IBM @server p5: A Highly Available Design for Business-Critical Applications

By: Jim Mitchell, Daniel Henderson, and George Ahrens

December 15, 2004

Introduction

The recently announced IBM POWER5™ processor-based servers support new levels of performance and virtualization. At a time when some UNIX® vendors are vigorously advocating "good enough" servers for "good enough" computing; that is, servers that tradeoff high availability features for lower cost, IBM recognizes that even small jobs, running on entry servers can be mission-critical. In addition, IBM Virtualization Engine™ system technologies, announced in the IBM @server® p5 family, can enable individual servers to run dozens or even hundreds of mission critical applications.

In IBM's view servers must be designed to avoid both planned and unplanned outages, and to maintain a focus on application uptime.

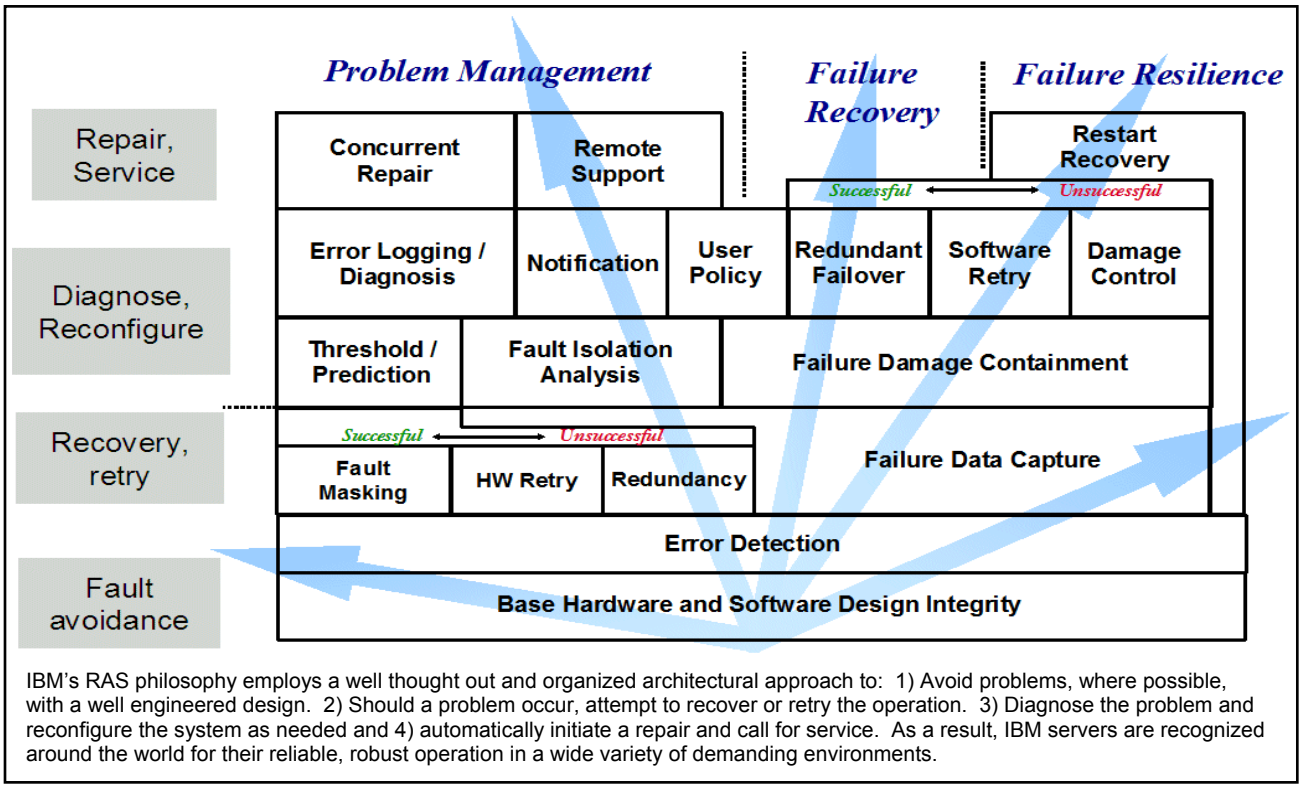
From a reliability, availability, and serviceability (RAS) standpoint, @server p5 servers include features designed to increase availability and to support the new levels of virtualization, building upon the leading-edge RAS features delivered in the IBM pSeries® family of servers.

IBM RAS engineers are constantly making incremental improvements in server design to help insure that IBM UNIX servers support high levels of concurrent error detection, fault isolation, recovery, and availability. Each successive generation of IBM UNIX and Linux® operating system (OS) -based servers is designed to be more reliable than the server family it replaces. IBM has spent years developing RAS capabilities for mainframes and mission-critical servers. The @server p5 family of servers builds on the reliability record of the POWER4™ processor-based offerings.

A RAS Design Philosophy

The overriding IBM @server p5 server design goal is simply stated:

Employ an architecture-based design strategy to devise and build IBM servers that can avoid unplanned application outages. In the unlikely event that a hardware fault should occur, the system must analyze, isolate, and identify the failing component so that repairs can be effected (either dynamically, through "self-healing," or via standard service practices) as quickly as possible – with little or no system interruption. This should be accomplished regardless of the size of the system or how the system is partitioned.



IBM's RAS philosophy employs a well thought out and organized architectural approach to: 1) Avoid problems, where possible, with a well engineered design. 2) Should a problem occur, attempt to recover or retry the operation. 3) Diagnose the problem and reconfigure the system as needed and 4) automatically initiate a repair and call for service. As a result, IBM servers are recognized around the world for their reliable, robust operation in a wide variety of demanding environments.

The core principles guiding IBM engineering design are reflected in the RAS architecture. The goal of any server design is to:

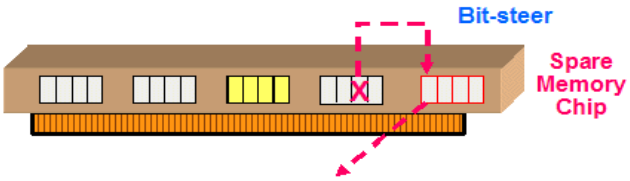
1. Achieve a highly reliable design through extensive use of highly reliable components built into a system package that supports an environment conducive to their proper operation.
2. Clearly identify, early in the server design process, those components which have the highest opportunity for failure. Employ a server architecture that allows the system to recover from intermittent errors in these components and/or fail-over to redundant components when necessary. A variety of redundancy strategies can be employed:
   - The server design can entirely duplicate a function, using, for example, dual I/O connections between the Central Electronics Complex (CEC) and an I/O drawer.
   - The redundancy can be of an N+1 variety. For example, the server can include multiple, variable speed fans. In this instance, should a single (or in some cases, even multiple failures can be tolerated) fan fail, the remaining fan(s) will automatically be directed to increase their rotational speed to maintain adequate cooling until a hot-plug repair can be effected.
   - Fine grained redundancy schemes can be used at subsystem levels. For example, extra or "spare" bits in a memory system (cache, main store) can be used to effect ECC (Error Checking and Correction) schemes.

   IBM engineers draw upon an extensive record of reliability data collected over decades of design and operation of high-end servers. Detailed component failure rate data is used to determine both what redundancy is needed to achieve high levels of system availability and what level of redundancy provides the most effective balance of reliable operation, server performance, and overall system cost.

   When the availability afforded by full redundancy is required, IBM and third party software vendors provide a number of high availability clustering solutions such as IBM High Availability Cluster Multiprocessing software (HACMP™ for AIX 5L™).
3. Develop server hardware that can detect and report on failures and impending failures.
   - IBM @server p5 servers employ a design methodology called First Failure Data Capture (FFDC). This methodology uses hardware-based fault detectors to extensively instrument internal system components. Each detector is a diagnostic probe capable of reporting fault details to a dedicated Service Processor. FFDC, when coupled with automated firmware analysis, is used to quickly and accurately determine the root cause of a fault the first time it occurs, regardless of phase of system operation and without the need to run "recreate" diagnostics. The overriding imperative is to identify *which component* caused a fault — *on the first occurrence of the fault* — and to prevent any reoccurrence of the error.
   - One key advantage of the FFDC technique is the ability to predict potentially catastrophic hardware errors before they occur. Using FFDC, a Service Processor in a POWER5 processor-based server has extensive knowledge of recoverable errors that occur in a system. Algorithms have been devised to identify patterns of recoverable errors that could lead to an unrecoverable error. In this case, the Service Processor is designed to take proactive actions to guard against the more catastrophic fault (system check-stop or hardware reboot).
4. Create server hardware that is self-healing, that automatically initiates actions to effect error correction, repair, or component replacement.
   - Striving to meet demanding availability goals, POWER5 processor-based systems deploy redundant components where they will be most effective. Redundancy can be employed at a functional level (as described above) or at a subsystem level. For example, extra data bit lines in memory can be dynamically activated before a non-recoverable error occurs or spare bit lines in a cache may be invoked after the fault has occurred.

     The goal of self-healing/sparing is to avoid faults by employing sparing where it can most effectively prevent an unscheduled outage.

- In some instances even scheduled outages may be avoided by "self-healing" a component. Self-healing concepts can be used to fix faults within a system without having to physically remove or replace a part. IBM's unique First Failure Data Capture methodology is used to accurately capture intermittent errors – allowing a Service Processor to diagnose potentially faulty components. Using this analysis, a server can "self-heal," effecting a repair before a system failure actually occurs.

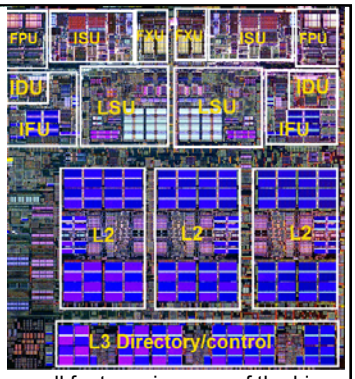


Should a main store memory location experience too many intermittent correctable errors, an IBM @server p5 server will automatically move the data stored at that location to a "back-up" memory chip. All future references to the original location will automatically be accessed from the new chip. Known as "bit-steering", this is an example of "self-healing." The system continues to operate will full performance, reliability, and no service call!

- The FFDC methodology is also used to predictively vary-off (deallocate) components for future scheduled repair. In this case, the system will continue to operate, perhaps in a degraded mode, avoiding potentially expensive unscheduled server outages. One example, available in IBM POWER™ servers since 2000, is processor run-time deconfiguration, the ability to dynamically (and automatically) take a processor off-line for scheduled repair before a potentially catastrophic system crash occurs.
- In those rare cases where a fault causes a partition or system outage, FFDC information can be used upon restart to deconfigure (remove from operation) a failing component, allowing the system or partition to continue operation, perhaps in a degraded mode, while waiting for a scheduled repair.

Reliability: Start with a Solid Base

The base reliability of a computing system is, at its' most fundamental level, dependent upon the intrinsic failure rates of the components that comprise it. Very simply, highly reliable servers are built with highly reliable components. This basic premise is augmented with a clear "design for reliability" architecture and methodology. Trained IBM RAS engineers use a concentrated, systematic, architecture-based approach designed to improve the overall server reliability with each successive generation of system offerings. At the core of this effort is an intensive focus on sensible, well-managed server design strategies that not only stress high system instruction execution performance but also require logic circuit implementations that will operate consistently and reliability despite potentially wide disparity in manufacturing process variance and operating environments. Intensive critical circuit path modeling and simulation procedures are used to identify critical system timing dependencies so that time-dependent system operations complete successfully under a wide variety of process tolerances.



The POWER5 chip features single- and simultaneous multi-threading execution. POWER5 maintains both binary and architectural compatibility with existing POWER4 processor-based systems to ensure that binaries continue executing properly and that application optimizations carry forward to newer systems. POWER5 technology supports additional enhancements such as virtualization, and improved reliability, availability, and serviceability at both chip and system levels. The chip includes approximately 276 M transistors.

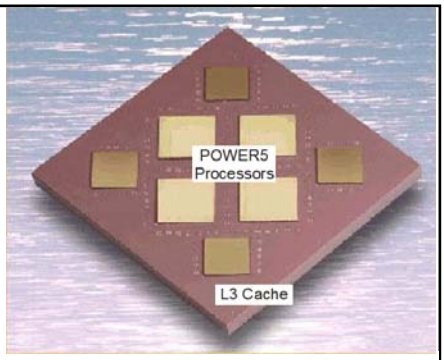
Given the large number of circuits and the small feature size, one of the biggest challenges in modern processor design is controlling chip power (heat). Unmanaged, the heat can significantly affect the overall reliable of a server. The introduction of simultaneous multi-threading in POWER5 allows the chip to execute more instructions per cycle per processor core, increasing total switching power. In mitigation, POWER5 chips use a fine-grained, dynamic clock-gating mechanism. This mechanism turns off clocks to a local clock buffer if dynamic management logic determines that a set of latches driven by the buffer will not be used in the next cycle. This allows substantial power saving with no performance impact.

During the system definition phase of the server design process, well before any detailed logic design is initiated, the IBM RAS team carefully evaluates system reliability attributes and calculates a server "reliability target." This target is primarily established by a careful analysis of the potentially attainable reliability (based on available components), and by comparison with current IBM server reliability statistics. In general, RAS targets are set with the goal of exceeding the reliability of currently available servers. For the past decade,

IBM RAS engineers have been systematically adding mainframe-inspired RAS technologies to IBM UNIX OS offerings, resulting in dramatically improved system designs.

In the "big picture" view, servers with fewer components, with fewer interconnects, have fewer chances to fail. Seemingly simple design choices, e.g. integrating two processor cores on a single POWER5 chip, can dramatically reduce the "opportunity" for server failure. In this case, a 64-way server will include half as many processor chips as with a single CPU per processor design. Not only will this reduce the total number of system components, it will reduce the total amount of heat generated in the design, resulting in an additional reduction in required power and cooling components. Finally, the biggest IBM @server p5 servers, the p5-590 and p5-595 use a very high degree of integration by employing IBM MCM (multi-chip module) designs using the same technology as is currently deployed in zSeries® (mainframe) servers. The POWER5 MCM uses a glass ceramic module that holds four POWER5 microprocessors (8-way module) and four L3 cache modules. All connections between the on-board components are included in wiring embedded in the substrate, all system connections are routed through the module to the system board. Not only does this result in a high-performance, highly scalable system package that carefully controls the interconnect speeds and manages the heat, but which also reduces the total number of supporting chips per 8-way module by eight chips over its POWER4 predecessor.

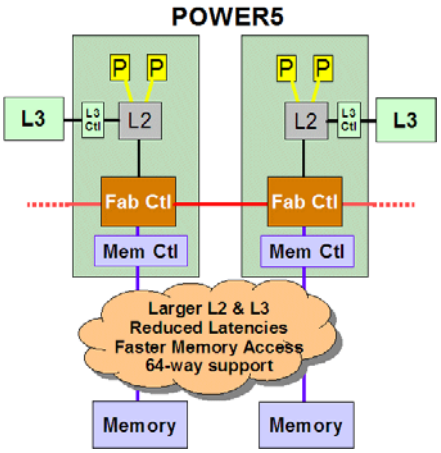
POWER5
- MCM Package
  - 4 POWER5 chips
  - 4 L3 cache chips
- 3.75" x 3.75"
  - 95mm x 95mm
- 4,491 signal I/Os
- 89 layers of metal



The POWER5 multichip module design uses proven mainframe packaging technology to pack four POWER5 chips (8 CPU's) and four L3 cache chips (36 MB each) on a single ceramic substrate. This results in a highly reliable, high performance system package for high capacity servers.

As has been illustrated, system packaging can have a significant impact on server reliability. Since the reliability of electronic components is directly related to their thermal environment (large decreases in component reliability can be measured due to relatively small increases in temperature), IBM servers are carefully packaged to insure adequate cooling. Critical system components (POWER5 processor chips for example) are positioned on printed circuit cards so that they receive "upstream" or "fresh" air, while less sensitive or lower power components like memory DIMMs are positioned "downstream." In addition, POWER5 processor-based servers are built with redundant, variable speed fans that are designed to automatically increase their output to compensate for increased heat in the central electronic complex.

The POWER5 chip maintains full binary and architectural compatibility with IBM's POWER4 processor, ensuring that application binaries will properly execute and that code optimizations will carry forward to the new generation of servers. POWER5 also offers a number of improvements including simultaneous multi-threading, the ability to execute two threads on the same CPU core at the same time (for improved performance), enhanced virtualization features, and improved data movement (reduced cache latencies and faster memory access).



The L3 cache, attached directly to the POWER5 processor using separate read and write busses (running at ½ processor speed), acts as a high speed buffer for data that doesn't "fit" in the L2 cache. This design significantly reduces L3 cache latency when compared to POWER4 designs.

Moving the memory controller function to the POWER5 chip reduces the number of chips needed to build a server, it also reduces main memory latency.

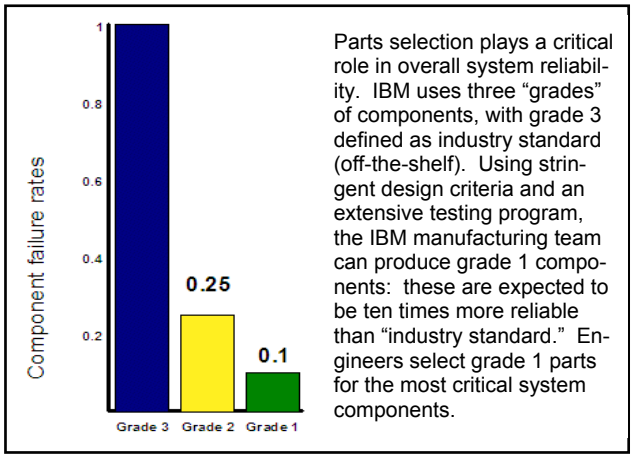
Restructuring the server reduces the load on the inter-processor "fabric" bus, allowing POWER5 processor-based servers to scale to 64-way systems. Fabric busses are protected with ECC, enabling the system to automatically correct data transmission errors and survive conditions that could have caused outages on earlier system designs. This is one example of the "continuous quality improvement" employed by IBM RAS engineers.

In each @server p5 product offering, from the p5-520 (1- to 2-way server), to the p5-595 (16- to 64-way server) the server packaging is designed to deliver both high-performance and high-reliability. In each case, IBM engineers perform an extensive "bottoms-up" reliability analysis using part level failure rate calculations for every part in the server. These calculations assist the system designers to select a package that best supports the design for reliability. For

example, while the p5-550 and the p5-570 servers are similarly packaged 19" rack offerings, they employ different CPU cards. The more robust p5-570 includes not only additional system fabric connections for performance expansion, but also the robust cooling components (heat sinks, fans) to compensate for the increased heat load of faster processors, larger memory, and bigger caches.

The detailed RAS analysis helps the design team to pinpoint those server features and design improvements that will have a significant impact on overall server availability. Thus the analysis enables IBM engineers to differentiate between "high opportunity" items (those that most affect server availability) that need to be protected with redundancy and fixed via concurrent repair, and "low opportunity" components (those which seldom fail or have low impact on system operation) that can be deconfigured and scheduled for deferred, planned repair.

Components that have the highest failure rate and/or highest availability impact are quickly identified and the system is designed to manage their impact to overall server RAS. For example selected @server p5 servers will include redundant, "hot-plug" fans and provisions for N+1 power supplies. Many CEC components are built using IBM "grade 1" components, parts that are designed and tested to be up to ten times more reliable than their "industry standard" counterparts. The @server p5 systems include measures that compensate for, or correct, errors received from components comprised of less extensively tested parts. For example, industry grade PCI adapters are protected by industry-first IBM PCI bus enhanced error recovery (for dynamic recovery of PCI bus errors) and, in most cases, support "hot-plug" replacement if necessary.



Parts selection plays a critical role in overall system reliability. IBM uses three "grades" of components, with grade 3 defined as industry standard (off-the-shelf). Using stringent design criteria and an extensive testing program, the IBM manufacturing team can produce grade 1 components: these are expected to be ten times more reliable than "industry standard." Engineers select grade 1 parts for the most critical system components.

Continuous Field Monitoring

Of course, setting failure rate reliability targets for component performance will help create a reliable server design. But simply setting targets is not sufficient.

IBM field engineering teams track and record repairs of system components covered under warranty or maintenance agreement. Failure rate information is gathered and analyzed for each part by IBM commodity managers, who track replacement rates. Should a component not be achieving its' reliability targets, the commodity manager will create an action plan and take appropriate corrective measures to remedy the situation.

Aided by the IBM FFDC methodology and the associated error reporting strategy, commodity managers build an accurate profile of the types of field failures that occur and initiate programs to enable corrective actions. In many cases, these corrections can be initiated without waiting for parts to be returned for failure analysis.

The IBM field support team also continually analyzes critical system faults, testing to determine if system firmware and maintenance procedures and tools are effectively handling and recording faults. This continuous field monitoring and improvement structure allows IBM engineers to ascertain, with some degree of certainty, how systems are performing in client environments, rather than just depending on projections. Not only will engineers undertake "in-flight" corrections to improve current products being deployed, but this structure also provides valuable information useful for planning and designing future server products.

Fault Detection and Isolation

While the ability to accurately detect hardware faults is certainly a critical requirement for any server, the real challenge is to accurately diagnose and identify the true source of the error. Determining that there is corrupt data on a PCI bus will insure "application data integrity" since the server will not allow this data to be used, but may not be helpful in repairing the server if the source of the corruption can not be clearly

identified. So fault detection *and isolation* are essential elements of high availability and serviceable server design.

In the @server p5 systems, three key elements; the POWER Hypervisor™, Service Processor (SP), and Hardware Management Console (HMC), cooperate in detecting, reporting, and managing hardware faults.

Service Processor

The Service Processor is a separately powered microprocessor, separate from the main @server p5 instruction processing complex. The Service Processor enables POWER Hypervisor and Hardware Management Console surveillance, selected remote power control, environmental monitoring (only critical errors are supported under Linux®), reset, and boot features, remote maintenance and diagnostic activities, including console mirroring. On systems without a hardware management console, the Service Processor can place calls to report surveillance failures with the POWER Hypervisor, critical environmental faults, and critical processing faults even when the main processor is inoperable. The Service Processor provides services common to modern computers:

1. Environmental monitoring
   - The Service Processor monitors the server's built-in temperature sensors, sending instructions to the system fans to increase rotational speed when the ambient temperature is above the normal operating range.
   - Using an architected operating system interface, the Service Processor notifies the operating system of potential environmental related-problems (for example, air conditioning and air circulation around the system) so that the system administrator can take appropriate corrective actions before a critical failure threshold is reached.
   - The Service Processor can also post a warning and initiate an orderly system shutdown for a variety of other conditions:
     – When the operating temperature exceeds the critical level.
     – When the system fan speed is out of operational specification.
     – When the server input voltages are out of operational specification.
2. Mutual Surveillance
   - The Service Processor monitors the operation of the POWER Hypervisor firmware during the boot process and watches for loss of control during system operation. It also allows the POWER Hypervisor to monitor Service Processor activity. The Service Processor can take appropriate action, including calling for service, when it detects the POWER Hypervisor firmware has lost control. Likewise, the POWER Hypervisor can request a Service Processor repair action if necessary.
3. Availability
   - The auto-restart (reboot) option, when enabled, can reboot the system automatically following an unrecoverable firmware error, firmware hang, hardware failure, or environmentally induced (AC power) failure.
4. Fault Monitoring
   - BIST (built-in self-test) checks processor, L3 cache, memory, and associated hardware required for proper booting of the operating system, when the system is powered on at the initial install or after a hardware configuration change (e.g., an upgrade). If a non-critical error is detected or if the error occurs in a resource that can be removed from the system configuration, the booting process is designed to proceed to completion. The errors are logged in the system nonvolatile random access memory (NVRAM). When the operating system completes booting, the information is passed from the NVRAM into the system error log where it is analyzed by error log analysis (ELA) routines. Appropriate actions are taken to report the boot time error for subsequent service if required.
   - Disk drive fault tracking can alert the system administrator of an impending disk failure before it impacts client operation.
   - The AIX 5L or Linux log (where hardware and software failures are recorded) is analyzed by ELA routines which warn the system administrator about the causes of system problems.

POWER Hypervisor

The advanced virtualization techniques available in the @server p5 servers require a powerful management interface for allowing a system to be divided into multiple partitions, each running a separate operating system image instance. This is accomplished using firmware known as the POWER Hypervisor. The POWER Hypervisor provides software isolation and security for all partitions.

The POWER Hypervisor is active in all systems, even those containing just a single partition. The POWER Hypervisor helps to enable IBM Virtualization Engine™ systems technology options including:

- Micro-Partitioning technology, allowing creation of highly granular dynamic LPARs or virtual servers as small as 1/10th of a processor, in increments as small as 1/100th of a processor. A fully configured @server p5 595 or 590 server can run up to 254 partitions.
- A shared processor pool, providing a pool of processing power that is shared between partitions, helping to improve utilization and throughput.
- Virtual I/O, supporting sharing of physical disk storage and network communications adapters, and helping to reduce the number of expensive devices required, improve system utilization, and simplify administration.
- Virtual LAN, enabling high-speed, secure partition-to-partition communications to help improve performance.

Elements of the POWER Hypervisor are used to manage the detection and recovery of certain errors, especially those related to the I/O hub (including the GX+ bus adapter and the "I/O-planar" circuitry that handles I/O transactions). The POWER Hypervisor communicates with both the Service Processor, to aggregate errors, and the Hardware Management Console.

The POWER Hypervisor can also reset and reload the Service Processor. It will automatically invoke a reset/reload of the SP if an error is detected. If the SP doesn't respond and the reset/reload threshold is reached, the POWER Hypervisor will initiate an orderly shutdown of the system. In the first half of 2005, a downloadable no charge firmware update will be provided to enable redundant Service Processor failover in p5-595 and p5-590 servers. Once installed, if the error threshold for the failing SP is reached, the system will initiate a failover from one Service Processor to the backup. Types of SP errors:

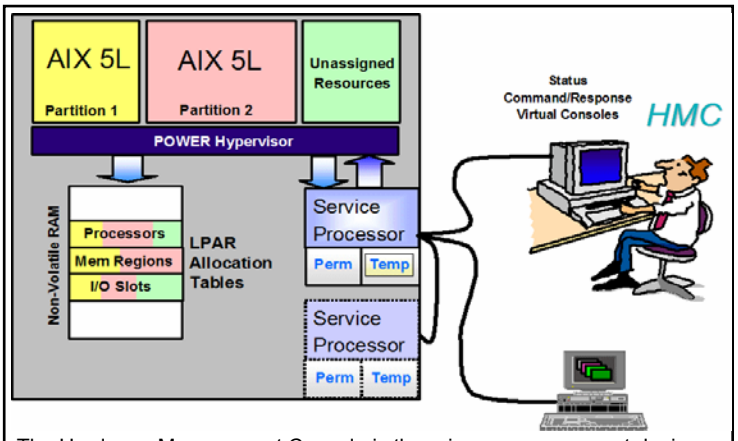
- Configuration I/O failure to the SP
- Memory-mapped I/O failure to the SP
- SP PCI-X/PCI bridge freeze condition

A Service Processor reset/reload is not disruptive and will not impact system operation. SP resets can be initiated by either the POWER Hypervisor or the SP itself. In each case, the system, if necessary, will initiate a smart dump of the SP control store to assist with problem determination if required.

Hardware Management Console

The Hardware Management Console is used primarily by the system administrator to manage and configure the virtualization technologies available for all @server p5 servers. The RAS team uses the HMC as an integrated service focal point, to consolidate and report error messages from the system. The Hardware Management Console is also an important component for concurrent maintenance activities. Key HMC functions:

- Logical partition configuration and management
- Dynamic Logical Partitioning
- Capacity and resource management
- Management of the HMC (for example, microcode updates, access control)
- System status



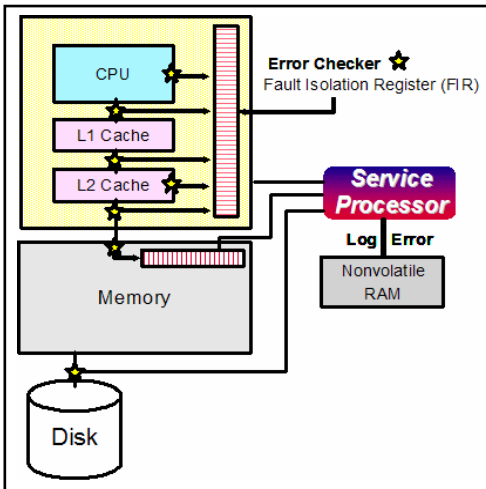
The Hardware Management Console is the primary management device for the POWER5 virtualization technologies (LPAR, Virtual I/O). It is also used to provide a service focal point to collect and manage all RAS information from the server. Two HMCs may be attached to a server to provide backup if desired.

- Service functions (for example, microcode updates, “call home” capability, automated service, and Service Focal Point)
- Remote HMC interface
- Capacity on Demand options

First Failure Data Capture, a “Full Spectrum” Diagnostics Strategy

Diagnosing problems in a computer is a critical requirement for autonomic computing. The first step to producing a computer that truly has the ability to “self-heal” is to create a highly accurate way to identify and isolate hardware errors. IBM has implemented a server design which “builds-in” thousands of hardware error check stations that capture and help to identify error conditions within the server. The @server p5 595 server, for example, includes almost 80,000 checkers to help capture and identify error conditions. These are stored in over 29,000 Fault Isolation Register (FIR) bits. Each of these checkers is viewed as a “diagnostic probe” into the server, and, when coupled with extensive diagnostic firmware routines, allows quick and accurate assessment of hardware error conditions at run-time.

Named “First Failure Data Capture” (FFDC), this proactive diagnostic strategy is a significant improvement over less accurate “reboot and diagnose” service approaches. Using projections based on IBM internal tracking information it’s possible to predict that high impact outages would occur 2 to 3 times more frequently without a FFDC capability. In fact, without some type of pervasive method for problem diagnosis, even simple problems which behave intermittently can be a cause for serious and prolonged outages.



First Failure Data Capture, first deployed by IBM POWER servers in 1997, plays a critical role in delivering servers that can self-diagnose and self-heal. Using thousands of checkers (diagnostic probes) deployed at critical junctures throughout the server, the system effectively “traps” hardware errors at system run time.

The separately powered Service Processor is then used to analyze the checkers and perform problem determination. Using this approach IBM no longer has to rely on an intermittent “reboot and retry” error detection strategy, but knows with some certainty which part is having problems.

In this automated approach, run-time error diagnostics can be deterministic, so that for every check station, the unique error domain for that checker is defined and documented. Ultimately the error domain becomes the FRU (part) call, and manual interpretation of the data is not normally required.

This architecture is also the basis for IBM’s predictive failure analysis, since the Service Processor can now count, and log, intermittent component errors and can deallocate or take other corrective actions when an error threshold is reached.

Integrated hardware error detection and fault isolation has been a key component of pSeries server design strategy since 1997. FFDC “check stations” are carefully positioned within the server logic and data paths to insure that potential errors can be quickly identified and accurately tracked to an individual Field Replaceable Unit (FRU). These checkers are collected in a series of Fault Isolation Registers, where they can easily be accessed by the Service Processor. All communication between the SP and the FIR is accomplished “out of band.” That is, operation of the error detection mechanism is transparent to an operating system. This entire structure is “below the architecture” and is not seen, nor accessed, by system level activities.

In this environment, strategically placed error checkers are continuously operating to precisely identify error signatures within defined hardware fault domains. IBM servers are designed so that in the unlikely event that a fatal hardware error occurs, FFDC, coupled with extensive error analysis and reporting firmware in the Service Processor, should allow IBM to isolate a hardware failure to a single Field Replaceable Unit (FRU). In this event, the FRU part number will be included in the extensive error log information captured by the server. In select cases, a set of FRUs will be identified when the fault is on an interface between two or more FRUs. For example, three FRUs may be “called out” when the system cannot differentiate between a failed driver on one component, the corresponding receiver on a second, or the interconnect fabric. In either case, it is IBM’s standard maintenance practice for the @server p5 systems to replace the identified components as a group. Meeting rigorous goals for fault isolation requires a RAS methodology that carefully instruments the entire system logic design with meticulously placed error checkers.

All hardware error checkers have distinct attributes:

1. Their primary purpose is to provide data integrity.
2. Checkers are used to initiate a wide variety of recovery mechanisms designed to correct the problem. POWER5 processor-based servers include extensive hardware (ranging from bus retry based on parity error detection, to ECC correction on caches and system busses) and firmware recovery logic.
3. Checkers deterministically isolate physical faults based on run-time detection of each unique failure.

Error check signals are captured and stored in hardware Fault Isolation Registers. Associated circuitry, called "who's on first" logic, is used to limit the domain of an error checker to the first checker that encounters the error. In this way, run-time error diagnostics can be deterministic, so that for every check station, the unique error domain for that checker is defined and documented. Ultimately the error domain becomes the FRU call, and manual interpretation of the data is not normally required.

This type of automated error capture and identification is especially useful in allowing quick recovery from unscheduled hardware outages. This data not only provides a basis for failure analysis of the component, it can also be used to improve the reliability of the part and be incorporated into design improvements in future systems.

IBM RAS engineers can use specially designed logic circuitry to create faults that can be detected and stored in FIR bits, simulating internal chip failures. This technique, called error injection, is used to validate server RAS features and diagnostic functions in a variety of operating conditions (power-on, boot, and operational run-time phases). Error injection is used to confirm both execution of appropriate analysis routines and correct operation of fault isolation procedures that report to upstream applications (the POWER Hypervisor, operating system, and Service Focal Point and Service Agent applications). Further, this test method verifies that recovery algorithms are activated and system recovery actions take place. Error reporting paths for client notification, pager calls, and call home to IBM for service are validated and RAS engineers substantiate that correct error and extended error information is recorded. A test servicer, using the maintenance package, then "walks through" repair scenarios associated with system errors, helping to insure that all the pieces of the maintenance package work together, and that the system can be restored to full functional capacity. In this manner, RAS features and functions, including the maintenance package, are verified for operation to design specifications.

IBM uses the client impact of a part failure as the measure of success of the availability design. This metric is defined in terms of application, partition, or system downtime. IBM traditionally classifies hardware error events multiple ways:

1. **Repair Actions (RA)** are related to the industry standard definition of Mean Time Between Failure (MTBF). A RA is any hardware event that requires service on a system. These include incidents that effect system availability along with incidents that are concurrently repaired.
2. **Unscheduled Incident Repair Action (UIRA).** A UIRA is a hardware event that causes the system to be rebooted in a full or degraded mode. The system will experience an unscheduled outage. The restart may include some level of capability degradation but remaining resources are available for productive work.

   The IBM @server p5 servers include a modification to the platform reboot function that reduces the system recovery/reboot time after system checkstop, bypassing system Initial Program Load (IPL) diagnostic testing. In cases where the FFDC mechanism has identified a single failing component, the system is rebooted without running further extensive diagnostic testing, resulting in faster recovery from unscheduled (and potentially costly) outages.
3. **High Impact Outage (HIO).** A HIO is a hardware failure that causes a system crash that is not recoverable by immediate reboot. This is usually as a result of a failure of a component that is critical to system operation and is, in some sense, a measure of system single points of failure. HIOs result in the most significant availability impact on the system, since repairs cannot be effected without a service call.

   IBM's high-end @server p5 servers are designed with improved redundancy and recovery over predecessor systems to help avoid high impact outages. These include redundant fans, bulk power supplies, power regulators, memory bit lines, and L2 cache slices. In the first half of 2005, IBM intends to offer a no charge firmware update enabling redundant Service Processors on selected systems. Two system clocks and two Service Processors are required in all p5-595 and p5-590 configurations. If a system clock fails, a server reboot is required.

4. In partitioned environments, an additional metric is required: **Single Partition Impact Event**, referring to the instances where a fault occurs that causes the termination of one partition in a system, but allows the POWER Hypervisor and other partitions in the system to remain active.

   IBM @server p5 servers can be enabled to reboot a partition automatically on an abnormal OS termination event (for a hardware or OS kernel error). Since the POWER Hypervisor remains active during the reboot of a single partition, the reboot is a rapid event largely governed by the time it takes for an operating system to reload and become available. It should be noted that the POWER Hypervisor is active even in systems with a single partition. Therefore, even large, single system image servers that terminate due to hardware faults that impact the partition but not the POWER Hypervisor can also be swiftly rebooted.

Phases of Problem Determination

@server p5 servers employ three main techniques for fault isolation

1. Performing power-up testing for validation of correct system operation at startup (Platform IPL).
2. Monitoring the system during normal operation via FFDC strategies.
3. Employing Operating System (OS) -based monitoring and error handling for conditions not contained in FFDC error domains (e.g., PCI adapters, I/O drawers).

Platform Initial Program Load

At system power-on, the Service Processor initializes the system hardware. IPL testing employs a multi-tier approach for system validation. Servers include Service Processor managed low-level diagnostics supplemented with system firmware initialization and configuration of I/O hardware, followed by OS-initiated software test routines.

As part of the initialization, the Service Processor can assist in performing a number of different tests on the basic hardware. These include:

1. Built-in-Self-Tests (BIST) for both logic components and arrays. These tests deal with the internal integrity of components. The Service Processor assists in performing tests capable of detecting errors within components. These tests can be run for fault determination and isolation, whether or not system processors are operational, and they may find faults not otherwise detectable by processor-based Power-on-Self-Test (POST) or diagnostics.
2. Wire-Tests discover and precisely identify connection faults between components. For example, between processors and memory or I/O hub chips.
3. Initialization of components. Initializing memory, typically by writing patterns of data and letting the server store valid ECC for each location (and detecting faults through this process) is an example of this operation.

Faulty components detected at this stage can be:

1. Repaired where built in redundancy allows (e.g., fans, power supplies, spare cache bit lines).
2. Deallocated to allow the system to continue booting in a degraded mode (e.g., processors, sections of memory, I/O adapters).

   If a faulty CPU is detected, and a Capacity Upgrade on Demand (CUoD) replacement is available, the system will “vary on” the spare component using Dynamic Processor Sparing. All the physical installed memory in a server is made available to the POWER Hypervisor by the Service Processor. The POWER Hypervisor will allow a system administrator to configure only the amount of the physical memory purchased in the system configuration. If some physical memory has been marked as bad by the SP, the POWER Hypervisor will automatically use CUoD memory, if available, at the next server IPL to provide a full system configuration. Repair of the faulty processor, I/O adapter, or memory can be scheduled at a later date.

In all cases, the problem will be logged and reported for repair.

Finally, a set of OS diagnostic routines will be employed during an OS IPL stage to both configure external devices and to confirm their correct operation. These tests are primarily oriented to I/O devices (disk drives, PCI adapters, I/O drawers).

Run-time Monitoring

All IBM @server p5 servers include the ability to monitor critical system components during run-time and to take corrective actions when recoverable faults occur (power supply and fan status, environmental conditions, logic design). The hardware error check architecture supports the ability to report non-critical errors in an "out of band" communications path to the Service Processor, without impacting system performance

The Service Processor includes extensive diagnostic and fault analysis routines developed and improved over many generations of IBM POWER processor-based servers that allow quick and accurate predefined responses to actual and potential system problems.

The Service Processor correlates and processes error information, using error "thresholding" and other techniques to determine when action needs to be taken. Thresholding is the ability to use historical data and engineering know-how to count recoverable errors and accurately predict when corrective actions should be initiated by the system. These actions can include:

1. Requests for a part to be replaced.
2. Dynamic (on-line) invocation of built-in redundancy to automatically replace a failing component.
3. Dynamic deallocation of failing components so that system availability is maintained.

While many hardware faults are discovered and corrected during system boot time via diagnostics other (potential) faults can be detected, corrected or recovered during run-time.

Run-time Diagnostics and Error Log Analysis

During operation, the system will employ Operating System-specific diagnostics to determine problems, primarily with I/O devices. In these cases, the OS device driver will often work in conjunction with I/O device microcode to isolate and/or recover from problems.

For example, most direct access storage devices (DASD) implement some form of predictive failure analysis. Potential problems are reported to an OS device driver, which logs the error. An Error Log Analysis program is then used to analyze the error to determine if it is a serviceable action. The ELA application will perform client notification on serviceable events. Serviceable events on systems using an HMC will forward service information to the Service Focal Point (SFP) application on the HMC. The SFP collects and correlates error log data from the multiple partitions, consolidating error reports and initiating the "call home" service process when remote support available. Error analysis and reporting is discussed in detail in the serviceability section of this document.

I/O devices may also include specific exercisers that can be invoked by the diagnostic facilities.

Unrecoverable Fault

A system encounters an unrecoverable fault when the full system, or some subsystem, must be terminated because computational data integrity can not be assured.

An example of an unrecoverable fault is a multi-bit data error that can not be corrected by Error Correction Code routines.

The amount of affected hardware depends on the nature of the error and the element using the data. Regardless of the severity of the fault however, the Service Processor and FFDC techniques can be used in an IBM @server p5 system to determine the cause of the fault and to invoke recovery techniques including:

1. Activating redundancy to self-heal elements (such as using a spare bit in an array).
2. Deconfiguring a faulty component to allow the rest of the system to run without a reoccurrence of the uncorrectable error (persistent deallocation).
3. Reporting the problem for service clearly identifying the failing elements (with no further requirement to recreate the event to diagnose the problem).

The primary goal of the FFDC approach is to identify the root cause of the failure and to initiate corrective action, without the need to run "recreate" diagnostics. This technique is used to remedy the problem so that a second failure will not occur.

Servers Designed for Improved Availability

The extensive system of FFDC error checkers also supports a strategy of predictive failure analysis; the ability to track “intermittent” correctable errors and to vary components off-line before they reach the point of “hard failure” causing a crash.

This methodology supports IBM’s autonomic computing initiative. The primary RAS design goal of any @server p5 server is to prevent unexpected application loss due to unscheduled server hardware outages. In this arena, the ability to self-diagnose and self-correct during run time, to automatically reconfigure to mitigate potential problems from “suspect” hardware, and the ability to “self-heal,” to automatically substitute good components for failing components, are all critical attributes of a quality server design.

System Deallocation of Failing Elements

Persistent Deallocation of Components

To enhance system availability, a component that is identified for deallocation or deconfiguration on an @server p5 server will be flagged for persistent deallocation. Component removal can occur either dynamically (while the system is running) or at boot-time (IPL), depending both on the type of fault and when the fault is detected.

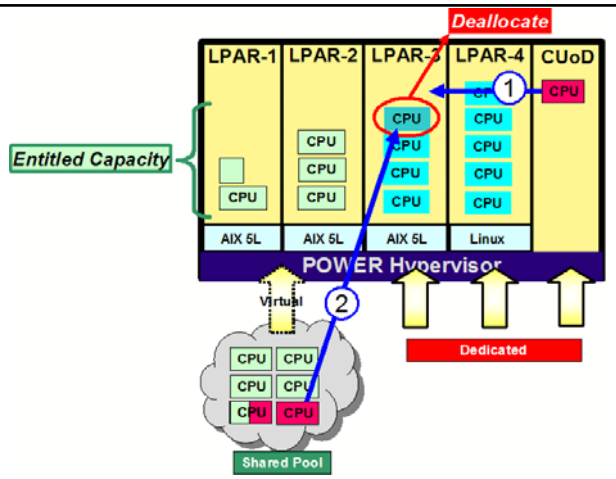
Run-time correctable/recoverable errors are monitored to determine if there is a pattern of errors or a “trend towards uncorrectability”. Should these components reach a predefined error limit, the Service Processor will initiate an action to deconfigure the “faulty” hardware, helping avoid a potential system outage, and enhancing system availability. Error limits are preset by IBM engineers based on historic patterns of component behavior in a variety of operating environments. Error thresholds are typically supported by algorithms that include a time-based count of recoverable errors; that is, the Service Processor responds to a condition of too many errors in a defined time span.

In addition, run-time unrecoverable hardware faults can be deconfigured from the system after the first occurrence. The system can be rebooted immediately after failure and resume operation on the remaining good hardware. This prevents the same “faulty” hardware from impacting the system operation again while the repair action is deferred to a more convenient, less critical time for the user operation.

Dynamic Processor Deallocation and Dynamic Processor Sparing

First introduced with the IBM RS/6000® S80 server and supported by AIX® Version 4.3.3 and AIX 5L, Dynamic Processor Deallocation is the ability for a system to automatically deconfigure an error-prone processor before it causes an unrecoverable system error (unscheduled server outage). Dynamic Processor Deallocation relies on the Service Processor’s ability to use FFDC generated recoverable-error information and to notify the AIX 5L or Linux operating system when the processor reaches its’ predefined error limit. The OS will then “drain” the run-queue for that CPU, redistribute the work to the remaining CPUs, deallocate the offending CPU, and continue normal operation, although potentially at a lower level of system performance. While AIX V4.3.3 precluded the ability for a SMP server to revert to a uniprocessor (i.e., a 2-way to a 1-way configuration), this limitation was lifted with the release of AIX 5L Version 5.1.

AIX 5L Version 5.2 support for dynamic Logical Partitioning allowed additional system



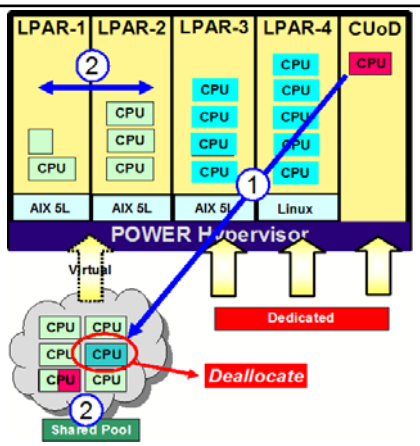
Should a POWER5 processor in a dedicated partition reach a predefined recoverable error threshold, the server can automatically substitute a spare processor before the faulty processor crashes. The spare CPU is logically moved to the target system partition; the POWER Hypervisor moves the workload and deallocates the faulty CPU for deferred repair.

Capacity Upgrade on Demand processors will always be selected first by the system for this process. As a second alternative, the POWER Hypervisor will check to see if there is sufficient capacity in the shared processor pool to make a CPU available for this operation.

availability improvements. An @server p5 server that includes an unlicensed CPU (an unused CPU included in a "Capacity Upgrade on Demand" system configuration) can be configured for Dynamic Processor Sparing. In this case, as a system option, the unlicensed CPU can automatically be used to "back-fill" for the deallocated bad processor. In most cases, this operation is transparent to the system administrator and to end users. The spare CPU is logically moved to the target system partition, the POWER Hypervisor moves the workload, and the failing processor is deallocated. The server continues normal operation with full functionality and full performance. The system generates an error message for inclusion in the error logs calling for deferred maintenance of the faulty component.

POWER5 technology and AIX 5L Version 5.3 introduce new levels of virtualization, supporting Micro-Partitioning™ technology, allowing individual processors to run as many as ten copies of the operating system. These new capabilities allow improvements in the Dynamic Processor Sparing strategy. POWER5 chips will support both dedicated processor logical partitions and shared processor dynamic LPARs. Dedicated processor partitions, supporting AIX 5L Version 5.2 and V5.3, operate like POWER4 processor-based system logical partitions. In a dedicated processor LPAR, one or more physical CPUs are assigned to the partition.

In shared processor partitions, supported by AIX 5L V5.3 and Linux, a "shared pool" of physical processors is defined. This shared processor pool consists of one or more physical processors. Up to ten logical partitions can be defined for every physical processor in the pool. Thus a six processor shared pool can support up to sixty logical partitions. In this environment, partitions are defined to include virtual processor and processor entitlements. Entitlements can be considered to be performance equivalents; for example, a logical partition can be defined to include 1.7 POWER5 processors worth of performance.



Dynamic Processor Deallocation from the shared pool uses a similar strategy (but may affect up to ten AIX 5L or Linux partitions). First, look for available CUoD processor(s). If not available, determine if there is one CPU's worth of performance available in the pool. If so, rebalance the pool to allocate the unused resource.

If the shared pool doesn't have enough available resource, query the partitions and attempt to reduce entitled capacities to obtain the needed performance.

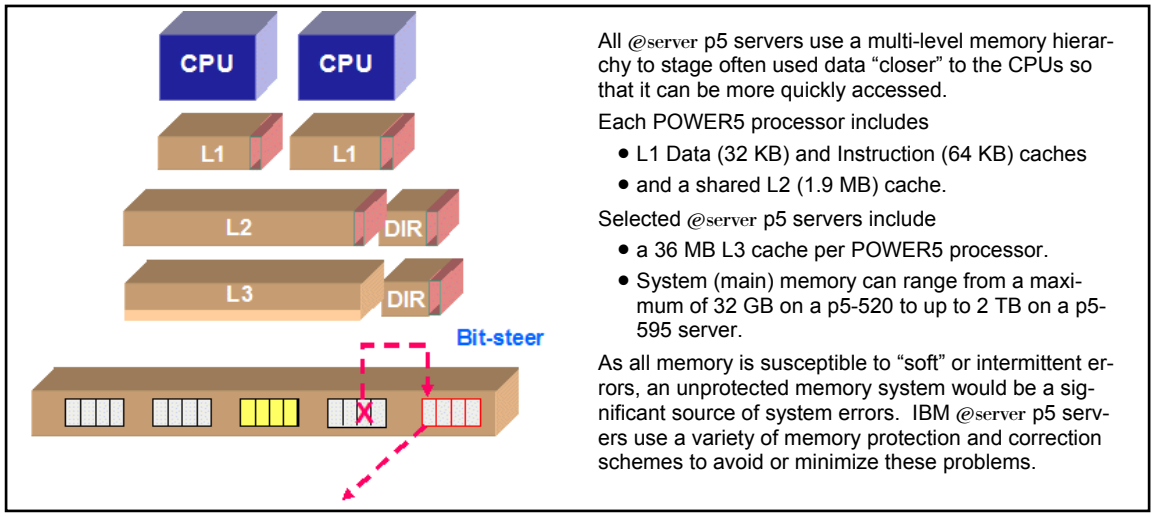
In dedicated POWER5 processor partitions, Dynamic Processor Sparing is transparent to the operating system. When a CPU reaches its error threshold, the Service Processor notifies the POWER Hypervisor to initiate a deallocation event.

- If a CUoD processor is available, the POWER Hypervisor automatically substitutes it for the faulty processor and then deallocates the failing CPU.
- If no CUoD processor is available, the POWER Hypervisor checks for excess processor capacity (capacity available because processors are unallocated or unlicensed). The POWER Hypervisor substitutes an available processor for the failing CPU.
- If there are no available processors for sparing, the operating system is asked to deallocate the CPU. When the operating system finishes the operation, the POWER Hypervisor stops the failing CPU.

Dynamic Processor Sparing in shared processor partitions operates in a similar fashion as in dedicated processor partitions. In both environments, the POWER Hypervisor is notified by the Service Processor of the error. As previously described, the system first uses any CUoD processor(s). Next, the POWER Hypervisor determines if there is at least 1.00 processor unit's worth of performance capacity available, and if so, stops the failing processor and redistributes the workload.

If the requisite spare capacity is not available, the POWER Hypervisor will determine how many processor capacity units each partition will need to relinquish to create at least 1.00 processor capacity units. The POWER Hypervisor uses an algorithm based on partition utilization and the defined partition minimum and maximums for CPU equivalents to calculate capacity units to be requested from each partition. The POWER Hypervisor will then notify the operating system (via an error entry) that processor units and/or virtual processors need to be varied off. Once a full processor equivalent is attained, the CPU deallocation event occurs. The deallocation event will not be successful if the POWER Hypervisor and OS cannot create a full processor equivalent. This will result in an error message and the requirement for a system administrator to take corrective action. In all cases, a log entry will be made for each partition that could use the physical processor in question.

Protecting Data in Memory Arrays



All @server p5 servers use a multi-level memory hierarchy to stage often used data "closer" to the CPUs so that it can be more quickly accessed.

Each POWER5 processor includes

- L1 Data (32 KB) and Instruction (64 KB) caches
- and a shared L2 (1.9 MB) cache.

Selected @server p5 servers include

- a 36 MB L3 cache per POWER5 processor.
- System (main) memory can range from a maximum of 32 GB on a p5-520 to up to 2 TB on a p5-595 server.

As all memory is susceptible to "soft" or intermittent errors, an unprotected memory system would be a significant source of system errors. IBM @server p5 servers use a variety of memory protection and correction schemes to avoid or minimize these problems.

Modern computers offer a wide variety of memory sizes, access speeds, and performance characteristics. System design goals dictate that some optimized mix of memory types be included in any system design so that the server can achieve demanding cost and performance targets.

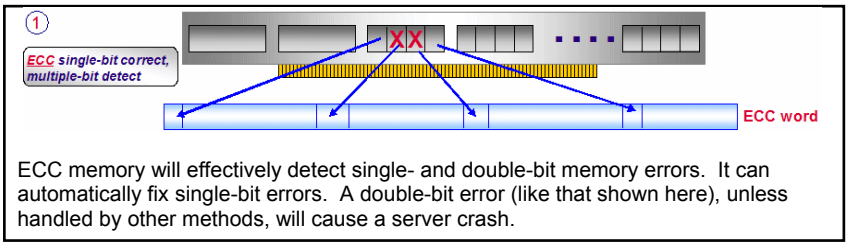
Powered by IBM's most advanced 64-bit POWER microprocessor, POWER5, @server p5 systems are designed to deliver extraordinary power and reliability, include simultaneous multi-threading, which makes each processor look like two to the operating system, increasing commercial performance and system utilization over servers without simultaneous multi-threading capabilities. To support these characteristics, IBM @server p5 systems employ a multi-tiered memory hierarchy with L1, L2, and L3 caches all staging main memory data for the processor. Each of these memory arrays will generate a different set of memory challenges for the RAS engineer.

Memory and cache arrays are comprised of data "bit lines" that feed into a memory word. A memory word is addressed by the system as a single element. Depending on the size and addressability of the memory element, each data bit line may include thousands of individual bits (memory cells). For example:

- A single memory module on a memory DIMM (Dual Inline Memory Module) may have a capacity of 1 Gbits, and supply 8 "bit lines" of data for an ECC word. In this case, each bit line in the ECC word holds 128 Mbits behind it (this corresponds to more than 128 million memory cell addresses).
- A 32 KB L1 cache with a 16-byte memory word, on the other hand, would only have 2 Kbits behind each memory bit line.

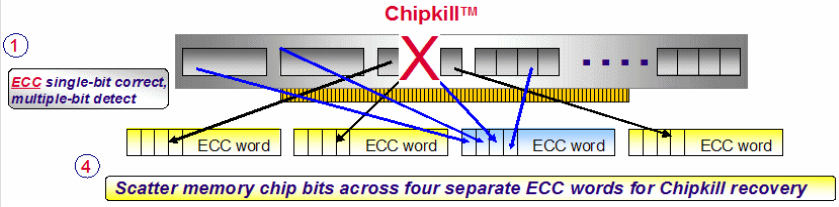
A memory protection architecture that provides good error resilience for a relatively small L1 cache may be totally inadequate for protecting the much larger system main store. Therefore, a variety of different protection schemes are used to avoid uncorrectable errors in memory. Memory protection plans must take into account many factors including size, desired performance, and memory array manufacturing characteristics.

One of the simplest memory protection schemes uses parity memory. A parity checking algorithm adds an extra memory bit (or bits) to a memory word. This additional bit holds information about the data that can be used to detect at least a single-bit memory error but usually doesn't include enough information on the nature of the error to allow correction. In relatively small memory stores (caches for example) that allow incorrect data to be discarded and replaced with correct data from another source, parity with retry (refresh) on error may be a sufficiently reliable methodology.



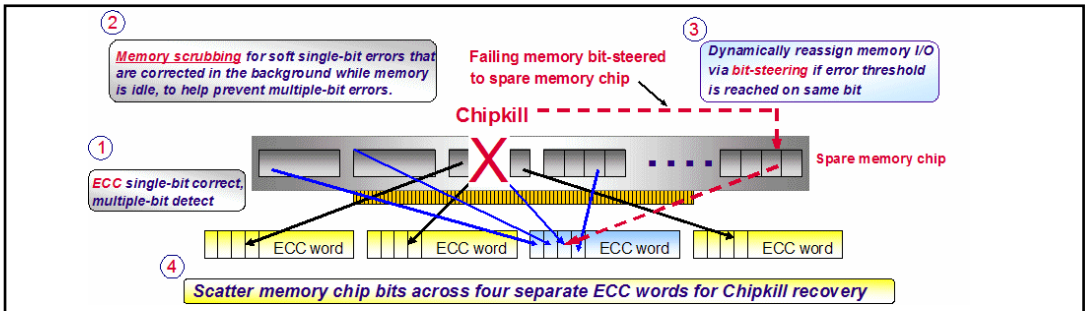
ECC memory will effectively detect single- and double-bit memory errors. It can automatically fix single-bit errors. A double-bit error (like that shown here), unless handled by other methods, will cause a server crash.

Error Correction Code (ECC) is an expansion and improvement of parity since the system now includes a number of extra bits in each memory word. The additional saved information allows the system to detect single- and double-bit errors. In addition, since the bit location of a single-bit error can be identified, the memory subsystem can automatically correct the error (by simply "flipping" the bit from "0" to "1" or vice versa). This technique provides an in-line mechanism for error detection and correction. No "retry" mechanism is required. A memory word protected with ECC can correct single-bit errors without any further degradation in performance. ECC provides adequate memory resilience, but may become insufficient for larger memory arrays, such as those found in main system memory. In very large arrays, the possibility of failure is increased by the potential failure of two adjacent memory bits or the failure of an entire memory chip.



IBM Chipkill memory can allow a server continue to operate without degradation after even a full memory chip failure.

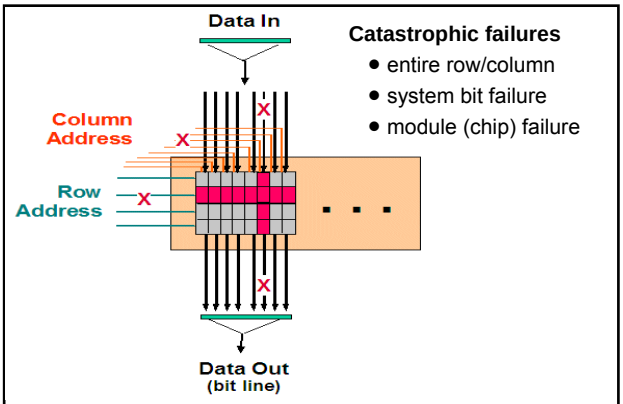
IBM engineers deployed a memory organization technique that spreads out the bits (bit lines) from a single memory chip over multiple ECC checkers (ECC words). In the simplest case, the memory subsystem distributes each bit (bit line) from a single memory chip to a separate ECC word. The server can automatically correct even multi-bit errors in a single memory chip. In this scheme, even if an entire memory chip fails, its errors are seen by the memory subsystem as a series of correctable single-bit errors. This has been aptly named Chipkill detection and correction. This means that an entire memory module can be bad in a memory group, and if there are no other memory errors, the system can run correcting single-bit memory errors with no performance degradation.



The highly resilient system memory in @server p5 systems includes multiple memory availability technologies: (1) ECC, (2) memory scrubbing, (3) dynamic bit-steering, and (4) Chipkill memory.

Transient or soft memory errors (intermittent errors caused by noise or other cosmic effects) that impact a single cell in memory can be corrected by parity with retry or ECC without further problem. IBM @server p5 systems proactively attempt to remove these faults using a hardware-assisted "memory scrubbing" technique where all the memory is periodically addressed and any address with an ECC error is rewritten with the faulty data corrected. Memory scrubbing is the process of reading the contents of memory through the ECC logic during idle time and checking and correcting any single-bit errors that have accumulated. In this way, soft errors are automatically removed from memory, decreasing the chances of encountering multi-bit memory errors.

However, even with ECC protection, intermittent or solid failures in a memory area can present a problem if they align with another failure somewhere else in an ECC word. This condition can lead to an uncorrectable memory error.



Catastrophic failures at a memory location can result in unrecoverable errors since this bit line will encounter a solid error. Unless this bit position is invalidated (by a technique like dynamic bit steering) any future solid or intermittent error at the same address will result in a system uncorrectable error and could cause a system crash.

Several error types can occur:

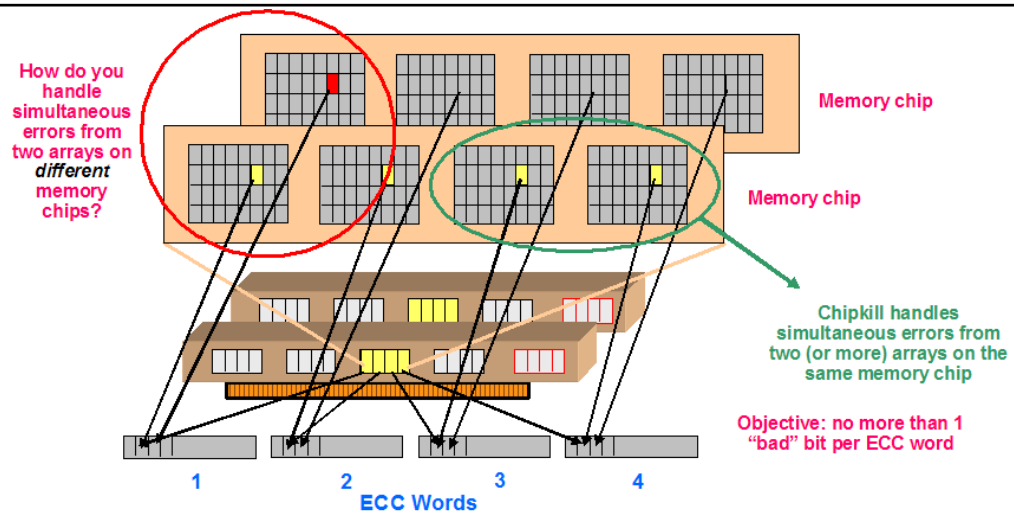
1. Errors confined to a single cell in memory (e.g., a data bit fault at single address.)
2. Faults in the memory device logic causing bad data to be delivered to a data bit for many or all of the addresses in the memory module.

Depending on the size of the memory, the latter case, while statistically a much rarer occurrence, can have a much bigger impact on server reliability than the former. Consider, for instance, a memory module with 128 Mbits behind each data line. Suppose that two separate data bits encounter a single memory cell failure. A potentially uncorrectable error would occur only if the bits have the same address. The odds of this randomly occurring are extremely low. On the other hand if there were a single cell failure in one module and a catastrophic failure of a bit line in another, this would result in a double-bit error every time.

To avoid uncorrectable errors in memory, IBM uses a dynamic spare memory scheme called “redundant bit-steering.” IBM main store includes spare memory bits with each ECC word. If a memory bit line is seen to have a solid or intermittent fault (as opposed to a transient error) at a substantial number of addresses within a bit line array, the system moves the data stored at this bit line to the spare memory bit line. IBM @server p5 systems can automatically and dynamically “steer” data to the redundant bit position as necessary during system operation.

This level of protection guards against the most likely uncorrectable errors within the memory itself: an alignment of a bit line failure with a future bit line failure as well as the alignment of a bit line failure with a memory cell failure (transient or otherwise) in another memory module.

For correctable errors, memory need not be replaced unless all of the redundant bits have been used and the recoverable threshold is reached. In this case, the server will generate a system error message calling for deferred maintenance. The server will continue to operate normally and the memory board can be replaced during some future scheduled maintenance activity.



IBM Chipkill memory has shown to be more than 100 times more reliable than ECC memory alone. The next challenge in memory design is to handle multiple-bit errors from different memory chips. Dynamic bit-steering resolves many of these errors.

Finally, should an uncorrectable error occur, the system can deallocate the memory group associated with the error on all subsequent system reboots until the memory is repaired. This is intended to guard against future uncorrectable errors while waiting for parts replacement.

Uncorrectable Error Handling

While it's a rare occurrence, an uncorrectable data error can occur in memory or a cache, despite all precautions built into the server. In older generations of servers (prior to IBM POWER4 processor-based offerings), this type of error would eventually result in a system crash. The IBM @server p5 systems extend the POWER4 technology design and include techniques for handling these types of errors.

On these servers, when an uncorrectable error (UE) is identified at one of the many checkers strategically deployed throughout the system's central electronic complex, the detecting hardware modifies the ECC word associated with the data, creating a special ECC code. This code indicates that an uncorrectable error has been identified at the data source and that the data in the "standard" ECC word is no longer valid. The check hardware also signals the Service Processor and identifies the source of the error. The Service Processor then takes appropriate action to handle the error. This technique is called Special Uncorrectable Error (SUE) handling.

Simply detecting an error does not automatically cause termination of a system or partition. In many cases, a UE will cause generation of a synchronous machine check interrupt. The machine check interrupt occurs when a processor tries to load the bad data. The firmware provides a pointer to the instruction that referred to the corrupt data, the system continues to operate normally, and the hardware observes the use of the data. The system is designed to mitigate the problem using a number of approaches:

1. If, as may sometimes be the case, the data is never actually used, but is simply over-written, then the error condition can safely be voided and the system will continue to operate normally.
2. If the data is actually referenced for use by a process (in AIX 5L 5.2 or greater) then the OS is informed of the error. The OS will terminate only the specific user process associated with the corrupt data.
3. If the data were destined for an I/O subsystem managed by the AIX 5L kernel, or for the kernel itself, then only the partition associated with the data would be rebooted. All other system partitions would continue normal operation.
4. If the data is to written to disk, the I/O hardware detects the presence of SUE and the I/O transaction is aborted
5. Finally, only in the case where the corrupt data is used by the POWER Hypervisor would the entire system be terminated and automatically rebooted, preserving overall system integrity.

Memory Deconfiguration and Sparing

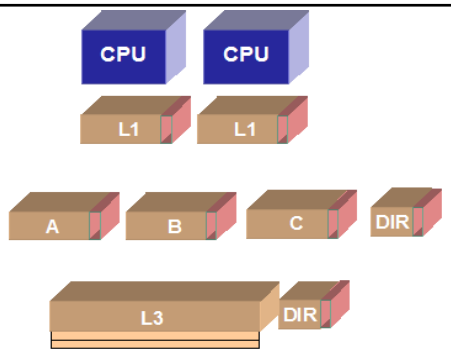
Defective memory discovered at IPL time will be switched off by a server.

1. If a memory fault is detected by the Service Processor at boot time, the affected memory will be marked as bad and will not be used on this or subsequent IPLs (Memory Persistent Deallocation).
2. All the physical installed memory in a server is made available to the POWER Hypervisor by the Service Processor at boot time. If the SP identifies faulty memory in a server that includes CUoD memory, the POWER Hypervisor counts the failed memory towards the unlicensed resources on the system. All resources are created equal; the POWER Hypervisor decides which memory to make available for server use and which to put in the unlicensed/spare pool based on system performance implications. As faulty resources are automatically "demoted" to the system's unlicensed resource pool, working resources are included in the active memory space.
3. If no spare memory is available, then POWER Hypervisor will reduce the capacity of one or more partitions. The HMC receives notification of the failed component, triggering a service call.

L3 Cache

The L3 cache is protected by ECC and Special Uncorrectable Error handling. The L3 cache also incorporates technology to handle memory cell errors via a special cache line delete algorithm.

During CEC IPL, if a solid error is detected during L3 initialization, a full L3 cache line will be deleted. During system run-time, a correctable error is reported as a recoverable error to the Service Processor. If an individual cache line reaches its predictive error threshold, it will be dynamically deleted. The state of L3 cache line delete will be maintained in a "deallocation record" and will persist through system IPL.



The L1 I-cache, L1 D-cache, L2 cache, L2 directory and L3 directory all contain additional or "spare" redundant array bits. These bits can be accessed by programmable address logic during system IPL. Should an array problem be detected, the Array Persistent Deallocation feature will allow the system to automatically "replace" the failing bit position with an available spare.

In addition, during system run-time, a correctable L3 error is reported as a recoverable error to the Service Processor. If an individual cache line reaches its predictive error threshold, it will be dynamically deleted. @server p5 servers can dynamically delete up to ten cache lines. It is not likely that deletion of a couple of cache lines will adversely affect server performance.

This insures that cache lines "varied offline" by the server will remain offline should the server be rebooted. These "error prone" lines can not then cause system operational problems. In the @server p5 product family, the server can dynamically delete up to ten cache lines. It is not likely that deletion of a couple of cache lines will adversely affect server performance. If this total (10) is reached, the L3 is marked for persistent deconfiguration on subsequent system reboots until repair.

Array Recovery and Array Persistent Deallocation

The L1 instruction cache (I-cache), directory, and instruction effective to real address translation (I-ERAT) are protected by parity. If a parity error is detected, it is reported as a cache miss or ERAT miss. The cache line with parity error is invalidated by hardware and the data is refetched from the L2 cache. If the error reoccurs (the error is solid) or if the cache reaches its soft error limit, the processor is dynamically deallocated and an error message for the FRU is generated.

While the L1 data cache (D-cache) is also parity checked, it gets special consideration when the threshold for correctable errors is exceeded. The error is reported as a synchronous machine check interrupt. The error handler for this event is executed in the POWER Hypervisor. If the error is recoverable the POWER Hypervisor invalidates the cache (clearing the error). If additional soft errors occur, the POWER Hypervisor will disable the failing portion of the L1 D-cache when the system meets its error threshold. The processor continues to run with degraded performance. A service action error log is created so that when the machine is booted, the failing part can be replaced. The data ERAT and TLB (translation look aside buffer) arrays are handled in a similar manner.

The L2 cache is protected by ECC. The ECC codes provide single-bit error correction and double-bit error detection. Single-bit errors will be corrected before forwarding to the processor. Corrected data is written back to L2. Like the other data caches and main memory, uncorrectable errors are handled during run-time by the Special Uncorrectable Error handling mechanism. Correctable cache errors are logged and if the error reaches a threshold, a Dynamic Processor Deallocation event is initiated.

Array Persistent Deallocation refers to the fault resilience of the arrays in a POWER5 microprocessor. The L1 I-cache, L1 D-cache, L2 cache, L2 directory and L3 directory all contain redundant array bits. If a fault is detected, these arrays can be repaired during IPL by replacing the faulty array bit(s) with the built-in redundancy, in many cases avoiding a part replacement.

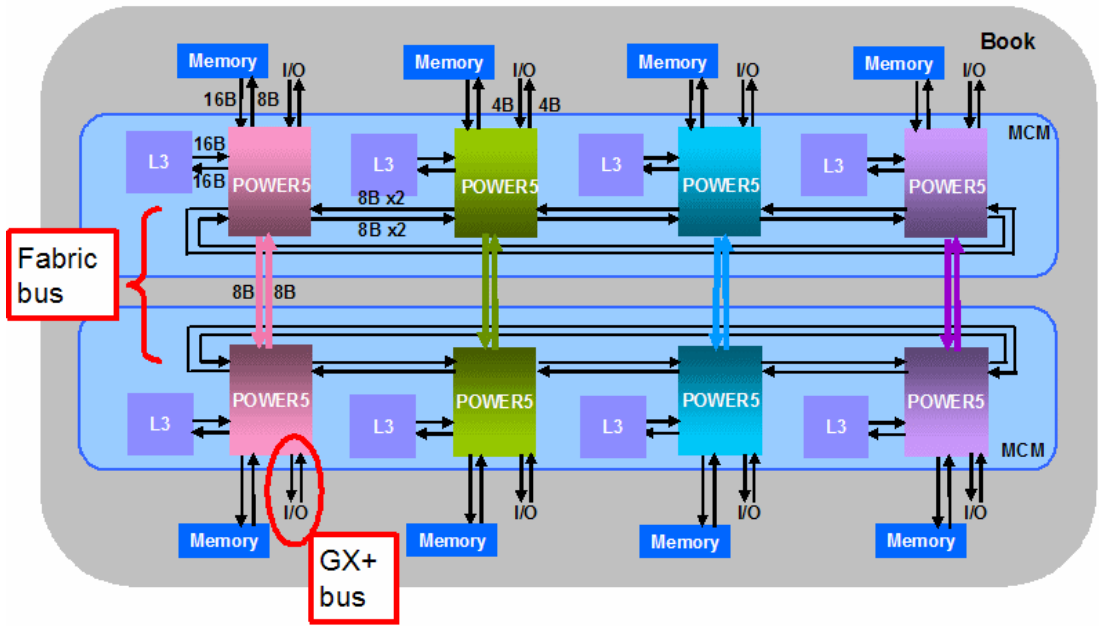
The initial state of the array "repair data" is stored in the FRU Vital Product Data (VPD) by manufacturing. During the first server IPL, the array "repair data" from the VPD is used for initialization. If an array fault is detected in an array with redundancy by the Array Built-In-Self-Test diagnostic, the faulty array bit is replaced. Then the updated array "repair data" is stored in the Service Processor persistent storage as part of the "deallocation record" of the processor. This repair data is used for subsequent system boots.

During system run time, the Service Processor monitors recoverable errors in these arrays. If a predefined error threshold for a specific array is reached, the Service Processor tags the error as "pending" in the deallocation record to indicate that the error is repairable by the system during next system IPL. The error is logged as a predictive error, repairable via re-IPL, avoiding a FRU replacement if the repair is successful.

For all processor caches, if "repair on reboot" doesn't fix the problem, the processor containing the cache can be deconfigured.

The Input Output Subsystem

A Server Designed for High Bandwidth and Reduced Latency



All IBM @server p5 servers use a unique "distributed switch" topology providing high bandwidth data busses for fast efficient operation. The high-end @server p5 590 and 595 servers use a 16-way building block. System interconnects scale with processor speed. Intra-MCM busses run at processor speed, Inter-MCM busses at ½ processor speed. Data movement on the fabric is protected by a full ECC strategy. The GX+ bus is the primary I/O connection path and operates at 1/3 processor speed.

This system topology results in reduced latency to the L3 cache and main memory and reduces the number of chips necessary to build a system when compared to earlier IBM offerings. Offering other innovations like simultaneous multi-threading processor cores; these servers deliver exceptional performance in both transaction processing and numeric intensive applications. The result is higher level of SMP scaling. IBM @server p5 systems support up to 64 physical processors.

I/O Drawer Redundant Connections

All @server p5 systems support integrated I/O devices (disk drives, PCI cards). The standard server I/O capacity can be significantly expanded in the rack-mounted offerings by attaching optional I/O drawers using IBM RIO-G busses. A remote I/O (RIO) loop includes two separate cables providing high speed drawer attachment. Should an I/O cable become inoperative during normal system operation, the system can automatically reconfigure to use the second cable for all data transmission until a repair can be made.

GX+ Bus Adapters

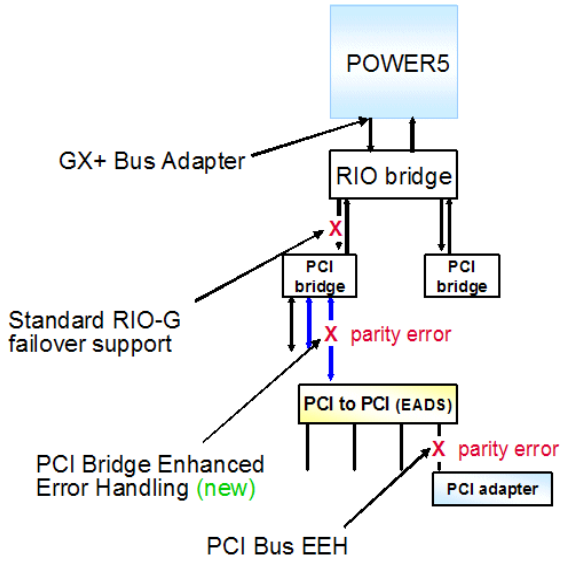
The GX+ bus provides the primary high bandwidth path for RIO connection to the system CEC. Errors in a GX+ bus adapter will be flagged by system "persistent deallocation" logic so that the adapter will be varied offline upon a server reboot for future repair.

PCI Bus Error Recovery

PCI adapters can account for a significant portion of the hardware based error opportunity on a large server, IBM estimates up to 25% of expected outages on fully configured servers. While servers that rely on "boot time" diagnostics can identify failing components to be replaced by "hot-swap" and reconfiguration, run time errors pose a more significant problem.

PCI adapters are generally complex designs involving extensive "on-board" instruction processing, often on embedded microcontrollers. Since these are generally cost sensitive designs, they typically use industry standard grade components, avoiding the more expensive but higher quality parts used in other parts of the server. As a result, they may encounter internal microcode errors, and/or many of the hardware errors described for the entire server. In general, these problems are handled through adapter internal error reporting and recovery techniques in combination with AIX 5L device driver management and diagnostics. In addition, an error in the adapter may cause transmission of bad data on the PCI bus itself, resulting in a hardware detected parity error (and causing a global machine check interrupt, eventually requiring a system reboot to continue). IBM introduced a methodology in the POWER4 processor-based servers that uses a combination of system firmware and new "Extended Error Handling" (EEH) device drivers to allow recovery from intermittent PCI bus errors (through recovery/reset of the adapter) and to initiate system recovery for a permanent PCI bus error (to include hot-plug replace of the failed adapter).

IBM has long built servers with redundant physical I/O paths using CRC checking and failover support to protect RIO server connections from the CEC to the I/O drawers. IBM extended data protection in the pSeries servers, introducing first-in-the-industry Extended Error Handling to allow recovery from PCI-bus error conditions. The @server p5 systems add additional recovery features to handle potential errors in the Processor Host Bridge (PCI bridge), and the GX+ bus adapter. These features provide improved diagnosis, isolation, and management of errors in the server I/O path and new opportunities for concurrent maintenance – to allow faster recovery from I/O path errors, often without impact to system operation.



POWER5 processor-based servers extend the capabilities of the EEH methodology. Generally, on @server p5 platforms, all PCI adapters controlled by operating system device drivers are connected to a PCI secondary bus created through an IBM designed PCI-PCI bridge. This bridge isolates the PCI adapters and supports "hot-plug" by allowing program control of the "power state" of the I/O slot. PCI bus errors related to individual PCI adapters under partition control can be transformed into a PCI slot freeze condition and reported to the EEH device driver for error handling. Errors that occur on the interface between the PCI-PCI bridge chip and the Processor Host Bridge (the link between the processor remote I/O bus and the primary PCI bus) result in a "bridge freeze" condition, effectively stopping all of the PCI adapters attached to the bridge chip. An operating system may recover an adapter from a bridge freeze condition by using POWER Hypervisor functions to remove the bridge from freeze state and resetting or reinitializing the adapters.

Miscellaneous Redundancy and Availability

POWER Hypervisor

Since the availability of the POWER Hypervisor is crucial to overall system availability, great care has been taken to design high quality, well tested code. Coding errors are significantly different from hardware errors. In general, a hardware system will see a higher than normal error rate when first introduced and/or when first installed in production. These types of errors are mitigated by strenuous engineering and manufacturing verification testing and using methodologies such as "burn in," designed to catch the fault before the server is shipped. At this point, hardware failures typically even out at relatively low, but fairly constant, error rates. This phase can last for many years. At some point, however, hardware failures may again increase as parts begin to "wear out." Clearly the "design for availability" techniques discussed here will help mitigate these problems.

However, unlike hardware, code can display a variable rate of failure. New code typically has a higher failure rate and older more seasoned code a very low rate of failure. Code quality will continue to improve as bugs are discovered and fixes installed. Although the POWER Hypervisor provides important system

functions, it is limited in size and complexity when compared to a full operating system implementation, and therefore can be considered to be better "contained" from a design and quality assurance viewpoint. As with any software development project, the IBM firmware development team writes code to strict guidelines using well-defined software engineering methods. The overall code architecture is reviewed and approved and each developer schedules a variety of peer code reviews. In addition, all code is strenuously tested, first by “visual” inspections, looking for logic errors, then by simulation and operation in actual test and production servers. Using this structure approach, most coding error are caught, and fixed, early in the design process.

The POWER Hypervisor is a converged design based on code used in iSeries™ and pSeries POWER4 processor-based servers. The development team selected the best firmware design from each platform for inclusion in the POWER Hypervisor. This not only helps reduce coding error, it also delivers new RAS functions that can improve the availability of the overall server. For example, the pSeries firmware had excellent, proven support for processor error detection and isolation and included support for dynamic CPU deallocation and sparing. The iSeries firmware had first-rate support for I/O recovery and error isolation and included support for errors like “cable pulls” (handling bad I/O cable connections).

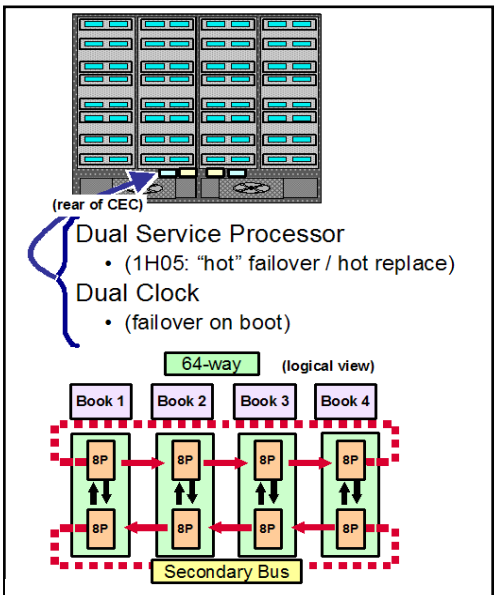
An inherent feature of the POWER Hypervisor is that the majority of the code runs in the protection domain of a hidden system partition. Failures in this code are limited to this system partition. Supporting a very robust tasking model, the code in the system partition is segmented into critical and non-critical tasks. If a non-critical task fails, the system partition is designed to continue to operate, albeit without the function provided by the failed task. Only in a rare instance of a failure to a critical task in the system partition would the entire POWER Hypervisor fail.

The resulting code provides not only advanced features but also superb reliability. It is used in the new @server i5 and @server p5 platforms and in the recently announced IBM TotalStorage® DS8000 series products. It has therefore been strenuously tested under a wide ranging set of system environments and configurations. This process has delivered a quality implementation that includes enhanced error isolation and recovery support when compared to POWER4 processor-based offerings.

Service Processor and Clocks

A number of availability improvements have been included in the Service Processor in the @server p5 servers. Separate copies of Service Processor microcode and the POWER Hypervisor code are stored in discrete Flash memory storage areas. Code access is CRC protected. The Service Processor performs low level hardware initialization and configuration of all processors. The POWER Hypervisor performs higher level configuration for features like the virtualization support required to run up to 254 partitions concurrently on the IBM @server p5 590 and 595 servers. The POWER Hypervisor enables many advanced functions; including sharing of processors, virtual I/O, and high-speed communications between partitions using Virtual LAN. AIX 5L, Linux, and i5/OS™ are supported. The servers also support dynamic firmware updates[1], in which applications remain operational while IBM system firmware is updated for most operations. Maintaining two copies insures that the Service Processor can run even if a Flash memory copy becomes corrupted, and allows for redundancy in the event of a problem during the upgrade of the firmware.

In addition, if the Service Processor encounters an error during run-time, it can reboot itself while the server system stays up and running. There will be no server application impact for Service Processor transient errors.



p5-595 and p5-590 servers support redundant clocks and service processors (located in the rear of the central electronics complex). Redundant system clocks are activated at server reboot. In a future firmware update (1H05), the system will allow dynamic failover of service processors at run-time.

If the Service Processor encounters a code “hang” condition, the POWER Hypervisor can detect the error and direct the Service Processor to reboot, avoiding other outage.

Two system clocks and two Service Processors are required in all p5-595 and p5-590 configurations.

1. A no charge firmware update will be provided in the first half of 2005 to enable redundant Service Processors. This firmware update will require a system reboot. Once installed, the firmware will allow an error condition in the primary Service Processor to be automatically detected. If the error threshold for the failing SP is reached, the system will initiate a failover from one Service Processor to the backup. Failovers can occur dynamically during run-time.
2. An IPL time failover will occur if a system clock should fail.

These and other internal enhancements in multiple building block servers are intended to make it nearly always the case that at least one building block will be available to IPL the server, regardless of the nature of a fault in any one system building block.

Availability in a Partitioned Environment

IBM's dynamic Logical Partitioning (LPAR) architecture has been extended with Micro-Partitioning technology capabilities. These new features are provided by the POWER Hypervisor and are configured using management interfaces on the HMC. This very powerful approach to partitioning maximizes partitioning flexibility and maintenance. It supports a consistent partitioning management interface just as applicable to single (full server) partitions as to systems with hundreds of partitions.

These new LPAR capabilities not only provide fine-grained resource allocation, but all the servers in IBM @server p5 product line include the underlying capability to individually assign any resource (processor, memory segment, I/O slot) to any partition in any combination. Not only does this allow exceptional configuration flexibility, it enables many high availability functions like:

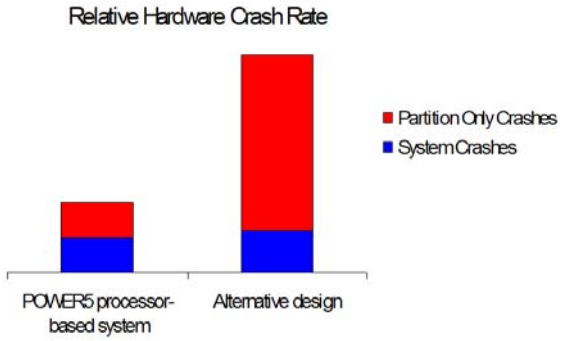
- Resource sparing (CPU Dynamic Deallocation and Dynamic Processor Sparing).
- Automatic redistribution of capacity on N-1 (automated shared pool redistribution of partition entitled capacities for Dynamic Processor Sparing).
- LPAR configurations with redundant I/O (across separate processor host bridges or even physical drawers) allowing system designers to build configurations with improved redundancy for automated recovery.
- The ability to reconfigure a server "on the fly." Since any I/O slot can be assigned to any partition, a system administrator can "vary off" a faulty I/O adapter and "back fill" with another available adapter, without waiting for a spare part to be delivered for service.
- Automated scale-up of high availability backup servers as required (via dynamic LPAR).
- Serialized sharing of devices (optical, tape) allowing "limited" use devices to be made available to all the partitions.
- Shared I/O devices through I/O server partitions. A single I/O slot can carry transactions on behalf of several partitions, potentially reducing the cost of deployment and improving the speed of provisioning of new partitions (new applications). Multiple I/O server partitions can be deployed for redundancy, giving partitions multiple paths to access data and improved availability in case of an adapter or I/O server partition outage.

In a logically partitioning architecture, all of the server memory is physically accessible to all the processors and all of the I/O devices in the system, regardless of physical placement of the memory or where the logical partition operates. The POWER4 and POWER5 Hypervisor mode with Real Memory Offset Facilities enables the POWER Hypervisor to insure that *any* code running in a partition (operating systems and firmware) only has access to the physical memory allocated to the dynamic logical partition. POWER4 and POWER5 processor-based systems also have IBM-designed PCI-to-PCI bridges that enable the POWER Hypervisor to restrict DMA (Direct Memory Access) from I/O devices to memory owned by the partition using the device. The single memory cache coherency domain design is a key requirement for delivering the highest levels of SMP performance. Since it is IBM's strategy to deliver hundreds of dynamically configurable logical partitions, allowing improved system utilization and reducing overall computing costs, these servers must be designed to avoid or minimize conditions that would cause a full server outage.

IBM's availability architecture provides a high level of protection to the individual components making up the memory coherence domain; including the memory, caches, and fabric bus. It also offers advanced

techniques designed to help contain failures in the coherency domain to a subset of the server. Through careful design, in many cases failures are contained to a component or to a partition, despite the shared hardware system design. Many of these techniques have been described in this document.

IBM's approach can be contrasted to alternative designs which group sub-segments of the server into isolated, relatively inflexible "hard physical partitions." Hard partitions are generally tied to CPU and memory board boundaries. Physically partitioning cedes flexibility and utilization for the "promise" of better availability, since a hardware fault in one partition will not normally cause errors in other partitions. Thus the user will see a single application outage, not a full system outage. However, if a system uses physical partitioning primarily to eliminate system failures (turning system faults into partition-only faults), then it's possible to have a very low system crash rate, but a high individual partition crash rate. This will lead to a high application outage rate, despite the physical partitioning approach. Many clients will hesitate to deploy "mission-critical" applications in such an environment.



IBM @server p5 systems are designed to minimize both faults which impact the entire system and faults that take down just an individual partition. While "hard partitioning" in servers may provide some protection against full system outages (assuming that single points of failure or repair are effectively eliminated), if the hard partitioning architecture is used to mitigate poor hardware reliability or a poor availability design, then the server may experience increased application outages.

IBM RAS engineers have built a RAS architecture that factors all aspects of availability design into a coherent, repeatable policy for building highly available servers. The new @server p5 systems build upon the extremely reliable POWER4 processor-based servers which have set a very high standard for system availability.

System level availability (in any server, no matter how partitioned) is a function of the reliability of the underlying hardware and the techniques used to mitigate the faults that do occur. The availability design of @server p5 systems minimizes system failures and localizes potential hardware faults to single partitions in multi-partition systems. In this design, while some hardware errors may cause a full system crash (causing loss of all partitions), since the rate of system crashes is very low, the rate of partition crashes is also very low.

The reliability and availability characteristics described in this document show how this "design for availability" approach is consistently applied throughout the system design. IBM believes this is the best approach to achieving partition level availability while supporting a truly flexible and manageable partitioning environment.

In addition, to achieve the highest levels of system availability, IBM also offers clustering solutions, including HACMP which allow for failover from one system to another, even geographically dispersed systems.

Serviceability

The Service strategy for the IBM @server p5 and @server i5 product families evolves from, and improves upon, the service architecture deployed on pSeries and iSeries Servers. The service team has enhanced the base service capability and continues to implement a strategy that incorporates best-of-breed service characteristics from various IBM @server systems including the xSeries®, iSeries, pSeries, and high-end zSeries systems.

The Service goal is to *provide the most efficient service environment* by designing a system package that incorporates:

- easy access to service components,
- on demand service education,
- an automated/guided repair strategy using common service interfaces for a converged service approach across multiple IBM server platforms.

The aim is to deliver faster and more accurate repair while reducing the possibility for human error.

The strategy contributes to higher systems availability with reduced maintenance costs. In many of the entry level systems, the server design supports client install and repair, allowing maximum client flexibility for controlling all aspects of their systems operations. This notion of client control of the service environment was extended to firmware maintenance on all of the IBM @server p5 Systems. When taken together, these factors can deliver increased value to the end user.

Service Environments

The IBM @server p5 and @server i5 systems support two main service environments:

1. Servers which do not include a Hardware Management Console. This is the default configuration for entry and mid-range systems.
2. Server configurations that include attachment to one or multiple HMCs. This is the default configuration for high-end systems and any server which supports logical partitions. In this case, all servers have at least one logical partition.

While some configurations will require an HMC, any server in the IBM @server p5 or @server i5 product line may optionally be connected to an HMC. This configuration delivers a variety of additional service benefits as described in the section discussing HMC based service.

Elements of Non-HMC Service Environment

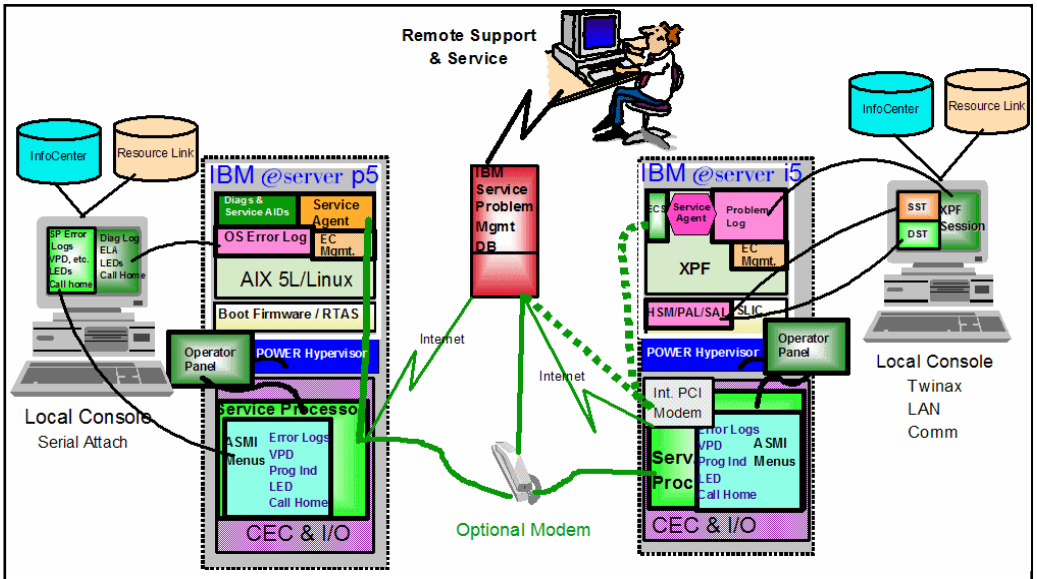
Both the IBM @server p5 and @server i5 system designs incorporate the same service elements for handling platform errors. These are defined to be errors related to:

- The Central Electronics Complex (CEC); that part of the server comprised of the central processor units, memory, storage controls and the I/O Hubs.
- The power and cooling subsystems.
- The firmware used to initialize the system and diagnose errors.

In addition, a common mechanism is used to handle base I/O errors out to, but not including the PCI adapters and devices.

While these servers generally employ a common design, some brand-unique service structures are incorporated to manage I/O subsystem service in support of I/O adapters and devices, since the device drivers reporting service events are operating system specific.

Service on non-HMC attached systems begins with the operating system service tools described below. If an error prevents the operating system from booting, the Systems Service Representative (SSR) will analyze the Service Processor and Operator Panel error logs. An IBM service application called “InfoCenter” guides an IBM SSR, or a third party servicer, through problem determination and problem resolution procedures. When the problem is isolated and repaired, InfoCenter will help the servicer verify correct system operation and close the service call. These components are described in more detail below.



Service on non-HMC attached systems begins with operating system service tools. If an error prevents the operating system from booting, the SSR will analyze the Service Processor and Operator Panel error logs. An IBM service application called "InfoCenter" guides the servicer through problem determination and problem resolution procedures. When the problem is isolated and repaired, InfoCenter will help the servicer verify correct system operation and close the service call. Other types of errors are handled by Service Processor tools and Operator Panel Messages.

IBM Support can be also enlisted to help using a remote service link.

Operator Panel

The Operator Panel on the IBM select POWER5 processor-based systems is a four row by sixteen element LCD display used to present boot progress codes indicating advancement through the system power-on and initialization processes. The Operator Panel is also used to display error and location codes when an error occurs that prevents the system booting. It includes several push-buttons allowing the SSR or the client to change various boot time options and a variety of other limited service functions.

Service Processor

As described earlier in this document, the Service Processor in the IBM @server p5 and @server i5 product families is an improved design when compared to the Service Processor that was available in the IBM POWER4 processor-based systems. The new Service Processor incorporates enhanced hardware functions such as an Ethernet service interface, additional serial port communications ports (these serial ports are not operational when an HMC is attached), and larger storage capacity. All of these new features support improved functions for service.

One important Service Processor improvement gives the system administrator or servicer the ability to dynamically access the Advanced Systems Management Interface (ASMI) menus. In previous generations of servers these menus were only accessible when the system was in standby power mode. Now the menus are now available from any Web browser-enabled console attached to the Ethernet service network concurrent with normal system operation. A user with the proper access authority and credentials can now dynamically modify service defaults, interrogate Service Processor progress and error logs, set and reset guiding light LEDs, indeed, access all Service Processor functions without having to power-down the system to the standby state.

On non-HMC based systems, on the rare occurrence of an uncorrectable checkstop error, the Service Processor will analyze the error, post an error message to the operator panel and initiate a call home for service if this function is enabled.

The Service Processor also manages the interfaces for connecting Uninterruptible Power Source (UPS) systems to the IBM POWER5 processor-based systems, performing Timed Power-On (TPO) sequences, and interfacing with the power and cooling subsystem.

Analyzing Errors: Operating System Logs, Error Log Analysis, and Service Agent

Since the Service Processor monitors the hardware environmental and FFDC (FIR bits) activities, it is the primary "catcher" of platform hardware errors and is used to begin analysis and processing of these events. The Service Processor will identify and sort errors by type and criticality. In effect, the Service Processor performs triage; initiating a preliminary error analysis to categorize events into specific categories:

1. Errors that are recoverable but should be recorded for threshold monitoring. These events do not require immediate service but should be logged and tracked to look for, and effectively respond to, future problems.
2. Fatal system errors (initiate server reboot/IPL, error analysis, and call home if enabled).
3. Recoverable errors that require service either because an error threshold has been reached or a component has been taken "off-line (even if a redundant component has been used for sparing).

When a recoverable and serviceable error (type 3 above) is encountered, the Service Processor notifies the POWER Hypervisor which places an entry into the operating system error log. The operating system log contains all recoverable error logs. These logs represent either recoverable platform errors or errors detected and logged by I/O device drivers. Operating System Error Log Analysis routines monitor this log, identify serviceable events (ignoring information-only log entries), and copy them to a diagnostic event log. At this point the operating system will send an error notification to a client designated user (by default, the root user). This action also invokes the Service Agent application which initiates appropriate system serviceability actions.

- On servers which do not include an HMC, the Service Agent notifies the system operator of the error condition and, if enabled, also initiates a call for service. The Service call can be directed to the IBM support organization, or to a client identified pager or service provider identified and set-up to receive service information.
- On servers equipped with an HMC, Service Agent forwards the results of the diagnostic error log analysis to the Service Focal Point application running on the HMC. The Service Focal Point consolidates and reports errors to IBM or a user designated system or pager.

In either case, failure information including:

- the source of error,
- the part numbers of the components needing repair,
- the location of those components,
- and any available extended error data

is sent back to IBM Service for parts ordering and additional diagnosis if required. This detailed error information enables IBM Service representatives to bring along probable replacement hardware components when a service call is placed, minimizing system repair time.

In a multi-system configuration, any HMC-attached @server p5 or @server i5 server can be configured to forward call home requests to a central Service Agent Gateway (SAG) application on a HMC which owns a modem and performs the call home on behalf of any of the servers. Similarly, Service Agent client code in a non-HMC attached system can be configured to forward call home requests to a central Service Agent Gateway on another non-HMC system which owns a modem for call home processing. At this time, HMC and non-HMC systems may not share a common Service Agent Gateway.

Converged Service Architecture

The IBM @server p5 and @server i5 systems represent a significant convergence of platform service architectures, merging the best characteristics of the iSeries and pSeries product offerings. This union allows similar maintenance approaches and common service user interfaces. A servicer can be trained on the maintenance of the base hardware platform, service tools, and associated service interface and be proficient in problem determination and repair for either POWER5 processor-based platform offering. In some cases, additional training may be required to allow support of I/O drawers and adapters and devices.

The convergence plan incorporates critical service topics.

- Identifying the failing component through architected error codes.

- Pinpointing the faulty part for service using location codes and LEDs as part of the guiding light diagnostic strategy.
- Ascertaining part numbers to quickly and efficiently order replacement components.
- Collecting system configuration information using common Vital Product Data which completely describes components in the system, to include detailed information such as their point of manufacture and Engineering Change (EC) level.
- Enabling service applications, such as Firmware and Hardware EC Management (described below) and Service Agent, to be portable across the multiple hardware and operating system environments.

The resulting commonality makes possible reduced maintenance costs and lower total cost of ownership for IBM @server p5 and @server i5 systems. This core architecture provides consistent service interfaces and a common approach to service, enabling owners of selected @server p5 or @server i5 servers to successfully perform set-up, manage and carry out maintenance, and install server upgrades; all at their own schedule and without available IBM support personnel.

IBM Service Problem Management Database

System error information can be transmitted electronically from the Service Processor for unrecoverable errors or from Service Agent for recoverable errors that have reached a Service Action Point. It can also be manually communicated by the client where no electronic call home capability is enabled. At the IBM support center this data is entered into an IBM Service and Support Problem Management database. All of the information related to the error, along with any service actions taken by the servicer are recorded for problem management by the support and development organizations. The problem is then tracked and monitored until the system fault is repaired.

When service calls are placed electronically, product application code on the front end of the problem management database searches for known firmware fixes (and for @server i5 systems, Operating System PTFs). If a fix is located, the system will download the updates for installation by the client. In this way, known problems with firmware or i5/OS fixes can be automatically sent to the system without the need for replacing hardware or dispatching a service representative.

Diagnostics

Because of the First Failure Data Capture technology employed in the IBM @server p5 and @server i5 systems, recreate diagnostics for CEC failures have been eliminated. As previously explained, the Service Processor working in conjunction with the FFDC technology provides the automatic detection and isolation of errors without having to recreate the failure. This means that solid and intermittent errors will be correctly detected and isolated at the time of the failure occurrence.

Diagnostics are provided for industry standard adapters and devices utilized in these systems.

InfoCenter

IBM @server Hardware Information Center (InfoCenter) is a repository of client and servicer related product information. The latest version of the documentation is accessible through the Internet; however, a CD-ROM based version is also available.

The purpose of InfoCenter, in addition to providing client related product information, is to provide softcopy service procedures to guide the servicer through various error isolation and repair procedures. Because they are electronically maintained, changes due to updates or addition of new capabilities can be used by servicers immediately.

InfoCenter also provides the capability to embed education-on-demand modules for the servicer to reference. The education-on-demand modules encompass information from detailed diagrams to movie clips showing specialized repair scenarios. Servicers can reference this material in the course of providing service to insure that the repair scenario is completed to proper specifications.

Resource Link

Similar to InfoCenter, Resource Link™ is electronic information repository. Resource Link provides on-line training and educational material; allowing service qualification for the various IBM @server p5 and @server i5 systems. Courseware can be downloaded and completed at any time. Using Resource Link, servicers can train for a new product or refresh their skills on specific systems without being tied to rigid classroom schedules that are dependent on instructor and class availability.

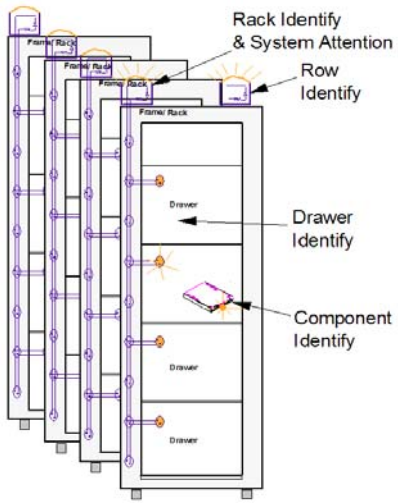
Guiding Light Diagnostics

Guiding light diagnostics are similar in concept to the lightpath diagnostics used in the xSeries server family to improve problem determination and isolation. Lightpath diagnostics use a series of LEDs (Light Emitting Diodes) to quickly guide a client or Service Support Representative to a failed hardware component so that it can be repaired or replaced. Guiding light LEDs support a similar system that is expanded to encompass the service complexities associated with high-end servers. Because of the inherent RAS features of the IBM @server p5 and @server i5 systems with capabilities like redundant power and cooling, redundant PCI adapters and devices, or Capacity Upgrade on Demand resources utilized for spare service capacity, it is technically feasible to have more than one error condition on a server at any point in time and still have the system be functional from a client and application point of view.

In the guiding light LED implementation, when a fault condition is detected on the IBM @server p5 or @server i5 system, an amber System Attention LED will be illuminated. Upon arrival at the server, a SSR or service provider sets the identify mode, selecting a specific problem to be identified for repair by the Guiding light method. The guiding light system pinpoints the exact part by flashing the amber identity LED associated with the part to be replaced.

The system can not only clearly identify components for replacement by using specific component level indicators, but can also "guide" the servicer directly to the component by signaling (causing to flash) the Rack/Frame System Identify indicator and the Drawer Identify indicator on the drawer containing the component. The flashing identify LEDs direct the servicer to the correct system, the correct enclosure, and the correct component.

In large multi-system configurations, optional row identify beacons can be added to indicate which row of racks contains the system to be repaired. Upon completion of the service event, the servicer resets the Identify LED indicator and the remaining hierarchical identify LEDs are automatically reset. If there are additional faults requiring service, the system attention LED will still be illuminated and the servicer can choose to set the identify mode and select the next component to be repaired. This provides a consistent unambiguous methodology for servicers to visually identify the component for repair in the case of multiple faults on the system. At the completion of the service process, the servicer resets the system attention LED indicating that all events requiring service have been repaired or acknowledged. Some service action requests may be scheduled for future deferred repair.



Guiding light diagnostics use a series of LEDs to lead a servicer directly to a component in need of repair. Using this technology, an IBM SSR or, in some cases, a client can select an error from the HMC. Rack, drawer, and component LEDs will "blink" to guide the servicer directly to the correct part. This technology can speed accurate and timely repair.

Blind-swap PCI Adapters

"Blind-swap" PCI adapters, first introduced in selected pSeries and iSeries servers in 2001, represent significant service and ease-of-use enhancements in I/O subsystem design. "Standard" PCI designs supporting "hot-add" and "hot-replace" require top access so that adapters can be slid to the PCI I/O slots vertically. This approach generally requires an I/O drawer to be slid out of its' rack and the drawer cover to be removed to provide component access for maintenance. While servers provided features such as cable management systems (cable guides) to prevent inadvertent accidents such as "cable pulls," this approach required moving an entire drawer of adapters and associated cables to access a single PCI adapter.

Blind-swap adapters mount PCI I/O cards in a carrier that can be slid into the rear of a server or I/O drawer. The carrier is designed so that the card is "guided" into place on a set of rails and seated in the slot, completing the electrical connection, by simply shifting an attached lever. This capability allows the PCI adapters to be concurrently replaced without having to put the I/O drawer into a service position. Since first delivered, minor carrier design adjustments have improved an already well-thought out service design. This technology has been incorporated in IBM @server p5 and @server i5 servers and I/O drawers. In addition, these features allow servicers to quickly and easily add additional I/O capacity, rebalance existing capacity, and effect repairs on PCI adapters.

Remote Support Capability

Dumps

In some cases, valuable problem determination and service information can be gathered using a system "dump" (for the POWER Hypervisor, memory, or Service Processor). Dumps can be initiated, automatically or "on request," for interrogation by IBM service and support or development personnel. Data collected by this operation can be transmitted back to IBM, or in some instances, can be remotely viewed utilizing special support tools if a client authorizes a remote connection to their system for IBM support personnel.

Firmware and Hardware Engineering Change (EC) Level Management

To improve system EC management, a variety of applications are run on the operating system and the HMC to survey the firmware and microcode levels installed on a system. These can generate a report to be sent back to the IBM support organization. A survey program informs the client of modules that are down level and attempts to retrieve the latest version available for download. In addition, an application tracks server hardware or configuration changes, gathering Vital Product Data on each component and transmitting it to an inventory tracking database at IBM for Service and Support use. IBM Support can access this database to proactively respond should a component technology problem ever arise. Using VPD information, IBM can determine what components should be selected for repair or replacement (based on manufacturing lot number, date of manufacture, or other specialized criteria), and quickly plan fixes for the specific subset of servers affected.

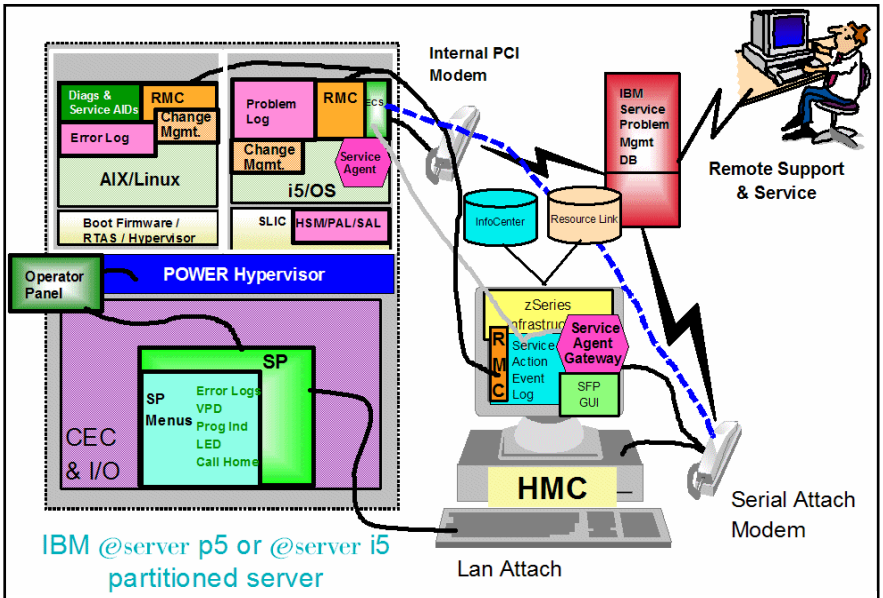
HMC-based Service Elements

In the pSeries family of servers, partitioning introduced new challenges and complexities to the UNIX server arena. The addition of Micro-Partitioning on the IBM @server p5 and @server i5 systems brought not only increased RAS capabilities in the hardware and platform firmware, but also new levels of service complexity and function. Each partition is treated as an independent operating environment. Common platform resources, such as a System Processor, can affect multiple partitions in the rare occurrence of a failure. Even failures in non-critical system resources (e.g., an outage in an N+1 power supply) require warnings to be presented to every operating system partition for appropriate notification and error handling.

The IBM @server p5 and @server i5 systems expand upon the leading-edge service capability of the pSeries servers, combining the Service Focal Point concept with a zSeries mainframe service infrastructure. The improved capabilities inherent in this combination facilitate new levels of service supporting capabilities such as automated maintenance and autonomic service on-demand ─ using excess Capacity Upgrade on Demand resources for service.

Hardware Management Console

All multi-partitioned IBM @server p5 and @server i5 systems require a Hardware Management Console. As has been seen, the HMC is an independent workstation used by system administrators to setup, manage, configure, and boot IBM @server p5 or @server i5 servers. The HMC for POWER5 processor-based servers includes improved performance, enabling system administrators to define and manage Micro-Partitioning capabilities and virtual I/O features; advanced connectivity; and sophisticated firmware performing a wide variety of systems management and service functions.



Logical Partitioning introduces a wide variety of new system management, consolidation, and virtualization options… it also generates new challenges and complexities for the UNIX server arena. The addition of Micro-Partitioning brings not only increased RAS capabilities in the hardware and platform firmware, but also new levels of service complexity and function. Even failures in non-critical system resources (e.g., an outage in an N+1 power supply) require warnings to be presented to every operating system partition for appropriate notification and error handling. The Hardware Management Console supports a wide variety of sophisticated diagnosis and service tools to simplify the service process and improve overall system availability. Because of the diversity of RAS tools available, some clients will choose to deploy HMC management even on servers that do not mandate its' use.

One significant improvement on the IBM @server p5 and @server i5 system HMC is to replace the "serial attachment" method used on predecessor consoles, with a LAN interface allowing high bandwidth connections to servers. Administrators can choose to establish a private service network, connecting all of their POWER5 processor-based servers and management consoles. Or they can include their service connections in their standard operations network. The Ethernet LAN interface also allows the HMC to be placed physically farther away from managed servers, though for service purposes it is still desirable to install the HMC in close proximity to the systems it manages.

The HMC comes with an install wizard to assist with installation and configuration of the HMC itself. This wizard helps to reduce user errors by guiding administrators through the configuration steps required to successfully install the HMC operating environment.

Service Focal Point

The service application has taken on an expanded role in the Hardware Management Console.

- The zSeries service framework has been incorporated, providing expanded basic service functions.
- The Service Focal Point graphical user interface has been enhanced to support a common service interface common across all HMC managed @server p5 systems.

A key service requirement in a partitioned system implementation is to insure that no error is lost before being reported for service and an error should only be reported once, regardless of how many partitions view (experience the potential effect of) the error.

Logically partitioned servers can encounter two types of errors: global and local. Local errors are faults in resources that are owned by only one partition. Examples include PCI adapters or devices assigned to a single partition. If a failure occurs to one of these resources, only a single operating system partition need be informed. Once notified, the OS diagnostic subsystem uses the Remote Management and Control Subsystem (RMC) to relay error information to the Service Focal Point application running on the HMC.

Global errors can occur in resources that may affect multiple partitions. Examples are power supplies, fans and blowers, processors, memory, and storage controllers. When a failure occurs in these components, the POWER Hypervisor notifies each partition to execute any required precautionary actions or recovery methods. In turn, each operating system environment will forward the error event through the RMC network to the SFP application. The Service Processor will also forward error notification of these events to the HMC, providing a redundant error reporting path in case of errors in the RMC network.

The Service Focal Point application logs the first occurrence of each failure type, filters, and keeps a history of repeat reports from other partitions or the Service Processor. The SFP, looking across all active service event requests, analyzes the failure to ascertain the root cause and, if enabled, initiates a call home for service. This methodology insures that all platform errors will be reported through at least one functional path either in-band through the operating system or out-of-band through the Service Processor interface to the SFP application on the HMC.

The Service Focal Point application is also the starting point for all service actions on HMC attached systems. The servicer begins the repair with the SFP application, selecting the “Repair Serviceable Events” view from the SFP Graphical User Interface (GUI). From here, the servicer selects a specific fault for repair from a list of open service events; initiating automated maintenance procedures specially designed for the IBM @server p5 and @server i5 systems. Those components that are concurrently maintainable are supported by the new automated processes.

Automating various service procedural tasks, instead of relying on servicer training, can help remove or significantly reduce the likelihood of servicer induced errors. Many service tasks can be automated. For example, the HMC can guide the servicer to:

- Interpret error information.
- Prepare components for removal or initiate them after install.
- Set and reset system identify LEDs as part of the guiding light service approach.
- Automatically link to the next step in the service procedure based on input received from the current step.
- Update the service history log, indicating the service actions taken as part of the repair procedure. The history log helps to retain an accurate view of the service scenarios in case future actions are needed.

zSeries Infrastructure

Borrowing heavily from lessons learned in mainframe service scenarios, the IBM @server p5 and @server i5 systems have added a number of significant new functions to the SFP application by incorporating a zSeries service framework. New or enhanced functions include:

- Root cause error analysis based on dissimilar reported events
- Automated maintenance
- Special support allowing IBM Product or Development Engineers to actively guide servicers while they perform preventative repairs
- Service History log automation

“Health Check” Scheduled Operations

Correct operation of call home service function is verified as an integral part of the overall server system installation process. During this process, the system generates a pseudo service event through the HMC interface to the IBM Service and Support organization, verifying the error reporting path and the call home capability. Once established, this path is checked on a periodic basis to insure that it is available and functioning should it be needed to transmit a service event. If the server misses too many periodic health check notifications, the IBM Service and Support organization will be notified. IBM personnel will contact

the client to substantiate that the system is still functional and is properly enabled for remote error reporting.

This service link is also used to report system configuration changes when features are added or removed from the server. These changes are stored in an IBM managed back-end database repository, available to the service organization should the need arise to carry out a preventative service action. This information is also helpful when Miscellaneous Equipment Specification (MES) orders for additional hardware are received; allowing the support person to decide if additional drawers or racks are required or if adequate reserve is available in the existing system configuration.

Remote Management and Control

The Remote Management and Control application is delivered as part of the base operating system. RMC provides a secure transport mechanism across the LAN interface between the operating system and the HMC and is used to by the operating system diagnostic application for transmitting error information. It performs a number of other functions as well, but these are not used for the service infrastructure.

HMC Enhanced Service Capabilities

The Hardware Management Console provides a number of RAS features to the servers it manages. While some IBM @server p5 or @server i5 server configurations do not mandate an HMC, any of these servers can optionally be attached to an HMC should these improved RAS capabilities be desired.

Automated Install/Maintenance/Upgrade

As previously discussed, the HMC provides a variety of automated maintenance procedures to assist in problem determination and repair. The Hardware Management Console extends this innovative technology, providing automated install and automated upgrade assistance. These procedures are expected to reduce or help eliminate servicer induced failures during the install or upgrade processes.

Concurrent Maintenance and Upgrade

All IBM POWER5 processor-based servers provide at least the same level of concurrent maintenance capability as was available in their predecessor pSeries (POWER4) servers. Components such as power supplies, fans, blowers, disks, HMCs, PCI adapters and devices can be repaired concurrently (“hot” service and replace).

The HMC, however, also supports many new concurrent maintenance functions in IBM @server p5 and @server i5 servers. Future offerings, like dynamic firmware update (1H05) will also be available to HMC attached systems.

Dynamic Firmware Maintenance or Update

Firmware on the IBM @server p5 and @server i5 systems is planned[1] to be released (1H05) in a cumulative sequential fix format packaged in RPM formats for concurrent application and activation. The objective is that the majority of firmware updates will be able to be installed and activated without having to cycle power or reboot the system. This is accomplished by loading the new firmware image on the HMC from any of the following methods:

1. IBM distributed media (such as CD-ROM)
2. A Problem Fix distribution from the IBM Service and Support repository
3. Download from the IBM Web site (http://techsupport.services.ibm.com/server/mdownload)
4. FTP from another server

IBM will support multiple firmware releases in the field so a server can run on an existing firmware release, using concurrent firmware updates to stay up-to-date with the current patch level. Under normal operating conditions, clients should plan for one disruptive upgrade per 12 month period. It is expected that this update will be needed to stay on a supported firmware release. In addition to concurrent and disruptive firmware updates, IBM will also offer delayed updates. Delayed patches require a system reboot to be activated, but the server will operate normally and can continue to install concurrent fixes until a sys-

tem reboot can be scheduled. If a delayed fix does not apply to the client's platform or configuration, there is no need to schedule a system reboot.

Once the firmware image designated as concurrently installable is loaded on the HMC, the Concurrent Microcode Management application on the HMC is used to flash the system and instantiate the new code without the need for a power cycle or system reboot. A backup copy of the current firmware image is maintained in Flash memory and can be used if necessary. Once normal system operation on the upgraded firmware is validated, the system administrator may replace the backup version with the new code image.

Service Summary

The IBM RAS Engineering team has planned, and is delivering, a roadmap of continuous service enhancements in IBM server offerings. The service plan embraces a strategy that shares "best of breed" service capabilities developed in IBM @server product families such as the xSeries and zSeries servers, and adds groundbreaking service improvements described in this document, specifically tailored to the IBM @server p5 and @server i5 product lines. The Service Team worked directly with the server design and packaging engineering teams, insuring that their designs supported efficient problem determination and service. This close coordination of the design and service teams has led to system service capabilities unique for the UNIX and Linux systems. Offerings such as automated install, upgrade, and maintenance improve the efficiency of our skilled IBM SSRs. These same methods are also modified and linked to client capabilities, allowing users to effectively perform diagnosis and repair services on many of our entry and mid-range system offerings. These can include:

- Increased client control of their systems
- Reduced repair time
- Minimized system operational impact
- Higher availability
- Increased value of their servers to clients and better tracking, control, and management by IBM

IBM @server p5: A Highly Available Design for Business-Critical Applications

The IBM @server p5 and @server i5 product families are engineered for reliability, availability, and serviceability using an architecture-based strategy designed to avoid unplanned outages. These servers include a wide variety of features to automatically analyze, identify, and isolate failing components so that repairs can be made as quickly and efficiently as possible

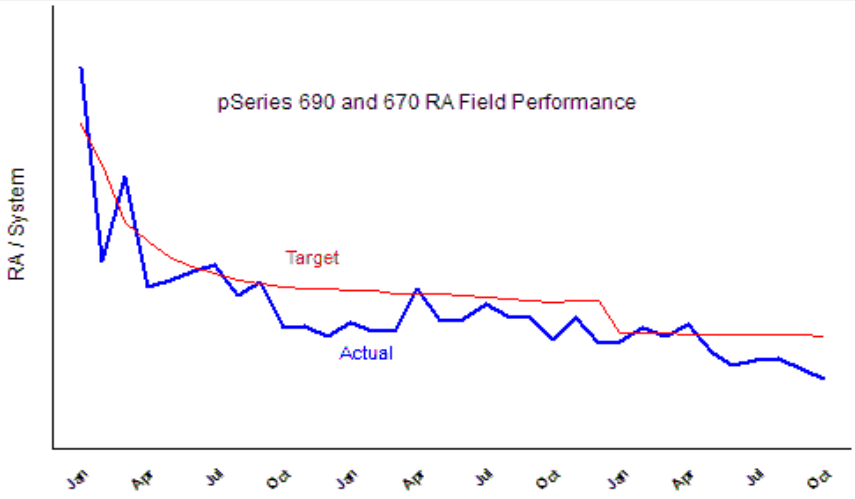
System design engineers incorporated state-of-the-art components and advanced packaging techniques, selecting parts with low intrinsic failure parts rates, and surrounding them with a server package that supports their reliable operation. Care has been taken to deliver rugged and reliable interconnects, and to include features that ease service; like card guides, PCI adapter carriers, cable straps, and "positive retention" connectors. This analytical approach identifies "high opportunity" components: those whose loss would have a significant effect on system availability. These receive special attention and may be duplicated (for redundancy), may be higher grade, or may include special design features to compensate for projected failure modes (or, of course, may receive all three improvements).

Should a hardware problem actually occur, these servers have been designed to be fault resilient, to continue to operate despite the error. Every server in the @server p5 product family includes advanced availability features like Dynamic Processor Deallocation and Dynamic Processor Sparing, PCI bus error recovery, Chipkill memory, memory bit-steering, L3 cache line delete, dynamic firmware update, redundant hot-plug cooling fans, and hot-plug N+1 power, power regulators, and power cords (optional in some configurations).

Many of these functions rely on IBM First Failure Data Capture technology, which allows the server to efficiently, capture, diagnose, and respond to hardware errors ─ the first time that they occur. Based on experience with servers implemented without the run time first failure diagnostic capability (using an older "recreate" strategy), it is possible to project that high impact outages would occur 2 to 3 times more fre-

quently without this capability. FFDC also provides the core infrastructure supporting predictive failure analysis techniques, allowing parts to automatically be deallocated from a server before they ever reach a failure that could cause a server outage. The IBM design objective for FFDC is to correctly identify a hardware failure to a single part in 96% of the cases, and to several parts the remainder of the time.

These availability techniques are backed up by service capabilities unique in the UNIX and Linux systems. Offerings such as automated install, upgrade, and maintenance can be employed by IBM SSRs or IBM clients (for selected models), allowing servicers from either organization to effectively diagnose and repair faults on these systems.



IBM Field Engineering teams track and record repairs of system components covered under warranty or maintenance agreement. Failure rate information is gathered and analyzed for each part by IBM commodity managers, who track replacement rates. Should a component not be achieving its' reliability targets, the commodity manager will create an action plan and take appropriate corrective measures to remedy the situation. Failure tracking data is also reviewed on a regular basis by IBM executives responsible for the design and reliability of the server family.

The IBM field support team continually analyzes critical system faults, testing to determine if system firmware and maintenance procedures and tools are effectively handling and recording faults.

The POWER4-base pSeries family of servers has established an exceptional record of reliability and availability. The biggest servers in the family, as shown here, are performing better in the field than their rigorous design targets. In fact, data gathered by the field engineering team[2], measuring installed server systems, indicates that these high-end servers are achieving a Mean Time Between (MTB) unscheduled hardware outages (MTBUIRA) of more than 21 years and an astounding MTBHIO of more than 130 years.

The IBM POWER engineering team is justifiably proud of the record of service for these systems.

The POWER4 processor-based pSeries offerings have demonstrated a superb record of reliability and availability in the field. As has been demonstrated in this white paper, the IBM @server p5 servers build upon this solid base, making RAS improvements in all major server areas: the CEC, the memory hierarchy, and the I/O subsystem.

The new POWER Hypervisor not only provides fine grained allocation of system resources supporting new levels of virtualization for UNIX and Linux servers, it also delivers many availability improvements. The POWER Hypervisor enables: resource sparing, automatic redistribution of capacity on N-1, redundant I/O across LPAR configurations, the ability to reconfigure a system "on the fly," automated scale-up of high availability backup servers, serialized sharing of devices , and sharing of I/O devices through I/O server partitions.

The Hardware Management Console supports the IBM virtualization strategy and includes a wealth of improvements for service and support including automated install and upgrade, and concurrent maintenance and upgrade for hardware and firmware. The HMC also provides a focal point for service receiving, logging, tracking system errors and, if enabled, forwarding problem reports to IBM Service and Support organizations. While the HMC is an optional offering for some configurations, it may be used to support any server in the IBM @server p5 product family.

Borrowing heavily from predecessor system designs in both the iSeries and pSeries, adding popular client set up and maintenance features from the xSeries, and incorporating many advanced techniques pioneered in IBM mainframes, the IBM @server p5 and @server i5 servers are designed to deliver leading-edge reliability, availability, and serviceability.

Appendix A: Operating System Support for Selected RAS Features

RAS Feature	AIX 5L V5.2	AIX 5L V5.3	i5/OS	RHEL AS 3	SLES 9
System Deallocation of Failing Components					
Dynamic Processor Deallocation	X	X	X		X
Dynamic Processor Sparing					
• Using CUoD processors	X	X	X	X	X
• Using capacity from spare pool	X	X	X	X	X
Persistent processor deallocation	X	X	X	X	X
GX+ bus persistent deallocation	X	X	X		
PCI bus extended error detection	X	X	X	X	X
PCI bus extended error recovery	X	X	X		Limited
PCI-PCI bridge extended error handling	X	X	X		
Redundant RIO link	X	X	X	X	X
PCI card hot-swap	X	X	X		X
Dynamic SP failover at run-time	1H05	1H05	1H05		
Memory sparing with CUoD at IPL time	X	X	X	X	X
Clock failover at IPL	X	X	X	X	X
Memory Availability					
ECC Memory, L2, L3 cache	X	X	X	X	X
Dynamic bit-steering (spare memory in main store)	X	X	X	X	X
Memory scrubbing	X	X	X	X	X
Chipkill memory	X	X	X	X	X
L1 parity check plus retry	X	X	X	X	X
Improved L3 cache line delete	X	X	X	X	X
Array Recovery and Array Persistent Deallocation ─ (spare bits in L1 and L2 cache; L1, L2, and L3 directory)	X	X	X	X	X
Special uncorrectable error handling	X	X	X	X	X
Fault Detection and Isolation					
Platform FFDC diagnostics	X	X	X	X	X
I/O FFDC diagnostics	X	X	X		X
Run-time diagnostics	X	X	X	Limited	Limited
Error log analysis	X	X	X	X	X
Service Processor support for:					
• Built-in-Self-Tests (BIST) for logic and arrays	X	X	X	X	X
• Wire tests	X	X	X	X	X
• Component initialization	X	X	X	X	X
Serviceability					
Boot-time progress indicators	X	X	X	Limited	Limited
Firmware error codes	X	X	X	X	X
Operating system error codes	X	X	X	Limited	Limited
Inventory collection	X	X	X	X	X
Environmental and power warnings	X	X	X	X	X
Hot-plug fans, power supplies, power regulators	X	X	X	X	X
Extended error data collection	X	X	X	X	X
SP “call home” on non-HMC configurations	X	X	X	X	X
I/O Drawer redundant connections	X	X	X	X	X
SP mutual surveillance w/ POWER Hypervisor	X	X	X	X	X
Dynamic firmware update with HMC	1H05	1H05	1H05	1H05	1H05
Service Agent	X	X	X		X
Guiding light LEDs	X	X	X	X	X
System dump for memory, POWER Hypervisor, SP	X	X	X	X	X
InfoCenter service publications	X	X	X	X	X
Resource Link education	X	X	X	X	X
Operating system error reporting to HMC SFP application	X	X	X	X	X
RMC secure error transmission subsystem	X	X	X	X	X
Health check schedule operations with HMC	X	X		X	X
Operator panel	X	X	X	X	X
Redundant HMCs	X	X	X	X	X
Automated server recovery/restart	X	X	X	X	X
High availability clustering support	X	X	X		

About the authors:

Jim Mitchell is an IBM Senior Engineer. He has worked in microprocessor design and has managed an operating system development team. An IBM patent holder, Jim has published numerous articles on floating-point processor design, system simulation and modeling, and server system architectures. Jim is currently assigned to the staff of the Austin Executive Briefing Center.

Daniel Henderson is an IBM Senior Technical Staff Member. He has been a part of the design team in Austin since the earliest days of RISC based products and is currently the lead availability system designer for IBM @server p5 and @server i5 systems.

George Ahrens is an IBM Senior Engineer. He has been responsible for the Service Strategy and Architecture of the POWER4 and POWER5 processor-based systems. He has published multiple articles on RAS modeling as well as several whitepapers on RAS design and Availability Best Practices. He holds numerous patents dealing with RAS capabilities and design on partitioned servers. George currently leads a group of Service Architects responsible for defining the service strategy and architecture for IBM Systems and Technology Group products.

.

Information concerning non-IBM products was obtained from the suppliers of these products. Questions on the capabilities of the non-IBM products should be addressed with the suppliers.

Many of the IBM @server p5 features described in this document are operating system–dependent and may not be available on Linux. For more information, please visit ibm.com/servers/eserver/pseries/linux/whitepapers/linux_pseries.html.

All performance information was determined in a controlled environment. Actual results may vary. Performance information is provided "AS IS" and no warranties or guarantees are expressed or implied by IBM.

[1] Dynamic firmware update capability is planned to be available in 1H05.

[2] This data represents IBM field measurements of unscheduled hardware/microcode outages for the p670 and the p690 for all IBM installed p670 and p690 systems from December 2001 to October 2004. Loss of system availability due to other causes such as software or scheduled outages is not included. The actual availability that any user will experience will vary. No assurance can be given that a user will achieve system availability equivalent to the numbers stated here. This statement is subject to change or withdrawal without notice.

Information concerning non-IBM products was obtained from the suppliers of these products or other public sources. Questions on the capabilities of the non-IBM products should be addressed with the suppliers.

Photographs show engineering and design models. Changes may be incorporated in production models.

IBM Corporation
Systems and Technology Group
Route 100
Somers, New York 10589

Produced in the United States of America
December 2004

RHEL AS 3 = Red Hat Enterprise Linux AS 3 for POWER, Update 43.
More information is available at: http://www.redhat.com/software/rhel/as/.

SLES 9 = SUSE LINUX Enterprise Server 9 for POWER. More information is available at: http://www.novell.com/products/linuxenterpriseserver/.

The IBM home page on the Internet can be found at: http://www.ibm.com.

The IBM @server p5 page can be found at http://www.ibm.com/eserver/pseries.

The AIX 5L home page on the Internet can be found at: http://www.ibm.com/servers/aix.