made that affect compatibility between modes. These changes are described in the *IBM Enterprise Systems Architecture/ 390: Principles of Operation.*

3.1.2 Features and Assists

Millennium Global Servers implement the standard ESA/ 390 architecture as described in *IBM Enterprise Systems Architecture/390: Principles of Operation* (SA22-7201-02). Millennium servers also implement the supplemental architectural features given in table 3-1.

Millennium Global Servers do not support the Interpretive Execution Facility (IEF) nor any additional components of the architecture designed to support or complement that facility. Support for the Virtual Machine/Enterprise Systems Architecture (VM/ESA) SCP will not be provided with the initial Millennium offering.

3.2 System Control Programs

Millennium Global Servers support the execution of the following SCPs:

- UTS
- OS/390
- VSE/ESA
- MVS/ESA System Product Versions 3, 4, and 5
- Transaction Processing Facility (TPF) 4.1
- TPF High Performance Option (HPO) 4.1

Amdahl periodically announces which specific SCP releases are supported on its processors.

3.3 Model-Dependent Instructions

Certain instructions produce model-dependent results on various processors. These include Diagnose and Store CPU ID (STIDP). Diagnose is a supervisor-state instruction that provides an interface for diagnostics.

Store CPU ID stores model-dependent processor information at the doubleword addressed by the second operand. The STIDP instruction stores a doubleword with four information fields. The format of the stored information is given in table 3-2.

Table 3-2. Results of the Store CPU ID Instruction

Version Code	
Bits 0-3:	X'0' for model GS422 X'1' for all other models
Bits 4-7:	Number of Physical CPUs
Other bit configurations are reserved.	
CPU ID Number	
Bits 8-15:	Logical Processor Identification (LPID)
Bits 8-11:	Logical CPU Address for Logical Processor
Bits 12-15:	Domain Identifier
Bits 16-31:	Processor serial number
Model Number	
Bits 32-47:	For Millennium Global Server 500 Series: Set to X'0500' For Millennium Global Server 400 Series: Set to X'0400'
Machine Check Extended Logout (MCEL) Length	
Bits 48-63:	Set to X'0000'

All CPUs and logical processors within a Millennium Global Server have the same processor serial number. (If more than one server is installed in the same frame, each server has its own serial number.) Each logical processor in a domain has a unique default CPU ID constructed by prefixing the Logical Processor Identification (LPID) to the processor serial number. The customer may substitute a different LPID for the default LPID of a domain, as long as all LPIDs within that domain are unique.

3.4 I/O Configuration

A Millennium Global Server can have up to four IOCDS files. One IOCDS can be active at a time, defining the I/O configuration for all active domains.

Four starter IOCDS files and corresponding input I/O control source (IOCS) files are provided with a Millennium server. These starter IOCDS files can be customized using console screens to edit the IOCS input files. The starter IOCDS files range from small one-partition systems to a full sized 10-domain configuration.

A Millennium system includes a standalone I/O configuration program (IOCP) that accepts input compatible with IBM IOCP source statement formats. The Amdahl IOCP can be run from a Millennium operator station or, if the Millennium server has HMCIF installed and is connected to an HMC, from an IBM HMC. In addition, a Millennium IOCDS can be defined from a program running under an SCP or generated via IBM's Hardware Configuration Definition (HCD), which invokes the Amdahl IOCP.

3.5 Remote IOCDS and Remote IPL

If an IOCDS has been generated on a Millennium server by a program running under an SCP or via IBM's HCD, it can be transmitted to another system that is connected to the same HMC network, assuming both systems are defined to the HMC. Similarly, a Millennium system can receive, or be the target system for, an IOCDS generated on another system in the same HMC network. A Millennium server can provide protection for an IOCDS using IBM's HCD, including support for IOCDS PROT, IOCDS UNPROT, and LIST IOCDS commands.

If a program running under an SCP generates IPL information (including IPL address and parameters) for a target system, a Millennium server can read, update, and transmit this IPL information to another system in the same HMC network or be the target for such IPL information.

3.6 Model-Dependent Programming Support

Amdahl provides changes to MVS SCPs in areas where MVS requires model-dependent information. The purpose of these changes is to identify the Millennium Global Server and its capabilities to MVS. Amdahl also provides changes to TPF and TPF/HPO to correctly identify the Millennium Global Server and its capabilities to the TPF system. These changes add Millennium support, but do not affect the functioning of the software on other processors.

4 Multiple Domain Feature

The Multiple Domain Feature (MDF) provides logical partitioning for Millennium Global Servers and is a standard feature. Amdahl introduced logical partitioning in 1985 with the first release of MDF. By integrating special hardware for each logical partition or domain, Amdahl processor complexes could run multiple operating systems at close to native performance.

Since its introduction, Amdahl has continually enhanced MDF with System/390 compatibility features in response to customer needs. MDF provides reporting compatibility with PR/SM, including LPAR CPU performance reporting, and compatibility with system control program (SCP) functions that work in an LPAR environment, such as Automatic Reconfiguration Facility (ARF), Dynamic Reconfiguration Management (DRM), and ESCON Multiple Image Facility (EMIF).

For Millennium MDF, compatibility is extended to operational and systems management to support centralized and automated data center operations. MDF can be defined, managed, and operated in an environment compatible with logical partitioning methods from other vendors. Operator screens are broadly compatible with PR/SM. This expanded envelope of compatibility facilitates the integration of Millennium servers into automated operation environments and into Parallel Sysplex environments operated through an IBM Hardware Management Console (HMC).

Millennium MDF domains also have independent clock settings that allow Year 2000 testing to proceed in one or more domains while production systems execute in other domains.

4.1 The Domain Environment

An MDF system essentially provides a processor complex kit from which one or more logical processing systems can be constructed. In MDF, a logical processing system is called a domain. The domain is the most basic processing environment; the processor complex exists only as a pool of resources to be allocated to domains. There is always at least one domain.

A domain is the functional and operational equivalent of a separate processor complex. An SCP in a domain directly executes the processor instruction set, services its own I/O, and handles its own interrupts. Like physical processor complexes, domains have their own channels, main storage, and expanded storage.

MDF is a native execution environment on Millennium servers. Domains are integral to Amdahl processor complexes; logical partitioning is not executed through an interpreted interface. Millennium servers have special hardware that minimize overhead when control of a CPU is switched from one domain to another, including:

- Additional registers and instructions to communicate efficiently between MDF and the hardware about the state of the domain logical processors.
- Additional identifiers in the caches, translation lookaside buffer (TLB), and access register translation lookaside buffer (ALB) to identify domains uniquely, eliminating the need to purge whole buffers and thus reducing the overhead of sharing buffers among domains.

4.2 Allocating Resources to Domains

The pool of system resources—CPUs, main storage, expanded storage, and channels—can be allocated to a single domain or can be divided among up to 10 domains. These resources are assigned through the domain definition process. Domain configurations can be defined from the primary COS using GUI-type definition screens or can be defined under HMC control. I/O devices are assigned indirectly, by assigning their channels to a domain via the IOCP. When a domain is activated, MDF allocates resources according to the parameters of the domain's activation profile.

CPU capacity. A domain either owns at least one CPU exclusively (dedicated allocation) or receives a portion of the capacity of the shared CPU pool (shared allocation). A domain can also own one or more CPUs exclusively while sharing the CPU pool with other domains (mixed allocation).

Processor storage. Each domain owns exclusively the main storage and expanded storage assigned to it.

I/O facilities. I/O facilities for a domain, including parallel channels, ESCON channels, OSA-features, and coupling links, are defined in an LPAR-compatible IOCDS, which specifies channel paths and devices for the domain. A domain owns the physical channels assigned to it exclusively, unless EMIF is used to share ESCON channels, coupling links, and OSA features with other domains. Devices are shared between domains in the same way they are shared between processor complexes.

There is no need to allocate consoles, as all domain operations are directed through the primary customer operator station.

4.3 CPU Management

Physical CPUs are represented in a domain by logical processors. There must be one physical CPU for each logical processor a domain uses at any one time. A Millennium domain can have up to eight logical processors to accommodate the maximum Millennium CPU configuration.

Defining a domain includes specifying how many logical processors the domain may have and whether the logical processors will own their CPUs exclusively or share a pool of resources with logical processors of other domains. Dedicating a CPU to a logical processor provides the lowest overhead of any of the ways that CPU resources can be allocated. Sharing CPUs among logical processors has greater overhead, but can be a more effective use of CPU resources in that a domain's CPU share can be more precisely matched to its needs at that instant. With MDF, there are three options for allocating CPU resources:

- Domains with dedicated CPUs. CPUs can be dedicated to each of a domain's logical processors. Because logical processors do not share their CPUs with other logical processors, partitioning overhead is at an absolute minimum.
- Domains with shared CPUs. Logical processors share a pool of CPUs with other logical processors according to customer-specified weights.
- Domains with both dedicated and shared CPUs. MDF allows dedicated and shared allocations to be combined in the same domain. One or more CPUs can be dedicated to specific logical processors in the domain, while the domain's other logical processors contend for CPU resources according to their weights. The capability to mix shared and dedicated CPU allocations within a domain has very important benefits in terms of overhead. A domain that has CPUs dedicated to some logical processors enjoys scheduling overhead that is negligible for these logical processors, while retaining the advantage of dynamic expansion of CPU share through its other logical processors (figure 4-1).

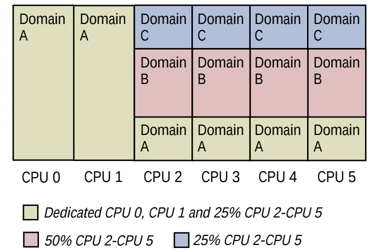


Figure 4-1. Mixed allocation example.

Processing Weights. In a shared CPU environment, the share of the CPU pool given to any one domain depends mainly on customer-supplied processing weights. Weights are assigned by domain and apply to all logical processors in the domain that share the CPU pool with other domains, but are irrelevant to a logical processor with a dedicated CPU.

For example, in a three domain configuration, if Domains 1, 2, and 3 have weights of 10 each, they each get a third of the CPU resources in the CPU pool. However, if the weight of Domain 1 is increased to 20, while the other two remain at 10, then Domain 1 gets 50 percent of the CPU pool and the others 25 percent each. These weights determine CPU share relative to other active domains. If a domain becomes inactive, the shares given to the remaining domains increase.

The share actually given to a domain can increase and decrease dynamically depending on the CPU demands of the various domains relative to their nominal shares. If three domains have weights giving each a third of the CPU pool, but one domain does not need that much, the other domains' actual share may increase beyond their nominal shares, while the relatively inactive domain's actual share declines.

Capping. While varying a domain's CPU share dynamically usually makes best use of a CPU, in particular cases customers may wish to rigorously limit the amount of CPU time that a domain can receive in order to, for example, reserve capacity for the expansion of a new application about to come online. To limit dynamic expansion of CPU share, the weight of a domain can be "capped" to not expand when the domain would normally be eligible for more CPU time. For example, if three domains have equal weights, they all get a third of the CPU resources in the CPU pool. If one domain of the three is capped (and the other domains are not), then its share cannot expand with increased demand beyond a third of the CPU pool while the shares of the other two can.

Max Capping. To provide an absolute upper bound on CPU usage, MDF offers the Max Capping option. Max Capping guarantees that a specified percentage of the CPU will not be exceeded, regardless of the weights of the other active domains. This is an advantage over capping, because capping cannot absolutely guarantee a limit to the CPU usage of a domain. To continue the previous example, if three domains have equal weights, they each receive a third of the CPU pool. If one of these domains is deactivated, the two remaining domains each get half of the CPU pool. Capping a domain would prevent its allocation from expanding beyond a third of the CPU pool in the first case and half in the second. In other words, capping provides a limit relative to other active domains. Max Capping, however, provides an absolute limit; the domain could be limited to a third of the CPU pool regardless of the number and weights of other domains.

CPU Scheduling. When domains share a CPU pool, the share that any one domain gets is determined by customer-supplied domain weights and the current CPU demand generated by each active domain. When “Do Not End Timeslice” is specified on the reset profile, domains are scheduled using timeslices based on their calculated share of the CPU pool. In this case, a domain’s logical processors do not relinquish control of their CPUs when they go into a wait state; they only give up control when their calculated timeslice ends.

When “Do Not End Timeslice” is not specified, a domain’s logical processors relinquish control of their CPUs when they enter a wait state, making their CPUs available to other domains. However, if the domain has not received up to its share, the scheduler may preempt other domains when the domain is ready to process again. This option is more event-driven than the first option and can be more responsive to peak demands, while incurring somewhat more overhead than the other method. The size of the timeslice used in either case can be determined dynamically or can be specified by the customer.

4.4 I/O Management

MDF uses one system-wide, LPAR-compatible IOCDS that defines the I/O configuration for all domains that can be active at one time. (See also section 3.4, “I/O Configuration.”) MDF supports Dynamic Reconfiguration Management (DRM), with which any IOCDS can be dynamically altered, even the active IOCDS with active domains. MDF also supports EMIF, allowing all active domains to share ESCON channels, coupling links, and OSA features with other domains.

4.5 Storage Management

All Millennium 500 Series models offer up to 8 GB of processor storage, and all Millennium 400 Series models offer up to 4 GB of processor storage. Processor storage can be configured completely as main storage or can be configured as a combination of main and expanded storage. Although current SCPs do not support more than 2 GB of main storage, with MDF, larger main storage configurations can be utilized by running multiple domains and partitioning the storage among them.

Main storage is assigned to a domain in increments of 4 MB and is allocated contiguously. Expanded storage can be allocated to domains in increments of 64 MB. MDF allocates storage when it activates a domain and releases storage when it deactivates a domain.

Domain storage is secure; hardware ensures that a domain cannot access the storage of another domain.

MDF Dynamic Storage Reconfiguration (DSR) allows unused main and expanded storage to be reallocated to an active domain. Usually all available main and expanded storage will have been allocated among a server’s domains, but when a domain is deactivated, its main storage and expanded storage can be reassigned to another domain. The main storage being reassigned from the deactivated domain must be contiguous with and higher in physical addressing than the main storage already assigned to the receiving domain. Similarly, the expanded storage being reallocated must be contiguous with and higher in physical addressing than the expanded storage already assigned to the receiving domain. Figure 4-2 shows this process.

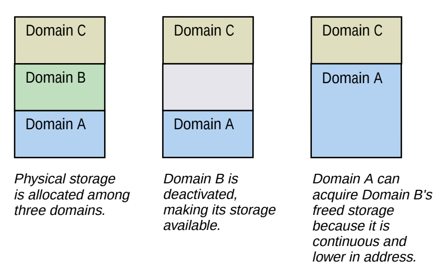


Figure 4-2. Dynamic Storage Reconfiguration example.

MDF supports ARF, which helps automate the movement of memory among domains and is particularly useful for managing system outages within a domain.

4.6 Applications for Logical Partitioning

As the pioneering logical partitioning system, MDF enabled the development of many applications that are now widely used in the industry. These fall into several broad categories.

Processor consolidation. The consolidation of multiple processor complexes into one processor complex is the most widely used application of logical partitioning. With up to 8 CPUs, 8 GB of combined main and expanded storage, and 256 channels, a Millennium Global Server has the capacity to consolidate workloads from smaller systems into separate Millennium domains.

Consolidation of multiple workloads onto one processor complex provides economies of scale, easier management, and potentially large software cost savings. The more workloads that are consolidated, the greater the cost savings, especially for software. The capacity of the consolidated system can be utilized much more efficiently and flexibly than can the capacity of separate physical systems because the reserve capacity of the consolidated system can be applied as needed to any of its workloads. Thus, the formerly separate systems can have, in effect, elastic capacity to draw upon for peak loads, requiring less aggregate reserve capacity.

Test, development, and migration. Setting up separate logical partitions for test and development and for software and application migration is another significant application group. The resource isolation and security attributes of MDF allow testing and development to occur side by side with production systems. MDF's flexible management and resource reallocation features make conversion from test to production easier.

Contingent processes. The ability of MDF to dynamically modify CPU allocations enables a class of applications that must respond to contingencies. Communications management configuration systems that must respond to network changes are one example. Typically such systems require little CPU resource most of the time but have dramatically larger resource needs at critical junctures. Parallel Sysplex environments, where one node may have to rapidly absorb work from another node, and hot standby systems are other examples.

Utility systems. In utility domains, devices can be tested before they are transferred to a production system, or operators can be trained using hands-on procedures in isolation from production systems.

4.7 Year 2000 Application Testing

The clock for each Millennium domain can be set completely independently of the clocks of other domains. A domain can have a clock set to the year 2000 or later for application testing under conditions of the transition from year 1999 to 2000. Applications can thus be tested for year 2000 date rollover in one or more domains while other production domains run with current clocks.

4.8 Performance and Tuning

Part of managing a processor complex is performance measurement, tuning, and capacity planning. MDF provides facilities to support these functions.

4.8.1 Machine Readable Scheduler Data

MRSD recording is a facility used for the retrieval of processor utilization data on Amdahl processors. System performance monitors can collect performance and configuration data through MRSD. MRSD returns the following data:

- Timeslice data (contains operator-specified timeslice value, if specified)
- Number of physical processors
- Number of configured domains
- Domain information including:
  - Domain number and name
  - Number of logical processors
  - General logical processor data, such as CAP and WEIGHT
  - Storage configuration information
  - CPU busy information
  - Channel busy information

MRSD data can be retrieved through the Request Machine Interface (RMI) instruction. Similar performance data is available through the LPAR Interface Facility, which provides an LPAR-compatible means for retrieving information on domains. SCPs and performance monitors can issue the same Diagnose instructions used in an LPAR environment to get information about domain operation.

4.8.2 Amdahl Performance Analysis Facility

APAF reports on the full set of MRSD information. APAF collects and reports configuration and performance information on Millennium servers and other Amdahl processor complexes. It supports both short-term tuning and long-term capacity planning. APAF is supported for the MVS/ESA, OS/390, and UTS operating systems.

Data recording and filtering. APAF collects information on domain and system activity through the RMI instruction and records data by storing records in its own private data file or, under MVS/ESA, in a System Management Facility (SMF) data set. Records can be filtered for both recording and reporting using simple Boolean operations or can be combined into summary data.

Data reporting and viewing. APAF data can be viewed in real time with an online monitor or in a report. The APAF Monitor (APAFMON) allows remote online display of all MRSD data. APAFMON can display current processor and domain usage, as well as current configuration, scheduling information, and channel activity.

APAF/MX. APAF/MX is a user component of the Computer Associates MICS system, providing the interface between APAF and MICS. MICS saves data from several sources, including Resource Measurement Facility (RMF) and SMF, as well as from other subsystems, such as IMS/VS and CICS. With APAF/MX, the unique data provided by APAF can be integrated with resource and utilization data from these other sources, all of them stored and managed in the MICS database.

5 System Design and Technology

Millennium servers scale well and deliver predictable performance for applications as CPUs are added to the server by combining fast CPUs with a sophisticated storage system. Each CPU has large Level 1 (L1) and Level 2 (L2) caches, and all CPUs can exchange L2 cache data simultaneously. The large caches and cache data exchange facility greatly reduce demands on processor storage. The large caches not only improve individual CPU performance, but also make performance stable at high rates of CPU utilization by minimizing storage queuing effects. Because storage queuing effects are a major component of multiprocessing overhead, the large caches also improve multiprocessing performance.

Millennium servers employ a 0.35-micron CMOS technology and are consequently very compact with low power and cooling requirements. CMOS technology is also highly reliable, a characteristic described in detail in chapter 6, “Availability.”

5.1 Millennium Performance

Millennium individual CPU performance derives from advanced CMOS technology, from system design, including multiple pipelines, pipeline aides such as branch instruction target prediction, and, in particular, from large L1 caches and very large L2 caches that minimize the need for storage access outside of the CPU. Each Millennium CPU has a 256-KB L1 cache and an 8-MB L2 cache. All CPU caches together form an access layer that is searched before storage is accessed. When storage requests cannot be satisfied in a CPU’s own caches, they will usually be satisfied from other CPUs’ caches rather than from main storage. On a Millennium GS585, which has an aggregate of 64 MB of L2 cache, typically three out of four storage requests outside a CPU are satisfied this way. Consequently, the load on storage is very low, only about a tenth of that on Amdahl’s 5995M processor. The low storage load results in low queuing effects on memory, which in turn provide several benefits for Millennium performance. In particular, Millennium servers have:

- Stable performance at high processor utilization
- Good scalability
- Highly effective individual CPU performance

5.1.1 Stable Performance at High Utilizations

In multiprocessing environments, as CPU utilization increases, typically so do storage queuing delays and overhead to manage the multiprocessing environment and ensure data integrity. These effects are reflected in increased CPU time per transaction, whether it is charged to the task or not, and in degraded transaction performance. Thus, many data centers limit processor utilization to maintain committed service levels. Because of performance degradation at high utilization, a CPU’s effective capacity under load is typically less than its nominal or rated capacity (usually expressed in millions of instructions per second or MIPS).

Performance at different levels of processor utilization can be measured by internal throughput rate (ITR), which is defined as the number of units of work processed divided by processor utilization or busy time*. To gauge the degree of performance degradation as utilization increases, ITR can be restated in terms of throughput performed per CPU second at different levels of CPU utilization. In most systems, ITR is usually stable up to a certain level, but then decreases as storage queuing and other multiprocessing overhead increases.

Amdahl processing systems have typically provided stable performance at high levels of processor utilization, but Millennium servers take this characteristic a step further. Millennium servers have ITRs that decline very little even at the highest server utilizations (figure 5-1). Thus, customers can operate Millennium servers at high

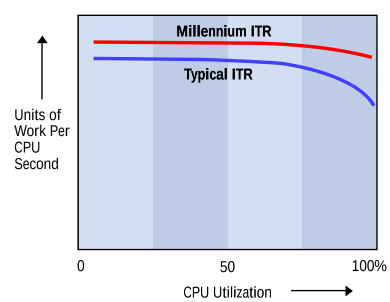


Figure 5-1. Internal Throughput Rate (ITR).

utilizations without suffering transaction degradation due to processor bottlenecks. This charac-teristic has the practical effect of making more of the capacity of the server usable. The Millennium storage system is the most important factor in sustaining stable ITR as CPU utilization increases.

* *IBM-Large Systems Performance Reference.* SC28-1187-01, page 12.

5.1.2 Scalability

The storage system also reduces variability in performance as more CPUs are added, yielding good multiprocessing factors. (The ratio of system processing capacity to that of a uniprocessor is called the multiprocessing factor.) Good multiprocessing factors facilitate capacity planning; the capacity gain for each additional CPU is higher and performance is more stable than in processors with less sophisticated storage systems.

5.1.3 Individual CPU Performance

Because single-task performance is limited by the capacity of a single CPU, effective CPU capacity is very important to tasks with heavy CPU demands.* Examples include:

- CICS regions running OLTP transactions, which often consume an entire CPU's capacity
- IMS/VS control regions using nearly the capacity of an entire CPU
- Batch jobs that must process during an increasingly narrow batch window

These tasks can benefit from the highly effective performance of Millennium CPUs.

The storage system improves the performance of individual Millennium CPUs in several ways. Large caches improve individual-CPU performance because more storage requests can be satisfied in a CPU's own caches. They are particularly beneficial to the performance of OLTP and database workloads. However, because the storage system provides good multiprocessing factors and stable performance at high utilizations, in effect, more of the capacity of the individual CPU is usable.

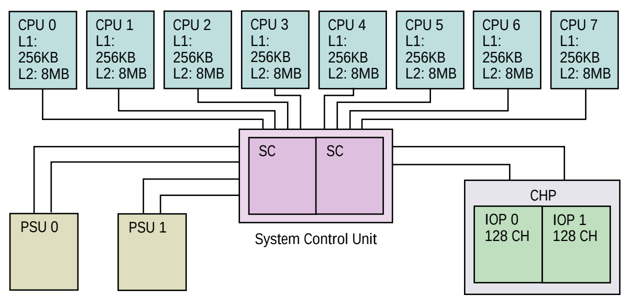
The remainder of this chapter describes Millennium technology and system design in detail.

5.2 Functional Unit Summary

Millennium Global Servers have the following functional units (figure 5-2):

- Central processing units (CPUs) process instrucons and interrupts. Each Millennium CPU has both a large L1 cache and very large L2 cache.
- Processor storage units (PSUs) control access to processor storage, which is divided between main and expanded storage. Each PSU has a maximum capacity of 4 GB of storage.
- Channel processors (CHPs) execute I/O instructions received from the CPU and perform all transfer of data between the server and I/O devices.
- The service processor (SVP) is connected to all other units through the system console interface (SCI) and provides an interface for operator control of the server. The SVP is the focal point for maintaining the server and for collecting information about its operation. Operator stations are connected to the SVP through a LAN.
- The system control unit (SCU) coordinates all communication and data transfer among CPUs, and in particular among the L2 caches of CPUs, as well as communication among the PSUs, the SVP, and CHPs.
- MDF controls logical partitioning, domain scheduling, and the coordination of error recovery operations with SCPs running in domains.

CHP = Channel Processor
CPU = Central Processing Unit
IOP = I/O Processor
L1 = Level 1 Cache
L2 = Level 2 Cache
PSU = Processor Storage Unit
SC = System Controller
SDS = System Data Switch



MDF is described in chapter 4, "Multiple Domain Feature." Section 2.7, "Operational Features," describes operator stations, the Hardware Management Console Interface Facility, and other operational features. Other functional units are described in this chapter.

5.3 Technology

Millennium servers use CMOS technology as both their primary and secondary logic technologies.

5.3.1 Multichip Modules

The primary Millennium logic technology is 0.35-micron CMOS, which allows up to 800,000 gates to be integrated on a single chip with a 16 mm^2 (0.63 in^2) base. The average gate delay time is 190 picoseconds. On-chip RAM has an access time of 2.7 ns and is used for L1 cache and control storage.

Logic chips are mounted on an air-cooled multichip module (MCM), which is attached to cooling fins (figure 5-3). The MCM is fabricated with advanced

* IBM-Washington Systems Center FLASH 9505, "MVS Performance, Capacity Planning Considerations for 9672-RXX Processors," February 1995.

high-density multilevel thin film MCM-D technology. It has 6 thin film layers on an aluminum nitride substrate. An MCM generally contains a functional unit. Each CPU, for example, is contained on an MCM. The size of an MCM varies according to the number of signal pins and chips that it contains. The largest MCMs are the two MCMs that make up the SCU, each measuring 97.88 mm^2 (3.9 in^2). One, the system controller (SC), is for control, and the other, the system data switch (SDS), is for data transfer. The SDS has 5 logic chips, but 2,150 I/O signal pins, the largest number of pins of any MCM. The CPU MCM has 8 chips, the largest number of any MCM, but only 550 pins. It measures 67.4 mm^2 (2.7 in^2).

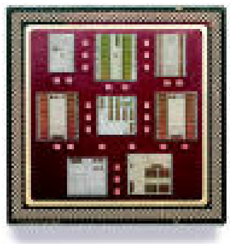


Figure 5-3. Multichip Module, mounting side up.

5.3.2 Storage Technology

On 500 Series servers with 4 GB or less of processor storage and on 400 Series servers with 2 GB or less, storage is composed of 16-megabit (Mb) synchronous dynamic RAMs (SDRAMs) with an access time of 60 ns. Forty RAMs, 20 to a side, are mounted on each processor storage card, comprising 64 MB of storage with its error-checking and correction (ECC). On 500 Series servers with greater than 4 GB of processor storage and on 400 Series servers with greater than 2 GB of processor storage, storage is composed of 64-Mb SDRAMs with an access time of 60 ns. Twenty 64-Mb SDRAMs are mounted on a processor storage card, comprising 128 MB of storage with its ECC.

L2 cache memory is composed of two-megabit self-timed RAMs with an access time of 15 ns. Twenty RAMs are mounted on each L2 cache card to make 4 MB of L2 cache with its ECC. Each L2 cache for a CPU consists of 2 L2 cache cards.

5.3.3 System Board

MCMs, processor storage cards, and L2 cache cards are mounted on the system board. The system board is a high-density, 31-layer, polyimide printed circuit board measuring 470 mm by 540 mm (18.5 in by 21.3 in) with a thickness of 5.4 mm (0.21 in). It attaches up to 14 MCMs, 64 processor storage cards (providing either 4 GB or 8 GB of processor storage), and 16 L2 cache cards (each CPU has two cards or 8 MB) on a single side of the board (figure 5-4).

There is one system board per server. On the 500 Series, either five or eight CPUs, and on the 400 Series, two CPUs, along with their L2 caches, processor storage, and most other functional units are mounted on a single system board. Processor storage cards are mounted on the system board in four sections on the 500 Series, which have two PSUs, and in two sections on the 400 Series, which have one PSU. To minimize access time, L2 cache cards are mounted on the system board among processor storage cards to be in close proximity to their respective CPUs. The system board also mounts alternate storage RAM modules (see section 6.3.2, “Patrol: Maintaining Memory in an Error-Free Condition”) and key storage RAM. Figure 5-4 shows the system board with MCMs, L2 cache, and storage mounted.

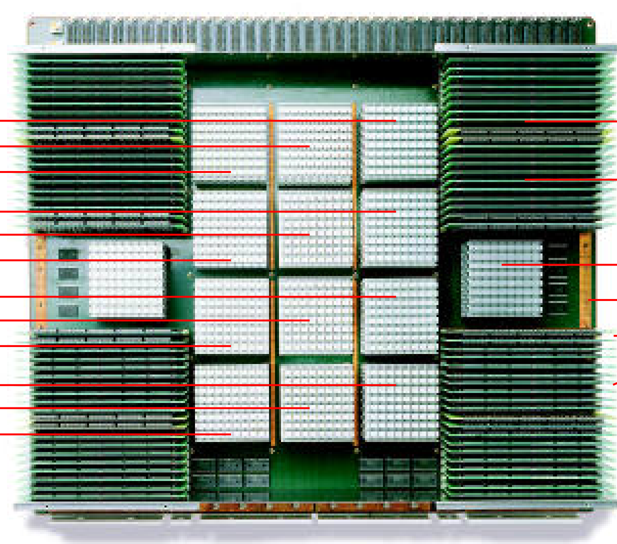


Figure 5-4. System Board Layout.

5.3.4 Secondary Logic Technology

The technology used for logic chips on channel cards and in the SVP and SCI is 0.5-micron CMOS with a maximum integration of 650,000 gates on a 16 mm^2 (0.63 in^2) base. The average loaded gate delay is 260 ps.

5.4 Instruction Processing

A Millennium CPU has four units (figure 5-5):

- The instruction unit (I-unit) manages instruction and interrupt processing.
- The execution unit (E-unit) executes operations under direction of the I-unit.
- The storage unit (S-unit) reads and writes L1 cache as requested by the I-unit.
- The L2 cache unit reads and writes L2 cache as requested by the S-unit or SCU.

Most instructions are executed in the I-unit and E-unit. I/O instructions are executed by the channel processor (CHP).

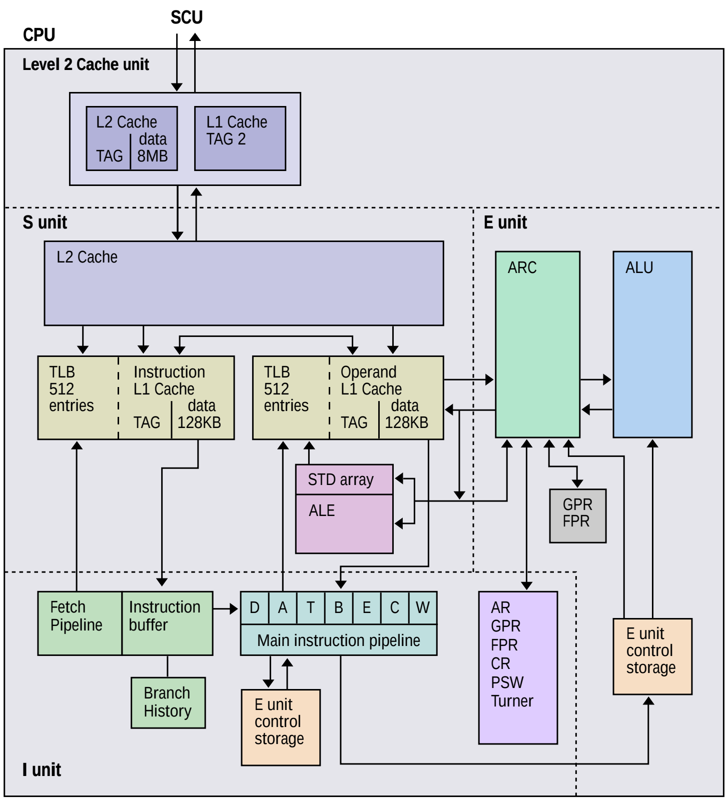
5.4.1 Instruction Unit

The I-unit manages instruction and interrupt processing, controlling the processing sequence for the entire CPU. It requests instructions and operands from the S-unit and sends instruction results as operands back to the S-unit. The I-unit delegates execution of most instructions to the E-unit.

The I-unit has two pipelines—the instruction fetch pipeline and the main instruction pipeline—that run asynchronously to one another, but are adjuncts to each other. The object of the instruction fetch pipeline is to always have an instruction ready for the main instruction pipeline. The function of the main instruction pipeline is to maximize the rate of instruction execution. (These are the most important pipelines in the server and the only ones described in this document. However, pipelines are used throughout the server to improve performance. The SCU, for example, has two pipelines.)

Instruction Fetch Pipeline. The instruction fetch pipeline requests instructions from the S-unit. The instruction fetch pipeline anticipates the three most likely paths that instruction execution will take and prefetches instructions into its *instruction buffer*, which accordingly has three 32-byte sections. One section contains the next sequential instructions, which are executed if no branches are taken. The other two sections have instructions that are the targets of branch instructions in the instruction stream. If a branch is taken to an instruction in one of these two sections, that section becomes the next sequential instruction stream, and new target instructions are prefetched into the other two sections. Thus, the instruction buffer section that is currently presenting instructions to the main instruction pipeline switches as branches are taken.

The instruction fetch pipeline determines whether to prefetch a particular branch's target instructions with the aid of *branch history*, a table with 1,024 entries. When a branch is taken, the major part of the branch instruction's address is stored in branch history. When a branch instruction is encountered in the instruction stream and its address is found in branch history, the instruction fetch pipeline anticipates that this branch will be taken and prefetches its target instructions into one of the instruction buffer sections. If a branch recorded in branch history is



ALB = Access Register Translation Lookaside Buffer
ALU = Arithmetic Logic Unit
AR = Access Register
ARC = Addressable Register Coupler
CR = Control Register
FPR = Floating Point Register
GPR = General-Purpose Register
PSW = Program Status Word
SCU = System Control Unit
STD = Segment Table Destination
TLB = Translation Lookaside Buffer

Figure 5-5. Central Processing Unit.

encountered three times in succession, but not taken, its branch history entry is invalidated.

Main Instruction Pipeline. The main instruction pipeline fetches instructions from the instruction buffer and processes them in seven stages.

1. Decode (D): The I-unit decodes the instruction, reads control storage to set up execution of the instruction, and reads the operand registers.

2. Address (A): The I-unit calculates the instruction operand addresses.

3. Translate (T): The S-unit references the TLB and L1 cache address section (TAG) to check whether the instruction's operands are in the L1 operand cache.

4. Buffer (B): The S-unit reads the operands from L1 operand cache and sends them to the E-unit.

5. Execution (E): The E-unit executes the instruction. This stage has a stack structure that allows three floating-point instructions to be in overlapped execution in the E-stage simultaneously, improving floating-point performance.

6. Check (C): Checking is done before results are written. Interrupts are processed in this stage.

7. Write (W): Results are written.

In pipelined operation, pipeline stages are overlapped for successive instructions, and seven stages are likely to be active at once for seven different instructions. As one instruction moves from the D to the A stage, another instruction enters the D stage, and so on. If one stage is delayed—for example, if execution takes more than one cycle or operand data is not in L1 cache—then the pipeline completes the current stage for each instruction before advancing all instructions in the pipeline to the next stage.

Bypassing. When an instruction uses the result of another instruction preceding it in the pipeline, the pipeline may pass the result directly back to the requesting instruction. Because the result is directly available, the requesting instruction does not have to wait for the results to be written to cache or registers and read back again. Bypassing these operations improves the effective speed of the pipeline.

Other facilities. The I-unit also has control registers (CRs), access registers (ARs), the program status word (PSW), and CPU timing facilities and retains copies of the general registers (GRs) and floating-point registers (FPRs), which are resident in the E-Unit.

5.4.2 Execution Unit

The E-unit is a microprogrammed unit that executes operations specified by the I-unit. The E-unit has the following components:

Control storage. The E-unit microcode is divided between two types of control storage. The kernel control storage contains the microprogram that controls E-unit execution, controlling data paths and directing the arithmetic units. Service control storage contains the error recovery microcode, including the command retry and the E-unit internal control storage patrol function.

Arithmetic units. The E-unit has six types of arithmetic units:

- A general-purpose 8-byte adder-subtracter
- A general-purpose 8-byte shifter
- A high-speed decimal arithmetic unit
- An 8-byte floating-point unit
- A high-speed multiplier to do 8-byte fixed-point or floating-point multiplication and to provide for pipelining of arithmetic operations
- A high-speed divider for 8-byte fixed-point or floating-point division

Register stack. The E-unit contains the general registers, the floating-point registers, and other internal registers.

Addressable register queues. There are eight addressable register queues (ARQs) used for operating on data and as buffers to stage data between the E-unit and I-unit. The I-unit or S-unit sends operand data to an ARQ, where it is processed according to the operation specified by the I-unit during the E-stage of the main instruction pipeline. Results are again stored in an ARQ and returned to the I-unit.

5.5 Storage System

Instruction execution is supported by the Millennium storage system, which includes the L1 and L2 caches of each CPU, processor storage, and the functional units of the storage system.

Level 1 cache. Each CPU has 256 KB of L1 cache divided into 128 KB of L1 instruction cache, which provides instructions to the instruction fetch pipeline, and 128 KB of operand cache, which supports the main instruction pipeline. This specialization improves the effectiveness of L1 cache.

Level 2 cache. Each CPU has 8 MB of L2 cache that is searched to satisfy a data request when the data is not in its L1 cache. If a CPU's L2 cache does not contain the requested data, the SCU determines if the data is in another CPU's cache before main storage is accessed. In the aggregate, a Millennium GS585 has 64 MB of L2 cache. These advanced cache designs ensure the consistent performance and scalability of Millennium servers.

Processor storage. A PSU can have up to 4 GB of processor storage. Servers in the 400 Series each have one PSU, and servers in the 500 Series each have two PSUs, for a maximum storage capacity of 8 GB.

5.5.1 Storage System Functional Units

The storage system is composed of the following functional units (figure 5-6):

- The S-unit, which contains L1 cache and is a component of each CPU
- The L2 cache unit, which contains L2 cache and is a component of each CPU
- The SCU, which coordinates L2 cache and processor storage access among all CPUs and I/O processors
- The PSUs, which control access to main and expanded storage

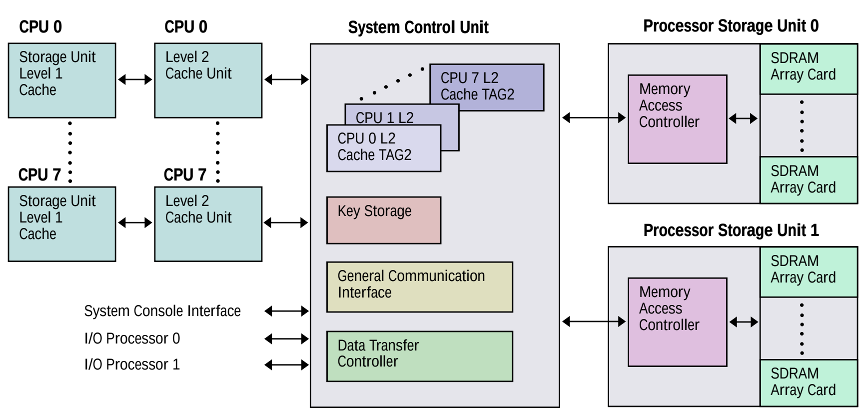


Figure 5-6. Storage System Functional Units.

Storage Unit. The S-unit provides instructions and operands to the I-unit and writes operands for the I-unit. The S-unit satisfies storage requests from L1 cache, if possible, or makes requests of the L2 cache unit, if data cannot be found in L1 cache. The S-unit has several buffers to stage data between the S-unit and the L2 cache unit and buffer invalidation requests from the L2 cache unit.

The S-unit also contains dynamic address translation facilities. To avoid retranslating access registers, the S-unit retains the segment table designation (STD) for address spaces in the ALB. The S-unit retains translated page addresses in the TLB. As with L1 cache, the TLB is divided into an instruction TLB and an operand TLB, each with 512 entries.

Level 2 Cache Unit. The level 2 cache unit manages the CPU's L2 cache, which is an extra level of cache between the CPU's L1 cache and main storage. If data requested by the S-unit is not in the CPU's own L2 cache, the L2 cache unit passes the request to the SCU.

System Control Unit. The SCU interconnects all functional units. Data transfer among CPUs, IOPs, and processor storage and between main storage and expanded storage passes through the SDS section of the SCU and is controlled by the SC section of the SCU. The SCU also has the following sections.

The *data transfer controller* controls data transfer between processor storage addresses, whether the processor storage location is defined as main or expanded storage. Data transfer is performed synchronously, if initiated by a CPU, or asynchronously if initiated by an IOP. The data transfer controller also executes 4-KB page moves between main and expanded storage, which are both operator-defined sections of processor storage.

The *general communication interface* (GCI) controls communication between CPUs, between IOPs, and between CPUs and IOPs. Communications that use the GCI include SIGP, I/O instructions, I/O interrupts, external timer, and service signal control.

The SCU also contains *key storage*. There is a one-byte storage protection key for each 4 kilobytes of processor storage, and each key is triplicated for increased reliability.

Processor Storage Unit. There is one PSU in 400 Series servers and two in 500 Series servers. A PSU consists of a memory access controller, housed on a single MCM, and a storage array, consisting of up to 32 SDRAM cards. If 16-Mb SDRAMs are used, a PSU has up to 2 GB of storage. If 64-Mb SDRAMs are used, a PSU has up to 4 GB of storage. The PSU's memory access controller executes storage read and write requests from the SCU. (Processor storage ECC, storage patrol, and alternate storage RAMs are described in section 6.3, "Processor Storage Recovery.")

5.5.2 Storage Organization and Management

Processor storage organization. The storage array of each PSU has up to 16 sections or ways, for interleaving, providing overlapped access to storage. A PSU has up to 16 ways, providing up to 16-way interleaving for 400 Series servers. Interleaving crosses PSUs; therefore, 500 Series servers, which have two PSUs, have up to 32-way interleaving. Storage is always configured for maximum interleaving.

Cache mapping algorithm. The algorithm for mapping storage into L1 or L2 cache is set-associative, which means that a set of main storage addresses with certain address bits in common is associated with the same set of locations in the cache.

- The 8 MB of L2 cache in each CPU is divided into four 2-MB ways. Each way is arranged into 4,096 512-byte blocks, and each 512-byte block is divided into 64-byte subblocks. A main storage block can be stored in any one of these four ways at the same relative offset as from the start of its 2-MB main storage block.
- L1 instruction and L1 operand caches each have 2,048 64-byte lines divided into four 32-KB ways. A main storage subblock can be stored in any of one of these four ways at the same relative offset as from the start of a 32-KB storage section.

Data Transfer. Data transfer between L2 and L1 cache is in 64-byte subblocks. Data transfer between the SCU and main storage, between L2 cache and the SCU, and between one L2 cache and another can be in either 64-byte subblocks or 512-byte blocks.

Cache coordination. A CPU's L2 cache must be coordinated with its L1 cache, with other CPUs' L2 caches, and with main storage. To keep track of storage blocks and subblocks stored in cache, each cache has an address and control information section, called TAG. For example, an L2 cache unit has L2 cache TAG that holds the addresses of data in the L2 cache, as well as other control information, including whether the block is shared with other CPUs or is used exclusively by this CPU, whether the block has been changed and so must be moved out to storage when it is displaced by another block, and whether the block has been invalidated.

Coordinating a CPU's L1 and L2 caches is done by the L2 cache unit. Each L2 cache unit has a copy of the TAG for its corresponding L1 cache. (These copies are called L1 cache TAG2.)

Coordinating the L2 caches in different CPUs is an SCU function; to do so, the SCU retains copies of each CPU's L2 cache TAG. (These copies are called L2 cache TAG2.) When a CPU or IOP makes a storage request, the SCU uses its TAG2 copies to determine whether it can satisfy the request from a CPU's L2 cache or whether it must access main storage. Storage consistency and data integrity across each CPU's caches and main storage are maintained by the SCU in conjunction with the L2 cache units and S-units.

5.5.3 Storage Access

From the standpoint of a particular CPU, a storage search proceeds from its own L1 cache, to its own L2 cache, to the L2 caches of other CPUs, and finally to main storage. This search sequence is described below and shown in figure 5-7:

1. The S-Unit first tries to satisfy a storage request from the L1 instruction cache or the L1 operand cache. Most requests are satisfied at this level.

2. If the data is not in L1 cache, the S-unit requests it from the L2 cache unit. If the data is in L2 cache, it is sent to the appropriate L1 cache. When a CPU accesses its own L1 or L2 cache, it generates no multiprocessing effects. Moreover, the large caches minimize the need to access storage blocks outside of the CPU.

3. If a CPU's own L2 cache doesn't have the data, its L2 cache unit requests it from the SCU. The SCU references its L2 cache TAG2 copies to determine whether the block is in another CPU's L2 cache and, if so, retrieves it from there rather than from main storage. L2 cache data exchange can occur among all CPUs at once, because each CPU has its own data path to the

SCU. For multiprocessor models, if a block is not found in a CPU's own cache system, the most likely place to find it is in another CPU's cache system. On a Millennium GS585, when a block is not in a CPU's own L2 cache, it can typically be found three out of four times in another CPU's L2 or L1 cache.

4. If no L2 cache has the data, the SCU requests it from the PSU or main storage.

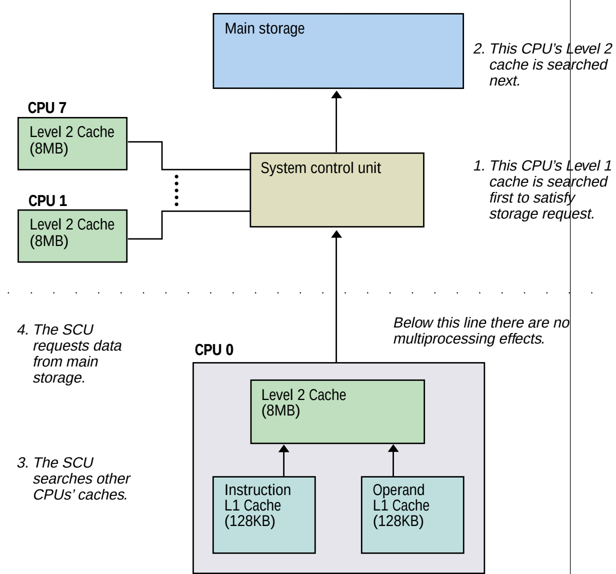


Figure 5-7. Storage hierarchy search sequence.

Cache store method. The store method for both L1 and L2 caches is store-in. Data is not written to main storage until it is necessary to do so, such as when a block or line is being displaced by another. The store-in method used in both caches reduces data traffic outside the CPU and is another factor reducing the load on main storage. When a new line is stored in a cache, the way that is selected is the one with the least recently used block of data of the four at that offset.

The large caches, the intercache transfer mechanism, and the store-in method combine to markedly reduce the number of accesses to main storage. The local cache system of each CPU improves its performance because so many storage lines are held in the caches that fewer accesses outside the CPU are needed than for CPUs with smaller caches. When access outside a CPU's own caches is required, the low storage load keeps queuing effects to a minimum.

5.6 I/O Processing

A CHP controls I/O instruction execution, interrupt processing, and the data transfer between I/O devices and main storage for up to 64 channels. I/O processing that is independent of channel type is done in an I/O processor (IOP), an MCM on the system board, which contains the general purpose logic to manage I/O operations for up to 128 channels. There is one IOP on servers in the 400 Series and two IOPs on servers in the 500 Series. An IOP has three main sections:

- An IOP has two *channel managers* (CHMs) that each process the I/O operations of up to 64 channels. Under request from the CPU, a CHM starts an I/O operation and selects the channel. A CHM also generates interrupts upon termination of the I/O operation at the device.
- An IOP uses a 32-kiloword *microprogram*, which is stored in the HSA section of main storage and executed in each CHM's control storage.
- A *transfer server* controls main storage access for both CHMs in an IOP.

A channel unit manages that part of the I/O operation that is channel-type dependent. It handles up to 64 channels and is located in the channel backpanels of the channel frame and channel expansion frame. (Channel backpanels attach IOPA cards and channel cards.) A channel unit has these components:

- One or two I/O processor adapters (IOPAs) that each connect up to 4 channel cards or up to 32 channels.
- Channel cards. The number of channels per card varies by the type of channel. Eight serial (ESCON) channels, four parallel channels, two coupling links, or two OSA features can be placed on one card. A channel is microprogrammed to be a specific channel type. For specifications of the types of channels supported and their characteristics, see section 2.3, "I/O Channels," section 2.4.1, "Coupling Links," and section 2.5.1, "Open Systems Adapter Feature."

A channel unit, a CHM, and a transfer server (which is shared between two CHPs) comprise a CHP (figure 5-8).

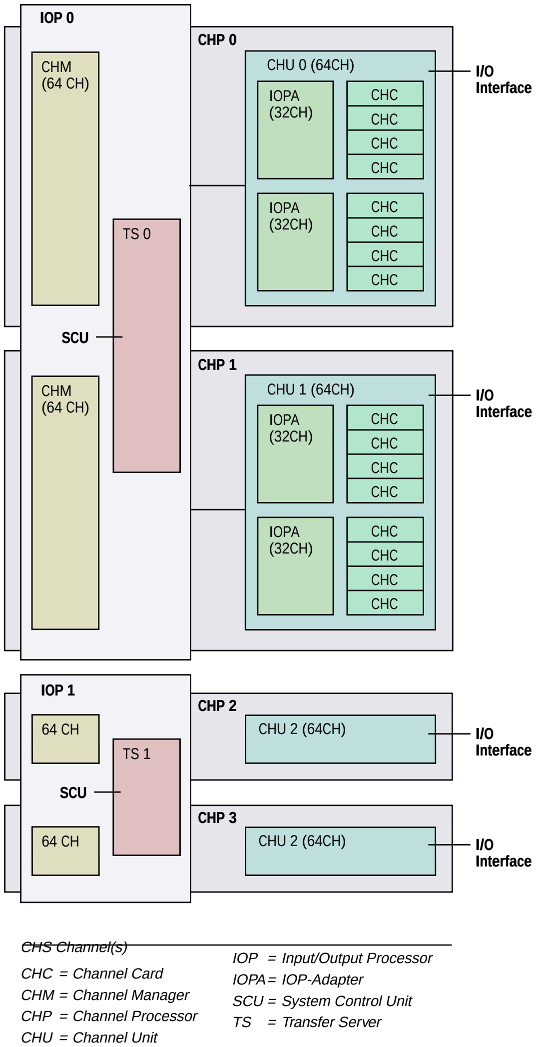


Figure 5-8. Channel processors.

5.7 Service Processor and System Console Interface

The SVP is a computer subsystem that is independent of the rest of the server. The SVP is a fast microprocessor with 16 MB of memory, 32 KB of cache, and two 512-MB hard disks. The first hard disk is used for system operation, the second is a mirrored disk for backup. The SVP has adapters to connect to floppy and hard disks, a display, the LAN for operator stations, the system console interface (SCI), and the system power controller (SPC).

The SVP communicates with other functional units over the SCI. It communicates control information to and receives status from other units, monitors certain functional units, and receives diagnostic information on failures through its scan facility. The SCI receives interrupts from other functional units, mainly related to completed actions performed by another unit at the request of the SVP, error conditions, logouts, and diagnostic information.

The SVP automatically receives scanned diagnostic information, and can automatically communicate them to the WWCSC. The SVP performs the following maintenance and diagnostic functions:

- Using proprietary diagnostic programs on its hard disk, the SVP performs an initial diagnostic check of the server after it is powered on and performs similar checks after maintenance is applied.
- The SVP monitors the power supplies, cooling system, and system environment through the system power controller (SPC).
- When a machine check occurs, the SVP scans the status of all functional units and stores it as part of the error logout on the hard disk.
- Using its automated scan-out analysis program (ASOA), it analyzes the logout and determines the failing component.
- It can automatically transmit the logout, or information about out-of-specification power or environmental conditions, to the WWCSC.

5.8 Power System

The SPC turns on and off system power and monitors power supply status and other system conditions. The system can be powered on or off from the system control panel or by command through the SVP. See also section 2.12, “Power and Cooling.”

5.9 Frames

A Millennium server has two or three frames, depending on channel capacity.

- The *system frame* contains the system board, power supplies, and fans (figure 5-9).
- The *channel frame* has the SVP and SCI, a channel backpanel, and the capacity to connect up to 128 channels (figure 5-9).
- The *channel expansion frame* has a channel backpanel and power supplies and the capacity to connect up to 128 additional channels.

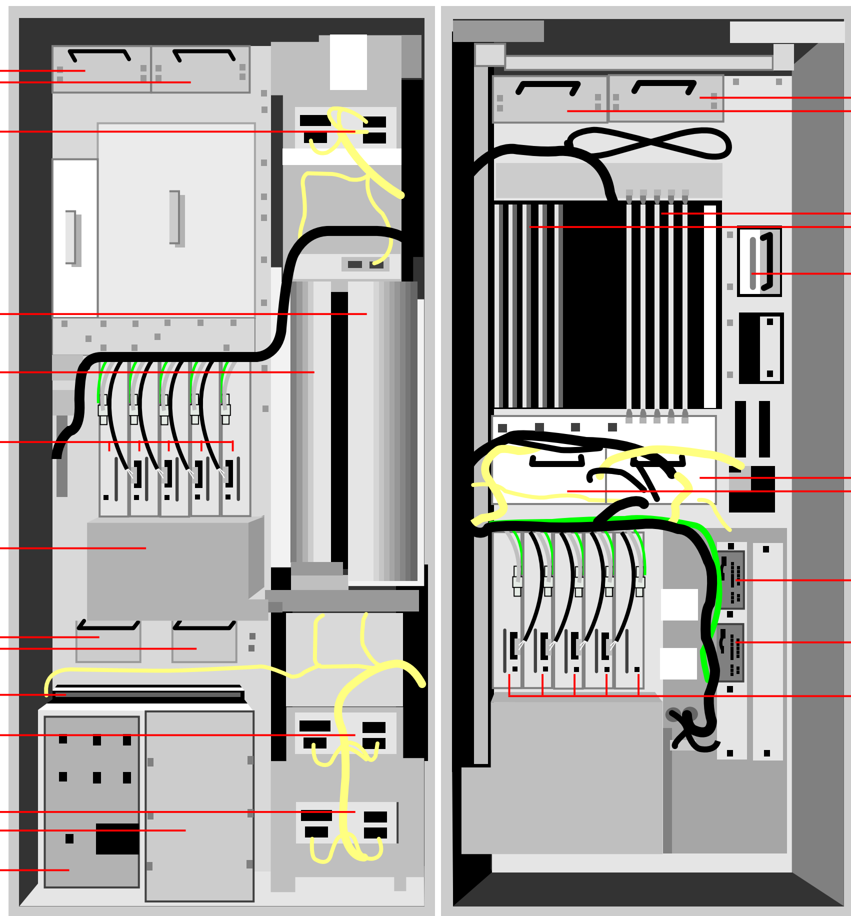


Figure 5-9a. System frame and channel frame, front.

If the system is a Dual Server, a second system board is installed in the system frame, a second SVP and SCI are installed in the channel frame, and the channel expansion frame is always present. See section 2.12, “Dual Server,” for more information.

If the system is configured with native Dual Power Feed, the channel expansion frame is always present. See section 2.14 “Power Source Options” for more information.

A Millennium server is very compact, occupying up to 92 percent less floor space than an ECL system of the same processing capacity. Figure 5-10 shows footprints for different frame configurations.

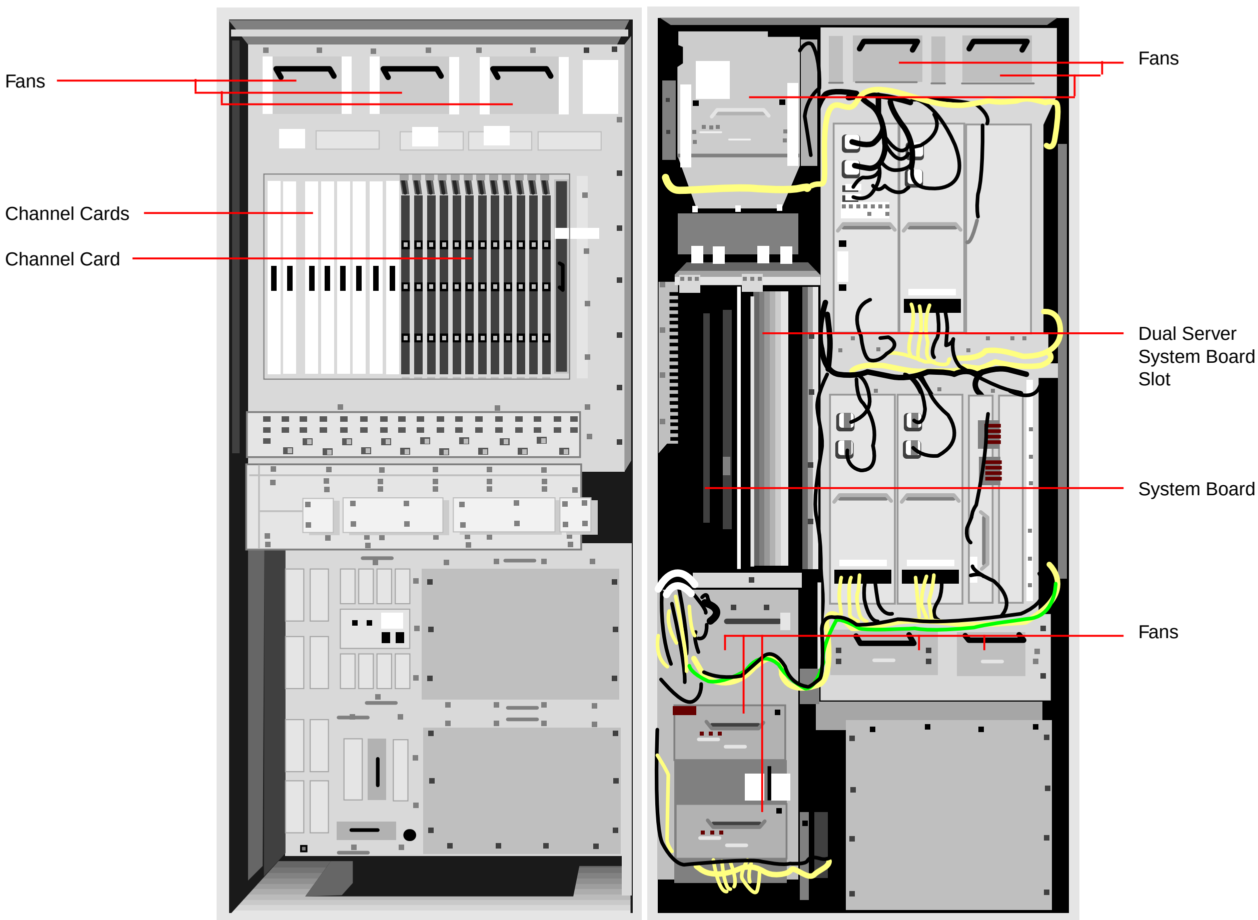


Figure 5-9b. System frame and channel frame, back.

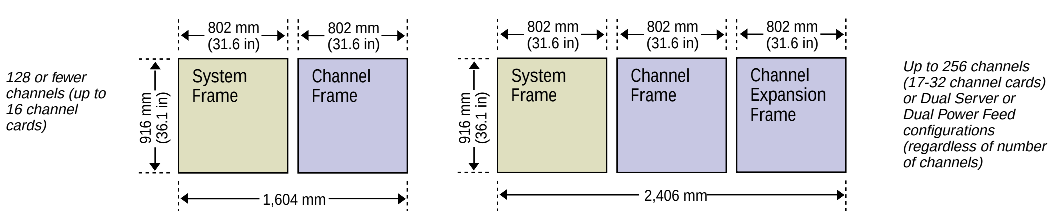


Figure 5-10. Frame footprints.

6 Availability

Millennium servers are distinguished, not simply in their performance and versatility, but also in the high availability they provide. The biggest factor in improving availability over earlier generations of S/390 processors is advanced CMOS technology. But using a reliable technology will not in itself produce the level of availability needed by mission-critical applications. Reliability must be combined with fault tolerance and serviceability features.

The reliability of Millennium's CMOS technology is complemented by mature recovery facilities evolved over several generations of Amdahl mainframes. Hardware errors will occur, particularly in memory systems, but because of the fault tolerance built into Millennium servers, hardware errors will seldom affect system operation.

Millennium server outages will typically be scheduled to perform system upgrades and routine or deferred maintenance. To reduce the incidence of scheduled outages, Millennium servers have ConServ and QuickAvailability features, which provide concurrent maintenance of the most frequently serviced or updated components and can also reduce the duration of scheduled outages for CPU upgrades and repairs.

6.1 Fault Avoidance: Reliability

In the System/390 marketplace, reliability is the term used for fault avoidance. The advanced CMOS technology used in Millennium servers significantly improves hardware reliability over previous generations of S/390 processors. The fundamental difference between CMOS and ECL technologies is that CMOS is an inherently “power off” technology and ECL an inherently “power on” technology. That is, in ECL technology all circuits are “on” all the time—there is always current going through each transistor. In CMOS, only circuits that are switching have current running through them—perhaps 10-12% of the circuits will be active in this way at one time. Thus, a CMOS chip generates less heat and operates at a lower temperature than an ECL chip, resulting in less transistor deterioration in the long run.

CMOS circuit density is much higher than ECL circuit density and this, in itself, leads to greater reliability by reducing the number of components and interconnections in the system. In general, given two comparably designed systems, the system using fewer components will be the more reliable. Whereas an ECL logic chip can have tens of thousands of circuits, a Millennium CMOS chip can have several hundred thousand circuits. For example, a Millennium server has only about a tenth of the number of components of an Amdahl 5995M processor.

Industry experience with CMOS already leads to the expectation of high reliability from Millennium technology. To confirm this expectation, Amdahl uses Accelerated Reliability Testing. To simulate the effects of a lifetime of use, chips are stressed with heat, humidity, and voltage bias. To test transient error recovery, chips are bombarded with alpha particles and neutrons. In these tests, Millennium CMOS technology produced less than one percent, and in some tests only a tenth of one percent, the number of transient errors generated in comparable testing of ECL technology. This testing strongly corroborates expectations of dramatically improved reliability for Millennium semiconductor technology.

6.2 Combining Fault Avoidance with Fault Tolerance

Reliable technology alone will not produce a server that meets the availability needs of mission-critical applications. Fault avoidance through reliability must be combined with fault tolerance that is achieved through recovery processes and redundancy. In designing a unit, both the reliability of component parts and fault tolerance of the component's environment are considered. Where intermittent failures are likely, recovery processes are made extremely robust.

6.3 Processor Storage Recovery

A good example of the complementary nature of Millennium reliability and recovery processes is found in processor storage. Although memory systems are highly reliable and reliability improves with each generation of RAMs, storage errors nonetheless comprise a high percentage of total hardware errors for two reasons. Memory chips are extremely dense and are more susceptible to random failures due to environmental radiation than are transistors in logic chips. The sheer volume of memory chips in the server also increases the likelihood of a memory error occurring. It is typical for memory errors to occur several times a year in a single system. Thus, to reinforce overall system availability, given the nature of memory chips, Millennium servers have particularly robust processor storage recovery, combining facilities from the Amdahl 5995A and 5995M processors.

6.3.1 Error-Checking and Correction

Processor storage has especially powerful ECC. Millennium storage is divided into 80-bit storage blocks, each with its own ECC code. Within each storage block there are 20 4-bit sections, where each 4-bit section is provided by a different storage RAM. Any combination of errors within a single 4-bit section, or nibble, can be corrected and any combination of errors in two 4-bit sections of a block can be detected. The execution of ECC code entails no performance penalty whatsoever. Any combination of errors within a single 4-bit section of a storage block can be detected and corrected repetitively without performance penalty, so long as there are no errors at the same time in any other 4-bit section of that block. Thus, a single RAM (which provides a 4-bit section of many storage blocks) can catastrophically fail without impacting system integrity or performance. Data on the failing chip is reconstructed from other RAMs and ECC on any fetch, and when newly stored, the data from the failing chip is accounted for by ECC.

6.3.2 Patrol: Maintaining Memory in an Error-Free Condition

The PSU has a patrol function whose object is to maintain processor storage in an error-free condition—in particular, to prevent either a transient error or a permanent error in a storage block from compounding with a later error in the same block to become uncorrectable by ECC. Although any error combination within a nibble (4 bits or one half of a byte) in a block is correctable, errors occurring simultaneously in two different nibbles in a block are not. While even a solidly failing storage chip in itself poses no data integrity problems (because of processor storage ECC), transient errors occurring in different sections of the same 80-bit block at the same time will result in an unrecoverable storage error. To avoid these errors, patrol corrects transient errors with ECC and replaces a permanently failing RAM with an alternate. Each PSU has one alternate 16-Mb RAM that patrol can use to replace a failing 16-Mb RAM or four alternate 16-Mb RAMs that are used to replace one failing 64-Mb RAM.

Patrol uses idle cycles to periodically read through processor storage and check for storage errors. When patrol finds an error in a storage block, it attempts correction using ECC and then writes back the corrected data. Patrol then reads the data again to determine if the error is a transient condition by noting whether its correction attempt succeeded. If the error persists, patrol considers it permanent and dynamically replaces the failing storage chip with an alternate, initializing the alternate chip as part of the replacement process. No system reset is required to effect the replacement, thus avoiding a server outage. By maintaining memory in an error-free condition, patrol greatly improves the chances that any later memory error will be correctable.

6.4 Fault Tolerance in Other Units

Each functional unit has methods of recovery unique to its environment. In general, a unit will correct its data environment and retry a failed operation, making recovery as specific as possible to the error condition and local environment, which helps to both contain the error and to correct it. As recovery techniques have evolved, techniques once used only at higher levels—such as ECC on main storage or instruction retry—have been incorporated at progressively lower levels, providing more precise localized recovery—for example, ECC on control storage and a subunit's operation retry.

6.4.1 Internal Storage Areas

Internal server storage areas are protected with a comprehensive infrastructure of error detection, error correction, operation retry procedures, and unit isolation techniques. The following are examples.

- Most control storage has ECC that detects double-bit errors and corrects single-bit errors. Other internal storage sections have parity error detection. Control storage ECC is supplemented in some cases by other recovery processes.
- The E-Unit and channel processor have control storage patrol functions similar to the processor storage patrol function. Patrol searches for and corrects single-bit control storage errors codes in the same manner as in processor storage. Patrol also reduces the likelihood that single-bit errors will compound into uncorrectable multiple-bit control storage errors.
- The E-Unit also has an exchange register that can replace a solidly failing control storage word in the same way that alternate chips can replace failing RAMs in processor storage.

- In case of an error in control storage or internal caches, some microprograms can be dynamically reloaded.
- The storage unit contains several buffers that speed data retrieval and address translation, such as L1 cache and the TLB. The recovery strategy for these buffers is to internally retry a request after invalidating or purging data from the buffers. If intermittent errors exceed a buffer's internal error threshold, a section of the buffer can be removed from use, allowing operation to continue with the remaining section or sections of the buffer. Similarly, a way in L2 cache can be disconnected when an error threshold is reached.
- The SVP also has single-bit ECC correction and double-bit error correction on internal storage units.
- The SCU has three copies of key storage and uses majority voting to determine the correct key value. An erroneous copy is corrected from the other two.

6.4.2 Retry

Typically operations are retried after a local environment has been corrected. Retry will first be as local as possible —for example, the E-Unit may retry a command or the S-Unit may retry an address translation—and will usually be successful, but if not, recovery can proceed to a higher unit, such as to the I-Unit for instruction retry. These low-level correction facilities are the lowest tiers in a hierarchy of recovery processes. Faults that cannot be recovered by local correction, operation retry, or instruction retry are escalated to MDF, which reflects these errors to the appropriate domain for SCP recovery. Another level of protection may be added to these recovery processes by Parallel Sysplex providing application or SCP recovery on another node.

6.4.3 Redundancy in Auxiliary Systems

Auxiliary systems are not a part of the instruction retry hierarchy. For power, cooling, and other electromechanical units, redundancy is typically used to ensure continuous operation.

- The power system has n + 1 redundancy. If one power supply module fails, the remaining modules supply sufficient power for continuous operation.
- An optional Auxiliary Power Source is available to enhance the availability of the power source. See section 2.13, “Auxiliary Power Source,” for more information.
- The cooling system has excess cooling capacity. Each set of fans includes at least one redundant fan, so that sufficient cooling for normal operation can be provided when one fan fails. Fans are dual speed and will operate at a higher speed to provide additional cooling if a fan fails or if the ambient room temperature increases beyond a specified threshold. If an entire set of fans fails, the remaining sets provide sufficient capacity to continue normal operation.
- The SVP has duplexed, or mirrored, hard disks, so that if one disk fails, the other can take over. A backup copy of SVP software is kept on this disk system. A removable storage device also provides for backup and restore functions.

6.5 QuickSwitch Feature

The QuickSwitch feature provides for the dynamic substitution of an unused CPU for a non-functioning CPU on the same system board. A system board has either five or eight CPUs on 500 Series models and two on 400 Series models. As described in section 2.9, “QuickCapacity Feature,” Millennium servers have CPU Fencing capability. For example, a Millennium GS415 and GS575 each have one CPU in the fenced reserve, whereas a Millennium GS422, GS425, and GS585 have none and therefore do not support the QuickSwitch feature. Under the QuickSwitch feature, CPUs in the fenced reserve are available as in-system backups for failing CPUs.

If a CPU fails and needs to be replaced and there is an available CPU in the fenced reserve, the fenced CPU can substitute for the failed CPU. No outage is required to perform the swap. Substitution is effected dynamically upon the recognition by firmware of a CPU failure. The results of a STORE CPU ID instruction are unchanged after the CPU replacement because the number of CPUs available for use remains the same.

6.6 Concurrent Maintenance

Because of the reliability and robust recovery facilities of Millennium servers, unscheduled outages caused by server failures are minimized. Rather, outages may be scheduled to perform system upgrades, routine server maintenance, or deferred maintenance. To keep scheduled outages to a minimum, the most frequently maintained Millennium components can be serviced concurrently with normal system operation.

Firmware Maintenance. The system component requiring by far the most maintenance and enhancement actions is firmware, which includes microcode, service processor software, and MDF. Firmware commonly accounts for the majority of maintenance actions. Consequently, concurrent maintenance of firmware offers the greatest potential for improving availability by avoiding scheduled outages for maintenance.

Millennium firmware can be concurrently maintained. When firmware is replaced in the server, the entire firmware package is generally replaced at once. This practice allows extensive testing of the firmware package and avoids problems of corequisite and prerequisite maintenance. When firmware maintenance is applied, the previous copy is retained so that the new firmware can be backed off in the unlikely event of problems.

Channel Maintenance. Upgrades and changes to the I/O subsystem are another frequently performed type of maintenance action. On Millennium servers, all maintenance to individual channels and coupling links can be done concurrently with normal operation. In addition, the installation of additional channels and coupling links can also be performed concurrently with normal operation. The number of channels or coupling links that can be added during normal operation depends upon the availability of channel positions in existing IOPAs. The capacity to add channels and coupling links concurrently with normal operation can be increased by installing additional IOPAs. See section 2.10, “QuickChannels Feature,” for additional information.

All channel boards have lights to show which paths are offline and which are online and to coordinate with the SVP which board should be changed. This is an important aid to the field engineer (FE) who is repairing or installing channels concurrently with normal operation.

Complementing the ability to physically add channels during normal operation is the ability to change the IOCDS dynamically, as noted in section 4.4, “I/O Management.”

Service Processor Maintenance. The installation of new SVP software requires a rebooting of the SVP, which occurs while the server continues operation. The previous version of SVP software is retained on the duplicate hard disk to allow the new software version to be replaced by the previous version in case of problems.

The mechanical parts of the SVP—the hard disks, the floppy disks, and fans—can be replaced concurrently with SVP and server operation.

The Millennium LAN and LAN interface can be serviced concurrently with normal operation of other components.

Customer Operator Station Maintenance. Each COS is completely independent of other COSs. Any maintenance to a COS can be done without affecting the operation of any other COS. When software for the HMCIF or other software in the COS must be replaced, the COS PC must be rebooted, but other operator stations and the Millennium server are unaffected. The HMC/NetView connection is unavailable during concurrent maintenance of the primary COS.

Cooling system. Any fan can be replaced concurrently with normal server operation. Air filters can be removed and replaced during normal operation.

6.7 Participation in a Parallel Sysplex

The concurrent maintenance and recovery characteristics of Millennium servers can be enhanced by participation in a Parallel Sysplex environment. For recovery, Parallel Sysplex technology potentially provides another level of protection beyond the SCP in a domain. Parallel Sysplex technology also potentially extends the concurrent maintenance capabilities of Millennium servers. Although Millennium servers allow concurrent maintenance on the most frequently upgraded components of the system, there are times when maintenance will require taking the server offline. By participating in a Parallel Sysplex environment, taking a Millennium Global Server offline can avoid loss of the processing environment to the end user.

6.8 Support and Service Tools

Amdahl or its authorized service provider supports every Millennium server as an integrated system, taking problem responsibility for all Millennium components and continuing the practice of regarding customer problems as Amdahl problems. Millennium customers have access to the Amdahl WWCSC, its network of problem specialists, standard problem escalation procedures, and a comprehensive spares system. In particular, Millennium servers are supported by the following support tools and services:

Worldwide Customer Support Center. The WWCSC provides 24-hour expert support coverage worldwide. The SVP has communications capability to allow communication with the WWCSC. This capability can be used to transmit diagnostic information, to download firmware updates, or to diagnose severe error conditions remotely from the WWCSC.

Diagnostic Programs. Millennium servers have many diagnostic programs that run from the SVP.

Collecting Information. The SVP collects information about the status of the server for both routine operations, such as environmental logs, and unusual events, such as hardware errors. When an error occurs (other than for a recoverable main storage error), error information is frozen and scanned out to the SVP.

"Phone-Home" Capability. Upon the occurrence of a significant error, the server can automatically transmit a log with pertinent information to the WWCSC, subject to customer authorization.

Expert System. When a log is transmitted to the WWCSC, it is stored in a log database and analyzed in real time by an expert system. The expert system determines whether a maintenance action is required and, if so, it initiates a service order. The system takes into account previous logs for this processor and supplements threshold counters in the server with its own threshold analysis. Thus, subsequent logs may cause the expert system to alter its service order, and, if necessary, escalate the urgency of the repair action.

Real-Time Alert System. The expert system notifies the local field engineer of service orders through the real-time alert system. The field engineer can be notified through voice or electronic mail or can be paged, according to criteria specific to that customer site. Service personnel can thus react to error conditions in real time.

Service Offerings. A range of service offerings are available to provide support appropriate to each customer's individual needs. For more information, refer to "Millennium Global Servers: Technical Support and Services" (MM002276).

Appendix. Millennium Coupling Servers

The Millennium Coupling Server is a System/390 Parallel Sysplex standalone coupling facility based on the same CMOS technology and system design as the Millennium Global Server. The Millennium Coupling Server offers high performance, availability, and flexibility to customers implementing mission-critical workloads in Parallel Sysplex environments. The Millennium Coupling Server can, as the primary coupling facility, interconnect appropriately configured Parallel Sysplex capable processors into a Parallel Sysplex environment. It can also serve as a backup coupling facility to an IBM 9674 Coupling Facility. The Millennium Coupling Server can be operated from an IBM HMC, if the Hardware Management Console Interface Feature (HMCIF) is installed.

Models

There are eight Millennium Coupling Server models. The smallest model has a single CPU to manage lighter coupling facility loads, while the largest model, with eight CPUs, can handle the high volume of message traffic found in the most demanding Parallel Sysplex environments. Each 500 Series model can have from 512 MB to 8 GB of processor storage and up to 32 coupling links. Each 400 Series model can have from 256 MB to 4 GB of processor storage and up to 32 coupling links. No other channel types can be attached. There is no current requirement for External Time Reference Attachments.

Millennium Coupling Server models can be upgraded to higher-capacity Millennium Coupling Server models. They can also be converted to higher-capacity Millennium Global Server models. Refer to the Millennium Coupling Servers 500 Series and Millennium Coupling Servers 400 Series specifications for more detail on upgrade and conversion paths.

Tables A-1 and A-2 list models with configuration options.

Table A-1. Millennium Coupling Server 500 Series Configurations

Model	CPUs	Processor Storage	Coupling Links
CS535	3	Minimum: 512 MB	Minimum: 1
CS545	4	Maximum: 8 GB	Maximum: 32
CS555	5		
CS565	6		
CS575	7		
CS585	8		

Table A-2. Millennium Coupling Server 400 Series Configurations

Model	CPUs	Processor Storage	Coupling Links
CS415	1	Minimum: 256 MB	Minimum: 1
CS425	2	Maximum: 4 GB	Maximum: 32

Multiple Coupling Facility Images

Coupling facility functionality on the Millennium Coupling Server is provided by Amdahl Coupling Control Code (ACCC), internal licensed code developed by Amdahl. ACCC runs in an MDF domain on the Coupling Server. Using MDF, customers can define up to 10 coupling facility images on their Millennium Coupling Server to support multiple Parallel Sysplex environments.

Dual Server

As a Dual Server, any two models of Millennium Coupling Servers can be installed in the same footprint, allowing a second dedicated coupling facility to be installed without having to make space for another system. The second server can be used to add extra capacity or as a backup coupling facility, or it can serve as a coupling facility for use in a completely separate Parallel Sysplex environment. Alternatively, the second system can be a Millennium Global Server, allowing the creation of a complete Parallel Sysplex within a single, compact footprint. Each member server within the Dual Server configuration functions as an independent system with its own operator stations, memory, channels, and coupling links. See section 2.11, “Dual Server,” for more information.

Flexibility

Millennium Coupling Servers can be easily adapted to changing data center requirements.

- The large number of coupling links makes the Millennium Coupling Server adaptable to the largest Parallel Sysplex configurations.
- The QuickChannel feature allows the coupling link configuration to be expanded during normal operation. The QuickCapacity feature provides for rapid model upgrades and the QuickMemory feature allows for quick memory upgrades.
- The Coupling Server can be converted from a coupling facility to a node in a Parallel Sysplex by conversion to a Millennium Global Server.

- The Dual Server feature provides easy addition of a Millennium Coupling Server to an existing Millennium Global Server configuration, providing a compact Parallel Sysplex configuration.

The flexibility to expand the server and readily convert it to other uses increases its usefulness and longevity.

Performance

The Millennium Coupling Server uses the same system design and technology as the Millennium Global Server. Millennium Coupling Servers have powerful CPUs for good response time and low multiprocessing overhead for stable performance as CPUs are added to handle growing coupling facility workloads. By providing rapid responses to coupling requests from nodes in the Parallel Sysplex environment, Millennium Coupling Servers minimize sharing overhead and enable the attached System/390 processors to complete more work. Consult chapter 5, “System Design and Technology,” for more information.

Availability

Millennium Coupling Servers benefit from the reliability of Millennium CMOS technology and from the availability and serviceability functions also found on Millennium Global Servers. With ConServ, firmware on Millennium Coupling Servers can be updated concurrently with normal operation. All maintenance to the cooling fans, many parts of the service processor, operator stations, and operator station LAN can also be performed concurrently with normal operation.

The Millennium QuickAvailability features are available on Millennium Coupling Servers. The QuickChannels feature allows additional coupling links to be installed concurrently with normal coupling facility operation, provided that they have been predefined in the I/O Configuration Data Set (IOCDS) for the Coupling Server. See section 2.10, “QuickChannels Feature,” for more information.

The QuickCapacity feature provides for the rapid addition of one or more unused CPUs to meet new demand and the QuickMemory feature allows for quick increase in memory capacity. The QuickSwitch feature allows a non-functioning CPU to be dynamically replaced by an unused CPU on the same board, avoiding a coupling facility outage. Both the QuickCapacity and QuickSwitch features require that unused CPUs (which are fenced off into a reserve) be present on the system board. Consult section 2.9 “QuickCapacity Feature,” section 2.11 “QuickMemory Feature,” and section 6.5 “QuickSwitch Feature,” for more information.

Millennium Coupling Servers can be configured with a local UPS to provide battery backup and a dual power feed. See section 2.14, “Power Source Options.” In addition to the basic battery backup capability for Millennium Global Servers, Millennium Coupling Servers have the ability to switch to Power Save state while running in non-volatility mode in order to reduce the demand for power when normal power is interrupted. Power Save state means that the system has ceased most operations and will draw just enough power to run minimal operations, including maintaining the contents of memory. All other operations are put on hold until normal power is restored. The contents of memory, including persistent cache structures, are preserved across power failures until such time as they can be recovered by the Parallel Sysplex failure management procedures. With the ability to define a Power Save state within the Millennium Coupling Server, this battery backup feature can preserve the contents of memory during an extended outage for up to serveral hours, depending on the server configuration as well as the number of optional battery units installed.

Operational Environments

The same operational environments possible for a Millennium Global Server apply to a Millennium Coupling Server. The minimum operational environment is one customer operator station (COS), used to control the server, and one service operator station (SOS), used by service personnel to maintain the server. Up to four additional COSs may be installed. One COS functions as the primary operator station and can be operated remotely using Distributed Console Access Facility (DCAF) or automatically from automated operation software programs using Simple Network Management Protocol (SNMP). If the Hardware Management Console Interface Feature (HMCIF) installed, the Millennium Coupling Server can be operated from an IBM Hardware Management Console (HMC). For more information, see section 2.7, “Operational Features.”

Practical Solutions for Parallel Sysplex

The Millennium Coupling Server is an element of the Amdahl’s Practical Solutions for Parallel Sysplex, a comprehensive suite of Amdahl products and services that implement and help customers manage their transition to a Parallel Sysplex environment. See section 2.4.4, “Practical Solutions for Parallel Sysplex,” for a summary of product offerings.

Bibliography

▼ ▼ ▼

Amdahl Publications

MM002205 *Millennium Global Servers (Comprehensive Brochure)*

MM002204 *Millennium Global Servers: Overview (Summary Brochure)*

MM002224 *Millennium Global Servers: 500 Series (Specification Sheet)*

MM002306 *Millennium Global Servers: 400 Series (Specification Sheet)*

MM002276 *Millennium Global Servers: Technical Support and Services*

MM002302 *The Ultimate Cache Machine, Millennium Global Servers System Design and Performance (White Paper)*

MM-143977 *Millennium Global Servers: Physical Planning Manual*

MC-143975 *Millennium Global Servers: Control System Reference*

MC-143973 *Millennium Global Servers: Control System Operations*

MC-143974 *Millennium Global Servers: Control System Messages and Codes*

MC-143972 *Millennium Global Servers: Control System Concepts*

MC-144004 *Millennium Global Servers: Request Machine Interface (RMI) Reference*

MC-144003 *IOCP User Guide*

MM002303 *Millennium Coupling Servers: 400 Series (Specification Sheet)*

MM002294 *Millennium Coupling Servers: 500 Series (Specification Sheet)*

MM002149 *Amdahl/Fujitsu: Creative Partners in Technology*

MM002151 *Amdahl System/390 Processors Fact Sheet*

MM001683 *Amdahl Processors Pocket Reference Guide*

MM002242 *Practical Solutions for Parallel Sysplex (Summary Brochure)*

MM002238-002 *Parallel Sysplex Bulletin: Issue 1*

MM002238-003 *Parallel Sysplex Bulletin: Issue 2*

MM002365-001 *Parallel Sysplex Bulletin: Issue 3*

IBM Publications

SA22-7201 *IBM Enterprise Systems Architecture/390: Principles of Operation*

SA22-7202 *IBM Enterprise Systems Architecture/390: Enterprise Systems Connection Architecture I/O Interface*

GA22-6974 *IBM System/360 and System/370 I/O Interface Channel to Control Unit Original Equipment Manufacturers' Information*

SC28-1187 *IBM Large Systems Performance Reference*

IBM Washington Systems Center FLASH 9505, "MVS Performance, Capacity Planning Considerations for 9672-RXX Processors," February 1995.

Glossary

▼ ▼ ▼

ACCC	Amdahl Coupling Control Code.
ALB	access register translation lookaside buffer.
APAF	Amdahl Performance Analysis Facility.
APAFMON	Amdahl Performance Analysis Facility Monitor.
APAF/MX	Amdahl Performance Analysis Facility for the Computer Associates MICS database.
APS	Auxiliary Power Source.
AR	access register.
ARF	Automatic Reconfiguration Facility.
ARQ	addressable register queue.
ASOA	automated scan-out analysis program.
availability	degree to which a system is available for productive operation.
cache	High-speed buffer storage containing frequently referenced instructions and/or data that is used to reduce access time to main storage data.
capping	In logical partitioning, setting a limit to a domain's share of the processor that is relative to other domains' shares. Compare to Max Capping, which sets an absolute limit to a domain's share.
CFR	Coupling Facility Receive.
CFS	Coupling Facility Send.
channel backpanel	panel located in the channel frame or channel expansion frame that attaches I/O processor adapters and channel cards.
channel card	printed circuit board that contains channels.
CHM	channel manager.
CHP	channel processor.
CMOS	complementary metal oxide semiconductor.
concurrent	system maintenance that can be maintenance performed concurrently with normal operation of the system.
COS	customer operator station.
Coupling Domain	MDF domain running Amdahl Coupling Control Code.
CPU	central processing unit.
CR	control register.
CTC	channel-to-channel adapter.
DCAF	Distributed Console Access Facility.
domain	in the Multiple Domain Feature, a logical processing system or logical partition; a processing environment with a specified architecture and assigned processing resources.
DRM	Dynamic Reconfiguration Management.
DSR	Dynamic Storage Reconfiguration.
ECC	error-checking and correction.
ECL	emitter-coupled logic.
EMIF	ESCON Multiple Image Facility.
EPO	Emergency Power Off.
ESA/390	Enterprise Systems Architecture/390.
ESCON	Enterprise Systems Connection Architecture.
ETRA	external time reference attachment.
E-unit	execution unit.
fault avoidance	Reliability; the state of having a low likelihood of failure.
fault tolerance	ability of a processing system to continue operation with a malfunctioning component.
firmware	microcode and machine language code.
FPR	floating-point register.
GB	gigabyte.
GUI	graphical user interface.
hardware system area	logical area of main storage not addressable by user system control programs that is used to store firmware and control information.
HCD	Hardware Configuration Definition.
HMC	Hardware Management Console.
HMCIF	Hardware Management Console Interface Feature.
HSA	hardware system area.
IEF	Interpretive Execution Facility.
IMPL	initial microprogram load.
I/O	input/output.
IOCDS	I/O configuration data set.
I/O configuration data set	data set that holds the output of the I/O configuration program.
I/O configuration program	program that defines all of the available channel paths and I/O devices for the system.
IOCP	I/O configuration program.

IOCS	I/O control source.
IOPA	I/O Processor Adapter.
IPL	initial program load.
ITR	internal throughput rate.
I-unit	instruction unit.
kBTU	kiloBritish Thermal Units.
kVA	kiloVolt Ampere.
kW	kiloWatts.
L1 cache	Level 1 cache.
L2 cache	Level 2 cache.
LAN	local area network.
logical processor	The representation of a physical CPU in a domain.
LPAR	logical partition.
LPID	Logical Processor Identification.
Mb	megabit(s).
MB	megabyte(s).
MCM	multichip module.
MDF	Multiple Domain Feature.
MIPS	millions of instruction per second.
model-dependent	relating to characteristics of a processor that vary from those of other processors that conform to the same architecture.
MRSD	Machine Readable Scheduler Data.
multichip module	air-cooled, six-layer, polyimide printed circuit board, with cooling fins, to which logic chips are attached.
multiprocessing effects	reduction in the effective performance of a multiprocessor principally due to contention for shared memory.
multiprocessing factor	effective performance or throughput of a multiprocessor expressed as a multiple of the performance or throughput of a uniprocessor CPU of the same type.
multiprocessor	a processing system with multiple processors that can access common main storage.
n + 1 redundancy	configuration of units performing comparable functions in which n units are required for normal operation, but having the surplus capacity of one additional unit to allow continuous operation in case one unit fails.
OLTP	online transaction processing.
OSA	open systems adapter.
PC	personal computer.
processor storage	on Millennium Global Servers, storage that can be defined as either main or expanded storage under operator control.
PR/SM	Processor Resource/Systems Manager.
PSLC	Parallel Sysplex License Charge.
PSMF	Parallel Sysplex Migration Facility.
PSU	processor storage unit.
PSW	program status word.
RAM	random access memory.
RAS	reliability, availability, serviceability.
reliability	absence of error, or the degree to which a system or component is free of error, regardless of whether the error is correctable.
RMI	Request Machine Interface.
SAM	System Activity Monitor.
SC	system controller.
scalability	degree to which adding a unit of processing capacity produces a proportionate increase in aggregate system throughput.
SCI	system console interface.
SCP	system control program.
SCU	system control unit.
SDRAM	synchronous dynamic random access memory.
SDS	system data switch.
self-timed RAM	Random access memory that has on-chip clocking.
serviceability	effective capability to do system maintenance, including concurrent maintenance capabilities, and to diagnose and repair any system failure.
service clearance	amount of space required to open doors and panels to a system or device in order to perform maintenance.
SMF	System Management Facility.
SNMP	Simple Network Management Protocol.
SOS	service operator station.
SPC	system power controller.
STD	segment table designation.
STIDP	Store CPU ID.
SVP	service processor.

synchronous DRAM	Dynamic random access memory with internal pipelining to allow multiple simultaneous memory accesses to be active in the same RAM.
system board	An air-cooled, 31-layer, polyimide printed circuit board that contains multichip modules, Level 2 cache, processor storage cards, and alternate RAMs.
S-unit	storage unit.
TAG	control section associated with a cache that contains main storage addresses of the data in that cache with related control information.
TCP/IP	Transmission Control Protocol/ Internet Protocol.
TLB	translation lookaside buffer.
TPF	Transaction Processing Facility.
TPF/HPO	Transaction Processing Facility High Performance Option.
TSCF	Target System Control Facility.
UPS	uninterruptible power supply.
WAN	wide area network.
way	in processor storage, a section of memory whose access can be overlapped with accesses in other ways. In cache, one of several sections in which a main storage block or subblock can be stored at the same relative offset as in other sections.
WWCSC	Worldwide Customer Support Center.

Worldwide Headquarters
Amdahl Corporation
1250 East Arques Avenue
Sunnyvale, CA 94088-3470
United States of America
Tel: 408 746 6000
Fax: 408 746 3243
Internet: www.amdahl.com

North America
United States of America
Tel: 800 223 2215
Fax: 800 452 3509
Canada
Tel: 416 510 3111
Fax: 416 510 3353

Europe
(covering Austria, Belgium, Denmark, France, Germany, Ireland, Italy, Luxembourg, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and the United Kingdom)

For your local Amdahl office, please access:
www.amdahl.com/europe/offices
or email:
Amdahl_In_Europe@amdahl.com

South Africa
Johannesburg
Tel: +27 11 467 0710
Fax: +27 11 467 0740

Asia
Amdahl Asia Services Sdn.Bhd
Tel: +60 3 794 7233
Fax: +60 3 794 9911

Australia and New Zealand
Amdahl Pacific Services Pty Ltd
ACN 003 839 503
Tel: +61 2 9561 9999
Fax: +61 2 9561 9811

For Y2K compliance status of this and any other Amdahl offerings, see URL: www.amdahl.com/y2k/

Not all Millennium server models are available in all countries. Availability dates may vary for some features and models. Please contact your local Amdahl representative regarding availability dates for specific models and features. The information in this document may be superseded by subsequent documents. For details regarding delivery of specific products, features, and services, contact your local Amdahl representative.

MM002388-002 [1:0] 2/99