Servaris ProServ 4680 Server Specification

Introduction

The main board provides most of the basic functions for the system. Nearly all of the boards from the boardset plug into or cable to the main board.

Main Board Block Diagram

The figure below provides a high-level diagram of the main board and an overview of how the main board fits together with the rest of the system.

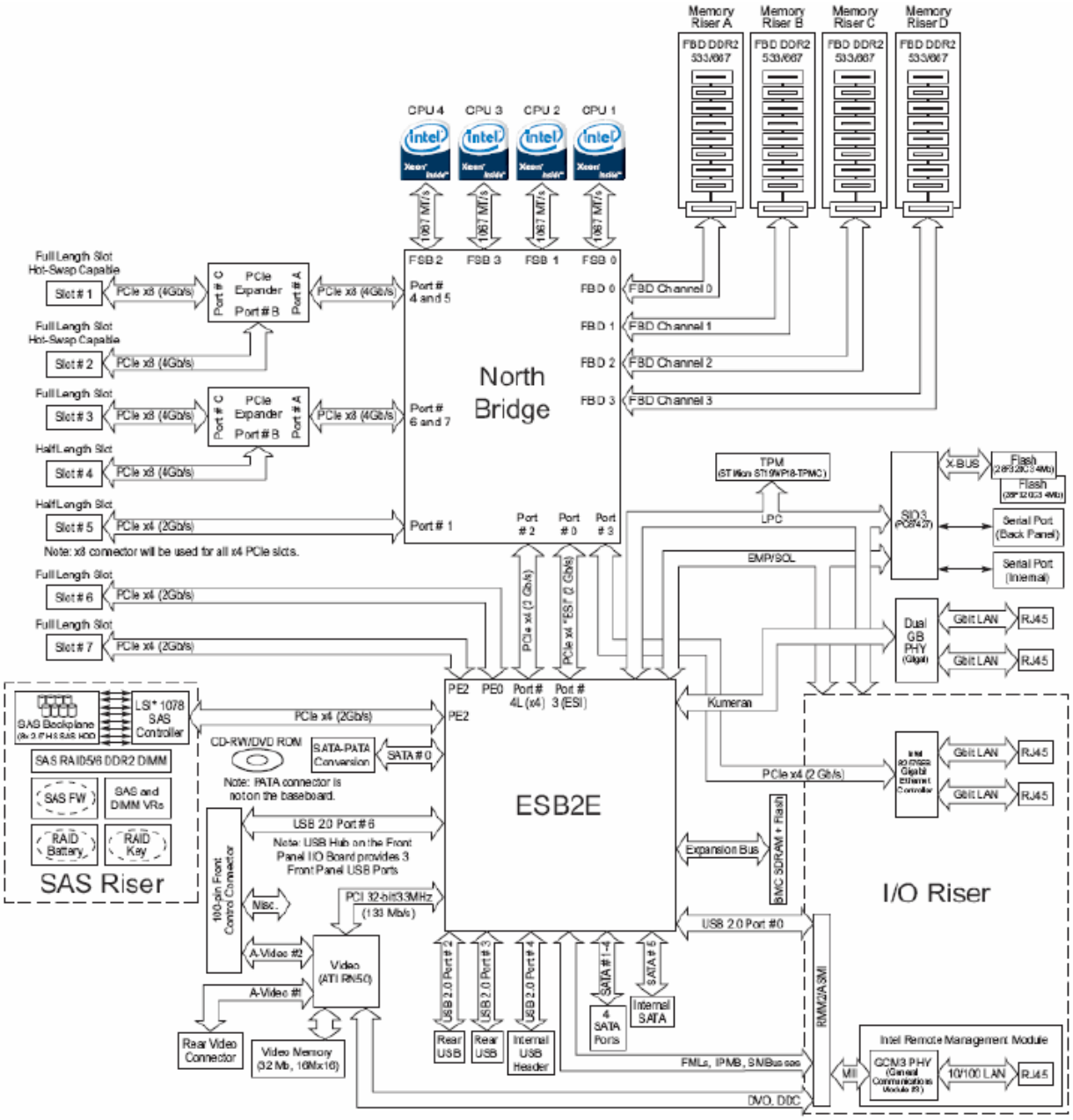


Figure 1. Main Board Block Diagram

The main board contains:
Chipset northbridge and southbridge components
CPU sockets
Video components
Trusted Platform Module
BIOS Flash components
Super I/O*
Seven PCI-Express* slots
Back-panel I/O connectors
Four memory riser connectors
The main board also contains many voltage regulators used by its components, as well as many of the primary rails used by the rest of the board set. The following chapters describe the main board in detail.

Functional Architecture
This section describes the primary functions, blocks, and components that reside on the main board. The section is laid out as follows:
Chipset components
Primary interfaces to the memory and I/O riser
Addition information regarding other functional blocks on the board

Intel® Xeon® Processors
The main board supports the Quad-Core Intel® Xeon® Processors 7300 Series or Dual-Core Intel Xeon Processors 7200 Series, which are based on the low-power next-generation Intel®Core™ micro-architecture. Several architectural and micro architectural enhancements have been added to this processor, including multiple processor cores.
The 64-bit Intel® Xeon® Processor MP includes the following advanced features:

- Intel® Extended Memory 64 Technology (Intel® EM64T) for executing both 32-bit and 64-bit applications simultaneously
- Next-generation Intel® Core™ microarchitecture
- Support for Enhanced Intel SpeedStep® Technology
- Execute-disable bit for hardware support of security features
- Intel® Virtualization Technology
- Enhanced power and thermal management
- Thermal Monitor 1 (TM1): Processor thermal monitoring and throttling.
- Thermal Monitor 2 (TM2) enhanced thermal management: Hardware controlled methods to reduce power consumption.

On-demand Mode: Software controllable method to reduce power consumption

- Platform Environment Control Interface (PECI)
- Streaming Single Instruction, Multiple Data (SIMD) Extensions 2 and 3 (SSE2, SSE3)
- 2.2.1.1 Processor Heatsink
- The main board uses the Common Enabling Kit (CEK) heatsink solution. The CEK design
- meets the 64-bit Intel® Xeon® Processors MP thermal performance targets. Each CEK heatsink

- consists of the following components:
- Passive heatsink (with captive standoff and screws)
- Thermal Interface Material (TIM-2) – to cover the entire processor Integrated Heat
- Spreader (I) and the heatsink base
- Hat spring/backplate – mounted on the backside of the main board
- 2.2.1.2 Processor Installation Order
- The four-processor sockets are on independent front side buses. Therefore, there are no installation order requirements enforced by the platform. The user is free to install supported processors in any configuration desired.

However, one logical order is as follows:
1. Populate the lower number processor sockets.
2. Move to the higher numbers.
3. Populate sockets in the following order: socket 1, socket 2, socket 3, and socket 4).

The board does not support mixing different processor models.

Intel® 7300 Chipset Memory Controller Hub

The Intel® 7300 Chipset Memory Controller Hub is the highest performance, most scalable chipset offering in the 64-bit Intel® Xeon® Processor MP family. The chipset represents an improvement to Intel's four-way multi-processor platform.

Intel® 7300 Chipset Features

The Intel® 7300 Chipset Memory Controller Hub is the center of the ProServ 4680 Server System architecture. This chipset is designed for multi-core processors and includes the advanced features detailed in the following bullets.

- Up to four 64-bit Intel® Xeon® Processors MP via independent 1067 MT/S FSBs optimized for server applications.
- Maintains coherency across four independent FSBs.
- Double-pumped 40-bit address buses with a total address bandwidth of 133 million
- addresses per second.
- Quad-pumped, 64-bit data bus providing a bandwidth of 8.5 GB/s per bus (1067 MT/S
- data rate)
- Uses a twelve-entry in-order queue depth
- Supports deferred reply responses to deferred transactions
- Four FBD channels supporting fully buffered DDRII DIMMs
- Up to eight FBD fully buffered DIMMs (FBD) can be linked on a channel (for up to a maximum of 32 DIMMs across all four channels).
- The four channels are paired into two lock-step branches.
- Channel 0 and 1 form the first branch. (On the main board, these channels are connected to memory risers A and B.)
- Channel 2 and 3 form the second branch. (On the main board, these channels are connected to memory risers C and D.)
- 24 lane serial bus providing 6.4 GB/s (using DDR2 533MHz memory) or 8GB/s (using DDR2 667MHz memory) peak theoretical bandwidth per Channel.

- Support for ECC and FBD Memory Failover.
- Seven x4 PCI Express* ports (compliant with the PCI Express* 1.0a specification)
- Specific x4 ports can be combined with other ports to form x8 ports.
- The main board utilizes four x4 ports and two x8 ports on the MCH. Figure 1 shows the assignment and further details are explained in Chapter 2.2.4 PCI-Express* Subsystem.
- Intel® I/O Acceleration Technology (Intel® I/OAT) enabled x4 link to MCH
- Peer-to-Peer memory mapped I/O, I/O, and configuration read/write is supported across these parts.
- Hot plug/hot swap is supported on each port with MSI and legacy ACPI protocol.
- ESI (x4 PCI Express*) port with 2GB/s aggregate bandwidth to communicate with the Enterprise Southbridge 2

Enterprise Southbridge 2

The Enterprise Southbridge2 works in conjunction with the Intel® 7300 Chipset Memory Controller Hub and Intel® Xeon® Processors MP to provide the latest I/O and Legacy I/O capabilities in Enterprise Server Platforms. The Enterprise Southbridge 2 is an integration of ICH6, BMC and LAN MAC, and PEXH PCI Express-to-PCI Express Bridge) components.

Some of the features of the Enterprise Southbridge 2 are not used in this board set. Those features of Enterprise Southbridge 2 that are used in the main board are as follows:

- Integrated Serial ATA (SATA) controller
- PCI 32-bit / 33MHz interface
- Five USB 2.0 ports (1 Front Panel, 2 Rear Panel, 1 Internal, 1 to Intel® Remote
- Management Module 2)
- Low Pin Count (LPC) Interface
- Support for FML/SMBUS for Sever Management operations
- General Purpose I/Os
- Kumeran high speed serial interface to external LAN PHY (Intel® 82563EB Gigabit Ethernet PHY)
- Intel® I/OAT enabled x4 link to MCH
- ESI x4 Link to MCH
- Two downstream PCI-Express* x4 links
- Four downstream PCI-Express* x1 links configured as a one x4 link
- 24 bit expansion bus for BMC memory devices
- EMP serial interface for server management operations
- ACPI power management logic support

PCI-Express* Subsystem

The PCI-Express* subsystem provides end-user support for nine PCI-Express* plug-in devices in a 4U chassis. Seven of the plug-in devices are available for standard PCI-Express* adapters. One device is for a custom I/O Riser, and the other is for a SAS Riser. The design enables easy use and replacement of four PCI-Express* hot-plug devices without powering down the system.

Table 1. PCI-Express* Expansion Slot Features

Segment	Slot	Hot-plug	Technology	Width	Bandwidth (GB/s)
PCI-Express* Expander 1 (x8 Port[C])	1	Yes	PCI-Express*	x8	4
PCI-Express* Expander 1 (x8 Port[B])	2	Yes	PCI-Express*	x8	4
PCI-Express* Expander 2 (x8 Port[C])	3	No	PCI-Express*	x8	4
PCI-Express* Expander 2 (x8 Port[B])	4	No	PCI-Express*	x8	4
MCH (x4 Port[1])	5	No	PCI-Express*	x4	2
Enterprise Southbridge 2 (x4 Port[0])	6	No	PCI-Express*	x4	2
Enterprise Southbridge 2 (x4 Port[2])	7	No	PCI-Express*	x4	2
MCH (x4 Port[3])	I/O Riser	No	PCI-Express*	x4	2
Enterprise Southbridge 2 (x4 Port[1])	SAS Riser	No	PCI-Express*	x4	2

PCI-Express* Interrupts

The PCI Express* interrupt model supports two mechanisms:

- INTx Emulation
- Message Signaled Interrupt (MSI) Support

INTx Emulation

For legacy compatibility, PCI-Express* provides a PCI INTx emulation mechanism to signal interrupts to the system interrupt controller (Enterprise Southbridge 2). This mechanism is compatible with existing PCI software.

The mechanism provides the same level and type of service as standard PCI interrupt mechanisms. However, it uses a different hardware implementation where physical PCI interrupt signals are virtualized as in-band PCI-Express* messages. This legacy compatibility allows boot device support without requiring complex BIOS-level interrupt configuration/control service stacks.

Message Signaled Interrupt (MSI) Support

In addition to PCI INTx compatible interrupt emulation, PCI-Express* requires support of Message Signaled Interrupt (MSI) mechanism. The PCI-Express* MSI mechanism is compatible with the MSI capability defined in the PCI 2.2 Specification.

PCI-Express* Expander/Switch

The PCI-Express* Expander/Switch IDT* 89HPES24N3A works in conjunction with the Intel®7300 Chipset. It utilizes a single x8 PCI-Express* upstream bus and expands it into two unique downstream x8 PCI-Express* buses. Some features of the PCI-Express* Expander used on the main board are as follows:

- 24 PCI Express* lanes (2.5 Gbps), three switch ports
- 6 GBps (48 Gbps) aggregate switching throughput
- Low latency cut-through switch architecture
- Supports 128 to 2048 byte maximum payload size
- One virtual channel
- Fully compliant with PCI Express Base Specification Revision 1.0a
- ACPI power management logic support
- Port arbitration schemes utilizing round robin or weighted round robin algorithms

- One port configurable as downstream port or non-transparent port
- Automatic per port link width negotiation to a x8, x4, x2 or x1
- Static lane reversal on all ports
- Polarity inversion
- Ability to load device configuration from serial EEPROM
- Internal end-to-end parity protection on all TLPs ensures data integrity even in systems
- that do not implement end-to-end CRC (ECRC)
- ECRC passed through in transparent and non-transparent modes
- Supports PCI Express* native hot-plug
- Compatible with hot-plug I/O expanders used on PC motherboards
- Hot-swap capable I/O
- Utilizes advanced low-power design techniques to achieve low typical power consumption
- Support PCI Express* Power Management Interface Specification (PCI-PM 1.1)
- Unused SerDes are disabled.
- Supports Advanced Configuration and Power Interface Specification, Revision 2.0
- (ACPI) supporting-active link state

PCI-Express* Errors

Errors are classified into three types:

- Correctable
- Uncorrectable
- Fatal

These error types are discussed in more detail in the following sections.

Correctable Errors

Correctable errors are corrected by the hardware such as single bit ECC

Uncorrectable Errors

Uncorrectable errors are not corrected by hardware or software. Operating system or other software layers may be able to recover. However, this may not always be the case. These errors leave the system in a functional state. An example is a multi-bit data error in an application. The application may crash but the system can continue.

Fatal Errors

Fatal errors may compromise the system integrity. An example of a fatal error is a protocol error. Errors detected by the PCI Expanders are routed to the MCH via in-band messages. The MCH can assert MCERR# or ERR [2:0] (if configured to do so) upon receiving an error message or detecting an error, itself.

PCI-Express* Hot-plug Support

PCI-Express* hot-plug is the concept of removing a standard PCI-Express* adapter card from a system without stopping the software or powering down the system as a whole.

The main board supports four hot plug capable PCI-Express* slots off the two PCI-Express * Expander components (IDT* 89HPES24N3A). The following slots are hot pluggable

- Slot 1
- Slot 2

The hot-plug implementation conforms to the PCI-Express* Base Specification, Revision 1.1.

PCI-Express* Hot-plug Power Controller

The main board uses the Texas Instruments* TPS2363 hot-plug controller. The Texas Instruments* TPS2363 is a dual slot PCI-Express* hot plug controller providing support for 12V, 3.3V and 3.3Vaux power control. This support includes the following:

- Current limiting
- Voltage supervision
- Fault indication
- Independent slot control

The main board contains buttons and LEDs to assist a user with hot-plug operations. The buttons can be used to initiate hot-plug events, while the LEDs provide slot power and hot-plug status. The LEDs have enough luminous intensity to pass through system-level light pipes and be visible at the top of a system. An attention button can be used to invoke a hot-plug sequence to remove or add an adapter without the use of an operating system/software interface.

Table 2. PCI Hot-plug LEDs

LED	State	Meaning
Power (green)	Off	Power off: All main rails have been removed from slot. Card can be inserted or removed.
	On	Power on: Slot is powered on. Card cannot be inserted or removed.
	Blinking	Power transition: Slot is in the process of changing state. Card cannot be inserted or removed.
Attention (amber)	Off	Normal: Normal operation.
	On	Attention: Power fault or operational problem at this slot.
	Blinking	Locate: Slot is being identified at the user's request.

PCI 32-Bit Subsystem

PCI Interrupts

Legacy PCI devices can deliver interrupts either by asserting external IRQ signals that are routed to the Enterprise Southbridge 2 or MSI via in-band messaging. In either case, the Enterprise Southbridge 2 forwards the interrupt to the NB as an inbound write for the processor to handle the event.

PCI INTx signals are mapped for the purpose of device interrupt load balancing. The specific requirement is to ensure the following:

- Interrupt controller receives INTx messages the represent the wire-ORed behavior and interrupt routing of legacy PCI implementations
- Software can determine at which interrupt controller input an interrupt is routed.

Table 3 describes how the interrupts for each of the PCI devices are mapped to the Enterprise Southbridge 2.

Table 3. PCI Interrupt Mapping

Device	Enterprise Southbridge 2 PCI Host Bridge			
	INTA#	INTB#	INTC#	INTD#
RN50 Video Controller		INTA# [ESB_PIRQB_N]		

PCI IDSEL Signal

The IDSEL signal is used as a chip-select for PCI32 devices during read/write transactions. The Enterprise Southbridge 2 PCI32 controller asserts a specific address bit on a given PCI bus to toggle the IDSEL signal to the PCI device. For the main board, the address bit to IDSEL mapping is shown in Table 4.

Table 4. IDSEL Mapping

Device	Device Number	IDSEL	Host Bridge
RN50 Video Controller	12	AD28	Enterprise Southbridge 2

PCI Bus Arbitration Signals

Request (REQ#) signals indicate to the bus arbiter that an agent/device desires the use of the bus. The Grant (GNT#) signal indicates to the agent/device that access to the bus has been granted. Every master has its own REQ#, which must be tri-stated while RST# is asserted. These are point-to-point signals, which are assigned to every bus master.

Table 5. Arbitration Connections

Device	REQ#	GNT#	Host Bridge
RN50 Video Controller	0	0	Enterprise Southbridge 2

Main board Memory Interface

The main board includes four 164 pin x16 PCI-Express* connectors that interface with up to four Memory Risers. Each of these Memory Riser connectors are individually connected to one of the four

MCH’s FBD channels.

Serial Presence Detect (SPD) side-band signals are also passed between the Memory Risers and the Intel® 7300 Chipset MCH.

The main board supports the following memory riser population configurations:

- Memory riser installed in slot A, with upto 8 DIMMs slot populated (uses the MCH’s single DIMM/single channel mode).
- Memory risers installed in both slot A and B together.
- Memory risers installed in all four-riser slots.

Other riser configurations are not supported because they will cause DIMM population violations and malfunctions in memory riser DIMM fault LED operation.

Main Board I/O Riser Interface
The main board includes a 280 pin PCI-Express* super-slot custom connector to interface with the I/O riser card. To communicate with the advanced firmware control (Intel® Remote Management Module 2) the I/O riser connector will have the following:

- Two FML buses
- One USB port
- A video DVO interface
- A LPC Bus
- A RS232 BMC Serial Bus

It also has the PCI-Express x4 Link and Gigabit Lan Link for the Intel® 82575EB Gigabit Ethernet Controller.

Main Board SAS Riser Interface

The main board includes a 98 pin x8 pin PCI-Express* connector to interface with the x4 PCIExpress* SAS Riser card. This PCI-Express* slot is meant for the SAS Riser and is not to be used with any other type of PCI-Express* Standard Adapter Card

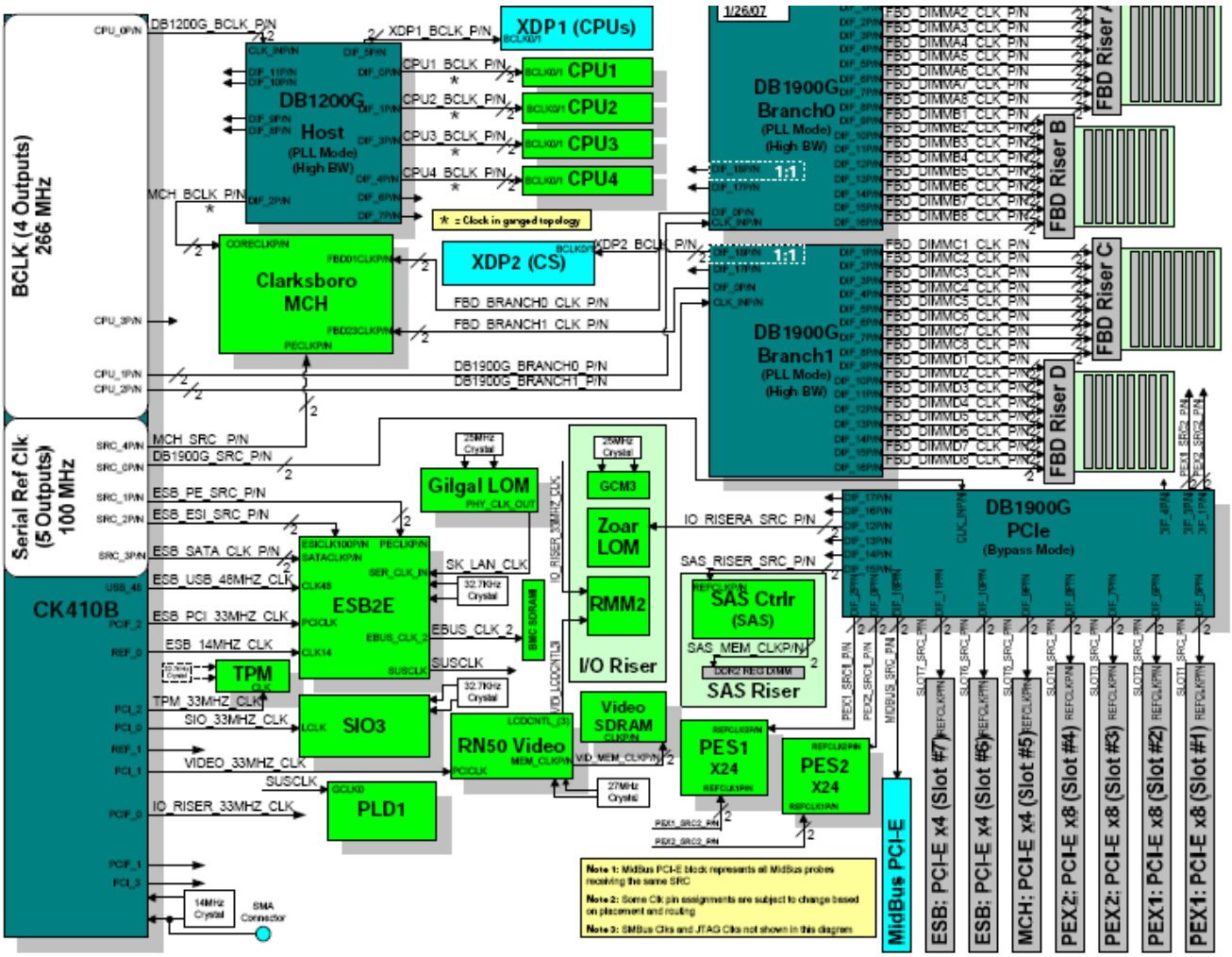
Clock Subsystem

This section describes the Clock Architecture/Sub-system for the main board.

Clock Overview

The main board clock tree is generated from a single CK410B with spread spectrum capability. The CK410B generates multiple copies of differential pair high-speed clocks (266MHz BCLK). DB1200G* (High BW / PLL mode) buffer generates additional BCLK copies for the CPUs, XDP1, and MCH core. The CK410B drives BCLKs to the two DB1900Bs (High BW / PLL mode) for FBD clocking. Each FBD branch clock input is fed by a DB1900G* buffer. 16 DIMMs are driven by each buffer (8 on each of two risers). The CK410B also generates 100MHz SRC clocks including an input to a DB1900G* buffer to I/O subsystems.

The figure below is the main board Clock Block Diagram.



CK410B (Clock Generator/Synthesizer)

The CK410B clock generator/synthesizer is the main clock source for most of the devices on the main board. Although it supports a maximum host clock frequency of 400 MHz, the main board will only support at most 266 MHz for its supported processors.

The CK410B supports SSC (Spread Spectrum Clocking), but only for FSB host clocks and SRCs (Serial Reference Clocks). The CK410B supplies the following:

- Four host clocks
- Five 100 MHz differential SRCs (Serial Reference Clocks)

- One 48 MHz USB clock
- Seven 33 MHz clocks
- Two 14 MHz clocks

On the main board, the CK410B is configured as follows:

- (1) BCLK differential pair to DB1200G* differential clock buffer driving Chipset
- components at 1:1 ratio. BCLK output frequencies can be set with stuffing options.

(Default = 266MHz)

- (2) BCLK differential pairs to two DB1900G* differential clock buffer parts (FBD memory
- risers Branch 1 and 2, and Chipset XDP).
- (5) SRC PCI-Express* clocks (100MHz) to MCH, DB1900G* differential clock buffer
- (PCI-Express*), and Enterprise Southbridge 2 (2 PCI-Express* and 1 SATA)
- (1) 48MHz clock for Enterprise Southbridge 2 USB controller
- (7 w/1 shared) 33MHz clocks to Enterprise Southbridge 2, TPM, SIO, Video, PLD, FWH
- (OEM), and I/O Riser. In order to support eight devices with seven clocks, one of the clocks is double loaded.
- (2) 14MHz clock to Enterprise Southbridge 2, and SIO (OEM).

DB1200G* (CPU Clock Buffer)

The DB1200G* is a differential clock buffer supporting the CPU, chipset, and CPU XDP clocks for the main board. It receives its input source from one of the CK410B processor host clocks. The DB1200G* provides 12 differential outputs with gear ratio capability.
The main board only uses six outputs at a gear ratio of 1:1. DB1200G* part was selected for its ability to take clock input w/ freq >200MHz. On the main board, the DB1200G* is configured as follows:

- PLL Mode / High BW
- Four host clock differential pairs to four CPUs
- One host clock differential pair to XDP1 (CPUs)
- One host clock differential pair to MCH (MCH)

Differential routing for the outputs for the CPUs are matched to within 5 mils from clock to clock. Routing for MCH clock is matched to the other CPUs lengths plus +0.7 inch. XDP clock routing has no clock-to-clock length matching requirements.

On the main board, breakout for differential pairs for the CPUs and the MCH are ganged together to reduce output-to-output skew.

DB1900G* (Memory and PCI-Express* Clock Buffers)

The DB1900G* is a differential clock buffer supporting the memory, XDP, and PCI-Express* clocks for the main board. There are two DB1900Gs* for memory and XDP, and one DB1900G* for PCI-Express* devices. The memory DB1900Gs* receive their input sources from the CK410B processor host clocks.

The memory DB1900G* provides 19 differential outputs. Eighteen outputs have gear ratio configuration capability to support processor to memory speed flexibility (see table below), and one output is set at 1:1 gear ratio for XDP. The PCI-Express* DB1900G* receives its input source from the

CK410B SRC (100 Mhz) output and does not need to use the DB1900Gs* gear ratio capability. On the main board, the three DB1900Gs* are configured as detailed in the following sections,

FBD Branch 0 DB1900G*

- PLL Mode / High BW
- BCLK input from CK410B
- (16) FBD clock differential pairs to support FBD memory risers 1 and 2 (eight DIMM slots each). Gear ratio set according to description below.
- (1) FBD clock differential pair to MCH. Gear ratio set according to description below.

FBD Branch 1 DB1900G*

- PLL Mode / High BW
- BCLK input from CK410G
- (16) FBD clock differential pairs to support FBD Memory Risers 3 and 4 (eight DIMM slots each). Gear ratio set according to description below.
- (1) FBD clock differential pair to MCH (Gear ratio set according to description below)
- (1) Host clock differential pair to XDP2 (Chipset) connector

PCI-Express* DB1900G*

- Bypass Mode (1:1 ratio) / BW N/A
- SRC input from CK410B
- Seven SRC differential pairs to seven PCI-Express* slots
- Five SRC differential pairs to I/O Riser (only one used by Intel® 82575EB Gigabit Ethernet Controller LOM, four are dedicated for OEM I/O riser)
- One SRC differential pair to SAS Riser
- Four SRC differential pairs to two X24 PCI-Express* expanders
- One SRC differential pair to Midbus LAI

Note: *FBD reference clock routing requires clock-to-clock length matching between FBD agents that are directly connected to each other (i.e. MCH-DIMM1, DIMM1-DIMM2, etc) to within five inches. PCI-Express* SRC does not have any clock-to-clock length matching requirements.*

Memory DB1900G* Gear Ratio Configuration

During memory initialization, BIOS reads the DIMM SPD (Serial Presence Detect) PROMs to gather critical information concerning the DIMM itself. Among this information is the DIMM's supported speed. Once the speed of the memory is determined, it is checked against that of the processor host clock speed. Recognizing the host clock speed, BIOS then configures the appropriate ratios using a gear ratio look up table to ultimately provide the correct memory clock speed.

Serial-ATA (SATA) Sub-system

The Enterprise Southbridge 2 provides six Serial-ATA (SATA) interface with a transfer rate of up to 3.0GB/s. The main board has two internal industry standard 7-pin vertical SATA connectors, which can be cabled directly to a SATA device. As an alternative to using a SAS Riser to support eight SAS drives, an internal x4 SFF 8087 SAS/SATA connector is provided to cable to the SAS Backplane to

support 4 SATA drives. SATA cables should be one meter (40 inches) or less in length.

Enterprise Southbridge 2 Port Configuration

The Enterprise Southbridge 2 Port configuration is as follows:

- SATA0 -> SATA connector1 (goes to SATA-to-PATA converter board, then to optical drive)
- SATA1 -> x4 connector Port0 (goes to SATA Drive 0 on SAS BP)
- SATA2 -> x4 connector Port1 (goes to SATA Drive 1 on SAS BP)
- SATA3 -> x4 connector Port2 (goes to SATA Drive 2 on SAS BP)
- SATA4 -> x4 connector Port3 (goes to SATA Drive 3 on SAS BP)
- SATA5 -> SATA connector2 (extra port, could potentially be used for a SATA Tape Drive)

Flash Devices

The main board has a combined total of 8MB flash memory that contains the system BIOS. The main system BIOS partition fits into 4MB of flash and the other 4MB of Flash is reserved for the “Rolling” (or “backup”) BIOS feature.

Video Subsystem

Feature Overview

A single ATI* RN50 video controller provides the onboard video interface. The ATI* RN50 features the following technologies:

- 2D/3D video accelerator
- Dual DAC for simultaneous port support (Front/Rear video support)
- Resolutions from VGA up to UXGA (1600x1200)
- 32MB Samsung* K4N56163QG-ZC2A DDRII Video Memory
- Digital Video Input/Output (DVI/DVO) interface goes to Intel® Remote Management Module 2 board for KVM support up to 165 MHZ
- 3.3V 32-bit / 33MHz PCI host interface

Video Disable Feature

BIOS can disable the video through a GPIO assigned to the Enterprise Southbridge 2. This disable GPIO is logically ORed with the Enterprise Southbridge 2 PCI reset output and can be driven hold the video chip in reset, effectively disabling it.

USB 2.0 Subsystem

The Enterprise Southbridge 2 provides EHCI host controllers and one EHCI host controller to support USB expansion. The main board utilizes five USB ports from the Enterprise Southbridge 2 USB. The port definitions are as follows:

- USB Ports 0 and 1 from Enterprise Southbridge 2 are routed to the I/O riser board.
- USB Port 2 (Lower) and Port 3 (Upper) from Enterprise Southbridge 2 are routed to the rear panel dual-stack USB connector on the main board.
- USB Port 4 from Enterprise Southbridge 2 is routed to an internal USB header on the main board. This can then be cabled to an optional 5.25” USB tape backup drive.
- USB Port 6 from Enterprise Southbridge 2 is routed to the front panel connector. This port then routes to a USB hub on the front panel I/O board, which then drives three USB ports to front

panel USB connectors.

Trusted Platform Module

The main board supports secure computing by providing an SC19WP18ET28PVEM ST Micro Trusted Platform Module (TPM) on the main board. The device is optional and can be enabled through BIOS and software.

The TPM is a security device that connects to the Enterprise Southbridge 2 LPC bus. This device allows private key generation and storage. In addition, it provides the ability to perform various platform trust metrics and authentication procedures, as outlined in the Trusted Platform Module Specification Version 1.2.

Serial Port Support

The SIO3 provides two serial communication ports:

- Serial A
- Serial B

Serial B is provided by Serial Interface2 to a DB9 connector on the rear panel of the main board. Serial A is provided by Serial Interface1 to an internal unshielded 9-pin shrouded header (2 x 5, with pin 10 removed for keying). The SIO3 also provides a serial interface that can be connected to the Enterprise Southbridge 2 serial port for manageability purposes. The serial port MUX can be configured within the SIO3 to monitor Serial B traffic or redirect serial traffic from the internal serial port connector directly to the Enterprise Southbridge 2.

Serial B is available as an Emergency Management Port (EMP) for remote server management. When used in this mode, it is unavailable to the BIOS/operating system. When server management is setup for Serial Over LAN (SOL) remote server management, Serial B is also unavailable to the BIOS/operating system. More information about the PC87427* SIO serial ports can be obtained from the Winbond* vendor website.

LAN on Motherboard 1 (LOM1)

The Main board LOM1 utilizes the Enterprise Southbridge 2 MAC and Intel® 82563EB Gigabit Ethernet PHY (Physical Layer). The Enterprise Southbridge 2 links to the Intel® 82563EB Gigabit Ethernet PHY through a high-speed serial interface called Kumeran.

The Kumeran interface consists of two sets of Tx/Rx pairs for a total of eight signals. Intel® 82563EB Gigabit Ethernet PHY outputs two Gbit LAN ports and will connect to a 1x2 RJ45 Gbit connector accessible at the rear of the chassis.

Post Code LEDs

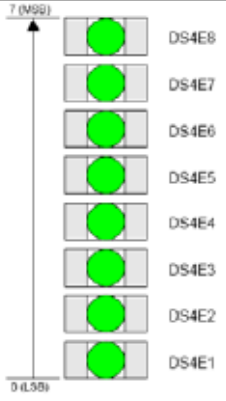
Eight light emitting diodes are used to indicate the raw binary output of BIOS POST codes. Although the value sent to the POST Code LEDs may be the same as the port 80h value at times during the POST process, it is not guaranteed. Table 6 shows the correlation the POST Code bit to LED reference designator.

Table 6. Binary Code Definition

Bit 3	Bit 2	Bit 1	Bit 0	Hexadecimal
0	0	0	0	0
0	0	0	1	1
0	0	1	0	2
0	0	1	1	3
0	1	0	0	4
0	1	0	1	5
0	1	1	0	6
0	1	1	1	7
1	0	0	0	8
1	0	0	1	9
1	0	1	0	A
1	0	1	1	B
1	1	0	0	C
1	1	0	1	D
1	1	1	0	E
1	1	1	1	F

The Post LEDs are situated as shown in the below table along with the corresponding reference designators.

Table 7. POST Code LED Definition

Post Code Bit	LED Reference Designator	POST Code LEDs
7 (MSB)	DS4E8	
6	DS4E7	
5	DS4E6	
4	DS4E5	
3	DS4E4	
2	DS4E3	
1	DS4E2	
0 (LSB)	DS4E1	

Programmable Logic Devices (PLDs)

The main board has two Programmable Logic Devices (PLDs) for fundamental logic on the main board. Due to the nature of these devices, they are not programmable by an end user.

Powergood / Reset

Powergood / Reset: The main board pwren / pwrgd chain begins with logic which checks for both power supplies' presence and power-ok input assertions. Based on these signals, PS_PWROK will assert to start the VR chain on the main board. (See Figure 5 for the VR sequence.)

- Upon assertion of the P1V5_PWRGD signal, VTT_PWREN signal will enable the VTT VR. VTT_PWRGD_3_3V signal from VTT VR to the PLD will enable the CPU#_VR_PWREN to all regulators of populated CPUs.
- After CPU and VTT VRs are enabled, as well as any memory riser presence signal asserted, a global VR enable is asserted for memory risers, SAS backpanel, and SAS Riser. An additional output for IO Riser power enable will be asserted at the same time as the other adapters in the system.
- A signal internal to the PLD representing a system-wide pwrgd signal will be asserted once all FRU pwrgd signals are asserted. This signal is inverted and used to enable clocks. The system powergood is delayed 100ms before the PLD asserts an output for the SYS_PWRGD_PLD signal.

Shifty Bus

The shifty bus provides a way to serially communicate status information to the BMC (Enterprise Southbridge 2). The BMC will assert a signal that will latch 27 bits of system status information present in the PLD. Then a clock signal (~1MHz), generated from the BMC, is used to synchronize a bit banging process that will transfer the information. This process will be repeated on a regular interval for the BMC to track system status.

The shifty bus will also ensure locking of shifty bus state at the moment failure is detected by the PLD. Locking is required because regularly scheduled BMC reads of the shifty bus are not frequent enough and may not allow for accurate failure detection. This will allow the BMC to identify the origin of failure within the system without capturing subsequent failures. CPU/FRU failure will be determined at falling edge of pwrgd inputs. Locking of shifty vector will be determined by the setting of a dirty bit. Once the dirty bit has been set no change in status of shifty vector bits will be recorded. The once the BMC has performed its scheduled shifty read, it is responsible for recognizing the failure status, logging, and shutting down system. The shifty dirty bit will not be cleared until a subsequent shifty bus read while in S5 (or AC cycle).

PCI-Express* Hot-plug:

The main board PLD will implement delay functions for PCI Express hot plug functionality:

- 100ms timer delay for the 3.3V powergood signal from the Texas Instruments* TPS2363 going to each HP-enabled PCI-Express* slot (1-4).
- Generate a 100ms delayed enable (based on slot's 3.3V STBY rail) to the hot-plug isolation logic for slot SMB and wake signals

Power Safe Monitoring

Power Safe Monitoring will monitor power supply status and utilization levels with respect to circuit break type setting to determine when a when a PS non-redundancy bit (shifty bus) should be asserted for reading by the BMC. The BMC will then be responsible for communicating with LM94 controllers to throttle CPUs when appropriate.

PS non-redundancy will be enabled under following conditions:
(Both PS Present AND ((CB_Type=Japan/Brazil AND Util = 37%) OR (CB_Type = Not Japan/Brazil AND Util = 45%))

OR

(One supply present AND ((CB_Type=Japan/Brazil AND Util = 74%) OR (CB_Type = Not Japan/Brazil) AND Util = 90%))

Normal operation mode occurs when a single power supply or both power supplies are present (redundant mode) and functioning. When utilization limits are crossed, the BMC is informed through the shifty bus bit. When two power supplies are present the BMC will check for PS_ACGOOD signals (through PSMI) and allow the system to continue running if both power supplies are functional. If either supply is not functional, or if a single power supply has crossed its utilization limit, the BMC will check whether the supply is low line (100VAC: AC_RANGE = 0), in which case processors will be throttled through FORCE_PR#.

Miscellaneous Functions

- Clock divider- the 32KHz clock input (SUSCLK) will be divided down into a 500Hz clock to be used in delays and a counter throughout the code.
- Debounce circuits- based on previous platforms, power supply PWROK signals and PRES signal require debounce functionality to avoid glitches.
- CPU_SKTOCC – 4 input NAND output using the 4 CPU#_SKTOCC signals going to PS_ON_N logic.

Interrupt and Error Logic Block Diagram

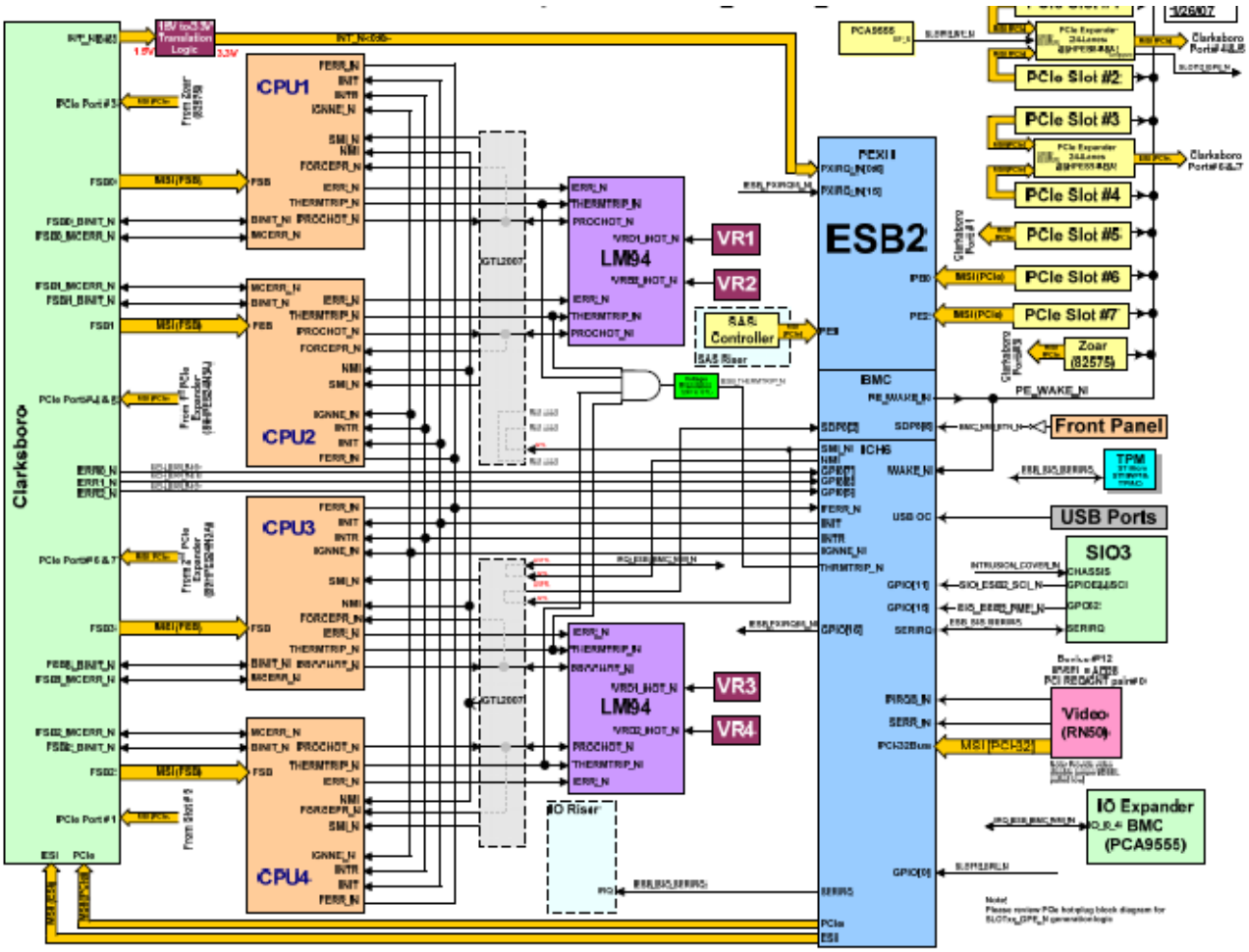


Figure 3. Interrupt and Error Logic Block Diagram

Circuit Breaker Type Jumper

Jumper J6F1 is used to set a threshold for power consumption when operating the server with a single power supply on a low-line 100/110/115/120/127VAC power circuit. This threshold is required to ensure the power consumption of the server does not exceed the power that can be supplied by a single AC power circuit. When the system has two power supplies installed, a separate AC power circuit is needed for each power supply to guarantee the AC power circuit capability is not exceeded.

When a server is connected to low-line power, the J6F1 jumper sets the following power consumption thresholds:

- Pins 1-2 covered: Sets the power consumption threshold to 1180 watts
- Pins 2-3 covered: Sets the power consumption threshold to 1030 watts

Power consumption is based on the power consumed within the system. Power factors for inefficiency are not included in the above figures.

Servers connected to high-line power (200/208/220/230/240VAC) do not have a power consumption threshold. Under these conditions, jumper J6F1 should be set as follows:

- 100/110VAC rated circuit: cover pins 2-3

- 115/120/127VAC rated circuit: cover pins 1-2
- 200/208/220/230/240VAC rated circuit: cover pins 1-2

The power consumption threshold is most likely to be exceeded when all of the following conditions are met:

- The server is connected to a low-line power circuit
- The server has a single power supply installed
- The server is fully configured with four processors, 16 x 4 GB DIMMs, and all PCI slots are filled
- The server is running at maximum performance

If the power consumption threshold is crossed, the hardware throttles the processors to reduce the power consumption to below the set threshold. The processor performance can be returned to the full performance level by power cycling the server.

When two power supplies are installed, the required power is divided between them. By using both circuits, the server can draw more power than the threshold limit for a single power supply.

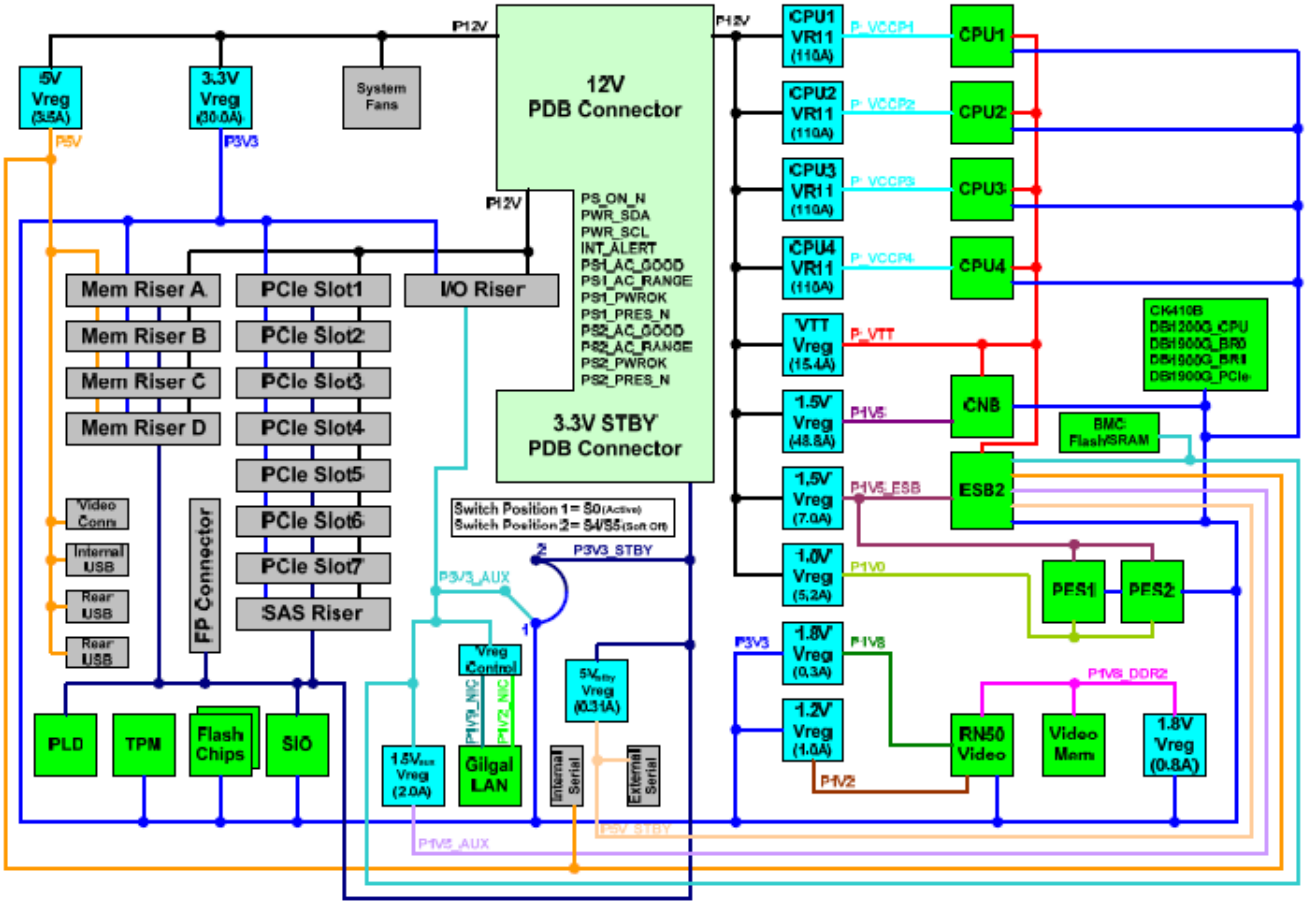
The hardware reduces the amount of power consumed if one of the power supplies fails. This ensures the system consumes less power than the threshold from the single operating power supply. When a failed power supply is replaced, the system is again able to share the power load and operate at full performance.

If the J6F1 jumper is set incorrectly, the following may occur:

- If the jumper is covering pins 1-2 on a 100/110VAC circuit, the server is allowed to consume up to 1180 watts. This setting may cause a circuit breaker to trip.
- If the jumper is covering pins 2-3 on a 115/120/127VAC circuit, the server power consumption threshold is set to 1030 watts. The lower power threshold may be exceeded, limiting system performance.

Power Delivery Block Diagram

The main board takes in P12V (+12V) and P3V3_STBY (3.3V Standby) voltage rails from the system power distribution board. These rails are used to generate the specialized power rails required by components on the main board and are distributed through the main board to other boards in the board set. Figure 4 shows the power delivery flow used on the main board.


12V PDB Connector
3.3V STBY PDB Connector
5V Vreg (3.5A)
3.3V Vreg (30.0A)
System Fans
P12V
P5V
P3V3
PS_ON_N
PWR_SDA
PWR_SCL
INT_ALERT
PS1_AC_GOOD
PS1_AC_RANGE
PS1_PWROK
PS1_PRES_N
PS2_AC_GOOD
PS2_AC_RANGE
PS2_PWROK
PS2_PRES_N
Mem Riser A
Mem Riser B
Mem Riser C
Mem Riser D
PCIe Slot1
PCIe Slot2
PCIe Slot3
PCIe Slot4
PCIe Slot5
PCIe Slot6
PCIe Slot7
SAS Riser
I/O Riser
Video Conn
Internal USB
Rear USB
Rear USB
FP Connector
PLD
TPM
Flash Chips
SIO
Switch Position 1= S0 (Active)
Switch Position 2= S4/S5 (Soft Off)
P3V3_STBY
P3V3_AUX
Vreg Control
P1V5_NIC
P1V2_NIC
1.5V Vreg (2.0A)
Gilgal LAN
5V Vreg (0.31A)
Internal Serial
External Serial
P5V_STBY
P1V5_AUX
CPU1 VR11 (110A)
CPU2 VR11 (110A)
CPU3 VR11 (110A)
CPU4 VR11 (110A)
P_VCCP1
P_VCCP2
P_VCCP3
P_VCCP4
CPU1
CPU2
CPU3
CPU4
VTT Vreg (15.4A)
P_VTT
1.5V Vreg (48.8A)
P1V5
CNB
1.5V Vreg (7.0A)
P1V5_ESB
ESB2
1.0V Vreg (5.2A)
P1V0
PES1
PES2
1.8V Vreg (0.3A)
P1V8
P3V3
1.2V Vreg (1.0A)
P1V2
P1V8_DDR2
RN50 Video
Video Mem
1.8V Vreg (0.8A)
CK410B
DB1200G_CPU
DB1900G_BR0
DB1900G_BR1
DB1900G_PCIe
BMC FlashSRAM

Reset and Powergood Diagram

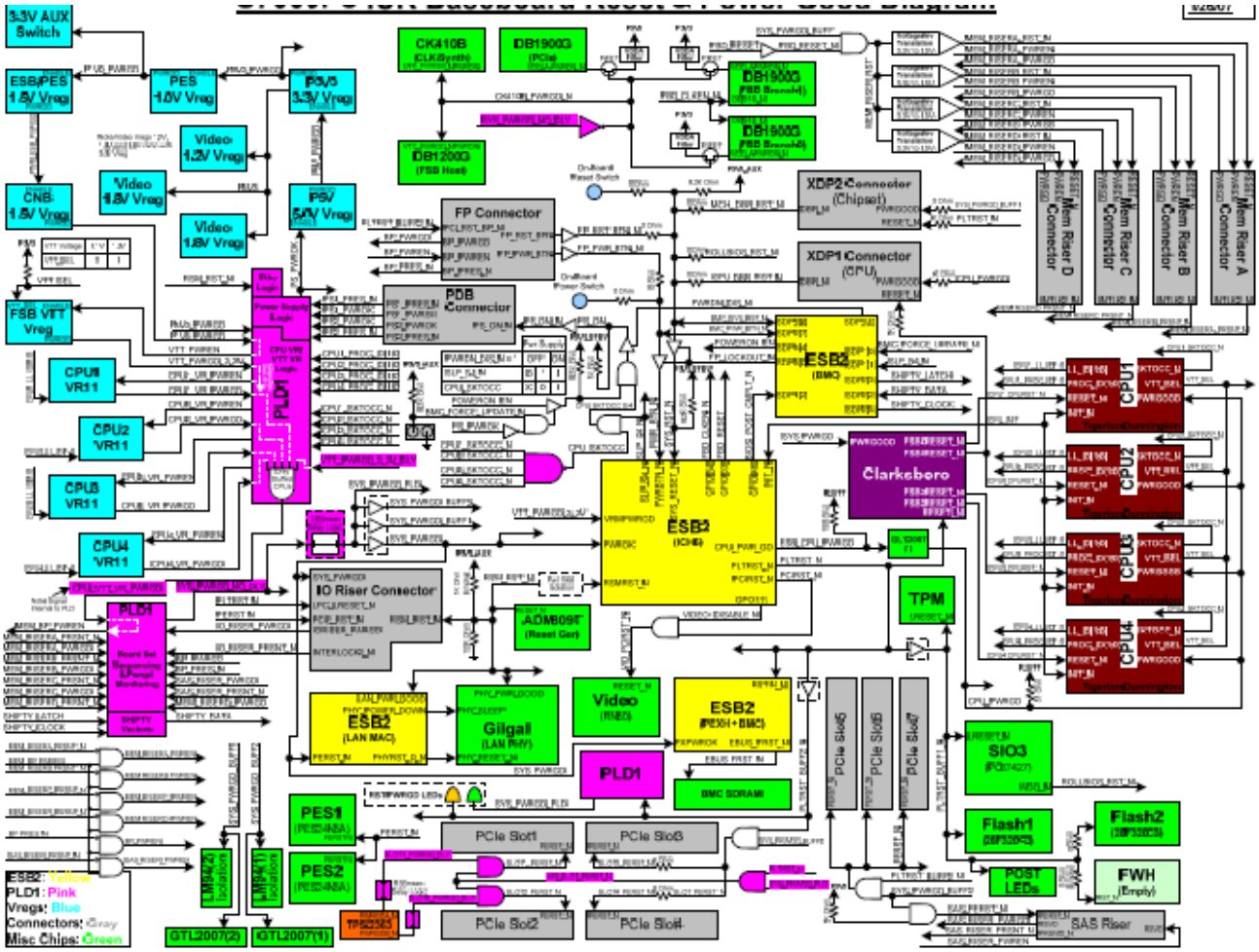


Figure 5. Main Board Reset and Powergood Block Diagram

Power-Up Sequencing Diagram

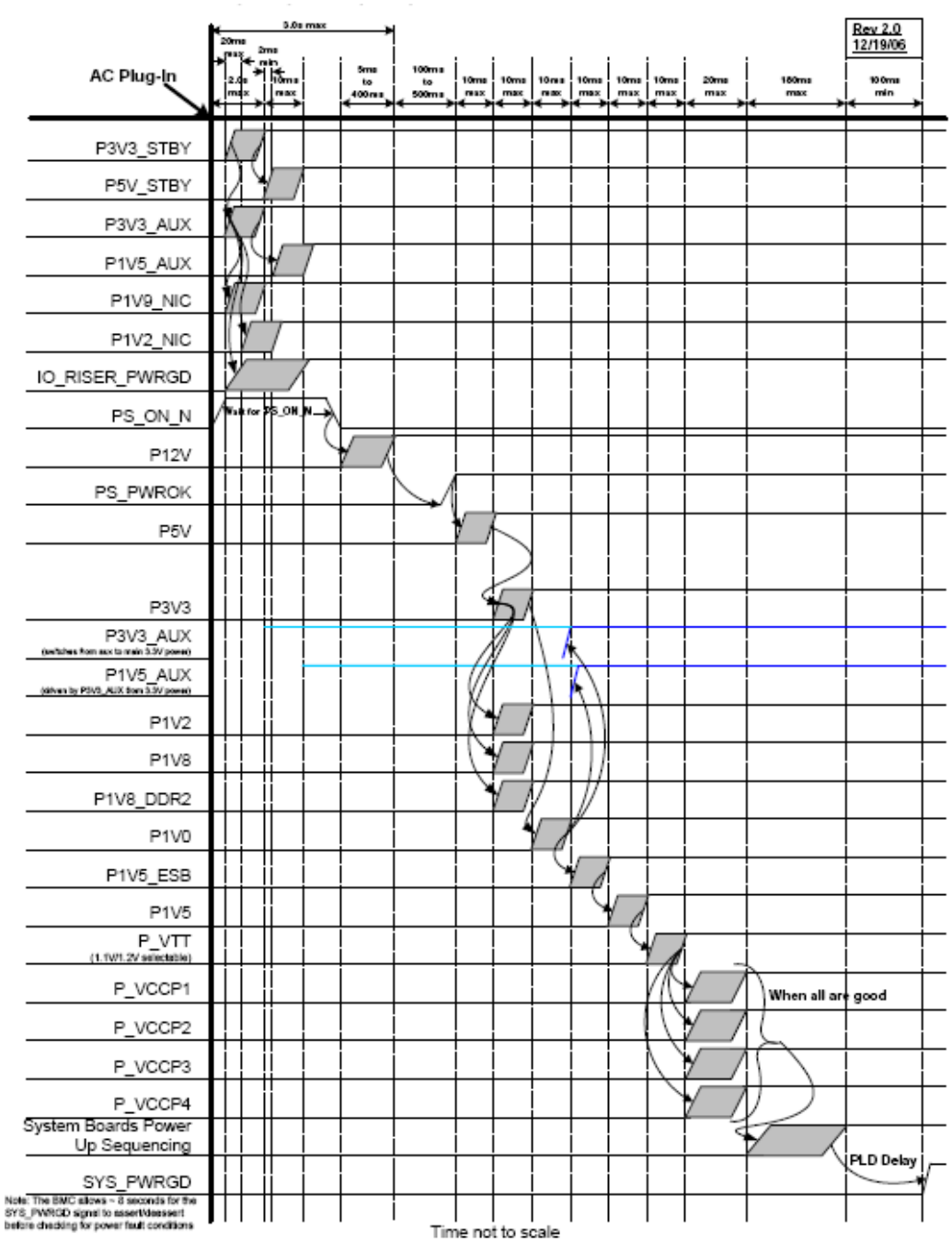


Figure 6. Main Board Power Sequencing Diagram

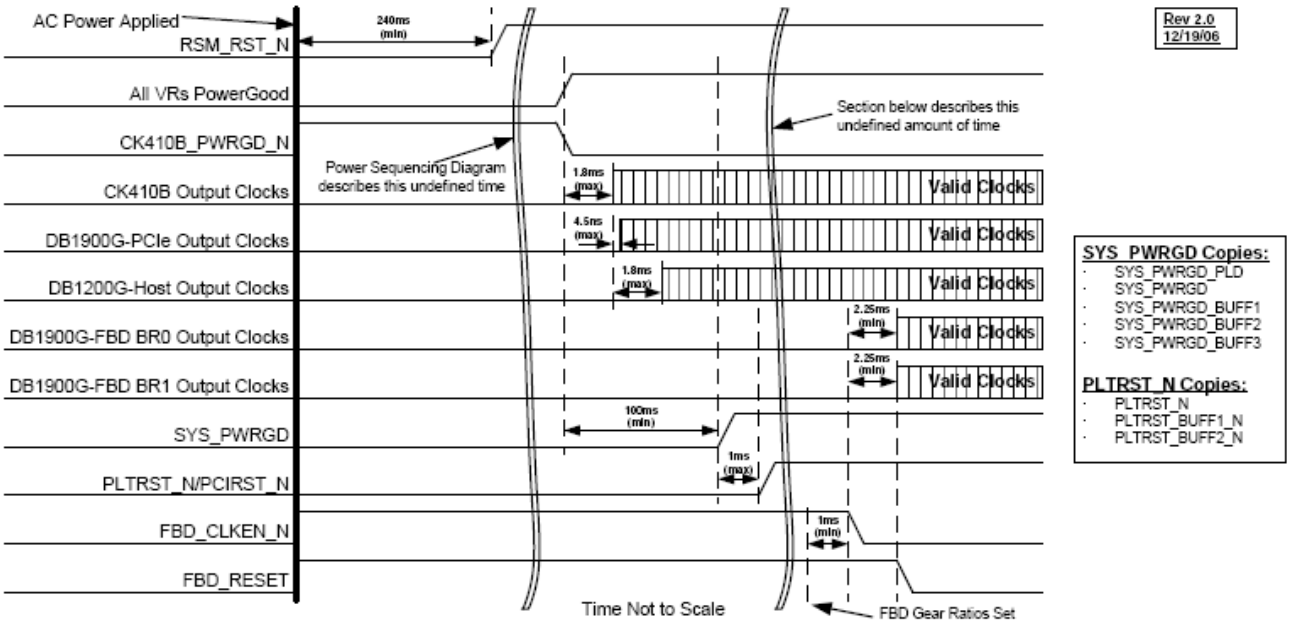


Figure 7. Main Board Reset Sequencing Diagram

Reset and Clock Enabling Sequence

1. Chipset (Intel® 7300 Chipset MCH and Enterprise Southbridge 2) and processors are powered as indicated by corresponding VR Power Goods.

2. On-die power-detect circuitry initiates PLL locking. However, the absence of a reference clock at PLL input triggers the low frequency detect circuit, which shuts the PLL off.

3. Once all VRs have achieved powergood, the PLD asserts CK410B_PWRGD_N low. This brings all of the clock chips out of their power down states.

4. After its PLL has locked (approximately 1.8ms maximum), the CK410 drives reference clocks to all of the downstream differential buffers.

5. Upon receiving a PCI-Express* reference clock, the DB1900G* PCI-Express* (in PLL Bypass Mode) immediately passes this to its outputs (approximately 2.5ns – 4.5ns).

6. Upon receiving a host reference clock, the DB1200G*-Host (in PLL Mode) takes approximately 1.8ms (maximum) before its PLL locks. Any clock chips in PLL Mode will drive output clocks only after their PLLs have locked. This is to ensure that unstable

clocks are not driven to downstream devices.

7. Although receiving an FBD reference clock, the FBD branch clock chips (DB1900G* FBD BR0 and DB1900G* FBD BR1) are held off by an Enterprise Southbridge 2(ICH6) GPO (GPIO[34], FBD_CLKEN_N) until the appropriate gear ratios can be set.

8. Presence of reference clocks at the chipset and processors (except for FBD devices) are detected by the low frequency detect circuits. PLL locking is then re-initiated.

9. Chipset and processor PLLs lock and some time later they receive an external SYS_PWRGD, allowing I/O transactions to commence.

10. FBDIMMs are still in reset and FBD clocks are still disabled. FBDIMMs are kept in reset by BIOS driving high ("1") on Enterprise Southbridge 2 (ICH6) GPO (GPIO[33], FBD_RESET). FBD clocks are kept disabled by BIOS driving high ("1") on Enterprise Southbridge 2(ICH6) GPO (GPIO[34], FBD_CLKEN_N).

11. Once BIOS is ready (undefined amount of time), it will detect the FSB host clock speed, and then detect the FBD clock speed from the FBDIMM SPDs.

12. Based on these speeds, BIOS will select the appropriate gear ratios in the DB1900G* FBD branch clocks. The internal PLLs will lock within 500 us of initial gear setting (regardless of OE# assertion). BIOS will provide a minimum 1ms delay before asserting the Clock Enable GPO low ("0"), thereby releasing the FBD memory clocks.

13. BIOS will then provide a minimum 2.25ms delay between the assertion of clock enable GPO and the de-assertion of the memory reset GPO to ensure that the memory clocks will be fully stable.

Thermal Specifications

The thermal solution designed to support the board must meet the following conditions:

Table 8. Thermal Specifications

Component	Target Velocity	Target Ambient	Temp Specification
Processors	See Processors Thermal Specifications		
Processor sockets	500 lfm	50 °C	100 °C, T_{socket}
CPU core VRDs	400 lfm	50 °C	90°C, T_{sink} @ MAXIMUM
Intel® 7300 Chipset MCH	400 lfm	50 °C	105 °C, T_{HIS}
IDT* 89HPES24N3A	400 lfm	50 °C	105 °C, T_{die}
Intel® 82563EB Gigabit Ethernet PHY	400 lfm	50 °C	105 °C, T_{case}
Enterprise Southbridge 2	400 lfm	50 °C	85 °C, T_{case}
ATI* RN50	400 lfm	50 °C	85 °C, T_{case}
Main board	400 lfm	50 °C	100 °C, T_{board}

Main Board Server Management

This chapter describes the server management related aspects. Server management consists of many embedded technologies. These technologies consist the following:

- Combination of board instrumentation
- Sensors
- Interconnects
- Server management controllers
- Firmware algorithms
- System BIOS

IPMI 2.0 Features

- Baseboard management controller (BMC).
- Watchdog timer.
- Messaging support, including command bridging and user/session support.
- Chassis device functionality, including power/reset control and BIOS boot flags support.
- Alert processing device including platform event trap (PET) and Simple Network Management Protocol (SNMP) alerts via LAN interfaces.
- Platform event filtering (PEF) device.
- Event receiver device: The BMC receives and processes events from other platform subsystems.
- Field replaceable unit (FRU) inventory device functionality: The BMC supports access to system FRU devices using IPMI FRU commands.
- System event log (SEL) device functionality: The BMC supports and provides access to a SEL.
- Sensor device record (SDR) repository device functionality: The BMC supports storage and access of system SDRs.
- Sensor device and sensor scanning/monitoring: The BMC provides IPMI management of sensors. It polls sensors to monitor and report system health.
- IPMI interfaces.
- Host interfaces include system management software (SMS) with receive message queue support, and server management mode (SMM).
- Terminal mode serial interface.
- PCI-SMBus interface that allows PCI cards to send commands to the BMC using an IPMB-like command protocol.
- IPMB interface.
- LAN interface that supports the IPMI-over-LAN protocol (RMCP, RMCP+).
- Serial-over-LAN (SOL).
- ACPI state synchronization: The BMC tracks ACPI state changes that are provided by the BIOS.
- BMC self test: The BMC performs initialization and run-time self-tests, and makes results available to external entities.

Non IPMI Features

- BMC firmware update using firmware transfer mode (FTM).
- BMC on-line update: BMC rolling update that supports a redundant firmware image.
- Fault resilient booting (FRB): FRB2 is supported by the watchdog timer functionality
- Chassis intrusion detection and chassis intrusion cable presence detection.

- Basic fan control using TControl version 2 SDRs.
- Fan redundancy monitoring and support.
- Power supply redundancy monitoring and support.
- Hot-swap fan support.
- Acoustic management: Support for multiple fan profiles.
- Alert Standard Forum (ASF) power-on self-test (POST) progress queuing: The BMC queues POST messages sent from the BIOS and makes these accessible through IPMI interfaces.
- Signal testing support: The BMC provides test commands for setting and getting platform signal states.
- The BMC generates diagnostic beep codes for fault conditions.
- System GUID storage and retrieval.
- Memory reliability, availability, and serviceability (RAS): The BMC provides sensors to track DIMM state and memory RAS redundancy state. The BMC provides IPMI OEM commands to enable the BIOS to push this information to the BMC.
- Front panel management: The BMC controls the system status LED and chassis ID LED. It supports secure lockout of certain front panel functionality and monitors button presses. The chassis ID LED is turned on using a front panel button or a command.
- Power state retention.
- Power fault analysis.
- Intel® Light-Guided Diagnostics.
- Power unit management: Support for power unit sensor. The BMC handles power-good dropout conditions.
- DIMM temperature monitoring: New sensors and improved acoustic management using closed-loop fan control algorithm taking into account DIMM temperature readings.
- Non-maskable interrupt (NMI): Provides commands to set/get NMI source. Supports generation of NMI due to watchdog timer, IPMI command, or front panel NMI button. Monitors system NMI signal.
- Address Resolution Protocol (ARP): The BMC sends and responds to ARP (supported on ESB2-embedded NICs)
- Dynamic Host Configuration Protocol (DHCP): The BMC performs DHCP (supported on ESB2-embedded NICs).
- BMC internal timeclock sync with SIO RTC: At BMC startup, the BMC reads the SIO RTC and updates its internal timeclock. The BMC updates the SIO RTC when it receives the Set SEL Time IPMI command from the BIOS.
- Chassis intrusion fan interactions: Fans go to high speed when the chassis intrusion signal is asserted.
- PCI-Express* link status monitoring support: The BMC maintains sensors for PCIExpress status and provides interface commands for the BIOS to push state information to the BMC.
- Intel® Remote Management Module 2 support: The BMC supports an add-in that uses its own dedicated NIC, the GCM3 NIC, to provide advanced server management features via Out of band.
- Platform environment control interface (PECI) thermal management support

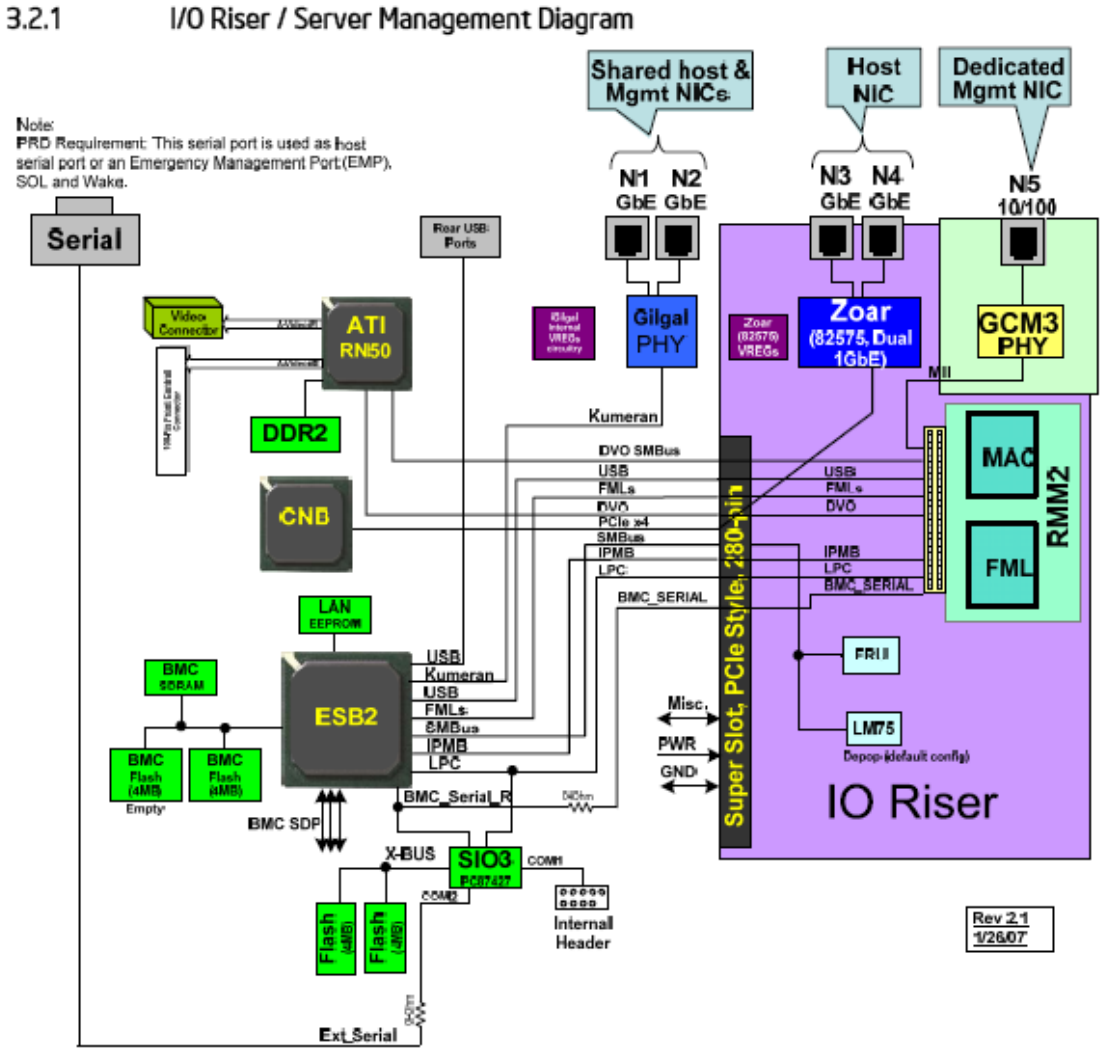


Figure 8. I/O Riser / Server Management Diagram

Sensor Data Record SDR (SDR) Repository

The BMC implements a logical Sensor Data Record (SDR) repository device. The SDR repository is accessible via all communication transports, even while the system is powered off.

Field Replaceable Unit (FRU) Inventory Devices

The BMC implements the interface for logical FRU inventory devices. This functionality provides commands used for accessing and managing FRU inventory information. These commands can be delivered via all interfaces.

The BMC provides FRU command access to its own FRU device, as well as to the FRU devices throughout the system. The FRU device ID mappings are shown in Table 9 and SMBus connectivity shown in Figure 9. The BMC controls the mapping of the FRU device ID to the physical device. Per the IPMI specification, FRU device 0 is always located on the main board. All Intel-designed server boards maintain onboard non-volatile storage to hold the FRU data.

Table 9. FRU Device Location and Size

FRU Device ID	SMBus Bus (* = CLK/DAT)	SMBus Address	Device	Read Only	Size (bytes)
0	SMB_BB_SENSOR_3V3SB_*	0xA0	Main board	RW	8K
1	SMB_SYS_BARD_*	0xA6	Front panel board	RW	256
2	SMB_SENSOR_SEG2_3V3SB_*	0xA8	I/O riser	RW	256
None	SMB_IPMB_*	None	Intel® Remote Management Module 2 (Not supported)	N/A	None
3	SMB_SYS_PWR_*	0xAA	Power distribution board	RW	256
4	SMB_SYS_PWR_*	0xAC	Power Supply Unit 1	RO	256
5	SMB_SYS_PWR_*	0xAE	Power Supply Unit 2	RO	256
6	SMB_SAS_3V3SB_*	0xA8	SAS riser	RW	256
VSC410	SMB_IPMB_5VSB_*	0xAC	SAS backplane	RW	8K
7	SMB_MEM_3V3_*	0xA4	Memory Riser A	RW	256
8	SMB_MEM_3V3_*	0xA6	Memory Riser B	RW	256
9	SMB_MEM_3V3_*	0xA2	Memory Riser C	RW	256
10	SMB_MEM_3V3_*	0xA0	Memory Riser D	RW	256

System Event Log (SEL)

The BMC allocates memory space for logging system events. SEL events can range from critical system errors to basic system monitoring reports. The SEL can be cleared in the system BIOS setup, or by using the SEL viewer utility or Intel® System Management application.

Real-Time Clock (RTC) Access

The SIO on this platform allows the BMC to have its own private RTC. In order for the BMC to remain in sync with the system RTC, the BIOS must send the Set SEL Time command with the current system time to the BMC during system boot and before system shut-down. If the time is modified through an OS interface, then the BMC's time is not synchronized until the next system reboot.

Rolling BIOS

The main board provides two flash chips that can contain two independent BIOS versions. This allows BIOS updates without a system reboot as well as failover to a good BIOS image in the event of BIOS corruption. Rolling BIOS support is controlled entirely by BIOS and requires no support from BMC Firmware.

BMC EBus Memory Sub-System

The main board provides two flash chips for BMC memory. The first flash chip represents the base BMC memory and the second represents the expansion BMC memory (Empty). This additional memory allows the BMC to store more complicated server management programs, as opposed to only using the BMC serial EEPROM.

The BMC memory sub-system also includes an SDRAM device, which allows the BMC to temporarily store and fetch data without using system memory. The BMC communicates with the flash and SDRAM memory through an Expansion Bus (Ebus).

Supported Features

Fan Speed Control

The BMC monitors and controls system fans, with each fan having a tachometer sensor used to determine cooling system health. The fan subsystem has three states:

- Sleep
- Nominal
- Boost

Nominal is the default state. In this state, fan speeds are based on the ambient system temperature. A system temperature threshold is set via an SDR. When the threshold is exceeded, it linearly ramps the fan speeds either until the fan speed reaches maximum saturation or the temperature reduces below the threshold.

If the system temperature stays below the threshold, fan speed ramps back to the default speed. If system temperature remains above the threshold, the system (through Closed Loop Thermal Throttling – CLTT) may throttle memory to reduce heat dissipation. Fan settings are configurable via SDRs to allow for the specific cooling requirements needed by system integrators. A test command can also be issued to manually force the fan speed to a selected value, overriding any other control or policy.

Hardware Monitoring (LM94*)

The main board platform design uses two LM94* hardware monitoring devices. The LM94*'s can be controlled/monitored by BMC over the System Management Bus (SMBus). Please see Figure 9 for LM94* SMBus connectivity. The following LM94* features are supported on the main board.

- CPU monitoring (PROCHOT#, THERMTRIP#, IERR#)
- CPU Vreg monitoring (VRHOT#)
- Fan control/monitoring (PWM, Tach)

PECI (ADT7490*)

The main board implements the Platform Environment Control Interface (PECI) to monitor processor temperature. PECI is a one-wire bus interface that provides a communication channel between an Intel processor (and potentially future chipset components) and an external monitoring device. In the case of main board, the monitoring device has been chosen to be the ADT7490*.

Note: *The ADT7490* will complement other fan control and temperature monitoring devices and is in addition to two National (Winbond*) LM94* devices.*

PECI uses a single wire for wake-up, self-clocking and data transfer. No additional control signals are required. Each bit transferred will begin with a driven, rising edge from an idle level near zero volts. The duration of the signal driven high depends on whether the bit value is a logic “0” or logic “1”. PECI also includes variable transfer rate established with every message.

Further details on the bus implementation are available in the Platform Environment Control Interface (PECI) Specification. The details are as follows:

- Electrical requirements
- Platform topologies
- Power management state handling
- Bus device enumeration, commands
- Address values

The PECI architecture includes system host and client devices. A system host, represented inmain board by the ADT7490*, is a special device with specific bus management duties.

Only one system host is allowed on a PECI physical layer and this device initiates all transactions. Client devices, represented by processors, are any other devices connected to PECI in the system that are not the system host. The main board PECI physical layer topology supports a 4-way symmetric multi-processor system. PECI devices are identified by their unique, fixed address.

All processor PECI devices are located in the address range of 0x30 to 0x33. Figure 11 provides an example implementation.

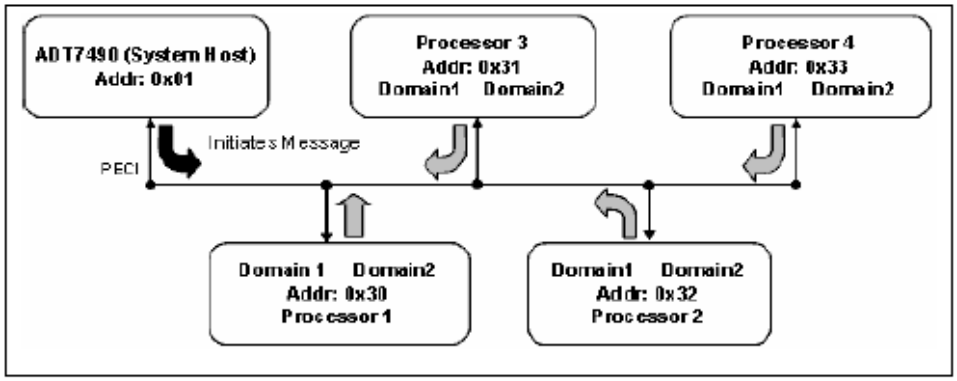


Figure 11. Main Board PECI Physical Layer Topology

The System Host, ADT7490*, supports three possible SMBus addresses, which include the following:

- 0x58
- 0x5A
- 0x5C

Address 0x5A has been assigned to this device. When the device is powered up with Pin 13 (PWM3/ADDREN) high, the ADT7490* has a default SMBus address of 1011100 or 0x5C. To change this address, the device is placed in ADDR SELECT mode by strapping Pin 13 low on power-up. The logic state of Pin 14 then determines the device's SMBus address. The logic of these pins is sampled on power-up. The device address is sampled on power-up and latched on the first valid SMBus transaction. More precisely the device is latched on the low-to-high transition at the beginning of the 8th SCL pulse, when the serial bus address byte matches the selected slave address. The selected slave address is chosen using the ADDREN pin/ ADDR SELECT pin. Any attempted changes in the address have no effect after this.

Within a single client device there may exist multiple PECI domains. A PECI domain is a subpartition that is served by an independent PECI client but shares the same device address.

The PECI GetTempx commands are used to retrieve temperature from a target PECI address. GetTemp0 specially refers to PECI domain 0 in processor's with multiple domains. GetTemp1 is used to read the temperature from the other domain.

PECI temperature data is returned as a negative value representing the number of degrees centigrade below the thermal control circuit activation temperature of the PECI device. Additionally, the associated Tcontrol value is also a negative number below the thermal control circuit activation temperature.

No thermal diode support is included for the processors. Therefore, server management must now manipulate relative numbers. Other components in the platform continue to provide absolute temperature readings.

PECI operates at the processor's P_VTT voltage level. All DC electrical specifications, including hysteresis, thresholds and current requirements are included in the Platform Environment Control Interface (PECI) Specification and applicable processor EMTS.

PECI System Requirements

To implement PECI enabled processors, the systems must perform various functions. These include, but are not limited to the following:

- Detecting the presence of PECI enabled processors
- Determining the number of PECI domains per physical package
- PECI Fault handling:

PECI Enabled Processor Presence Detection

The system must detect the presence of PECI enabled processors. BIOS or other system management / initialization software can determine if a particular processor supports a PECI interface by utilizing the processor's CPUID instruction to return the processor signature.

Number of PECI Domains Per Physical Package Determination:

The system must determine the number of PECI domains per physical package. Once software has established the presence of a PECI enabled processor, it must determine the number of PECI domains supported per physical package.

The number of PECI domains provides information to manageability hardware/software regarding how many temperature requests should be made to determine the temperature for the physical package.

PECI Fault Handling

The system must determine PECI Fault handling. System RESET#, both during system powerup and during a previously powered re-boot, produces a window of time during which PECI transactions cannot be guaranteed to complete or for data to be valid. Additionally, the PECI addressing in the processor may not resolve until after RESET# deassertion, which is critical for multiple processor-based systems to operate correctly.

Processor Throttling

Processor throttling is the ability of the processor to reduce core speed, and thereby its heat, when generated heat exceeds normal thermal thresholds. There are several ways a processor can be throttled:

Processor Method

If a processor reaches a certain temperature, it will assert PROCHOT_N. At the same time the processor will internally assert its own FORCEPR_N (Force Power Reduction) internally, thereby throttling the processor to reduce the temperature.

Processor Voltage Regulator (VR) Method

If a processor VR reaches a certain temperature, it will assert VRHOT_N to the LM94. This willcause the LM94 to assert its PROCHOT_N output to the GTL2007, which will then drive the FORCEPR_N input on that processor thereby throttling the processor to reduce the temperature.

BMC Method

If a preset power consumption threshold of the system is crossed, it will be detected by the BMC. The BMC can assert PROCHOT_N to the GTL2007 from the LM94 through use of the SMBus. This will cause the GTL2007 to assert FORCEPR_N on a processor thereby throttling the processor to reduce the temperature.

Memory Throttling

Memory throttling is the ability of the chipset to reduce bandwidth of the FBDIMMs when their generated heat exceeds the normal thermal threshold. Each FBDIMM has an internal temperature sensor on its Advanced Memory Buffer (AMB). Temperature readings from all FBDIMM AMBs are monitored by the BMC and used to drive a Closed Loop Thermal Throttling (CLTT) scheme.

As a fallback to CLTT, an Open Loop Thermal Throttling (OLTT) scheme is also available on the server system. Each Memory riser has a temperature-sensing device that provides the difference between the left and right sides of the DIMMs. This difference estimates the heat generated by the DIMMs and is continuously monitored by the BMC.

In both cases, depending on memory riser temperature readings, memory may be throttled back and fans nearby to the memory riser(s) may be boosted. Whenever this temperature reaches the upper critical threshold, the BMC requests the AMB on the DIMMs to enable DIMM throttling. Memory throttling is also enabled when the system chassis intrusion sensor is engaged and in the event of a

system fan failure or removal.

ACPI Power Control

The main board supports ACPI S0, S1, S4 and S5 system/sleep states.

- The S0 system state is the normal power on state and is required for normal system operation.
- The S1 sleep state is a lower power state where the processor stops executing instructions.
- The S4 state, also called Suspend to Disk, is a “soft-off” state (like S5), whereby all of the Operating System Contents are stored to the hard drive.
- The S5 system state is the normal power off state (or “soft-off”) and is required in order to perform certain maintenance tasks.

When the system is operating in ACPI mode, the operating system retains control of the power of the system. During ACPI mode, operating system policy determines the entry methods and wakeup sources for each system/sleep state. An ACPI-enabled operating system generates a System Management Interrupt (SMI) to request that the system enable ACPI support. The BIOS responds to the SMI by communicating to the BMC that ACPI support is required.

S1 Sleep State Support

During the S1 Sleep State, the following events take place.

- The front panel power LED blinks at a rate of 1 Hz with a 50% duty cycle (not controlled by the BMC).
- If enabled via the Set ACPI Configuration Mode command, the server board fans are set to sleep speed as specified in the associated OEM TControl SDR for each fan domain. Otherwise, fan control is the same as for ACPI S0 state. The DIMM temperature sensors do not contribute to the fan speed control algorithm.
- The watchdog timer is stopped.
- The power, reset, front panel NMI, and ID buttons are unprotected.

The BMC detects that the system has exited the ACPI S1 sleep state when it is notified by the BIOS SMI handler.

S4 and S5 System State Support

Network adapters hold the wake configuration state for Wake On LAN (WOL). This is typically configured by the operating system and is not cleared by a system reset. However, WOL date information should be cleared when going into S4/S5 system state. When a WOL Magic Packet* is received by the BMC, the system powers on if WOL is enabled in BIOS setup.

The WOL feature is supported for the onboard, IO Riser and PCI-Express* plug-in network adapters.

Wake On LAN

Legacy Wake On LAN (WOL) is supported by multiple devices within the board set.

On the main board, the 82563EB and Enterprise Southbridge 2 MAC support wake on LAN. The MAC will generate PE_WAKE# signal upon detecting a filter packet. PE_WAKE# is connected to WAKE# pin of the Enterprise Southbridge 2 core via the platform. Upon receiving WAKE#, the Enterprise Southbridge 2 suspend well logic will start resume power activity. It will commence S4/S5 Exit process by reversing the process of S4/S5 Entry.

On the I/O riser, the 82575EA Gigabit Ethernet controller also supports WOL.

The 82575EA Gigabit Ethernet controller, which is a PCI-Express* based device, can generate a wake event by asserting the WAKE# signal. The assertion of a WAKE# signal causes the system to return to the ACPI S0 sleep state. Following is a detailed description of the 82575EA controller wake operation.

“Advanced Power Management Wakeup”, or “APM Wakeup”, was previously known as “Wake on LAN”. It is a feature that has existed in the 10/100 Megabit NICs for several generations. The basic premise is to receive a broadcast or unicast packet with an explicit data pattern, and then to assert a signal to wake-up the system. In the earlier generations, this was accomplished by using a special signal that ran across a cable to a defined connector on the motherboard. The NIC would assert the signal for approximately 50ms to signal a wakeup.

The 82575EA Gigabit Ethernet controller uses (if configured to) an in-band PM_PME message for this. On power-up, Intel® 82575EB Gigabit Ethernet Controller will read the APM Enable bits from the EEPROM Initialization Control Word 2 into the APM Enable (APME) bits of the Wakeup Control Register (WUC). These bits control enabling of APM Wakeup. When APM Wakeup is enabled, Intel® 82575EB Gigabit Ethernet Controller checks all incoming packets for “Magic Packets*”.

Reference the definition of “Magic Packets*”. Once the 82575EA Gigabit Ethernet controller receives a matching Magic Packet*, it will:

- If the Assert PME On APM Wakeup (APMPME) bit is set in the Wake Up Control Register (WUC):

- Set the PME_Status bit in the Power Management Control / Status Register (PMCSR) and issue a PM_PME message (in some cases, this may require to assert the WAKE# signal first to resume power and clock to the PCI-Express* interface).

- Store the first 128 bytes of the packet in the Wake Up Packet Memory (WUPM).

- Set the Magic Packet Received bit in the Wake Up Status Register (WUS).

- Set the packet length in the Wake Up Packet Length Register (WUPL).

The 82575EA Gigabit Ethernet controller will maintain the first Magic Packet* received in the Wake Up Packet Memory (WUPM) until the driver writes a 1 to the Magic Packet* Received MAG bit in the Wake Up Status Register (WUS). “APM Wakeup” is supported in all power states and only disabled if a subsequent EEPROM read results in the APM Wake Up bit being cleared or the software explicitly writes a 0 to the APM Wake Up (APM) bit of the WUC register.

The NIC maintenance port associated with the Intel® Remote Management Module 2 does not support WOL.

There are no LAN devices on PCI32, therefore PME# is not utilized for Wake-On-LAN.

Secure Mode Operation

Since the BMC is not logically located between the power button and the chipset, additional front panel lockout buffers must be used on the ProServ 4680 Server System in order to support Secure Mode Operation. These lockout buffers allow the BMC to prevent the user from powering off or resetting the system.

The BMC logs secure mode violation events into the SEL when secure mode is enabled and a user presses front panel buttons that are in a protected state. Secure mode is cleared when the following occurs:

- AC power or system power is applied
- System reset occurs
- BMC reset occurs

Table 10. Secure Mode During ACPI States

ACPI System State	Power Switch	Reset Switch
S0 (On)	Protected	Protected
S1 (CPU Sleep)	Unprotected	Unprotected
S4/S5 (Off)	Unprotected	Unprotected

Intelligent Platform Management Bus (IPMB)

The IPMB is a communication protocol that utilizes a 100 KB/s I2C bus. The IPMB implementation in the BMC is compliant with the IPMB v1.0, revision 1.0, with the BMC having an IPMB slave address of 0x20.

The BMC both sends and receives IPMB messages over the IPMB interface. Non-IPMB messages received via the IPMB interface are discarded. In addition to the public IPMB, the BMC has five private I2C buses that extend throughout the system. Figure 9 shows all the I2C buses.

Serial Over LAN (SOL)

Serial Over LAN (SOL) provides bi-directional transport of system Serial B port data encapsulated in IPMI over LAN packets. This provides the following:

- Out-of-band LAN access to the BIOS console redirection
- Service partition application communication
- Operating system console interaction without the BIOS
- Software being LAN-enabled or aware of anything beyond a serial port interface

The console type is set to VT100+ and data bits are set to 8bits/charatecter, no parity and one stop bit as per IPMI messaging requirement.

The BMC supports the Intel proprietary SOL (now known as SOL 1.0) as well as the IPMI 2.0-defined SOL feature, implemented as a standard payload type over RMCP+. The boardset provides the SOL interface via the Intel® Remote Management Module 2 (RMM2) and RMM2 NIC port.

Emergency Management Port (EMP) Interface

The EMP interface is the Intel implementation of the IPMI 2.0 over serial feature, providing an out-of-band RS232 connection into the server management subsystem. This gives system administrators the ability to access low-level server management firmware functions by using commonly available tools. To make it easy to use and provide the most compatibility with LAN and IPMB protocols, the protocol adopts some features of both protocols.

Both the basic and PPP proxy modes of IPMI over serial are supported and are available regardless of the system DC power state. The following EMP features are supported:

- Hardware handshaking
- Data Carrier Detect (DCD) signals

The main board provides the EMP interface through the Serial B external RS232 connector. The BMC has control over which agent (BMC or system) has access to Serial B.

Chassis Intrusion

A three-pin chassis intrusion header is supported for the front of the system. This is intended tosupport a micro-switch that is normally open when the chassis cover is installed on the system.

The micro-switch will close and make a connection to ground when the chassis cover is opened or removed. Pin 3 of the chassis connector gives BIOS the ability of detecting if the chassis intrusion

cable is installed. If the cable is absent then BIOS will drive the fans to high speed. These intrusion signals are routed to the SIO3 Chassis Intrusion pin and GPIO pin (for Cable Present).

Memory Riser

This chapter describes the memory riser. Up to four memory risers plug vertically into the main board. The memory riser has the following features:

- Supports up to eight FBD Generation 1 DIMMs
- Supports FBD speeds of 533MT/s (4-4-4, 5-5-5 latencies) and 667MT/s (5-5-5 latency)
- Supports FBDIMM configurations of x8, x4, single, dual-rank DDR2 DRAMs
- Supports DDR2 DRAM technologies of 512Mbit, 1Gbit and 2Gbit
- Supports Closed Loop Thermal Throttling by using FBDIMM AMB temp sensors
- Supports Open Loop Thermal Throttling by using on board temp sensor (NE1617) -
- Optional PCI-Express* x16 card edge connector that plugs into the main board

LED fault indicators for each DIMM

- On-board voltage regulators for 0.9V (Vtt), 1.5V (Vcc), and 1.8V (Vdd)
- One Field Replaceable Unit (FRU) EEPROM
- Supports Memory Mirroring and Memory Sparing.

Functional Architecture

FBD Memory Sub-system Overview

The Intel® 7300 Chipset MCH on the main board supports a fully buffered DIMM (FBD) memory subsystem. The FBD interface consists of four channels (Ch A, B, C and D) routing to a total of four memory risers (1 channel per riser, with 8 FBD connectors per riser). This data bus is a 24 lane per channel (14 northbound and 10 southbound) point-to-point high-speed differential interface. For each channel, FBD transfers 144 bits every Northbound data frame, which is equivalent to the 18 byte data (16 bytes of data, 2 bytes of ECC) transfer of an ECC DDR DIMM in a single clock. So the peak theoretical throughput of the Northbound data connection is identical to that of a DDR2-533 subsystem (8 bytes x 533MT/s = 4.267GB/s). Since the southbound lanes include commands along with data, the peak theoretical throughput of the southbound data connection is equivalent to half a DDR2-533 (2.133GB/s). Therefore, the overall peak theoretical throughput of an FBD-533 (267MHz DDR2 reference clock) channel would be 6.4BGB/s (northbound plus southbound). The peak theoretical throughput would scale
to 8GB/s, as the DDR2 reference clock and FB DIMMs increase to 333MHz and FBD-DDR2-667, respectively. The peak theoretical bandwidth is actually limited by several factors including:

the number of DIMMs, memory capacity, bus utilization, and various sub-system latencies.

The following table lists the FBD maximum bandwidths supported by the memory sub-system.

Table 11. Memory Riser Max Memory Bandwidth

533 Channel BW Theoretical = 6.4GB/s	Max BW per Channel	Max BW* per DIMM
1 DIMMs/Channel	2.8GB/s	2.80GB/s
2 DIMMs/Channel	3.5GB/s	1.75GB/s
4 DIMMs/Channel	3.8GB/s	0.95GB/s
8 DIMMs/Channel	3.8GB/s	0.48GB/s
667 Channel BW Theoretical = 8.0GB/s	Max BW per Channel	Max BW* per DIMM
1 DIMMs/Channel	3.0GB/s	3.0GB/s
2 DIMMs/Channel	3.9GB/s	1.95GB/s
4 DIMMs/Channel	4.6GB/s	1.15GB/s
8 DIMMs/Channel	4.6GB/s	0.58GB/s

**Note: Assumes FBDIMMs of equal capacity*

The FBD memory subsystem can be broken into different segments (see following topology). The first segment, called the FBD southbound Channel, is a high-speed differential, point to point, frame-based interface from the Intel® 7300 Chipset MCH to an Advanced Memory Buffer (AMB) that resides on the DIMM module. The next segment, called the DDR2 Channel, represents the interface between the AMB and DDR2 DRAM devices. The AMB takes the frame-based data packets from the FBD southbound channel and translates them into standard JEDEC-DDR2 based data and commands. In addition, the AMB also repeats the southbound data packets from the MCH and sends them on to the second DIMM/AMB in the chain via its own southbound Channel. This southbound data transfer scheme is repeated for all DIMMs in the chain or channel (eight DIMMs per channel in the case of the memory riser). The other FBD segment, called the northbound FBD Channel, represents the path for all data coming from the DIMMs going northbound to the Intel® 7300 Chipset MCH. Similar to the southbound Channel (but in the opposite direction), the data on the northbound Channel gets repeated by each DIMM/AMB in the chain until it eventually arrives at the MCH. FBD Channels A and B combine to represent Branch 0, and FBD Channels C and D combine to represent Branch 1. Each branch controller and every DIMM will receive its own FBD reference clock. The FBD Channel reference clock is exactly half the DDR2 reference clock. The PLL within the AMB on the DIMM module will multiply the input reference clock input by two and deliver the appropriate clock speed to all of the DDR2 DRAM devices.

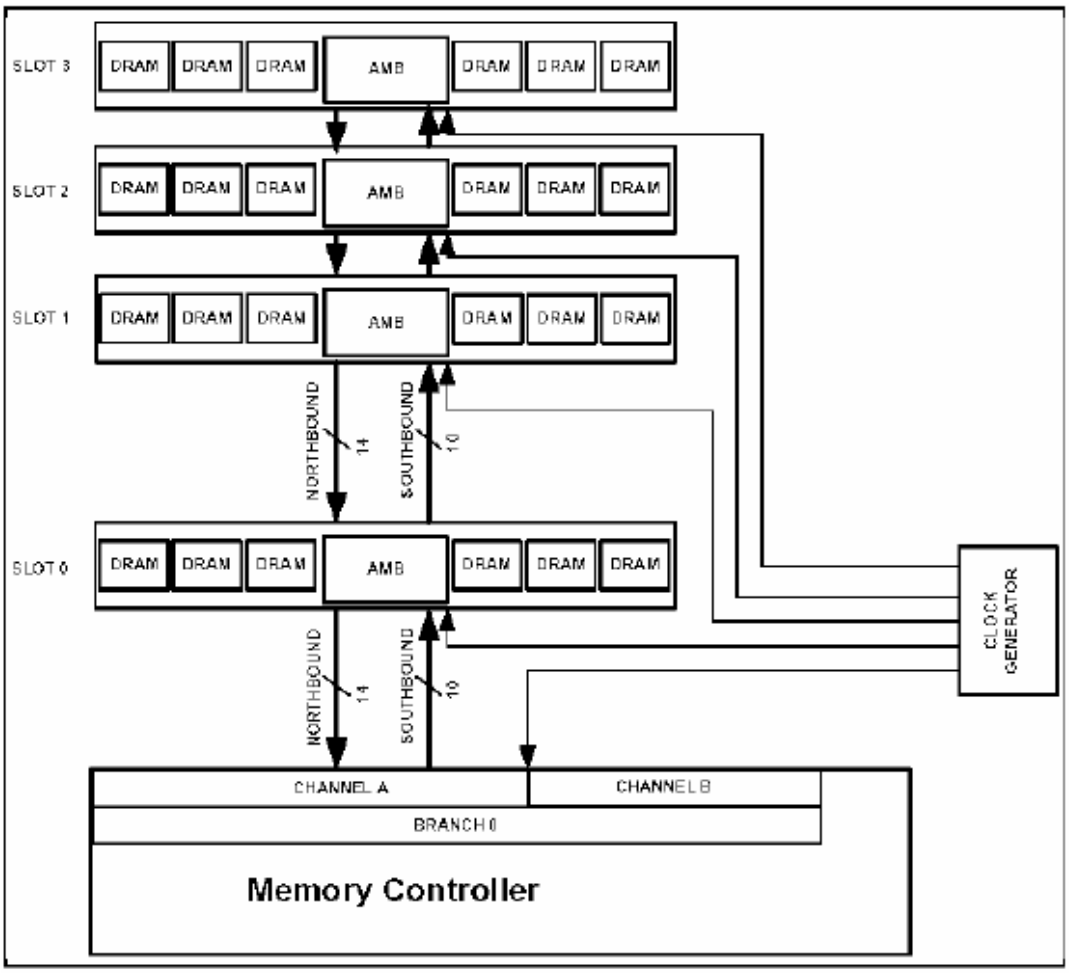


Figure 12. Fully-Buffered DIMM Topology (Generic)

Memory Riser Functional Diagram

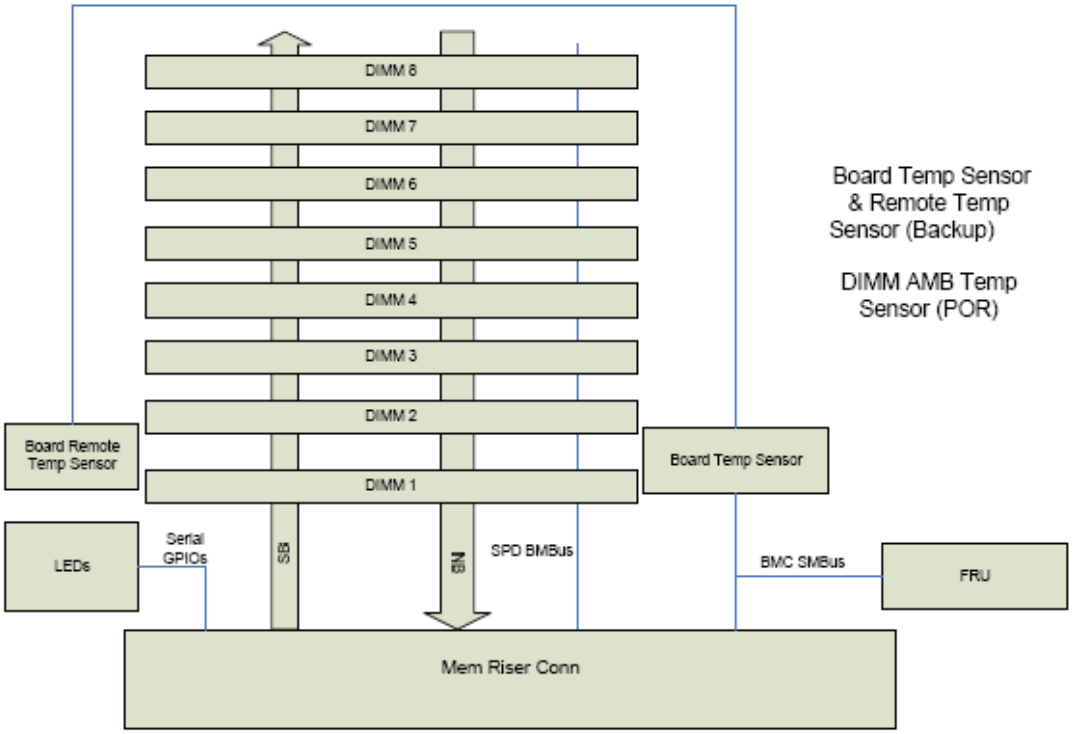


Figure 13. Memory Riser Functional Diagram

Supported Memory

The memory riser supports FBD Generation 1 DIMMs. FBD Generation 1 represents FBDDDR2-533 and FBD-DDR2-667 memory speeds. The maximum memory capacity supported per memory riser is 64GB, using eight sticks of 8-GB FBDIMMs (2-Gbit x4, stacked DRAM technology devices). The maximum memory capacity using four fully-loaded memory risers is 256 GB.

Notes:

1. 512Mb is the lowest technology point available from FBDIMM suppliers.
2. 2-Gb technology may not be available at the time of launch.

Table 12. Supported Fully-Buffered DIMMs

Technology	Organization	SDRAM chips/ DIMM	Capacity	Rank
512Mb	16M X 8 X 4bks	8	512MB	Single
	16M X 8 X 4bks	16	1GB	Dual
	32M X 4 X 4bks	16	1GB	Single
	32M X 4 X 4bks	32	2GB	Dual
1Gb	32M X 8 X 8bks	8	1GB	Single
	32M X 8 X 4bks	16	2GB	Dual
	64M X 4 X 8bks	16	2GB	Single
	64M X 4 X 8bks	32	4GB	Dual
2Gb	64M X 8 X 8bks	8	2GB	Single
	64M X 8 X 8bks	16	4GB	Dual
	128M X 4 X 8bks	16	4GB	Single
	128M X 4 X 8bks	32	8GB	Dual

Temperature Sensors, FRU, and SPD, BMC Bus

- Temperature Sensor: A two package temperature-sensing device provides a sensor at the left and right of the DIMM sockets. Server management sees this as one sensor, measuring the temperature drop across the board, which estimates the heat generated by the DIMMs.

- Field Replaceable Unit: An EEPROM device provides 256 bytes of programmableField-Replaceable Unit (FRU) space. This FRU is programmed during manufacturing to contain the board version and serial number but may also be programmed to meet other integrator-specific needs.

- Serial Presence Detect Bus: The Serial Presence Detect (SPD) bus is a low frequency I2C chain that is routed to each FBD memory channel. The Chipset acts as a master for the SPD bus.

- BMC Bus: The BMC Bus is a low frequency I2C bus that routed to FRU unit, and Optional Open Loop temperature sensor on the Memory Board.

Memory Riser LEDs

The following table describes all the LEDs on the memory riser.

Table 13. Memory Riser LED Descriptions

Name	Color	Description
FRU pwr good	Green	Memory Riser Power is good
01 LED	Amber	DIMM1 has had an error and needs to be replaced
02 LED	Amber	DIMM2 has had an error and needs to be replaced
03 LED	Amber	DIMM3 has had an error and needs to be replaced
04 LED	Amber	DIMM4 has had an error and needs to be replaced
05 LED	Amber	DIMM5 has had an error and needs to be replaced
06 LED	Amber	DIMM6 has had an error and needs to be replaced
07 LED	Amber	DIMM7 has had an error and needs to be replaced
08 LED	Amber	DIMM8 has had an error and needs to be replaced

Power Rails

The main board supplies 12V, 5V, 3.3V, and 3.3V_stby power to the memory riser. The Memory Riser has on-board regulators to generate 1.5V, 1.8V and 0.9V. The FBD AMB requires 1.5V, the FBD AMB and DDR2 DRAM require 1.8V, and the DDR2 termination requires 0.9V. The DIMM EEPROM and AMB SPD Bus require 3.3V. DIMM LEDs and control circuits require 3.3V_stby

I/O Riser

The I/O riser provides support for advanced server management with a dedicated maintenance Ethernet port and dual gigabit Ethernet ports on one vertical riser.

The advanced server management upgrade kit provides the Remote Management Module (Intel® RMM2) and a dedicated NIC port via the RMM2 NIC module. The RMM2 NIC module

contains the LAN PHY and an RJ45 jack for external access.

The Intel® RMM2 NIC provides an upgrade path to advanced server management capabilities. When the Intel RMM2 is plugged into the I/O riser, the original set of server management features continue to function and additional functionality is available. This functionality seamlessly integrates with respect to configuration functions and software support.

The Intel® RMM2 supports the keyboard, mouse, video redirect, and media redirect functionality. This enables the user to use a remote system to control the activity of the host server. See the ASMI external architecture specification for information about the Intel RMM2 and the RMM2 NIC.

The Intel® 82575 PCI Express*-based Ethernet controller provides advanced networking control and capability with dual-gigabit Ethernet ports. This controller hosts the Intel® I/O Acceleration Technology II (Intel® I/OAT2) capability that provides optimization of the TCP flow. The I/O riser provides an option to disable Gbit port A and / or port B in the BIOS. Server management traffic over these ports is not supported. For management traffic, use the main board LAN ports.

5.1 I/O Riser Features

- Intel® 82575 dual Gbit LAN controller with two dedicated Gbit ports
- PCI Express* interface (x4), 2.5 Gbps
- Supports sensor system management buses for field replaceable unit (FRU) and temperature sensor
- RMM2 NIC with 10 / 100 Mbs MII interface, with the Intel® LXT972 LAN controller for external manageability and access from the remote machine
- Intel® Remote Management Module 2 (Intel® RMM2) provides keyboard, video, mouse (KVM) redirect, and median redirect support for the server.
- Intel® RMM2 provides USB2.0 redirect support
- Intel RMM2 uses the IPMB bus interface to remotely shut down the host system through a median website.

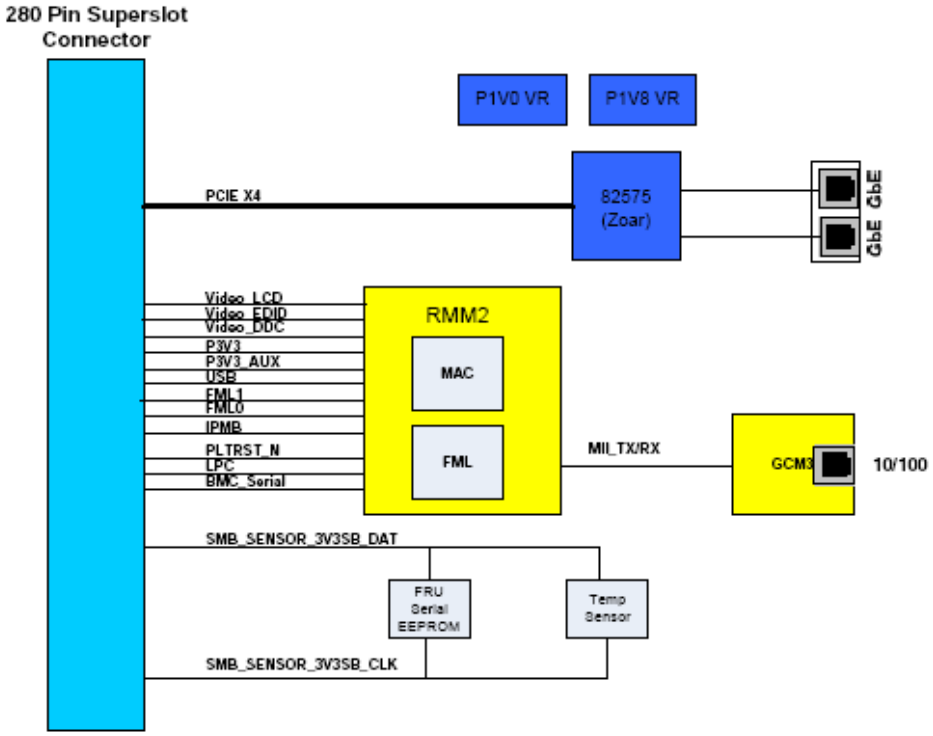


Figure 14. I/O Riser Block Diagram

SAS Riser

The SAS riser works in conjunction with the SAS backplane to give the end-user support foreight SAS hard drives in a 4U chassis. The SAS riser makes use of a dedicated PCI-Express* slot located in the front of the chassis to make cabling to the backplane convenient. The following block diagram, architectural overview, and placement diagram will provide a general idea of how the SAS riser works.

SAS Riser Features

- The SAS riser supports the following features:
- LSI1078 ROC (Raid-On-a-Chip) controller
- PCI-Express interface (x4), 2.5Gbps
- 8 channels of SAS/SATA at up to 3Gb/s
- SAS Integrated RAID (levels 0, 1, 1E)
- SAS HW RAID-on-Motherboard (ROMB) (levels 0, 1, 5, 6, 10, 50 and 60) through RAID key upgrade option.
- DDR2 Registered DIMM running at 667MHz for enhanced HW RAID performance
- Intel® RAID Smart Battery for DDR2 DIMM refresh support during power failure

- 8 MB flash component (Intel® TE28F640J3D-75) and a 32K non-volatile SRAM (STK14C88) store the code and hardware configuration information.
- SES (System Enclosure Spec) connectivity through I2C* cable or SGPIO
- UART and JTAG debug ports

Functional Architecture

Intel® SAS RAID-on-MotherBoard (ROMB)

The LSI1078* controller provides a SAS RAID-on-Motherboard (ROMB) solution which supports RAID levels 0, 1, 5, 6, 10, 50, and 60. An 8-MB flash component (Intel® TE28F640J3D-75) and a 32-K non-volatile SRAM (STK14C88) store the code and hardware configuration information.

Default RAID solution is IR (Integrated RAID) which supports RAID levels 0, 1, and 1E. IR RAID is activated through the LSI1078* F/W and does not utilize a DIMM or battery. To activate the optional HW RAID ROMB solution upgrade, a physical Intel® RAID Activation Key and DDR2 667MHz RAID DIMM must be placed in the SAS riser's right angle DIMM connector. The Intel® RAID Activation Key contains a registration code which is required to unlock the HW RAID engine in the LSI Mega RAID Controller. The MegaRAID solution supports RAID levels 0, 1, 5, 6, 10, 50, and 60. The DDR2 667MHz RAID DIMM serves as memory for the LSI1078* and as a disk cache to store data for the drives. In addition to these components, an Intel® RAID Smart Battery (iBBU) may also be installed to refresh the RAID DIMM when the system power drops below specifications.

After installing an Intel® RAID Activation Key and DDR2 667MHz RAID DIMM, and optional Intel® RAID Smart Battery, and upgrading to the HW RAID firmware, the system BIOS setup allows the user to enable the ROMB solution. During option ROM scan, an option to configure the RAID is displayed. The following three sections provide an overview of the Intel ROMB solution.

Intel® RAID Activation Key

The Intel® RAID Activation Key is a round one-wire serial EEPROM device programmed by LSI. This key has a registration code which is required to enable the LSI Mega RAID* solution. The RAID activation is an upgrade path from standard IR (Integrated RAID), which is the default RAID solution.

DDR2 RAID DIMM

The LSI1078* provides a DDR2 SDRAM interface, which significantly improves RAID system performance. The LSI1078* uses the high-speed DDR2 SDRAM interface to queue, organize, and track RAID accesses independent of the host processor. The host processor can continue to issue RAID writes for a given SAS Virtual Drive even when the SAS RAID channel is busy.

The LSI1078* queues the RAID write data in the DDR2 SDRAM until the connection is available, and then performs the write. The host does not need not wait for the SAS channel (only in Write Back mode). The ROMB solution only supports 667MHz registered ECC 240-pin DIMM. The LSI1078* supports single- or dual-ranked DIMMs but does not support quad-rank configurations. The SAS riser will only work with 667MHz DIMMs.

Intel® RAID Smart Battery

The Intel® RAID Smart Battery keeps the contents of the DDR2 667MHz RAID DIMM preserved if power drops below specifications. When the LSI1078* controller senses power has dropped below

specifications, it initiates a power fail sequence that safely puts the RAID DIMM into selfrefresh state. Standby power will be used to generate DIMM refresh power. In the absence of any power from the power distribution board, the NiMH battery cells will be used for DIMM refresh. The power subsystem generates enough of a delay to allow the LSI1078 to complete its power fail sequence, even in the event of total system power loss. After the power fail sequence is complete, additional logic keeps the RAID DIMM in self-refresh mode. When power is restored, data from the RAID DIMM is safely written to the disk array. The ideal data-retention time goal is 72 hours, but this may only be feasible using DIMMs with certain configurations.

FRU and SES

A 2K EEPROM (AT24C02) device provides 256 bytes of programmable Field-Replaceable Unit (FRU) space. Like all Intel® server boards, this FRU is programmed during manufacturing to contain the board version and serial number but may be programmed to meet integrator-specific needs.

The SAS riser will implement SES connectivity to the SAS backplane in several ways. The SFF8086 connectors will route the 2 SGPIO busses (SFF-8485) from the LSI1078* to the Vittesse* controller on the SAS backplane through the SAS cables. The SFF8086* cable headers will be strapped (SES signals: Contrl Type and Bckplne Type) to choose SGPIO over 2-wire method. In addition, a backup 3-pin header will allow cabling of dedicated SES SMB to the BP.

SMBus

The LSI 1078 is not configured as a slave within the boardset SMBus scheme. The SMBus segment from the system is connected to the SAS riser FRU through the connector. I2C0 port on the LSI1078 controls all SAS/ROMB related communication on the card. The bootstrap EEPROM, DDRII DIMM, IBBU, and SES cable connect to this port. A PCA9543 SMB mux is used to isolate the IBBU and SES cabling header in order to avoid periodic chatter from these interfaces which would interfere with power-on data retrieval from bootstrap EEPROM.

SAS Riser Power

The main board supplies 12V, 3.3V, and 3.3V_STBY power to the SAS riser. The memory riser board has on-board regulators to generate 1.5V and 1.8V. The LSI1078* requires 1.5V for core power and SAS interface. The DDRII interface (both LSI1078* and DDRII DIMM) requires 1.8V. A switching mechanism is used to change from main 1.8V (S0), Standby (1.8V Stby through iBBU)(S5), and 1.8V BAT (iBBU in AC OFF). To ensure correct Reset/PWRGD logic out of PWR FAIL mode, LSI* recommends generating 3.3V from 12V for proper sequencing in select logic.

SAS Backplane

The SAS backplane supports eight SAS hard drives in a ProServ 4680 Server System chassis. Without powering down the system, this design enables the following:

- Easy use of the SAS hard drives
- Easy replacement of the SAS hard drives

Figure 15 illustrates the general architecture for the SAS backplane.

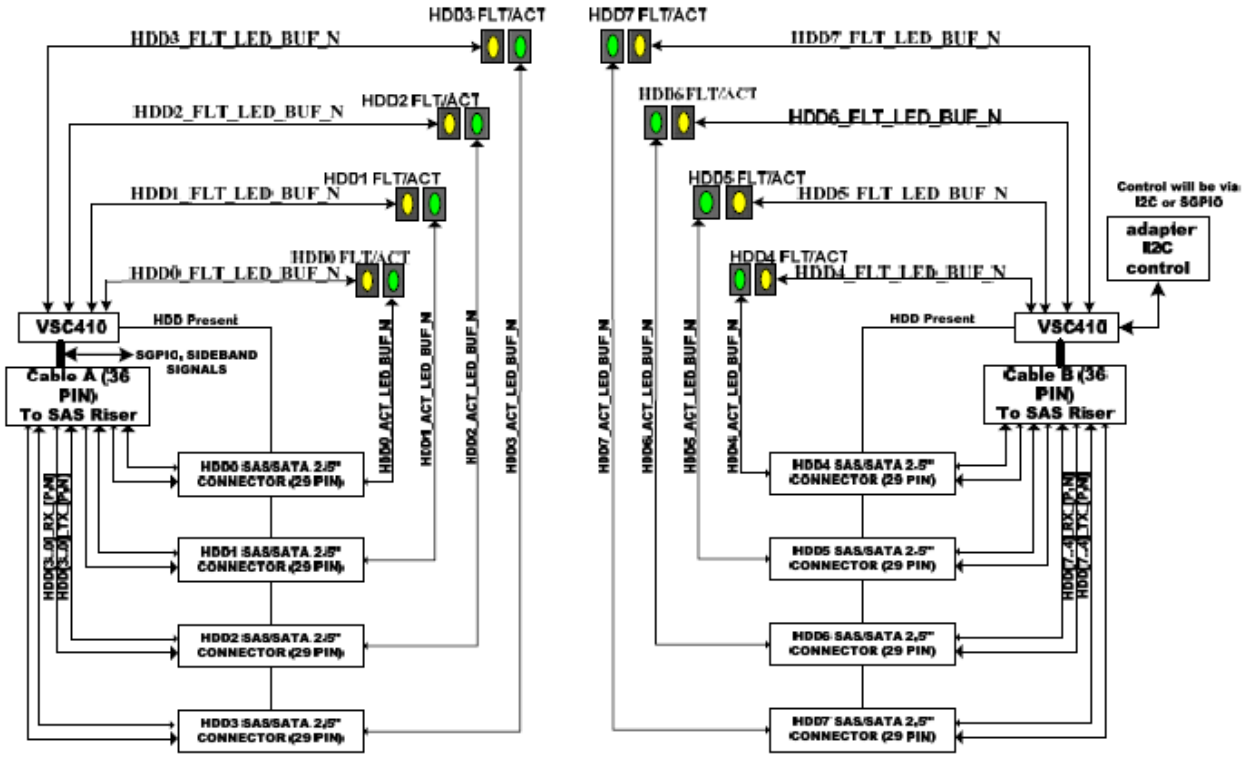


Figure 15. SAS Backplane System Block Diagram

Architectural Overview

The SAS backplane provides several main functions for the system.

3Gbit SAS Port Expanders

3Gbit SAS port expanders provide high-speed serial data paths. The paths are from the eight attached SAS hard drives to the server board.

SAS Data

SAS data between drives and server board are routed across two 4-port internal SAS cables. The SAS cables connect the SAS backplane to the PCI-Express* SAS Redundant Array of Independent Disks (RAID) controller card. In addition, the SAS controller card is plugged into the server board at one of the PCI-Express* slots.

SAS Controller

The eight drives are connected directly to the SAS controller. The SAS controller is used to control SAS traffic flow between drives and the SAS RAID controller card. A Vitesse* VSC410

module communicates the presence and fault signals via a SES-2 interface (I2C* cable) or via a Serial General Purpose Input/Output (SGPIO) interface thru the SAS cables.

Enclosure Management

The enclosure management is comprised of the following:

- SAS enclosure management per SES-2 or via SGPIO interface
- Fault and presence Light Emitting Diode (LED) control logic

Server Management I2C* Interface

The server management I2C* Interface is comprised of the following:

- Fan presence sensing and fan fault LED control
- I2C* serial EEPROM Field Replaceable Unit (FRU)
- Temperature sensor
- Expander controller firmware update capability

Voltage Regulators

The voltage regulators are as follows:

- 12VDC to 5VDC
- 5VDC to 3.3VDC

System Fan Control

The two fan assemblies are controlled by system fan control.

Power Functions

The last main functions for the system provided by the SAS backplane are as follows:

- Power connector for tape and DVD/CD Drive
- Power

Functional Architecture

This section provides a more detailed architectural description of the SAS backplane's functional blocks.

SAS Buses

The SAS buses are directly connected to the server board via the SAS RAID controller card that is plugged into a PCI-Express* slot on the server board. As a result, the SAS RAID controller provides all SAS functionality and interfacing to the SAS backplane.

SAS Backplane

The SAS backplane routes data to/from each of the eight internal SAS drives from/to the SAS controller on the adapter card. Data movement between the SAS drives on the SAS backplane and the SAS controller is achieved through two high-speed SAS cables. These cables connect the PCI-Express* card to the SAS backplane.

There are a total of eight separate SAS buses or lanes. These buses are contained within the two high-

speed cables. A Molex* SFF 8086 mini connector (or Molex* SFF 8484 x SFF 8086*) is used to terminate each end of the cable assembly to the SAS adapter card and the SAS backplane.

Full-duplex Serial Mode Operation

Each SAS lane operates in full-duplex serial mode. In addition, each lane contains dedicated transmit and receive differential pairs. The combination creates a total of four differential pairs on each of the two cables that routes data directly to the eight SAS drives that are attached to its ports.

SAS Controller

The SAS controller consists of eight identical SAS ports, two of which are connected to the Molex* SFF 8086 mini connector. The mini connector routes the data to/from the SAS controller card. All eight ports are used to connect directly to the eight hot-pluggable SAS (or SATA) drives with each drive having a dedicated port. All ports are used.

All SAS channels on the SAS backplane are capable of 3Gbps data transmission on either the transmition path or the reception path. During system power-on and hot drive insertion, SAS rates are negotiated to operate at either 1.5Gbps for SATA mode or 3.0Gbps for SAS mode.

Vitesse* VSC410 Controller Functionality

The Vitesse* VSC410 is a storage management controller with SCSI Accessed Fault-Tolerant Enclosure (SAF-TE) and SCSI enclosure services (SES). The figure below shows the Vitesse* VSC410 internal logic and external interfaces. Note that you need a SES I2C* cable which plugs into the SAS backplane and the SAS adapter, in order to use the SES functions. This communication can also be accomplished via the SGPIO interfaces.

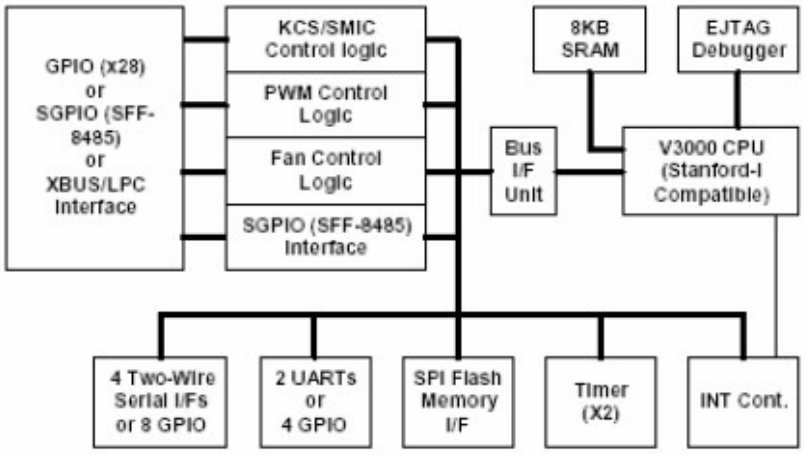


Figure 16. VSC410* Block Diagram

SPI Flash

Firmware for the Vitesse* VSC410 storage management controller is stored in an 8-Megabit (Mb) SPI Flash memory device. Each Flash device can be updated via the Intelligent Platform Management Bus (IPMB) bus.

SAS Drive Functionality
The SAS backplane provides connections for a maximum of eight SAS drives. Each drive can be inserted and removed while the system is powered-on and automatic detection and rate negotiation are performed after each insertion. The SAS backplane provides +5V and +12V to each drive connector and supports in-rush current limiting to 300mA during hot swapping.

Power Control Interlock
The power control interlock is part of the SAS specification. This prevents drives from powering on at the same time. Since only one drive can power on at once the board power requirements can be kept lower.

System Status Notification
Internal SAS drive status information is collected by the Vitesse* VSC410 storage management controller. The information can be monitored by accessing the VSC410's serial port. Output drive strength and input pre-emphasis may also be controlled via the serial port. In addition, any drive data can be routed to the server management via the IPMB.

SAS Status LEDs
The status LEDs gives the user a visual indication of the drives' condition. There is a single green LED (activity - left) and a single amber LED (fault - right) for each drive. The LEDs use a combination of color and blinking frequency to indicate multiple conditions.
The hard drive status LEDs are located on the SAS backplane and projected out the front of the carrier via light pipes. The states of the LEDs are described in Table 14.

Table 14. SAS Hard Drive LED Details

LED State	Description
Green On	The hard drive is configured and ready for access.
Green Blinking	The hard drive is active.
Amber On	Hard drive/slot failure.
Amber Slow Blinking (~1Hz)	A predictive hard drive/slot failure or rebuild in process
Amber Fast Blinking (~2.5Hz)	Hard drive rebuild interrupted, rebuild on empty slot, or identify slot

SAS Enclosure Management
SAS enclosure management allows the SAS backplane to report SAS drive status and backplane temperature readings. A SAS RAID controller will interface with the enclosure management. The SAS enclosure management subsystem consists of one VSC storage management controller, and the associated serial peripheral interface (SPI) Flash and electrically erasable programmable read-only memory (EEPROM) memory devices.

Server Management Interface
The SAS backplane will support the following server management features:

- Two SGPIO Interfaces
- Hot-swap controller (HSC) Secure Digital Input/Output (SDIO) Interface
- UART Serial Interface
- Local I2C* Interface
- System I2C* Interface

- Local I2C* Bus
- Isolated Global I2C* Bus IPMB

Two SGPIO Interfaces
There are two SGPIO interfaces for cable A (drives 0-3) and cable B (drives 4-7). The interfaces communicate the fault LED and present information.

HSC SDIO Interface
This interface communicates with the M25P80 flash module to access the board firmware.

UART Serial Interface
There is one serial port interface.

Local I2C* Interface
The local I2C* interface is as follows:
- SAS backplane FRU
- SAS backplane temperature sensor
- One temperature sensor is attached to the local I2C* bus of each of the two expanders.
- Micro-controller interface

Local I2C* Bus
The bus A local I2C* bus connects the DS75* thermal sensor and Atmel* AT24C64N (or equivalent) serial EEPROM (with FRU data) to the Vitesse* VSC410 7.3.9.6 Isolated Global I2C* Bus (IPMB) The global I2C* bus connects the on-board SAS expanders to the system. The IPMB bus also goes thru the front panel control connector for use with a liquid crystal display (LCD).

I2C* I/O Bus
The SMB_SYS_PWR_SCL/SDA bus connects the system server management controller to the PCA9554* device used for fan sensing and LED control. The bus address is listed in Table 17.

I2C* Addresses
Two I2C* devices and their addresses are listed in Table 15 and one in Table 16.

Table 15. I²C* Local Bus Addresses

Device	Address	Bus	Description
AT24C64*	0xAC	VSC local bus	Private SAS backplane FRU EEPROM
DS75*	0x90	VSC local bus	Private SAS backplane temperature sensor

Table 16. Global I²C* bus Addresses (IPM Bus)

Device	Address	Bus	Description
VSC410*	NA	IPMB system interface	VSC410* controller public IPMB bus

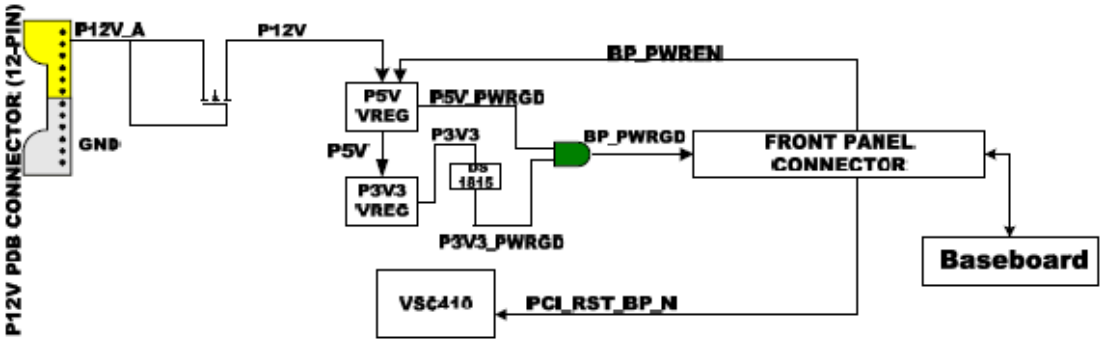
Table 17. I²C* I/O Bus Addresses

Device	Address	Bus	Description
PCA9554*	0x42	SMB_SYS_PWR_SDA/SCL	Micro controller public I/O bus

Resets

The principal reset for logic on the SAS backplane is supplied by the PCI_RST_BP_N signal from the server board via the 100-pin connector.

The PCA9554* device being used to control the fans, has an internal power-on reset that configures all its I/O pins as inputs.
See the diagram below for reset flow.



Connector Interlocks

In the sections below, the connector interlocks are described.

Server Board Cable Connector

The SAS backplane has an interlock on the 100-pin connector. This allows the server board to detect its presence.

Clock Generation

The SAS backplane requires one 4MHz crystal for the VSC410* controller.

Programmed Devices

There are two programmed devices on the SAS backplane.

Flash Memory

The Flash memory device contains program code. The code is run by the VSC410* controller.

- Memory configuration: 64Mb SPI

Field Replaceable Unit (FRU)

The FRU is a serial EEPROM programmed at automated test equipment (ATE).

- Memory Configuration: 64Kb serial

System Overview

The ProServ 4680 is a 4U, high-density, rack-mount server system with support for one to four processors and up to 256GB FBDIMM memory. Figure 18 and Figure 19 show the front and rear views of the system.

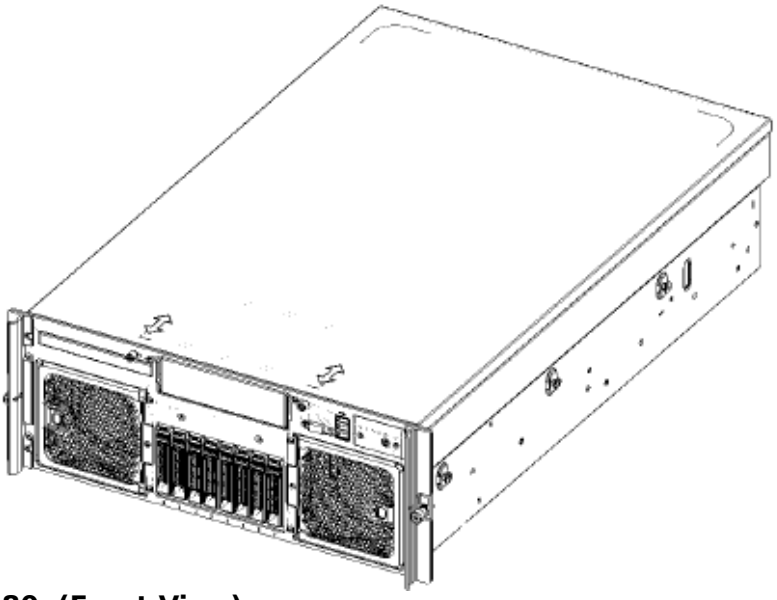


Figure 18. ProServ 4680 (Front View)

The ProServ 4680 is based on the Intel® 7300 Chipset. The main board attaches to a sheet metal tray for structural support. The tray and board assembly is installed along the bottom of the chassis. The memory boards plug vertically to the main board and contain eight FBDIMM slots each. With all four memory boards installed, the system supports up to 256GB of memory (using 8GB DIMMs).

The hard drive bay located at the front of system holds eight 2.5-inch SAS hard disk drives. The SAS hard drives plug into a horizontal SAS backplane at the rear of hard drive bay. One slimline (½-inch) optical drive bay and a half-height 5.25-inch tape backup bay are also located at the front of the system. The cooling subsystem utilizes eight hot-swap system fans. All 8 fans are required for operation in a redundant cooling configuration. In a non-redundant cooling configuration 2 of the rear fans are

removed. Each system fan contains a status LED, illuminating in the event of a fan failure. The fans are accessible from the front and the inside of the system. The front panel provides a user interface for system monitoring and management via buttons and status LEDs. The optional front bezel has snap-on features for ease of installation. It can be customized to meet Original Equipment Manufacturer (OEM) industrial design requirements. The bezel design allows adequate airflow to cool system components

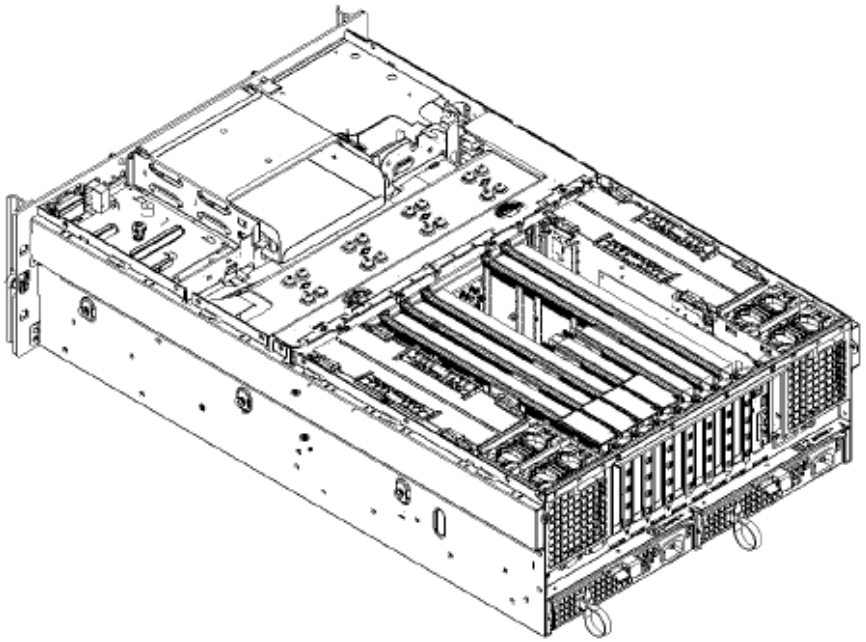


Figure 19. ProServ 4680 Server System (rear view with top cover removed)

The power supply modules are located at the rear of system under the main board. The modules plug directly into connectors on the power distribution board. The system supports up to two hot-swap power supply modules in a 1+1 redundant configuration.

Upon removal of the top cover, the user has access to the memory boards, memory fans and PCI-Express* adapters.

8.1 External Chassis Features – Front

Figure 20 shows the front view of the system with the front bezel removed. The front provides access to the following components:

- Front panel buttons and LEDs (with optional LCD)
- User accessible video and USB connectors
- Hard drive, optical drive, and tape backup drive
- Hot-swap fans

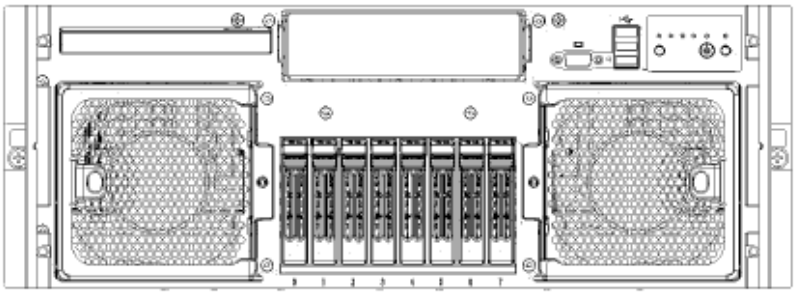


Figure 20. Front View (without bezel)

Front Panel

The front panel contains the following:

- System control buttons and LED status indicators
- One video connector
- Three USB 2.0 ports
- One system speaker

The front bezel must be removed to access the front panel switches and connectors. All LEDs are visible with the front bezel installed. Refer to Section 12 for a detailed description of the front panel boards.

Hard Drive and Peripheral Device Bays

The hard drive and peripheral device bays can accommodate the following devices:

- Eight hot-swap hard disk drives
- One ½-inch optical drive
- One half-height 5.25-inch tape backup device

The SAS backplane supports both 2.5-inch SATA and SAS drives.

The optical drive and tape backup drive are not hot-swap devices. System power must be turned off when installing or removing these drives.

Hard disk drives have different cooling, power, and vibration characteristics; therefore, Freedom Technologies will validate specific hard disk drive types in the ProServ 4680 Server System. Refer to the Tested Hardware and Operating System List for of the qualified drives.

The hard drive carriers supplied with the system accommodate only 2.5-inch disk drives. The hard drive is attached to the carrier with four Phillips* head screws. The carrier is retained in the chassis by

a locking handle.

The SAS backplane contains two LEDs for each hard drive to display status. The LED signal is transmitted to the front of the system via a light pipe integrated in the hard drive carrier.

The signals for the optical drive are carried from the main board via a signal lane SATA cable. There is a SATA-to-PATA converter board. The converter board plugs into the optical drive.

External Chassis Features – Rear

Figure 21 shows the rear view of the system. The user-accessible connectors, PCI-Express* slots, and power supply modules are located at the rear of the system. They are described in the following sections.

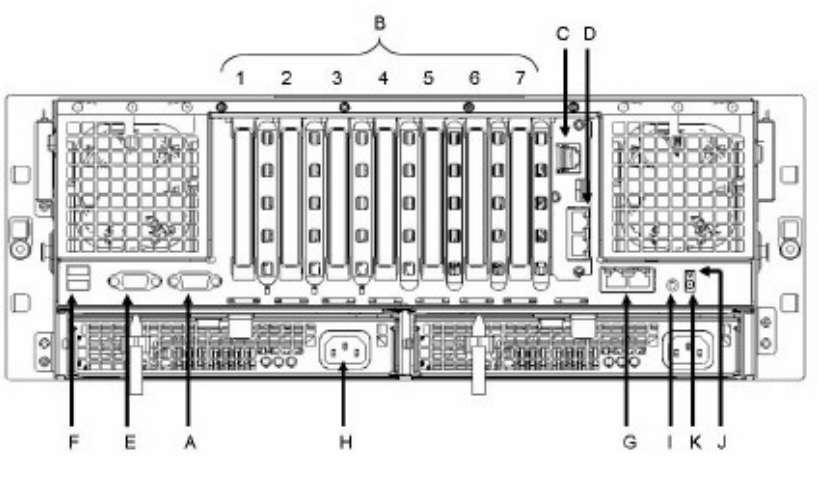


Figure 21. Rear View

8.2.1 User-Accessible Connectors, PCI Slots and LEDs

Table 18. lists the user-accessible connectors at the rear of the system.

Table 18. User-Accessible Connectors, PCI Slots, and LEDs

Item	Description	
A	Serial port connector	
B	PCI-Express* Slots	
	Slot 1	PCI-Express* x8 (hot-plug)
	Slot 2	PCI-Express* x8 (hot-plug)
	Slot 3	PCI-Express* x8
	Slot 4	PCI-Express* x8
	Slot 5	PCI-Express* x4
	Slot 6	PCI-Express* x4
	Slot 7	PCI-Express* x4
C	I/O Riser Management Ethernet Port	
D	I/O Riser Dual Gigabit Ethernet Ports	
E	Video port, standard VGA compatible, 15-pin connector	
F	Two USB 2.0 ports (USB1 on top, USB2 on bottom)	
G	Two LAN ports, RJ45 connector (LAN1 on left, LAN2 on right) LAN port LEDs:	
	Status LED (Green)	On – ethernet link is detected Off – no ethernet connection Blinking – ethernet link is active
	Speed LED (Green/Amber dual color)	Off – 10 Mbps Green On – 100 Mbps Amber On – 1000 Mbps
H	AC input power connectors	

Item	Description
I	System ID button
J	DC Jack (not used)
K	System ID LED (blue)

Power Distribution Board (PDB)

The power distribution board is located below the main board in the chassis. It has two connectors for hot-swap power supply modules. It also routes 12V and standby power and signals to the main board and SAS backplane. Refer to Section 11 for a detailed description of the PDB.

Front Panel I/O Board

The front panel I/O board communicates with the main board via a cable with a 100-pin connector. The board contains the following:

- Video connector
- Three USB ports
- NMI button

Refer to Section 12 for a detailed description of this board.

Front Panel Control Board

The front panel control board connects to the front panel I/O board via a cable. It houses the buttons and LEDs described in Section 12.

SATA-to-PATA Converter Board

The SATA-to-PATA converter board receives the SATA signal from the main board via a x1 SATA cable and converts it to IDE signals routed to the optical drive.

Intel® Remote Management Module 2 (Intel® RMM2)

The Intel® RMM2 contains the remote server management support. The module plugs into a connector

provided on the I/O riser board.

Power Subsystem

The power subsystem supports up to two power supplies. The hot-swap power supply modules are rated at 1570W over an input range of 200-240 VAC.
The total power requirement exceeds the 240 VA energy hazard limit that defines an operatoraccessible area. As a result, only qualified technical personnel should access the processor, memory, and non-hot-swap areas while the system is energized.
The power subsystem can be configured as follows:

- With two power supply modules installed, a fully configured system has (1+1) power redundancy
- With one power supply module installed, the system does not have redundant power.

At 200-240VAC input, one power supply module is capable of handling the maximum power requirements for a fully configured ProServ 4680 Server System, which includes the following:

- Four processors
- 256GB of memory
- Seven PCI-Express* add-in cards
- Eight hard disk drives
- One optical drive
- One tape drive

When the system is configured with two power supply modules, the hot-swap feature allows the user to replace a failed power supply module without affecting the system functionality.

The power subsystem receives AC power through two power cords. When two power supply modules and two power cords are installed, the system supports (1+1) power cord redundancy. This feature allows the system to be powered by two separate AC sources. In this configuration, the system will continue to function without interruption if one of the AC sources fails.

A 3 volt lithium battery provides power to the RTC when the Main Board is powered down. The expected battery life is greater than 5 years.

Cooling Subsystem

The server system contains three cooling fan zones comprising a total of eight system fans. Four fans are located in the front and four in the rear of the chassis. The zones are designed to be redundant in order to system coo maintain ling in the event of fan failure. To maintain system performance, only one of the eight fans can fail at any one time.

Note: *The cooling system is non redundant in a non redundant power supply system configuration.*

Each fan assembly has a single LED to indicate its status. In the event of a fan failure, the LED will illuminate amber. Failed front fans can be hot-swapped out the front of the chassis. Failed

rear fans can be hot-swapped from the inside of the chassis with the cover removed. The maximum time limit to perform a fan hot-swap operation is two minutes before impacting system performance.

For systems not configured with four processors, the processor heat sink fillers must be installed to maintain proper cooling.

Specifications

Environmental Specifications

The production system will be tested to the environmental specifications as indicated in Table 19.

Table 19. Environmental Specifications Summary

Environment	Specification
Temperature operating	10°C to 35°C (50°F to 95°F)
Temperature non-operating	-40°C to 70°C (-40°F to 158°F)
Altitude	-30 to 1,500 m (-100 to 5,000 ft)
Humidity non-operating	95%, non-condensing at temperatures of 25°C (77°F) to 30°C (86°F)
Vibration non-operating	2.2 Grms, 10 minutes per axis on each of the three axes
Shock operating	Half-sine 2 G, 11 ms pulse, 100 pulses in each direction, on each of the three axes
Shock non-operating	Trapezoidal, 25 G, two drops on each of six faces V : 175 inches/sec on bottom face drop, 90 inches/sec on other 5 faces
Safety	UL60 950, CSA60 950, AS/NZS 3562, GB4943-1995, EN60 950 and 73/23/EEC, IEC 60 950, EMKO-TSE (74-SEC) 207/94, GOST-R 50377-92
Emissions	Certified to FCC Class A; tested to CISPR 22 Class A, EN 55022 Class A and 89/336/EEC, VCCI Class A, AS/NZS 3548 Class A, ICES-003 Class A, GB9254-1998, MIC Notice 1997-42 Class A, GOST-R 29216-91 Class A, BSMI CNS13438
Immunity	Verified to comply with EN55024, CISPR 24, GB9254-1998, MIC Notice 1997-41, GOST-R 50628-95
Electrostatic discharge	Tested to ESD levels up to 15 kilovolts (kV) air discharge and up to 8 kV contact discharge without physical damage
Acoustic	▪ Sound power: < 7.0 BA at ambient temperature < 23° C measured using the Dome Method ▪ GOST MSanPiN 001-96

Physical Specifications

Table 20 describes the physical specifications of the system.

Table 20. Physical Specifications

Specification	Value
Height	6.8 inches (173 mm)
Width	17.6 inches (447 mm)
Depth	27.8 inches (706 mm)
Front clearance	3 inches (76 mm)
Side clearance	1 inch (25 mm)
Rear clearance	6 inches (152 mm)
Weight[1]	90 lbs (40 kg)

Note: 1. The system weight listed above is an estimate for a fully configured system and will vary depending on number of peripheral devices and add-in cards, as well as the number of processors and DIMMs installed in the system.

System Chassis and Sub-Assemblies

This section describes the system chassis and its sub-assemblies.

Base Chassis and Top Covers

Base Chassis

The system fits into a standard 19-inch EIA rack and is 4U high x 28-inches deep. The 4U height is defined by standard EIA rack units where 1U = 1.75-inches. The depth, as measured from the front mounting flange to the back of the PCI slots, does not include cables or a bezel.

The chassis is modular for ease of serviceability and manufacturability. All major modules in the chassis are easily accessible. Hot-swap component replacement capability is provided for:

- System fans
- Hard drives
- Memory riser boards
- PCI slots
- Power supplies

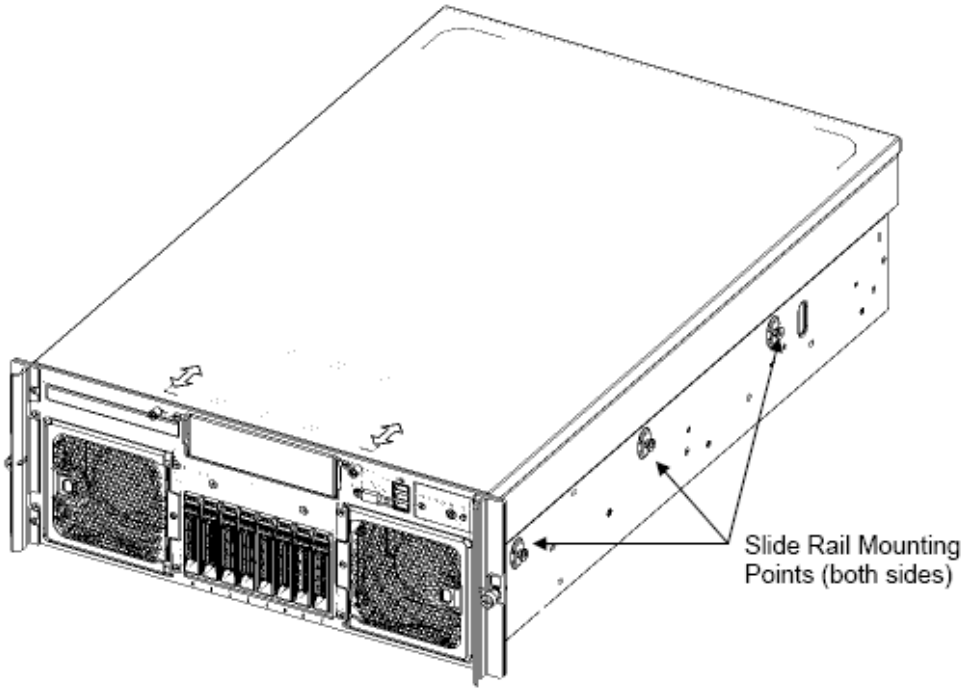
All system FRUs can be replaced without any loose hardware.

Top Cover

The top cover is a single-piece design. It attaches to the chassis with a series of slot features in the sides of the chassis that mate with features in the top cover.

Slide Rails

The server chassis accommodates slide rails for mounting the chassis into standard 19-inch racks. The keyhole features on the slide rails attach to studs on the sides of the chassis. No tools or screws are needed.



Power and Fan Subsystems

Power Subsystem

The power bay provides space for two power supply modules and the power distribution board (PDB). The dimensions of the power supply are 7.75-inches (W) x 14.5-inches (D) x 1.47-inches (H).

The PDB distributes the power in two ways. There is a connector on the back edge of the board that mates to the power supplies. In addition, there are cables that route power up to the main board and to the SAS backplane. The AC power is filtered with a combination 15A power plug integrated with a filter.

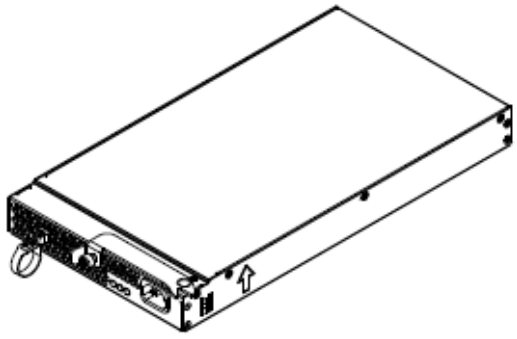


Figure 23. Power Supply Module

Fan Subsystem

Two fan assemblies are located at the front of the chassis and are removed from the front (See Figure 9). Each assembly contains two fans. The fans are in a sheetmetal enclosure with a plastic bezel mounted to the front. The assembly has an integrated amber LED wired to the front of the plastic bezel. The LED will turn on when either fan is not functioning within specifications. The fan connector extends from the rear of the fan assembly.

Four additional fans are located at the rear of the chassis and are removed from the top (See Figure 24). The fans are assembled into a plastic enclosure. The fan has an integrated amber LED that will turn on when the fan is not functioning within specifications. The fan connector extends from the bottom of the fan assembly.

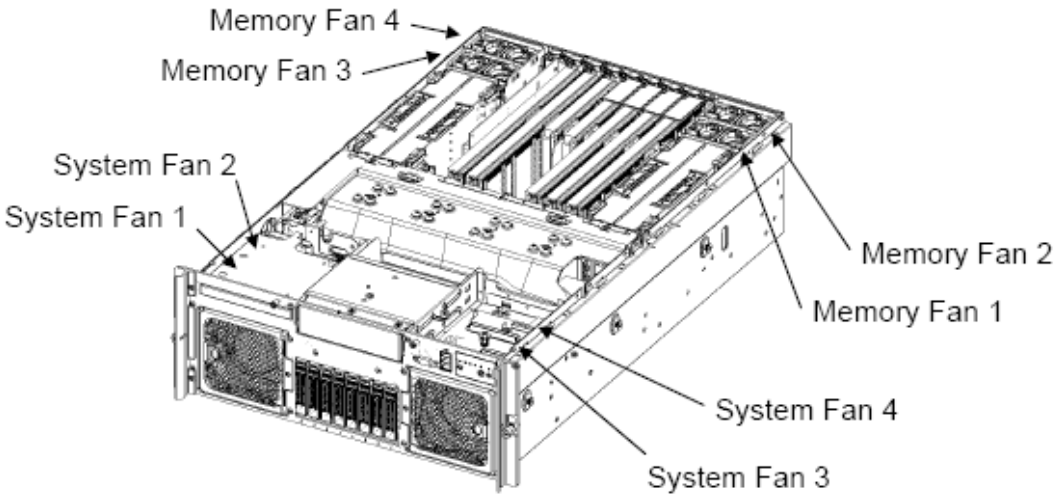


Figure 24. Fan Locations

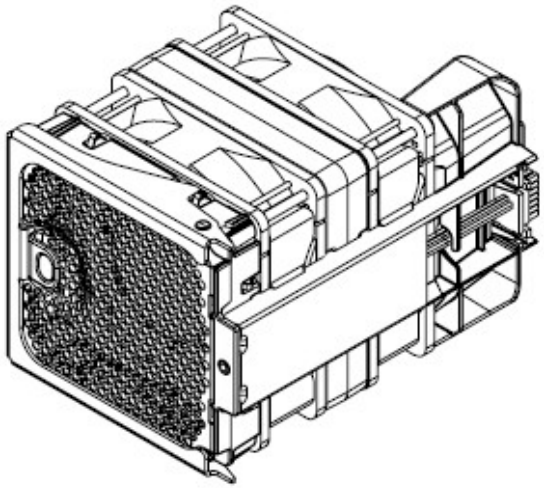


Figure 25. Front Fan Assembly

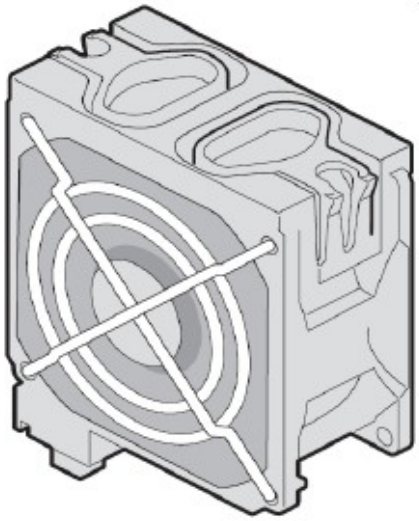


Figure 26. Rear Fan

Main Board Subsystem

The main board mounts to a sheet metal tray with four metal brackets from the Component Enabling Kit (CEK) and four loose screws. The main board assembly is mounted in the chassis via slot and tab hooks and secured by a single captive plunger.

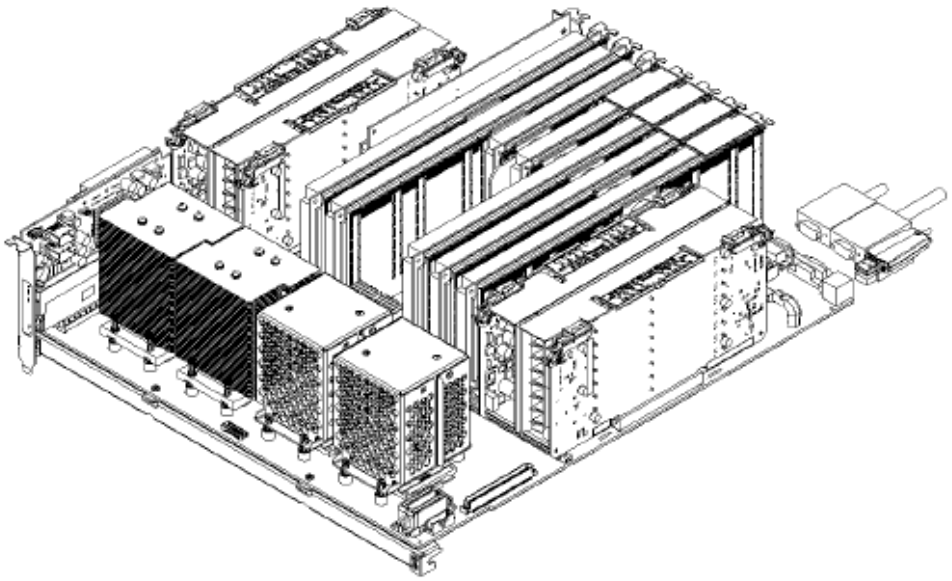


Figure 27. Main Board and Sheetmetal Tray

Peripheral Bay Subsystem

The peripheral bay is a sheet metal enclosure with features to mount the hard drives, half height 5.25-inch tape drive, and optical drive. The SAS backplane is installed by sliding the slots on the board onto hooks on the peripheral bay. One integral latch is used to secure the SAS backplane into the bay.

9.4.1 Hard Drive Carrier

The hard drive carrier is an assembly that provides guidance for hot swapping. It contains two integrated light pipes to transfer the LED indicator light from the SAS backplane to the front, and an insertion/extraction mechanism that includes a hard drive bezel.

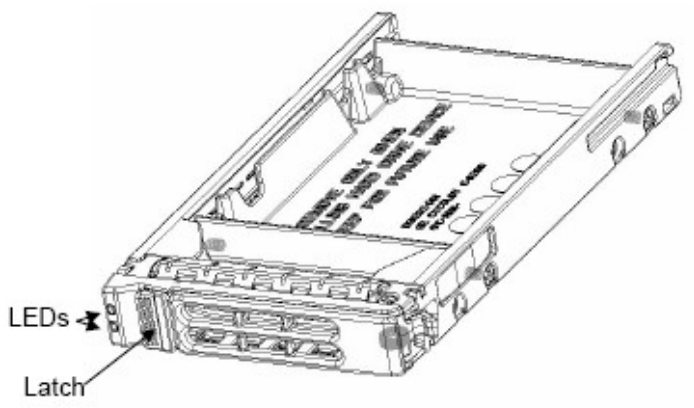


Figure 28. Hard Drive Carrier

Optical Drive Carrier
The optical drive is installed in a sheet metal tray and then installed in the chassis. The SATA-to-PATA converter board is then plugged in.

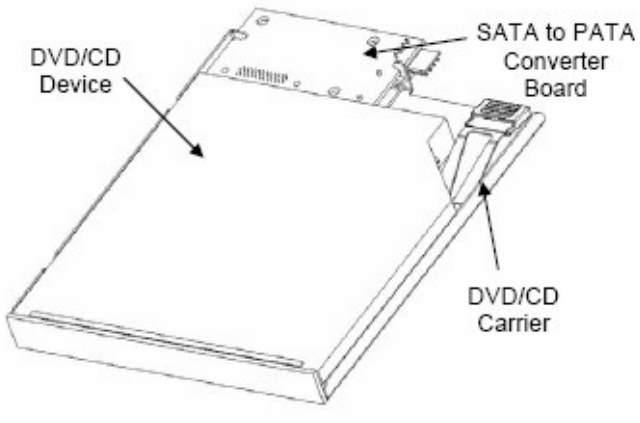


Figure 29. Optical Drive Carrier with Converter Board

Front Bezel
The front bezel assembly is a single-piece design that attaches to features on the front of the chassis and covers the hard drives, peripheral device, and front panel buttons/connectors. The front panel LEDs are visible through the bezel.

There are two black plastic handles. These handles cover the EIA mounting flanges that are used to pull the chassis from a rack.

1570W Power Supply
This section describes some of the power supply features. It is a current sharing power supply

with auto ranging input. The dimensions of the power supply are 7.75-inches (W) x 14.5-inches (D) x 1.47-inches (H).
(DI)
The output rating of the power supply is 1570W when operated between 200VAC and 240VAC. The system can run with only one power supply installed. For redundancy, two power supplies must be installed.

AC Input Requirement

AC Input Voltage Specification
The power supply will operate over the range and limits shown in Table 21.

Table 21. AC Input Rating

Parameter	Minimum	Nominal	Maximum	Unit
Voltage (115)	90	100-127	140	VAC_{rms}
Voltage (220)	180	200-240	264	VAC_{rms}
Frequency	47	50/60	63	Hz
I_{in} (90VAC)			14.4	A_{rms}
I_{in} (103.5VAC)			14.4	A_{rms}
I_{in} (180VAC)			11.1	A_{rms}
Vin (turn-on)	81		89	VAC_{rms}
Vin (turn-off)	70		80	VAC_{rms}

The main outputs of the power supply will turn off per VIN (turn-off). Any standby outputs may continue to operate at input AC voltages below VIN (turn-off).

Efficiency
The power supply will have a minimum efficiency of 80% when operated under the maximum loading conditions at 90VAC-240VAC.

Input Over-Current Protection
The power supply has internal primary over-current protection. A normal-blow (fast blow), highbreaking-capacity fuse is placed in the input circuit.

Inrush Current
When input power is applied to the power supply, any initial current surge or spike of 10ms or less will not exceed 55A peak. Any additional inrush current surges or spikes in the form of AC cycles or multiple AC cycles greater than 10ms, and less than 150ms, will not exceed 25A peak. For any conditions during turn-on, the inrush current will not open the primary input fuse or damage any other components.

Auto Restart
Although the power supply may power off under the conditions mentioned it is capable of restarting, either automatically or under program control after the disturbance. In addition, the power supply will not be in a latched state such that any of the operator buttons/buttons do not operate correctly after the disturbance. At no time will the AC power cord have to be removed to clear an error condition. Auto restart conditions are tested from -40% to -100% AC under-voltage conditions for time intervals ranging from 25ms to 2sec. For each time interval, all of the under-voltage conditions

listed below will be tested. These tests are performed at both the lowest and highest nominal operating voltages of the power supply.

- Time intervals: 25ms, 40ms, 60ms, 90ms, 130ms, 200ms, 280ms, 400ms, 600ms, 900ms, 1.3sec, and 2.0sec
- Under-voltage deviation from nominal AC voltage: -40%, -50%, -60%, -70%, -80%, -90%, -100%

Power Factor Correction (PFC)
The power factor is greater than 0.99 at 100VAC to 127VAC input voltages.

AC Input Connector
The AC input receptacle is an IEC-320* C14 15A rated for 250VAC minimum.

DC Output Requirements
The DC output voltages will remain within the regulation ranges shown in the following table when measured at the load end of the connector.

Table 22. DC Output Voltage Regulation Limits

Output Level	Minimum (V)	Nominal (V)	Maximum (V)
+12V	11.40	12.00	12.60
+3.3V standby	3.20	3.30	3.46

Hot Swap Functionality
Hot swapping is the process of inserting and extracting a power supply from an operating power bay. During this process, the output voltages will remain within the limits specified in Table 22, and the system will continue to operate normally.

Output Current Rating
The combined continuous output power for all outputs will not exceed 1570W. Each output has a maximum and minimum current rating shown in Table 23.

Table 23. 1570W Load Ratings

Output Level	Minimum[1]	Nominal[1]	Maximum[1]	Peak[1]
+12V	0A		130.8A	144.0A
+3.3V standby	0.1A		6.0A	

Note: 1. Values are at the system level. For 1+1 redundant systems, the load each power supply will provide will be based on its current sharing accuracy.

Over- and Under-Voltage Protection
The power supply will provide latch mode over and under voltage protection as defined in the following table. A fault on any output will cause the rest of the outputs to latch off. (In addition, see note 3 in the following table.)

Table 24. Over- and Under-Voltage Limits

Level	Under-Voltage		Over Voltage	
Output Level	Minimum	Maximum	Minimum	Maximum
+3.3V standby [1,2,3]	2.77V	3.00V	3.76V	4.3V
+12V	10.5	11.0	13.5	15.0

Notes:

1. In standby mode, the power supply will not latch off due to an under-voltage condition.
2. In standby mode, the power supply may or may not latch off due to an over-voltage condition.
3. A fault on any output other than +3.3V standby will not cause the +3.3V standby to turn off. A fault on +3.3V standby will cause the other outputs to turn off.

OverCurrentProtection

Over-current is a fault condition defined as a 10A/s current ramp starting from full load applied to the output under test. A fault on any output will cause the rest of the outputs to latch off. (See note 4.)

Table 25. Over-Current Protection Limits

Output Level	Input Voltage	Minimum[2,4]	Maximum[2,4]
+3.3V standby[1,3,4,]		7 Amps	10 Amps
+12V	90 – 140 VAC	115 Amps	132 Amps
+12V	180 – 264 VAC	151 Amps	173 Amps

Notes:

1. Output is Level III SELV and non-energy hazard complaint
2. The above current limits will be satisfied throughout the entire operating temperature range
3. A fault on any output other than +3.3V standby will not cause the +3.3V standby to turn off. A fault on +3.3V standby will cause the other outputs to turn off.
4. Dynamic loading must not cause a false over current when two supplies are in parallel.
5. The +3.3V standby output will not latch off. It must return to normal operation once the fault is removed. Current foldback method is preferred.

Short Circuit Protection

A short circuit, which is defined as an impedance of 0.1 ohms or less, applied to any output during start-up or while running will not cause any damage to the power supply (connectors, components, PCB traces, etcetera).

When the +3.3VSB is shorted the output may go into “hiccup mode.” When the +3.3VSB attempts to restart, the maximum peak current from the output must be less than 8.0A. The maximum average current, taking into account the “hiccup” duty cycle, must be less than 4.0A.

Reset After Shutdown

If the power supply latches into a shutdown state due to a fault condition on any output, the power supply will return to normal operation only after the fault has been removed and the power supply has been power-cycled. Power cycling is defined as either:

- Removing AC input power, waiting for +3.3V standby to drop below 1.0V, then reapplying AC power. (The time it takes for +3.3V standby to drop below 1.0V must not exceed 15 seconds.)
- Cycling the state of PS_ON from on to off to on. (The minimum cycle time will be 1mS.)

Current Sharing
Outputs of two (or more) supplies connected in parallel must meet the regulation requirements of a single supply. Under normal operation with two (or more) supplies running in parallel the following outputs must share load current.

If one of the supplies fails, the remaining supply (supplies) must pick up the entire load without any of the outputs dropping out of regulation. A defective supply that is connected to the output voltage bus will have no adverse effect on the operation of the remaining functional supply (supplies).

Table 26. Output Current Sharing

Output Level	Output Sharing
+3.3V standby	Not required
+12V	Active

I2C Devices
All I2C devices will be powered from the cathode side of the +3.3V standby OR'ing diode. This will allow the status and FRU data to be read from a power supply that is not powered on or has some other fault. Protection is provided so if a fault within the power supply occurs it does not take down the +3.3V standby bus.

Address locations will be determined by external settings through P1, pin A5. The A1 and A2 address will be wired high on the power supply. (NE1617A* does not have an A2 address). The alert signal from (only) the I/O port will be through P1, pin D5.

FRU Data
The power supply contains a serial EEPROM. The address will be either AC or AE depending on address bit A0.

Temperature Sensors
A temperature sensor, Philips* NE1617A or equivalent, will be located near the air inlet of the supply. The address will be either 34 or 9C depending on address bit A0. The second sensor location will be next to the exhaust outlet.

Power Supply Module LED indicators

Power Supply Fail
This amber LED is driven by internal circuitry and will illuminate when a power rail has failed. The LED should not be illuminated if the supply turns off due to PS_KILL. The LED will illuminate even if the power supply is in a latched state. The only time (during a fault) when it will not illuminate is when the +3.3VSB is lost.

Power Good
This green LED is driven by internal circuitry and will illuminate whenever PWRGD is asserted.

AC OK
This green LED is driven by internal circuitry and will illuminate whenever VIN_GOOD is asserted.

Regulatory Agency Requirements
The power supply must have UL recognition, CSA or cUL certification to Level 3, or any NORDIC CENELEC-certified (such as SEMKO, NEMKO or SETI) markings demonstrating compliance. The power supply must also meet FCC Class B, VDE 0871 Level B, and CISPR Class B requirements.

Power Distribution Board
The power distribution board provides docking connectors for the hot-swappable power supply modules and distribute power to the main board and SAS backplane. A group of comparators on the PDB supplies total power consumption information to the system main board. The board also contains EEPROM FRU information storage. The PDB has no logic on it; it is essentially a pass-through board.

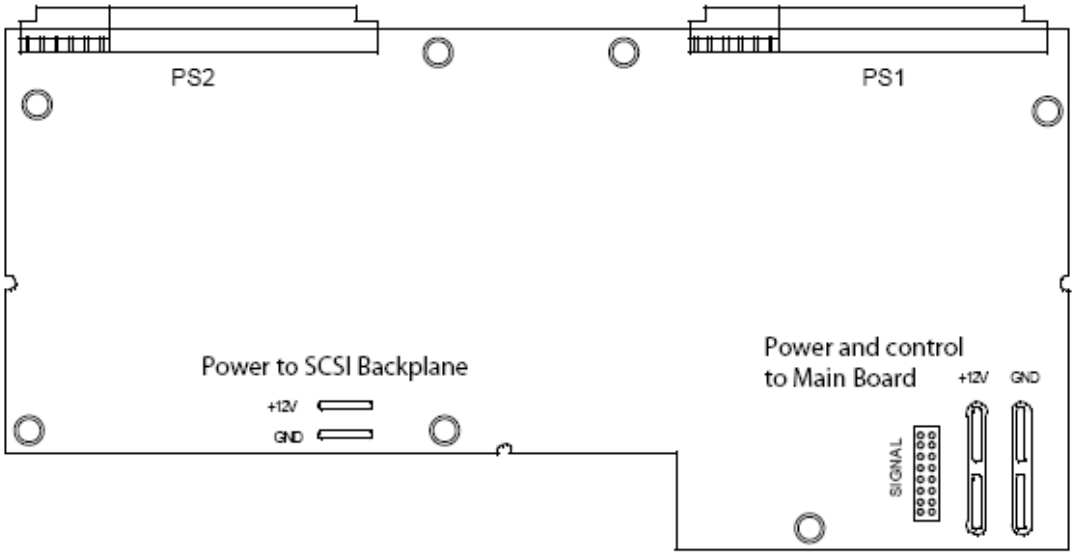


Figure 30. Power Distribution Board Layout

Remote On/Off (-PS_ON)
The power supply DC outputs will be enabled when this signal is pulled low, below 0.8V. In the low state, the input will not source more than 1mA of current. The DC outputs will be disabled when the input is driven higher than 2.4V, or open circuited.

Provisions for de-bouncing will be included in the –PS_ON circuitry to prevent the power supply from oscillating on/off at startup.

POWER GOOD SIGNAL (POK, or P_GOOD)
A power good signal will be asserted, driven high, by the power supply to indicate that all outputs are valid. If any of the outputs fails then this output will be driven low.

In the event AC main power is lost, or a fan has failed, this signal must be driven low at least 1ms before any of the outputs go out of regulation.

The output will be an open collector/drain. It will be capable of driving the output below 0.4V

with a load of 4mA. The output will have an internal pull-up resistor of 1K between the output and +3.3V standby. The pull-up will be connected to the anode side of the +3.3V standby OR'ing diode.

This output also goes to I2C port P5.

VIN_GOOD
This signal will be asserted, driven high, by the power supply to indicate that the input voltage meets the minimum requirements of the input voltage range. Within 12ms after falling outside the input voltage requirements, the output must be driven low.

Front Panel I/O and Control Boards
The front panel I/O board gives the end user access to the system video and USB interfaces. It also interfaces with the front panel control module that contains the buttons and LEDs.
Circuitry on the I/O board consists of the following:

- USB hub controller
- Thermal sensor
- FRU information EEPROM
- Speaker
- Miscellaneous circuitry to support the front panel control module

Architectural Overview

- The front panel I/O board provides five main functions for the system.
- Main board to SAS backplane signal interconnects
- Fan tach (RPM) signals
- Fan PWM speed control
- Reset logic
- I2C bus
- Backplane D2D enable
- Backplane power good signal
- USB hub
- External front panel connector for three USB 2.0 ports
- High-speed hub controller to support the port, above
- Required safety fusing and EMI filtering for the hub
- Video Output
- External front panel 15-pin VGA connector
- Required safety fusing and EMI filtering
- Speaker
- Audible beep-code and alarm speaker
- Speaker drive circuitry
- Miscellaneous
- FRU file storage
- Thermal sensor connected to I2C bus
- NMI button

12.2 **Functional Architecture**

This section provides a more detailed architectural description of the front panel I/O board's functional blocks.

VGA

The front panel I/O board passes the VGA video signals from the main board 100-pin connector to the external video connector.

USB

The front panel I/O board contains a high-speed USB 2.0-compliant controller hub that provides three external USB ports. The controller is accessible through the I2C bus at address HEX 5A.

FRU

The front panel I/O board contains FRU information EEPROM, accessible through the I2C bus at address HEX A6.

Thermal Sensor

A thermal sensor is located on the front panel I/O board to sense ambient air temperature. It is located on the I2C bus at address HEX 98.

50-pin Control Panel Connector

A separate front panel control module interfaces to the front panel board through a 50-pin connector. The following input functions are routed to the main board: reset, power and chassis ID. The following LEDs are activated through the connector:

- Hard drive status
- NIC1 and NIC2 activity
- System status
- Power and system ID

Speaker

The front panel board contains the drive circuitry and the system speaker.

NMI Button

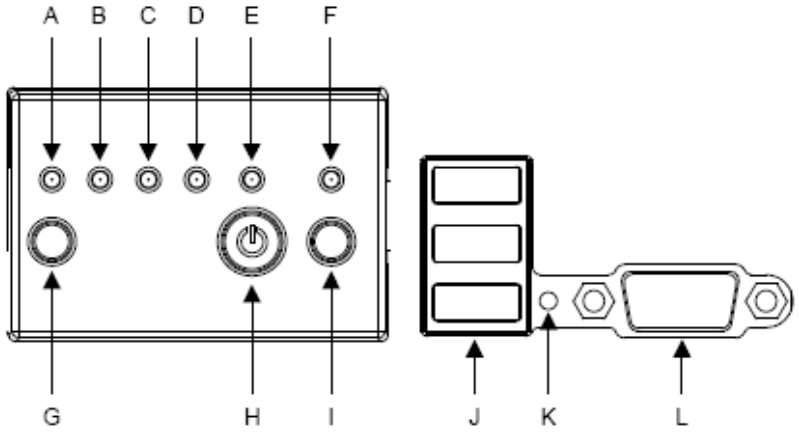
The front panel I/O board has the system NMI button and routes the signal to the 100-pin connector.

Main Board and SAS Backplane Connectors

The front panel I/O board contains the 100-pin connector that interfaces with the main board. Required signals are routed to a 34-pin connector that interfaces to the SAS backplane.

Front Panel Control Module

The front panel control module provides button inputs and LED indicators for the system. It snap-fits into the system front plate and connects to the front panel I/O board through a 50-pin connector.



System ID Buttons and LEDs

The system contains two system ID buttons and two blue system ID LEDs. One button/LED pair is located on the front panel and a second button/LED pair is located at the rear of the system. The system ID LEDs can be activated either by the system ID buttons or remotely via server management software to easily locate/identify the system.

Table 27. System ID LED Details

LED State	Description
Off	System ID inactive.
On	System ID active via button.
Blinking	System ID active via remote command

Pressing a button turns the LEDs on Solid. Pressing a button again turns them off.

If the LEDs were activated by a button, they cannot be turned off remotely. If the LEDs were activated remotely, the buttons cannot turn them off.

Figure 31. Non-LCD Control Module

Basic Input/Output System (BIOS)

This chapter describes the functionality of the Basic Input Output System (BIOS) for the ProServ 4680 Server System. It is written for persons involved in design, validation, integration, manufacture, and

support. It is assumed that the reader is familiar with Intel® processors and the standards that define server architecture.

BIOS Architecture
The BIOS is implemented as firmware that resides in the Flash ROM. It provides hardwarespecific initialization algorithms, standard PC-compatible basic input / output (I/O) services, and standard Intel® server board features. The Flash ROM contains firmware for certain embedded devices. These images are supplied by the device manufacturers and are not specified in this document.

The BIOS implementation is based on the Intel® Platform Innovation Framework for EFI architecture and is compliant with all Intel Platform Innovation Framework for the EFI architecture specifications specified in the Extensible Firmware Interface Reference Specification, Version 1.1. The Intel Platform Innovation Framework for EFI is referred to as “Framework” in this document.

Data Structure Descriptions
Data structures in this document are described in “little endian” format. This means that the loworder byte of a multi-byte data item in memory is at the lowest address, while the high-order byte is at the highest address. In some memory layout descriptions, certain fields are marked reserved. Software must initialize such fields to zero, and ignore them when read. On an update operation, software must preserve any reserved field.

BIOS Identification String
The BIOS Identification string is used to uniquely identify the revision of the BIOS being used on the server. The string is formatted as follows:
BoardFamilyID.OEMID.MajorRev.MinorRev.BuildID.BuildDateTime
Where:

- BoardFamilyID = String name for this board family.
- OEMID = Three-character OEM ID. “86B” is used for Intel EPSD.
- MajorRev = Two decimal digits
- MinorRev = Two decimal digits
- BuildID = Four decimal digits
- BuildDateTime = Build date and time in MMDDYYYYHHMM format:
  – MM = Two-digit month
  – DD = Two-digit day of month

- YYYY = Four-digit year
- HH = Two-digit hour using 24 hour clock
- MM = Two-digit minute

For example, BIOS Build 1, generated on September 18, 2006 at 05:56 AM has the following BIOS ID string that is displayed in the POST diagnostic screen:
86B.01.00.0001.091820060556
The BIOS version in the BIOS Setup utility is displayed as:
86B.01.00.0001
The BIOS ID is used to identify the BIOS image. The board ID is available in the SMBIOS Type 2 structure. The board ID is also available in BIOS Setup. The BIOS ID is available in Setup and the SMBIOS Type 0 structure.

BIOS Initialization

Processors

Multiple Processor Initialization

IA-32 processors have a microcode-based Boot Strap Processor (BSP) arbitration protocol. The system BSP starts executing from the reset vector (F000:FFF0h). Any processor not performing the role of system BSP is called an application processor (AP).

The Memory Controller Hub (MCH) supports four processor front side bus (FSB) units, each accommodating one Quad-Core Intel® Xeon® processor. At reset, hardware arbitration chooses one system BSP from the available processor cores on each FSB. The BIOS Power-on Self Test (POST) code requires only one processor for execution. This requires the BIOS to elect a system BSP using registers in the MCH. The BIOS cannot guarantee which processor will be the system BSP, only that a system BSP will be selected.

In the remainder of this document, the system BSP is referred to as the BSP. The BSP executes the BIOS POST and prepares the server to boot the operating system. At boot time, the server is in virtual wire mode and the BSP alone is programmed to accept local interrupts via the INTR signal driven by the 8259 Programmable Interrupt Controller (PIC) and non-maskable interrupt (NMI) logic.

As a part of the boot process, the BSP wakes each AP. All AP Memory Type Range Register (MTRR) sets are then programmed identically to the BSP. All APs then execute a halt instruction with their local interrupts disabled.

The BSP executes CPUID instructions to determine the supported BSP feature set. If the BSP determines that an AP exists that is a lower-featured processor or that has a lower value returned by the CPUID function, then the BSP switches to the lowest-featured processor in the server.

The System Management Mode (SMM) handler expects all processors to respond to a System Management Interrupt (SMI).

Processor Built-In Self Test (BIST)

The BIOS does not support processor BIST. The BIOS leaves all processor BIST settings in the processor and chipset set to power on default values.

Processor Cache

The BIOS enables all levels of processor cache. There are no user options to modify the cache configuration, size, or policies. All detected cache sizes are reported in the SMBIOS Type 7 structures. The largest and highest-level cache detected is reported in the BIOS Setup utility.

Microcode Update

IA-32 processors can correct specific errata by loading an Intel-supplied data block, known as a microcode update. The BIOS stores the update in non-volatile memory and loads it into each processor during POST. The BIOS allows microcode updates to be stored in the flash. This is

limited by the amount of free space available.

The BIOS supports variable size microcode updates. The BIOS performs the recommended update signature verification before storing the update in Flash. The system BIOS supports the real mode Interrupt 15h, Function D042h interface for updating microcode updates in flash memory.

Enhanced Intel SpeedStep® Technology
Intel® Xeon® processors support the Enhanced Intel SpeedStep® Technology. This feature enables the operating system and applications to place processors in various performance states (P-states) as described in the Advanced Configuration and Power Interface Specification, Revision 3.0. P-state changes allow processors to operate in various core-speed ratio and voltage levels.

Thermal Monitor Technology
The BIOS enables both Thermal Monitor (TM) and Thermal Monitor 2 (TM2) as supported by processor hardware. If the processor supports both TM and TM2, then the BIOS enables the Extended Throttle Enable feature. This feature first attempts to reduce processor-operating temperature by activating TM2.

Thermal Monitor 2 (TM2)
TM2 throttles voltage and processor core-to-bus frequency to reduce processor temperature. If this does not reduce processor temperature within approximately 10 ms then the processor activates TM in addition to TM2.

Thermal Monitor (TM)
TM throttles the processor clock according to a duty cycle (e.g. 50%) thus reducing processor throughput and thermal output.

Thermal Control Circuit
If TM and TM2 fail to cool the processor and the temperature continues to rise, eventually the processor Thermal Control Circuit (TCC) engages, causing the processor to enter a shutdown state. This is a Thermal Trip event. When this occurs, system hardware brings the system to a power-off state.

Intel® Extended Memory 64 Technology (Intel® EM64T)
The system BIOS enables Intel® EM64T mode using the following steps:

- Detects whether the processor is Intel® Extended Memory 64 Technology capable
- Initializes the SMBASE for each processor
- Detects the appropriate SMRAM State Save Map used by the processor
- Enables Intel® EM64T during memory initialization if necessary

The BIOS does not activate Intel® EM64T mode. The system is in IA-32 compatibility mode when booting to an operating system.

See the Intel® Extended Memory 64 Technology BIOS Writer's Guide for more information about activating and deactivating Intel® EM64T mode.

Execute Disable Bit Feature

The Execute Disable Bit feature (XD bit) is an enhancement to the Intel® IA-32 architecture. An IA-32 processor supporting the Execute Disable Bit feature can prevent data pages from being used by malicious software to execute code. An IA-32 processor with the XD bit feature can provide memory protection in either of the following modes:

- Legacy protected mode if Physical Address Extension (PAE) enabled.
- IA-32e mode when 64-bit extension technology is enabled. (Entering IA-32e mode requires enabling PAE.)
-

The XD bit does not introduce any new instructions. It requires operating systems to operate in a PAE-enabled environment and establish a page-granular protection policy for memory. The XD bit can be enabled and disabled in BIOS Setup. The default behavior is enabled.

Enhanced Halt State (C1E)

All processors support the Halt State (C1) through the native processor instructions HLT and MWAIT. Some processors implement an optimization of the C1 state called the Enhanced Halt State (C1E) to further reduce the total power consumption while in C1.

When C1E is enabled, and all logical processors in the physical processors have entered the C1 state, the processor reduces the core clock frequency to system bus ratio and VID. The transition of the physical processor from C1 to C1E is accomplished similar to an Enhanced Intel SpeedStep® Technology transition. If the BIOS determines all the system processors support C1E, then it is enabled.

Hardware Prefetch

The automatic hardware prefetch unit operates transparently without requiring programmer's intervention. It is triggered by regular access patterns and helps predict future access thereby overlapping memory latency with computation. By enabling concurrency between memory accesses and computation, the computational benefit of higher processor frequencies is maximized.

Adjacent Cache Line Prefetch

Cache lines can be fetched one at a time, or by enabling Adjacent Cache Line Prefetch the cache lines are fetched in pairs. This can be helpful if the data would continue to the next cache line, causing less cache misses to maximize throughput. When the data is not in adjacent lines, then performance can be slowed due to more cache misses and more time spent filling the cache lines.

Intel® Core Multi-Processing (CMP)

Intel® CMP architecture divides the processor core into multiple parts, allowing separate control of each execution unit and the package with regard to power management. A CMP processor core consists of independent execution cores and shared logic block.

Execution Core Contents

The execution cores consists of:

- Instruction fetch/dispatch units
- Level 1 cache controllers

- Integer/floating point execution units

The shared logic block contains L2 cache controller, bus interface logic and power management logic.

CMP Support

The BIOS takes these steps to support CMP:

- Initializes all processor cores
- Installs NMI handlers for all multi-core processors
- Leaves each AP in CLI/HLT loop after completing processor initialization
- Initializes stack for all APs

The BIOS performs these actions when CMP is enabled:

- The BIOS POST diagnostic screen displays the total number of logical processors.
- Creates a separate ACPI MADT table entry for each logical processor. This causes Windows Device Manager to display a separate processor icon for all logical processors.
- Creates a separate Multiprocessor Specification, Revision 1.4, May 1997, Intel Corporation table entry for each logical processor
- SMBIOS Type 4 structure shows only the physical processors installed. It does not describe the virtual processors.

Intel® Virtualization Technology

Intel® Virtualization Technology supports multiple software environments sharing the same hardware resources. Each software environment may consist of an operating system and applications. Intel® Virtualization Technology can be enabled or disabled in the BIOS Setup utility. The default behavior is disabled. The BIOS runtime error handling is identical regardless of whether Intel® Virtualization Technology is enabled. Operating system error handling is performed only by the primary operating system. The guest operating system(s) are not exposed to any error conditions.

Note: *If the Setup options are changed to enable or disable the Intel® Virtualization Technology Setting in the processor, the user must be fully powered off and powered back on again before the changes take effect.*

“Fake MSI” Support

In PCI compatible INTx mode, the chipset supports a maximum of four unique interrupts. If more than four unique interrupts are used by devices behind the chipset root ports, it could result in a potential interrupt scaling problem due to sharing of interrupts. On a platform that supports eight processor cores, the configuration allows for interrupt distribution to all eight cores.

Since the available number of unique interrupts (4) for this system is less than the number of available cores, the system cannot take advantage of all the available cores for interrupt distribution. The chipsets provides an interrupt scaling feature called “Fake MSI” to mitigate this problem.

All PCI Express* devices must support MSI (Message Signaled Interrupt). In this scheme, the device causes an interrupt by writing the value of the MSI data register to the address in the MSI address register. The resulting memory write transaction is translated through chipset logic into an interrupt transaction for the appropriate target processor core(s). The MSI scheme requires support in the operating systems, which is not widely available in available operating systems. The “Fake MSI”

scheme allows PCI Express devices running on a legacy operating system to use the MSI mechanism to generate INTx compatible interrupts. This is accomplished by targeting the MSI memory write to an I/OxAPIC.

Under the “Fake MSI” scheme, PCI Express devices are programmed to enable MSI functionality, and given a write-path directly to the pin assertion register (PAR) of an I/OxAPIC already present in the platform. The targeted I/OxAPIC now generates an APIC interrupt message in response to a memory write to the PAR, thus providing equivalent functionality to a virtual (edge-triggered) wire between the PCI Express endpoint and the I/OxAPIC. The chipsets ensure that PCI ordering rules are maintained for the “Fake MSI” memory write.

When Fake MSI is enabled, the PCI Express devices generate a memory transaction with an address equal to I/OxAPIC_MEM_BAR + 0x20 (PAR) and a 32-bit data equal to the interrupt vector number corresponding to the device. This information is stored in the device's MSI address and data registers, and would be initialized by the system firmware (BIOS) prior to booting a non-MSI aware operating system.

Limitations

- The “Fake MSI” scheme can only be used by I/O devices that support MSI capability. AllPCI Express* devices must support either MSI or MSI-X.
- The “Fake MSI” scheme cannot be used with a device that supports MSI-X1 The devicesupports MSI-X only and does not support MSI.
- The “Fake MSI” scheme can be used with MSI capable devices only. It cannot be used with a device that only supports MSI-X.
- The I/OxAPIC interrupt used for “Fake MSI” cannot be shared because MSI is an edgetriggered mechanism and sharing results in loss of interrupts.
- Even if the I/O device is multiple-message capable, firmware must program the device to allocate one vector only. The “Fake MSI” scheme cannot support MSI multiple messages. This is required to ensure that the device-function does not modify any bits of the message data field.
- Each I/O device that intends to use the “Fake MSI” scheme should be programmed to a unique MSI data value corresponding to a unique I/OxAPIC input. The MSI address remains the same as we are targeting the PAR of the ESB2 I/OxAPIC.
- If the I/O device generates interrupts for multiple internal events, the device driver ISR must check for all internal events on each interrupt2. Otherwise, overrun situations are possible.
- In case of multi-function devices, the “Fake MSI” scheme can be used to support up tofour functions only. This is because interrupt routing of devices using the “Fake MSI” scheme are exposed to the operating system using MPS1.4 or _PRT table; these firmware tables are limited to four unique interrupts per device as required by the PCI Specification.

Direct Cache Access (DCA)

Direct cache access (DCA) is a component of Intel® I/O Acceleration Technology (Intel® I/OAT).

The DCA mechanism is a system-level protocol in a multi-processor system to improve I/O network performance, thereby resulting in higher system performance. The basic idea is to minimize cache misses when a demand read is executed. This is accomplished by placing the data from the I/O devices directly into CPU cache through hints to the processor to perform a data prefetch and install it in its

local caches.

The BIOS enables DCA by default.

1 MSI-X requires BAR registers to be initialized to locate the MSI-X table in MMIO space. Since legacy operating systems could potentially reconfigure the device and its BARs, in the case of “Fake MSI”, there is a risk of losing the MSI-X programming done by the BIOS.

2 The concern here is that a device driver written with level triggered semantics in mind may dismiss the interrupt with processing all the internal events associated with the interrupt because it is assured that the interrupt will be reasserted as long as internal events are pending.

Snoop Filter
The chipset, in conjunction with the processor, supports the snoop filter feature. The snoop filter stores tags and coherency information for all cache lines. The snoop filter is used to determine if a cache line associated with an address is cached, and if so, where.

Platform Environmental Control Interface (PECI)
PECI is a new thermal management interface. It uses a wire bus interface to provide a communication channel between an Intel processor and an external monitoring device (PECI host controller). The PECI host controller for this system is the ADT7490*. The processors provide processor temperature via PECI interface. The PECI feature configuration and support is controlled via the processor PECI_CTL MSR register (Offset 5A0h Bit0). Installed processors must support PECI. The BIOS polls all installed processors for PECI support. If this is true then the BIOS enables PECI by setting the PECI_CTL MSR Bit 0.

Processor Temperature Monitoring and Management
BMC firmware monitors and manages the processor temperature and enables the PECI polling circuitry.

Physical Layer Topology
The server system PECI physical layer topology supports a 4-way symmetric multi-processor system. PECI devices are identified by their unique, fixed address. All processor PECI devices are located in the address range of 0x30 to 0x33.

The following figure shows an example implementation. More PECI implementation information, including platform topologies, commands, and address values, is in the Platform Environment Control Interface (PECI) Specification.

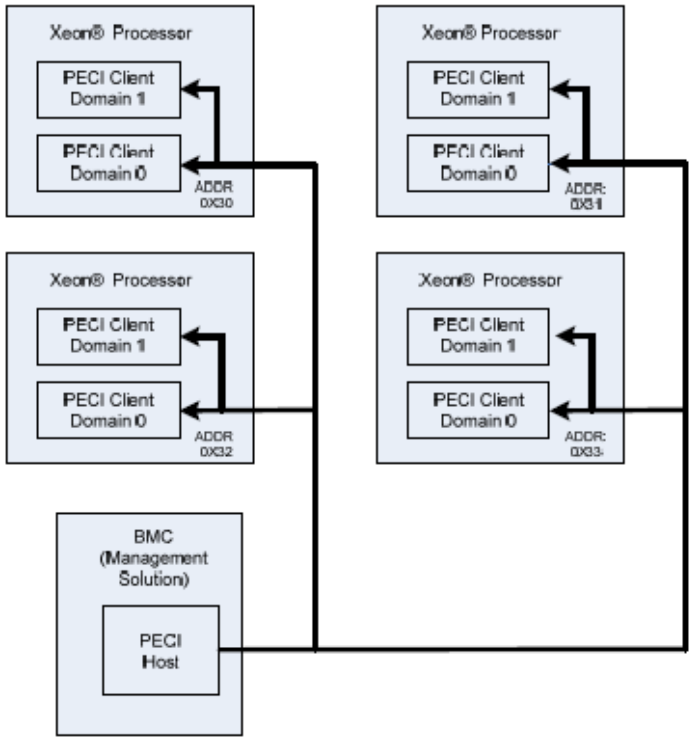


Figure 32. PECI Conceptual Quad Socket Block Diagram

Processor Configuration Errors

Mixed Processor Steppings
For optimum performance, only identical processors should be installed. Processor stepping within a common processor family can be mixed as long as it is listed in the processor specification updates from Intel Corporation.

The BIOS does not check for mixed processor steppings. See the Intel® Xeon® Processor Specification Update for supported mixed processor steppings.

Mixed Processor Families
Processor families cannot be mixed. If this condition is detected, the system responds to the error as described in Table 28.

Mixed Processor FSB Speeds
Processors with different FSB speeds cannot be mixed in a system. If this condition is detected, the system responds to the error as described in Table 28.

Mixed Processor Speeds
Processors with different speeds can be mixed in a system. If this condition is detected, all processor speeds are set to the lowest common denominator (highest common speed) if possible. See Table 28 for more information.

Mixed Processor Cache Sizes
The size of all cache levels must match between all installed processors. Processor cache size mismatches are reported as an error. See Table 28.

Microcode Update Not Available
If the system BIOS detects a processor for which a microcode update is not available, the BIOS reports an error as described in Table 28.

Mixed Processor Configuration
The following table describes mixed processor conditions and recommended actions.

Table 28. Mixed Processor Configurations

Error	BIOS Response
Processor family not Identical	▪ Logs the error into the System Event Log (SEL) ▪ Alerts the BMC of the configuration error with the IPMI Set Processor State command indicating a configuration error for all mismatched processors. ▪ Does not disable the processor ▪ Displays "0194: Processor family mismatch detected" message in the POST Error Manager. ▪ Halts the system
Processor cache not identical	▪ Logs the error into the SEL ▪ Alerts the BMC of the configuration error with the IPMI Set Processor State command indicating a configuration error for all mismatched processors. ▪ BIOS does not disable the processor ▪ Displays "0192: Cache size mismatch detected" message in the POST Error Manager. ▪ Halts the system
Processor frequency (speed) not identical	▪ Adjusts all processor frequencies to lowest common denominator ▪ Continues to boot the system successfully If the frequencies for all processors cannot all be adjusted to be identical, then the BIOS: ▪ Logs the error into the SEL ▪ Alerts the BMC of the configuration error with the IPMI Set Processor State command indicating a configuration error for all mismatched processors. ▪ Displays "0197: Processor speeds mismatched" message in the POST Error Manager. ▪ Halts the system

Error	BIOS Response
Processor microcode missing	▪ Logs the error into the SEL ▪ Alerts the BMC of the configuration error with the IPMI Set Processor State command indicating a configuration error for all mismatched processors. ▪ Does not disable processor ▪ Displays "816x: Processor 0x unable to apply microcode update" message in the POST Error Manager. ▪ Pauses the system for user intervention
Processor FSB speeds not identical	▪ Logs the error into the SEL ▪ Alerts the BMC of the configuration error with the IPMI Set Processor State command indicating a configuration error for all mismatched processors. ▪ Does not disable processor ▪ Displays "0195: Processor Front Side Bus speed mismatch detected" message in the POST Error Manager. ▪ Halts the system

14.2 **Memory**

The chipset Memory Controller Hub (MCH) supports fully-buffered DIMM (FBDIMM) technology. The integrated MCH on the chipset divides the FBDIMMs on the board into two autonomous sets called branches.

Each branch has two channels. In dual-channel mode, FBDIMMs on adjacent channels work in lockstep to provide the same cache line data and a combined ECC. In the single-channel mode only Channel 0 is active.

The BIOS dynamically configures the memory controller in accordance with the available FBDIMM population and the selected Reliability, Availability, and Serviceability (RAS) mode of operation.

Memory Sub-System Nomenclature

The server system complies with the memory subsystem nomenclature guidelines. The guidelines are in the sections below.

Memory Riser Boards

The server system supports four removable memory riser boards. The memory riser board connectors are silk screened on the main board as follows:

- MEM A
- MEM B
- MEM C
- MEM D

14.2.1.2 **FBDIMM Sockets**

Each memory riser board supports eight FBDIMM sockets for a total of 32 FBDIMM sockets.

The memory riser board FBDIMM sockets are silk screened from DIMM_1 to DIMM_8 sequentially from the top to bottom of the board.

14.2.1.3 **DIMM Fault LED**

Each memory riser board supports one DIMM Fault LED for each FBDIMM that is used to report DIMM failures and error conditions. There are no other LEDs supported on the memory riser boards.

Memory Branches

The chipset supports two memory branches, referred to as Branch 0 and Branch 1.

Memory Channels

Each memory branch consists of two channels. The memory channels are identified as Channel 0, 1, 2, and 3.

- Each memory riser board slot is connected to a memory channel.
- The Memory Riser Board A connector on the main board is routed to Branch 0, Channel 0.
- The Memory Riser Board B connector on the main board is routed to Branch 0, Channel 1.

- The Memory Riser Board C connector on the main board is routed to Branch 1, Channel 2.
- The Memory Riser Board D connector on the main board is routed to Branch 1, Channel 3.

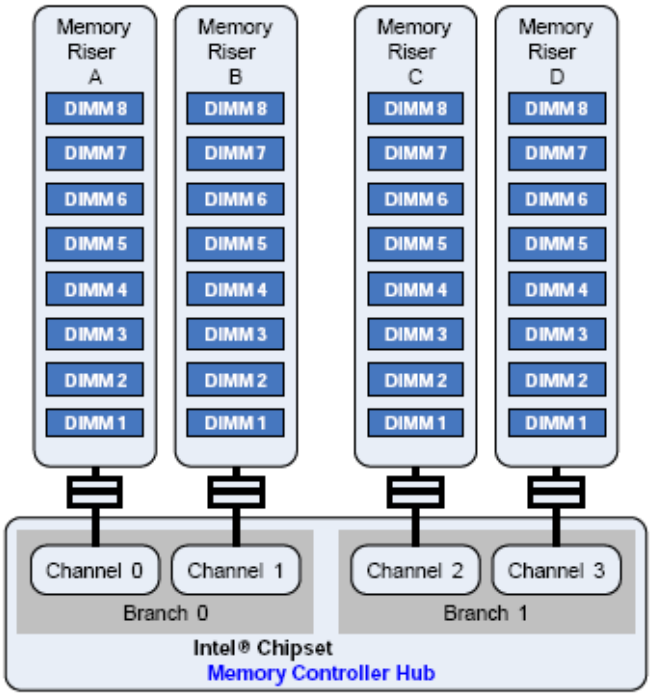


Figure 33. Memory Subsystem Layout

Memory Population Table

The following table describes the memory configurations that are fully supported by the BIOS. The system may work in other configurations, but the behavior may be unpredictable or abnormal. These notations are used in the table:

- SC: Single-channel mode
- SCn: Single-channel mode with n FBDIMMs
- DCm_ABCD): Dual-channel mode with m FBDIMMs per channel, and boards A, B, C, or D populated
- M: Memory Mirroring is possible
- S: DIMM Sparing is possible
- ×: Not possible for RAS mode, or not present for the memory riser board

Table 29. Memory Population

Configuration	RAS	Branch 0		Branch 1	
		Board A	Board B	Board C	Board D
SC1	×	DIMM_1	×	×	×
SC2	×	DIMM_1 DIMM_2	×	×	×
SC4	×	DIMM_1 DIMM_2 DIMM_3 DIMM_4	×	×	×
SC8	×	DIMM_1 DIMM_2 DIMM_3 DIMM_4 DIMM_5 DIMM_6 DIMM_7 DIMM_8	×	×	×
DC1_AB	×	DIMM_1	DIMM_1	×	×
DC1_ABCD	M	DIMM_1	DIMM_1	DIMM_1	DIMM_1
DC2_AB	S	DIMM_1 DIMM_2	DIMM_1 DIMM_2	×	×
DC2_ABCD	S, M	DIMM_1 DIMM_2	DIMM_1 DIMM_2	DIMM_1 DIMM_2	DIMM_1 DIMM_2
DC3_AB	S	DIMM_1 DIMM_2 DIMM_3	DIMM_1 DIMM_2 DIMM_3	×	×
DC3_ABCD	S, M	DIMM_1 DIMM_2 DIMM_3	DIMM_1 DIMM_2 DIMM_3	DIMM_1 DIMM_2 DIMM_3	DIMM_1 DIMM_2 DIMM_3
DC4_AB	S	DIMM_1 DIMM_2 DIMM_3 DIMM_4	DIMM_1 DIMM_2 DIMM_3 DIMM_4	×	×
DC4_ABCD	S, M	DIMM_1 DIMM_2 DIMM_3 DIMM_4	DIMM_1 DIMM_2 DIMM_3 DIMM_4	DIMM_1 DIMM_2 DIMM_3 DIMM_4	DIMM_1 DIMM_2 DIMM_3 DIMM_4
DC6_AB	S	DIMM_1 DIMM_2 DIMM_3 DIMM_4 DIMM_5 DIMM_6	DIMM_1 DIMM_2 DIMM_3 DIMM_4 DIMM_5 DIMM_6	×	×

Configuration	RAS	Branch 0		Branch 1	
		Board A	Board B	Board C	Board D
DC6_ABCD	S, M	DIMM_1	DIMM_1	DIMM_1	DIMM_1
		DIMM_2	DIMM_2	DIMM_2	DIMM_2
		DIMM_3	DIMM_3	DIMM_3	DIMM_3
		DIMM_4	DIMM_4	DIMM_4	DIMM_4
		DIMM_5	DIMM_5	DIMM_5	DIMM_5
		DIMM_6	DIMM_6	DIMM_6	DIMM_6
DC8_AB	S	DIMM_1	DIMM_1	×	×
		DIMM_2	DIMM_2		
		DIMM_3	DIMM_3		
		DIMM_4	DIMM_4		
		DIMM_5	DIMM_5		
		DIMM_6	DIMM_6		
		DIMM_7	DIMM_7		
		DIMM_8	DIMM_8		
DC8_ABCD	S, M	DIMM_1	DIMM_1	DIMM_1	DIMM_1
		DIMM_2	DIMM_2	DIMM_2	DIMM_2
		DIMM_3	DIMM_3	DIMM_3	DIMM_3
		DIMM_4	DIMM_4	DIMM_4	DIMM_4
		DIMM_5	DIMM_5	DIMM_5	DIMM_5
		DIMM_6	DIMM_6	DIMM_6	DIMM_6
		DIMM_7	DIMM_7	DIMM_7	DIMM_7
		DIMM_8	DIMM_8	DIMM_8	DIMM_8

All dual-channel configurations that involve Memory Riser Board A and B require that adjacent DIMMs, Board A {DIMM_m} and Board_B {DIMM_m} are identical in size, organization, timing and electrical characteristics.

All dual-channel configurations that involve Memory Riser Board C and D require that adjacent DIMMs, Board C {DIMM_m} and Board_D {DIMM_m} are identical in size, organization, timing and electrical characteristics.

Modes of Operation

The BIOS configures the system memory into the best possible configuration after comparing the current FBDIMM population with the desired memory configuration selected by the user in BIOS Setup. Possible configurations are:

- Dual-channel Mode (Maximum Performance Mode): The default setting providing the highest system performance and increased FBD bandwidth. This requires each lockstepped pair of FBDIMMs on a branch to be identical. A lock-stepped FBDIMM pair is defined as the FBDIMMs installed in identically numbered FBDIMM sockets on both memory riser boards (channels) on a given Memory Branch.
- Single-channel Mode: A failsafe mode when the installed memory configuration is incompatible with dual-channel operation. In single-channel mode, only Branch 0, Channel 0 is operational with all other FBDIMMs disabled automatically.
- DIMM Sparing Mode: Only supported in a lock-stepped (dual-channel) configuration. DIMM Sparing is the use of a lock-stepped FBDIMM rank on a memory branch to provide a backup in case any other lock-stepped FBDIMM rank on the same branch exceeds a user-selectable Memory ECC Correctable Error threshold in a fixed time period. This failure prediction mechanism allows the system to automatically:
    - Copy the contents of a failing FBDIMM rank to a backup or spare FBDIMM rank

- Disable the failing FBDIMM rank

These actions are completed before the FBDIMM rank begins to generate more serious memory ECC uncorrectable errors that would bring down the system by corrupting memory.

- Memory Mirroring Mode: Memory Mirroring is a high availability mode providing a redundant image of the system memory. This image allows the system to continue operating if memory ECC uncorrectable errors would otherwise bring down the system in another memory configuration.

See RAS Features for details about the DIMM Sparing and Memory Mirroring features.

Single and Dual Channel Configuration Population Rules

- Memory must be populated beginning with Memory Riser Board A, Slot 1.
- Memory riser boards must be populated beginning with DIMM Slot 1 and proceeding with consecutively numbered slots.
- Corresponding, identically-numbered FBDIMM sockets for both memory riser boards (channels) on a branch must be populated with identical FBDIMMs in terms of timing, technology, and size for the branch to operate in lock step (dual-channel) mode.
- FBDIMMs installed in different socket positions (numbers) on a memory riser board do not need to be identical for dual-channel operation.
- Branch 0 is always given precedence over Branch 1 in determining the system memory mode.
- The BIOS uses FBDIMM in Slot 1 on both Memory Riser Board A and B (Branch 0) to determine the memory mode.
- The BIOS always selects the mode of operation that best matches the requirements of Memory Riser Board A, DIMM Slot 1, such that it is always enabled and used for runtime memory.

The BIOS automatically disables any FBDIMM that fails to conform to the rules during the POST memory initialization.

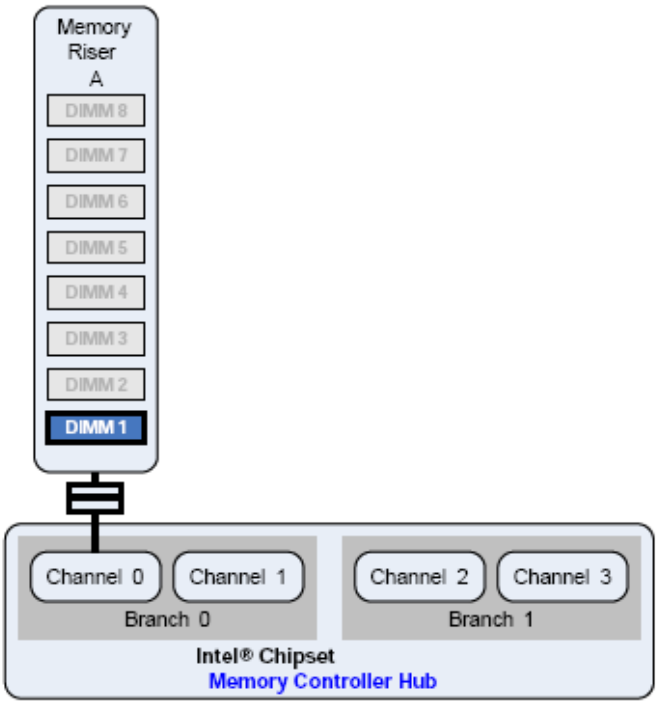


Figure 34. Memory Population for Single-Channel with Minimal Upgrade

This configuration is a minimal system memory configuration, provides the lowest performance and is a fail-safe mode.

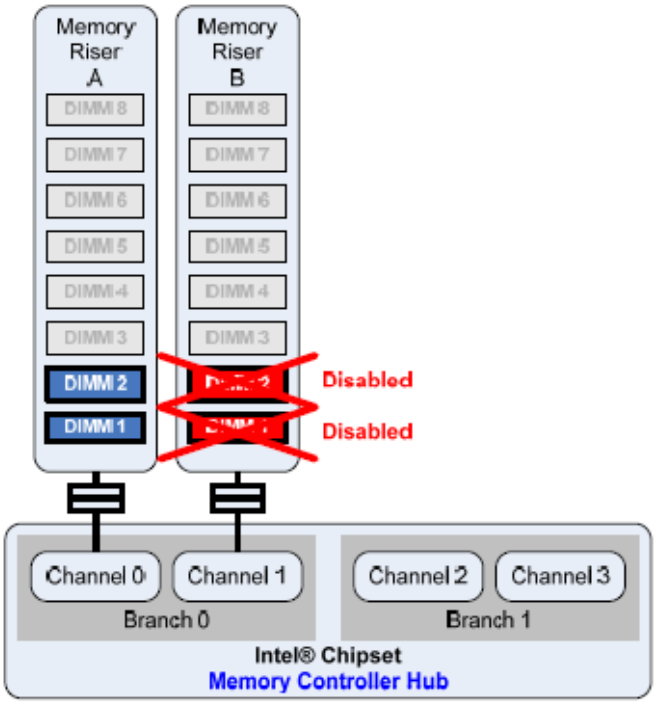


Figure 35. Memory Population for Single-Channel with Multiple FBDIMMs

This configuration:

- Low-performance mode, and is a fail-safe mode when the DIMMs cannot be lockstepped and the system has multiple DIMMs on Memory Riser Boards A and B.
- Default mode if only Memory Riser Board A DIMM slots are populated. See Table 29.

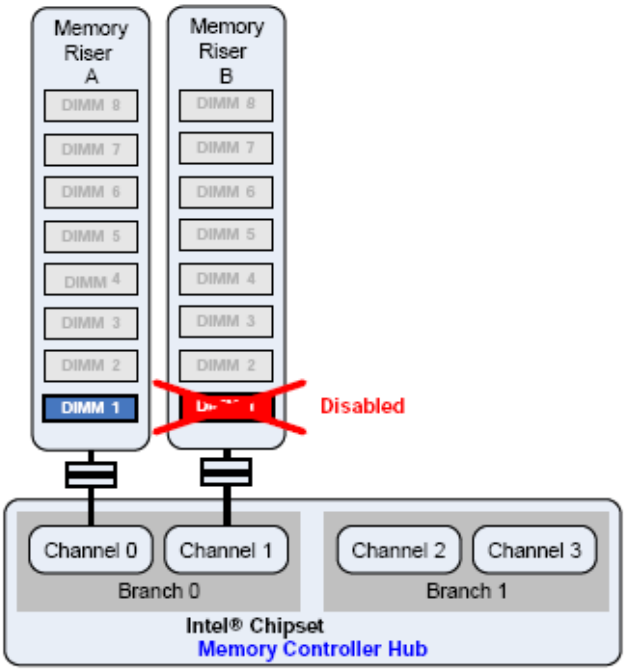


Figure 36. Memory Population for Single-Channel Failsafe

This configuration:

- The FBDIMMs installed in the DIMM_1 sockets on Memory Riser Boards A and B are not identical in organization, size, and speed.
- Does not meet the requirements for dual-channel mode.
- The BIOS selects the operating mode best matching the requirements of Memory Riser Board A, DIMM_1 such that it is always enabled. Therefore, the BIOS disables Memory Riser Board B, DIMM_1 and configures the system for single-channel mode as a failsafe.

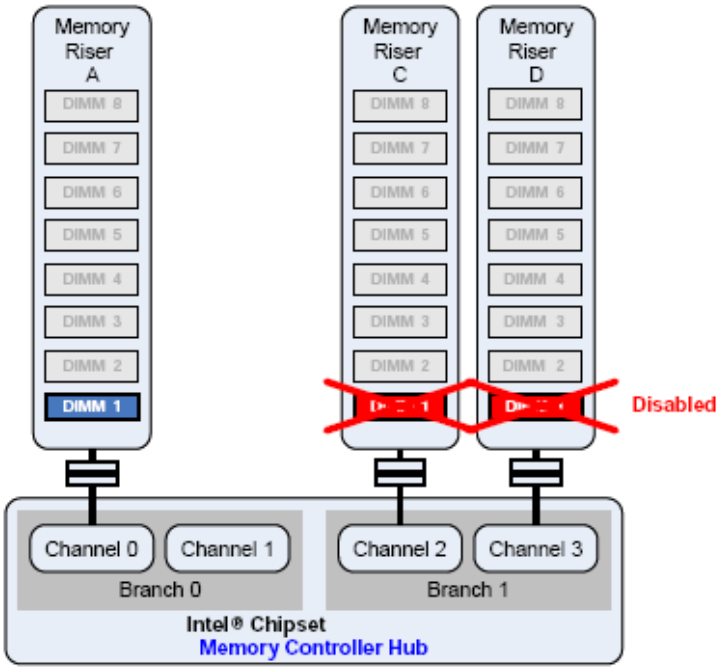


Figure 37. Memory Population for Single-Channel Failsafe

This configuration:

- The population in Branch 0 meets the requirements for single-channel mode only.
- The population in Branch 1 meets the requirements for dual-channel mode because the FBDIMMs installed in the DIMM_1 sockets on Memory Riser Boards C and D are identical in terms of organization, speed, and size.
- The BIOS uses the FBDIMM population of Slot 1 on both Memory Riser Board A and B (Branch 0) to determine the memory-operating mode. Therefore, the BIOS disables all FBDIMMs except Memory Riser Board A, DIMM_1 and configures the system for singlechannel mode.

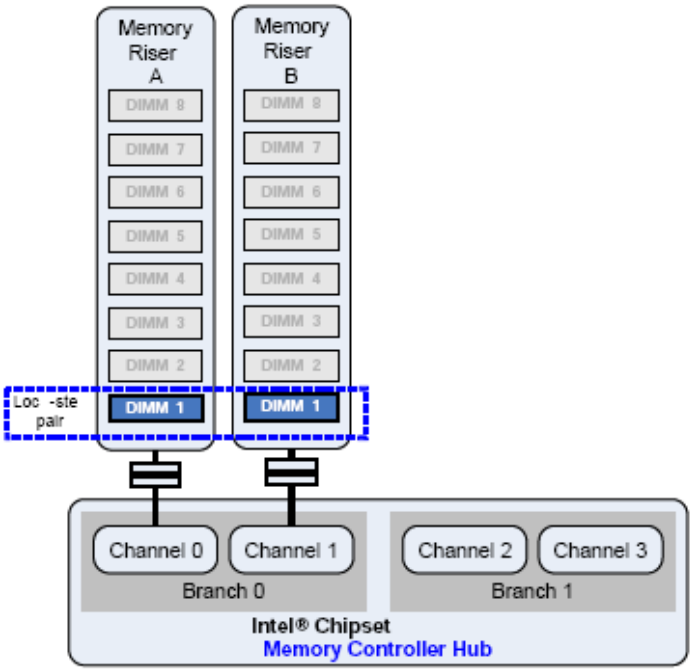


Figure 38. Memory Population for Dual-Channel Configuration on One Branch

This configuration:

- Minimal dual-channel, single branch configuration is one lock-step FBDIMM pair in the DIMM_1 socket on Memory Riser Boards A and B. Additional lock-step pairs can be populated to increase system memory as desired.
- Less efficient in performance than the dual-channel, dual-branch configuration below because the load is all on one branch.
- FBDIMM lock-stepped pair in identically-numbered DIMM slots on both memory riser boards must be identical in organization, size, and speed for dual-channel operation for the system to operate in dual-channel mode.
- The BIOS enables all installed FBDIMM modules and configures the system for dual channel mode.

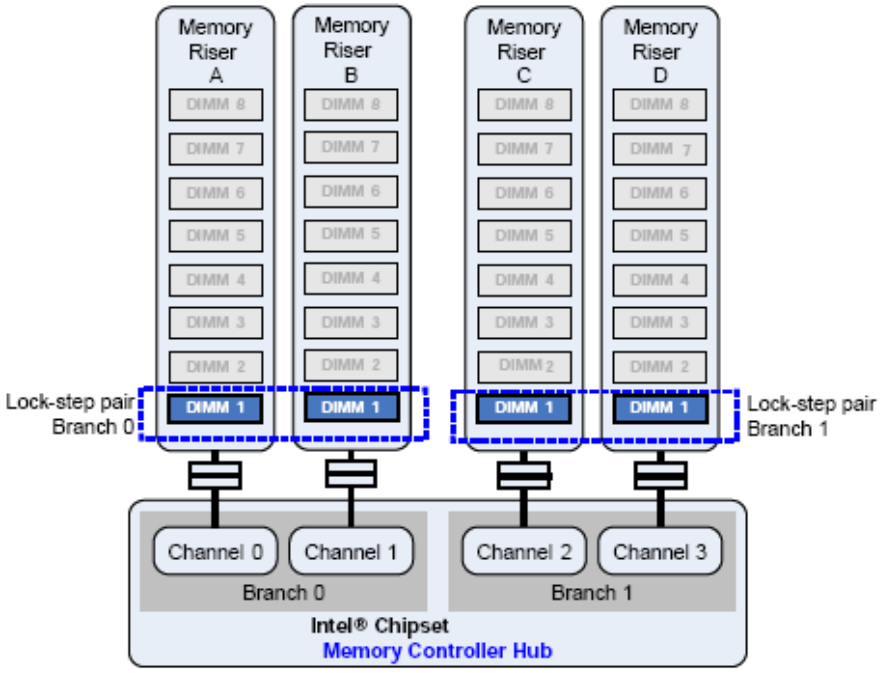


Figure 39. Memory Population for Dual-Channel Configuration on Both Branches

This configuration:

- Minimal dual-channel, dual branch configuration is one lock-step FBDIMM pair in the DIMM_1 socket on Memory Riser Boards A and B and an additional lock-step FBDIMM pair in the DIMM_1 socket on Memory Riser Boards C and D. Additional lock-step pairs can be populated to increase system memory as desired.
- Provides the highest system performance because using both memory branches doubles FBDIMM bandwidth.
- FBDIMM lock-stepped pair in identically numbered DIMM slots on both memory riser boards on a branch must be identical in organization, size, and speed for dual-channel operation.
- FBDIMM lock-stepped pairs in identically numbered DIMM slots do not have to be identical between branches. However, this is recommended to improve memory performance.
- The BIOS enables all installed FBDIMM modules and configures the system for dual channel mode.

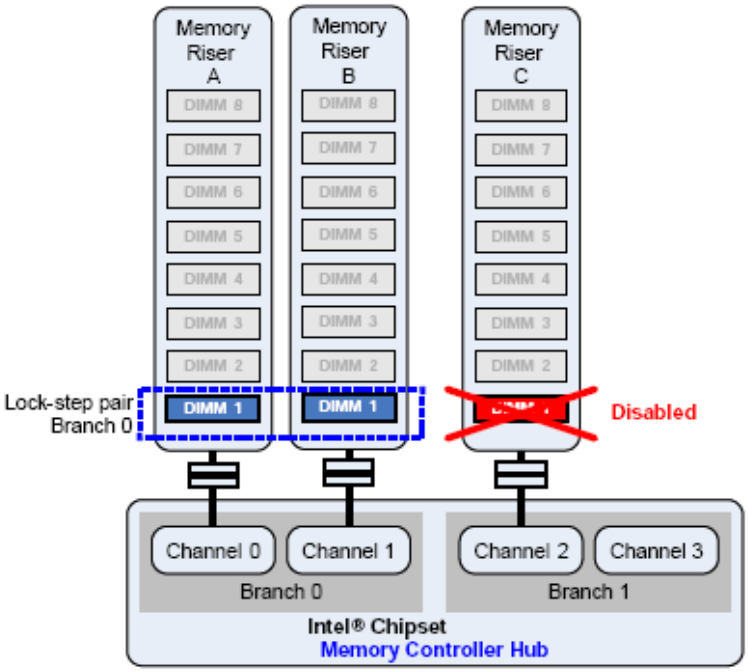


Figure 40. Memory Population for Dual-Channel Failsafe

This configuration:

- The population in Branch 0 meets the requirements for dual-channel mode because the FBDIMMs installed in the DIMM_1 sockets on Memory Riser Boards A and B are identical in terms of organization, speed, and size.
- The population in Branch 1 does not meet the requirements for dual-channel mode because only Memory Riser Board C, DIMM_1 is installed.
- The BIOS uses the FBDIMM population of Slot 1 on both Memory Riser Board A and B (Branch 0) to determine the memory-operating mode.
- The BIOS configures the system for dual-channel mode on one branch and disables Branch 1.

DIMM Sparing Population Rules

FBDIMM Sparing relies on dedicating the largest available FBDIMM rank as a spare in the event of a pending FBDIMM failure. Sparing is only supported in dual channel mode and requires a minimum of two lock-stepped ranks of memory. This requires either two lock-stepped pairs of single-ranked FBDIMM modules or one lock-stepped pair of dual ranked FBDIMM modules. The chipset supports FBDIMM Sparing on each branch independently.

The BIOS Setup utility provides an option to enable sparing. When sparing is selected, the BIOS attempts to enable the feature on both branches, but the actual configuration for a given branch depends

on the population of FBDIMMs on that branch.

Note: The FBDIMM rank(s) allocated for sparing do not contribute to available physical memory since they are reserved as a backup to replace failing FBDIMM ranks. The Effective Memory field in the BIOS Setup utility indicates this absence of memory for the sparing operation.

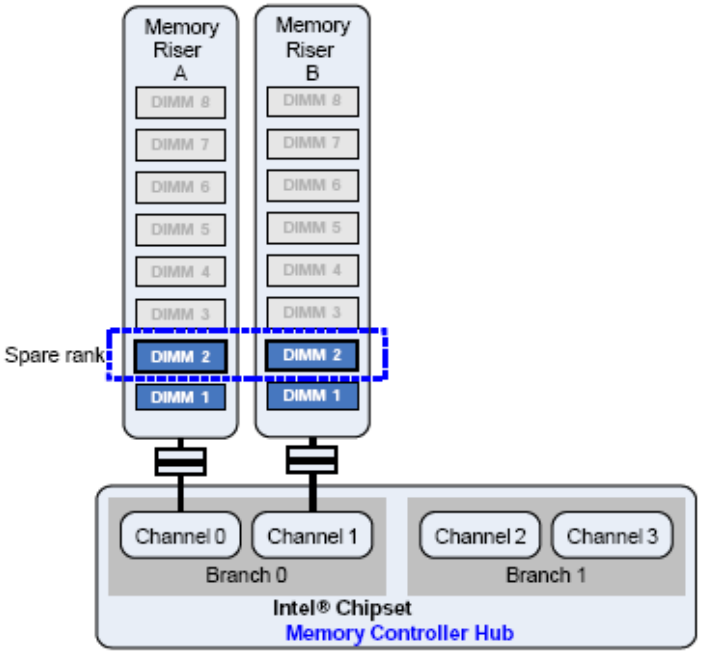


Figure 41. Memory Population for Dual-Channel on One Branch with Sparing

This configuration:

- Less efficient than the dual-channel, dual-branch configuration below because the load is all on one branch.
- Minimal configuration consists of either two single-ranked FBDIMM pairs (four FBDIMMs total) or one dual-ranked FBDIMM pair (two FBDIMMs total).
- Additional FBDIMM lock-stepped pairs can be added to increase system memory as desired.
- The FBDIMM lock-stepped pair in identically numbered DIMM slots on both memory riser boards on a branch must be identical in organization, size, and speed for dualchannel operation.
- User must set Memory RAS to Sparing in the BIOS Setup utility.
- The largest lock-stepped FBDIMM rank is selected for sparing.

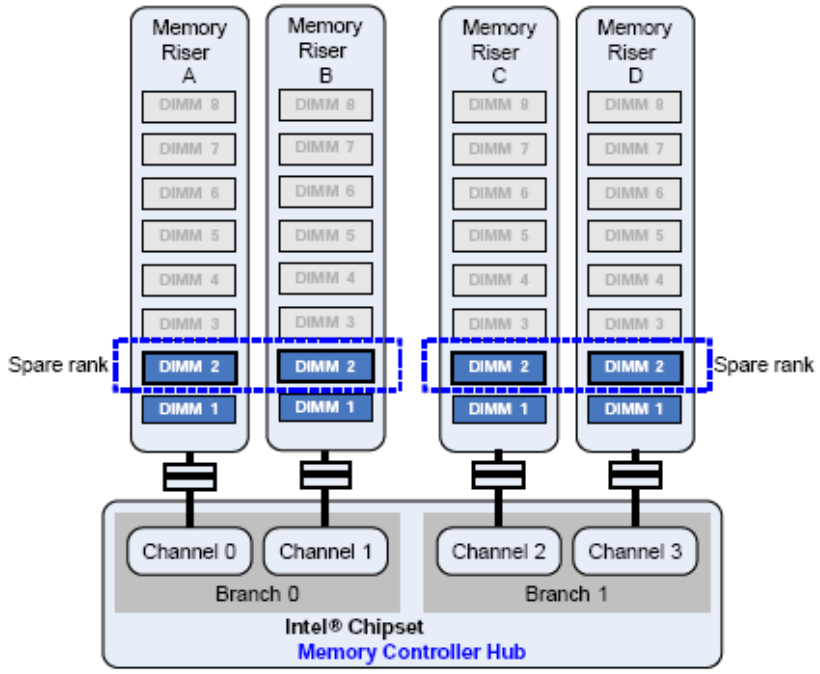


Figure 42. Memory Population for Dual-Channel on Both Branches with Sparing

This configuration:

- This configuration provides the highest system performance because the use of both memory branches doubles FBD bandwidth.
- Minimal configuration is either two single-ranked FBDIMM pairs per branch (eight FBDIMMs total) or one dual-ranked FBDIMM pair per branch (four FBDIMMs total).
- The FBDIMM lock-stepped pairs installed in identically numbered DIMM slots across branches do not have to be identical in organization, size, and speed. However, it is recommended identically numbered DIMM slots be populated with identical FBDIMMs to improve memory performance.
- User must Configure Memory RAS to Sparing in BIOS Setup.
- Sparing is configured independently on each branch.
- The single largest lock-stepped FBDIMM rank on each branch is selected as the spare unit. The spare unit may be located in different FBDIMM slots for the memory riser boards on each branch.

Mirroring Population Rules

Memory mirroring relies on the dual-channel mode of operation with both branches enabled. The two branches provide mirror copies of each other for redundancy. Therefore, the system must operate in dual-channel mode with an identical memory configuration between channels on one branch (for dual-channel operation) and an identical memory configuration between branches (for mirror support).

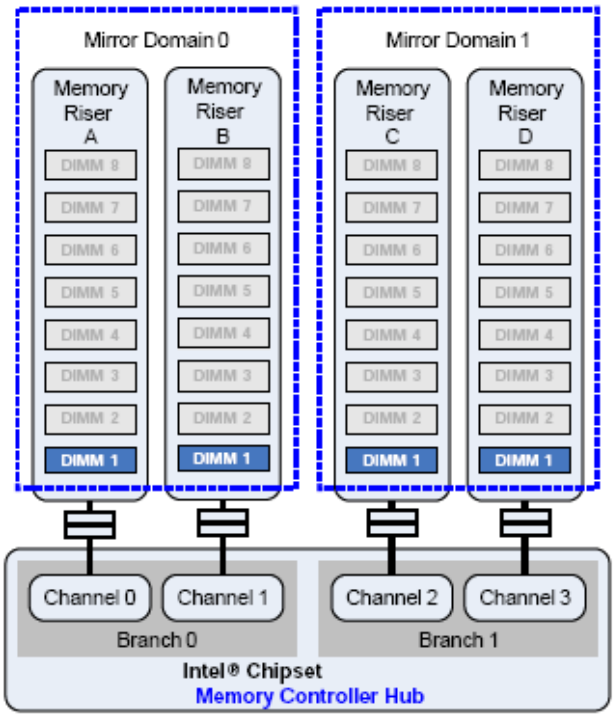


Figure 43. Memory Population for Mirroring

This configuration:

- The minimal mirroring configuration is identical FBDIMMs installed in the DIMM_1 slot of all four memory riser boards. The user can populate additional slots to increase system memory as desired.
- FBDIMMs installed in identically numbered DIMM slots on ALL memory riser boards must be identical in organization, size, and speed for dual-channel operation.
- The User must Configure Memory RAS to Mirroring in BIOS Setup.
- Branch 0 (Memory Riser Boards A and B) is mirrored with Branch 1 (Memory Riser Boards C and D).

Note: The BIOS Setup Memory page displays the state of the FBDIMMs.

Memory Sizing and Configuration
The BIOS supports various memory module sizes and configurations. These combinations of sizes and configurations are valid only for FBDIMMs approved by Intel. The BIOS reads the Serial Presence Detect (SPD) SEEPROMs on each installed FBDIMM to determine the supported size and timing characteristics.

The memory-sizing algorithm then determines the cumulative size of each FBDIMM rank. The BIOS programs the MCH accordingly, such that the range of memory accessible from the processor is mapped into the correct FBDIMM or set of FBDIMMs.

The minimum memory configuration is 512MB for single-channel mode and 1024MB for dualchannel mode. All four channels support up to thirty-two DIMM ranks.

Support for Mixed-Speed Memory Modules
The BIOS supports memory modules of mixed speed by automatic selection of the highest common frequency of all memory modules (FBDIMM). This section describes the expected outcome on installation of FBDIMMs of different frequencies.

FBDIMM Characteristics
To program a FBDIMM to function correctly for a given frequency, the BIOS queries each FBDIMMs Serial Presence Detect (SPD) data store. The SPD contains the frequency characteristics of the FBDIMM, which are measured in terms of the following parameters:

- CAS Latency (CL)
- Common clock frequency
- Additive Latency (AL)
- Buffer Read Delay (BRD)

The CAS latency and the additive latency are configurable parameters detected by the BIOS by reading the SPD data of the FBDIMMs. The BRD is the average inherent delay that is caused by the finite time that the FBDIMM Advanced Memory Buffer (AMB) consumes in buffering the data read from the individual DRAM devices on the module before forwarding data on the Northbound or Southbound path.

Host Frequency and Gear Ratio
The host frequency is the speed of the memory interface of the chipset. This frequency determines the speed at which the chipset completes a memory transaction. The gear ratio determines the relative speed between the processor interface and the memory interface.

The BIOS supports the following two frequencies:
533 MHz
667 MHz

The BIOS automatically selects and configures the host frequency and gear ratio.

During memory discovery, the BIOS keeps track of the minimum latency requirements of each installed FBDIMM by recording relevant latency requirements from each FBDIMMs SPD data. The BIOS then arrives at a common frequency that matches the requirements of all components and configures both the MCH and individual FBDIMMs with that common frequency.

Memory Reservation for Memory-Mapped Functions
Memory address space starting at 4 GB and extending downward is reserved for various system BIOS, chipset, and PCI memory resource requirements. The starting address of this memory hole is controlled by the BIOS Setup PCI Memory Mapped I/O Space menu item. The user can configure the starting address of the memory hole to the following:

- 1.0GB

- 2.0GB (default)
- 3.0GB

The user may wish to select a lower memory address space starting address (larger memory hole size). This increases the amount of memory available for allocation to PCI devices. The chipset provides the High-Memory Reclaim feature. This feature allows the BIOS to remap the physical memory behind this memory hole back into the system memory address space above the 4GB boundary. The BIOS always enables High-Memory Reclaim if it discovers any installed physical memory behind (overlapping) the memory hole address space. See the chipset technical documentation for more details regarding this Memory Mapped Configuration Region.

Operating systems must support Physical Address Extensions (PAE) or Intel® EM64T technology to utilize memory mapped above the 4GB boundary and recapture this memory for operating system and application use. Most operating systems support this feature. See the relevant operating system manuals for details.

Memory Interleaving
In general, to optimize memory accesses, the BIOS enables Branch Interleaving. This allows the chipset to interleave data for successive cache-lines between the autonomous branches.

Additionally, the chipset MCH also provides interleaving across logical memory devices called ranks. A pair of single-ranked lock-stepped FBDIMMs constitutes a memory rank. Interleaving effected between ranks allows the chipset to interleave cache-line data between participant ranks, and the process is called Rank Interleaving.

The BIOS by default enables 4:1 Rank Interleaving, in which four ranks participate in one cache-line access. For more details, see the chipset documentation.

POST Memory Test
The server system supports the Fully-Buffered DIMM (FBD) Memory BIST (MemBIST) engine to initialize and test system memory during the BIOS POST. This MemBIST feature provides extensive coverage of memory errors at both the memory cell level and the data paths emanating from the FBDIMMs.

The BIOS uses this in-built MemBIST engine to perform two specific operations:
- ECC fill operation
- Extensive FBDIMM testing

ECC Fill Operation
An ECC fill operation allows the memory contents set to a known state. This provides a bare minimal error detection capability and is referred to as the Basic Memory Test algorithm.

Extensive FBDIMM Testing
This extensive FBDIMM testing searches for memory errors on both the memory cells and data paths. This is referred to as the Comprehensive Memory Test algorithm. The MemBIST engine replaces the traditional BIOS-based software memory tests and is much faster than the traditional memory tests

thereby reducing overall memory initialization time.

Publishing System Memory
The BIOS notifies the user of installed memory as follows:

- The BIOS displays the “Total Memory” of the system during POST if Quiet Boot is disabled in BIOS Setup. This is the total size of memory discovered by the BIOS duringPOST representing the sum of the individual sizes of FBDIMMs.
- The BIOS displays the “Effective Memory” of the system in the BIOS Setup. This is the total size of all FBDIMMs that are active (not disabled) and not used as redundant units.
- The BIOS provides the total memory of the system in the main page of BIOS Setup. This figure is identical to the “Total Memory” figure reported during POST.
- The BIOS provides the total amount of memory by supporting the EFI Boot Service function GetMemoryMap().
- The BIOS provides the total amount of memory by supporting the Interrupt 15h, Function E820h memory map. See the Advanced Configuration and Power Interface Specification, Revision 3.0 for details.

Note: *The Interrupt 15h, Function E820h system address map reports the memory hole region as reserved (not available memory). Any physical memory behind the hole is relocated above the 4GB boundary and is available to operating systems that support PAE or Intel® EM64T.*

Memory Reliability, Availability, Serviceability (RAS)

RAS Features
The following memory RAS features are supported:
Memory scrub engine
Memory sparing
Memory mirroring
Automatic thermal throttling

Memory Scrub Engine
The chipset MCH incorporates a memory scrub engine. This integrated component, when enabled, performs periodic checks on the memory cells. The scrub engine can identify and correct Memory ECC Single-Bit Errors (SBE). The scrub engine can also identify ECC Multi-Bit Errors (MBE). However, these errors are uncorrectable.

Types of Scrubbing Operations
Two types of scrubbing operations are possible:

- Demand scrubbing – executes when an error is encountered during a normal read/write of data.
- Patrol scrubbing – proactively walks through populated memory space seeking soft errors

.
The BIOS enables both demand scrubbing and patrol scrubbing by default. Demand scrubbing is not possible when memory mirroring is enabled. Therefore, the BIOS disables it automatically if the memory is configured for mirroring.

Memory Sparing

The chipset MCH provides memory sparing capabilities. Sparing is a RAS feature. It involves placing a FBDIMM rank in reserve. This allows it to be used to replace a failing FBDIMM rank at runtime without bringing the system down.

Sparing is only supported in dual-channel mode and is configured on each branch independently. The BIOS assigns one FBDIMM rank per branch to act as a spare (reserve) rank.

Spared Memory Configurations
Spared memory configurations do not provide redundant copies of memory , In addition, the system cannot continue to operate when an ECC Uncorrectable Error / Multi-Bit Error (UE/MBE) occurs. The purpose of memory sparing is to provide runtime failure prediction for FBDIMM ranks exceeding a specified frequency of ECC Correctable Error events in a given time period.

The underlying assumption is that FBDIMMs generating increasing numbers of ECC Correctable Errors are eventually prone to ECC Uncorrectable Errors. These FBDIMMs should be removed from service prior to causing a system crash.

Once a FBDIMM rank exceeds the specified frequency of ECC Correctable Errors the contents of the failing FBDIMM rank are copied to the spare (reserved) FBDIMM rank. Hardware then isolates and removes the failing FBDIMM rank from the set of active FBDIMM ranks. These actions prevent future memory errors and maintain system integrity.

Note: *The DIMM sparing feature requires that the spare FBDIMM rank be at least the size of the largest primary FBDIMM rank in use. When sparing is enabled, the BIOS selects the spare rank automatically during POST. No manual configuration of this feature is required beyond turning on the feature in BIOS Setup. With sparing enabled, the total effective memory size is reduced by the size of the spare FBDIMM rank(s).*

Dual-Ranked DIMM Sparing
When a dual-ranked FBDIMM is used as spare, the BIOS can independently select a physical rank on that FBDIMM as the spare unit and utilize the other physical rank as a normal unit. This selective sparing ensures maximization of available memory while still providing RAS.

Note: *Populating differently ranked FBDIMMs for sparing is not a good practice and may yield unpredictable results.*

Memory Mirroring
The chipset MCH component provides the ability to configure the available set of FBDIMMs in the mirrored configuration. Unlike memory sparing, the mirrored configuration is a redundant image of the memory. In addition, the system can generally continue to operate despite the presence of sporadic ECC Memory Uncorrectable Errors.

Memory mirroring is a RAS feature in which two identical images of memory data are maintained, providing maximum redundancy. Mirroring is achieved across Branch 0 and Branch 1 such that one of these branches is the primary image and the other the secondary. The memory controller alternates between both branches for read transactions. Write transactions are issued to both branches under normal circumstances.

Due to the available system memory being divided into a primary image and a copy of the image, the effective system memory is reduced by one-half. For example, if the system is operating in memory mirroring mode and the total size of the FBDIMMs is 1 GB, the effective size of the memory is 512 MB because half of the FBDIMMs are the secondary images.

For memory mirroring to work, memory riser boards must be installed in pairs and all DIMMs with the same slot number must match. For e.g. Memory Riser Board A DIMM slot 1 must be the same as Memory Riser Board B DIMM slot 1. It is not required to match DIMMs between different slot numbers. DIMMS installed must be the same number of ranks, timing, and size.

The BIOS provides a Setup option to enable memory mirroring. When memory mirroring is enabled, the BIOS attempts to configure the memory system accordingly. If the FBDIMM population is not suitable for mirroring, the BIOS disables mirroring. It then reverts to the default non-RAS mode with maximum interleave or to the single channel mode based on the system memory configuration. BIOS Setup then displays the selected memory configuration on the next boot.

Memory Sub-System Errors
This section describes the BIOS and chipset policies used for handling and reporting errors occurring in the memory sub-system.

The BIOS handles memory errors using a variety of platform-specific policies. Each of these policies is aimed at providing comprehensive diagnostic support to the system administrator towards system recovery following the failure.

Memory Error Classification
The BIOS classifies memory errors into the following categories:

- Memory Controller Fatal and Uncorrectable Errors
- Memory Controller Recoverable Errors
- Memory Controller Correctable Errors

Memory Controller Fatal and Uncorrectable Errors

- This category includes all memory related errors classified as Fatal or Uncorrectable in the chipset datasheet.
- Both of these categories are handled by the BIOS using an NMI to halt the system and prevent silent data corruption and/or erratic system behavior.
- This category also includes persistent Memory ECC Uncorrectable Errors (UE). The chipset ECC engine can detect these errors but cannot correct them.

Memory Controller Recoverable Errors

- This category includes all memory related errors classified as Recoverable in the chipset datasheet.
- This category includes several errors that can potentially be successfully retried.
- If the retry is successful, then the system continues operation.
- If the retry is unsuccessful, then the chipset reports a fatal or uncorrectable error.

Memory Controller Correctable Errors:

- This category includes all memory related errors classified as Correctable in the chipset datasheet.
- BIOS does not respond to any errors in this category except for Memory ECC Correctable Errors (CE) which are errors involving only single-bit data corruption.

Memory BIST (MemBIST)

The BIOS enables the MemBIST hardware engine during POST memory initialization on every boot. The MemBIST hardware engine isolates failed FBDIMMs. The BIOS then completes the following actions:

- Marks those FBDIMMs as failed
- Takes them off-line
- Displays an error message in the POST Error Manager

FBDIMM Channel Failure

The chipset supports FBD Fail-over mode operation for both southbound and northbound interfaces. For southbound lanes, it supports 10-bit lanes with fail-over to nine lanes. For northbound lanes, it supports 14-bit lanes with fail-over to 13 lanes (14-lane fail-over mode).

POST

During POST, failed lanes are detected during channel initialization after reset. If the chipset detects a lane has failed during channel initialization, it configures the interface for fail-over mode.

In general, when a fatal link failure occurs, the BIOS disables all FBDIMMs on that link. If all FBDIMMs are present on the same faulty link, the BIOS generates POST code 0xE1 to indicate that the system has no usable memory, and then halt the system.

Runtime

During runtime, failed lanes are detected during FBDIMM fast reset that can be triggered by alerts or recoverable memory sub-system errors (e.g. multi-bit ECC errors). If the chipset detects a lane has failed during fast reset, it configures the interface for fail-over mode.

If a fatal link failure occurs during normal operation at runtime (after POST), the BIOS signals a fatal error. It then performs policies related to fatal error handling.

Memory ECC Errors

Memory ECC errors are handled by the BIOS at runtime.

Error Counters and Thresholds

The BIOS uses error counters on the chipset and internal software counters to track the number of runtime ECC correctable errors. The chipset increments these counter registers when an error occurs. The count also decays at a given rate programmable by the BIOS. Due to the particular nature of the counters, they are termed Leaky Bucket Counter (LBC) registers.

LBC Registers

The LBC registers provide a measurement of the frequency of errors. The BIOS configures and uses the leaky bucket counters and the decay rate such that it can be notified of a failing FBDIMM. A degrading DRAM typically generates errors faster over time, which is detected by the leaky bucket algorithm.

The BIOS initializes the LBC for memory ECC correctable errors to a value of 10. These counters are on a per-rank basis. See the Glossary section for a definition of the term rank.

Error Period
The error period, or decay rate, defines the rate at which the leaky bucket counter values are decremented. The decay period is the time period for the leaky bucket count to decay to 0. The expected Memory ECC Correctable Error frequency is directly related to the FBDIMM size. Therefore, BIOS should only report a DIMM Sparing failover if the FBDIMMs on a branch exceed the sparing threshold of 10 errors within the time period indicated below:

Table 30. Leaky Bucket Counter Error Decay Periods

FBDIMM Size	Decay Period (Approximate Duration)
512 MB	672 hours
1 GB	336 hours
2 GB	168 hours
4 GB	84 hours
8 GB	42 hours

Recoverable Error Handling in Non-Redundant Mode
In the event of any recoverable memory sub-system error (e.g. Multi-Bit ECC) the chipset issues an AMB fast reset to all FBDIMMs on the branch. It then retries the memory transaction. If either the fast reset or the retry transaction fails BIOS logs a SEL entry for Uncorrectable ECC Memory Error and halts the system with an NMI. If the fast reset and transaction retry are both successful no SEL entry is logged and the system continues operation.

See the chipset technical documentation for a detailed description regarding the FBDIMM Error Recovery Scheme. This documentation includes detailed information for both memory read and memory write transactions in non-redundant mode.

Recoverable Error Handling in Redundant / Mirror Mode
Memory write cycles are issued to both mirror domains (i.e. MCH branches). Memory read cycles can be issued to either mirror domain / branch.

In the event of a memory subsystem recoverable error, the chipset hardware attempts to issue an AMB fast reset to both branches. In addition, it may retry the failing transaction depending on whether the error occurs on a memory write or memory read transaction.

AMB Fast Reset Fails
If the AMB fast reset fails on both branches or on the alternate branch then the following is determined:

- There is generally no possibility for a recovery for either memory write or read transactions.

- The BIOS attempts to log a SEL entry for Uncorrectable ECC Memory error and generates an NMI to halt the system.

If the AMB fast reset fails on the branch generating the error then the following actions result:

- The chipset disables the branch.
- The BIOS logs SEL entries indicating a Memory ECC Uncorrectable Error and a transition to non-redundant mode.
- On memory write transactions, the data has also been written to the alternate domain and the system continues operation.
- On memory read transactions, the system retries the transaction on the alternate branch in non-redundant mode. If the error persists on retry, the BIOS attempts to log another SEL entry for Uncorrectable ECC Memory error. In addition it generates an NMI to halt the system. If the retry is successful, the system continues operation in non-redundant mode.

AMB Fast Reset Succeeds

If the AMB fast reset succeeds on both branches then the following actions result:

- The chipset retries the memory transaction.
- If the retry is successful, the system continues operation. On subsequent read cycles to the same location if the c/s detects another Uncorrectable ECC Error then the branch is disabled and system transitions to non-redundant mode.
- If a memory write retry fails, the chipset disables the branch. The BIOS reports a SEL entry indicating an Uncorrectable ECC Memory Error and a transition to non-redundant mode before continuing operation.
- If a memory read retry fails (at this point both branches have failed), the BIOS reports a SEL entry indicating an Uncorrectable ECC Memory Error. It then generates an NMI to halt the system.

Memory Error Handling

Memory errors are reported through a variety of platform-specific elements, as described in this section

Table 31. Memory Error Reporting Agent Summary

Platform Element	Description
POST Error Manager	Memory errors found during the POST MemBIST are reported in the BIOS Error Manager.
POST Beep Codes	The BIOS emits beep code for the following cases: ▪ Where the system has no memory, or ▪ When a link failure is detected during memory discovery, causing all memory to be mapped out.
BIOS Setup Screen	When FBDIMMs fail MemBIST, or RAS configuration errors occur, the FBDIMM status is captured in the Advanced | Memory screen in BIOS setup.
System Event Logging (SEL)	When a memory error occurs at runtime, the BIOS logs the error into the Baseboard Management Controller (BMC) System Event Log (SEL) repository in compliance with the formats described in the Error Handling section.
Memory Riser Board DIMM Fault Indicator LEDs	A set of DIMM Fault Indicator LEDs is on each memory riser board. One LED is provided for each FBDIMM socket.
Front Panel System Fault/Status LED	There is a Front Panel System Fault Status LED. The BIOS controls the behavior of this LED using BMC commands in response to certain memory sub-system errors.
BMC Memory RAS	The BIOS sends the memory riser board and FBDIMM status to the BMC during POST and in response to certain memory subsystem error conditions at runtime. The user can retrieve this information from the BMC.
NMI Generation	The BIOS triggers / initiates an NMI to halt the system when a critical error occurs.

The tables on the following pages describe the platform response to various memory subsystem errors both during POST and runtime.

The BIOS issues a Set Fault Indication command to the BMC to request a change to the Front Panel System Fault Status LED. The table below indicates the desired behavior of the LED after this command.

The BMC maintains an internal state machine to manage requests from different sources so the final behavior of the LED may differ from what the BIOS requests.

Table 32. Memory Error Handling — POST

Error Scenario	POST Behavior	System Event Log (SEL)	DIMM Fault LED System Fault LED	IPMI Memory RAS Behavior	System Operation
Intel® MemBIST Uncorrectable Error (UE / hard error)	Uncorrectable Error SEL message identifying FBDIMM location(s) POST Error Manager 0x852x, 0x853x	UE POST code DIMM Failed POST code SEL messages identifying FBDIMM location(s)	DIMM LED: ▪ On for the failed FBDIMM only System Fault LED: ▪ No change	Set DIMM State: ▪ DIMM failure status = Y ▪ DIMM disabled status = Y Set Memory RAS Redun. State Set RAS Config Information: ▪ No change	The system continues to boot if good memory is found. If no good memory is found, the system emits a beep code, displays a POST diagnostic LED message, and halts the system.
Intel® MemBIST Channel Error	Uncorrectable Error SEL message identifying FBDIMM location(s)	UE POST code DIMM Failed POST code SEL messages identifying FBDIMM location(s)	DIMM LED: ▪ On for all affected FBDIMMs System Fault LED: ▪ No change	Set DIMM State: ▪ DIMM failure status = Y ▪ DIMM disabled status = Y Set Memory RAS Redun. State Set RAS Config Information: ▪ No change	The system: ▪ Disables all FBDIMMs on the FBDIMM channel that failed. ▪ Continues to function normally if there are good FBDIMMs to be found on the other channel or branch ▪ Lights fault LEDs for all FBDIMMs, starting from the first, that failed MemBIST irrespective of whether these DIMM sockets are populated or not. This is to indicative a broader-level channel or branch failure.

Table 33. Memory ECC Error Handling — Runtime, Non-Redundant Configuration

Error Scenario	System Event Log (SEL)	DIMM Fault LED1 System Fault LED	IPMI Memory RAS Behavior	System Operation
Correctable Errors CE < Threshold2	CE SEL message	DIMM LED: ▪ No change System Fault LED: ▪ No change	Set DIMM State Set Memory RAS Redun. State Set RAS Config Information: ▪ No change	The system continues to operate.
CE = Threshold	CE SEL message CE Threshold Reached SEL message CE Logging Stopped SEL message	DIMM LED: ▪ On for the failed FBDIMM only System Fault LED: ▪ Amber blink: More than one FBDIMM installed. ▪ Amber on: One FBDIMM installed.	Set DIMM State: ▪ DIMM failure status = Y ▪ DIMM disabled status = Y Set Memory RAS Redun. State Set RAS Config Information: ▪ No change	The system continues to operate normally, but masks all correctable memory errors.
CE > Threshold	No action	DIMM Fault LED: ▪ No change System Fault LED: ▪ No change	Set DIMM State Set Memory RAS Redun. State Set Memory RAS Configuration: ▪ No change	Operating system continues to operate normally.
UE	UE SEL message identifying the FBDIMM location	DIMM Fault LED: ▪ On for the lock stepped pair or for a single FBDIMM, depending upon the mode of operation System Fault LED: ▪ Amber on	Set DIMM State: ▪ DIMM failure status = Y ▪ DIMM disabled status = Y Set Memory RAS Redun. State Set RAS Config Information: ▪ No change	Chipset initially reports recoverable error. After initial recoverable error is reported, the chipset issues AMB fast reset and retries the memory transaction. If either the fast reset or retry fails. The BIOS logs a SEL record for uncorrectable ECC memory error and halts the system with an NMI. See Section 14.2.14.4.3 for more information.

Notes:

1. When an FMDIMM pair is operating in lock-stepped mode and one of the FBDIMMs fails, the BIOS lights the DIMM Fault LED of both FBDIMM modules because the failure cannot be isolated at the individual FBDIMM level in this mode.
2. The correctable Error logging threshold for non-redundant configurations = Ten Correctable Errors.

Table 34. Memory ECC Error Handling — Runtime, Redundant Configuration

Error Scenario	System Event Log (SEL)	DIMM Fault LED System Fault LED	IPMI Memory RAS Behavior	System Operation
Config = Sparing CE < Threshold1	CE SEL message identifying FBDIMM location	DIMM LED: ▪ No change System Fault LED: ▪ No change	Set DIMM State: Set Memory RAS Redun. State: Set Memory RAS Configuration: ▪ No change	The system continues to operate normally.
Config = Sparing CE = Threshold	CE SEL message identifying FBDIMM location CE Threshold Reached SEL message identifying FBDIMM location CE Logging Stopped SEL message	DIMM Fault LED: On for failed FB-DIMM only System Fault LED: ▪ Green blink.	Set DIMM State: ▪ DIMM failure status = Y ▪ DIMM disabled status = Y ▪ DIMM sparing status = N Set Memory RAS Redundant State: ▪ Domain = 0000b Sparing ▪ State = 01b Non-redundant, sufficient Set Memory RAS Configuration: ▪ Sparing Domain Enable = 0	The system continues to operate normally. System transitions to non-redundant mode. The BIOS masks all correctable memory errors.
Config = Sparing CE > Threshold	No action.	DIMM Fault LED: ▪ No change System Fault LED: ▪ No change	Set DIMM State Set Memory RAS Redun. State Set Memory RAS Configuration: ▪ No change	Operating system continues to operate normally.
Config = Sparing Pre-SFO3 UE	UE SEL message identifying FBDIMM location	DIMM Fault LED: ▪ On for the failed pair of FBDIMMs in lock-step mode or for the failed FBDIMM in single-channel mode. System Fault LED: ▪ Amber on	Set DIMM State: ▪ DIMM failure status = Y Set Memory RAS Redun. State Set Memory RAS Configuration: ▪ No change	The system asserts an NMI.

Config = Sparing Post-SFO4 UE	UE SEL message identifying FBDIMM location	DIMM Fault LED: ▪ On for the failed pair of FBDIMMs in lock-step mode or for the failed FBDIMM in single-channel mode. System Fault LED: ▪ Amber on.	Set DIMM State: ▪ DIMM failure status = Y Set Memory RAS Redun. State: Set Memory RAS Configuration: ▪ No change	The system asserts an NMI.
Config = Mirror Current State: Redundant CE < Threshold2	CE SEL message identifying FBDIMM location	DIMM Fault LED: ▪ No change System Fault LED: ▪ No change	Set DIMM State: Set Memory RAS Redun. State Set Memory RAS Configuration: ▪ No change	Operating system continues to operate normally. The BIOS masks all correctable errors.
Config = Mirror Current State: Redundant CE = Threshold	CE SEL message identifying FBDIMM location CE Threshold Reached SEL message identifying FBDIMM location	DIMM Fault LED: On for the failed FB-DIMM only System Fault LED: ▪ Green blink	Set DIMM State: ▪ DIMM failure status = Y ▪ DIMM disabled status = Y Set Memory RAS Redun. State Set Memory RAS Configuration: ▪ No change	Operating system continues to operate normally. The BIOS does not respond to further correctable errors on this lock-stepped FBDIMM pair.
Config = Mirror Current State: Redundant CE > Threshold	No action.	DIMM Fault LED: ▪ No change System Fault LED: ▪ No change	Set DIMM State: Set Memory RAS Redun. State Set Memory RAS Configuration: ▪ No change	Operating system continues to operate normally.
Config = Mirror Current State: Redundant UE Transition to non-redundant mode	UE SEL message identifying FBDIMM location Redundancy Loss	DIMM Fault LED: ▪ On for the failed pair System Fault LED: ▪ Green blink	Set DIMM State: ▪ DIMM failure status = Y ▪ DIMM disabled status = Y ▪ For all FBDIMMs on the failed branch / group Set Memory RAS Redun. State: Set Memory RAS Configuration: ▪ Redundancy Lost	The system transitions to non-redundant mode in the following circumstances: ▪ AMB fast reset fails on branch generating the error. ▪ AMB fast reset succeeds on both branches but memory write retry fails. ▪ See Section 14.2.14.4.4 for more information.

Config = Mirror Current State: Redundant UE Fatal error	UE SEL message identifying FBDIMM location	DIMM Fault LED: ▪ On for the failed pair System Fault LED: ▪ Amber on	Set DIMM State: ▪ DIMM failure status = Y ▪ DIMM disabled status = Y ▪ For all FBDIMMs on the failed branch / group Set Memory RAS Redun. State: Set Memory RAS Configuration: ▪ No change	The system asserts an NMI when: ▪ AMB fast reset fails on both branches. ▪ AMB fast reset fails on alternate branch. ▪ AMB fast reset succeeds on both branches but memory read fails on both mirror domains / branches. ▪ See Section 14.2.14.4.4 for more information.

Notes:
1. The Correctable Error logging threshold for sparing configurations = Ten Correctable Errors.
2. The Correctable Error logging threshold for mirror configurations = Ten Correctable Errors.
3. Pre SFO indicates the memory state before a Spare Fail-Over event. In other words, the active DIMM ranks have not yet accumulated sufficient ECC Correctable Errors to cross the sparing threshold and activate the hardware sparing copy engine.
4. Post SFO indicates the memory state after a Spare Fail-Over event. In other words, after the sparing copy hardware engine has copied the contents of the active DIMM rank that has reached the Correctable Error sparing threshold to the spare rank.

Server Management Aspects of Memory and Memory RAS
The BIOS is responsible for communicating the current memory and memory RAS configuration to the BMC. There are three separate IPMI Intel OEM commands for this purpose:

- Set DIMM State

- Set Memory RAS Configuration
- Set Memory RAS Redundancy State

The BIOS is responsible for issuing these commands to the BMC during POST in order to describe the status of system memory after memory initialization has been completed. BIOS should describe the physical memory installed rather than the state of memory recognized by software.

The BIOS is also responsible for re-issuing these commands to the BMC at runtime after memory errors as specified in the Error Handling section of this document. The BMC then stores this information until subsequently updated in response to another memory error or during the next boot cycle. The current state of system memory can then be queried at any time by the user or an application using the following, complementary BMC commands:

- Get DIMM State
- Get Memory RAS Redundancy State
- Get Memory RAS Configuration

IPMI Command Definitions

These commands are described in this document to indicate platform-specific bit field interpretations for these commands.

Table 35. Memory RAS Baseboard Management Controller Commands

Command	Request / Response Data	Description
Set DIMM State	Request: Byte 1 — DIMM Group Selector ▪ [7:1]: Group Id ▪ [0]: Presence (1 = group present) Byte 2 — Bit Mask of DIMM sockets Byte 3 — Bitmap of DIMM failure state Byte 4 — Bitmap of DIMM disabled state Byte 5 — Bitmap of DIMM sparing state Byte 6 — Bitmap of DIMM presence state Response: Byte 1 — Completion code	This command allows the state of a set of FBDIMMs to be set. Byte 1 Presence bit: ▪ 1 = Memory riser board present ▪ 0 = Memory riser board not present Group ID: ▪ 1 = Memory Riser Board A ▪ 2 = Memory Riser Board B ▪ 3 = Memory Riser Board C ▪ 4 = Memory Riser Board D **Note:** Group ID signifies the memory riser board number not a bitmap format.

Command	Request / Response Data	Description
		For Bytes 2–6 a bitmap format is supported where each bit corresponds to one FBDIMM socket on the memory riser board indicated in Byte 1 Group ID field: ▪ [0] = DIMM_1 ▪ [1] = DIMM_2 ▪ [2] = DIMM_3 ▪ [3] = DIMM_4 ▪ [4] = DIMM_5 ▪ [5] = DIMM_6 ▪ [6] = DIMM_7 ▪ [7] = DIMM_ 8 Byte 2 ▪ 1 = Slot/Socket exists ▪ 0 = Slot/Socket does not exist **Note:** This byte must always supply the physical slot/socket bitmap regardless of whether they are currently populated with FBDIMM modules. **Note:** This byte should always be 0xFF on since each memory riser board supports 8 physical FBDIMM sockets. Byte 3 ▪ 1 = FBDIMM marked as failed ▪ 0 = FBDIMM is good Byte 4 ▪ 1 = FBDIMM is disabled ▪ 0 = FBDIMM is good Byte 5 ▪ 1 = FBDIMM is marked as spare ▪ 0 = FBDIMM is primary FBDIMM, not spare Byte 6 ▪ 1 = FBDIMM is present ▪ 0 = FBDIMM is not present **Note:** Set DIMM State accepts a minimum of 1 byte when setting the group to not Present, otherwise all 6 bytes are needed. **Example:** Byte 1 = 0x03 and Byte 6 = 0x01 means Memory Riser Board A is present. Only DIMM_1 is populated on Memory Riser Board A.

Command	Request / Response Data	Description
Set Memory RAS Redundancy State	Request: Byte 1 — RAS Domain Selector ▪ [7:4]: Domain Type 0000b = Sparing 0001b = Mirroring 0010b:1111b = Reserved ▪ [3:0]: Domain Instance (1-based) Byte 2 — RAS Domain State ▪ [7:2]: Reserved ▪ [1:0]: Specific State 00b = Redundant 01b = Non-redundant, sufficient resources 10b = Non-redundant, insufficient resources 11b = Reserved Response: Byte 1 — Completion code	This command is used by BIOS to inform the BMC of Memory RAS redundancy state. Domain Instance is: ▪ Bit 0 for Branch 0 ▪ Bit 1 for Branch 1 ▪ Bit 0 for mirroring enabled between Branch 0 and Branch 1 – this is the only possible configuration for mirroring. RAS Domain State definition is: ▪ Redundant = Domain is redundant and working properly. ▪ Non-redundant, sufficient resources = Domain has a failure, but is still operational. For example, a FBDIMM has failed and the spare FBDIMMs are being used. ▪ Non-redundant, insufficient resources = Domain has a failure and is now unable to operate. For example, a sparing domain in non-redundant mode with sufficient resources. A spare FBDIMM fails, causing the system to be inoperable.
Set Memory RAS Configuration	Request: Byte 1 – Sparing domain enable mask ▪ [7:0] – Bit set indicates associated domain enabled Byte 2 – Mirroring domain enable mask ▪ [7:0] – Bit set indicates associated domain enabled Response: Byte 1 – Completion code	This command is used by the BIOS to inform the BMC of the Memory RAS redundancy state. Each domain has an associated BMC entity presence sensor whose state is controlled by its associated mask bit state. Sparing Domain Enable Mask is: ▪ [0] = Memory Branch 0 has DIMMs marked for sparing ▪ [1] = Memory Branch 1 has DIMMs marked for sparing Mirroring Domain Enable Mask is: ▪ [0] = Memory Branches 0 and 1 are mirrored (only possible configuration on the chipset MCH).

Subsystem

I/O Subsystem Specification Compliance

- PCI Express Base Specification, Revision 1.1.
- PCI Local Bus Specification, Revision 2.3, PCI Special Interest Group (PCI-SIG).
- PCI-X Express Specification, Revision 2.0.
- PCI to PCI Bridge Architecture Specification, Revision 1.2, PCI Special Interest Group (PCI-SIG)
- Universal Serial Bus v1.1 Specification and Universal Serial Bus Revision 2.0 Specification
- Serial ATA Revision 2.5 Specification, Serial ATA International Organization (SATA-IO)
- Intel Low Pin Count Interface Specification, Revision 1.1. Intel Corporation.

API Specification Compliance

- BIOS Boot Specification Version 1.01. Compaq Computer Corporation, Phoenix Technologies Ltd., Intel Corporation. 1996]
- PCI BIOS Specification, Revision 2.1, PCI Special Interest Group (PCI-SIG)
- Plug and Play BIOS Specification, Revision 1.0a, Relevant portions of the Initial Program Load Device (IPLD) sections only

14.3.2 **Crystal Beach Technology**

The Crystal Beach technology is a component of Intel® I/O Acceleration Technology (Intel® I/OAT) as described in Section 20.1. The Crystal Beach technology offers higher network performance. In addition, it accelerates TCP/IP Offload Engine (TOE) performance.

The chipset and server hardware support Crystal Beach technology. Support is provided by dedicated Direct Memory Access (DMA) engines on two of the PCI Express* ports.

These ports are connected to the onboard LAN controllers (Intel® 82563EB and Intel® 82575EB). The BIOS enables Crystal Beach technology by default.

Peripheral Component Interconnect (PCI) Bus

Table 36. Supported On-board PCI Devices

Bus	PCI Device	Comments
PCI 32-bit	Onboard ATI* RN50 Video	
PCI Express* x4	Onboard Intel® Gigabit Ethernet	Intel® 82563EB
PCI Express* x4 Dedicated Slot	I/O riser board	Supports Intel® 82575EB Gigabit Ethernet device
PCI Express* x4 Dedicated Slot	SAS riser board	Supports LSI* 1078 SAS Controller
PCI Express* x8	Slot 1	Hot-plug Capable

Bus	PCI Device	Comments
PCI Express* x8	Slot 2	Hot-plug Capable
PCI Express* x8	Slot 3	
PCI Express* x8	Slot 4	
PCI Express* x4	Slot 5	
PCI Express* x4	Slot 6	
PCI Express* x4	Slot 7	

Resource Allocation
The BIOS assigns PCI bus numbers and base addresses for I/O, prefetch memory, and nonprefetch memory resources in ascending order of bus number, device number, and function number.

PCI Express* Configuration Space
PCI Express* configuration space is assigned starting at the top of lower memory (HECBASE). The user can select the starting address of this PCI Memory Mapped I/O Space in BIOS Setup.

PCI memory resources are assigned starting at the bottom of this memory hole upwards to the upper boundary of PCI Express* configuration space at 0xFE000000.

Interrupt Allocation
The BIOS assigns interrupts for devices connected to the North Bridge to the ESB2 PXH I/O Advanced Programmable Interrupt Controller (IOAPIC) [0..6]. Interrupts for ESB2 internal devices and devices connected to the ESB2 are assigned to the ICH IOAPIC [0..7].

Interrupt Delivery
Interrupt delivery is via PCI Express* default ASSERTx/DEASSERTx messages delivered to the appropriate I/OAPIC unit. Final delivery to the CPU complex is via Programmable Interrupt Controller (PIC) by default or I/OAPIC by operating system switch. The BIOS does not support Message Signaled Interrupts (MSI) or user interrupt selection.

“Fake MSI” Support
“Fake MSI” is enabled only for the BnB PCI-Express ports 4 though 7. The respective slots using these ports would be benefited by using “Fake MSI”.

Option ROM Support
BIOS Setup options are provided to allow for user control of Option ROM Enable/Disable for onboard devices and adapters installed in PCI Express* slots. This allows for utilization control over limited Legacy Shadow RAM region resources in the C000h and D000h segments. In most cases, Legacy Option ROM execution is not required unless its device is an Initial Program Load (IPL) device for a legacy operating system.

Table 37. PCI Onboard Device Option ROM List

Option ROM Provider	Option ROM Functionality
ATI	RN50 Video (embedded video controller)
Intel	SATA Advanced Host Controller Interface
LSI	SATA Software RAID
LSI	SAS Integrated RAID (IR) Mode
LSI	SAS Software RAID
Intel	PXE (Combined Option ROM for Intel® 82563EB and Intel® 82575EB support)
Intel	iSCSI (Combined Option ROM for Intel 82563EB and Intel 82575EB support)

Note: *The LSI* SAS IR and Software RAID mode Option ROM images are integrated directly on the SAS riser board instead of the System BIOS image.*

EFI PCI APIs
The BIOS provides standard PCI protocols as described in the Extensible Firmware Interface Reference Specification, Version 1.1.

Legacy PCI APIs
In legacy mode, the system BIOS supports the Interrupt 1Ah, Function B1h functions as defined in the PCI BIOS Specification, Revision 2.1, PCI Special Interest Group (PCI-SIG). The system BIOS supports the real mode interface.

Dual Video
The BIOS supports single and dual video modes. Dual video mode is disabled by default.

- In single mode, the onboard video controller is disabled when an add-in video card is detected.
- In dual mode, the onboard video controller is enabled as the primary video device. The external video card is allocated resources and is considered the secondary video device.

PCI Express* Hot-plug
The server features the following PCI Express* Hot-plug Capable (HPC) slots compliant with the PCI Express Base Specification, Revision 1.1:

- Slot 1 (PCI Express* x8)
- Slot 2 (PCI Express* x8)

A PCI adapter can be added, removed, or replaced in these slots without shutting down the system. The BIOS provides ACPI hot-plug methods to support this functionality. The HPC follows the PCI Express Base Specification, Revision 1.1 requirements. When using an operating system that has native support for PCI Express* hot plug, the operating system invokes the ACPI OSHP method to transfer control of the HPC from the firmware to the operating system.

PCI Hot-plug Controller Initialization in Pre-Boot Phase
The Hot-plug Controllers (HPC) come up in an un-initialized state after a power-off or reset. As a result, the hot-plug slots come up in an un-powered state. The BIOS initializes the hot-plug controllers during the boot process. Hot-plug controller initialization involves applying power to the hot-plug slots and detecting any error conditions during link initialization.

The IDT PCI Express* Expander devices provide Hot-plug Controller support compliant with the PCI Express Base Specification, Revision 1.1. This support allows the following actions to occur without powering down the system:

- PCI card removal
- PCI card replacement
- PCI card addition

PCI Hot-plug Resource Padding in Pre-Boot Phase
The BIOS over-allocates resources to PCI slots during boot process. This is called resource padding. The over-allocation is done at the PCI-PCI bridge level. Over-allocation of resources allows a limited number of add-in cards to be hot-plugged into a PCI bus without disturbing the allocation to the rest of the busses.
By default the BIOS reserves the following set of resources for each hot-pluggable slot

regardless of the presence of a card in the slot:

- One bus number
- 4KB I/O
- 64 MB non-prefetch memory below 4 GB
- 64 MB prefetch memory below 4 GB

Direct Effect of Resource Padding

The direct effect of the resource padding is as follows:

- The total memory allocated by the BIOS to the PCI subsystem increases.
- The main memory available to the operating system in the 0 to 4 GB range decreases according to the lower bounds of the memory addresses allocated to PCI.

The PCI Express Base Specification, Revision 1.1 defines a standard usage model for PCI Hotplug. The usage model specifies several elements including the state of the indicator LEDs' interlocking switches. The server provides these hardware elements and the BIOS supports them in accordance with this specification

Operating System Interfaces

ACPI Control Methods

PCI hot-plug requires an ACPI hot-plug aware operating system. The ACPI methods provided in the BIOS support the standard hot-plug controller usage model.

PCI Hot-plug Usage Model

This section describes the hot-plug usage model for PCI Express* slots. Each of the hot-plug slots has a status LED and an attention button. The attention button is used to invoke a hot-plug sequence to remove or add an adapter without having to use the software interface. The status LED is lit green for power indication and amber for attention indication

Table 38. PCI Hot-plug Power Indication

Power Indication	Status	Definition
Off	Power Off	▪ Power has been removed from the slot. ▪ The card can be inserted or removed.
Green On	Power On/Normal	▪ Power is applied to the slot and operating normally. ▪ The card cannot safely be inserted or removed.
Green Blinking	Power Transition	▪ The slot is in the process of powering-up or powering-down. ▪ The card cannot be safely inserted or removed.

Table 39. PCI Hot-plug Attention Indication

Attention Indication	Status	Definition
Amber On	Attention	Power fault or operational problem of this slot.
Amber Blinking	Locate	Slot is being identified at the user's request.

When a new card is added to the slot, the ACPI BIOS completes the following actions:

- Applies power to the adapter card
- Initiates PCI Express* link training
- Negotiates link width with the adapter

The BIOS then checks for the following:

- PCI Express* link training errors (trained link width less than supported width)
- PCI Express* negotiated link width

Hot-plug Controller (HPC) Interface

The PCI Express Base Specification, Revision 1.1 standardizes the register level programmatic interface of the PCI Express* hot-plug controllers. If the operating system supports this specification and the HPCs comply with the specification, the BIOS-provided control methods are not used.

One such control method is powering on/off the slot. The BIOS provides other control methods that describe platform hardware. However, the operating system can control the power and the other hot-plug elements directly.

PnP ISA

Although the platform does not support add-in ISA devices, some embedded Super I/O legacy I/O devices require ISA resources. As needed, the BIOS assigns the following from the system resource pool to the embedded PnP Super I/O devices:

Memory
I/O
Direct Memory Access (DMA) channels
IRQ

A BIOS Setup option controls the I/O address and interrupt assignment of Serial Port A and Serial Port B. All other legacy addresses are fixed.

Keyboard / Mouse

The BIOS supports only USB keyboards and mice. The system can boot without a keyboard or mouse attached. If present, the BIOS detects the keyboard during POST and displays the message “Keyboard Detected” on the POST Screen.

Universal Serial Bus (USB)

Native USB Support

The BIOS initializes and configures the USB subsystem during POST according to the Extensible Firmware Interface Reference Specification, Version 1.1. The BIOS is capable of initializing and using the following types of USB devices:

- USB Specification compliant keyboards
- USB Specification compliant mice
- USB Specification compliant storage devices utilizing the bulk-only transport mechanism USB devices are scanned to determine if they are required for booting.

The BIOS supports Universal Serial Bus Revision 2.0 Specification mode of operation, and as such supports Universal Serial Bus Revision 1.1 Specification and Universal Serial Bus

Revision 2.0 Specification compliant devices and host controllers.

During the pre-boot phase, the BIOS automatically supports the hot addition and hot removal of USB devices and a short beep is emitted to indicate such. For example, if a USB device is hot plugged, the BIOS detects the device insertion, initializes the device, and makes it available to the user. During POST, when the USB Controller is initialized, a short beep is emitted for each attached USB device (includes front panel hub).

Only on-board USB controllers are initialized by BIOS. This does not prevent the operating system from supporting any available USB controllers, including on add-in cards.

Legacy USB Support
The BIOS supports PS/2 emulation of USB keyboards and mice. During POST, the BIOS initializes and configures the root hub ports and then searches for a keyboard and/or a mouse on the USB hub and then enables them.

Serial ATA (SATA) Support
The BIOS supports and initializes SATA IDE devices only. Hot plugging SATA drives during the boot process is not supported by the BIOS and may result in undefined behavior.

Removable Media Drives
The BIOS supports removable media devices in accordance with the Tested Hardware and Operating System List.

The server system supports and initialize an ATAPI CD-ROM / DVD ROM optical drive through an on-board SATA to PATA converter. From the BIOS standpoint, the CD-ROM / DVD-ROM drive looks like a SATA drive and BIOS treats it as SATA device.

In addition, the BIOS supports USB mass storage devices such as CD-ROM / DVD-ROM drive, floppy drive, and USB flash drive devices. The BIOS supports booting from these removable devices.

The BIOS supports the Universal Serial Bus Revision 2.0 Specification compliant media storage devices that are backward compatible to the Universal Serial Bus Revision 1.1 Specification. The BIOS does not support ISA floppy drives.

Flash ROM
The BIOS supports the Intel® Advanced+ Boot Block Flash Memory 28F320C3B part of size 4MB. There are two such flash parts providing a combined flash size of 8MB. This capacity is used to store two BIOS images as required to support the Rolling BIOS feature. Each BIOS image contains:

- System initialization routines
- BIOS Setup utility
- Runtime support routines

The exact layout is subject to change as determined by Intel. A 64 KB block is available for

storing OEM custom logos.

The flash ROM contains the necessary drivers for onboard peripherals including the following:

- Ethernet
- SATA
- SAS
- Video controllers

The Flash Memory Update utility loads the BIOS image into the flash. The complete ROM is visible, starting at physical address 4 GB minus the size of the ROM, which on this server board is 8 MB. Due to shadowing, none of the flash blocks are visible at the aliased addresses below 1 MB.

Fan Speed Control and Thermal Management
The BIOS and BMC software work cooperatively to implement system thermal management support. This is accomplished with a combination of memory and processor thermal management as described in the sub-sections below.

FBDIMM Thermal Management
The BIOS implements support for Static Closed Loop Thermal Throttling (CLTT) in conformance with the requirements indicated in the Common Fan Speed Control and Thermal Management Platform Architecture Specification (PAS).

Fan Profile Option
BIOS Setup provides a fan profile option allowing the user to influence the system acoustic profile. This menu item provides two options:

- The performance option results in higher system performance at the expense of a slightly increased acoustic signature.
- The acoustic option results in a reduced acoustic signature using more aggressive memory throttling.

BMC Get Thermal Profile Data Command
The BIOS issues the BMC Get Thermal Profile Data command during early POST to retrieve the appropriate thermal profile as indicated by the user in BIOS Setup (performance or acoustic). If the BIOS cannot retrieve the thermal parameters from the BMC, it uses the Memory Reference Code (MRC) default settings for the chipset and the FBDIMM thermal throttling configuration.

Altitude information is combined with the thermal profile data to program the memory throttling characteristics of the chipset and FBDIMMs.

Table 40. Thermal Profile Data SDR Record Format

Byte		Name	Description
0:2		OEM ID	Intel manufacturers ID – 157h, little endian
3		Record Subtype	Value 0Bh
Thermal Profile Data Record	4	Throttling Mode	Indicates throttling mode. 01: Open Loop 02: Closed Loop.
	5	Profile Support Bitmap	[7:4] reserved (Initialized to 0) [3:0] Fan Control Profile Support. This is a bitmask that indicates which fan control profiles that this record is valid for. Bit 3 1 = valid for profile 3 Bit 2 1 = valid for profile 2 Bit 1 1 = valid for profile 1 Bit 0 1 = valid for profile 0
	6:23	Thermal Profile Data	Byte format defined by thermals team and/or BIOS. May be formalized later If the number of bytes in this field changes, the "Record Length" field and "_REC_LEN" tag above the SDR data must be adjusted to match. Values must be Hexadecimal. Data words must be represented in little endian. The maximum supported value of n is 63, for a maximum SDR body size of 64 bytes. This is the largest SDR size that can be retrieved over the KCS interfaces. Values of n larger than 23 will prevent the SDRs from being retrieved over IPMB, but this is not an expected usage model other than during development.
	6	TempInlet	Temperature at chassis inlet in units of 0.5 degrees Celsius (90 units = 45 degrees Celsius) applied to SYS.TempInlet, data format is hexadecimal.
	7	TempRise	Temperature rise from chassis inlet to FBDIMM local ambient in units of 0.5 degrees Celsius (90 units = 45 degrees Celsius) applied to SYS.TempRise, data format is hexadecimal.
	8:9	AirFlow	Average air flow velocity in dim channel in units of mm/sec (1500 units = 1.5 m/s) applied to SYS.AirFlow. Data format is in Hex, little endian.
	10:11	DimmPitch	The pitch between DIMMs in units of 1/1000 inch (410 units = 0.41 in) applied SYS.DimmPitch. Data format is in Hex, little endian.
	12:13	ThrtLoRatio	Ratio of BNB.THRTLOW / BNB.THRTHI in percentage. "0" = unlimited BW. Data format is in Hex, little endian.
	14:15	ThrtMidRatio	Ratio of BNB.THRTMID / BNB.THRTHI in percentage. "0" = unlimited BW. data format is in Hex, little endian.
	16	TempMidGb	Temperature guard band to AMB.TEMPMID in units of 0.5 degrees C (90 = 45 C) applied to SYS.TempMidGb (this is offset from the AMB max temp at max BW "Z"). Data format is hexadecimal.
	17	TempLoGb	Temperature guard band to AMB.TEMPLO in units of 0.5 degrees C (90 = 45 C) applied to SYS.TempLoGb ((this is offset from the AMB max temp at max BW "Z"). Data format is hexadecimal.
	18	GlobalActRatio	Ratio of BNB.GBLACT / BNB.THRTHI in percentage. "0" = unlimited BW, data format is hexadecimal.

Byte		Name	Description
	20	Bit[0] – GlobalThrtEn	"1" enables global throttling, GLOBTHRT_EN. Data format is hexadecimal.
		Bit[1] – ThermalThrtEn	"1" enables closed-loop throttling, THERMTHRT_EN. Data format is hexadecimal.
		Bit[2] – HighTempEn	"1" enables high DRAM temp operation when FBDIMM is capable, HIGHTEMP_EN. Data format is hexadecimal.
		Bits[3:7] – Reserved	Reserved for future use.
	21:24	Reserved	Reserved for future use.

Static Open Loop Thermal Throttling (OLTT) Operation

If for any reason the system cannot be successfully configured for Static CLTT operation, the

BIOS programs the system for Static Open Loop Thermal Throttling (OLTT) operation instead.

If the system faults to programming for OLTT, then an Altitude setup option is made configurable on the System Acoustic and Performance Configuration screen in BIOS setup. The BIOS subsequently issues the BMC Set Fan Control Configuration command to inform the BMC of the Fan Profile (acoustic or performance) selected by the user in BIOS Setup.

Processor Thermal Management
The processors implement a methodology for managing processor temperatures through processor throttling. There are two components to the temperature calculation used to regulate the processor temperature:

- Tcontrol offset (BIOS sends this value to BMC)
- Tcontrol base (BMC retrieves this value from SDR)

The BIOS retrieves the Tcontrol offset from a processor Model Specific Register (MSR) and sends this value to the Baseboard Management Controller (BMC) by issuing the BMC Set Processor Tcontrol command.

The BMC uses these two Tcontrol values to regulate processor thermal characteristics according to the user-selected Fan Profile

BIOS User Interface

Splash Logo / Diagnostic Screen

The BIOS displays one of two screens during POST:

- Splash logo
- Diagnostic screen

BIOS Setup provides the Quiet Boot option that controls splash logo display.

Splash Logo Screen

The BIOS displays the splash logo image during POST by default. The splash logo is a graphic image stored on the flash ROM in BMP format. Resolution up to 800 x 600 in any color depth is supported. A standard Intel Splash Logo is included in the flash ROM. An OEM can load a customized splash logo..
The BIOS may display the diagnostic screen instead of the splash logo screen if:

- The BIOS Setup Quiet Boot option is disabled.
- The BIOS cannot locate a splash logo bitmap file in the flash ROM.
- Remote terminal display when console redirection is enabled.
- User presses <Esc> or <TAB> during POST while the splash logo is displayed. This causes BIOS to immediately suppress the splash logo and display the diagnostic screen instead.

Diagnostic Screen
The diagnostic screen includes:

- BIOS ID.
- Platform name
- Total memory detected (total size of all installed FBDIMMs)
- Processor information (Intel branded string, speed, and number of physical processors)
- Flash bank used to boot the system
- Keyboard device(s) detected, if any
- Mouse device(s) detected, if any

Note: *Only USB keyboard and mouse devices are supported.*

15.2 **BIOS Setup Utility**

The BIOS Setup utility is a text-based utility that allows the user to configure the system and view current settings and environment information for the platform devices. The Setup utility controls the platform's built-in devices, the Boot Manager, and the Error Manager. The BIOS Setup utility interface consists of a number of pages or screens. Each page contains information or links to other pages. The Setup Advanced screen displays a list of general categories as links. These links lead to pages containing a specific category's configuration.

15.2.1 **Features**

- Localization: English only.
- BIOS Setup is functional via console redirection over various terminal emulation standards. This may limit some functionality for compatibility (e.g. usage of colors, certain keys or key sequences, or support of pointing devices).

Page Layout

The BIOS Setup page layout is divided into functional areas. Each occupies a specific area of the screen and supports a dedicated function. The following table lists and describes each functional area.

Table 41. BIOS Setup — Page Layout

Functional Area	Description
Title Bar	The title bar is located at the top of the screen and displays the title of the form (page) the user is currently viewing. It may also display navigational information.
Setup Item List	The Setup Item List is a set of controllable and informational items. Each item in the list occupies the left column of the screen. A Setup item may also open a new window with more options for that functionality on the board.
Item Specific Help Area	The Item Specific Help area is located on the right side of the screen and contains help text for the highlighted Setup Item. Help information may include the meaning and usage of the item, allowable values, effects of the options, etcetera.
Keyboard Command Bar	The Keyboard Command Bar is located at the bottom right of the screen and continuously displays help for keyboard special keys and navigation keys. The keyboard command bar is context-sensitive. It displays keys relevant to current page and mode.

Entering BIOS Setup

The BIOS displays a "Press <F2> to enter setup" message during POST. The message is displayed on

the POST diagnostic screen if Quiet Boot is disabled or under the Splash Screen if Quiet Boot is enabled.

Keyboard Commands

The bottom right portion of the Setup screen provides a list of commands that are used to navigate through the Setup utility. These commands are displayed at all times. The Keyboard Command Bar supports the following:

Table 42. BIOS Setup — Keyboard Command Bar

Key	Option	Description
<Enter>	Execute Command	The <Enter> key allows the user to display sub-menus and drop down lists as well as to select sub-fields for multi-valued features like time and date. Pressing the <Enter> key while a drop down list is displayed selects the currently highlighted option. The user is then returned to the parent menu.
<Esc>	Exit	The <Esc> key provides a mechanism for backing out of any field. Pressing <Esc> while a drop down list is displayed returns the user to the parent menu without making any changes. Pressing <Esc> while a sub-menu is displayed exits the sub-menu and returns the user to the parent menu. Pressing <Esc> in any major menu displays the following dialog box: Quit Without Saving? Yes No Selecting Yes and pressing <Enter> exits Setup without affecting any existing settings and continues booting the system. Selecting No and pressing <Enter> or pressing <Esc> clears the dialog box and returns to the major menu previously displayed.
	Select Item	The up and down arrow keys allow the user to scroll through menu items or drop down list items.
	Select Menu	The left and right arrow keys allow the user to scroll through top-level menus. These keys have no effect when a sub-menu or drop down list is displayed.
<Tab>	Select Field	The <Tab> key move between fields. For example, <Tab> can be used to move from hours to minutes in the time item in the main menu.
-	Change Value	The minus key allows the user to scroll through drop down list values in descending order without displaying the full list.
+	Change Value	The plus key allows the user to scroll through drop down list values in ascending order without displaying the full list. On 106-key Japanese keyboards, the plus key has a different scan code than the plus key on the other keyboard. However, it has the same effect.
<F9>	Setup Defaults	The <F9> key is a shortcut key for loading BIOS Setup factory default settings. The following dialog box is displayed in response to the <F9> key: Load Factory Defaults? Yes No Selecting Yes and pressing <Enter> loads all Setup fields to their factory default settings. Selecting No and pressing <Enter> or pressing <Esc> clears the dialog box without affecting any existing settings.
<F10>	Save and Exit	The <F10> key is a shortcut to Save and Exit Setup. The following dialog box is displayed in response to the <F10> key: Save configuration and reset? Yes No Selecting Yes and pressing <Enter> saves all changes and resets the system. Selecting No and pressing <Enter> or pressing <Esc> clears the dialog box without affecting any existing settings.

Menu Selection Bar
The Menu Selection Bar is located at the top of the BIOS Setup Utility screen and it displays the major menu selections available to the user. Use the left and right arrow keys to select a menu item. Some menus are hidden and become available by scrolling off the left or right of the current selections.

BIOS Setup Utility Screens
The sections below describe the screens available for the configuration of a server platform. In these sections, tables are used to describe the contents of each screen. These tables conform to the following guidelines:

- The text and values in the Setup Item, Options, and Help Text columns in the tables are displayed on the BIOS Setup utility screens.
- Bold text in the Options column of the tables indicates default values. These values are not displayed in bold on the Setup screen and are for reference only.
- The Comments column provides more information where it may be helpful. This information does not appear in the BIOS Setup utility screens.
- Information in the Options column enclosed in brackets (< >) using black text indicates information dependent on the system configuration that the user can modify. For example, the <Current Date> field is populated with the actual current date.
- Information in the Options column enclosed in brackets (< >) using light grey text indicates Information Only fields that the user cannot modify.

Information in square brackets ([ ]) indicates fields in which the user needs to type in text instead of selecting from a provided option.

Menu Selection Bar
The Menu Selection Bar is located at the top of the BIOS Setup Utility screen and it displays the major menu selections available to the user. Use the left and right arrow keys to select a menu item. Some menus are hidden and become available by scrolling off the left or right of the current selections.

BIOS Setup Utility Screens
The sections below describe the screens available for the configuration of a server platform. In these sections, tables are used to describe the contents of each screen. These tables conform to the following guidelines:

- The text and values in the Setup Item, Options, and Help Text columns in the tables are displayed on the BIOS Setup utility screens.
- Bold text in the Options column of the tables indicates default values. These values are not displayed in bold on the Setup screen and are for reference only.
- The Comments column provides more information where it may be helpful. This information does not appear in the BIOS Setup utility screens.
- Information in the Options column enclosed in brackets (< >) using black text indicates
- information dependent on the system configuration that the user can modify. For example, the <Current Date> field is populated with the actual current date.
- Information in the Options column enclosed in brackets (< >) using light grey text indicates Information Only fields that the user cannot modify.

- Information in square brackets ([ ]) indicates fields in which the user needs to type in text instead of selecting from a provided option.

Main Screen

The Main screen is the first screen displayed after entering BIOS Setup unless an error has occurred. If an error has occurred, the Error Manager screen is displayed instead.

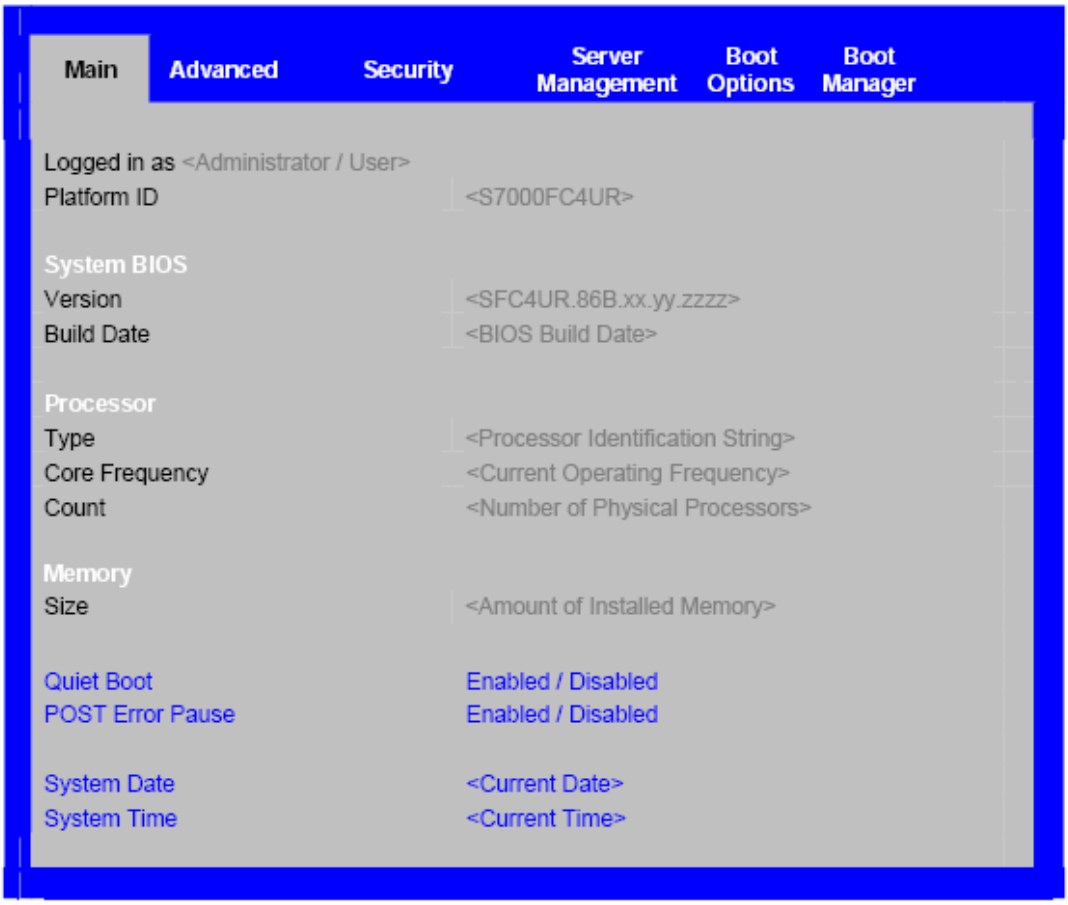


Figure 44. Setup Utility — Main Screen Display

Table 43. Setup Utility — Main Screen Fields

Setup Item	Options	Help Text	Comments
Logged in as <Administrator / User>			**Information only** Displays the security account used to enter BIOS Setup Utility. Administrator is the security account default if an Administrator password has not been set.
Platform ID	<S7000FC4UR>		**Information Only**
System BIOS Version	<SFC4UR.86B.xx.yy.zzzz>		**Information only** Displays the current BIOS version: ▪ xx = Major version ▪ yy = Minor version ▪ zzzz = Build number
System BIOS Build Date	<BIOS Build Date>		**Information only** Displays the current BIOS build date.
Processor Type	<Processor Identification String>		**Information only** Displays Intel processor name and the speed of the CPU. This information is retrieved from the processor.
Processor Core Frequency	<Current Operating Frequency>		**Information only** Displays the current speed of the boot processor in GHz or MHz
Processor Count	<Number of Physical Processors>		**Information only** Displays the number of physical processors detected.
Memory Size	<Amount of Installed Memory>		**Information only** Displays the total physical memory installed, in MB or GB. The term physical memory indicates the total memory discovered in the form of installed FBDIMMs.
Quiet Boot	**Enabled** Disabled	[Enabled] – Display the logo screen during POST. [Disabled] – Display the diagnostic screen during POST.	When the logo is displayed, no other data is viewable in POST, this means Option ROMS are also run under the logo and cannot be accessed.

Setup Item	Options	Help Text	Comments
POST Error Pause	Enabled **Disabled**	[Enabled] – System will enter the Error Manager for critical POST errors. [Disabled] – System will continue to boot bypassing the Error Manager for critical POST errors.	Determines whether the BIOS enters the POST Error Manager to display Major errors See Section 19.3.3 for more information.
System Date	[Day of Week MM/DD/YYYY]	System Date has configurable fields for Month, Day, and Year. Use [Enter] or [Tab] key to select the next field. Use [+] or [-] key to modify the selected field.	Valid Month values are 1 to 12. Valid Day values are 1 to 31. Valid Year values are 1998 to 2099.
System Time	[HH:MM:SS]	System Time has configurable fields for Hours, Minutes, and Seconds. Hours are in 24-hour format. Use [Enter] or [Tab] key to select the next field. Use [+] or [-] key to modify the selected field.	Valid Hours values are 0 to 23. Valid Minutes values are 0 to 59. Valid Seconds values are 0 to 59.

Advanced Screen
The Advanced screen provides the user several different sub-menus for various categories of configuration options.

From the Main screen, select Advanced to access this screen.

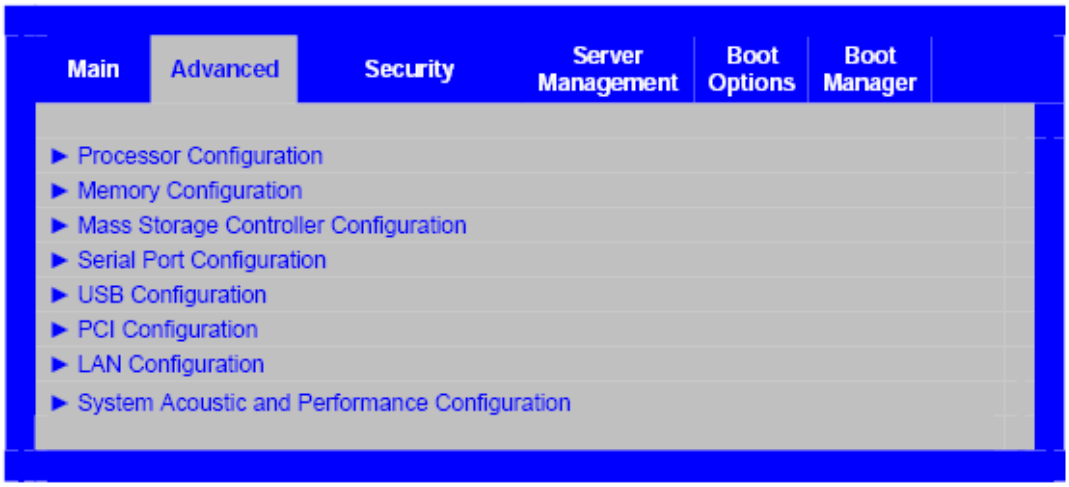


Figure 45. Setup Utility — Advanced Screen Display

15.2.3.2.1 **Processor Configuration Screen**
The Processor Configuration screen provides the ability for a user to view the processor core frequency, system bus frequency, and configure several processor options. The user can also
select an option to view information about any of the processors installed.

From the Main screen select Advanced | Processor Configuration to access this screen.

Figure 46. Setup Utility — Processor Configuration Screen Display

Table 44. Setup Utility — Processor Configuration Screen Fields

Setup Item	Options	Help Text	Comments
Core Frequency	<Current Processor Frequency>		**Information only** Frequency at which processors currently run
System Bus Frequency	<Current FSB Frequency>		**Information only** Current frequency of the processor front side bus

Setup Item	Options	Help Text	Comments
Enhanced Intel SpeedStep® Tech	**Enabled** Disabled	Enhanced Intel SpeedStep® Technology allows the system to dynamically adjust processor voltage and core frequency, which can result in decreased average power consumption and decreased average heat production. Contact your OS vendor regarding OS support of this feature.	
Intel® Virtualization Technology	Enabled **Disabled**	Intel® Virtualization Technology allows a platform to run multiple operating systems and applications in independent partitions. Note: A change to this option requires the system to be powered off and then back on before the setting takes effect.	
Execute Disable Bit	**Enabled** Disabled	Execute Disable Bit can help prevent certain classes of malicious buffer overflow attacks. Contact your OS vendor regarding OS support of this feature.	
Hardware Prefetcher	**Enabled** Disabled	Hardware Prefetcher is a speculative prefetch unit within the processor(s). Note: Modifying this setting may affect system performance.	
Adjacent Cache Line Prefetch	**Enabled** Disabled	[Enabled] – Cache lines are fetched in pairs (even line + odd line) [Disabled] – Only the current cache line required is fetched. Note: Modifying this setting may affect system performance.	
Processor Retest	Enabled **Disabled**	Activate and retest all processors during next boot only. Note: This option will automatically reset to [Disabled] on the next boot, after all processors are retested.	Since no Processors can be disabled in this system, this option is only used to clear a persistent error message (i.e.: Thermal Trip).
Processor <n> Information		View Processor <n> information.	To view information about processor select <n>. (Takes the user to a sub-menu screen) This option is repeated for each physical processor socket.

Processor <n> Information Screen

The Processor # Information screen provides the ability for a user to view information about a specific processor.

From the Main screen select Advanced | Processor Configuration | Processor # to access this screen, where # is the number of the processor you want to view.

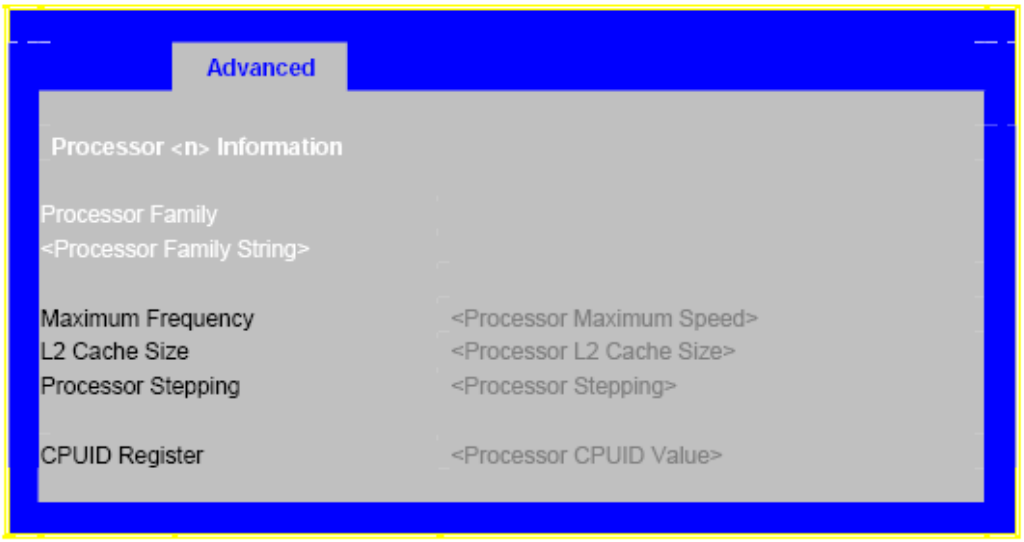


Figure 47. Setup Utility — Specific Processor Information Screen Display

Table 45. Setup Utility — Specific Processor Information Screen Fields

Setup Item	Options	Help Text	Comments
Processor Family <Processor Family String>			**Information only** Identifies family or generation of the processor
Maximum Frequency	<Processor Maximum Speed>		**Information only** Maximum frequency supported by the processor core.
L2 Cache Size	<Processor L2 Cache Size>		**Information only** Size of the processor L2 cache
Processor Stepping	<Processor Stepping>		**Information only** Stepping number of the processor
CPUID Register	<Processor CPUID Value>		**Information only** CPUID register value identifies details about the processor family, model, and stepping.

Memory Configuration Screen

The Memory Configuration screen provides the ability for a user to view details about system memory configuration. The user can also select options to open the Configure and View Memory RAS screen or the memory riser board Information screens.

From the Main screen select Advanced | Memory Configuration to access this screen.

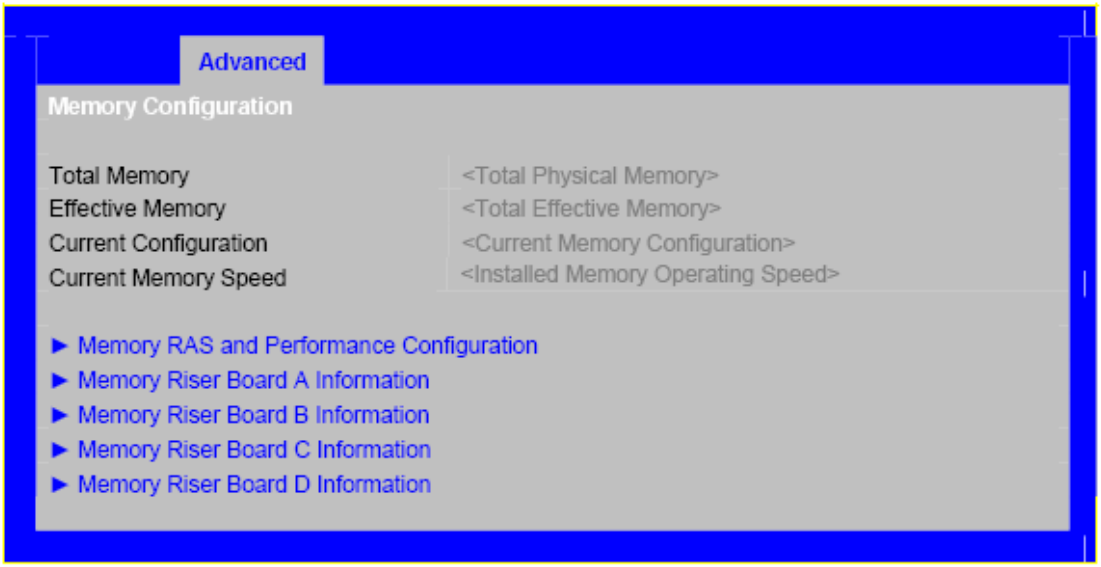


Figure 48. Setup Utility — Memory Configuration Screen Display

Table 46. Setup Utility — Memory Configuration Screen Fields

Setup Item	Options	Help Text	Comments
Total Memory	<Total Physical Memory>		**Information only** The amount of memory available in the form of installed FBDIMMs, in units of MB or GB.
Effective Memory	<Total Effective Memory>		**Information only** The amount of memory available to the operating system in MB or GB The Effective Memory is the difference between Total Physical Memory and the sum of all memory reserved for internal usage, RAS redundancy and SMRAM. This difference includes the sum of all FBDIMMs that failed MemBIST during POST, or were disabled by the BIOS during memory discovery phase in order to optimize memory configuration.

Setup Item	Options	Help Text	Comments
Current Configuration	<Current Memory Configuration>		**Information only** Displays one of the following: ▪ Maximum Performance Mode: System memory is configured for optimal performance and efficiency. No RAS features are enabled. ▪ Single Channel Mode: System memory is functioning in a reduced efficiency failsafe mode. ▪ Memory Mirroring Mode: System memory is configured for maximum reliability in the form of memory mirroring. ▪ Dual-DIMM Sparing Mode: System memory is configured for optimal performance and efficiency. Sparing is also enabled.
Current Memory Speed	<Installed Memory Operating Speed>		**Information only** Displays the speed the memory is currently running at: ▪ 533 MT/s (266MHz): System memory is configured to operate at 266MHz frequency. ▪ 667 MT/s (333MHz): System memory is configured to operate at 333MHz frequency.
Memory RAS and Performance Configuration		Configure memory RAS (Reliability, Availability, and Serviceability) and view current memory performance information and settings.	Select to configure Memory RAS and performance. Takes the user to a sub-menu screen.
Memory Riser Board <n> Information		View Memory Riser Board and associated FB-DIMM information.	Select to view information about the specific memory riser board. Takes the user to a sub-menu screen.

Configure Memory RAS and Performance Screen

The Configure Memory RAS and Performance screen provides fields to customize several memory configuration options, such as whether to use Memory Mirroring or Memory Sparing. From the Main screen select Advanced | Memory Configuration | Configure Memory RAS and Performance Configuration to access this screen.

Figure 49. Setup Utility — Configure and View Memory RAS Screen Display

Table 47. Setup Utility — Configure and View Memory RAS Screen Fields

Setup Item	Options	Help Text	Comments
Memory Mirroring Possible	<Yes / No>		**Information only** Only displayed on systems with chipsets that are capable of Memory Mirroring
Memory Sparing Possible	<Yes / No>		**Information only**
Select Memory RAS Configuration	**RAS Disabled** Sparing Mirroring	Available modes depend on the current memory population. [RAS Disabled] – Optimizes system performance [Mirroring] - Optimizes reliability by using half of physical memory as a backup [Sparing] – Improves reliability by reserving memory for use as a replacement in the event of DIMM failure	Provides options for configuring Memory RAS. This option is only displayed if the current layout and positioning of the FBDIMMs on the board supports one or more Memory RAS modes. The BIOS dynamically configures the drop-down options to display only those RAS features that can currently be supported. The possible options for this menu item are: ▪ RAS Disabled: The default in normal mode of operation. In this mode, no Memory RAS is supported. ▪ Sparing: Available and displayed only when the FBDIMM population can support memory sparing.

Setup Item	Options	Help Text	Comments
			▪ Mirroring: Available and displayed only when the FBDIMM population is capable of supporting memory mirroring. When this option is available and selected, the BIOS reconfigures memory in the mirroring mode on the next boot.
Snoop Filter	**Enabled** Disabled	The Snoop Filter component monitors and controls the data transactions between memory and the processor(s).	

Memory Riser Board <n> Information Screens

The memory riser board Information screens provide the ability for a user to view details about the memory riser boards and associated FBDIMMs installed.

From the Main screen select Advanced | Memory Configuration | Memory Riser Board <n> Information to access these screens.

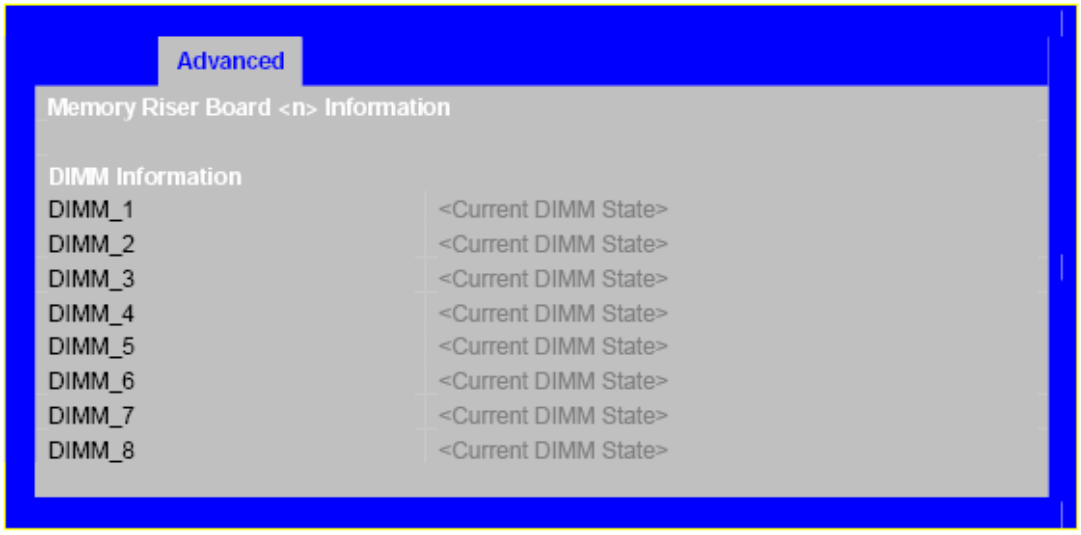


Figure 50. Setup Utility — Memory Riser Board Information Screen Display

Table 48. Setup Utility — Memory Riser Board Information Screen Fields

Setup Item	Options	Help Text	Comments
DIMM_#	<Current DIMM State>		**Information only** Displays the state of each DIMM socket present on the board. Each DIMM socket field reflects one of the following possible states: ▪ Installed: There is a FBDIMM installed in this slot. ▪ Not Installed: No FBDIMM is installed in this slot. ▪ Disabled: The FBDIMM installed in this slot has been disabled by the BIOS in order to optimize memory configuration. ▪ Failed: The FBDIMM installed in this slot is faulty or malfunctioning. ▪ Spare Unit: The FBDIMM is functioning as a spare unit for Memory RAS purposes.

Mass Storage Controller Configuration Screen

The Mass Storage Controller Configuration screen provides configuration options and informational for Serial Attached SCSI (SAS) and Serial ATA (SATA) devices.

From the Main screen select Advanced | Mass Storage Controller Configuration to access this screen.

Table 49. Setup Utility — Mass Storage Controller Configuration Screen Fields

Setup Item	Options	Help Text	Comments
AHCI Option ROM	**Enabled** Disabled	Enable or Disable the onboard Advanced Host Controller Interface (AHCI) option ROM. Note: For AHCI capability in EFI, the AHCI Legacy Option ROM should be set to [Disabled].	
SAS Option ROM	**Enabled** Disabled	Enable or Disable the onboard Serial Attached SCSI (SAS) Controller option ROM.	This option should only be displayed when the SAS riser board is installed.
SATA Mode	**IDE** AHCI SW RAID	[IDE] – Supports up to 4 SATA ports with Parallel ATA emulation [AHCI] – Supports all SATA ports using the Advanced Host Controller Interface [SW RAID] – Supports configuration of SATA ports for RAID via RAID configuration software.	Allows the user to configure the system as follows: ▪ IDE: SATA Legacy IDE compatibility mode emulates a traditional Parallel ATA (IDE) based configuration by reporting up to four SATA devices to the user and the operating system as if they were IDE / Parallel ATA devices. The devices attached to SATA ports 0 and 2 are reported as IDE Primary channel master and slave devices. The devices attached to SATA ports 1 and 3 are reported as IDE Secondary channel master and slave devices. ▪ AHCI: The Advanced Host Controller Interface (AHCI) Option ROM identifies and configures AHCI devices on all six SATA ports. Note: The AHCI Option ROM supports only Hard Disk Drives and CD-ROM devices. Other types of devices are not supported by the AHCI Option ROM (i.e. tape drives). If the connected device is not supported, an error message is displayed informing the user that the AHCI OPROM cannot handle this device, and the device is not accessible in pre-OS. Once the system boots to an OS, OS drivers must be installed for these devices to be functional during OS runtime. ▪ SW RAID: The RAID for Serial ATA Option ROM enumerates and configures SATA devices as a Redundant Array of Independent Disks (RAID). BIOS Setup does not report device information for any drives configured by the Option ROM as part of a RAID volume.

Setup Item	Options	Help Text	Comments
SATA Port 0	< Not Installed / Drive Information >		**Information only** This field displays information for the device connected to SATA Port 0. This field is not displayed if the device has been configured as part of a RAID volume.
SATA Port 1	< Not Installed / Drive Information >		**Information only** This field displays information for the device connected to SATA Port 1. This field is not displayed if the device has been configured as part of a RAID volume.
SATA Port 2	< Not Installed / Drive Information >		**Information only** This field displays information for the device connected to SATA Port 2. This field is not displayed if the device has been configured as part of a RAID volume.
SATA Port 3	< Not Installed / Drive Information >		**Information only** This field displays information for the device connected to SATA Port 3. This field is not displayed if the device has been configured as part of a RAID volume.
SATA Port 4	< Not Installed / Drive Information >		**Information only** This field displays information for the device connected to SATA Port 4. This field is not displayed if the SATA controller has been configured for Legacy IDE emulation or if the device has been configured as part of a RAID volume.
SATA Port 5	< Not Installed / Drive Information >		**Information only** This field displays information for the device connected to SATA Port 5. This field is not displayed if the SATA controller has been configured for Legacy IDE emulation or if the device has been configured as part of a RAID volume.

Serial Port Configuration Screen

The Serial Port Configuration screen provides configuration options for Serial Port A and Serial Port B.

From the Main screen select Advanced | Serial Port Configuration to access this screen.

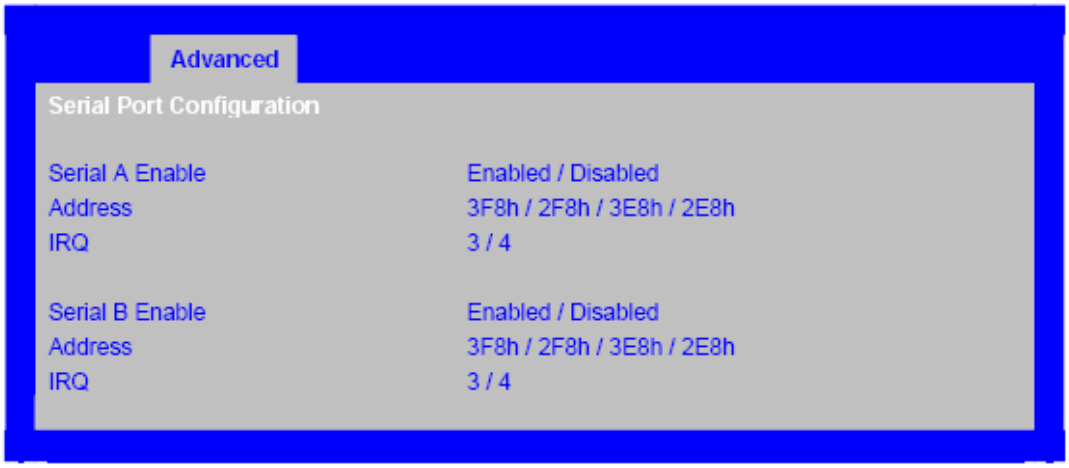


Figure 52. Setup Utility — Serial Port Configuration Screen Display

Table 50. Setup Utility — Serial Ports Configuration Screen Fields

Setup Item	Options	Help Text	Comments
Serial A Enable	**Enabled** Disabled	Enable or Disable serial port A.	
Address	**3F8h** 2F8h 3E8h 2E8h	Select serial port A base I/O address.	
IRQ	3 **4**	Select serial port A interrupt request (IRQ) line.	
Serial B Enable	**Enabled** Disabled	Enable or Disable serial port B.	Serial B is no longer available as a serial port when SOL or EMP mode are in effect.
Address	3F8h **2F8h** 3E8h 2E8h	Select serial port B base I/O address.	
IRQ	**3** 4	Select serial port B interrupt request (IRQ) line.	

Note: *Serial ports cannot be assigned identical I/O Addresses or IRQ assignments. BIOS Setup does not allow the user to exit this screen if both Serial Ports assigned the same I/O Address or IRQ values.*

USB Configuration Screen

The USB Configuration screen provides configuration options for the USB host controllers.

From the Main screen select Advanced | USB Configuration to access this screen.

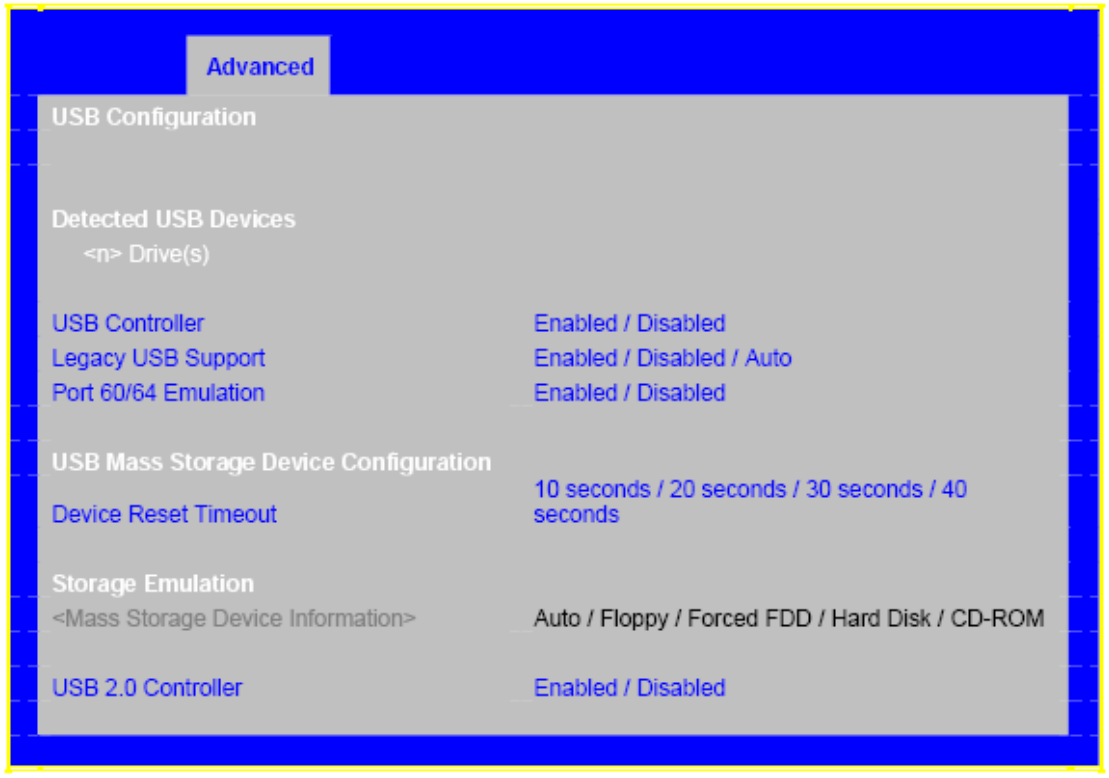


Figure 53. Setup Utility — USB Controller Configuration Screen Display

Table 51. Setup Utility — USB Controller Configuration Screen Fields

Setup Item	Options	Help Text	Comments
Detected USB Devices			Information only Shows the number of USB devices.
USB Controller	Enabled Disabled	[Enabled] – All onboard USB controllers will be turned on and accessible by the OS. [Disabled] – All onboard USB controllers will be turned off and inaccessible by the OS.	
Legacy USB Support	Enabled Disabled Auto	USB device boot support and PS/2 emulation for USB keyboard and USB mouse devices. [Auto] – Legacy USB support will be enabled if a USB device is attached.	
Port 60/64 Emulation	Enabled Disabled	I/O port 60h/64h emulation support. Note: This may be needed for legacy USB keyboard support when using an OS that is USB unaware.	
Device Reset Timeout	10 seconds 20 seconds 30 seconds 40 seconds	USB mass storage device start unit command timeout. Setting to a larger value provides more time for a mass storage device to be ready, if needed.	

Table 51. Setup Utility — USB Controller Configuration Screen Fields

Setup Item	Options	Help Text	Comments
Detected USB Devices			Information only Shows the number of USB devices.
USB Controller	Enabled Disabled	[Enabled] – All onboard USB controllers will be turned on and accessible by the OS. [Disabled] – All onboard USB controllers will be turned off and inaccessible by the OS.	
Legacy USB Support	Enabled Disabled Auto	USB device boot support and PS/2 emulation for USB keyboard and USB mouse devices. [Auto] – Legacy USB support will be enabled if a USB device is attached.	
Port 60/64 Emulation	Enabled Disabled	I/O port 60h/64h emulation support. Note: This may be needed for legacy USB keyboard support when using an OS that is USB unaware.	
Device Reset Timeout	10 seconds 20 seconds 30 seconds 40 seconds	USB mass storage device start unit command timeout. Setting to a larger value provides more time for a mass storage device to be ready, if needed.	

<Mass Storage Device Information>	Auto Floppy Forced FDD Hard Disk CD-ROM	[Auto] – USB devices less than 530 MB will be emulated as floppy. [Forced FDD] – HDD formatted drive will be emulated as FDD (e.g. ZIP drive).	A separate line is displayed for each USB mass storage emulation device detected. This screen can show a maximum of eight devices. If more than eight devices are installed, the 'USB Devices Enabled' shows the correct count, but only the first eight devices can be displayed here. Setting the device to a specific setting assumes that your device is formatted correctly to operate as the choosen device, else the boot from that device fails: [Floppy] attempts to boot the device as a floppy drive. [Forced FDD] emulates a FDD boot when device is formatted as a HDD. This option works only for drives formatted with FAT12, FAT16 or FAT32, and have a single partition on the device. [Hard Disk] boots as a hard disk, This option works only for drives formatted with FAT12, FAT16 or FAT32, [CD-ROM] boots as specified by the "El Torito" Bootable CD-ROM Format Specification Version 1. Once one of these settings are chosen, the system remembers that setting even when the device is removed and a different device is inserted in its place. The user must reset this option to Auto if they wish the device to be auto detected.
USB 2.0 Controller	Enabled Disabled	Onboard USB ports will be enabled to support USB 2.0 mode. Contact your OS vendor regarding OS support of this feature.	

PCI Configuration Screen

The PCI Configuration screen provides configuration options for PCI devices including PCI adapters and embedded devices.

From the Main screen select Advanced | PCI Configuration to access this screen.

Figure 54. Setup Utility — PCI Configuration Screen Display

Table 52. Setup Utility — PCI Configuration Screen Fields

Setup Item	Options	Help Text	Comments
Memory Mapped I/O Start Address	1.5GB 1.75GB **2.00GB** 2.25GB 2.5GB	Select the start of the reserved memory region for PCI memory mapped I/O space that ends at 4GB. Warning: Depending on the system configuration, this option may impact the amount of system memory detected by an OS without Physical Address Extension (PAE) support.	For all PAE (Physical Address Extension) aware operating systems, 2.0GB should be selected. The system remaps memory and the operating system detects all memory installed. If the installed operating system does not support PAE, the maximum physical memory size detected is equal to the value selected for this Setup option. For example, if 1.0GB is selected then only 1.0GB of physical memory is detected and reported by the operating system.
Memory Mapped I/O above 4GB	Enabled **Disabled**	Enable or disable memory mapped I/O of 64-bit PCI devices to 4GB or greater address space.	
Onboard Video	**Enabled** Disabled	Onboard video controller Warning: System video will be completely disabled if this option is disabled and an add-in video adapter is not installed.	When disabled, the system requires an add-in video card in order for video to be seen.
Dual Monitor Video	Enabled **Disabled**	Both the onboard video controller and an add-in video adapter will be enabled for system video. The onboard video controller will be the primary video device.	
Slot <x> ROM	**Enabled** Disabled	Controls execution of the add-in adapter option ROM during POST This setting only takes effect if an adapter with an option ROM is installed in the slot. Warning: If [Disabled] is selected, the adapter may not be used to boot the system.	One Setup option is provided for each of the seven PCI Express* slots. Controls dispatching of the Option ROM for the PCI Express* adapter in the slot

LAN Configuration Screen

The LAN Configuration screen provides configuration options for LAN controllers on the main board and I/O riser board.

From the Main screen select Advanced | LAN Configuration to access this screen.

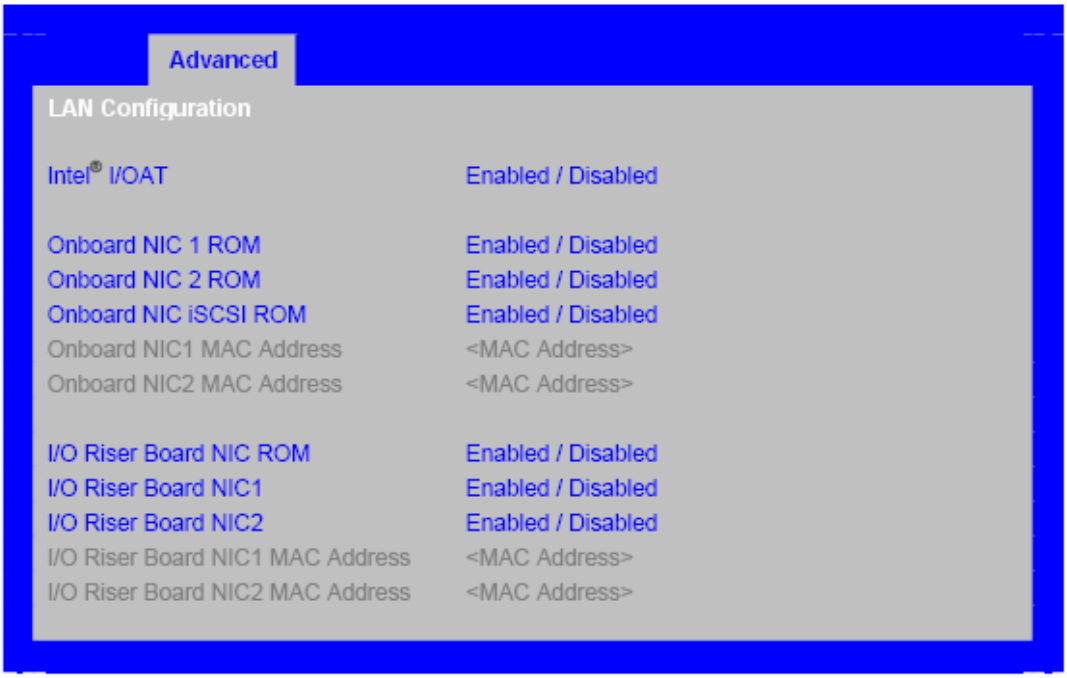


Figure 55. Setup Utility — LAN Configuration Screen Display

Table 53. Setup Utility — LAN Configuration Screen Fields

Setup Item	Options	Help Text	Comments
Intel® I/OAT	**Enabled** Disabled	Intel® I/O Acceleration Technology (Intel® I/OAT) accelerates TCP/IP processing for onboard NICs, delivers data-movement efficiencies across the entire server platform, and minimizes system overhead.	

Setup Item	Options	Help Text	Comments
Onboard NIC 1 ROM	**Enabled** Disabled	Load the embedded option ROM for the onboard network controller. Warning: If [Disabled] is selected, NIC1 can not be used to boot or wake the system.	This corresponds to the main board Intel® 82563EB Ethernet device.
Onboard NIC 2 ROM	**Enabled** Disabled	Load the embedded option ROM for the onboard network controller. Warning: If [Disabled] is selected NIC2 can not be used to boot or wake the system.	This corresponds to the main board Intel® 82563EB Ethernet device.
Onboard NIC iSCSI ROM	Enabled **Disabled**	Load the embedded Internet SCSI option ROM for the onboard network controller.	This corresponds to the main board Intel 82563EB Ethernet device.
Onboard NIC1 MAC Address	<MAC Address>		**Information only** This corresponds to the main board Intel 82563EB Ethernet device. 12 hex digits of the MAC address.
Onboard NIC2 MAC Address	<MAC Address>		**Information only** This corresponds to the main board Intel 82563EB Ethernet device. 12 hex digits of the MAC address.
I/O Riser Board NIC ROM	**Enabled** Disabled	Load the embedded option ROM for the I/O Riser Board network controllers. Warning: If [Disabled] is selected, the I/O Riser Board NIC1 and NIC2 can not be used to boot or wake the system.	This corresponds to the Intel I/O riser board Intel 82575EB Ethernet device. This menu item should be suppressed if the I/O riser board is not installed.
I/O Riser Board NIC1	**Enabled** Disabled	Enables or disables the I/O Riser Board network controller.	This corresponds to the Intel I/O riser board Intel 82575EB Ethernet device. This menu item should be suppressed if the I/O riser board is not installed. If the user selects disabled, the BIOS hides the PCI configuration space for this device and the device is disabled.

Setup Item	Options	Help Text	Comments
I/O Riser Board NIC2	**Enabled** Disabled	Enables or disables the I/O Riser Board network controller.	This corresponds to the Intel I/O riser board Intel 82575EB Ethernet device. This menu item should be suppressed if the I/O riser board is not installed. If the user selects disabled, the BIOS hides the PCI configuration space for this device and the device is disabled.
I/O Riser Board NIC1 MAC Address	<MAC Address>		**Information only** This corresponds to the Intel I/O riser board Intel 82575EB Ethernet device. This menu item should be suppressed if the I/O riser board is not installed. 12 hex digits of the MAC address.
I/O Riser Board NIC2 MAC Address	<MAC Address>		**Information only** This corresponds to the Intel I/O riser board Intel 82575EB Ethernet device. This menu item should be suppressed if the I/O riser board is not installed. 12 hex digits of the MAC address.

System Acoustic and Performance Configuration

The System Acoustic and Performance Configuration screen provides configuration options for system thermal characteristics and behavior.

From the Main screen select Advanced | System Acoustic and Performance Configuration to access this screen.

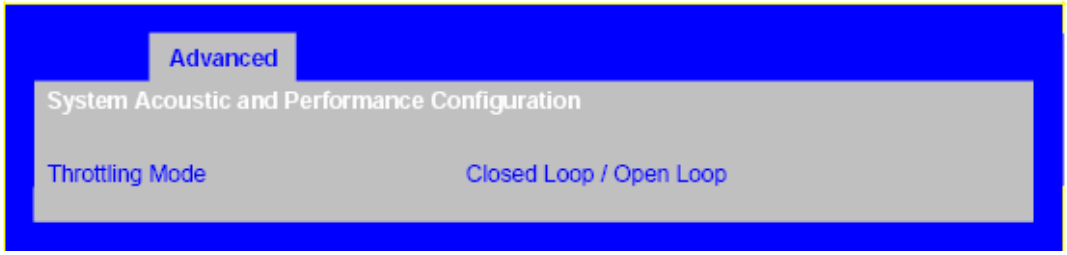


Table 54. Setup Utility — System Acoustic and Performance Configuration Screen Fields

Setup Item	Options	Help Text	Comments
Throttling Mode	**Closed Loop** Open Loop	Open Loop does not rely on a thermal sensor on the board and sets up a static level which equates to a fixed bandwidth. Closed Loop will allow the system to achieve higher performance by monitoring system temps and adjusting bandwidth.	Closed Loop is CLTT mode, Open Loop is OLTT mode.

Security Screen

The Security screen provides configuration options for BIOS Security features. This screen allows the user to set administrative and/or user passwords and to lockout front panel buttons so they cannot be used.

From the Main screen select Security to access this screen.

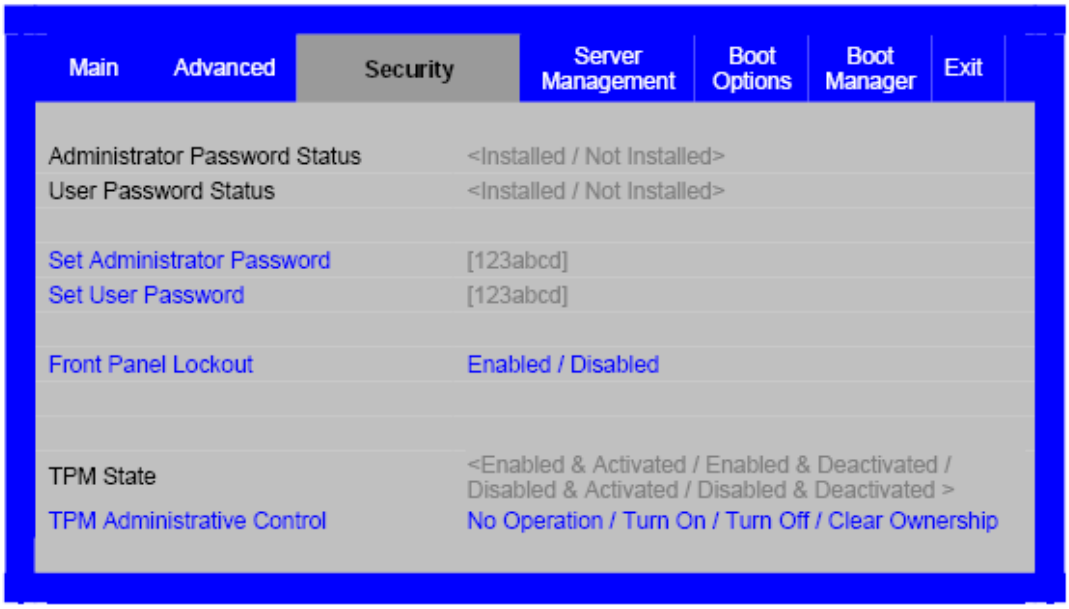


Figure 57. Setup Utility — Security Configuration Screen Display

Table 55. Setup Utility — Security Configuration Screen Fields

Setup Item	Options	Help Text	Comments
Administrator Password Status	<Installed / Not Installed>		**Information only** Indicates the status of the administrator password
User Password Status	<Installed / Not Installed>		**Information only** Indicates the status of the user password
Set Administrator Password	[123abcd]	The Administrator password is used to control change access in BIOS Setup Utility. Only alphanumeric characters can be used. Maximum length is 7 characters. It is case sensitive. Note: Administrator password must be set in order to use the user account.	This option is only to control access to BIOS Setup. The administrator has full access to all BIOS Setup items. Clearing the administrator password clears the user password, too.

Setup Item	Options	Help Text	Comments
Set User Password	[123abcd]	User password is used to control entry access to BIOS Setup Utility. Only alphanumeric characters can be used. Maximum length is 7 characters. It is case sensitive. Note: Removing the administrator password will also automatically remove the user password,.	Available only if the Administrator Password is installed This option is only to control access to BIOS Setup. The user password provides only limited access to BIOS Setup configuration options.
Front Panel Lockout	Enabled **Disabled**	Locks the power button and reset button on the system's front panel. If [Enabled] is selected, power and reset must be controlled via a system management interface.	
TPM State	< Enabled & Activated, Enabled & Deactivated, Disabled & Activated, Disabled & Deactivated >		**Information Only.** Shows the current TPM device state. A disabled TPM device does not execute commands that use TPM functions and TPM security operations are not available. An enabled and deactivated TPM is in the same state as a disabled TPM except setting of TPM ownership is allowed if not present already. An enabled and activated TPM executes all commands that use TPM functions and TPM security operations are available.
TPM Administrative Control	**No Operation** Turn On Turn Off Clear Ownership	[No Operation] – No changes to current state. [Turn On] – Enables and activates TPM. [Turn Off] – Disables and deactivates TPM. [Clear Ownership] – Removes the TPM ownership authentication and returns the TPM to factory default state. Note: BIOS setting will return to [No Operation] on every boot cycle by default.	

Server Management Screen

The Server Management screen provides configuration options for several server management features. It also provides an access point to the screens for configuring console redirection and displaying server management related system information.

From the Main screen select Server Management to access this screen.

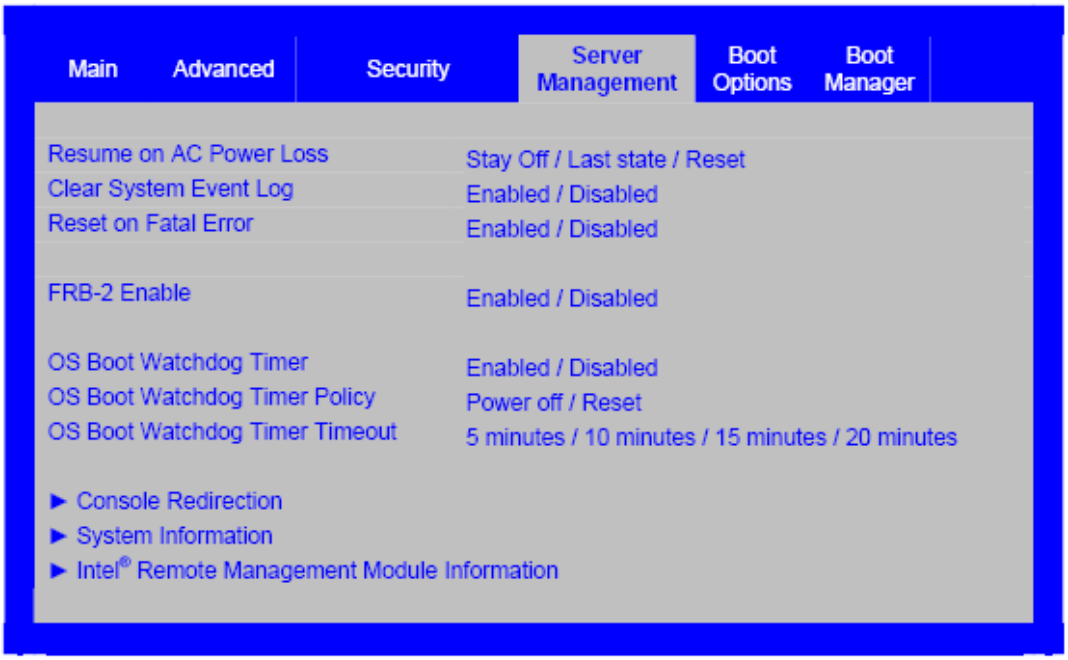


Figure 58. Setup Utility — Server Management Configuration Screen Display

Table 56. Setup Utility — Server Management Configuration Screen Fields

Setup Item	Options	Help Text	Comments
Resume on AC Power Loss	**Stay Off** Last state Reset	System action to take on AC power loss recovery. [Stay Off] – System stays off. [Last State] – System returns to the same state before the AC power loss. [Reset] – System powers on.	
Clear System Event Log	Enabled **Disabled**	Clears the System Event Log. All current entries are lost. Note: This option will be reset to [Disabled] after a reboot.	

Setup Item	Options	Help Text	Comments
Reset on Fatal Error	Enabled **Disabled**	[Enabled] – System will trigger a reset on fatal errors. [Disabled] – System will trigger a Non-Maskable Interrupt on fatal errors.	The system normally generates a NMI in response to fatal errors at runtime. The operating system typically halts the system in response to NMI thereby containing potential data corruption. The user can configure this option to Enabled for Operating Systems that do not respond to an NMI with a system halt. This ensures the corrupted data is contained by the reset.
FRB-2 Enable	**Enabled** Disabled	Fault Resilient Boot (FRB). BIOS programs the BMC watchdog timer for approximately 6 minutes. If BIOS does not complete POST before the timer expires, the BMC will reset the system.	
OS Boot Watchdog Timer	Enabled **Disabled**	BIOS programs the watchdog timer with the timeout value selected. If the OS does not complete booting before the timer expires, the BMC will reset the system and an error is logged. Requires OS support or Intel® System Management Software.	
OS Boot Watchdog Timer Policy	**Power Off** Reset	If the OS watchdog timer is enabled, this is the system action taken if the watchdog timer expires. [Reset] – System performs a reset. [Power Off] – System powers off.	
OS Boot Watchdog Timer Timeout	5 minutes **10 minutes** 15 minutes 20 minutes	If the OS watchdog timer is enabled, this is the timeout value BIOS will use to configure the watchdog timer.	
Console Redirection		View/Configure console redirection information and settings.	Takes user to Console Redirection screen.
System Information		View system information.	Takes user to System Information screen.
Intel® Remote Management Module Information		View Intel® RMM2 information.	Takes user to Intel® Remote Management Module Information screen. This menu item should be suppressed if the I/O riser board is not installed.

Console Redirection Screen

The Console Redirection screen provides configuration options for console redirection and associated connectivity options.

From the Main screen select Server Management | Console Redirection to access this screen.

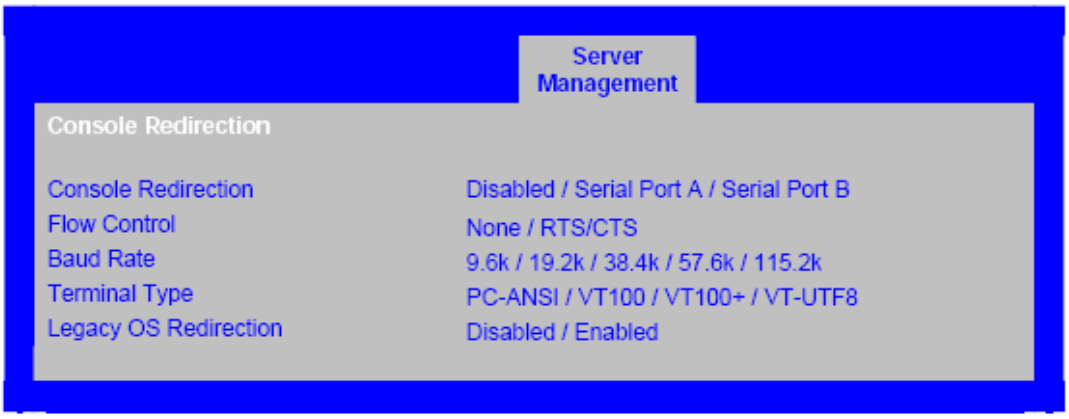


Figure 59. Setup Utility — Console Redirection Screen Display

Table 57. Setup Utility — Console Redirection Configuration Fields

Setup Item	Options	Help Text	Comments
Console Redirection	**Disabled** Serial Port A Serial Port B	Console redirection allows a serial port to be used for server management tasks. [Disabled] – No console redirection [Serial Port A] – Configure serial port A for console redirection. [Serial Port B] – Configure serial port B for console redirection.	
Flow Control	**None** RTS/CTS	Flow control is the handshake protocol. Setting must match the remote terminal application. [None] – Configure for no flow control. [RTS/CTS] – Configure for hardware flow control.	
Baud Rate	9.6K 19.2K 38.4K 57.6K **115.2K**	Serial port transmission speed: Setting must match the remote terminal application.	

Setup Item	Options	Help Text	Comments
Terminal Type	PC-ANSI **VT100** VT100+ VT-UTF8	Character formatting used for console redirection. Setting must match the remote terminal application.	
Legacy OS Redirection	**Disabled** Enabled	This option will enable legacy OS redirection (i.e. DOS) on serial port. If it is enabled the associated serial port is hidden from the legacy OS.	

Server Management System Information Screen

The Server Management System Information screen provides options to review information regarding part numbers, serial numbers, and firmware revisions.

From the Main screen select Server Management | System Information to access this screen.

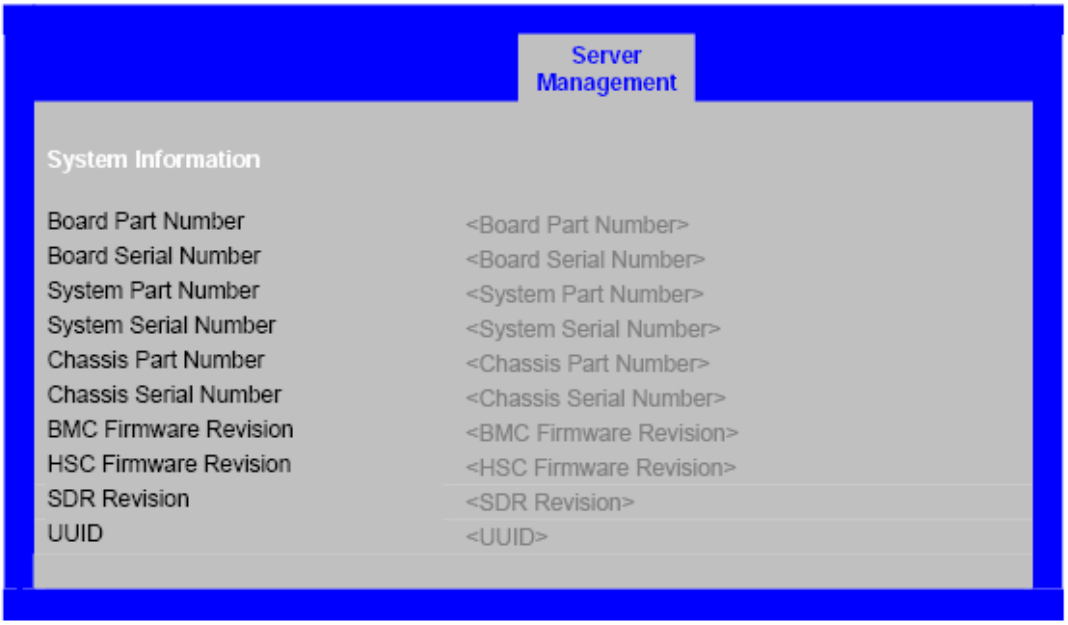


Figure 60. Setup Utility — Server Management System Information Screen Display

Table 58. Setup Utility — Server Management System Information Fields

Setup Item	Options	Help Text	Comments
Board Part Number	<Board Part Number>		**Information only**
Board Serial Number	<Board Serial Number>		**Information only**
System Part Number	<System Part Number>		**Information only**
System Serial Number	<System Serial Number>		**Information only**
Chassis Part Number	<Chassis Part Number>		**Information only**
Chassis Serial Number	<Chassis Serial Number>		**Information only**
BMC Firmware Revision	<BMC Firmware Revision>		**Information only**
HSC Firmware Revision	<HSC Firmware Revision>		**Information only**
SDR Revision	<SDR Revision>		**Information only**
UUID	<UUID>		**Information only**

Intel® Remote Management Module Information Screen

The Intel® Remote Management Module Information screen provides options to review information regarding firmware revisions and Intel GCM network devices.

From the Main screen select Server Management | Intel® Remote Management Module Information to access this screen.

Note: *This sub-menu should only be displayed if the I/O riser board is installed.*

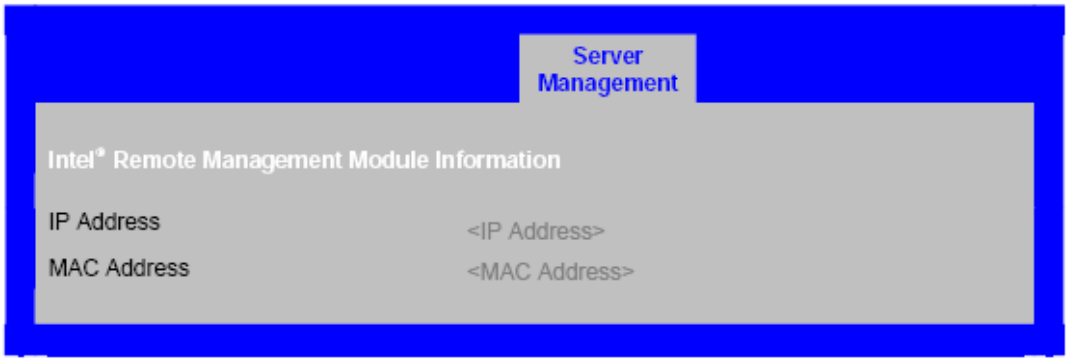


Figure 61. Setup Utility — Intel® Remote Management Module Information Screen Display

Table 59. Setup Utility — Intel® Remote Management Module Information Fields

Setup Item	Options	Help Text	Comments
IP Address	<IP Address>		**Information only** Displays the Intel GCM3 Ethernet device IP Address information This is obtained from BMC.
MAC Address	<MAC Address>		**Information only** Displays the Intel GCM3 Ethernet device MAC Address information This is obtained from BMC.

Boot Options Screen

The Boot Options screen displays any bootable media encountered during POST and allows the user to configure their desired boot device.

From the Main screen, select Boot Options to access this screen.

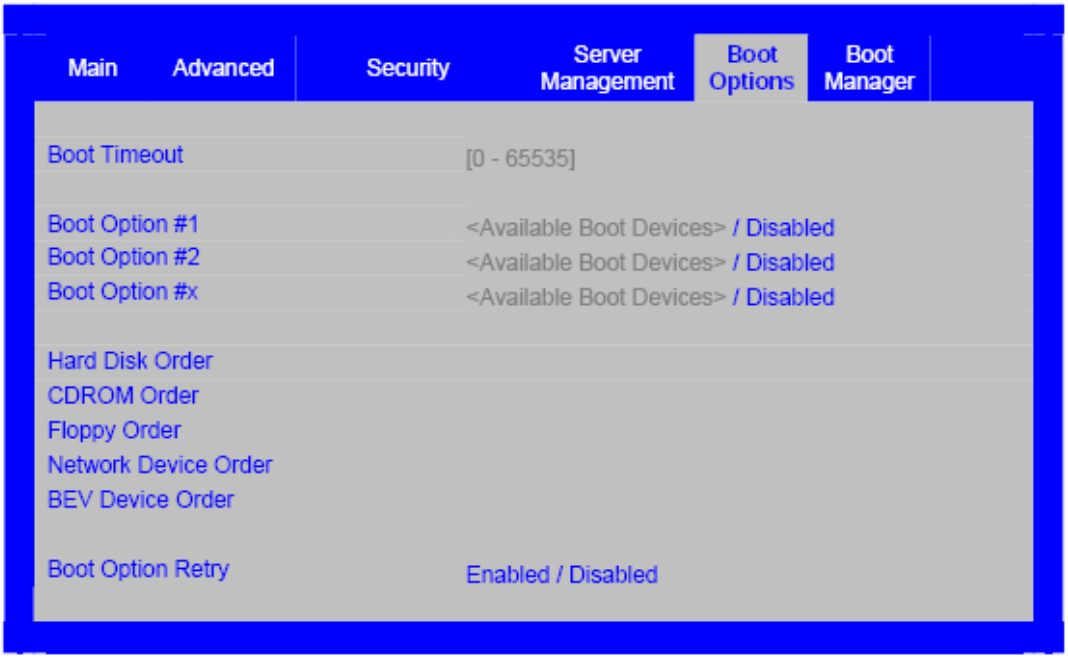


Figure 62. Setup Utility — Boot Options Screen Display

Table 60. Setup Utility — Boot Options Screen Fields

Setup Item	Options	Help Text	Comments
Boot Timeout	<XXXXX>	The number of seconds BIOS will pause at the end of POST to allow the user to press the [F2] key for setting in in seconds	After entering the desired timeout, press enter to register the new timeout value to the system. This

Setup Item	Options	Help Text	Comments
Hard Disk Order		Set hard disk boot order by selecting the boot option for this position.	This option is displayed when more than one hard disk drive has been detected. Takes the user to a sub-menu screen
CDROM Order		Set CDROM boot order by selecting the boot option for this position.	This option is displayed when more than one CDROM drive has been detected. Takes the user to a sub-menu screen
Floppy Order		Set floppy disk boot order by selecting the boot option for this position.	This option is displayed when more than one floppy disk drive has been detected. Takes the user to a sub-menu screen
Network Device Order		Set network device boot order by selecting the boot option for this position. Add-in or onboard network devices with a PXE option ROM are two examples of network boot devices.	This option is displayed when more than one network devices has been detected. Takes the user to a sub-menu screen
BEV Device Order		Set the Bootstrap Entry Vector (BEV) device boot order by selecting the boot option for this position. BEV devices require their own proprietary method to load an OS using a bootable option ROM. BEV devices are typically found on remote program load devices.	This option is displayed when more than one BEV device has been detected. Takes the user to a sub-menu screen
Boot Option Retry	Enabled **Disabled**	This will continually retry NON-EFI based boot options without waiting for user input.	

Hard Disk Order Screen

The Hard Disk Order screen provides a way to control the hard disk boot order.

From the Main screen select Boot Options | Hard Disk Order to access this screen.

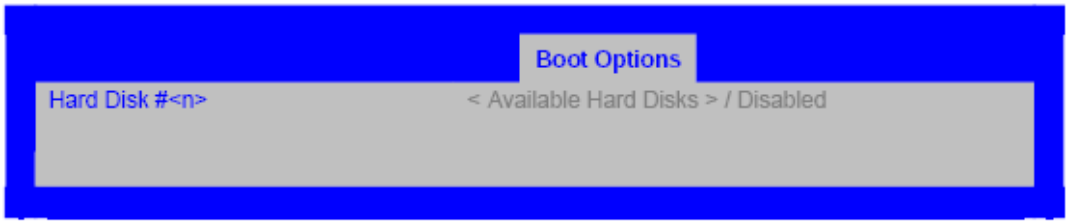


Figure 63. Setup Utility — Hard Disk Order Screen Display

Table 61. Setup Utility — Hard Disk Order Fields

Setup Item	Options	Help Text	Comments
Hard Disk #<n>	<Available Hard Disks> Disabled	Set hard disk boot order by selecting the boot option for this position.	A separate line is displayed for each Hard Disk detected. Disabled means the device is skipped.

CDROM Order Screen

The CDROM Order screen provides a way to control the CDROM device boot order.

From the Main screen select the Boot Options | CDROM Order option to access this screen.

Figure 64. Setup Utility — CDROM Order Screen Display

Table 62. Setup Utility — CDROM Order Fields

Setup Item	Options	Help Text	Comments
CDROM #<n>	<Available CDROM Drives> Disabled	Set CD-ROM boot order by selecting the boot option for this position.	A separate line is displayed for each CD-ROM drive detected. Disabled means the device is skipped.

Floppy Order Screen

The Floppy Order screen provides a way to control the floppy drive boot order.

From the Main screen select Boot Options | Floppy Order to access this screen.

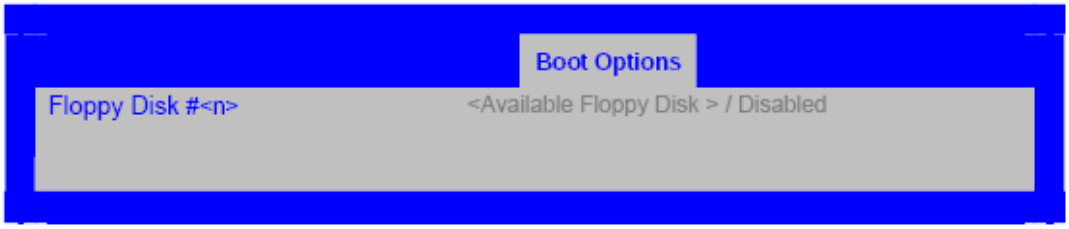


Figure 65. Setup Utility — Floppy Order Screen Display

Table 63. Setup Utility — Floppy Order Fields

Setup Item	Options	Help Text	Comments
Floppy Disk #<n>	<Available Floppy Disks> Disabled	Set floppy disk boot order by selecting the boot option for this position.	A separate line is displayed for each Floppy Disk drive detected. Disabled means the device is skipped.

Network Device Order Screen

The Network Device Order screen provides a way to control the network bootable device boot order.

From the Main screen select Boot Options | Network Device Order to access this screen.

Figure 66. Setup Utility — Network Device Order Screen Display

Table 64. Setup Utility — Network Device Order Fields

Setup Item	Options	Help Text	Comments
Network Device #<n>	<Available Network Devices> Disabled	Set network device boot order by selecting the boot option for this position. Add-in or onboard network devices with a PXE option ROM are two examples of network boot devices.	A separate line is displayed for each network device detected. Disabled means the device is skipped.

BEV Device Order Screen

The BEV Device Order screen provides a way to control the BEV bootable devices boot order. Bootstrap Entry Vector (BEV) devices do not support the standard Interrupt 13h boot support routines in their Option ROM images and require proprietary methods to load an operating system. Devices utilizing the BEV method are typically remote program load devices such as network cards. See the BIOS Boot Specification Version 1.01. Compaq Computer Corporation, Phoenix Technologies Ltd., Intel Corporation 1996 for more information.

From the Main screen select Boot Options | BEV Device Order to access this screen.

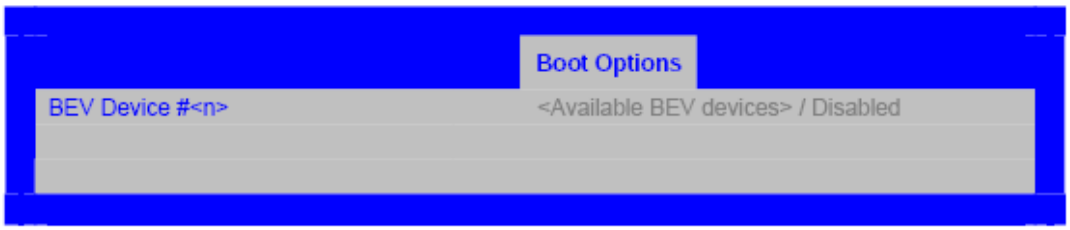


Figure 67. Setup Utility — BEV Device Order Screen Display

Table 65. Setup Utility — BEV Device Order Fields

Setup Item	Options	Help Text	Comments
BEV Device #<n>	<Available BEV Devices> Disabled	Set the Bootstrap Entry Vector (BEV) device boot order by selecting the boot option for this position. BEV devices require their own proprietary method to load an OS using a bootable option ROM. BEV devices are typically found on remote program load devices.	A separate line is displayed for each BEV device detected. Disabled means the device is skipped.

Boot Manager Screen

The Boot Manager screen displays a list of devices available to boot from and allows the user to select a boot device for the current boot.

From the Main screen select Boot Manager to access this screen.

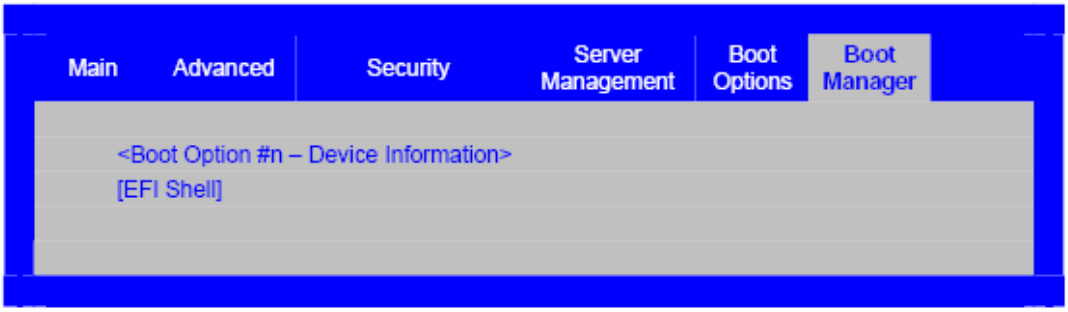


Figure 68. Setup Utility — Boot Manager Screen Display

Table 66. Setup Utility — Boot Manager Screen Fields

Setup Item	Options	Help Text	Comments
<Boot Option #n – Device Information>		Select this option to boot now. Note: This list is not the system boot option order. Use the Boot Options menu to view and configure the system boot option order.	A separate line is displayed for each boot device.
[EFI Shell]		Select this option to boot now. Note: This list is not the system boot option order. Use the Boot Options menu to view and configure the system boot option order.	

Error Manager Screen

The Error Manager screen displays any errors encountered during POST.

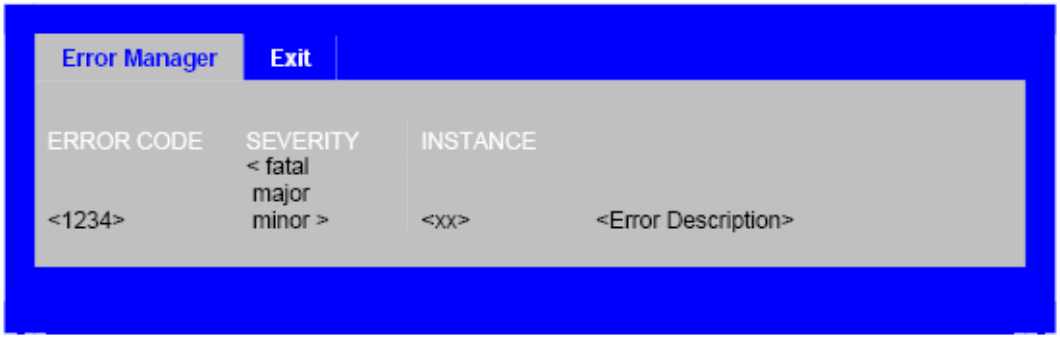


Figure 69. Setup Utility — Error Manager Screen Display

Table 67. Setup Utility — Error Manager Screen Fields

Item	Attribute	Comment
Error Code	3 to 4 digit Error Codes	**Information only** The error code value identifies the error. Refer to Section 19.3 for the complete list of Error Codes.
Severity	Major/ Minor/ Fatal	**Information only** Major severity requires user intervention but does not stop system boot. Minor severity do not require user intervention or stop the booting of the system. Fatal severity requires user intervention and prohibits system boot.
Instance		**Information only** The instance value identifies the component in error.
Description		**Information only** Brief description of the error

Exit Screen

The Exit screen provides several user options related to BIOS Setup menu item changes in the current session as well as the loading of BIOS factory default values.

- If Load Default Values is selected, the default settings, noted in bold in the tables in this chapter, are applied.
- If Load User Default Values is selected, the system is restored to the default values that the user saved earlier, instead of being restored to the factory defaults.

From the Main screen select Exit to access this screen.

Figure 70. Setup Utility — Exit Screen Display

Table 68. Setup Utility — Exit Screen Fields

Setup Item	Options	Help Text	Comments
Save Changes and Exit		Exit BIOS Setup Utility after saving changes. The system will reboot if required. The [F10] key can also be used.	
Discard Changes and Exit		Exit BIOS Setup Utility without saving changes. The [Esc] key can also be used.	
Save Changes		Save changes without exiting BIOS Setup Utility. Note: Saved changes may require a system reboot before taking effect.	
Discard Changes		Discard changes made since the last save changes operation was performed.	
Load Default Values		Load factory default values for all BIOS Setup Utility options. The [F9] key can also be used.	User is prompted for confirmation.

Setup Item	Options	Help Text	Comments
Save as User Default Values		Save current BIOS Setup Utility values as custom user default values. If needed, the user default values can be restored via the Load User Default Values option below. Note: Clearing CMOS or NVRAM will cause the user default values to be reset to the factory default values.	User is prompted for confirmation.
Load User Default Values		Load user default values.	User is prompted for confirmation.

Loading BIOS Defaults
Different mechanisms exist for restoring the system configuration to default values. The BIOS loads the default system configuration values during the next POST after the request is received.

A user can restore the BIOS factory default configuration as follows::

- The BIOS Setup Utility <F9> hotkey or Exit menu Load Default Values option
- The main board NVRAM Clear jumper

- Power down the system, but do not remove AC power.
- Configure the NVRAM Clear jumper to enabled/clear CMOS, connecting pins 2 and 3.
- Configure the NVRAM Clear jumper to default/normal mode, connecting pins 1 and 2.
- Power on the system.

Note: *The server system does not support any other mechanisms to clear NVRAM.*

16. **BIOS Update Support**
The server system supports these flash update utilities:

- iFlash32: EFI Shell, Windows PE
- Intel® One Boot Flash Update

Rolling BIOS
The Rolling BIOS feature provides a fault tolerant BIOS update mechanism. The BIOS relies on specialized hardware and additional flash space to support the Rolling BIOS feature.

The server system supports two physical 4 MB flash parts, each storing a separate BIOS image. The physical flash part containing the BIOS image used to boot the system is referred to as the Primary Flash Bank.

The physical flash part containing the alternate or backup BIOS image is referred to as the Secondary Flash Bank. All BIOS flash updates are made to the Secondary Flash Bank.

Operating Modes
The Rolling BIOS jumper has two operating modes.

- Pins 2-3 Connected – Normal Mode (Default Setting)

- Pins 1-2 Connected – Force Other Bank

Rolling BIOS behavior is described below for each of these jumper settings.

Rolling BIOS Jumper Behavior (Normal Mode)

The user should perform all BIOS flash updates with the Rolling BIOS Jumper configured for Normal Mode (default jumper setting). The following steps describe the BIOS flash update process:

1. Configure the Rolling BIOS Jumper to Normal Mode (Pin 2 and 3).
2. Boot the system.
3. Update the BIOS using the EFI Flash or Intel® One Flash Update (OFU) utility.
4. Reset the system.
5. The Rolling BIOS feature automatically performs the following steps:
   --Boots the system using the old BIOS image.
   – Validates the new BIOS image.
6. If the new image flashed successfully then the BIOS automatically resets the system and boots the new BIOS image.
7. If the new BIOS image failed to boot, the Rolling BIOS feature automatically restores the system to a known-good state by booting the old (known good) BIOS image.

Rolling BIOS Jumper Behavior (Force Other Bank)

There are several possible scenarios where the user may need to manually rollback a BIOS flash update to boot from the old (known good) BIOS image:

- The user has successfully updated the BIOS. The user subsequently learns the new BIOS image does not provide the desired functionality.
- The user has previously updated the BIOS successfully. The user subsequently changes the system configuration in some manner and the new BIOS stops working in response.
- A power failure occurs during the flash update. The user has reset the system and attempted to boot the new BIOS unsuccessfully.

This can be accomplished with the steps below:

1. Configure the Rolling BIOS Jumper to Force Other Bank mode (Pin 1 and 2).
2. Boot the system.
3. The system boots from the BIOS image stored on the Secondary Flash Bank.

OEM Binary

The BIOS supports an OEM Firmware Volume (OEM FV). The size of the OEM FV is 192 KB and the OEM FV can be updated independently of other firmware volumes. The OEM FV hosts a firmware file system. The OEM FV is used to contain the OEM Splash Logo as well as optional OEM strings.

OEM Splash Logo

The OEM FV may contain an optional OEM Splash Logo for display during POST.
The Change Logo utility allows users to replace the standard Intel Splash Logo with a customized OEM Splash Logo. This utility supports BMP files at resolutions up to 800x600 in any color depth.

OEM Strings

The OEM FV may contain optional strings intended for population in SMBIOS Type 11 OEM Strings data structures.

The DMIEDIT utility allows users to update several SMBIOS fields including the OEM Strings.

Operating System Boot, Sleep, and Wake

Boot Device Selection
The Boot Options menu allows the user to modify the boot order by selecting the order of various boot devices. Any boot order modifications can then be saved and remain persistent across boot cycles.

The Boot Manager menu allows the user to override the current boot order and select a specific device to boot the system.

Note: Any selection in the Boot Manager applies only to the current boot.

Server Managment Boot Device Control
The Intelligent Platform Management Interface Specification, Version 2.0, Intel Corporation specification includes provisions for server management devices to set certain boot parameters by setting boot flags. Among the boot flags (parameter #5 in the IPMI specification), the BIOS checks data 1-3 for forced boot options.

The BIOS supports forced booting from the following media devices:

- PXE
- iSCSI
- HDD (USB, SATA, and SAS)
- USB FDD
- USB Key Fob
- DVD/CD-ROM

On each boot, BIOS invokes the Get System Boot Options command to determine what changes to boot options have been set. The BIOS takes the appropriate action and clears these settings.

USB Device Booting
USB boot devices can take on many different characteristics. The following rules are used in detecting and booting from a USB device:

- If the block size of the media is greater than 512 bytes then the emulation type is assumed as CDROM (El Torito formatted).
- If the device has a partition table with more than one partition then the emulation type is assumed as "Hard Disk".
- If the device has a partition table with only one partition then the device is emulated as "Forced FDD".
- All the other devices that do not follow the above rules are assumed as "Floppy Disk".

Additionally a USB HDD can only emulate a FDD (Forced FDD) if there is only a single partition on the HDD. If the HDD has multiple partitions, it can only be set to Auto or Hard Disk and cannot emulate a FDD. The Force FDD option only works for devices formatted with FAT12, FAT16 or

FAT32.

USB Boot Device Reordering
In order to facilitate priority boot of various external USB boot devices & media without the need to enter the Setup Utility and reconfigure the saved Boot Options, BIOS adjusts Boot Options for bootable USB devices. This automatic reordering of USB boot devices only occurs when a USB device is newly detected and not found in the previous configured boot order. When that USB boot device is removed, the configured order of Boot Options is restored.

If a standard boot device of the same type (Hard Disk, CDROM, Floppy) is already present in the configured Boot Options, then the USB boot device of that type is given priority and moved to the top of that device type boot order to boot before other devices of the same type. However, the position of that device type in the Boot Manager order is not changed to preserve the configured boot device type order. If a standard boot device of the same type is not already present in the configured Boot Options, then that type is given priority and moved to the top position in the Boot Manager order to boot before other device types already configured.

If the USB boot device is not intended for a one-time boot and remains configured, then the boot order including the USB device can still be configured and saved in the Setup Utility and is preserved as a permanent change to the boot order.

For security reasons, this USB boot device reordering does not occur if the User Password has been installed via the Security Configuration Screen in the Setup Utility.

Operating System Support

Microsoft Windows* Compatibility
Intel Corporation and Microsoft Corporation co-author design guides for system designers using Intel® processors and Microsoft* operating systems. The Hardware Design Guide for Microsoft Windows 2000 Server, Version 3.0, Microsoft Corporation is intended for systems designed to work with Microsoft Windows Server* class operating systems. The specification further classifies the systems and includes sets of requirements based on the intended usage for that system. For example, a server system that is used in small home / office environments has different requirements than one used for enterprise applications.

This product supports the Hardware Design Guide for Microsoft Windows 2000 Server, Version
3.0 Microsoft Corporation enterprise requirements.
3.1

Advanced Configuration and Power Interface (ACPI)
The BIOS performs several functions to support ACPI:

- Configure the system for ACPI mode when requested by the operating system.
- Implement the Interrupt 15h, Function E820h memory map interface.
- Provide ACPI table support

As described in the ACPI specifications, an ACPI-aware operating system generates an SMI. The SMI

requests that the system be switched into ACPI mode. The BIOS responds completing the following actions:

- Configuring the system as required to support ACPI,
- Issuing the appropriate command to the BMC to enable ACPI mode
- Setting the SCI_EN bit as defined by the ACPI specification

The system automatically returns to legacy mode after hard reset or power-on reset. The BIOS supports the Interrupt 15h, Function E820h system memory map interface as described in the Advanced Configuration and Power Interface Specification, Revision 3.0. The memory used for the ACPI tables is marked as “reserved” in the memory map to prevent memory use conflicts with a non-ACPI-aware operating system.

The primary role of the ACPI BIOS is to supply the ACPI tables. The BIOS creates the ACPI tables during POST. The RSDP table is located in the F000h segment in compliance with the Advanced Configuration and Power Interface Specification, Revision 3.0 with all other tables in extended memory (above 1 MB).

The BIOS supports the following the Advanced Configuration and Power Interface Specification, Revision 2.0, July 2000 and the Advanced Configuration and Power Interface Specification, Revision 3.0 tables:

Table 69. Supported ACPI Tables

ACPI Table	Table Description	ACPI v2.0 Compliant	ACPI v3.0 Compliant
DBGP	Debug Port Table	Yes	Yes
DSDT	Differentiated System Description Table	Yes	Yes
FADT	Fixed ACPI Description Table	Yes	Yes
FACS	Firmware ACPI Control Structure	Yes	Yes
HPET	High Precision Event Timer Table	No	Yes
MADT	Multiple APIC Description Table	Yes	Yes
MCFG	Memory Mapped Configuration Space Base Address Description Table	No	Yes
RSDT	Root System Description Table	Yes	Yes
SLIC	Software Licensing Description Table	No	Yes
SPCR	Serial Port Console Redirection Table	Yes	Yes
SSDT	Secondary System Description Table	Yes	Yes
TCPA	Trusted Computing Platform Alliance Capabilities Table	No	Yes

The format and location of these tables is documented in the public the Advanced Configuration and Power Interface Specification, Revision 2.0, July 2000 and the Advanced Configuration and Power Interface Specification, Revision 3.0 specifications.

Front Control Panel Support

In the control panel, the platform supports:

- Power button
- Reset button
- NMI button

Power Button

The BIOS supports a front control panel power button. Pressing the power button initiates a request that is forwarded directly to the chipset.

Power Button — Off to On
The power button press is forwarded directly to the chipset. Since the processors are not executing, the BIOS does not participate in this sequence.

Power Button — On to Off (operating system absent)
The System Control Interrupt (SCI) is masked. The BIOS sets up the power button event to generate an SMI and checks the power button status bit in the ACPI hardware registers when an SMI occurs. If the status bit is set, the BIOS sets the ACPI power state of the machine in the chipset to the OFF state.

Power Button — On to Off (operating system present)
If an ACPI operating system is running, pressing the power button switch generates a request via SCI to the operating system to shutdown the system. The operating system retains control of the system and operating system policy determines the sleep state into which the system transitions, if any. Otherwise, the BIOS turns off the system.

Reset Button
The platform supports a front control panel reset button. Pressing the reset button initiates a request that is forwarded directly to the chipset. The BIOS does not affect the behavior of the reset button.

NMI Button
The BIOS supports a front control panel non-maskable interrupt (NMI) button. Pressing the NMI button initiates a request that is forwarded directly to the chipset.

Note: *The Front Panel NMI Button is recessed to prevent accidental triggering.*

Sleep and Wake Support

System Sleep States
The server system supports these ACPI system sleep states:

- ACPI S0 state – working state
- ACPI S1 state – low latency sleep state
- ACPI S4 state – hibernate
- ACPI S5 state – soft-off state

The ACPI specification requires the system to support at least one sleep state. S1 is considered a sleep state. The S5 state is equivalent to operating system shutdown. No system context is saved when going into S5.

Supported Wake Events
The ACPI operating system controls the system wake policy. The role of the BIOS is limited to describing the supported wake devices to the operating system via the ASL code in the Differentiated System Description Table (DSDT). The BIOS has no direct control over the wakeup sources when an

ACPI operating system is loaded.

Supported Wake Devices
The hardware and BIOS supports these wake devices in the ACPI environment:

- USB devices including USB mice and keyboards connected to any port can wake the system from the S1 sleep state.
- The Serial Port B can be configured to wake the system from the S1, S4 or S5 sleep states.
- PCI cards, such as LAN cards, can wake the system from the S1, S4 or S5 sleep states. The PCI card must have the necessary hardware and be configured correctly for this to work.
- As required by the ACPI specification, the power button can wake the system from all supported sleep states.

Wake On LAN Support
The BIOS supports Wake On LAN (WOL) as follows:

- WOL supported from ACPI S1 sleep state.
- WOL supported for onboard/embedded NIC devices and add-in NIC devices from ACPI S5 soft off state without A/C power loss as long as they are configured correctly.
- WOL supported for onboard/embedded NIC devices only from ACPI S5 soft off state after A/C power loss.

The wake sources for S1 are configured entirely by the operating system using information in the ACPI DSDT table. PCI devices (Intel® 82563EB NIC) assert PME# signal for WOL when enabled. PME# assertion causes an Out-Of-Band (OOB) wake event. PCI Express* cards also cause wake event for WOL via PME messages or EXP_WAKE signal. PCI Express* cards should support PME or EXP_WAKE assertion on WOL in order to wake the system.
The BIOS supports WOL from S5 and AC power loss through Onboard NICs only. The following usage scenario is supported for WOL from S5:

- System is powered up, booted to the operating system and it operates normally.
- After sometime, the system is shutdown gracefully.
- Then, the system is powered down to S5 state.
- A\C power may be completely removed after the S5 transition.
- At a later time, A\C power is restored. The system then resumes to h/w standby or S5.
- At this point, the system is not powered up and is on stand by. The user wants to be able to WOL this server.

The BIOS needs to properly configure wake events for WOL before entering S5 state for correct wake operation. WOL is supported by platform hardware.

On-board NIC devices must also enable WOL and PME in default configuration at power-up to properly perform WOL after A/C power loss.

Non-Maskable Interrupt (NMI) Handling
Non-maskable Interrupt (NMI) events are generated by two sources:

- Front Panel NMI Button press
- BIOS / Software Generated NMI

The Front Panel NMI Button is described in Section 17.3.3. The BIOS generates Software NMI events during POST and runtime to halt the system in response to unrecoverable system errors.

The BIOS installs a default NMI handler to respond to NMI events during POST including the EFI Shell environment. This handler detects the NMI source and displays an error message as described in the table below before halting the system.

Operating systems typically install their own NMI handler at boot. Operating system response to NMI events is therefore vendor specific.

Table 70. NMI Error Messages

NMI Source	System Error Message
Front Panel NMI Button press	Front Panel NMI activated - System Halted
BIOS / Software NMI	NMI has been received - System Halted

BIOS Role in Server Management
The BIOS supports many standards-based server management features and several proprietary features. The Intelligent Platform Management Interface (IPMI) is an industry standard and defines standardized, abstracted interfaces to platform management hardware.

The BIOS implements many proprietary features that are allowed by the IPMI specification. However, these features are outside the scope of the IPMI specification. This chapter describes the implementation of the standard and proprietary features.

IPMI
Intelligent platform management refers to autonomous monitoring and recovery features that are implemented in platform hardware and firmware. Platform management functions such as the following:

- Inventory
- Event log
- Monitoring
- System health reporting

These functions are available without help from the host processors and when the server is in a powered down state, as long as AC power is attached. The Baseboard Management Controller (BMC) and other controllers perform these tasks independently of the host processor. The BIOS interacts with the platform management controllers through standard interfaces.

The BIOS enables the system interface to the BMC in early POST. The BIOS logs system events and POST error codes during the system operation. The BIOS logs a boot event to BMC early in POST. The events logged by the BIOS comply with the Intelligent Platform Management

Interface Specification, Version 2.0, Intel Corporation requirements.
IPMI defines the required use of all but two bytes in each event log entry, called Event Data 2 and Event Data 3. An event generator can specify that these bytes contain OEM-specified values.

Console Redirection
The BIOS supports both video and keyboard redirection via a serial link (serial port). When console redirection is enabled, the local (host server) keyboard input and video output are passed both to the local keyboard and video connections, and to the remote console through the serial link. Keyboard inputs from both sources are considered valid and video is displayed to both outputs.

As an option, the system can be operated without a host keyboard or monitor attached to the system and run entirely via the remote console. Utilities that can be executed remotely include BIOS Setup.

The BIOS console redirection feature can support both legacy 80x25 text mode as well as EFI graphics console support used to display BIOS Setup and graphics based character display. Redirection of the splash logo image is not supported.

Console redirection ends at legacy operating system boot (Interrupt 19h) or when an EFI-aware operating system calls EFI Boot Services using the ExitBootServices command. The operating system is responsible for continuing the redirection from that point.

Keystroke Mappings
During console redirection, the remote terminal sends keystrokes to the local server. The remote terminal can be a dumb terminal with a direct connection and running a communication program. The keystroke mappings follow VT-UTF8 format with the following extensions.

Standalone <Esc> Key for Headless Operation
The Microsoft Headless Design Guidelines describe a specific implementation for the <Esc> key as a single standalone keystroke:

- <Esc> followed by a two-second pause must be interpreted as a single escape.
- <Esc> followed within two seconds by one or more characters that do not form a sequence described in this specification must be interpreted as <Esc> plus the character or characters, not as an escape sequence.

The escape sequence in the following table is an input sequence. This means it is sent to the BIOS from the remote terminal.

Table 71. Console Redirection Escape Sequences for Headless Operation

Key	PC-ANSI	VT100 / VT100+ / VTUTF8
UP	ESC [A	
DOWN	ESC [B	
RIGHT	ESC [C	
LEFT	ESC [D	
HOME	ESC [H	ESC h
END	ESC [K	ESC k
INSERT	ESC [L	ESC +
DELETE	ESC [P	ESC -
PG UP	ESC [?	ESC ?
PG DOWN	ESC [/	ESC /
F1	ESC [O P	ESC 1
F2	ESC [O Q	ESC 2
F3	ESC [O w	ESC 3
F4	ESC [O x	ESC 4
F5	ESC [O t	ESC 5
F6	ESC [O u	ESC 6
F7	ESC [O q	ESC 7
F8	ESC [O r	ESC 8
F9	ESC [O p	ESC 9
F10	ESC [O M	ESC 0
Remote Console Reset	ESC R ESC r ESC R	ESC R ESC r ESC R
Force Serial Port B MUX to BMC	ESC [O 9	

Interface to Server Management
If the BIOS determines that console redirection is enabled, it reads the current baud rate and pass this value to the appropriate management controller via the Intelligent Platform Management Bus (IPMB).

IPMI Serial Interface
The system shares a communication serial port with the BMC. A multiplexer, controlled by the BMC, determines if the Serial Port B external connector is electrically connected to the BMC or to the standard serial port of the Super I/O. See the Intelligent Platform Management Interface Specification, Version 2.0, Intel Corporation Section 14 “IPMI Serial/Modem Interface” for information about these features.

Channel Access Modes
The BIOS supports the four different channel access modes that are described in the Intelligent Platform Management Interface Specification, Version 2.0, Intel Corporation Table 6-4.

Interaction with BIOS Console Redirection
BIOS Console Redirection accomplishes the implementation of VT-UTF8 console redirection support in Intel’s server BIOS products. This implementation meets the functional requirements set forth in the Microsoft Windows 2003* WHQL requirements for headless operation of servers. It also maintains a necessary degree of backward compatibility with existing Intel server BIOS products.

The BIOS has a console that interacts with a display and keyboard combination. The BIOS instantiates sources and sinks of input / output data in the form of the following:

- BIOS Setup screens
- Boot Manager screens
- Power On Self Test (POST) informational messages
- Hot-key / escape sequence action requests

Output is displayed locally at the computer on video display devices. This is limited to VGA displays in text or graphics mode. Local input may come from a USB keyboard. Mouse support is not available.

The use of serial port console redirection allows a single serial cable to be used for each server system. The serial cables from a number of servers can be connected to a serial concentrator or to a switch. This allows access to each individual server system. The system administrator can remotely switch from one server to another to manage large numbers of servers.

BIOS console redirection supports an extra control escape sequence to switch Serial Port B to BMC control. The character sequence that switches the multiplexer to the BMC serial port is “ESC O 9” (denoted as ^[O9). This key sequence is above the normal ANSI function keys and is not used by an ANSI terminal.

After this command is sent, the Serial Port B attaches to the BMC Channel Access serial port and Super I/O Serial Port B data is ignored. This feature allows a remote user to monitor the status of POST using the standard BIOS console redirection features and then take control of the system reset or power using the Channel Mode features. If a failure occurs during POST, a watchdog time-out feature in the BMC automatically takes control of the Serial Port B.

18.3.3 Serial Over LAN

The server system must support console redirection via SOL during the entire BIOS POST process with these requirements:

- No loss of characters in the remote console during POST redirection
- No garbage characters should be seen.
- User should be able to clearly see the POST messages.
- Should be able to enter BIOS Setup, select and modify options, and save or discard the change.
- Function keys should work just like the local console.
- No momentary loss even during PXE\DHCP negotiation
- Should support console redirection (PC-ANSI terminal modes).
- Should support all baud rates — 9.6, 19.2, 36K, 56K, 115K.
- Performance should be comparable to the local console with no redirection.
- The platform supports SOL on Serial Port B that also provides EMP and standard serial functionality as an external serial port. Since this port is not dedicated for SOL BIOS Setup options should be provided for configuration when not operating in SOL mode.
- Serial Port B configuration when SOL is enabled should all be done via BMC configuration utilities. The BIOS determines if SOL has been enabled during POST. If SOL has been enabled then the BIOS enables console redirection on Serial Port B and sets the baud rate from BMC. The BIOS also sets the term-type to PC-ANSI and flow control to CTS-RTS.

In addition, operating system console redirection must be supported with SOL for all P1 operating

systems with the following requirements:

- For example, the user should be able to run shell commands under Linux and should be able to see the redirected console.
- No loss of characters in the remote console during operating system redirection
- If the EFI shell is running, to see the EFI Shell prompt on the SOL console, an <Enter> key must be pressed.
- No garbage characters should be seen.

Through the redirection capabilities of the BMC on Intel platforms, the Serial Port B (UART) input/output stream can be further redirected and sent over a platform LAN device as a packetized serial byte stream. This BMC function is called Serial over LAN (SOL) .In addition, space requirement and server management compatibility are further optimized. BIOS starts the console redirection on Serial Port B automatically when it detects SOL is enabled in BMC. BIOS sets Serial Port B flow control and baud rate from the BMC IPMI Serial/Modem configuration. Data bits are set to 8 bits / character, no parity and one stop bit as per IPMI messaging requirement. The console type is set to VT100+.

Wired For Management (WFM)

Wired for Management is an industry-wide initiative to increase overall manageability and reduce total cost of ownership. WFM allows a server to be managed over a network. The system BIOS supports the System Management BIOS Reference Specification, Version 2.5 to help higher-level instrumentation software meet the Wired For Management Baseline Specification, Revision 2.0.

PXE BIOS Support

The BIOS supports the EFI PXE implementation as specified in the Extensible Firmware Interface Reference Specification, Version 1.1 Chapter 15. To utilize this, the user must load EFI Simple Network Protocol driver and the UNDI driver specific for the network interface card being used. The UNDI driver should be included with the network interface card. The Simple Network Protocol driver is available at http://developer.intel.com/technology/framework

The BIOS supports legacy PXE Option ROMs in legacy mode and includes the necessary PXE ROMs in the BIOS image for the onboard controllers. The legacy PXE ROM is required to boot a non-EFI operating system over the network.

System Management BIOS (SMBIOS)

The BIOS provides support for the System Management BIOS Reference Specification, Version 2.5 to create a standardized interface for manageable attributes. These attributes are expected to be supported by DMI-enabled computer systems. The BIOS provides this interface via data structures through which the system attributes are reported. Using SMBIOS, a system administrator can obtain the following: server component information:

- Types
- Capabilities
- Operational status
- Installation date
- Other information

Access Methods
Two access methods are defined for the SMBIOS structures.

- The first method provides the SMBIOS structures through a PnP function interface. This method is not supported by the BIOS.
- The second method is the table convention. The table convention allows the SMBIOS structures to be accessed under 32-bit protected-mode operating systems.

The total number of structures can be obtained from the SMBIOS entry-point structure. The system information is presented to an application as a set of structures that are obtained by traversing the SMBIOS structure table referenced by the SMBIOS entry-point structure.

Structure Table Entry Point
The SMBIOS entry point structure can be located by application software by searching for the structure information in the following table within the physical memory address range of 000F0000h to 000FFFFFh. The SMBIOS tables are located above physical address 100000h.

Table 72. SMBIOS Table Entry Point Structure

Offset	Name	Length	Value	Description
00h	Anchor String	Four bytes	"_SM_"	
04h	Entry Point Structure Checksum	Byte	Varies	Checksum of whole structure.
05h	Entry Point Length	Byte	1Fh	
06h	SMBIOS Major Revision	Byte	02h	
07h	SMBIOS Minor Revision	Byte	05h	
08h	Maximum Structure Size	Word	Varies	Size of largest supported structure.
0Ah	Entry Point Revision	Byte	00h	
0Bh	Formatted Area	Five bytes		
10h	Intermediate Anchor String	Five bytes	"_DMI_"	
15h	Intermediate Checksum	Byte	Varies	Checksum from 10h to end of structure.
16h	Structure Table Length	Word	Varies	Total length in bytes of structure pointed at by offset 18h.
18h	Structure Table Address	DWord	Varies	32-bit physical starting address of the SMBIOS structure table.
1Ch	Number of SMBIOS Structures	Word	Varies	Total number of SMBIOS structures present.
1Eh	SMBIOS BCD Revision	Byte	24h	Bits 7:4 - Major revision Bits 3:0 - Minor revision

OEM Modification
An OEM customer may use the DMIEDIT utility to modify certain SMBIOS fields. See the DMIEDIT User Guide for more details including a list of the specific SMBIOS fields which can be modified with the utility.

SMBIOS Stuctures Supported

The BIOS supports the following structure types. A detailed table is provided for those structure types requiring specific clarification or implementation detail for this server system.

Type 0 Structure — BIOS Information

The Type 0 structure contains information about the BIOS revision ID, BIOS build date and the technologies supported by the BIOS. Only one structure exists to describe the BIOS. No structures are present to describe Option ROMs.

Table 73. SMBIOS Type 0 Structure — BIOS Information

Offset	Name	Length	Value	Description
00h	Type	Byte	0	BIOS information indicator.
01h	Length	Byte	18h	
02h	Handle	Word	Varies	
04h	Vendor	Byte	String	String number of the BIOS Vendors Name. "Intel Corporation".
05h	BIOS Version	Byte	String	String number of the BIOS Version. Contains full Intel BIOS ID string.
06h	BIOS Starting Address Segment	Word	Varies	
08h	BIOS Release Date	Byte	String	String number of the BIOS Release Date. Date is in mm/dd/yyyy format.
09h	BIOS ROM Size	Byte	Varies (n)	Size (n) where 64K*(n+1) is the size of the BIOS flash part. This should be equivalent to 4MB (3Fh).
0Ah	BIOS Characteristics	QWord	Bit Field	See the System Management BIOS Reference Specification, Version 2.5 Section 3.3.1.1 for enumeration of values.
12h	BIOS Characteristics Extension Bytes	Word	Bit Field	See the System Management BIOS Reference Specification, Version 2.5 Section 3.3.1.2 for enumeration of values. Byte 1 and Byte 2 are supported.
14h	System BIOS Major Release	Byte	Varies	Major version of BIOS Identifies the major release of the System BIOS; for example, the value is 0Ah for revision 10.22 and 02h for revision 2.1. This field and/or the System BIOS Minor Release field is updated each time a System BIOS update for a given system is released. If the system does not support the use of this field, the value is 0FFh for both this field and the System BIOS Minor Release field.

Offset	Name	Length	Value	Description
15h	System BIOS Minor Release	Byte	Varies	Minor version of BIOS Identifies the minor release of the System BIOS; for example, the value is 16h for revision 10.22 and 01h for revision 2.1.
16h	Embedded Controller Firmware Major Release	Byte	Varies	Major version of BMC firmware Identifies the major release of the embedded controller firmware; for example, the value is 0Ah for revision 10.22 and 02h for revision 2.1. This field and/or the Embedded Controller Firmware Minor Release field is updated each time an embedded controller firmware update for a given system is released. If the system does not have field upgradeable embedded controller firmware, the value is 0FFh.
17h	Embedded Controller Firmware Minor Release	Byte	Varies	Minor version of BMC firmware Identifies the minor release of the embedded controller firmware; for example, the value is 16h for revision 10.22 and 01h for revision 2.1. If the system does not have field upgradeable embedded controller firmware, the value is 0FFh.

Type 1 Structure — System Information

The SMBIOS Type 1 record is populated by obtaining information from the product area of BMC FRU and from the Vital Product Data area. The information obtained from the product area of the BMC FRU can be customized.

Table 74. SMBIOS Type 1 Structure — System Information

Offset	Name	Length	Value	Description
00h	Type	Byte	1	System information indicator.
01h	Length	Byte	1Bh	
02h	Handle	Word	Varies	
04h	Manufacturer	Byte	String "Intel"	String number of the Manufacturer Obtained from FRU "Product Manufacturer" field
05h	Product Name	Byte	String "This Products name"	String number of the Product Name Obtained from FRU by concatenation of the "Product Name" and "Product Part Number" fields with a space character between the two strings
06h	Version	Byte	String	String number of the Product Version Obtained from FRU "Product Version" field
07h	Serial Number	Byte	String	String number of the Serial Number Obtained from FRU "Product Serial Number" field
08h	UUID	16 bytes	Varies	This is from the value stored in non-volatile RAM (either BIOS Flash or BMC).
18h	Wakeup Type Interrupt Info	Byte	Enum	See the System Management BIOS Reference Specification, Version 2.5 Section 3.3.2.1 for meaning.

Offset	Name	Length	Value	Description
19h	SKU Number	Byte	String	String number of the SKU Number Number of Null terminated string. This text string is used to identify a particular computer configuration for sale. It is sometimes also called a product ID or purchase order number. This number is frequently found in existing fields, but there is no standard format. Typically for a given system board from a given OEM, there are tens of unique processor, memory, hard drive, and optical drive configurations.
1Ah	Family	Byte	String	String number of the Family Number of Null terminated string. This text string is used to identify the family a particular computer belongs to. A family refers to a set of computers that are similar but not identical from a hardware or software point of view. Typically, a family is composed of different computer models, which have different configurations and pricing points. Computers in the same family often have similar branding and cosmetic features.

Type 2 Structure — Base Board Information

The SMBIOS Type 2 structure is populated by obtaining information from the product area of the BMC FRU. The information obtained from this area can be customized. See the IPMI Platform Management FRU Information Storage Definition, Version 1.0 for more information.

Type 3 Structure — System Enclosure or Chassis

The SMBIOS Type 3 structure is populated by obtaining information from the product area of the BMC FRU. The information obtained from this area can be customized. See the IPMI Platform Management FRU Information Storage Definition, Version 1.0 for more information.

Type 4 Structure — Processor Information

The SMBIOS Type 4 structure describes the attributes of a single physical processor. The SMBIOS Table Structure contains one Type 4 structure for each physical processor socket in the server.

Table 75. SMBIOS Type 4 Structure — Processor Information

Offset	Name	Length	Value	Description
00h	Type	Byte	4	Processor information indicator.
01h	Length	Byte	28h	
02h	Handle	Word	Varies	
04h	Socket Designation	Byte	String	Number of null terminated string. Contains the reference designator on the silkscreen of the processor socket
05h	Processor Type	Byte	B3h	B3h = Intel® Xeon processor family.
06h	Processor Family	Byte	Enum	See the System Management BIOS Reference Specification, Version 2.5

Offset	Name	Length	Value	Description
				Section 3.3.5.2 for values.
07h	Processor Manufacturer	Byte	String	Number of null terminated string. String contains "Intel Corporation".
08h	Processor ID	QWord	Varies	Contains the results of the CPUID instruction with EAX = 1 as follows: ▪ Offset 08h-0Bh: EAX ▪ Offset 0Ch-0Fh: EDX
10h	Processor Version	Byte	String	Number of Null terminated string that describes the processor. This string is returned from the processor.
11h	Voltage	Byte	Varies	▪ Bit 7 - 1 ▪ Bits [6:0] ▪ Current processor voltage * 10 - 1.8V = 92h
12h	External Clock	Byte	Varies	External clock speed in MHz.
14h	Max Speed	Word	Varies	Maximum internal processor speed in MHz.
16h	Current Speed	Word	Varies	Current internal processor speed in MHz.
18h	Status	Word	Varies	Bit 7 ▪ 0 = Reserved Bit 6 ▪ 0 = Socket unpopulated ▪ 1 = Socket populated Bits 5:3 ▪ 0 = Reserved Bits 2:0 ▪ 0h = Unknown ▪ 1h = CPU enabled ▪ 2h = CPU disabled by user ▪ 3h = CPU disabled by BIOS ▪ 4h = CPU idle, waiting to be enabled ▪ 5h, 6h = Reserved ▪ 7h = Other
19h	Processor Upgrade	Byte	04h	04h - ZIFF socket.
1Ah	L1 Cache Handle	Word	Varies	Handle of the cache information structure for L1 cache for this processor. Set to 0FFFFh if the cache information structure is not supported.
1Ch	L2 Cache Handle	Word	Varies	Handle of the cache information structure for L2 cache for this processor. Set to 0FFFFh if cache information structure is not supported.
1Eh	L3 Cache Handle	Word	Varies	Handle of cache information structure for L3 cache for this processor. Set to 0FFFFh if cache information structure is not supported.
20h	Serial Number	Byte	String	String number for the serial number of this processor. This value is set by the manufacturer and normally not changeable.
21h	Asset Tag	Byte	String	String number for the asset tag of this processor.

Offset	Name	Length	Value	Description
22h	Part Number	Byte	String	String number for the part number of this processor. This value is set by the manufacturer and normally not changeable.
23h	Core Count	Byte	Varies	Number of cores per processor socket. If the value is unknown, the field is set to 0. See the *System Management BIOS Reference Specification*, Version 2.5, Section 3.3.5.6 for more information.
24h	Core Enabled	Byte	Varies	Number of enabled cores per processor socket. If the value is unknown, the field is set 0. See the *System Management BIOS Reference Specification*, Version 2.5, Section 3.3.5.7 for more information.
25h	Thread Count	Byte	Varies	Number of threads per processor socket. If the value is unknown, the field is set to 0. See the *System Management BIOS Reference Specification*, Version 2.5, Section 3.3.5.8 for more information.
26h	Processor Characteristics	Word	Bit Field	Defines which functions the processor supports.

Type 7 Structure — Cache Information

The SMBIOS Type 7 structure describes the attributes of the processor cache device(s) in the server. The BIOS dynamically creates one structure per cache device present in the server. For example, the BIOS creates six Type 7 structures if two processors are installed each supporting three levels of cache.

Table 76. SMBIOS Type 7 Structure — Cache Information

Offset	Name	Length	Value	Description
00h	Type	Byte	7	Cache information indicator.
01h	Length	Byte	13h	
02h	Handle	Word	Varies	
04h	Socket Designation	Byte	String	String number for Reference Designation Same as associated processor
05h	Cache Configuration	Word	Varies	Bits 15:10 ▪ 0 = Reserved Bits 9:8 ▪ 00b = Write through ▪ 01b = Write back ▪ 10b = Varies with memory address ▪ 11b = Unknown Bit 7 ▪ 0b = Disabled at boot time ▪ 1b = Enabled at boot time Bits 6:5 ▪ 00b = Internal Bit 4

Offset	Name	Length	Value	Description
				▪ 0 = Reserved Bit 3 ▪ 1 = Socketed Bits 2:0 ▪ Cache level, zero-based
07h	Maximum Cache Size	Word	Varies	Bit 15 ▪ 0 = 1K granularity ▪ 1 = 64K granularity Bits 14:0 ▪ Max size in granularity
09h	Installed Size	Word	Varies	Bit 15 ▪ 0 = 1k granularity ▪ 1 = 64k granularity Bits 14:0 ▪ Installed size in granularity ▪ Set to 0 if no cache or processor installed.
0Bh	Supported SRAM Type	Word	Bit Field	See the System Management BIOS Reference Specification, Version 2.5 Section 3.3.8.1 for values.
0Dh	Current SRAM Type	Word	Bit Field	See the System Management BIOS Reference Specification, Version 2.5 Section 3.3.8.1 for values.
0Fh	Cache Speed	Word	Varies	In nanoseconds. Set to 0 if unknown.
10h	Error Correction Type	Byte	Enum	See the System Management BIOS Reference Specification, Version 2.5 Section 3.3.8.2 for values.
11h	System Cache Type	Byte	05h	05h = Unified cache
12h	Associativity	Byte	Enum	See the System Management BIOS Reference Specification, Version 2.5 Section 3.3.8.4 for values.

Type 8 Structure — Port Connector Information
The SMBIOS Type 8 structure provides the attributes for all internal and external ports or connectors in the server. There is one type 8 structure for each port / connector.

Type 9 Structure — System Slots
The SMBIOS Type 9 structure describes the attributes of the expansion slots in the server. One Type 9 structure is present for each slot in the server.

Type 10 Structure — Onboard Devices Information
The SMBIOS Type 10 structure defines the attributes of devices integrated into the server board. One Type 10 structure is present for each integrated device on the server board.

Type 11 Structure — OEM Strings
The SMBIOS Type 11 structure is a free-form string area in which OEMs can store string data. This can be optionally constructed via the OEM binary.

The strings are stored sequentially following Offset 04h in a null-terminated format as described in the System Management BIOS Reference Specification, Version 2.5. The BIOS supports a maximum OEM string length of 128 characters. These strings can be altered by an OEM using the DMIEDIT utility.

Table 77. SMBIOS Type 11 Structure — OEM Strings

Offset	Name	Length	Value	Description
00h	Type	Byte	11	OEM strings indicator
01h	Length	Byte	05h	
02h	Handle	Word	Varies	
04h	Count	Byte	Varies	Number of strings

Type 12 Structure — System Configuration Options

The SMBIOS Type 12 structure contains strings describing the configuration settings of all jumpers and switches on the server board.

Type 13 Structure — BIOS Language Information

The SMBIOS Type 13 Structure describes the available languages and current configured language for the BIOS user displays.

Table 78. SMBIOS Type 13 Structure — BIOS Language Information

Offset	Name	Length	Value	Description
00h	Type	Byte	13	Language information indicator.
01h	Length	Byte	16h	
02h	Handle	Word	Varies	
04h	Installable Languages	Byte	1	Only 1 language (English) is supported.
05h	Flags	Byte	Bit Field	Bits 7:1 – Reserved Bit 0 – 0 = ISO 639 / ISO 3166
06h	Reserved	15 bytes	0	
015h	Current Language	Byte	String	String number of current language (1-based).

Type 16 Structure — Physical Memory Array

This structure describes a collection of memory devices that operate together to form a memory address space.

The BIOS reports one Type 16 structure for each memory riser board (memory channel) with one or more physical FBDIMMs installed.

Table 79. SMBIOS Type 16 Structure — Physical Memory Array

Offset	Name	Length	Value	Description
00h	Type	Byte	16	Physical memory array type.
01h	Length	Byte	0Fh	
02h	Handle	Word	Varies	
04h	Location	Byte	09h	Proprietary add-on card
05h	Use	Byte	03h	System memory
06h	Memory Error Correction	Byte	06h	This field is set to Multi-bit ECC. See the System Management BIOS Reference Specification, Version 2.5, Section 3.3.17.3.
07h	Maximum Capacity	DWord	4000000h	The maximum memory capacity in KB Each memory riser board supports 8 GB FBDIMM x 8 slots = 64 GB maximum
0Bh	Memory Error Handle Information	Word	0FFFEh	Type 18 error information not supported.
0Dh	Number of Memory Devices	Word	8	The total number of physical FBDIMM sockets available on each memory riser board..

Type 17 Structure — Memory Device

This structure describes a single memory device that is part of a larger Physical Memory Array.

The BIOS reports one Type 17 structure for every physical FBDIMM socket on each memory riser board (MCH memory channel) with one or more physical FBDIMMs installed.

Table 80. SMBIOS Type 17 Structure — Memory Device

Offset	Name	Length	Value	Description
00h	Type	Byte	17	Memory device type.
01h	Length	Byte	Varies	Length varies, minimum of 15h.
02h	Handle	Word	Varies	
04h	Memory Array Handle	Word	Varies	Handle for the Type 16 Structure describing the memory riser board supporting the FBDIMM socket.
06h	Memory Error Information Handle	Word	0FFFEh	Type 18 error information not supported.
08h	Total Width	Word	72	FBDIMM total width including any check or ECC bits (in bit units).

Offset	Name	Length	Value	Description
0Ah	Data Width	Word	64	FBDIMM data width (in bit units).
0Ch	Size	Word	Varies	The size of the memory device Granularity depends on most-significant bit (bit15). For this BIOS, Bit15 = 1 and the size of the memory device is indicated in MB units. If no FBDIMM is installed in the socket, this field should be 0.
0Eh	Form Factor	Byte	09h	Form factor = DIMM
0Fh	Device Set	Byte	Varies	Identifies when the memory device is one of a set of memory devices that must be populated with all devices of the same type and size, and the set to which this device belongs. A value of 0 indicates that the device is not part of a set; a value of FFh indicates that the attribute is unknown. The device sets are based on the lock-stepped operation concept. In other words, lock-stepped FBDIMM pairs belong to the same Device Set value (e.g. Memory Riser Board A, DIMM_1 and Memory Riser Board B, DIMM_1 should have matching device set value). Note: A device set number must be unique within the context of the memory array containing this memory device.
10h	Device Locator	Byte	String	The string number of the string that identifies the server board silk screen label for the socket or board position where the memory device is located. The string should be in the format indicated by the following example: Memory Riser Board A, DIMM_1
11h	Bank Locator	Byte	NULL	Memory bank concept is not supported
12h	Memory Type	Byte	13h	Memory Device Type = DDR2.
13h	Type Detail	Word	0080h	Type Detail: Bit 7 = 1 (Synchronous)
15h	Speed	Word	Varies	Identifies the frequency: 215h = 533Mhz 29Bh = 667MHz
17h	Manufacturer	Byte	String	The string number manufacturer of this memory device.
18h	Serial Number	Byte	String	The string number for the serial number of this memory device. This value is set by the manufacturer and normally not changeable.
191h	Asset Tag	Byte	String	The string number for the asset tag of this device.
1Ah	Part Number	Byte	String	The string number for the part number of this memory device. This value is set by the manufacturer and normally not changeable.

Type 24 Structure — Hardware Security

The SMBIOS Type 24 structure describes the current states of the password and front panel security features.

Type 32 Structure — System Boot Information

The SMBIOS Type 32 structure is utilized by the client's Pre-eXecution Environment (PXE) to identify the reason why the PXE was initiated.

Type 38 Structure — IPMI Device Information

The SMBIOS Type 38 structure describes the attributes of the embedded IPMI controller on the server board. In addition to the System Management BIOS Reference Specification, Version 2.5 requirements,

two bytes have been appended to the Type 38 structure to provide the following:

- Information about the interrupt used by embedded IPMI controller
- More information about the IPMI base address

Table 81. SMBIOS Type 38 Structure — IPMI Device Information

Offset	Name	Length	Value	Description
00h	Type	Byte	38	IPMI device information structure indicator.
01h	Length	Byte	12h	
02h	Handle	Word	Varies	
04h	Interface Type	Byte	01h	01h = KCS Interface.
05h	IPMI Specification Revision	Byte	20h	IPMI Specification, Version 2.0.
06h	I²C Slave Address	Byte	Varies	Slave address of the I²C bus
07h	NV Storage Device Address	Byte	Varies	Bus ID of the non-volatile storage device.
08h	Base Address	QWord	Varies	The base address for the BMC's system interface. The field can describe both I/O mapped and memory-mapped base addresses. The least significant bit indicates whether the base address is an I/O address or a memory address. The most significant 63 bits holds the most significant 63 bits (bits 63:1) of a 64-bit address. The least significant bit (bit 0) of the base address is kept in the Base Address Modifier field.
10h	Base Address Modifier / Interrupt Info	Byte	Varies	Base address modifier: Bit 7:6 Register spacing ▪ 00b = interface registers are on successive byte boundaries ▪ 01b = interface registers are on 32-bit boundaries ▪ 10b = interface registers are on 16-byte boundaries ▪ 11b = reserved Bit 5 ▪ Reserved: Return as 0b Bit 4 LS-bit for addresses ▪ 0b = Address bit 0 = 0b ▪ 1b = Address bit 0 = 1b ▪ Interrupt information identifies the type and polarity of the interrupt associated with the IPMI system interface, if any.

Offset	Name	Length	Value	Description
				Bit 3 ▪ 1b = interrupt information specified ▪ 0b = interrupt information not specified Bit 2 ▪ Reserved: Return as 00000b. Bit 1 Interrupt polarity ▪ 1b = active high ▪ 0b = active low Bit 0 Interrupt Trigger Mode ▪ 1b = level ▪ 0b = edge
11h	Interrupt Number	Byte	Varies	Interrupt number for IPMI system interface. 00h = Unspecified / unsupported

Type 126 Structure – Inactive
The System Management BIOS Reference Specification, Version 2.5 indicates any static structure describing a feature(s) not supported in the current system configuration may be marked as Type 126 Inactive.

The BIOS marks any static structures describing any system attributes not supported by the current system configuration as Inactive in accordance with the System Management BIOS Reference Specification, Version 2.5.

Type 127 Structure — End-of-Table
The SMBIOS Type 127 structure identifies the end of the structure table that might be earlier than the last byte within the buffer specified by the structure. To ensure backward compatibility with management software written to previous versions of the SMBIOS specification, the structure table is still reported as a fixed-length. The entire length of the table can still be indexed. If the end-of-table indicator is used in the last physical structure in a table, the field's length is encoded as four.

Security
The BIOS uses passwords to prevent unauthorized tampering with the server setup. Both user and administrator passwords are supported by the BIOS.

Password
The maximum length of the password is seven characters. The password cannot have characters other than alphanumeric (a-z, A-Z, 0-9). It is case sensitive. Once set, a password can be cleared by changing it to a null string.

An Administrator password must be entered in order to set the user password. If only one password is set, this password is required to enter BIOS Setup.

Entering the user password allows the user to modify only the following:

- BIOS Setup time
- Date
- Boot manager
- User password menu items

Other Setup fields can be modified only if the administrator password is entered.

The administrator has control over all fields in BIOS Setup, including the ability to clear the user password.

If the user or administrator enters an incorrect password, an incorrect password dialog box is presented and the password must be re-entered. If an incorrect password is entered three times in a row during the boot sequence, the system is placed into a halt state. A system reset is required to exit out of the halt state. This feature makes it difficult to break the password by guessing at it.

Password Clear Jumper
If the user and/or administrator password is lost or forgotten, both passwords may be cleared by

moving the main board Password Clear jumper into the clear/enabled position. The BIOS determines if the Password Clear jumper is in the clear/enabled position during BIOS POST and clears any passwords if required. The Password Clear jumper must be restored to its original position before any new passwords stay set.

BMC Timestamp Synchronization

The BMC maintains a 4-byte internal timestamp clock used by the SEL and SDR subsystems. This clock is read and set using the Get SEL Time and Set SEL Time commands, respectively. The Get SDR Time command can also be used to read the timestamp clock. These commands are specified in the Intelligent Platform Management Interface Specification, Version 2.0.

The BIOS programs the BMC with the current RTC time during POST via the Set SEL Time command. The BIOS also sends a Timestamp Clock Sync System event immediately before and immediately after the Set SEL Time command. The following table describes the event format.

When the user changes the real-time clock (RTC) during operation, SMS is responsible for keeping the BMC and system time in sync.

Table 82. Timestamp Clock Sync Format

Offset	Value	Description
1	03h	Generator ID
2	04h	Event Message Revision
3	12h	System Event
4	83h	Boot Event Sensor

Offset	Value	Description
5	6Fh	Event Type
6	05h	Boot Event
7		[7] – First / second ▪ 0b = Event is first of the pair ▪ 1b = Event is second of the pair [6:4] Reserved [3:0] Timestamp clock sync ▪ 0b = SEL Timestamp clock updated ▪ 1b = SDR Timestamp clock updated
8	FFh	No data

BIOS Error Handling

This chapter defines the following error handling features:

- Fault Resilient Booting
- Error Handling and Logging
- Error Messages and Beep Codes

Fault Resilient Booting

Fault Resilient Booting (FRB) is an Intel-specific feature that detects and handles errors during the system boot process. The FRB feature provides a recovery mechanism in the event of a system hang during BIOS POST.

There are several possible failures during the booting process that can be detected and handled by the BIOS and BMC. The server system implements this FRB support:

- BSP POST Failure (FRB-2)
- Operating System Load Failure

Note: *The server system does not support recovery from processor BIST (FRB-1) or BSP reset (FRB-3) failure.*

BSP POST Failure (FRB-2)

FRB-2 involves the use of a watchdog timer, which can be configured to reset the system after a specified amount of time in the event of a POST hang. The BIOS sets the FRB-2 timer to 6 minutes if the BIOS Setup FRB-2 Enable menu item is enabled.

The BIOS disables the watchdog timer before prompting the user for a boot password (user password), while scanning for Option ROM, and when the user enters BIOS Setup. If the system hangs during POST anytime before the BIOS disables the FRB-2 timer, the BMC generates an Asynchronous System Reset (ASR) as soon as the timer expires.

The BMC retains status bits after the ASR that the BIOS can read on the subsequent boot cycle to log the appropriate event into the System Event Log (SEL) and display an appropriate error message to the user in the POST Error Manager.

Operating System Load Failure (OS Boot Timer)

The BIOS provides an additional watchdog timer to provide fault resilient booting to the operating system. This timer option is disabled by default. The timeout value and the option to enable the timer are configured in BIOS Setup.

When enabled, the BIOS enables the operating system Boot Timer in the BMC. It is the responsibility of the operating system or an application to disable this timer once the operating system has successfully loaded.

Warning: *Enabling this option without having an operating system or server management application installed that supports this feature causes the system to reboot when the timer expires. See the application or operating system documentation to confirm this feature is supported for your operating system environment.*

Operating System Watchdog Failure
If an operating system device driver is using the watchdog timer to detect software or hardware failures and that timer expires, an Asynchronous Reset (ASR) is generated, which is equivalent to a hard reset. The POST portion of the BIOS can query the BMC for watchdog reset event as the system reboots, and logs this event in the SEL.

Boot Event
The BIOS downloads the system date and time to the BMC during POST. The BIOS then logs a boot event. Software that parses the event log should not treat the boot event as an error.

Error Handling and Logging
This section defines how errors are handled by the system BIOS. The role of the BIOS in error handling and the interaction between the BIOS, platform hardware, and server management firmware with regard to error handling is discussed. In addition, error-logging techniques are described and beep codes for errors are defined.

In the case of fatal and non-recoverable errors, the continued execution of the asynchronous SMI error handler cannot be guaranteed due to the possibility for either catastrophic data corruption compromising the integrity of the handler and/or system hardware reliability issues.

The handler records the error to the system event log only if the system has not experienced a catastrophic failure that compromises the integrity of the handler.

Runtime Error Handler
A System Management Mode (SMM) runtime handler is used during runtime to handle and log system level events that are not visible to the server management firmware. The runtime handler pre-processes all system errors, even those normally considered to generate an NMI.

The runtime handler sends a command to the BMC to log the event and provides the data to be logged. For example, the BIOS programs the hardware to generate an SMI on a single-bit memory error and logs the location of the failed FBDIMM in the system event log. System events handled by the BIOS generate an SMI. After the BIOS finishes logging the error it asserts the NMI if needed.

The BIOS normally generates a NMI event in response to fatal and uncorrectable errors to prevent continued system operation with corrupted data. Most operating systems halt the system in response to NMI. However, certain Linux releases do not halt the system in response to an NMI event and therefore do not provide effective containment of data corruption. BIOS Setup provides an option to either reset the system or assert NMI in response to a PCI System error. This option should be reconfigured to reset the system in response to fatal and uncorrectable errors for these Linux releases.

Error Sources and Types
One of the major server management requirements is to correctly and consistently handle system errors. System errors that can be enabled and disabled individually or as a group can be categorized as:

- Processor errors
- Memory errors
- Legacy PCI and PCI-X* errors

- PCI Express* errors
- Sensor events / errors

Processor Errors
The BIOS enables the error correction and detection capabilities of the processors by setting appropriate bits in the processor Model Specific Register (MSR) set and the appropriate bits inside the chipset.

In the case of unrecoverable errors on the host processor bus, proper execution of the asynchronous SMM error handler cannot be guaranteed and the handler cannot be relied upon to log such conditions. The handler records the error to the system event log only if the system has not experienced a catastrophic failure that compromises the integrity of the handler.

Internal Error (IERR) and Thermal Trip
The BIOS contains no runtime handlers for processor IERR or thermal trip events. The system relies on the BMC to detect and log these errors at runtime. The BIOS subsequently determines the processor status during POST using the BMC Get Processor State command.

If the BMC reports either an IERR or thermal trip event on the previous boot, then the BIOS displays an error message in the POST Error Manager and continues normal operation.

If a persistent status sensor needs to be cleared (such as the Thermal Trip sensor), the user needs to select "Processor Retest" in the BIOS Setup utility Advanced | Processor page. The BIOS then instructs the BMC to re-arm its sensors.

Machine Check Errors
The BIOS clears all machine check error status banks on a power good reset and enables all machine check errors during POST.

The BIOS installs a default machine check handler during POST for legacy operating systems. This default handler resets the system in response to machine check events during runtime. It is assumed most operating systems install their own machine check handler. The BIOS does not report machine check errors.

Memory Errors
The hardware generates an SMI on both uncorrectable and correctable data errors in the memory array. Uncorrectable errors may corrupt the contents of SMRAM. The BIOS SMI handler logs the error and the failing FBDIMM number to the BMC if the SMRAM contents are still valid. The ability to isolate the failure down to a single FBDIMM may not be available on certain errors and/or during early POST.

Legacy PCI and PCI-X* Errors
The traditional PCI bus is a parallel bus mechanism that provides two sideband signals for error reporting. The PERR# signal reports parity errors and the SERR# signal reports all other system errors.

PCI data parity errors are not considered intrinsically fatal because the PCI bus master has the option to retry the offending transaction. The BIOS correspondingly logs a PERR SEL entry but does not halt the system. If the bus master cannot retry or if the retry fails, then the hardware escalates the error to a fatal SERR# event. All other PCI-related errors are considered fatal and reported by SERR#. The BIOS

handles SERR events by generating a SERR SEL entry and then triggers either a Non-Maskable Interrupt or system reset based on the BIOS Setup utility option ‘Reset on Fatal Error’.

The BIOS configures all PCI-to-PCI bridges so they generate SERR# on the primary interface whenever an SERR# assertion is detected on the secondary/downstream side. The server system does not support 32-bit PCI slots. The only traditional, 32-bit PCI device is the ATI embedded video* on a dedicated 32-bit legacy PCI bus controlled that the Intel® ESB2 controls. Video parity errors are not generally considered critical so the server system wires the PERR# signal on this bus to a pull-up connector providing a no connect functionality. PERR is not reported on the legacy PCI bus.

The server system does not have PCI-X* slots or embedded devices. However, it is expected that the system may be used with first generation PCI Express* adapters that are commonly organized as a PCI-X device behind a PCI Express to PCI-X bridge.

PCI Express* Errors
The server system supports a PCI Express*-based topology with all PCI devices downstream from the root ports. PCI Express devices report errors with an in-band messaging scheme based on these categories:

- Uncorrectable, fatal errors signaled with an ERR_FATAL message.
- Uncorrectable, non-fatal errors signaled with an ERR_NONFATAL message.
- Correctable errors signaled with an ERR_COR message.

Legacy Error Reporting Scheme
The BIOS supports a legacy error reporting scheme based on SERR and PERR reporting only. The BIOS configures all PCI Express* root ports and downstream devices such that all PCI Express uncorrectable fatal and uncorrectable non-fatal error messages are simultaneously reported as SERR in the standard configuration space PCI Status register SERR reporting bit for the device. In other words, fatal and non-fatal error messages are both considered critical errors requiring a system halt or reset to provide containment.

Error-handling Algorithm
The BIOS error-handling algorithm scans from the chipset PCI Express* root ports recursively downstream through all bridges identifying if any downstream device has flagged an SERR or PERR event. The handler logs a SEL entry for each PCI device reporting the error using the IPMI Critical Interrupt SERR or PERR sensor offset. Thus, the BIOS handler reports a chain of error events starting with the highest-level device reporting an error and proceeding through all intermediate reporting agents (PCI bridge devices) to the lowest level device reporting the error.

The BIOS does not report PCI Express* errors flagged in either the baseline capability structure or the optional Advanced Error Reporting (AER) structure directly as the Intelligent Platform Management Interface Specification, Version 2.0, Intel Corporation provides no support for reporting PCI Express* errors. Any PCI Express uncorrectable, fatal or uncorrectable, non-fatal errors are propagated to SERR so they are captured by our legacy error handler.

Parity Error Reporting
Parity error reporting is a legacy concept based on parallel bus implementations in the PCI Local Bus

Specification. PCI Express* is based on CRC protected, serial communications therefore the entire concept of parity error reporting is invalid. Consequently, the BIOS is not expected to encounter any parity error events directly on PCI Express bus topologies.

One case in which a PCI parity error may be relevant in a PCI Express system is in the event of an embedded traditional PCI device or PCI-X* device located behind a PCI Express to PCI or PCI-X bridge device. In this event, the downstream PCI / PCI-X device reports a PERR event, and then the PCI Express bridge device converts the signal into an uncorrectable error message on the primary interface of the bridge.

This uncorrectable error message is propagated upstream to the root port and signaled as an SERR in the PCI Status register on the primary side of bridge device. Our PCI error handler detects the PERR event on the downstream device and an SERR event on the primary side of the bridge and all upstream agents to the root port.

The server system does not support embedded PCI Express to PCI-X* bridges or PCI-X slots. The only case in which this situation might be encountered is for a PCI Express* adapter card using a PCI Express* to PCI-X* bridge onboard the adapter with a PCI-X* device behind the bridge. This is quite common in early generation PCI Express* devices. PCI Express* correctable error messages are not reported by the BIOS. These errors are scrubbed by chipset and/or device hardware and there is no need to report them in the absence of any frequency monitoring software that provides fault prediction analysis.

Sensor Events/Errors
The BMC manages sensors. It can receive event messages from individual sensors and logging system events.

System Event Logging (SEL) Format Conventions
The BIOS complies with the logging format defined in the Intelligent Platform Management Interface Specification, Version 2.0. See Table 32-1, SEL Event Records. This section describes the format used by the BIOS to create SEL entries for reporting certain system errors and events.

The following table indicates standard header fields common to all SEL entries that the BIOS logs:

Table 83. SEL Entry Format — Generic Fields

Byte	Field	IPMI Description	BIOS Implementation
1 2	Record ID	ID used for SEL record access.	The BMC logs unique Record ID.
3	Record Type	Bit[7:0] = Record Type ▪ 02h = system event record ▪ C0h–DFh = OEM, Byte 8-16 ▪ E0h–FFh = OEM, Byte 4-16	The BIOS sends this to the BMC: ▪ Bit[7:0] Record Type = 02h ▪ Currently all SEL entries logged as standard IPMI System Event Record entries
4 5 6 7	Timestamp	▪ Time when event was logged ▪ LS byte first	The BMC logs timestamp value.
8 9	Generator ID	Byte 8 ▪ Bit[7:1] = System software ID or IPMB slave address. ▪ Bit[0] ▪ 0b = IPMB slave address ▪ 1b = System Software ID Byte 9	The BIOS sends this to the BMC: Byte 8 = 0x33: ▪ Bit[7:4] System Software ID (SSID) ▪ 0011b = BIOS SMI Error Handler ID ▪ Bit[3:1] Custom sub-field

Byte	Field	IPMI Description	BIOS Implementation
		▪ Bit[7:4] = Channel number ▪ Bit[3:2] = reserved (00b) ▪ Bit[1:0] = IPMB device LUN or 00h	▪ 001b = Event Data format revision 1 ▪ Bit[0] ▪ 1b = System Software ID Byte 9 = 0x00: ▪ Bit[7:4] = 0h (channel = system interface) ▪ Bit[3:2] = 00b ▪ Bit[1:0] = 00b (Byte 8 is SSID)
10	EvM Rev	Event Message format version	The BIOS sends this to the BMC: 0x04 = as per the *Intelligent Platform Management Interface Specification*, Version 2.0, Intel Corporation requirement
13	Event Dir | Event Type	Bit[7] = Event Direction ▪ 0b = assertion event ▪ 1b = deassertion event Bit[6:0] = Event Type Code ▪ See the *Intelligent Platform Management Interface Specification*, Version 2.0, Table 42-1	The BIOS sends this to the BMC: ▪ Bit[7] = 0b – Assertion event ▪ Bit[6:0] = 0x6F – sensor specific type code The BIOS logs all events using the sensor-specific type code category and discrete event trigger/sensor class.

SEL formats for system errors are created using pre-defined formats for specific the Intelligent Platform Management Interface Specification, Version 2.0-compliant Sensor Types. The following list describes the various the Intelligent Platform Management Interface Specification, Version 2.0-compliant Sensor Types used to report different types of system errors:

- IPMI Memory Sensor Events

- Memory ECC Correctable Errors
- Memory ECC Uncorrectable Errors
- Memory Correctable ECC Memory Error Logging Limit Reached

- IPMI System Firmware Progress Events
  - System Firmware Error (POST Error) on FRB-2 events
- IPMI Event Logging Disabled Events
  - Correctable Memory Error Logging Disabled
- IPMI Critical Interrupt Sensor Events
  - PCI Bus Legacy PERR events
  - PCI Bus Legacy SERR events
- Software NMI Events

Note: *Technically, the FRB-2 event is not logged by the SMI handler, but it uses the same Generator ID range as memory errors. This makes it easier for the BIOS and the event log parser code.*

IPMI Sensor Type Events — Memory
The BIOS is responsible for logging SEL entries for the following events according to the format described in the table below:

- Sensor Offset 00h — Correctable ECC Memory Error
- Sensor Offset 01h — Uncorrectable ECC Memory Error
- Sensor Offset 05h — Correctable ECC Memory Error Logging Limit Reached

See the Intelligent Platform Management Interface Specification, Version 2.0, Table 32-1 SEL Event Records and Table 42-3 Sensor Type Codes for the following implementation details:

Table 84. SEL Entry Format — Memory Sensor Type

Byte	Field	IPMI Description	BIOS Implementation
11	Sensor Type	See the *Intelligent Platform Management Interface Specification*, Version 2.0, Table 42-3 Sensor Type Code for allowable values.	The BIOS sends this to the BMC: ▪ 0x0C = Memory
12	Sensor Number	Number of sensor that generated this event.	The BIOS sends this to the BMC: ▪ 0x08 = Memory ECC Correctable Error ▪ 0x08 = Memory ECC Uncorrectable Error ▪ 0x09 = Correctable Error Logging Limit Reached
14	Event Data 1 (ED1)	▪ Bit [7:6] 00b = ED2 unspecified 10b = ED2 contains OEM value ▪ Bit [5:4] 00b = ED3 unspecified 10b = ED3 contains OEM value (The BIOS does not use encodings 01b or 11b for errors discussed in this document.). ▪ Bit [3:0] See the *Intelligent Platform Management Interface Specification*, Version 2.0, Table 42-3 Sensor Specific Offset appropriate for the event.	The BIOS sends this to the BMC: ▪ Bit[7:6] = 00b - ED2 unspecified ▪ Bit[5:4] = 10b - ED3 contains OEM value ▪ Bit[3:0] - Supported Sensor Offsets 0x0 - Correctable ECC Memory Error 0x1 - Uncorrectable ECC Memory Error 0x5 - Correctable ECC Error Threshold Reached
15	Event Data 2 (ED2)	Bit [7:0] OEM value or unspecified	The BIOS sends this to the BMC: ▪ 0xFF = ED2 unspecified
16	Event Data 3 (ED3)	Bit [7:0] OEM value or unspecified	The BIOS sends this to the BMC: ▪ Memory module/device identification ▪ Bit[7:6] = Index into SMBIOS Type 16 entry. ▪ This shall be the zero-based memory riser board number. ▪ Bit[5:0] = Index into SMBIOS Type17 entry for the failed FBDIMM.

Table 85. SEL Entry Format — Memory Sensor Type Examples

Error Type	Event Data 1	Event Data 2	Event Data 3
Correctable ECC Memory Error Memory Riser Board A, FBDIMM 6	0x20	0xFF	0x06 Bit [7:6] = 00 Bit [5:0] = 06
Uncorrectable ECC Memory Error Memory Riser Board B, FBDIMM 5	0x21	0xFF	0x45 Bits [7:6] = 01 Bits [5:0] = 05

IPMI Sensor Type Events — System Firmware Progress

The BIOS is responsible for logging SEL entries for BIOS POST errors using the IPMI System Firmware Progress sensor type. BIOS only supports the following sensor offset for this sensor type:

- Sensor Offset 00h — System Firmware Error (POST Error)

See the Intelligent Platform Management Interface Specification, Version 2.0, Intel Corporation Table 42-3 for the following implementation details:

Table 86. SEL Entry Format — System Firmware Progress Sensor Type

Byte	Field	IPMI Description	BIOS Implementation
11	Sensor Type	See the *Intelligent Platform Management Interface Specification*, Version 2.0, Table 42-3 for allowable values.	The BIOS sends this to the BMC: ▪ 0x0F = System Firmware Progress
12	Sensor Number	Number of sensor that generated this event.	The BIOS sends this to the BMC: 0x06 = POST Error (BIOS)
14	Event Data 1 (ED1)	Bit[7:6] ▪ 00b = ED2 unspecified ▪ 10b = ED2 contains OEM value Bit[5:4] ▪ 00b = ED3 unspecified ▪ 10b = ED3 contains OEM value (The BIOS does not use encodings 01b or 11b for errors discussed in this document.). Bit [3:0] See the *Intelligent Platform Management Interface Specification*, Version 2.0, Table 42-3 for the event.	The BIOS sends this to the BMC: ▪ Bit[7:6] = 10b - ED2 contains OEM value ▪ Bit[5:4] = 10b - ED3 contains OEM value ▪ Bit[3:0] - Supported Sensor Offsets 0x0 — System Firmware Error (POST Error)
15	Event Data 2 (ED2)	Bit [7:0] ▪ OEM value or unspecified	The BIOS sends this to the BMC: ▪ Bit[7:0] - Least Significant Byte of the 16-bit POST error code
16	Event Data 3 (ED3)	Bit [7:0] ▪ OEM value or unspecified	The BIOS sends this to the BMC: ▪ Bit[7:0] - Most Significant Byte of the 16-bit POST error code

Table 87. SEL Entry Format — System Firmware Progress Sensor Type Examples

Error Type	Event Data 1	Event Data 2	Event Data 3
POST Error Code 8190h	0xA0	0x90	0x81

IPMI Sensor Type Events — Event Logging Disabled

The BIOS logs SEL entries for the following events according to the format described in the table below:

- Sensor Offset 00h — Correctable Memory Error Logging Disabled

Note: *SEL record logging for ALL other Event Logging Disabled Sensor Offsets is the responsibility of the BMC (i.e. BIOS is not responsible for SEL records indicating SEL Full or Log Area Reset/Cleared).*

See the Intelligent Platform Management Interface Specification, Version 2.0, Table 42-3 for the following implementation details:

Table 88. SEL Entry Format — Event Logging Disabled Sensor Type

Byte	Field	IPMI Description	BIOS Implementation
11	Sensor Type	See the *Intelligent Platform Management Interface Specification*, Version 2.0, Table 42-3 for allowable values.	The BIOS sends this to the BMC: ▪ 0x10 = Event Logging Disabled
12	Sensor Number	Number of sensor that generated this event.	The BIOS sends this to the BMC: ▪ 0x08 = Correctable Memory Error Logging Disabled
14	Event Data 1 (ED1)	Bit [7:6] ▪ 00b = ED2 unspecified ▪ 10b = ED2 contains OEM value Bit [5:4] ▪ 00b = ED3 unspecified ▪ 10b = ED3 contains OEM value (The BIOS does not use encodings 01b or 11b for errors discussed in this document.). Bit [3:0] See the *Intelligent Platform Management Interface Specification*, Version 2.0, Table 42-3 for the event.	The BIOS sends this to the BMC: Bit[7:6] = 10b - ED2 contains OEM value Bit[5:4] = 00b - ED3 unspecified Bit[3:0] - Supported Sensor Offsets ▪ 0x0 - Correctable Memory Error Logging Disabled
15	Event Data 2 (ED2)	Bit [7:0] ▪ OEM value or unspecified	The BIOS sends this to the BMC: ▪ Memory module/device identification ▪ Bit [7:6] - Index into SMBIOS Type 16 entry ▪ This shall be the zero-based memory riser board number.

Byte	Field	IPMI Description	BIOS Implementation
			▪ Bit [5:0] - Index into SMBIOS Type17 record for the failed FBDIMM
16	Event Data 3 (ED3)	Bit [7:0] ▪ OEM value or unspecified	BIOS sends following information to BMC: ▪ 0xFF = ED3 unspecified

The following examples are provided to clarify SEL entry Event Data values for various events.

Table 89. SEL Entry Format — Event Logging Disabled Sensor Type Examples

Error Type	Event Data 1	Event Data 2	Event Data 3
Correctable Memory Error Logging Disabled Memory Riser Board A, FBDIMM 6	0x80	0x06	0xFF

IPMI Sensor Type Events — Critical Interrupt

The BIOS is responsible for logging SEL entries for the following events according to the format described in the table below:

- Sensor Offset 04h — PCI PERR Event
- Sensor Offset 05h — PCI SERR Event
- Sensor Offset 0Ah — Fatal NMI (port 61h, bit 7) Event

See the Intelligent Platform Management Interface Specification, Version 2.0, Intel Corporation Table 42-3 for the following implementation details:

Table 90. SEL Entry Format — Critical Interrupt Sensor Type

Byte	Field	IPMI Description	BIOS Implementation
11	Sensor Type	See the *Intelligent Platform Management Interface Specification*, Version 2.0, Intel Corporation Table 42-3 Sensor Type Code for allowable values.	The BIOS sends this to the BMC: ▪ 0x13 = Critical Interrupt
12	Sensor Number	Number of sensor that generated this event.	The BIOS sends this to the BMC: ▪ 0xEA = PCI PERR ▪ 0xEB = PCI SERR
14	Event Data 1 (ED1)	Bit [7:6] ▪ 00b = ED2 unspecified ▪ 10b = ED2 contains OEM value Bit [5:4] ▪ 00b = ED3 unspecified ▪ 10b = ED3 contains OEM value (The BIOS not use encodings 01b or 11b for errors discussed in this document.). Bit [3:0] See the *Intelligent Platform Management Interface Specification*, Version 2.0, Table 42-3 for the event.	The BIOS sends this to the BMC: Bit[7:6] = ED2 encoding varies by event Bit[5:4] = ED3 encoding varies by event Bit[3:0] - Supported Sensor Offsets ▪ 0x4 — PCI PERR ▪ 0x5 — PCI SERR ▪ 0x9 — Fatal NMI (port 61, bit7)
15	Event Data 2 (ED2)	Bit [7:0] ▪ OEM value or unspecified	The BIOS sends this to the BMC: For PCI PERR and SERR events: ▪ PCI bus number on which the reporting device resides ▪ If the error handler is unable to determine the PCI address of the device reporting the error, then ED1 reports ED2 as unspecified and ED2 contains 0xFF. For Fatal NMI events: ▪ 0xFF = ED2 unspecified
16	Event Data 3 (ED3)	Bit [7:0] ▪ OEM value or unspecified	The BIOS sends this to the BMC: For PCI PERR and SERR events: PCI device and function address of the reporting device in the following format: If the error handler is unable to determine the PCI address of the device reporting the error, then ED1 reports ED3 as unspecified and ED3 contains 0xFF. ▪ Bit[7:3] - Device number of the reporting PCI device ▪ Bit[2:0] - Function number. (Will always contain a zero if the device is not a multifunction device.) For Fatal NMI events: ▪ 0xFF = ED2 unspecified

The following examples are provided to clarify SEL entry Event Data values for various events.

Table 91. SEL Entry Format — Critical Interrupt Sensor Type Examples

Error Type	Event Data 1	Event Data 2	Event Data 3
PCI PERR, failing device is not known	0x04	0xFF	0xFF
PCI SERR, failing device is not known	0x05	0xFF	0xFF
PCI PERR on PCI Bus 5, Device 3, Function 1	0xA4	0x05	0x19 Bit[7:3] = 03 Bit[2:0] = 01
PCI SERR on PCI Bus 0 with device and function unknown	0x85	0x00	0xFF
Software NMI generated by SMI handler on fatal error	0x09	0xFF	0xFF

POST Progress Codes and Errors

The system BIOS complies with the EFI Framework POST Progress Code specification by reporting 32-bit status codes at various points during POST that contain class, subclass, and operation information. The class and subclass fields describe the type of hardware that is being initialized. The operation field represents the specific initialization activity.

The system BIOS generates the following types of notifications during POST:

- POST Error beep codes for fatal errors in early POST prior to video initialization
- POST codes displayed on the system board diagnostic LED array
- POST Error Manager

POST error codes are also logged in the BMC System Event Log (SEL).

POST Error Beep Codes

The BIOS notifies the user of fatal error conditions in early POST prior to system video initialization using the following beep codes:

Table 92. POST Error Beep Codes

Beeps	Error Message	Description
3	Memory error	System halted because a fatal error related to the memory was detected.
6	BIOS rolling back error	The system has detected a corrupted BIOS in the flash part, and is rolling back to the last good BIOS.

POST Codes

The system BIOS truncates 32-bit EFI POST Progress Codes to 8-bit values for display on the system board Diagnostic LED array.

The resulting 8-bit POST code is displayed on the system board POST Code Diagnostic LED array at the start of each configuration process. This information can be used to assist with debugging system hangs during POST by identifying the last POST process initiated by the BIOS.

Table 93. Post Codes and Messages

Progress Code	Progress Code Definition
Host Processor	
0x10	Power-on initialization of the host processor (Boot Strap Processor)
0x11	Host processor cache initialization (including AP)
0x12	Starting Application processor initialization
0x13	SMM initialization
Chipset	

Progress Code	Progress Code Definition
0x21	Initializing a chipset component
Memory	
0xE1	No memory available (system halted)
0xE4	BIOS cannot communicate with FBDIMM (serial channel hardware failure)
0xE6	FBDIMM(s) failed Memory iBIST or Memory Link Training failure
0xEB	FBDIMM with corrupted SPD data detected (system halted)
0x22	Reading configuration data from memory (SPD on FBDIMM)
0x23	Detecting presence of memory
0x24	Programming timing parameters in the memory controller
0x25	Configuring memory parameters in the memory controller
0x26	Optimizing memory controller settings
0x27	Initializing memory, such as ECC init
0x28	Testing memory
PCI Bus	
0x50	Enumerating PCI buses
0x51	Allocating resources to PCI buses
0x52	Hot-plug PCI controller initialization
0x53-0x57	Reserved for PCI Bus
USB	
0x58	Resetting USB bus
0x59	Reserved for USB devices
ATA / ATAPI / SATA	
0x5A	Resetting SATA bus and all devices
0x5B	Reserved for ATA
SMBUS	
0x5C	Resetting SMBUS
0x5D	Reserved for SMBUS
Local Console	
0x70	Resetting the video controller (VGA)
0x71	Disabling the video controller (VGA)
0x72	Enabling the video controller (VGA)
Remote Console	
0x78	Resetting the console controller
0x79	Disabling the console controller
0x7A	Enabling the console controller
Keyboard (only USB)	
0x90	Resetting the keyboard

Progress Code	Progress Code Definition
0x91	Disabling the keyboard
0x92	Detecting the presence of the keyboard
0x93	Enabling the keyboard
0x94	Clearing keyboard input buffer
0x95	Instructing keyboard controller to run Self Test (PS2 only)
Mouse (only USB)	
0x98	Resetting the mouse
0x99	Detecting the mouse
0x9A	Detecting the presence of mouse
0x9B	Enabling the mouse
Fixed Media	
0xB0	Resetting fixed media device
0xB1	Disabling fixed media device
0xB2	Detecting presence of a fixed media device (hard drive detection, etc.)
0xB3	Enabling/configuring a fixed media device
Removable Media	
0xB8	Resetting removable media device
0xB9	Disabling removable media device
0xBA	Detecting presence of a removable media device (CDROM detection, etc.)
0xBC	Enabling/configuring a removable media device
Boot Device Selection	
0xDy	Trying boot selection y (where y = 0 to F)
Pre-EFI Initialization (PEI) Core	
0xE0	Started dispatching early initialization modules (PEIM)
0xE2	Initial memory found, configured, and installed correctly
0xE1,0xE3	Reserved for initialization module use (PEIM)
Driver eXecution Environment (DXE) Core	
0xE4	Entered EFI driver execution phase (DXE)
0xE5	Started dispatching drivers
0xE6	Started connecting drivers
DXE Drivers	
0xE7	Waiting for user input
0xE8	Checking password
0xE9	Entering BIOS Setup
0xEA	Flash Update
0xEE	Calling Int 19. One beep unless silent boot is enabled.
0xEF	Unrecoverable Boot failure

Progress Code	Progress Code Definition
Runtime Phase / EFI Operating System Boot	
0xF4	Entering sleep state
0xF5	Exiting sleep state
0xF8	Operating system has requested EFI to close boot services ExitBootServices ( ) has been called
0xF9	Operating system has switched to virtual address mode SetVirtualAddressMap ( ) has been called
0xFA	Operating system has requested the system to reset ResetSystem () has been called
Pre-EFI Initialization Module (PEIM) / Recovery	
0x30	Crisis recovery has been initiated because of a user request
0x31	Crisis recovery has been initiated by software (corrupt flash)
0x34	Loading crisis recovery capsule
0x35	Handing off control to the crisis recovery capsule
0x3F	Unable to complete crisis recovery.

POST Error Manager Messages and Handling
The POST Error Manager displays error messages reported by the system BIOS during POST.

The system BIOS truncates the 32-bit EFI POST Progress Code associated with the error to 16-bit values for display in the POST Error Manager.

The POST Error Manager behavior in response to the error is defined by the error severity reported by the BIOS. Errors are categorized in one of three severity levels. The system behavior in response to severity level of fatal, major, or minor

Fatal
The system behavior in response to fatal error is described below:

- BIOS logs an error to the POST Error Manager.
- BIOS logs an error message to the BMC System Event Log (SEL).
- BIOS unconditionally enters POST Error Manager to display error message.
- BIOS halts the system to prevent boot.
- The user needs to replace the faulty part and restart the system.

Major
The system behavior in response to major error is described below:

- BIOS logs an error to the POST Error Manager.
- BIOS logs an error message to the BMC System Event Log (SEL).
- The BIOS continues booting in a degraded state by default (i.e. BIOS does not automatically enter the POST Error Manager to display the error message).
- The user can override this default behavior by configuring the BIOS Setup POST Error Pause

option to Enabled. This forces the system to enter the POST Error Manager and display the error message before booting.

- The user can choose to take immediate corrective action or continue booting.

Minor

The system behavior in response to minor error is described below:

- BIOS logs an error to the POST Error Manager.
- BIOS continues booting with a degraded state (i.e. BIOS does not automatically enter the POST Error Manager to display the error message).
- The user may want to replace the erroneous unit.

The POST Error Manager reports a maximum of 500 errors on any single boot cycle. Errors are automatically cleared from the Error Manager on each boot.

Table 94. POST Error Manager Messages and Handling

POST Error Code	POST Error Manager Message	Error Severity
0012	CMOS date / time not set	Major
004C	Keyboard / interface error	Major
0108	Keyboard component encountered a locked error.	Minor
0109	Keyboard component encountered a stuck key error.	Minor
0113	Fixed Media The SAS RAID firmware can not run properly. The user should attempt to reflash the firmware.	Major
0140	PCI component encountered a PERR error.	Major
0141	PCI resource conflict	Major
0146	PCI out of resources error	Major
0192	Cache size mismatch	Fatal
0194	CPUID, processor family are different	Fatal
0195	Front side bus mismatch	Fatal
0196	Processor Model mismatch	Major
0197	Processor speeds mismatched	Fatal
0198	Processor family is unsupported.	Major
019A	Processor voltage mismatch detected	Fatal
5220	CMOS/NVRAM Configuration Cleared	Major
5221	Passwords cleared by jumper	Major
5224	Password clear Jumper is Set.	Major
8110	Processor 01 Internal Error (IERR) on last boot	Major
8111	Processor 02 Internal Error (IERR) on last boot	Major
8112	Processor 03 Internal Error (IERR) on last boot	Major
8113	Processor 04 Internal Error (IERR) on last boot	Major

POST Error Code	POST Error Manager Message	Error Severity
8120	Processor 01 thermal trip error on last boot	Major
8121	Processor 02 thermal trip error on last boot	Major
8122	Processor 03 thermal trip error on last boot	Major
8123	Processor 04 thermal trip error on last boot	Major
8130	Processor 01 disabled	Minor
8131	Processor 02 disabled	Minor
8132	Processor 03 disabled	Minor
8133	Processor 04 disabled	Minor
8160	Processor 01 unable to apply microcode update	Major
8161	Processor 02 unable to apply microcode update	Major
8162	Processor 03 unable to apply microcode update	Major
8163	Processor 04 unable to apply microcode update	Major
8180	Processor 01 microcode update not found	Minor
8181	Processor 02 microcode update not found	Minor
8182	Processor 03 microcode update not found	Minor
8183	Processor 04 microcode update not found	Minor
8190	Watchdog timer failed on last boot	Major
8198	Operating system boot watchdog timer expired on last boot	Major
8300	Baseboard management controller failed self-test	Major
8305	Hot swap controller failed	Major
84F2	Baseboard management controller failed to respond	Major
84F3	Baseboard management controller in update mode	Major
84F4	Sensor data record empty	Major
84FF	System event log full	Minor
8500	Memory component could not be configured in the selected RAS mode.	Major
8520	Memory failed Self Test (BIST). Memory Board A, DIMM_1.	Major
8521	Memory failed Self Test (BIST). Memory Board A, DIMM_2.	Major
8522	Memory failed Self Test (BIST). Memory Board A, DIMM_3.	Major
8523	Memory failed Self Test (BIST). Memory Board A, DIMM_4.	Major
8524	Memory failed Self Test (BIST). Memory Board A, DIMM_5.	Major
8525	Memory failed Self Test (BIST). Memory Board A, DIMM_6.	Major
8526	Memory failed Self Test (BIST). Memory Board A, DIMM_7.	Major
8527	Memory failed Self Test (BIST). Memory Board A, DIMM_8.	Major
8528	Memory failed Self Test (BIST). Memory Board B, DIMM_1.	Major
8529	Memory failed Self Test (BIST). Memory Board B, DIMM_2.	Major
852A	Memory failed Self Test (BIST). Memory Board B, DIMM_3.	Major
852B	Memory failed Self Test (BIST). Memory Board B, DIMM_4.	Major
852C	Memory failed Self Test (BIST). Memory Board B, DIMM_5.	Major
852D	Memory failed Self Test (BIST). Memory Board B, DIMM_6.	Major
852E	Memory failed Self Test (BIST). Memory Board B, DIMM_7.	Major
852F	Memory failed Self Test (BIST). Memory Board B, DIMM_8.	Major
8530	Memory failed Self Test (BIST). Memory Board C, DIMM_1.	Major
8531	Memory failed Self Test (BIST). Memory Board C, DIMM_2.	Major
8532	Memory failed Self Test (BIST). Memory Board C, DIMM_3.	Major

POST Error Code	POST Error Manager Message	Error Severity
8533	Memory failed Self Test (BIST). Memory Board C, DIMM_4.	Major
8534	Memory failed Self Test (BIST). Memory Board C, DIMM_5.	Major
8535	Memory failed Self Test (BIST). Memory Board C, DIMM_6.	Major
8536	Memory failed Self Test (BIST). Memory Board C, DIMM_7.	Major
8537	Memory failed Self Test (BIST). Memory Board C, DIMM_8.	Major
8538	Memory failed Self Test (BIST). Memory Board D, DIMM_1.	Major
8539	Memory failed Self Test (BIST). Memory Board D, DIMM_2.	Major
853A	Memory failed Self Test (BIST). Memory Board D, DIMM_3.	Major
853B	Memory failed Self Test (BIST). Memory Board D, DIMM_4.	Major
853C	Memory failed Self Test (BIST). Memory Board D, DIMM_5.	Major
853D	Memory failed Self Test (BIST). Memory Board D, DIMM_6.	Major
853E	Memory failed Self Test (BIST). Memory Board D, DIMM_7.	Major
853F	Memory failed Self Test (BIST). Memory Board D, DIMM_8.	Major
8540	Memory Board A, DIMM_1 Disabled.	Major
8541	Memory Board A, DIMM_2 Disabled.	Major
8542	Memory Board A, DIMM_3 Disabled.	Major
8543	Memory Board A, DIMM_4 Disabled.	Major
8544	Memory Board A, DIMM_5 Disabled.	Major
8545	Memory Board A, DIMM_6 Disabled.	Major
8546	Memory Board A, DIMM_7 Disabled.	Major
8547	Memory Board A, DIMM_8 Disabled.	Major
8548	Memory Board B, DIMM_1 Disabled.	Major
8549	Memory Board B, DIMM_2 Disabled.	Major
854A	Memory Board B, DIMM_3 Disabled.	Major
854B	Memory Board B, DIMM_4 Disabled.	Major
854C	Memory Board B, DIMM_5 Disabled.	Major
854D	Memory Board B, DIMM_6 Disabled.	Major
854E	Memory Board B, DIMM_7 Disabled.	Major
854F	Memory Board B, DIMM_8 Disabled.	Major
8550	Memory Board C, DIMM_1 Disabled.	Major
8551	Memory Board C, DIMM_2 Disabled.	Major
8552	Memory Board C, DIMM_3 Disabled.	Major
8553	Memory Board C, DIMM_4 Disabled.	Major
8554	Memory Board C, DIMM_5 Disabled.	Major
8555	Memory Board C, DIMM_6 Disabled.	Major
8556	Memory Board C, DIMM_7 Disabled.	Major
8557	Memory Board C, DIMM_8 Disabled.	Major
8558	Memory Board D, DIMM_1 Disabled.	Major
8559	Memory Board D, DIMM_2 Disabled.	Major
855A	Memory Board D, DIMM_3 Disabled.	Major
855B	Memory Board D, DIMM_4 Disabled.	Major
855C	Memory Board D, DIMM_5 Disabled.	Major
855D	Memory Board D, DIMM_6 Disabled.	Major
855E	Memory Board D, DIMM_7 Disabled.	Major

POST Error Code	POST Error Manager Message	Error Severity
855F	Memory Board D, DIMM_8 Disabled.	Major
8560	Memory Board A, DIMM_1 Component encountered a Serial Presence Detection (SPD) fail error.	Major
8561	Memory Board A, DIMM_2 Component encountered a Serial Presence Detection (SPD) fail error.	Major
8562	Memory Board A, DIMM_3 Component encountered a Serial Presence Detection (SPD) fail error.	Major
8563	Memory Board A, DIMM_4 Component encountered a Serial Presence Detection (SPD) fail error.	Major
8564	Memory Board A, DIMM_5 Component encountered a Serial Presence Detection (SPD) fail error.	Major
8565	Memory Board A, DIMM_6 Component encountered a Serial Presence Detection (SPD) fail error.	Major
8566	Memory Board A, DIMM_7 Component encountered a Serial Presence Detection (SPD) fail error.	Major
8567	Memory Board A, DIMM_8 Component encountered a Serial Presence Detection (SPD) fail error.	Major
8568	Memory Board B, DIMM_1 Component encountered a Serial Presence Detection (SPD) fail error.	Major
8569	Memory Board B, DIMM_2 Component encountered a Serial Presence Detection (SPD) fail error.	Major
856A	Memory Board B, DIMM_3 Component encountered a Serial Presence Detection (SPD) fail error.	Major
856B	Memory Board B, DIMM_4 Component encountered a Serial Presence Detection (SPD) fail error.	Major
856C	Memory Board B, DIMM_5 Component encountered a Serial Presence Detection (SPD) fail error.	Major
856D	Memory Board B, DIMM_6 Component encountered a Serial Presence Detection (SPD) fail error.	Major
856E	Memory Board B, DIMM_7 Component encountered a Serial Presence Detection (SPD) fail error.	Major
856F	Memory Board B, DIMM_8 Component encountered a Serial Presence Detection (SPD) fail error.	Major
8570	Memory Board C, DIMM_1 Component encountered a Serial Presence Detection (SPD) fail error.	Major
8571	Memory Board C, DIMM_2 Component encountered a Serial Presence Detection (SPD) fail error.	Major
8572	Memory Board C, DIMM_3 Component encountered a Serial Presence Detection (SPD) fail error.	Major
8573	Memory Board C, DIMM_4 Component encountered a Serial Presence Detection (SPD) fail error.	Major
8574	Memory Board C, DIMM_5 Component encountered a Serial Presence Detection (SPD) fail error.	Major
8575	Memory Board C, DIMM_6 Component encountered a Serial Presence Detection (SPD) fail error.	Major
8576	Memory Board C, DIMM_7 Component encountered a Serial Presence Detection (SPD) fail error.	Major
8577	Memory Board C, DIMM_8 Component encountered a Serial Presence Detection (SPD) fail error.	Major
8578	Memory Board D, DIMM_1 Component encountered a Serial Presence Detection (SPD) fail error.	Major

POST Error Code	POST Error Manager Message	Error Severity
8579	Memory Board D, DIMM_2 Component encountered a Serial Presence Detection (SPD) fail error.	Major
857A	Memory Board D, DIMM_3 Component encountered a Serial Presence Detection (SPD) fail error.	Major
857B	Memory Board D, DIMM_4 Component encountered a Serial Presence Detection (SPD) fail error.	Major
857C	Memory Board D, DIMM_5 Component encountered a Serial Presence Detection (SPD) fail error.	Major
857D	Memory Board D, DIMM_6 Component encountered a Serial Presence Detection (SPD) fail error.	Major
857E	Memory Board D, DIMM_7 Component encountered a Serial Presence Detection (SPD) fail error.	Major
857F	Memory Board D, DIMM_8 Component encountered a Serial Presence Detection (SPD) fail error.	Major
8580	Memory Board A, DIMM_1 Correctable ECC error encountered.	Minor/Major after 10
8581	Memory Board A, DIMM_2 Correctable ECC error encountered.	Minor/Major after 10
8582	Memory Board A, DIMM_3 Correctable ECC error encountered.	Minor/Major after 10
8583	Memory Board A, DIMM_4 Correctable ECC error encountered.	Minor/Major after 10
8584	Memory Board A, DIMM_5 Correctable ECC error encountered.	Minor/Major after 10
8585	Memory Board A, DIMM_6 Correctable ECC error encountered.	Minor/Major after 10
8586	Memory Board A, DIMM_7 Correctable ECC error encountered.	Minor/Major after 10
8587	Memory Board A, DIMM_8 Correctable ECC error encountered.	Minor/Major after 10
8588	Memory Board B, DIMM_1 Correctable ECC error encountered.	Minor/Major after 10
8589	Memory Board B, DIMM_2 Correctable ECC error encountered.	Minor/Major after 10
858A	Memory Board B, DIMM_3 Correctable ECC error encountered.	Minor/Major after 10
858B	Memory Board B, DIMM_4 Correctable ECC error encountered.	Minor/Major after 10
858C	Memory Board B, DIMM_5 Correctable ECC error encountered.	Minor/Major after 10
858D	Memory Board B, DIMM_6 Correctable ECC error encountered.	Minor/Major after 10
858E	Memory Board B, DIMM_7 Correctable ECC error encountered.	Minor/Major after 10
858F	Memory Board B, DIMM_8 Correctable ECC error encountered.	Minor/Major after 10
8590	Memory Board C, DIMM_1 Correctable ECC error encountered.	Minor/Major after 10
8591	Memory Board C, DIMM_2 Correctable ECC error encountered.	Minor/Major after 10

POST Error Code	POST Error Manager Message	Error Severity
8592	Memory Board C, DIMM_3 Correctable ECC error encountered.	Minor/Major after 10
8593	Memory Board C, DIMM_4 Correctable ECC error encountered.	Minor/Major after 10
8594	Memory Board C, DIMM_5 Correctable ECC error encountered.	Minor/Major after 10
8595	Memory Board C, DIMM_6 Correctable ECC error encountered.	Minor/Major after 10
8596	Memory Board C, DIMM_7 Correctable ECC error encountered.	Minor/Major after 10
8597	Memory Board C, DIMM_8 Correctable ECC error encountered.	Minor/Major after 10
8598	Memory Board D, DIMM_1 Correctable ECC error encountered.	Minor/Major after 10
8599	Memory Board D, DIMM_2 Correctable ECC error encountered.	Minor/Major after 10
859A	Memory Board D, DIMM_3 Correctable ECC error encountered.	Minor/Major after 10
859B	Memory Board D, DIMM_4 Correctable ECC error encountered.	Minor/Major after 10
859C	Memory Board D, DIMM_5 Correctable ECC error encountered.	Minor/Major after 10
859D	Memory Board D, DIMM_6 Correctable ECC error encountered.	Minor/Major after 10
859E	Memory Board D, DIMM_7 Correctable ECC error encountered.	Minor/Major after 10
859F	Memory Board D, DIMM_8 Correctable ECC error encountered.	Minor/Major after 10
85A0	Memory Board A, DIMM_1 Uncorrectable ECC error encountered.	Major
85A1	Memory Board A, DIMM_2 Uncorrectable ECC error encountered.	Major
85A2	Memory Board A, DIMM_3 Uncorrectable ECC error encountered.	Major
85A3	Memory Board A, DIMM_4 Uncorrectable ECC error encountered.	Major
85A4	Memory Board A, DIMM_5 Uncorrectable ECC error encountered.	Major
85A5	Memory Board A, DIMM_6 Uncorrectable ECC error encountered.	Major
85A6	Memory Board A, DIMM_7 Uncorrectable ECC error encountered.	Major
85A7	Memory Board A, DIMM_8 Uncorrectable ECC error encountered.	Major
85A8	Memory Board B, DIMM_1 Uncorrectable ECC error encountered.	Major
85A9	Memory Board B, DIMM_2 Uncorrectable ECC error encountered.	Major
85AA	Memory Board B, DIMM_3 Uncorrectable ECC error encountered.	Major
85AB	Memory Board B, DIMM_4 Uncorrectable ECC error encountered.	Major
85AC	Memory Board B, DIMM_5 Uncorrectable ECC error encountered.	Major
85AD	Memory Board B, DIMM_6 Uncorrectable ECC error encountered.	Major
85AE	Memory Board B, DIMM_7 Uncorrectable ECC error encountered.	Major
85AF	Memory Board B, DIMM_8 Uncorrectable ECC error encountered.	Major
85B0	Memory Board C, DIMM_1 Uncorrectable ECC error encountered.	Major
85B1	Memory Board C, DIMM_2 Uncorrectable ECC error encountered.	Major
85B2	Memory Board C, DIMM_3 Uncorrectable ECC error encountered.	Major
85B3	Memory Board C, DIMM_4 Uncorrectable ECC error encountered.	Major

POST Error Code	POST Error Manager Message	Error Severity
85B4	Memory Board C, DIMM_5 Uncorrectable ECC error encountered.	Major
85B5	Memory Board C, DIMM_6 Uncorrectable ECC error encountered.	Major
85B6	Memory Board C, DIMM_7 Uncorrectable ECC error encountered.	Major
85B7	Memory Board C, DIMM_8 Uncorrectable ECC error encountered.	Major
85B8	Memory Board D, DIMM_1 Uncorrectable ECC error encountered.	Major
85B9	Memory Board D, DIMM_2 Uncorrectable ECC error encountered.	Major
85BA	Memory Board D, DIMM_3 Uncorrectable ECC error encountered.	Major
85BB	Memory Board D, DIMM_4 Uncorrectable ECC error encountered.	Major
85BC	Memory Board D, DIMM_5 Uncorrectable ECC error encountered.	Major
85BD	Memory Board D, DIMM_6 Uncorrectable ECC error encountered.	Major
85BE	Memory Board D, DIMM_7 Uncorrectable ECC error encountered.	Major
85BF	Memory Board D, DIMM_8 Uncorrectable ECC error encountered.	Major
85FC	Closed Loop Thermal Throttling could not be configured, defaulting to Open Loop.	Major
85FD	Memory was not configured for the selected Memory RAS Configuration.	Minor
8601	System booting from the other bank. Recovery Jumper is set to Recovery mode.	Minor
8602	WatchDog timer expired (secondary BIOS may be bad!)	Minor
8603	Secondary BIOS checksum fail	Minor
9000	Unspecified processor component has encountered a non specific error.	Major
9223	Keyboard component was not detected.	Minor
9226	Keyboard component encountered a controller error.	Minor
9243	Mouse component was not detected.	Minor
9246	Mouse component encountered a controller error.	Minor
9266	Local Console component encountered a controller error.	Minor
9268	Local Console component encountered an output error.	Minor
9269	Local Console component encountered a resource conflict error.	Minor
9286	Remote Console component encountered a controller error.	Minor
9287	Remote Console component encountered an input error.	Minor
9288	Remote Console component encountered an output error.	Minor
92A3	Serial port component was not detected	Major
92A9	Serial port component encountered a resource conflict error	Major
92C6	Serial Port controller error	Minor
92C7	Serial Port component encountered an input error.	Minor
92C8	Serial Port component encountered an output error.	Minor
94C6	LPC component encountered a controller error.	Minor
94C9	LPC component encountered a resource conflict error.	Major
9506	ATA/ATPI component encountered a controller error.	Minor
95A6	PCI component encountered a controller error.	Minor
95A7	PCI component encountered a read error.	Minor
95A8	PCI component encountered a write error.	Minor
9609	Unspecified software component encountered a start error.	Minor
9641	PEI Core component encountered a load error.	Minor
9667	PEI module component encountered a illegal software state error.	Fatal
9687	DXE core component encountered a illegal software state error.	Fatal

POST Error Code	POST Error Manager Message	Error Severity
96A7	DXE boot services driver component encountered a illegal software state error.	Fatal
96AB	DXE boot services driver component encountered invalid configuration.	Minor
96E7	SMM driver component encountered a illegal software state error.	Fatal
A000	TPM device not detected.	Minor
A001	TPM device missing or not responding.	Minor
A002	TPM device failure.	Minor
A003	TPM device failed self test.	Minor
A022	Processor component encountered a mismatch error.	Major
A027	Processor component encountered a low voltage error.	Minor
A028	Processor component encountered a high voltage error.	Minor
A421	PCI component encountered a SERR error.	Fatal
A500	ATA/ATPI ATA bus SMART not supported.	Minor
A501	ATA/ATPI ATA SMART is disabled.	Minor
A5A0	PCI Express component encountered a PERR error.	Minor
A5A1	PCI Express component encountered a SERR error.	Fatal
A5A4	PCI Express IBIST error.	Major
A6A0	DXE boot services driver Not enough memory available to shadow a legacy option ROM.	Minor

New Technologies

Intel® I/O Acceleration Technology (Intel® I/OAT)
The server system supports Intel® I/O Acceleration Technology (Intel® I/OAT) version 1.0 to improve network I/O performance. Intel® I/OAT support consists of both processor-based Direct Cache Access (DCA) and chipset-based Crystal Beach Technology.

Intel® I/OAT requires BIOS and operating system software support. The BIOS enables both DCA and Crystal Beach Technology during POST.

Trusted Platform Module (TPM) Security
The Trusted Platform Module (TPM) is a hardware-based security device that addresses the growing concern on boot process integrity and offers better data protection. TPM protects the system start-up process by ensuring that they are tamper-free before releasing system control to the operating system. A TPM device provides secured storage to store data, such as security keys and passwords. In addition, a TPM device has encryption and hash functions. The ProServ 48680 Server System implements TPM as per TPM PC Client specifications revision 1.2 by the Trusted Computing Group (TCG).

A TPM device is affixed to the motherboard of the server and is secured from external software attacks and physical theft. A pre-boot environment, such as the BIOS and operating system loader, use the TPM to collect and store unique measurements from multiple factors within the boot process to create a

system fingerprint. This unique fingerprint remains the same unless the pre-boot environment is tampered with. Therefore, it is used to compare to future measurements to verify the integrity of the booting process.

After the BIOS completes measurement of its boot process, it hands off control to the operating system loader and in turn to the operating system. If the operating system is TPM enabled, it compares the BIOS TPM measurements to those of previous boots to make sure that the system has not been tampered with before continuing the operating system boot process. Once the operating system is in operation, it optionally uses TPM to provide additional system and data security (for example, Microsoft Vista* supports Bitlocker drive encryption).

TPM Security BIOS
The BIOS TPM support conforms to the TPM PC Client Specific - Implementation Specification for Conventional BIOS, version 1.2, and to the TPM Interface specification, version 1.2. The BIOS adheres to the Microsoft Vista* BitLocker requirement. The role of the BIOS for TPM security includes the following:

- Measures and stores the boot process in the TPM microcontroller to allow a TPM enabled operating system to verify system boot integrity.
- Produces EFI and legacy interfaces to a TPM enabled operating system for utilizing TPM.
- Produces ACPI TPM device and methods to allow a TPM enabled operating system to send TPM administrative command requests to the BIOS.
- Verifies operator physical presence. Confirms and executes operating system TPM administrative command requests.
- Provides BIOS Setup options to change TPM security states and to clear TPM ownership.

See the TCG PC Client Specific Implementation Specification, the TCG PC Client Specific Physical Presence Interface Specification and the Microsoft BitLocker Requirement documents for more details.

Physical Presence
Administrative operations to the TPM require TPM ownership or the physical presence indication by the operator to confirm the execution of the administrative operations. The BIOS implements operator presence indication by verifying the setup Administrator password.

A TPM administrative sequence invoked from the operating system proceeds as follows:

- User makes a TPM administrative request through the operating system’s security software.
- The operating system requests the BIOS to execute the TPM administrative command through TPM ACPI methods, and then resets the system.
- The BIOS verifies the physical presence and confirms the command with the operator.
- The BIOS executes TPM administrative command(s), inhibits BIOS Setup entry and boots directly to the operating system which requested the TPM command(s).

TPM Security Setup Options
BIOS TPM Setup allows the operator to view the current TPM state and to carry out rudimentary TPM administrative operations. Performing TPM administrative options through BIOS setup requires TPM physical presence verification.

Using BIOS TPM Setup, the operator can turn ON or OFF TPM functionality and clear the TPM ownership contents. After the requested TPM BIOS Setup operation is carried out, the option reverts to “No Operation”.

BIOS TPM Setup also displays the current state of the TPM, whether TPM is enabled or disabled and activated or deactivated. Note that while utilizing TPM, a TPM enabled operating system or application may change the TPM state independent of BIOS setup. When an operating system modifies the TPM state, BIOS Setup displays the updated TPM state.

The BIOS Setup TPM Clear option allows the operator to clear the TPM ownership key and allows the operator to take control of the system with TPM. This option is used to clear security settings for a newly initialized system or to clear a system for which the TPM ownership security key has been lost.

Internet SCSI (iSCSI)
Internet SCSI (iSCSI) is an industry standard where a host platform sends SCSI commands over the internet to the remote target SCSI device. This feature can be supported in two ways: hardware-based iSCSI and software-based iSCSI.

The server system only supports the hardware-based iSCSI solution. In the hardware-based iSCSI solution, the entire iSCSI command engine is within the NIC adapter. The management of data over the internet via the TCP/IP protocol is done by the iSCSI initiator (in pre-boot space part of Option ROM) and operating system driver in operating system space.

The onboard NIC (Intel® 82563EB) acts as an iSCSI initiator. The storage sub-system attached through the Ethernet connection is referred to as the iSCSI Target.

The iSCSI solution uses the Intel® 82563EB network interface device behind the ESB2 South Bridge as an initiator by executing an Option ROM embedded into the system BIOS. This Option ROM contains 16-bit legacy code and is discovered during the PCI enumeration process.

BIOS Setup provides a menu item to enable or disable iSCSI support. By default, iSCSI is disabled. PXE and iSCSI functionality is mutually exclusive on the Intel® 82563EB Ethernet device because iSCSI or PXE configures access to all the available ports of the same network.

The user must therefore ensure the Intel 82563EB PXE Option ROM is disabled prior to enabling the Intel 82563EB iSCSI Option ROM.

The BIOS needs to provide boot parameters to the iSCSI target device used by the host system to boot from. Occasionally, these targets may have DHCP capability and their iSCSI OPROM image can acquire the IP for the initiator (itself) and setup its communication. Once the Option ROM is executed, the iSCSI device hooks up its drive information into the BBS table and the device appears in the BIOS Boot Manager menu during POST.

The Interrupt 13h interface is used to read the Master Boot Record and load the operating system when a boot attempt is made from the device attached to iSCSI.

The BIOS does not report any specific issues with iSCSI Option ROM initialization. This should be

handled by the iSCSI Option ROM itself similar to other Legacy OPROM images.

In case the BIOS runs out of available memory in the Option ROM shadow region while dispatching Option ROM, the issue is reported in the POST Error Manager.

Baseboard Management Controller (BMC)
This section describes functional and communication interfaces for the enterprise south bridge (ESB2) baseboard management controller (BMC) for the ProServ 4680 Server System. It describes the functional blocks of the BMC and the interactions among them. It describes the commands and codes necessary to access, control, and configure the BMC, and to create reliable communications with other controllers on the Intelligent Platform Management Bus (IPMB).

The Intelligent Platform Management Interface Specification v2.0 describes the communication interfaces.

ESB2 South Bridge
The ESB2 multi-function device merges four functions:

- Controller similar to the ICH6
- PCI-X* bridge
- GBe controller
- BMC

Each function has a set of configuration registers. Once configured, the server sees each register as a distinct hardware controller. The ESB2 provides the gateway to all PC-compatible I/O devices and features. The server board uses the following ESB2 features:

- PCI-X* bus interface
- Six-channel SATA interface with SATA-busy LED control
- Dual Gbe Media Access Control (MAC)
- Baseboard management controller
- Single ATA interface, with Ultra DMA 100 capability
- Universal Serial Bus (USB) 2.0 interface
- Low pin count (LPC) bus interface
- PC-compatible timer / counter and DMA controllers
- Advanced Programmable Interrupt Controller (APIC) and 8259-programmable interrupt controller
- Power management
- System real-time clock (RTC)
- General purpose I/O

Individual components of the ESB2 are referred to as separate components. Example: The ESB2 BMC component is referred to as the BMC, and the component that is similar to the I/O controller hub 6 (ICH6) is referred to as the I/O controller hub (ICH).

ESB2 Baseboard Management Controller Functionality
The BMC is provided by an embedded ARC* controller and associated peripheral functionality that is

required for IPMI-based server management.

The following is a summary of the ESB2 management hardware features utilized by the BMC:

- ARC4 processor with 16 Kbytes instruction cache (I-cache) and data cache (D-cache).
- 256 kbyte of internal SRAM with dual ports: one for code accesses and one for all other accesses.
- Expansion bus, allowing connection to asynchronous or synchronous external flash programmable read-only memory (PROM), external SRAM, or external SDRAM. Wait states are programmable.
- Serial flash interface.
- Five SMB ports, two that support fast management links (FML), either master or slave
- RS-232 serial port (UART).
- Cryptographic module, supporting
  - Advanced Encryption Standard (AES) algorithm
  - Rivest Cipher 4 (RC4) encryption algorithms
  - Secure Hash Algorithm 1 (SHA1) authentication algorithm with internal direct memory access (DMA) and raw checksum support.
  - Message Digest Algorithm 5 (MD5) authentication algorithm with internal direct memory access (DMA) and raw checksum support.
- Keyboard text (KT) interface.
- Universal host controller interface (UHCI): USB.
- KCS interface mapped to a PCI Express* function.
- Two additional KCS interfaces, controlled by the BIOS, and residing on the LPC bus.
- General-purpose input/output (GPIO) interface.
- Media Access Controller (MAC) control and status register (CSR) interface.
- Timer interface.
- Host DMA interface.

BMC Functional Specifications

Power System

The ESB2 BMC is not the power control path, but it can block power control actions that are caused from front panel power button presses or chipset-initiated power state changes. It can generate power state changes by simulating a front panel power button press. It monitors both the requested power state from the chipset and the power good state.

Power Supply Interface Signals

The ICH controls the POWER_ON signal. It connects to the chassis power sub-system and is used to request power state changes (asserted = request power on). The POWER_GOOD signal from the chassis power sub-system indicates the current power state (asserted = power is on).

To turn the system on, the BMC asserts the BMC_FP_POWER signal and waits for the POWER_GOOD signal to assert in response, indicating that DC power is on.

The POWER_GOOD signal is normally asserted within 1.5 seconds, but the timeout interval can be set

longer to add flexibility in manufacturing test environments. The POWER_GOOD signal must remain stable and not glitch when being asserted. The BMC uses the state of the

POWER_GOOD signal (indirectly monitored through another signal available to the BMC) to monitor whether the power supply is on and operational, and to confirm whether the system power state matches the intended system on / off power state that was commanded with the POWER_ON signal.

Power-Good Dropout
De-assertion of the POWER_GOOD signal generates an interrupt that the BMC uses to detect either power sub-system failure or loss of AC power. The BMC performs the power fault analysis according to section 22.19.4 to determine the cause of the POWER_GOOD signal dropout. A power-good dropout is defined as the POWER_GOOD signal de-asserting when the system should be in the DC power-on state as determined by the state of the POWER_ON signal. If the BMC detects a power-good dropout, the following occurs:

- The BMC powers down the system.
- The BMC asserts the Power Unit Failure offset of the Power Unit sensor and logs a SEL event.
- The BMC generates a beep code for a Power Fault.
- The BMC waits 10 seconds. If the power state retention feature is configured to power on the server after an AC loss, it attempts to power up the server. This is the case in which either AC dropped out momentarily, but not long enough to reset the BMC, or the power sub-system had a momentary failure that the BMC could not differentiate from a momentary AC loss.

The BMC responds to the power loss interrupt within 1-2 ms if it is in operational mode. The BMC does not respond to a power-good dropout if it is in firmware transfer mode.

Power-up Sequence
To power up the server, the BMC simulates the front panel power button being pressed for 8 seconds or until POWER_GOOD is asserted. If POWER_GOOD is not asserted within 8 seconds, then a fault is generated.

Power Down Sequence
To power down the system, the BMC simulates the front panel power button being pressed for 2 seconds or until POWER_GOOD is deasserted. If POWER_GOOD is not deasserted within 2 seconds, then a fault is generated.

Before initiating the system power down, the BMC stops scanning any sensors that should not be scanned in the powered-down state.

Power Control Sources
The following sources can initiate power-up and / or power-down activity.

Power Button Signal
The POWER_BUTTON signal toggles the system power. This signal is activated by a momentary contact switch on the front panel assembly and is routed to the BMC as a bidirectional signal. The BMC de-bounces and monitors the signal. The signal must be in a constant state for 50 ms before it is treated as asserted.

The signal is routed to the CHIPSET_PWR_BUTTON signal through blocking circuitry that allows the BMC to lock out the signal. The chipset responds to the assertion of the signal; it reacts to the press of the button, not the release of it. The chipset does not respond when secure mode is enabled and active.

Chipset Sleep S5
The BMC monitors the sleep S5 signal to provide an indication of power state change requests. The S5 signal is only used for monitoring S5 transitions because S4 is not supported. This signal is the same as the POWER_ON signal. It is routed to the power sub-system.

The BMC requires the sleep S5 signal to maintain its level for at least 15 ms to be recognized. This signal can change state as a result of the following events:

- Operating system request
- Real-time clock (RTC) alarm
- Chipset power button request response, including BMC-initiated power state changes

Power-On Enable
The BMC must assert the POWER_ON_ENABLE signal to enable the system to power-on. When AC power is applied, the BMC initializes this signal to a de-asserted state. After the BMC has completed a specific phase of its initialization it asserts this signal. This prevents potential race conditions between the BMC and the BIOS.

When the system transitions to the off state (S5), the BMC de-asserts the POWER_ON_ENABLE signal, blocking system power-on.

The BMC monitors the POWER_ON signal (SleepS5) to detect if the chipset is attempting to power on the system. If so, the BMC completes necessary transitional operations before asserting POWER_ON_ENABLE and allowing the system power-on to complete.

The BMC also uses the POWER_ON_ENABLE signal to enforce a minimum off-time whenever the system is shut down. This is currently set to 10 seconds.

Assertion of the FORCE_UPDATE jumper signal allows power on to occur. This handles the case in which the BMC operational code is not functional.

Power-down Disable
The BMC asserts POWER_DOWN_DISABLE to momentarily block system power-down while transitional operations are completed before allowing the system power-down to complete.

Power State Retention
The BMC persistently stores the latest power state that was attained do to a power state change initiator. This capability supports the power state restoration feature.

Power State Restoration
The BMC provides the ability to control the AC power-on behavior of the server. The Set Power Restore Policy command configures the BMC to restore the power state in one of three ways.

- Power always off – Leave power off when AC is restored.
- Power always on – Power server on when AC is restored.
- Restore power state – Restore power state to the state it was in when AC was lost.

When standby power returns after an AC power loss, the BMC activates the server power as directed by the configuration.

Wake-On-LAN (WOL)
The BMC does not directly participate in WOL. The NICs directly interact with the chipset to initiate the power on of the system. The BMC blocks power-on until it is ready for the power-on to occur. The BMC must detect when the system is trying to power on and assert the POWER_ON_ENABLE signal for a WOL-initiated DC power-on to occur.

Advanced Configuration and Power Interface (ACPI)
The BMC works with the ACPI BIOS and with the server board hardware.

Table 95. ACPI Power States

State	Supported	Description
S0	Yes	Working ▪ The front panel power LED is on (not controlled by the BMC). ▪ The fans spin at the normal speed, as determined by sensor inputs. ▪ Front panel buttons work normally.
S1	Yes	Sleeping. Hardware context maintained; equates to processor and chipset clocks stopped. ▪ The front panel power LED blinks at a rate of 1 Hz with a 50% duty cycle (not controlled by the BMC). ▪ If enabled via the Set ACPI Configuration Mode command, the server board fans are set to sleep speed as specified in the associated OEM TControl SDR for each fan domain. Otherwise, fan control is the same as for ACPI S0 state. The DIMM temperature sensors do not contribute to the fan speed control algorithm. ▪ The watchdog timer is stopped. ▪ The power, reset, front panel NMI, and ID buttons are unprotected. The BMC detects that the system has exited the ACPI S1 sleep state when it is notified by the BIOS SMI handler.
S2	No	Not supported
S3	No	Not supported
S4	No	Not supported
S5	Yes	Soft off. ▪ The front panel buttons are not locked. ▪ The front panel power LED is off ▪ The fans are stopped. ▪ The power up process goes through the normal boot process. ▪ The power, reset, front panel NMI, and ID buttons are unlocked.

ACPI Power Control
The chipset implements ACPI-compatible power control. Power control requests are routed to the power push-button input of the chipset, allowing the ACPI-compatible power push-button logic in the chipset to be used. To support secure mode, the BMC can block the power button signal.

ACPI State Synchronization

The BIOS keeps the BMC synchronized with the system ACPI state. The BIOS provides the ACPI state when the server transitions between the power and the sleep states. It uses the SMM interface to provide the ACPI state.

ACPI Power State Notify
If enabled through the Set ACPI Configuration Mode commands, the BMC sends the system's ACPI power state changes (S0, S1, and S5) to other management controllers by sending the Set ACPI Power State command on the IPMB as indicated by their SDR management device records. The command is sent whenever there is a power state transition.

System Reset Control

Reset Signal Output
The BMC simulates a press of the front panel reset button to perform a system reset. The ICH performs the rest of the system reset process. The BMC cannot hold the system in reset, and once started, the process is asynchronous with respect to BMC operation. The BMC provides system status indication through the front panel LEDs.

Reset Control Sources
All system resets result in the BMC running its sensor initialization agent service.

Table 96. System Reset Sources and Actions

Reset Source	System Reset?	BMC Reset
Standby power comes up	No (no DC power)	Yes
Main system power comes up	Yes	No
Reset button or in-target probe (ITP) reset	Yes	No
Soft reset / warm boot (DOS Ctrl-Alt-Del)	Yes	No
Hard reset	Yes	No
Command to reset the system	Yes	No
Set Processor State command	Yes	No
Watchdog timer configured for reset	Yes	No
PEF action	Optional	No
Exit BMC firmware update mode	No	Yes

Front Panel System Reset
The reset button is a momentary contact button on the front panel. Its signal is routed through the front panel connector to the BMC, which monitors and de-bounces it. The signal must be stable for at least 50 ms before a state change is recognized.

If secure mode is enabled or if the reset button is locked by the BMC, then the button does not reset the system. Instead a platform security violation attempt event message is generated if the reset button is pressed.

Soft Reset and Hard Reset
The BMC monitors an ICH6 signal called BIOS_POST_CMPLT_N, which deasserts at the beginning of POST and asserts at the end of POST. The signal deassertion indicates that a system reset has occurred. The BMC monitors this signal to detect both hard resets and soft resets. The BIOS converts

all INITs into hard resets. Therefore, INITs also result in the BMC sensing a system reset. The BMC detects these resets but does not participate in the reset mechanism.

BMC Command to Cause System Reset
Chassis Control is the primary command used to reset the system. The Set Processor State command, which is used by the BIOS during POST, can also cause a system reset.

Watchdog Timer Expiration
The watchdog timer can be configured to cause a system reset when the timer expires. See the Intelligent Platform Management Interface Specification, Version 2.0.

BMC Reset Control

BMC Exits Firmware Update Mode
The BMC firmware can be updated using firmware transfer commands through the LPC interface. The BMC enters firmware transfer mode if it detects that the Force Update signal is asserted during initialization or if the operation code checksum validation fails. When exiting firmware transfer mode, the BMC resets. The BMC re-synchronizes to the state of the processor and power control signals it finds when it initializes.

Standby Power Comes Up
The system has AC power applied, but the system is not up. The BMC resets the system when DC power output from the power supplies is available. The BMC re-synchronizes to the state of the processor and power control signals it finds when it initializes.

System Initialization
The following items are initialized by both the BIOS and the BMC during system initialization.

Processor TControl Setting
Processors used with this system implement a feature called Tcontrol, which provides a processor-specific value that can be used to adjust the fan control behavior to achieve optimum cooling and acoustics. The BMC cannot access these values. The BIOS reads the values during POST and communicates them to the BMC using the Set Processor Tcontrol command. The BMC uses these values as part of the fan speed control algorithm.

Fault Resilient Booting (FRB)
Fault resilient booting (FRB) is a set of BIOS and BMC algorithms and hardware support that allow a multiprocessor system to boot even if the bootstrap processor (BSP) fails. Only FRB2 is supported, using watchdog timer commands.

FRB2 refers to the FRB algorithm that detects system failures during the POST. The BIOS uses the BMC watchdog timer to back up its operation during POST. The BIOS configures the watchdog timer to indicate that the BIOS is using the timer for the FRB2 phase of the boot operation.

After the BIOS has identified and saved the BSP information, it sets the FRB2 timer use bit and loads the watchdog timer with the new timeout interval.

If the watchdog timer expires while the watchdog use bit is set to FRB2, the BMC (if so configured) logs a watchdog expiration event showing the FRB2 timeout in the event data bytes. The BMC then hard resets the system, assuming the BIOS selected reset as the watchdog timeout action.

The BIOS is responsible for disabling the FRB2 timeout before initiating the option ROM scan and before displaying a request for a boot password. If the processor fails and causes an FRB2 time-out, the BMC resets the system.

The BIOS gets the watchdog expiration status from the BMC. If the status shows an expired FRB2 timer, the BIOS enters the failure in the system event log (SEL). In the OEM bytes entry in the SEL, the last POST code generated during the previous boot attempt is written. FRB2 failure is not reflected in the processor status sensor value. The FRB2 failure does not affect the front panel LEDs.

Processor Presence and Population Check
When the BMC detects an empty processor socket, it sets the disable bit in the processor status sensor for that socket and clears the remaining status bits, including any persistent bits. The BMC checks for processor presence before the system is powered-on.

Processor Disabling
No processor error condition requires the BMC to disable a processor. There is no hardware support for the BMC to disable a processor.

BSP Identification
The BMC cannot indicate which processor is the BSP. Software that needs to identify the BSP should use the multiprocessor specification tables.

Integrated Front Panel User Interface
- The front panel has the following indicators:
- Power LED
- System status / fault LED
- Chassis ID LED

The front panel provides the following buttons:
- Reset button
- Power button
- System diagnostic interrupt button (NMI button)
- Chassis ID button

Power LED
The green power LED is active when system DC power is on. The power LED reflects a combination of the state of system (DC) power and the system ACPI state. The BIOS controls it.

Table 97. Power LED Indicator States

State	ACPI	Power LED
Power off	No	Off
Power on	No	Solid on
S4 / S5	Yes	Off
S1 Sleep	Yes	~1 Hz blink
S0	Yes	Solid on

System Status LED

Note: *The system status LED state shows the state for the current, most severe fault. Example: If there was a critical fault due to one source and a non-critical fault due to another source, the system status LED state would be the state for the critical fault.*

The system status / fault LED is a bicolor LED. Green (status) is used to show a normal operation state or a degraded operation. Amber (fault) shows the platform hardware state and over-rides the green status. The system status LED is mainly controlled by the BMC. Early in the startup boot process, the BIOS checks the chipset for any memory errors.

The BMC-detected states are included in the LED states. For fault states that are monitored by BMC sensors, the contribution to the LED state follows the associated sensor state, with priority given to the most critical asserted state.

When the server is powered down (transitions to the DC-off state or S5), the BMC is still on standby power and retains the sensor and front panel status LED state established before the power-down event. Note: System status LED will not update the status of the LED when the HSC events are generated.

Color	State	System Status	Description
Green	Solid on	Ok	System ready
Green	~1 Hz blink	Degraded	BIOS detected ▪ Unable to use all of the installed memory (more than one DIMM installed). [1] ▪ Correctable errors over a threshold of 10 and migrating to a spare DIMM (memory sparing). This indicates that the user no longer has spared DIMMs, indicating a redundancy lost condition. The corresponding DIMM LED lights. [1] ▪ In mirrored configuration, when memory mirroring takes place and system loses memory redundancy. [1] ▪ PCI Express* correctable link errors. BMC detected ▪ Redundancy loss such as power-supply or fan. Applies only if the associated platform sub-system has redundancy capabilities. ▪ CPU disabled ▪ Fan alarm / Fan failure. The number of operational fans should be more than minimum number needed to cool the system ▪ Non-critical threshold crossed – Temperature, voltage, power nozzle, power gauge, and PROCHOT1 (Therm Ctrl) sensors. ▪ Battery failure ▪ The system has lost redundancy due to predictive failure conditions on power supplies
Amber	~1 Hz blink	Non-fatal	Non-fatal alarm. The system is likely to fail. BIOS detected ▪ In non-sparing and non-mirroring mode if the threshold of 10 correctable errors is crossed within the window1 ▪ PCI Express* uncorrectable link errors. BMC detected ▪ Critical threshold crossed – Temperature, and PROCHOT (Therm Ctrl) sensors. ▪ VRD hot-asserted ▪ Minimum number of fans to cool the system not present or failed
Amber	Solid on	Fatal	Fatal alarm – system has failed or shutdown. BIOS detected ▪ DIMM failure when there is one DIMM present, no good memory present1 ▪ Run-time memory uncorrectable error in non-redundant mode1 ▪ CPU configuration error (for instance, processor stepping mismatch). BMC detected ▪ CPU IERR signal asserted ▪ No CPU present or CPU Configuration Errors ▪ CPU THERMTRIP ▪ No power good – power fault ▪ Power unit redundancy sensor – Insufficient resources offset. indicates not enough power supplies present.
Off	N/A	Not ready	AC power off

Note 1: *Support for upper non-critical is not provided in default SDR configuration. If a user enables this threshold in the SDR, then the system status LED behaves as described.*

Chassis ID LED

The chassis ID LED provides a visual indication of a system being serviced. The state of the chassis ID

LED is affected by three things:

- It is toggled by the chassis ID button
- It can be controlled by the Chassis Identify command (IPMI)
- It can be controlled by the Chassis Identify LED command (OEM)

Table 99. Chassis ID LED Indicator States

State	LED State
Identify active via button	Solid on
Identify active via command	~1 Hz blink
Off	Off

There is no precedence or lock-out mechanism for the control sources. When a new request arrives, previous requests are terminated. Example: If the chassis ID LED is blinking and the chassis ID button is pressed, then the chassis ID LED changes to solid on. If the button is pressed again, then the chassis ID LED turns off.

- Chassis ID led will maintain the state after system reset, power on/off but not on AC cycle.
- Chassis ID led will be on in Firmware transfer mode and cannot be changed.

Front Panel / Chassis Inputs

The BMC monitors the front panel buttons and other chassis signals. The front panel input buttons are momentary contact switches that are de-bounced by the BMC. The de-bounce time is 50 ms; the signal must be in a constant low state for 50 ms before it is treated as asserted. BMC debouncing does not affect the operation of the power or reset button, since the power and reset buttons are connected to the chipset. The debouncing is only for BMC monitoring.

Chassis Intrusion

Chassis intrusion detection is supported. The BMC monitors the state of the Chassis Intrusion signal and makes the status of the signal available via the Get Chassis Status command and the Physical Security sensor state. A chassis intrusion state change causes the BMC to generate a Physical Security sensor event message with a General Chassis Intrusion offset (00h).

The BMC boosts all fans when the chassis intrusion signal is active. Fans return to their previous level when the chassis intrusion signal is no longer active. This provides sufficient cooling during system servicing. The BMC monitors the chassis intrusion cable. If the cable is missing, the BMC logs a SEL event and sets the chassis intrusion status to the init-in-progress state.

The BMC detects chassis intrusion and logs a SEL event when the system is in the on, sleep, or standby state. Chassis intrusion is not detected when the system in an AC power-off state.

Reset Button

An assertion of the Front Panel Reset signal to the BMC causes the system to start the reset and reboot process, as long as the BMC has not locked-out this input. This assertion is immediate and without the cooperation of software or the operating system.

Diagnostic Interrupt (Front Panel NMI)

A diagnostic interrupt is a non-maskable interrupt or signal for generating diagnostic traces and core dumps from the operating system. The diagnostic interrupt button is connected to the BMC through a front panel connector. Once an NMI has been generated by the BMC, the BMC does not generate another until the system has been reset or powered down.

Chassis Identify
The front panel chassis identify button toggles the state of the chassis ID LED. If the LED is off, then pushing the button lights the LED. It remains lit until the button is pushed again or until a Chassis Identify or Chassis Identify LED command is received that changes the state of the LED.

Secure Mode and Front Panel Lock-out Operation
Secure mode protects the front panel buttons against unauthorized use. Secure mode is enabled and controlled by the Set Secure Mode Options command.

If a protected front panel button is pressed while the system is in secure mode, a secure mode violation event is generated. *Secure Mode Violation Attempt offset* (00h) of the BMC's Security Violation Attempt sensor is asserted.

Secure mode is cleared whenever AC power is applied, when a system reset occurs, or when a BMC reset occurs. The secure mode state includes the bits that specify which actions are to be taken when secure mode is active the Force Secure Mode On bit.

The Set Secure Mode Options command protects front panel buttons regardless of the secure mode state. This protection includes blocking the buttons and generating secure mode violation events if one of the buttons is pressed when in secure mode state. The front panel power and reset buttons must be protected as a unit. They cannot be individually locked.

The set of buttons that is protected when secure mode is active varies, depending on the system ACPI power state.

Table 100. Secure Mode vs. ACPI State

ACPI State	Power Button	Reset Button	Diagnostic Interrupt (Front Panel NMI) Button	ID Button
S0	Protected	Protected	Unprotected	Unprotected
S1	Unprotected	Unprotected	Unprotected	Unprotected
S5	Unprotected	Unprotected	Unprotected	Unprotected

Set Fault Indication Command
Satellite controllers and system management software use the Set Fault Indication command to communicate fan, temperature, power, and drive fault states to the BMC. The BMC consolidates the state with its own platform state when determining how to set the front panel indicator LED states and how to control other behavior, such as fan boosting.

The Set Fault Indication command has a source field that allows the BMC to track the fault states of multiple sources. Each source must use a separate unique source ID.

The fault state of each source is tracked independently. Whenever a source sets the fault state for a particular fault type, such as fan or power, the new state overrides the previous state. The tracked fault state is cleared when the server is powered up or reset.

Button Sensor
The BMC supports a button sensor that monitors the state of the front panel power and reset buttons. This sensor does not reflect button presses that are simulated by the BMC that the BMC may do when initiating a system power state change or a system reset.

Watchdog Timer
The BMC implements a fully IPMI 2.0-compatible watchdog timer. See the IPMI 2.0 specification. The NMI / diagnostic interrupt for an IPMI 2.0 watchdog timer is associated with an NMI. A watchdog pre-timeout SMI or equivalent signal assertion is not supported.

BMC Internal Timestamp Clock
The BMC maintains a four-byte internal timestamp clock that sub-systems, such as the SEL, use. The timestamp value is derived from an RTC element that is internal to the ESB2 BMC.

This internal timestamp clock is read and set using the Get SEL Time and Set SEL Time commands, respectively. The Get SDR Time command can also be used to read the timestamp

System Event Log (SEL)
The BMC implements the system event log as specified in the Intelligent Platform Management Interface Specification, Version 2.0. The SEL is accessible regardless of the system power state via the BMC's in-band and out-of-band interfaces.

The BMC allocates 65,536 bytes (64 KB) of non-volatile storage space to store system events. Each record is padded with a four-byte timestamp that indicates when the record was created, making each SEL record 20 bytes in size. The SEL timestamps might not be in order. Up to 3,276 SEL records can be stored at a time. An attempt to add records beyond the maximum results in a failure and the out-of-space completion code is returned.

Servicing Events
Events can be received while the SEL is being cleared. The BMC implements an event message queue to avoid the loss of messages. Up to three messages can be queued before messages are overwritten.

The BMC recognizes duplicate event messages by comparing sequence numbers and the message source. See the IPMI 2.0 specification. Duplicate event messages are discarded (filtered) by the BMC after they are read from the event message queue. This means the queue can contain duplicate messages.

SEL Entry Deletion
The Delete SEL command does not reclaim deleted SEL record space, but marks the records so they are not seen by the IPMI Get SEL Entry command. The BMC implements the Get All SEL Entry extension command that lists all SEL records, including those that have been deleted by the Delete SEL command. SEL erasure is the only way to reclaim SEL record space.

The Get All SEL Entry is an OEM-command and assumes a particular SEL implementation. If a software application that accesses the SEL is required to be IPMI compatible (it is expected to work with any IPMI-compliant platform), then this command should not be used.

SEL Erasure
SEL erasure is a background process. After initiating erasure with the Clear SEL command, additional Clear SEL commands must be executed to get the erasure status and determine when the SEL erasure is completed. This may take several seconds. SEL events that arrive during the erasure process are queued until the erasure is complete and then committed to the SEL.
SEL erasure generates an Event Logging Disabled (Log Area Reset / Cleared offset) sensor event. The BMC generates Event Logging Disable to indicate the SEL area has been erased.

Sensor Data Record (SDR) Repository
The BMC implements the sensor data record (SDR) repository as specified in the Intelligent Platform Management Interface Specification, Version 2.0. The SDR is accessible through the BMC's in-band and out-of-band interfaces regardless of the system power state The BMC allocates 65,536 bytes (64 KB) of non-volatile storage space for the SDR. See Table 38 for SDR command support.

SDR Repository Erasure
SDR repository erasure is a background process. After initiating erasure with the Clear SDR Repository command, additional Clear SDR Repository commands must be executed to get erasure status and determine when the SDR repository erasure is done. This may take several seconds. The SDR repository cannot be accessed or modified until the erasure is done.

Field Replaceable Unit (FRU) Inventory Device
The BMC implements the interface for logical FRU inventory devices as specified in the Intelligent Platform Management Interface Specification, Version 2.0. This functionality provides commands used for accessing and managing the FRU inventory information. These commands can be delivered via all interfaces.

The BMC provides only low-level access to the FRU inventory area storage. It does not validate or interpret the data that are written. This includes the common header area. Applications cannot relocate or resize any FRU inventory areas.

Note: *Fields in the internal use area are not for OEM use. Intel reserves the right to relocate and redefine these fields without prior notification. Definition of this area is part of the software design. The format in the internal use area may vary with different BMC firmware revisions.*

Diagnostics and Beep Code Generation
The BMC may generate beep codes upon detection of failure conditions. Beep codes are sounded each time the problem is discovered (Example: On each power-up attempt), but are not sounded continuously. Supported codes are in Table 101. Each digit in the code is represented by a sequence of beeps whose count is equal to the digit.

Table 101. BMC Beep Codes

Code	Reason for Beep	Associated Sensors	Supported
1-5-2-1	CPU: Empty slot / population error.	CPU missing error	Yes
1-5-4-2	Power fault: DC power unexpectedly lost (power good dropout)	Power unit – power unit failure offset	Yes
1-5-4-4	Power control fault (Power good assertion timeout)	Power unit – soft power control failure offset	Yes

NMI

On IA-32 platforms, the BMC has monitoring and signal generation functionality in regards to the NMI (non-maskable interrupt) signal. When the BMC generates a diagnostic interrupt, the NMI signal is pulsed. A front panel diagnostic interrupt sensor is used to log SEL events for assertion of the diagnostic interrupt.

Signal Generation

The BMC generates an NMI pulse under certain conditions. The BMC-generated NMI pulse duration is at least 30 ms. Once an NMI has been generated by the BMC, the BMC does not generate another until the system has been reset or powered down, unless an NMI Enable / Disable command is used to re-arm the NMI.

The BMC captures the NMI source(s) and makes that information available via a Get NMI Source command. Reading the NMI source information clears the information. The Set NMI Source command is available to other agents, such as the BIOS SMI handler, to register NMI sources when they detect NMI-generating errors. Operating system NMI handlers that save the system crash state can use the Get NMI Source command to determine and save the cause of the NMI.

The NMI Enable / Disable command disables BMC NMI generation. The default state is enabled. The enabled / disabled state is volatile; it is not saved across AC power cycles.

The following cause the BMC to generate an NMI pulse:

- Receiving a Chassis Control command to pulse the diagnostic interrupt. This command does not cause an event to be logged in the SEL.
- Detecting that the front panel diagnostic interrupt button has been pressed.
- A PEF table entry matching an event where the filter entry has the diagnostic interrupt action indicated.
- Watchdog timer pre-timeout expiration with NMI / diagnostic interrupt pre-timeout action enabled.
- A Set NMI Source command issued from a command interface.

The following table shows behavior regarding NMI signal generation and event logging by the BMC.

Table 102. NMI Signal Generation and Event Logging

Causal Event	NMI (IA-32 Only)	
	Signal Generation	Front Panel Diag Interrupt Sensor Event Logging Support
Chassis Control command (pulse diagnostic interrupt)	X	–
Front panel diagnostic interrupt button pressed	X	X
PEF matching event with diagnostic interrupt action selected [1]	X	–
Watchdog Timer pre-timeout expiration with NMI / diagnostic interrupt action	X	–
Set NMI Source command	X	–

General Sensor Behavior

Sensor Initialization

As part of the BMC initialization upon application of standby power or BMC reset, the BMC enables a subset of its sensors. This is done before loading any SDRs. This allows some amount of sensor functionality even if there are no SDRs present. Subsequent loading of SDRs may change the configuration of these sensors.

The sensors that are enabled independently of an SDR load include:

- Processor status sensors
- DIMM sensors
- PCI Express* link sensors

Processor Sensors

The BMC provides IPMI sensors for processors and associated components, such as voltage regulators and fans. The sensors are implemented on a per-processor basis.

Table 103. Processor Sensors

Sensor Name	Per-Proc Socket	Description
Processor Status	Yes	Processor presence and fault state
Digital Thermal Sensor	Yes	Relative temperature reading via PECI.
Processor VRD Over-temperature Indication	Yes	Discrete sensor that indicates a processor VRD has crossed an upper operating temperature threshold
Processor Voltage	Yes	Discrete sensor that indicates a processor power good states.
Processor Thermal Control (Prochot)	Yes	Percentage of time a processor is throttling due to thermal conditions

A processor's Digital Thermal Sensor and Processor Thermal Control sensors will be disabled whenever that processor is not physically present when the BMC's SDR initialization agent executes.

Processor Status Sensors

The BMC provides an IPMI sensor of type processor for monitoring status information for each processor slot. Except for the processor presence offset, if an event state (sensor offset) has been asserted, it remains asserted until one of the following happens:

- A Rearm Sensor Events command is executed for the processor status sensor.
- A Processor Retest command is executed. The BIOS sends this command to the BMC if a user chose Processor Retest from the BIOS Setup utility.
- AC power cycle occurs. This only clears persistent bits of the sensor if the processor is not present.
- DC power-on and system resets do not re-arm processor status sensors.

Processor Presence
When the BMC detects an empty processor socket, it sets the disable bit in the processor status for that socket and clears the remaining status bits, including any persistent bits. Upon BMC initialization, the processor presence offset is initialized to the de-asserted state. The BMC then checks to see if the processor is present, setting the offset accordingly. This state is updated at each DC power-on and at system resets. The net effect is that there should be one event logged for processor presence at BMC initialization for each installed processor, assuming the SDR is configured to generate the event. No additional events for processor presence are expected unless the sensor is manually re-armed using an IPMI command.

Thermtrip Monitoring
The BMC retains ThermTrip history for each processor. This history tracks whether the processor has had a ThermTrip since the last processor sensor re-arm or retest. When a ThermTrip occurs, the BMC polls the ThermTrip status for each processor and then the system begins the power down sequence. If the BMC detects that a ThermTrip occurred, then it sets the ThermTrip offset for the applicable processor status sensor.
IERR Monitoring
The BMC monitors the internal error (IERR) signal from each processor and maps it to the IERR offset of the associated processor status sensor.

Processor VRD Over-Temperature Sensor
- This sensor monitors a signal that indicates whether a processor VRD is running over temperature. The state of this signal is not an input into the system fan control subsystem, but it is an input into the LM94* devices, which asserts the associated Prochot signal and lowers the VRD temperature. This relationship is 1:1: if VRD-hot is asserted, then Prochot asserts.
- The Prochot assertion will affect the reading of the associated Prochot sensor and may result in additional SEL events being logged.

Digital Thermal Sensor
The Quad-Core Intel® Xeon® Processors 7300 Series or Dual-Core Intel Xeon Processors 7200 Series supports a digital thermal sensor that provides a relative temperature reading that is defined as the number of degrees below the processor's thermal throttling trip point, also called the PROCHOT threshold. When a processor reaches this temperature, the processor's PROCHOT signal asserts, indicating that one or more of the processor's built-in thermal control circuits (TCC) has activated to limit further increases in temperature by throttling the processor.

The digital thermal sensor reading value is always less than or equal to zero. A reading of zero indicates that the PROCHOT threshold has been reached. The reading remains at zero until the temperature goes back below the PROCHOT threshold.

The digital thermal sensors are located on the processor platform environment control interface (PECI) bus. The BMC does not access this bus directly, but communicates over a SMBus with a PECI-poller device that polls the digital thermal sensors over the PECI bus.

The IPMI sensor names associated with these devices are P1 Therm Margin, P2 Therm Margin, P3 Therm Margin and P4 Therm Margin. The default SDR configuration is to have no thresholds programmed or event generation enabled because the sensor is expected to reach its maximum value of zero during normal operation.

PECI Interface
The platform environment control interface (PECI) is a one-wire, self-clocked bus interface that provides a communication channel between Intel processors and chipset components to an external monitoring device. The PECI bus communicates environment information, such as the temperature data, between the managed components, referred to as the PECI client devices, and the management controller, referred to as the PECI system host. The PECI standard supersedes older methods, such as the thermal diode, for gathering thermal data.

The PECI interface consists of a microcontroller with PECI drive circuitry. The BMC monitors the processor temperature by reading the PECI controller over one of the ESB2's private I2C / SMBus.

Processor Thermal Control Monitoring (Prochot)
The BMC monitors processor thermal control monitoring for each processor. The LM94* provides this functionality by reading the percentage of time that the processor ProcHot signal is asserted over a given measurement window (set to 5.8 seconds).

The BMC implements this as a threshold sensor (IPMI sensor type = processor, sensor name = Therm Margin) on a per-processor basis. This sensor supports one threshold, the upper-critical, and it is set for 50% by default in the SDRs. The IPMI sensor names associated with these devices are P1-P4 Thermal ctrl %.
When the processors are throttled by the Power Safe feature, the Prochot sensors will show reading/state unavailable status to prevent spurious Prochot-related SEL events.

CPU Missing Sensor
The BMC verifies at least one processor is installed at start-up. The hardware does not allow the server to power up if no processor is installed. At BMC initialization, the CPU missing sensor is first set to a de-asserted state. The BMC then checks for a missing processor and sets the new value accordingly. If an error is detected and the SDR is so configured, a SEL event is logged. The BMC checks for this fault condition and updates the sensor state at each attempt to DC power-on the system. At each DC power-on attempt, a beep code is generated if this fault is detected. Beep codes are listed in Table 101.
The CPU missing sensor is an auto-re-arm sensor, but it is not re-armed at system DC poweron or for system resets. To clear the sensor:

1. AC power down the server.
2. Install a processor into any socket.
3. AC power on the server.

Standard Fan Management
The BMC controls and monitors the system fans. Each fan is associated with a fan speed sensor that detects fan failure and may also be associated with a fan presence sensor for hotswap support. For redundant fan configurations, the fan failure and presence status determines the fan redundancy sensor state.

The system fans are divided into fan domains, each of which has a separate fan speed control signal and a separate configurable fan control policy. A fan domain can have a set of temperature and fan sensors associated with it. These are used to determine the current fan domain state. A fan domain has three states: sleep, nominal, and boost. The sleep and boost states have fixed (but configurable via OEM SDRs) fan speeds associated with them. The nominal state has a variable speed determined by the fan domain policy.

An OEM SDR record is used to configure the fan domain policy. The Set SM Signal command can be used to manually force the fan domain speed to a selected value, overriding any other control or policy. The fan domain state is controlled by several factors. In order of precedence, high to low:

- Boost
  - Associated fan in a critical state or missing.
  - Any associated temperature sensor in a critical or non-recoverable state.
  - Chassis cover missing.
  - If any of the above conditions apply, the fans are set to a fixed boost state speed, specified in the Tcontrol OEM SDRs.
- Sleep
  - No boost conditions, system in ACPI S1 sleep state, and BMC configured to transition fan domains to sleep state via the Set ACPI Configuration Mode command. In this situation, fans are set to a fixed sleep state speed, specified in the Tcontrol OEM SDRs. The BMC can support normal fan speed control in the S1 sleep state, so the BIOS does not enable APCI fan control.
- Nominal

Hot Swap Fans
Hot-swap fans are supported. These fans can be removed and replaced while the system is powered on and operating. The BMC implements fan presence sensors (sensor type = Fan (04h), event / reading type = Sensor Specific (6Fh)) for each hot swappable fan.

When a fan is not present, the associated fan speed sensor is put into the reading/state unavailable state, and any associated fan domains are put into the boost state. The fans may already be boosted due to a previous fan failure or fan removal.

When a removed fan is inserted, the associated fan speed sensor is rearmed. If there are no other critical conditions causing a fan boost condition, the fan speed returns to the nominal state. Power-cycling or resetting the system rearms the fan speed sensors and clears fan failure conditions. If the failure condition is still present, the boost state returns once the sensor has reinitialized and the threshold violation is detected again.

Sleep State Fan Control

Using the Set ACPI Configuration Mode command, the BMC may be configured to set the fans to a fixed sleep state speed when the system is in the S1 sleep state.

Fan Redundancy Detection
The BMC supports redundant fan monitoring and implements a fan redundancy sensor. A fan redundancy sensor generates events when its associated set of fans transitions between redundant and non-redundant states, as determined by the number and health of the fans. The definition of fan redundancy is configuration dependent. The BMC allows redundancy to be configured on a per fan-redundancy sensor basis via OEM SDR records.

A fan failure, or removal of hot-swap fans up to the number of redundant fans specified in the SDR, in a fan configuration is a non-critical failure and is reflected in the front panel status as such. A fan failure or removal that exceeds the number of redundant fans is a non-fatal insufficient resources condition and is reflected in the front panel status as a non-fatal error.

Fan Domains
System fan speeds are controlled through pulse width modulation (PWM) signals, which are driven separately for each domain by auxiliary system management devices. Fan speed is changed by adjusting the duty-cycle, which is the percentage of time the signal is driven high In each pulse.

Nominal Fan Speed
A fan domain’s nominal fan speed can be configured as static (fixed value) or controlled by the state of one or more associated temperature sensors.

OEM SDR records are used to configure which temperature sensors are associated with which fan control domains and the algorithmic relationship between the temperature and fan speed. Multiple OEM SDRs can reference / control the same fan control domain, and multiple OEM SDRs can reference the same temperature sensors.

Acoustic Management (Acoustic Monitoring)
This feature refers to enhanced fan management to keep the system optimally cooled while reducing the amount of noise generated by the system fans. Aggressive acoustics standards might require a trade-off between fan speed and system performance parameters that contribute to the cooling requirements, primarily memory bandwidth. The BIOS, the BMC, and the SDRs work together to provide control over how this trade-off is determined.

Interactions with DIMM Thermal Management

Thermal Profile Data
The BIOS requires knowledge of characteristics to use as input into its calculations for DIMM throttling setup. This depends on which fan profile is enabled. The BIOS retrieves the information from the BMC at system boot.

The BMC supports this with thermal profile data SDRs, which allow this data to be stored in the BMC’s SDR repository and can be customized by fan domain. Each thermal profile SDR can apply to one or more profiles in a fan domain.

The BMC requires these SDRs to follow a standard structure and provides a mechanism to retrieve the data, but the BMC does not interpret or use the thermal profile data information in any manner.

The BMC expects that only one set of thermal profile data SDRs, such as for a board / chassis combination, is present in the SDR repository at one time, and that each profile is associated with only one SDR per supported throttling type. If multiple SDRs match one throttling type and profile, only the data from one of these SDRs is retrievable. The BMC does not guarantee the order in which the SDRs are processed.

AMB Aggregate Margin Temperature Sensors
The BMC implements one advanced memory buffer (AMB) aggregate margin temperature sensor for each rear fan domain.

ASHRAE Compliance
System requirements for ASHRAE compliance is defined in the ProServ 4680 Server System Fan Speed Control & Thermal Management Platform Architecture Specification. Altitude-related considerations are handled through the chipset throttling configuration.

Platform Configuration

Overview
The base fan speeds for each domain are determined by stepwise linear controls linked to the Front Panel ambient temperature sensor (32h). Each domain also has a stepwise linear domain max control based on that sensor.

Model 0 Considerations
The platform is required to support a “Model 0” fan configuration in which memory fans 2 and 4 are not installed. The Thermal Profile Data and Tcontrol SDR parameters are determined by the system thermals engineer to maintain sufficient but non-redundant cooling.

The determination of which set of parameters to use is made at the time of SDR installation. Sensor 6Dh exposes an aggregation of all of the fan module presence signals, and the FRUSDR utility’s DISCRETE_SENSOR probing feature is used to compare the sensor reading against various expected values:

- If all fans are present, the BMC installs all fan monitoring SDRs, the full configuration fan redundancy sensor and map SDRs, and the fan control SDRs associated with the full fan configuration. These are designated by the “FAN_FULL” tag in the SDR file.
- If either memory fan 2 or memory fan 4 is missing, the BMC installs all fan monitoring SDRs except the ones for those two fans and the fan control SDRs associated with the Model 0 fan configuration. Additionally, a special fan redundancy map SDR for sensor 6Eh is installed to ensure proper system fault LED behavior related to fan status. These SDRs are all designated by the “FAN_MODEL0” tag in the SDR file.
- If any other fan is missing, the BMC does not install any fan monitoring or control SDRs.

This is considered an unknown and unsupported configuration.

22.18 **Power Supply Management Interface (PSMI)**

The BMC supports a minimum set of PSMI 2.13 functionality. The BMC obtains the following power supply status information from the PSMI device.

- AC Lost Status
- Power Supply Temperature
- AC range
- Power Supply Failure Status

Power Unit Management

Power Off

The BMC asserts the Power Off offset whenever the system DC power is off.

Power Cycle

The Power Cycle offset is asserted when the system is DC power-cycled. This offset is for event generation only and does not remain asserted.

AC Lost

The BMC asserts the AC lost offset when AC power is applied to the system and the previous system power state was on. This offset is for event generation only and does not remain asserted.

Soft Power Control Fault

The BMC asserts the Soft Power Control Failure offset if the system fails to power-on within 2 seconds due to the following power control sources:

- Chassis control command
- PEF action
- BMC watchdog timer
- Power state retention
- The BMC provides system status indication via the front panel LEDs
- The BMC generates a beep code for Power Control Fault. See Table 101.

Power Unit Failure

The BMC asserts the Power Unit Failure offset of the Power Unit sensor for the following situations:

- Power-good dropout (see section 22.1.2).
- The system fails to power down: The POWER_GOOD signal fails to transition to the deasserted state within 2 second when any of the enabled power control sources attempt to transition the system to the power-off state.
- The system fails to power-on due to any enabled hardware power control source: The POWER_GOOD signal from the power sub-system fails to assert within 2 seconds in response to a chipset or front panel power button request to power on.
- The BMC provides system status indication via the front panel LEDs.
- The BMC generates a beep code for a power fault. See Table 101.

Power distribution board (PDB) failure is detected. This is supported only for power supply

configurations that have a PDB with a PSMI device.

Power Supply Status Sensors
The BMC processes the AC lost indication to avoid logging an AC lost event against multiple supplies when all supplies are unplugged simultaneously or AC power to the server is shut off. In these cases, the BMC only logs an AC lost event against the power unit sensor when AC is next applied to the system.

The PS1 status sensor SDR is always installed during the normal SDR installation process. The PS2 status sensor is only installed if that power supply is present at the time of SDR installation.

Power Unit Redundancy
The BMC supports redundant power sub-systems and implements a Power Unit Redundancy sensor per platform. A Power Unit Redundancy sensor is of sensor type Power Unit (09h) and reading type Availability Status (0Bh). This sensor generates events when a power sub-system transitions between redundant and non-redundant states, as determined by the number and health of the power sub-system's component power supplies. The BMC allows redundancy to be configured on a per power-unit-redundancy sensor basis via the OEM SDR records.

The power unit redundancy sensor SDR is only installed if both PS1 and PS2 are present at the time of SDR installation.

Power Fault Analysis
A single power good signal from each power unit components are monitored by the BMC. The BMC supports individual discrete sensors for the VR/D2D status, for the power fault analysis feature. The BMC monitors the power good signal of each power unit component via the Server Management Diagnostics bus. The BMC generates a SEL event and system beep codes, as shown in Table 101, if the VR/D2D fails due to power good de-assertion. The BMC also provides the failure indications to the front panel system status LED to indicate a critical fault.

Power Safe
The BMC provides the Power Safe feature support for the server system. The Power Safe feature prevents the server from shutting down, when one of the power supplies fails or is removed unexpectedly while the power utilization level exceeds the power capacity of a single power supply. The power utilization is monitored by the PLD, which provides the power utilization indication to the BMC via the PS_NON_REDUNANCY signal. The PS_NON_REDUNDANCY signal is asserted when the power supply utilization has reached or exceeded a safe point and CPU throttling may be required. The Power Safe feature is ONLY enabled if the input voltage range is less than 200 VAC. The Power Safe consists of three operational states, Redundant, Non-Redundant, and Throttled.
Redundant State: The power sub-system remains in this state when ALL of the following conditions applied.

- Two operational power supplies presence in the system.
- The PS_NON_REDUNDANCY signal is deasserted. This condition only applicable when the input voltage is less than 200VAC.

Non-Redundant State: The BMC will log a non-redundant SEL event for the Power Unit Redundancy sensor. The power sub-system will enter this state if ONE of the following conditions occurs.

- Only one power supply is operational.
- The PS_NON_REDUNDANCY signal is asserted. The BMC will turn on the PS_UTIL_LED on the main board for light guided diagnostic purposes. Note: This condition is ONLY applicable when the input voltage range is less than 200VAC.

Throttled State: The BMC will throttle all the CPUs to reduce the power utilization in the system, since the power utilization has exceeded the power capacity of a single power supply. The BMC will log a CPU throttling SEL event. The power sub-system will enter this state if ALL of the following conditions applied:

- The input voltage range is less than 200 VAC.
- Only one power supply is operational.
- The PS_NON_REDUNDANCY signal is asserted.

Note: The BMC gets the AC input range from the PSMI device, which is embedded in the power supply. Therefore, the BMC will disable the Power Safe feature if it detects problems with communicating to the PSMI device.

System Memory RAS and Bus Error Monitoring

System memory and bus error monitoring is done by the system BIOS. Early in the startup boot process, the BIOS checks the chipset for any memory errors. The BIOS updates the status of RAS configuration at startup and later at run time. BMC monitors and logs SEL events based on the SDR definitions. In addition, the BIOS help the BMC maintain the current DIMM presence and failure state and current memory RAS configuration, such as memory sparing and mirroring.

Support is provided for monitoring errors on system buses, such as front side bus (FSB) errors and PCI bus errors. The BIOS monitors these and generates critical interrupt sensor SEL events when an error is detected.

BMC Self Test

The BMC performs tests as part of its initialization. If a failure is determined, such as a corrupt BMC SDR, then the BMC stores the error internally. BMC or BMC sub-system failures detected during regular BMC operation may also be stored internally. Two commands may be used to retrieve the detected errors. The IPMI 2.0 Get Self Test Results command can be used to return the first error detected. The Read Self Test command can be used to sequentially read all the accumulated self test errors.

Field Replaceable Unit (FRU) / Fault LED Control

Several sets of FRU / POST / fault LEDs are supported. Some LEDs are owned by the BMC and some by the BIOS.

The BMC owns control of the following FRU / fault LEDs:

- Fan fault LEDs – A fan fault LED is associated with each fan. The BMC lights a fan fault LED if the associated fan tach sensor has a lower critical threshold event status asserted. Fan tach sensors are manual rearm sensors, therefore once the lower critical threshold is crossed, the LED remains lit until the sensor is rearmed. These sensors are rearmed at system DC power-on and system reset.
- CPU fault LEDs – A CPU fault LED is associated with each processor slot. The BMC lights a

CPU fault LED when the associated processor status sensor has either the configuration error or processor disabled offset asserted. Processor status sensors are manual rearm sensors, so if either of these offsets is asserted, the LED remains lit until the sensor is rearmed. These sensors are not rearmed at system DC power-on or system reset.

Hot-Swap Controller

Backplane Types
SAS / SATA backplanes are supported in the following configurations.

- Modular hot-swap controller (HSC) using Vitesse* 410: This configuration uses a modular board that plugs into SAS / SATA backplanes.

The Vitesse SEPs support the legacy BMC to SEP commands that were implemented on earlier server boards that used a Qlogic* GEM 424. These are supported via the IPMB interface. These commands are augmented with new commands capable of supporting up to 32 drives.

Intel® Remote Management Module 2 (Intel® RMM2)
This module is plugged into the Intel® RMM2 connector. This connector complies with a new management connector specification designed to last across multiple generations of servers. The Intel® RMM2 is an optional card that can be plugged into the server. The Intel® RMM2 and the Intel RMM2 NIC must both be installed to obtain the features provided by the Intel RMM2.

The Intel RMM2 and Intel RMM2 NIC are not hot pluggable, the server's AC power must be off when an Intel RMM2 and Intel RMM2 NIC are installed. The Intel® RMM2 contains it own stand alone firmware. This stand alone Intel RMM2 firmware can be updated using a Windows* or Linux* utility or logging into the Intel RMM2 card's WebServer.

The Intel® RMM2 provides keyboard, video, mouse (KVM) redirection capability, and other advanced functionality as follows.

- KVM redirection via the RMM2 dedicated NIC:
- USB – Media Redirection
- Embedded Web Server
- Update & Remote / Local Configuration Utility – Linux and Windows
- OEM Customization

Memory Region Temperature Monitoring
Temperature information from the memory region is a vital component of overall system cooling and acoustics management.

DIMM Temperature Monitoring
The FBDIMMS have an Advanced Memory Buffer (AMB) component on each module. The AMB chip has the ability to measure and report its own temperature. The DIMM module has been characterized so that the AMB temperature is an indication of the DRAM temperature. Due to the large number of DIMMs, the BMC does not maintain an IPMI temperature sensor for each DIMM slot. Instead, there is an IPMI temperature sensor for each AMB aggregate margin value calculated by the BMC. These temperature sensors are used for reporting temperature information, logging SEL entries, and as input

to the system fan management. If a DIMM is present and enabled and the system is powered-on, the BMC periodically accesses the AMB via the SMBus interface to the MCH (North Bridge) to retrieve the current temperature value as input to the aggregate margin IPMI sensor value calculations.

The BIOS provides the DIMM presence information to the BMC using the Set DIMM State command at each system boot. After each power-on or system reset, the BMC must wait to scan the AMB temperature for a slot until after the BIOS sends the Set DIMM State command to set the DIMM state as present and enabled for that slot.

Each AMB aggregate margin temperature sensor is shown as init-in-progress until the system is powered on and the BMC has successfully scanned the temperatures of all of AMB devices associated with the aggregate margin sensor. The BMC disables an AMB aggregate margin sensor if the DIMM state information indicates that no associated DIMMs are present and enabled.

Memory Riser Board Temperature Monitoring
Each memory riser board has locations for two physical temperature sensors, one at each end of the board. This section describes the support to be provided by the BMC, which can be enabled via SDRs if the physical sensor devices are populated.The BMC implements one IPMI sensor for each memory riser board representing the minimum of the readings from the two physical sensors on the board. The preliminary thermal assessment indicates that this sensor is required as an input to standard fan management regardless of the availability and accuracy of the AMB sensors described above.

The BMC disables a memory riser board's temperature sensors whenever that riser board is not physically present when the BMC's SDR initialization agent executes.

LAN Leash Event Monitoring
The Physical Security sensor is used for monitoring LAN link status and chassis intrusion status. This is implemented as a "LAN Leash" offset in this discrete sensor. This sensor monitors the link state of the two ESB2 embedded LAN channels. It does not monitor the state of the optional RMM2 dedicated NIC.

The "LAN leash lost" offset asserts when one of the two ESB2 LAN channels loses a previously established link. It deasserts when at least one LAN channel has a new link established after the previous assertion.

SMTP Alerting
SMTP alerting is implemented as an OEM alert type (OEM1) for LAN channels. The SMTP is only supported over the ESB2 embedded LAN channels. This alert type is Unacknowledged only. Each LAN Alert Destination can be configured as an SMTP alert. The LAN Alert Destination configuration determines the IP address of the SMTP server used to deliver the alert and the default gateway to use. All SMTP Alerts are sent to TCP port 25 on the destination machine. A separate LAN Channel OEM parameter defines the SMTP alert configuration associated with a LAN Alert Destination. Maximum of four SMTP configurations are supported.
SMTP alert configurations are accessed / defined via the Get/Set SMTP Alert Configuration Parameter commands. The supported parameters include:

- Number of supported alert configurations
- Email From: name (per configuration)

- Email To: name (per configuration)
- Email Subject: line (per configuration)
- Alerting machine name (identifies the managed server to the SMTP server, it is used with the SMTP HELO command)

Support for the feature for a LAN channel can be determined by the LAN Channel OEM Feature Support parameter.
The Message Content is a human readable version of the SEL event that triggered the alert.

23. **BMC Messaging Interfaces**

- This chapter describes the supported BMC communication interfaces:
- Host SMS Interface via low pin count (LPC) / keyboard controller style (KCS) interface
- Host SMM interface via low pin count (LPC) / keyboard controller style (KCS) interface
- Intelligent Platform Management Bus (IPMB) I2C interface
- Emergency management port (EMP) using the IPMI-over-serial protocols for serial remote access
- LAN interface using the IPMI-over-LAN protocols

Server Management Software (SMS) Interface
The SMS interface is the BMC host interface. The BMC implements the SMS KCS interface as described in the IPMI 2.0 specification. The BMC implements the optional Get Status / Abort transaction on this interface. Only logical unit number (LUN) 0 is supported on this interface. If so configured via the Set BMC Global Enables command, the BMC can generate an interrupt requesting attention when setting the SMS_ATN bit in the status register.

The SMS_ATN bit being set indicates one or more of the following:

- At least one message is in the BMC receive message queue
- An event is in the event message buffer
- Watchdog pre-timeout interrupt flag set

All conditions must be cleared and all BMC to SMS messages must be flushed for the SMS_ATN bit to be cleared.

The host I/O address of the SMS interface is nominally 0CA2h – 0CA3h, but this address assignment may be overridden. See the platform-specific information in the appendix.

The operation of the SMS interface is described in the Intelligent Platform Management Interface Specification.

SMM Interface
The SMM interface is a KCS interface that is used by the BIOS when interface response time is a concern, such as with the BIOS SMI handler. The BMC gives this interface priority over other communication interfaces. The BMC has limits on how many back-to-back transactions it can handle without loss in responsiveness. It must be able to handle up to 30 back-to-back commands from the BIOS.
The BMC implements the optional Get Status / Abort transaction on this interface. Only LUN 1 is

supported on this interface.
The event message buffer is shared across SMS and SMM interfaces.

The host I/O address of the SMM interface is nominally 0CA4h – 0CA5h, but this address assignment may be overridden by the platform-specific section of the appendix.

IPMB Communication Interface
The IPMB communication protocol uses the 100 KB/s version of an I2C bus as its physical medium. For more information on I2C specifications, see The I2C Bus and How to Use It. The IPMB implementation in the BMC is compliant with the IPMB v1.0, revision 1.0. The BMC IPMB slave address is 20h.

The BMC both sends and receives IPMB messages over the IPMB interface. Non-IPMB messages received via the IPMB interface are discarded.

Messages sent by the BMC can be either originated by the BMC, such as when due to initialization agent operation, or on behalf of another source, such as due to a Send Message command with IPMB channel number issued by SMS.

For IPMB request messages originated by the BMC, the BMC implements a response timeout interval of 60 ms and a retry count of 3.

IPMI Serial Feature
The IPMI 2.0 Intel implementation of IPMI-over-serial was known before IPMI 1.0 as the emergency management port (EMP) interface. The EMP nomenclature is no longer used. The primary goal of providing an out-of-band RS232 connection is to give system administrators the ability to access low-level server management firmware functions by using commonly available tools. To make it easy to use and to provide high-compatibility with LAN and IPMB protocols, this protocol design adopts some features of both the LAN and IPMB protocols. The Intel implementation shares EMP function with the platform's COM2 interface. The BMC has control over which agent (BMC or System) has access to COM2. Hardware handshaking is supported as are the Ring Indicate and Data Carrier Detect signals. See the IPMI 2.0 specification.

COM Port Switching
The SIO3 (formerly National Semiconductor* 87427) is used for Com port sharing. It has two legacy UARTS and a MUX switching arrangement that permits the BMC to monitor and intercept the serial traffic on serial port 2. If IPMI-over-serial is enabled, then the BMC watches the serial traffic when serial port 2, the COM2 port, is owned by the system. This is done to respond to in-band port switching requests.

Terminal Mode
The BMC supports terminal mode, as specified in the IPMI 2.0 specification. Terminal mode provides a printable ASCII text-based way to deliver IPMI messages to the BMC over the serial channel or any packet-based interface. Messages can be delivered in two forms:

- Via hex-ASCII pair encoded IPMI commands
- Via text SYS commands

The terminal mode interface supports a maximum IPMI message length of 40 bytes. The line continuation character is supported over the serial channel in terminal mode only. The line continuation character is supported for both hex-ASCII and text commands.

Input Restrictions

Maximum Input Length
The BMC supports up to 122 characters per line. The BMC stops accepting new characters and stops echoing input when the 122-character limit is reached. However, the <ESC>, <backspace> / <delete>, illegal, and input <newline> characters continue to be accepted and handled after the character limit is reached.

Invalid Passwords
If three successive invalid Activate Session commands are received on the EMP interface, the BMC delays 30 seconds before accepting another Activate Session command. The BMC logs an out-of-band access password violation event to the system event log each time an invalid Activate Session command is received.

LAN Interface
The BMC implements both the IPMI 1.5 and IPMI 2.0 messaging models. These provide out-ofband local area network (LAN) communication between the BMC and the external world.

The BMC supports a maximum of three LAN interfaces:

- Two of the LAN interfaces utilize the embedded ESB2 NICs (one channel per embedded NIC). There is no server management traffic supported over the IO riser LAN interfaces.
- One LAN interfaces utilizes an optional external NIC known as the GCM3. Use of this NIC requires the presence of the optional Intel® Remote Management Module add-in card.

See the IPMI 2.0 Specification for details about the IPMI-over-LAN protocol.
Run-time determination of LAN channel capabilities can be determined both by standard IPMI defined mechanisms and by an OEM configuration parameter that defines advanced feature support.

IPMI 1.5 Messaging
The communication protocol packet format consists of IPMI requests and responses encapsulated in an IPMI session wrapper for authentication, and wrapped in an RMCP packet, which is wrapped in an IP/UDP packet. Although authentication is provided, no encryption is provided, so administrating some settings, such as user passwords, through this interface is not advised.

Session establishment commands are IPMI commands that do not require authentication or an associated session.
The BMC supports the following authentication types over the LAN interface.

- None (no authentication)
- Straight password / key
- MD5

ESB2 Embedded LAN Channels

Even though the ESB2 embedded NICs are shared by the BMC and the server, sharing only means that both the BMC and the server use the same NIC. These shared NICs provide a dedicated MAC address solely for BMC use. As a result, in some ways these channels are more similar to a dedicated LAN channel than a shared channel. The IP address for the server is always different from the BMC IP address for an embedded NIC.

For these channels, support can be enabled for IPMI-over-LAN, ARP, and DHCP.

As an integral part of the ESB2, the BMC has a can access to and control the primary network interfaces. All LAN features for a given LAN channel are disabled unless the channel's access mode is set to Always Enabled. If an Intel® Remote Management Module is installed, then the ESB2 embedded LAN channels are configured differently.

Dedicated MAC Address
Each of the ESB2's two NIC channels has a unicast MAC filter reserved BMC use. These filters enable the BMC to receive network data streams that are logically separate from, and invisible to, operating systems and software running on the server, despite sharing the same physical LAN connections. This allows the BMC to support features beyond standard IPMI-over-LAN, such as DHCP, full ARP request / response, and ICMP, without requiring a separate Ethernet cable. This also allows an Intel® Remote Management Module add-in card to have nearly full access to the ESB2 embedded NICs in a shared NIC configuration. To prevent users from disrupting the BMC's ESB2 LAN configuration, the BMC treats LAN configuration parameter 5, MAC Address, as read-only for ESB2 NICs. Using the Set LAN Configuration Parameter command to attempt to change the MAC address on an ESB2 NIC has no effect, and the BMC returns error code 0x82, Attempt to write a read-only parameter, per IPMI errata E394.

Address Resolution Protocol (ARP)
The BMC can receive and respond to ARP requests on ESB2 NICs, and can also generate gratuitous ARPs. For gratuitous ARP support, the gratuitous ARPs bit (bit 0) in the Set LAN Configuration Parameter command needs to be set and the LAN interface needs to be active (IPMI-over- LAN).

For ARP response support, set the ARP response bit (bit 1) in the Set LAN Configuration Parameter command. IPMI-over-LAN does not need to be activated for ARP response to work. The BMC's default configuration on power on, or when the private store map changes or is corrupted, is for ARP generation to be disabled.

The Intel® Remote Management Module can override the BMC-generated ARP control configuration through the BMC LAN Configuration Override parameter. If the ARP response override bit is set in that parameter and filtered straight pass-through is active, the BMC behaves as though its BMC-generated ARP responses control bit is set to 0b: Disabled. If the ARP response override bit is clear, or SPT is inactive, the BMC honors the setting of the BMCgenerated ARP responses control bit.
The BMC does not support receiving or generating ARP packets on non-ESB2 NICs. Using the Set LAN Configuration Parameter command to attempt to enable ARP responses or gratuitous ARPs on such a NIC has no effect, and the BMC returns error code 0xCC, "Invalid data field in request."

Internet Control Message Protocol (ICMP)
The BMC supports the following ICMP message types targeting the BMC over ESB2 NICs:

- Echo request (ping): The BMC sends an Echo Reply
- Destination unreachable: If message is associated with an active socket connection within the BMC, the BMC closes the socket
- Redirect: The BMC updates its internal routing table
- Timestamp Request: The BMC sends a Timestamp Reply

Serial-over-LAN (SOL) 2.0
IPMI 2.0 introduced a standard serial-over-LAN feature. This is implemented as a standard payload type (01h) over RMCP+.
Three commands are implemented for SOL 2.0 configuration.

- Get / Set SOL 2.0 Configuration Parameters These commands are used to get and set the values of the SOL configuration parameters. The parameters are implemented on a per-channel basis.
- Activating SOL This command is not accepted by the BMC, but sent by the BMC is an active session when SOL is activated, to notify a remote client of the switch to SOL.

Event Filtering and Alerting

The BMC supports the following IPMI 2.0 alerting features.

- Platform event filtering (PEF)
- Dial paging
- Alert over LAN
- Alert over serial / point-to-point protocol (PPP).

Platform Event Filtering (PEF)
The BMC monitors platform health and logs failure events into the SEL. PEF provides a flexible, general mechanism that enables the BMC to perform selectable actions triggered by a configurable set of platform events. The BMC supports the following PEF actions:

- Power off
- Power cycle
- Reset
- Diagnostic interrupt
- OEM action
- Alerts

Alert-over-LAN
Standard and advanced PET alerts are supported over a LAN. Alert-over-LAN is used to notify remote system management application about PEF-selected events, regardless of the state of the operating system. LAN alerts can be sent over any of the LAN channels. The BMC implements three OEM PEF parameters associated with PET alerts over the LAN.

Alert Policies
Associated with each PEF entry is an alert policy that determines whether the alert is a dial page or a PPP alert, and over which IPMI channel the alert is to be sent. There is a maximum of 20 alert policy entries. There are no pre-configured entries in the alert policy table because the destination types and alerts may vary by user. Each entry in the alert policy table contains 4 bytes for a maximum table size

of 80 bytes.

MIB File
A modular information block (MIB) text file is provided with the BMC firmware to aide an SNMP browser in decoding the PETs generated by the BMC. The file provides information on PETs associated with the default PEF filter entries only. Users must extend the MIB file for any userconfigured PEF filters.

BMC Flash Update

Immediate Firmware Update
The BMC provides a firmware transfer mode that allows the BMC firmware to be updated. When in this mode, normal BMC functions, such as sensor monitoring and alert generation, are not performed. Only firmware update functions are provided and only the SMS interface is supported. Firmware transfer mode is entered in one of the following ways.

- An Enter Firmware Transfer Mode command is executed This command is available in both operational and firmware transfer modes. If executed over an interface other than the SMS interface, the client loses communication with the BMC.
- The BMC is reset while the Force Update signal is asserted. This signal is asserted using the Force Update jumper. The BMC only samples this signal at BMC startup, which occurs when AC power is applied or when the BMC exits Firmware Transfer Mode.
- The BMC is reset and the initialization code detects that the BMC firmware is corrupt (has an invalid checksum). This includes both operational code and platform information area (PIA) data corruption. The system may be permanently inoperable if the boot code is corrupted during a boot code update.

On entering firmware transfer mode, system power is not automatically applied; the user must manually power on the system using the front panel power button. Processor-enables are left in the state defined before entering firmware transfer mode (default: all processors enabled) and the system is allowed to boot (reset released). All front panel buttons are ignored. This mode is exited in two ways.

- An Exit Firmware Transfer Mode command is executed (assuming valid firmware).
- The BMC is reset and the initialization code detects valid BMC firmware.

The BMC firmware is divided into two main functional pieces:

- Operational code – The BMC application code used during normal operation.
- Boot block code – The BMC initialization code that checks for valid firmware.

The operational and boot block code can be updated and verified using firmware transfer commands delivered through the LPC / KCS SMS host interface. Firmware transfer commands allow any area of the BMC flash to be updated. Firmware transfer mode functions understand the block structure of the flash device used on the server board, so the update utility need not, and cannot, issue erase commands. Flash blocks are erased as necessary before the first write to a block.

The boot block area of the flash device is physically protected, either by the design of the flash component or by address-line decoding and write-enable gating. A boot code protection jumper enables

updating the boot block. The firmware transfer code cannot sense the state of this jumper, but if the jumper is not in the enabled position, boot block writes fail. Operational code image updates overwrite the last-known-good operational code image.

On-line Firmware Update

In addition to immediate firmware transfer mode updates, the BMC supports online updates. This feature allows a new BMC image to be copied into the inactive operational code bank while the BMC continues to operate. The operational code bank is opposite the executing operational bank. At the next system reset, the BMC hard-resets, verifies the new downloaded BMC image through the boot code, and transfers control to and executes the new image if verification succeeds. If the verification fails, the last known good operational code bank is automatically selected for execution. This prevents an inoperable or degraded server. The BMC can also save the previous firmware before updating with the new firmware. If the update fails, then the BMC can roll back to the previous version. A manual roll back process is also available.

Operational code, the platform information area (PIA), and the SDR firmware area can be updated. Boot code updates are not permitted online because a failure during a boot code update may leave the system permanently inoperable. Configuration settings are not updated or rolled back. Any one or all of the supported firmware area types (operational, PIA, and SDR) may be updated in one update operation.

The BMC switches between the new and old images by storing two operational code images, one image per operational code bank, and selecting the appropriate images with a hardwareassisted bank selection mechanism. The “staging image” consists of the inactive operational code bank (the operational code bank opposite the executing operational bank). The “execution image” and the “rollback image” are identical; they consist of the executing operational code image

BIOS-BMC Interactions

The BIOS-BMC interactions include the following:

- FRB2
- The BIOS uses the Get Self Test Results command to determine the health of the BMC.
- BMC time-stamp synchronization:
- The BIOS can use the BMC secure mode features to lock-out front panel access. BIOS control of this feature is configured in the BIOS Setup utility.
- System information for SMBIOS, the BIOS Setup utility and BIOS-BMC behavior: The BIOS retrieves sub-system inventory information to display it in the BIOS Setup utility, and to use it to populate certain SMBIOS fields. The BIOS has different behaviors with respect to BMC depending on system OEM type.
- Set system GUID: The BIOS initializes the BMC system globally unique ID (GUID) at every system boot. The system GUID and the BIOS Universal Unique Identifier (UUID) are the same. During manufacturing, the UUID is programmed into a reserved area of the BIOS flash.
- Set processor state: The BIOS uses the Set Processor State command to communicate fault states to the BMC.
- Boot control: A remote console application can set the boot options, then send a command to reset or power-cycle the server. The boot flags only apply for one system restart. The BIOS reads the boot flags from the BMC during the system boot.
- Serial MUX control: The BMC has a command that allows the BIOS to ‘force’ the serial

connector to be switched over to the server board serial controller. This command is provided to support the pre-boot only and disabled configurations of the EMP.

- Console redirection using SOL: The BIOS interacts with the BMC's SOL feature to provide console redirection via SOL.
- ACPI:
  - The BIOS keeps the BMC synchronized with the system ACPI state. This is necessary because some ACPI states cannot be sensed by the hardware signals that the BMC monitors. Synchronization is done with the Set ACPI Power State command.
  - The BIOS sets the BMC's ACPI configuration during boot using the Set ACPI Configuration command.
- Memory RAS and bus error monitoring: See section 22.20.
- ASF POST progress queuing.
- BIOS SEL logging: The BIOS logs SEL events using the Platform Event Message command.
- Disable PEF on entry to setup: Per the recommendation in the PEF Startup Delay section of the IPMI 2.0 specification, The BIOS disables PEF when the BIOS Setup Utility is run and restores it restores when the BIOS Setup utility is exited.
- Watchdog timer interactions:
  - During the boot process, the BIOS uses the BMC's Watchdog Timer for FRB2.
  - After the system boots, the BIOS starts the BMC's Watchdog Timer for "OS Load" usage. To prevent the timer from expiring, server management software agent turns off the timer after the operating system is successfully loaded.
- Processor TControl: See section 22.5.1.
- DIMM Throttling and Fan Management Interactions: See section 22.17.1.
- Synchronization of BIOS and BMC access to SIO3: The BMC and BIOS share access to the SIO3 device. This is done using the Acquire System Resource Semaphore command. At system boot the BMC automatically relinquishes ownership of the resource until BIOS returns it or there is a timeout waiting for BIOS to give back the semaphore.
- PCI Express* link status update

.

Sensors

Table 73 lists the sensor identification numbers and information regarding the sensor type, name, what thresholds are supported, assertion and deassertion information, and a brief description of what the sensor is used for. See the Intelligent Platform Management Interface Specification, Version 2.0, for sensor and event / reading-type table information.

- Sensor Type
  - The Sensor Type references the values enumerated in the Sensor Type Codes table in the IPMI specification. It provides the context in which to interpret the sensor, e.g., the physical entity or characteristic that is represented by this sensor.
- Event / Reading Type
  – The Event / Reading Type references values from the Event / Reading Type
  – Code Ranges and Generic Event / Reading Type Codes tables in the IPMI specification.

Digital sensors are a type of discrete sensors, which have only two states.

- Event Thresholds / Triggers
    - Event Thresholds are supported event generating thresholds for Threshold type sensors.
- [u,l][nr,c,nc] upper nonrecoverable, upper critical, upper noncritical, lowernonrecoverable, lower critical, lower noncritical
- uc, lc upper critical, lower critical
    - Event Triggers are supported event generating offsets for Discrete type sensors. The offsets can be found in the Generic Event / Reading Type Codes or Sensor Type Codes tables in the IPMI specification, depending on whether the sensor event / reading type is generic or a sensor specific response.
- Assertion / Deassertion
    - Assertions and Deassertion indicators reveals what type of events this sensorgenerates:
- As: Assertions
- De: Deassertion
- Readable Value / Offsets
    - Readable Value indicates the type of value returned for threshold and other nondiscretetype sensors.
    - Readable Offsets indicates the offsets for discrete sensors that are readable via the Get Sensor Reading command. Unless otherwise indicated, all Event Triggers are readable, i.e., Readable Offsets consists of the reading type offsets that do not generate events.
- Event Data
    - This is the data that is included in an event message generated by the associated sensor.
    - For threshold-based sensors, the following abbreviations are used:
- R: Reading value
- T: Threshold value
- Rearm Sensors
    - The rearm is a request for the event status for a sensor to be rechecked and updated upon a transition between good and bad states. Rearming the sensors can be done manually or automatically. This column indicates the type supported by the sensor. The following abbreviations are used in the comment column to describe a sensor:
- A: Auto-rearm
- M: Manual rearm
- Default Hysteresis
    - Hysteresis setting applies to all thresholds of the sensor. This column provides the count of hysterisis for the sensor, which can be 1 or 2 (positive or negative Hysteresis).
- Criticality
    - Criticality is a classification of the severity and nature of the condition. It also controls the behavior of the Front Panel Status LED (see Table 6, System Status LED Indicator States.
- Standby
    – Some sensors operate on standby power. These sensors may be accessed and / or generate events when the main (system) power is off, but AC power is present.

Table 104. Sensors

Sensor Name[3]	Sensor #	Sensor Type	Event / Reading Type	Event Offset Triggers	Contrib. To System Status	Assert / De-assert	Readable Value / Offsets	Event Data	Rearm	Stand-by
Power Unit Status	01h	Power Unit 09h	Sensor Specific 6Fh	Power down Power cycle A/C lost	OK	As and De	–	Trig Offset	A	X
				Soft power control failure Power unit failure	Fatal					
Power Unit Redundancy[4]	02h	Power Unit 09h	Generic 0Bh	Fully Redundant	OK	As and De	–	Trig Offset	A	X
				Redundancy lost	Degraded					
				Redundancy degraded	Degraded					
				Non-red: suff res from redund	Degraded					
				Non-red: suff from insuff resources	Degraded					
				Non-red: insufficient resources	Fatal					
				Redun degrade from fully redun	Degraded					
				Redun degrade from non-redundant	Degraded					
Watchdog	03h	Watchdog 2 23h	Sensor Specific 6Fh	Timer expired, status only Hard reset Power down Power cycle Timer interrupt	OK	As	–	Trig Offset	A	X

Sensor Name[a]	Sensor #	Sensor Type	Event / Reading Type	Event Offset Triggers	Contrib. To System Status	Assert / De-assert	Readable Value / Offsets	Event Data	Rearm	Stand-by
Platform Security Violation	04h	Platform Security Violation Attempt 06h	Sensor Specific 6Fh	Secure mode violation attempt	OK	As and De	–	Trig Offset	A	X
Physical Security	05h	Physical Security 05h	Sensor Specific 6Fh	Chassis intrusion	OK	As and De	–	Trig Offset	A	X
				LAN least lost	Degraded					
FP Diag Interrupt (NMI)	07h	Critical Interrupt 13h	Sensor Specific 6Fh	Front panel NMI / diagnostic interrupt Bus uncorrectable error	OK	As	–	Trig Offset	A	–
System Event Log	09h	Event Logging Disabled 10h	Sensor Specific 6Fh	Log area reset / cleared	OK	As	–	Trig Offset	A	X
Session Audit	0Ah	Session Audit 2Ah	Sensor Specific 6Fh	00h – Session activation 01h – Session deactivation	OK	As	–	As defined by IPMI	A	X
System Event ('System Event')	0Bh	System Event 12h	Sensor Specific 6Fh	00 – System reconfigured 04 – PEF action	OK	As	–	Trig Offset	A	X
BB Temp	30h	Temperature 01h	Threshold 01h	[u,l] [c,nc]	nc = Degraded c = Non-fatal	As and De	Analog	R, T	A	X
FP Temp	32h	Temperature 01h	Threshold 01h	[u,l] [c,nc]	nc = Degraded c = Non-fatal	As and De	Analog	R, T	A	X
Reserved for future use	33h	–	–	–	–	–	–	–	–	–

Sensor Name[3]	Sensor #	Sensor Type	Event / Reading Type	Event Offset Triggers	Contrib. To System Status	Assert / De-assert	Readable Value / Offsets	Event Data	Rearm	Stand-by
Mem Brd A-D Temp	40h – 43h	Temperature 01h	Threshold 01h	[u,l] [c,nc]	nc = Degraded c = Non-fatal	As and De	Analog	R, T	A	–
Mem A-B and C-D Aggregate Temps	48h – 49h	Temperature 01h	Threshold 01h	–	–	–	Analog	–	–	–
Tach Fan Sensors	50h – 57h	Fan 04h	Threshold 01h	[l] [c,nc]	nc = Degraded c = Non-fatal[2]	As and De	Analog	R, T	M	–
Fan Present Sensors	60h – 65h	Fan 04h	Generic 08h	Device present	OK	As and De	–	T	A	–
Reserved for BMC internal use	6Ch – 6Eh	–	–	–	–	–	–	–	–	–
Fan Redun-dancy[4]	6Fh	Fan 04h	Generic 0Bh	Fully redundant	OK	As and De	–	Trig Offset	A	X
				Redundancy lost	Degraded					
				Non-red: suff res from redund	Degraded					
				Non-red: suff from insuff resources	Degraded					
				Non-red: suff from insuff resources	Degraded					
				Non-red: insufficient resources	Non-fatal					
				Redun degrade from full	Degraded					
				Redun degrade from non-redundant	Degraded					
Power Supply Status[4, 1]	70h	Power Supply 08h	Sensor Specific 6Fh	Presence	OK	As and De	–	Trig Offset	A	X
				Failure	Degraded					
				A/C lost	Degraded					
				Configuration error	OK					

Sensor Name[3]	Sensor #	Sensor Type	Event / Reading Type	Event Offset Triggers	Contrib. To System Status	Assert / De-assert	Readable Value / Offsets	Event Data	Rearm	Stand-by
Power Supply Status[4] 2	71h	Power Supply 08h	Sensor Specific 6Fh	Presence	OK	As and De	–	Trig Offset	A	X
				Failure	Degraded					
				A/C lost	Degraded					
				Configuration error	OK					
PS1 Temp	72h	Temperature 01h	Threshold 01h	[u,l] [c,nc]	nc = Degraded c = Non-fatal	As and De	Analog	R, T	A	X
PS2 Temp	73h	Temperature	Threshold 01h	[u,l] [c,nc]	nc = Degraded c = Non-fatal	As and De	Analog	R, T	A	X
Mem Riser A PWRGD Fail	78h	Power Supply 08h	Digital Discrete 03h	PWR Good Fail	Fatal	As	01h	Trig Offset	M	–
Mem Riser B PWRGD Fail	79h	Power Supply 08h	Digital Discrete 03h	PWR Good Fail	Fatal	As	01h	Trig Offset	M	–
Mem Riser C PWRGD Fail	7Ah	Power Supply 08h	Digital Discrete 03h	PWR Good Fail	Fatal	As	01h	Trig Offset	M	–
Mem Riser D PWRGD Fail	7Bh	Power Supply 08h	Digital Discrete 03h	PWR Good Fail	Fatal	As	01h	Trig Offset	M	–
Baseboard PWRGD Fail	7Ch	Power Supply 08h	Digital Discrete 03h	PWR Good Fail	Fatal	As	01h	Trig Offset	M	–
SAS BP PWRGD Fail	7Dh	Power Supply 08h	Digital iscrete 03h	PWR Good Fail	Fatal	As	01h	Trig Offset	M	–
IO Riser PWRGD Fail	7Eh	Power Supply 08h	Digital Discrete 03h	PWR Good Fail	Fatal	As	01h	Trig Offset	M	X
System ACPI Power State	82h	System ACPI Power State 22h	Sensor Specific 6Fh	S0 / G0 S1 S3 S4 S5 / G2 G3 mechanical off	OK	As	–	Trig Offset	A	X

Sensor Name[a]	Sensor #	Sensor Type	Event / Reading Type	Event Offset Triggers	Contrib. To System Status	Assert / De-assert	Readable Value / Offsets	Event Data	Rearm	Stand-by
Button	84h	Button 14h	Sensor Specific 6Fh	Power button Reset button	OK	As	–	Trig Offset	A	X
SMI Timeout	85h	SMI Timeout F3h	Digital Discrete 03h	01h – State asserted	Fatal	As and De	–	Trig Offset	A	–
Chassis Intrusion Cable	86h	Cable / Interconnect 1Bh	Digital Discrete 08h	00h – Removed 01h -- Inserted	OK	As & De	–	Trig Offset	A	X
NMI Signal State	87h	OEM C0h	Digital Discrete 03h	01h – State asserted	OK	–	01h	–	–	–
SMI Signal State	88h	OEM C0h	Digital Discrete 03h	01h – State asserted	OK	–	01h	–	–	–
Proc 1 Status	90h	Processor 07h	Sensor Specific 6Fh	IERR	Fatal	As and De	–	Trig Offset	M	X
				Thermal trip	Fatal					
				Config error	Fatal					
				Presence	OK					
				Disabled	Degraded					
Proc 2 Status	91h	Processor 07h	Sensor Specific 6Fh	IERR	Fatal	As and De	–	Trig Offset	M	X
				Thermal trip	Fatal					
				Config error	Fatal					
				Presence	OK					
				Disabled	Degraded					
Proc 3 Status	92h	Processor 07h	Sensor Specific 6Fh	IERR	Fatal	As and De	–	Trig Offset	M	X
				Thermal trip	Fatal					
				Config error	Fatal					
				Presence	OK					
				Disabled	Degraded					
Proc 4 Status	93h	Processor 07h	Sensor Specific	IERR	Fatal	As and De	–	Trig Offset	M	X
				Thermal trip	Fatal					

Sensor Name[8]	Sensor #	Sensor Type	Event / Reading Type	Event Offset Triggers	Contrib. To System Status	Assert / De-assert	Readable Value / Offsets	Event Data	Rearm	Stand-by
			6Fh	Config error	Fatal					
				Presence	OK					
				Disabled	Degraded					
CPU Missing	94h	Processor 07h	Generic 03h	01h — State asserted	Fatal	As and De	–	R, T	–	X
CPU Throttled	95h	Processor 07h	Generic 03h	01h — State asserted	Fatal	As and De	–	R, T	A	–
P1 Therm Margin	98h	Temperature 01h	Threshold 01h	–	–	–	Analog	–	–	–
P2 Therm Margin	99h	Temperature 01h	Threshold 01h	–	–	–	Analog	–	–	–
P3 Therm Margin	9Ah	Temperature 01h	Threshold 01h	–	–	–	Analog	–	–	–
P4 Therm Margin	9Bh	Temperature 01h	Threshold 01h	–	–	–	Analog	–	–	–
PCI-Express Link Sensors	9Eh - AFh	Critical Interrupt 13F	Sensor Specific 6Fh	Bus correctable error	Degraded	As	–	See the BIOS EPS	A	–
				Bus uncorrectable error	Non-fatal					
P1 Therm Ctrl %	B0h	Temperature 01h	Threshold 01h	[u] [c,nc]	Non-fatal	As and De	Analog	Trig Offset	A	–
P2 Therm Ctrl %	B1h	Temperature 01h	Threshold 01h	[u] [c,nc]	Non-fatal	As and De	Analog	Trig Offset	A	–
P3 Therm Ctrl %	B2h	Temperature 01h	Threshold 01h	[u] [c,nc]	Non-fatal	As and De	Analog	Trig Offset	A	–
P4 Therm Ctrl %	B3h	Temperature 01h	Threshold 01h	[u] [c,nc]	Non-fatal	As and De	Analog	Trig Offset	A	–
Proc 1 VRD Hot	B8h	Temperature 01h	Digital Discrete 05h	01h – Limit exceeded	Non-fatal	As and De	–	Trig Offset	M	–

Sensor Name[8]	Sensor #	Sensor Type	Event / Reading Type	Event Offset Triggers	Contrib. To System Status	Assert / De-assert	Readable Value / Offsets	Event Data	Rearm	Stand-by
Proc 2 VRD Hot	B9h	Temperature 01h	Digital Discrete 05h	01h – Limit exceeded	Non-fatal	As and De	–	Trig Offset	M	–
Proc 3 VRD Hot	BAh	Temperature 01h	Digital Discrete 05h	01h – Limit exceeded	Non-fatal	As and De	–	Trig Offset	M	–
Proc 4 VRD Hot	BBh	Temperature 01h	Digital Discrete 05h	01h – Limit exceeded	Non-fatal	As and De	–	Trig Offset	M	–
Proc 1 PWRGD	BCh	Power Supply 08h	"Digital" Discrete 03h	PWR Good Fail	Fatal	As	01h	Trig Offset	M	–
Proc 2 PWRGD	BDh	Power Supply 08h	"Digital" Discrete 03h	PWR Good Fail	Fatal	As	01h	Trig Offset	M	–
Proc 3 PWRGD	BEh	Power Supply 08h	"Digital" Discrete 03h	PWR Good Fail	Fatal	As	01h	Trig Offset	M	–
Proc 4 PWRGD	BFh	Power Supply 08h	"Digital" Discrete 03h	PWR Good Fail	Fatal	As	01h	Trig Offset	M	–
MEM_A DIMM 1- 8	C0h – C7h	Slot Connector 21h	Sensor Specific 6Fh	Fault status asserted	Situation dependent[5]	As and De	–	Trig Offset	A	–
				Device installed	OK					
				Disabled	Situation dependent[5]					
				Sparing	OK					
MEM_B DIMM 1- 8	C8h – CFh	Slot Connector 21h	Sensor Specific 6Fh	Fault status asserted	Situation dependent[5]	As and De	–	Trig Offset	A	–
				Device installed	OK					
				Disabled	Situation dependent[5]					
				Sparing	OK					
MEM_C DIMM 1- 8	D0h – D7h	Slot Connector 21h	Sensor Specific 6Fh	Fault status asserted	Situation dependent[5]	As and De	–	Trig Offset	A	–
				Device installed	OK					

Sensor Name[a]	Sensor #	Sensor Type	Event / Reading Type	Event Offset Triggers	Contrib. To System Status	Assert / De-assert	Readable Value / Offsets	Event Data	Rearm	Stand-by
				Disabled	Situation dependent[5]					
				Sparing	OK					
MEM_D DIMM 1- 8	D8h – DFh	Slot Connector 21h	Sensor Specific 6Fh	Fault status asserted	Situation dependent[5]	As and De	–	Trig Offset	A	–
				Device installed	OK					
				Disabled	Situation dependent[5]					
				Sparing	OK					
Memory Error A - D	ECh – EDh	Memory 0Ch	Sensor Specific 6Fh	Uncorrectable ECC	OK[6]	As and De	–	Trig Offset	A	–
B0 DIMM Sparing Enabled	F0h	Entity Presence 25h	Sensor Specific 6Fh	Entity present	OK	As	–	Trig Offset	A	–
B0 DIMM Sparing Redun-dancy	F1h	Memory 0Ch	Discrete 0Bh	Fully redundant	OK	As	–	Trig Offset	A	–
				Non-red: suff res from redund	Degraded					
				Non-red: Insuff res	Fatal					
B1 DIMM Sparing Enabled	F2h	Entity Presence 25h	Sensor Specific 6Fh	Entity present	OK	As	–	Trig Offset	A	–
B1 DIMM Sparing Redun-dancy	F3h	Memory 0Ch	Discrete 0Bh	Fully redundant	OK	As	–	Trig Offset	A	–
				Non-red: suff res from redund	Degraded					
				Non-red: insuff res	Fatal					
B01 DIMM Mirroring Enabled	F4h	Entity Presence 25h	Sensor Specific 6Fh	Entity present	OK	As	–	Trig Offset	A	–
B01 DIMM Mirroring Redun-dancy	F5h	Memory 0Ch	Discrete 0Bh	Fully redundant	OK	As	–	Trig Offset	A	–
				Non-red:suff res from redund	Degraded					

Sensor Name[a]	Sensor #	Sensor Type	Event / Reading Type	Event Offset Triggers	Contrib. To System Status	Assert / De-assert	Readable Value / Offsets	Event Data	Rearm	Stand-by
				Non-red: insuff res	Fatal					

Note 1: Not supported except for ESB2 embedded NICs
Note 2: For system with redundant cooling capability, the contribution to system status is determined by the fan redundancy sensor.
Note 3: Sensor name strings in SDR may vary from the names in this table.
Note 4: Sensor only present on systems that have applicable redundancy (for instance, fan or power supply).
Note 5: The BMC does not provide a direct contribution to overall system status due to the DIMM sensors. BIOS determines contribution depending on failure scenario and uses the Set Fault Indication command to provide this information to the BMC. See Table 6, System Status LED Indicator States.
Note 6: Error logging for this sensor is due to port-mortem memory error scan after an SMI Timeout has occurred. Contribution to system status is determined by
the SMI Timeout sensor.

Hot-Swap Controller (HSC) Architecture

The HSC uses a VSC410* SAF-TE enclosure processor (SEP). This microcontroller employs a v3000 RISC CPU, 8 KB of internal SRAM, GPIO, SGPIO, two general purpose UARTs, one SPI, and four I2C compatible interfaces.

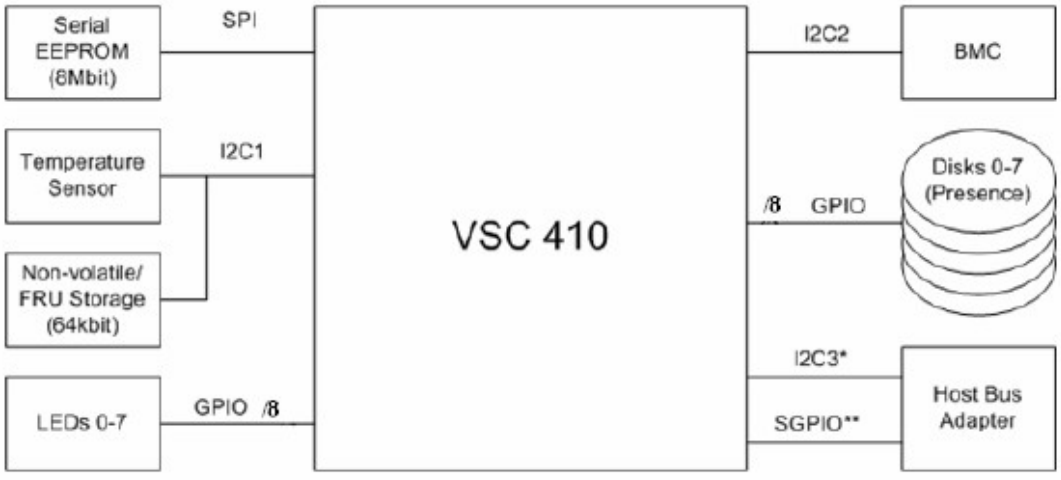


Figure 71. HSC Interface Routing

* If present, SGPIO is disconnected.

** If present, I2C3 is disconnected.

I2C Interfaces

The VSC410 supports four I2C compatible serial interfaces. These multi-master interfaces are configured in firmware to operate at 100 KHz. Optional support functions, such as I2C bus cleanups, can be configured in firmware.

Table 105. I²C Bus Assignements

I²C Bus Number	Connection Protocol	Connected Device(s)
I2C0	Reserved	None
I2C1	Master / slave I²C (private bus)	Temperature sensor, NV/FRU EEPROM
I2C2	IPMB	Baseboard management controller
I2C3	SES2-over-I²C, SAFTE	Host bus adapter

Serial Peripheral Interface (SPI)

The VSC410 SPI accesses operational code in a separate SPI-compatible EEPROM device. This interface is private and can only be accessed by the HSC to retrieve or update firmware.

GPIO Pins

Twenty GPIO pins are on the VSC410:

- Eight for drive presence detection
- Eight are for LED control
- One for write protection control for both the SPI and I2C EEPROM devices
- Two for SFF-8087 cable detection via side-band ground pins

Serial General-purpose Input / Output (SGPIO)

The VSC410 supports serial GPIO (SGPIO). This four-wire bus provides the status for up to 32 disks via a series of fault / locate / active bits. The hot-swap controller supports two SGPIO interfaces (SGPIO0 and SGPIO1), according to the SFF-8485 specification. Each SGPIO interface provides disks status for four disks. This implementation supports only SGPIO communication from the host bus adapter (HBA) to the HSC (simplex).

HSC Functional Specifications

Platform Determination

The HSC provides a unique platform identifier through several management interfaces. The table below shows the identifiers returned by the interfaces on the backplane. The I2C identification is returned as part of the IPMI Get Device ID response. The SAFTE and SES responses are part of the inquiry data. The firmware BootInfo identifier is embedded in the firmware image header.

Table 106. Platform Identification

Interface	Identifier
I2C/IPMB	0A0Dh
I2C/SAF-TE	SCA HSBP M12....
I2C/SES	SCA HSBP M12....
Firmware BootInfo	SCA HSBP M12

Auto Detection of Platform Type

The HSC firmware is shared by both server systems, but the HSC communication through the SGPIO differs. The HSC firmware configures to the appropriate bus adapter type by detecting a unique data pattern on the SLoad line provided by BIOS during the first 20 seconds of POST. The table shows the unique SGPIO data pattern for the ESB2 configuration. If the pattern is not seen during the first 20 seconds of POST, the HSC will assume default the SGPIO mode.

Table 107. Bus Adapter Identification

Signal	Value
SLOAD	0x0C
SDATA0_in	0xB6D

System Initialization

Non-Volatile Setting Initialization

Upon initialization, the HSC reads non-volatile settings from its I2C EEPROM. These settings include initial sensor configuration values and FRU/sensor record integrity headers. If an I2C EEPROM cannot be found, then default values are used.

Sensor Initialization

The HSC receives sensor initialization values from the baseboard management controller (BMC). The BMC sends IPMI sensor initialization values to the HSC during IPMI initialization agent runtime.

Cable Detection

The HSC detects the presence of the SFF-8087 cables upon firmware initialization. The detection is done via active-low GPIO signals routed from out-of-band signal ground pins. Each SFF-8087 connection corresponds to four disk drive connections. Depending on the combination of presence signals, the HSC configures itself for four- or eight-disk management as shown in Table 108. After self-configuration, the HSC only acknowledges either four or eight disks in management responses. In a four disk configuration (cable A only) the HSC reports four disk slots in IPMI, SAFTE, and SES responses. All LEDs remain accessible via IPMI and SGPIO, regardless of the number of cables detected.

Table 108. Cable Detect Configuration

Cable A Detected	Cable B Detected	Configuration
No	No	Invalid Configuration: will use 8 disks by default.
No	Yes	Invalid configuration: will use 8 disks by default.
Yes	No	Will use 4 disks.
Yes	Yes	Will use 8 disks.

System Event Log (SEL)

The VSC410 controller does not implement a system event log. Instead, SEL entries are maintained by the BMC. If the BMC is unable to accept platform event messages, the VSC410 does not cache the entries.

Sensor Data Repository (SDR)

The VSC410 controller does not implement a sensor data repository. Instead, the BMC maintains the HSC SDR entries.

Field Replaceable Unit (FRU) Inventory Device

The VSC410 supports an I2C-compatible EEPROM for FRU storage located at address 0xAC.The EEPROM is on a private I2C bus, and is accessible only by the HSC, or through Master write-read I2C commands.

The FRU storage contains:

- Common header
- Internal use area
- Board information area 00h. The FRU file must be uploaded to the FRU EEPROM using an Intel FRUSDR utility.

HSC FRU Format
The FRU inventory area format follows the Platform Management FRU Information Storage Definition. See Platform Management FRU Information Storage Definition, Version 1.0. The HSC provides only low-level access to the FRU inventory area storage. It does not validate or interpret the data written to the FRU, including the common header area. Applications cannot relocate or resize FRU inventory areas.

The HSC provides 256 bytes of non-volatile storage to hold the serial number, part number, and other FRU inventory information about the hot-swap backplane. The HSC implements commands that allow this private FRU data to be written or read via the IPMB.

Note: *Fields in the internal use area are not for OEM use. Intel reserves the right to relocate and redefine these fields without prior notification. Definition of this area is part of the software design. The format in the internal use area may vary with different BMC firmware revisions.*

Temperature Monitoring
The VSC410 HSC supports an I2C-compatible temperature sensor located at address 0x90 for backplane temperature monitoring. This sensor is on a private I2C bus that is shared with the FRU storage device.

The HSC monitors and reports the temperature using values that the BMC provides during initialization. The HSC supports reporting lower critical (lc), lower non-critical (lnc), upper noncritical (unc), and upper critical (uc) thresholds. Threshold values are reported as going high or going low, depending on the direction of change. The HSC supports hysteresis values.

Disk Management

Drive Fault Light Control
The HSC activates and deactivates drive fault LEDs according to the states received via SAFTE or SES pages, or the SGPIO bus. Only the host bus adapter can change the state of a disk. IPMI commands can be used to toggle the drive fault LEDs for diagnostic purposes. The HSC does not have control of the green drive ready / activity LEDs. Disk hardware controls these LEDs.

Drive Presence Detection
The HSC detects drive presence and makes this information available via SAF-TE, SES2, and IPMI. It is the HSC firmware’s responsibility to make sure that the drive presence signals have been properly de-bounced.

Enclosure Temperature Sensing
A temperature sensor device is connected to the HSC via a private I2C bus. This device monitors the

enclosure temperature. The temperature can be read via SAF-TE, SES2, and IPMI commands. Programmable temperature thresholds are provided via IPMI commands. The HSC can be configured to issue an event message on the IPMB when a temperature threshold is crossed.

Slot Status to Fault Light State Mapping

The fault light state for each internal drive slot state is maintained by the hot-swap controller. The HSC supports various OEM LED models.

Table 109. Slot Status to Fault Light State Mapping

Slot Status						Fault Light State	
Device Inserted	Identify	Device Rebuilding	Rebuild Stopped	Device Faulty	Predicted Fault	Fault Light	Indicated Condition
X	0	0	0	0	0	Off	No errors
X	0	0	0	0	1	Slow blink	Predicted Fault
X	0	0	0	1	X	Steady on	Faulted Slot
1	0	1	0	X	X	Slow blink	Rebuild
0	0	1	0	X	X	Fast blink	Rebuild on empty slot
X	0	X	1	X	X	Fast blink	Rebuild Interrupted
X	1	X	X	X	X	Fast blink	Identify Slot

X = don't care

Fast Blink = ~2.5 Hz

Slow Blink = ~1 Hz.

HSC IPMB Application and Sensors

This section presents the additional specifications required for the HSC's implementation as an IPMI controller. See the Intelligent Platform Management Interface Specification for more information.

LUNs

The HSC accepts Intelligent Platform Management Bus requests directed to its LUN 00. There are no restrictions on the LUNs that the HSC uses when sending requests or responses to other controllers.

Sensors

The HSC implements the same basic sensor model that is utilized by the other management controllers in the system. Sensor model information is in the Intelligent Platform Management Interface Specification. A common set of IPMI commands configures the sensors and returns the threshold status. The following table specifies the sensor numbers and thresholds for the sensors implemented by the HSC.

Sensor initialization is handled as follows: The BMC implements the internal sensor initialization agent functionality specified in the Intelligent Platform Management Interface Specification. When the BMC initializes, it walks the sensor data records and configures the IPMB devices that have the Init Required

bit set in their SDRs. This includes setting sensor thresholds, enabling/disabling sensor event message scanning, and enabling/disabling sensor event messages, as indicated.

Table 110. HSC Sensor / Event Message Source Numbers

Sensor Name	Sensor #	Sensor Type (Hex)	Event / Reading Type Code	Event Data	Re-arm	Event / Threshold Trigger
Backplane Temperature	01h	Temp. (01h)	01h	Reading thresh. Value	auto	Assertion or deassertion of transition to uc, unc, lnc, lc
Drive Slot 0 Status	02h	Drive Slot (0Dh)	6Fh	Status	auto	Device Rebuilding. Device Faulty.
Drive Slot 1 Status	03h	Drive Slot (0Dh)	6Fh	Status	auto	Device Rebuilding. Device Faulty
Drive Slot 2 Status	04h	Drive Slot (0Dh)	6Fh	Status	auto	Device Rebuilding. Device Faulty
Drive Slot 3 Status	05h	Drive Slot (0Dh)	6Fh	Status	auto	Device Rebuilding. Device Faulty

Sensor Name	Sensor #	Sensor Type (Hex)	Event / Reading Type Code	Event Data	Re-arm	Event / Threshold Trigger
Drive Slot 4 Status	06h [3]	Drive Slot (0Dh)	6Fh	Status	auto	Device Rebuilding. Device Faulty
Drive Slot 5 Status	07h [3]	Drive Slot (0Dh)	6Fh	Status	auto	Device Rebuilding. Device Faulty
Drive Slot 6 Status	08h [3]	Drive Slot (0Dh)	6Fh	Status	auto	Device Rebuilding. Device Faulty
Drive Slot 7 Status	09h [3]	Drive Slot (0Dh)	6Fh	Status	auto	Device Rebuilding. Device Faulty
Drive Slot 0 Presence	0Ah	Drive Slot (0Dh)	08h	Presence [2]	auto	dev. remove, dev. Inserted
Drive Slot 1 Presence	0Bh	Drive Slot (0Dh)	08h	Presence [2]	auto	dev. remove, dev. Inserted
Drive Slot 2 Presence	0Ch	Drive Slot (0Dh)	08h	Presence [2]	auto	dev. remove, dev. Inserted
Drive Slot 3 Presence	0Dh	Drive Slot (0Dh)	08h	Presence [2]	auto	dev. remove, dev. Inserted
Drive Slot 4 Presence	0Eh [3]	Drive Slot (0Dh)	08h	Presence [2]	auto	dev. remove, dev. Inserted
Drive Slot 5 Presence	0Fh [3]	Drive Slot (0Dh)	08h	Presence [2]	auto	dev. remove, dev. Inserted
Drive Slot 6 Presence	10h [3]	Drive Slot (0Dh)	08h	Presence [2]	auto	dev. remove, dev. Inserted
Drive Slot 7 Presence	11h [3]	Drive Slot (0Dh)	08h	Presence [2]	auto	dev. remove, dev. Inserted

Notes:

1. Event messages are not generated for this sensor.
2. See Table 111.
3. Only available when HSC is in eight-disk mode.

Digital and Discrete Sensor Formats

Drive slot sensors have unique, device-specific formats.

Table 111. Sensor Formats

Sensor Name	Sensor #	Format (2-bytes)
Drive Slot Status	02h-09h	Bit 15:13: Reserved. Bit 12: Identify asserted. Bit 11: Prepared for Operation. Bit 10: Ready for Insertion/Removal. Bit 09: Device Inserted. Bit 08: Rebuild stopped. Bit 07: Hot Spare. Bit 06: Un-configured. Bit 05: Predicted Fault. Bit 04: Parity Check. Bit 03: In Critical Array. Bit 02: In Failed Array. Bit 01: Device Rebuilding. Bit 00: Device Faulty.
Drive Slot Presence	0Ah-11h	Bit 15:2: Reserved Bit 01: Device Inserted/Device Present Bit 00: Device Removed/Device Absent.

Event Message Generation
Specified sensor events that are detected by the HSC cause a corresponding event message to be sent out on the IPMB. Event message generation is configured via IPMI commands. The format for event messages is in the *Intelligent Platform Management Interface Specification*.

HSC Firmware Update
The HSC firmware is stored in a separate SPI-compatible EEPROM module. This EEPROM is only accessible by the HSC to read or write operational code. The HSC reads code actively from the SPI EEPROM, which can contribute to increased execution times.

HSC Update Over IPMB
Firmware updates primarily take place via the IPMB. This method requires a firmware update utility and an Intel hex-format image.

The HSC firmware EEPROM is divided into primary and secondary areas. The primary area holds operational code that is in use by the HSC. The secondary area stores an incoming firmware image. The transition between primary and secondary area is handled internally to the HSC firmware and is transparent to other management controllers. The following sections explain the IPMI commands used to update the firmware image.

Entering Firmware Transfer Mode
Firmware transfer / update mode can be entered at any time using the Enter Firmware Transfer Mode command to the HSC. Of the firmware transfer mode commands, only the Enter Firmware Transfer Mode command is executable from operational mode. The other firmware transfer commands are recognized only in firmware transfer mode.

Exiting Firmware Transfer Mode

This command causes firmware transfer mode to be exited. If the request data byte is not present, then the HSC immediately considers it an abort and returns to operational mode. When the command provides a 01h as request data, the HSC burns the new code, and initiates a hard reset. Sensor data is not retained across this reset and the controller initializes as if a power on reset occurred. The HSC provides an additional response byte indicating expected firmware burn and reboot time in seconds (0-255).

Firmware Transfer Version
The Get Device ID command returns the version number of the firmware. The HSC returns the device ID information from the primary code area, regardless of whether it is in firmware update mode or operational mode. When in firmware transfer mode, the HSC responds to Get Device ID with a short response. The auxiliary firmware revision data is truncated and the device available bit is set to 1.

Verifying Entry Into Firmware Transfer Mode
It is possible to verify that the HSC is in firmware transfer mode by sending an IPMI Get Device ID request. If the HSC responds with a truncated response (missing the auxiliary firmware revision) and the device available bit is set to 1, then it is in firmware transfer mode.

Set Program Segment Command
This command sets the upper 16 bits of the address for the Firmware Read, Firmware Program, and Get Firmware Range Checksum commands.

FLASH Erase and Sequential Programming
There is no explicit erase command. Flash blocks are erased as needed when the Exit Firmware Transfer Mode command is executed. Therefore, firmware updating proceeds sequentially from the beginning of the operational code.

The HSC ignores all interfaces during a flash erase. Firmware transfer applications should have their time-outs and retries designed accordingly. The worst-case flash erase time is one-half second.

Access to Operational Mode Commands
Except for Get Device ID, non-firmware transfer network functions and their associated responses are not recognized in firmware transfer mode. Firmware transfer ,ode must be exited before issuing non-firmware transfer commands, such as application or event message commands.

Glossary

Word / Acronym	Definition
16-bit Legacy	The traditional personal computer environment. Includes legacy Option ROMs and legacy 16-bit code.
ACPI	Advanced Configuration and Power Interface. ACPI is an open industry specification proposed by Intel, Microsoft and Toshiba. ACPI enables and supports reliable power management through improved hardware and operating system coordination. For more information, see ACPI_1.0b or ACPI_2.0.
AES	Advanced Encryption Standard
AL	After Life. Component of Intel® Platform Innovation Framework for EFI architecture.
AMB	Advanced Memory Buffer. Used on FBDIMMs.
AML	ACPI Machine Language.
ANSI	American National Standards Institute.
AP	Application processor.
API	Application Programming Interface. A software abstraction provided by the BIOS to applications and/or the Operating System.
APIC	Advanced Programmable Interrupt Controller
ARP	Address Resolution Protocol
ASCII	American Standard Code for Information Interchange. An 8-level code (7 bits plus parity check) widely used in data processing and data communications systems.
ASF	Alert Standards Forum
ASIC	Application specific integrated circuit
ASL	ACPI Source Language
ASR	Asynchronous System Reset.
Asserted	Active-high (positive true) signals are asserted when in the high electrical state (near power potential). Active-low (negative true) signals are asserted when in the low electrical state (near ground potential).
Asynchronous	An event that causes a change in state with no timing relationship to a particular timing reference (such as a clock signal).
BAR	Base Address Register. Device configuration registers that define the start address, length and type of memory space required by a device.
BDS	Boot Device Select. Component of Intel® Platform Innovation Framework for EFI architecture.
BEV	BootStrap Entry Vector.
BIOS	Basic Input/Output System. The firmware (Software embedded into hardware) that is used to boot the processors and initialize the chipset and I/O of the system prior to handing off control of execution to the operating system.
BIST	Built-in self test
BMC	Baseboard Management Controller. A management microcontroller on the baseboard that is used for voltage, temperature, and fan failure sensing. The device also serves as an I2C Master controller, and as the system's Event Receiver.
Bridge	Circuitry connecting one computer bus to another, allowing an agent on one to access the other.
BSP	Boot Strap Processor. The processor selected at boot time to be the primary processor in a multi-processor system.

CBC	Chassis bridge controller. A microcontroller connected to one or more other CBCs. Together they bridge the IPMB buses of multiple chassis.
CD	Compact Disk.
CE	Community European or Memory ECC Correctable Error.
CISPR	International Special Committee on Radio Interference
CLI	command-line interface
CLTT	Closed Loop Thermal Throttling for FB-DIMMs
CMOS	Complementary Metal-Oxide Semi-Conductor. In terms of this specification, this describes the PC-AT compatible region of battery-backed 128 bytes of memory on the server board.
CMP	Core Multi-Processing.
COM	Communications
CPD	Component Data Sheets
CRTM	Core Rott of Trust Measurement
CRU	Customer Replaceable Unit
Crystal Beach	A component of Intel® I/O Acceleration Technology (Intel® I/OAT) which provides improved Network I/O and RAID performance through a combination of chipset hardware support as well as BIOS and operating system driver support.
CSA	Canadian Standards Organization
CSR	control and status register
D2D	DC-to-DC converter
DB	Data Bus
dBA	deciBel Acoustic
DCA	Direct Cache Access. A component of Intel® I/O Acceleration Technology (Intel® I/OAT) which provides improved I/O network performance.
D-cache	Data cache. Processor-local cache dedicated for memory locations explicitly loaded and stored by running code.
DDR	Double Data-Rate memory.
DDR2	Double Data Rate DRAM devices (2nd generation)
Deasserted	Signals are deasserted when they are not asserted.
DHCP	Dynamic Host Configuration Protocol
DIMM	Dual In-line Memory Module. A packaging arrangement of memory devices on a socketable substrate.
DMA	Direct Memory Access
DPC	Direct Platform Control. Remote server management over the network or serial port regardless the status of the server's operating system or power state.
DPS	Distributed Power supply
DSDT	Differentiated System Description Table. An OEM must supply a DSDT to an ACPI-compatible operating system. The DSDT contains the Differentiating Definition Block, which supplies the implementation and configuration information about the base system.
DSS	Decision Support System
DT	Double Transition
DVI/DVO	Digital Video Input/Ouput.
DWord	Double Word, a 32-bit quantity.
DXE	Driver Execution Environment. Component of Intel® Platform Innovation Framework for EFI architecture.

ECC	Error Correction Code. Refers to a memory system that has extra bit(s) to support limited detection/correction of memory errors.
EEPROM	Electrically erasable programmable read-only memory
EFI	Extensible Firmware Interface. A new hardware/operating system interface for the BIOS to utilize in the bootup of the system.
EMI	Electromagnetic Interference
EMP	Emergency Management Port.
EPS	External Product Specification
ESB2	Enterprise South Bridge 2
ESD	Electro Static Discharge
Event Receiver	Term from the Intelligent Management Bus Communications Protocol. Refers to a device that can receive Event Messages via the Intelligent Management Bus. The BMC is typically the Event Receiver. The BMC can generate a system interrupt upon receipt of an Event Message, triggering the system's Critical Event Logging routines.
FBD	Fully buffered DIMM. Also referes to the FBDIMMs High speed serial interface.
FBD Channel	One electrical interface to one or more Fully Buffered DDRII DIMM.
FBDIMM	Fully Buffered DIMM.
FCC	Federal Communications Commission
Firmware	Any software that is permanently stored in an integrated circuit and used by the hardware for basic initialization or operation.
Flash	Also Flash ROM. A type of fast write Electrically Erasable Programmable ROM. FLASH ROM varies from what is typically referred to as EEPROM in that typical EEPROM is reprogrammable on a per byte basis, while Flash is byte-writable, but to be re-written must be block erased. The advantages of Flash include fast read/write speed and significantly lower cost per bit than EEPROM.
FML	Fast Management Link.
FNI	fast management link network interface
Formset	Framework term for display pages, which includes Setup pages.
FRB	Fault Resilient Booting. A hardware/firmware method to boot past any fault that would not prevent that system from otherwise operating.
FRU	Field Replaceable Unit. This also refers to an EEPROM device that stores part number and serial number information about the board.
FSB	Front Side Bus. Also known as System Bus. The processor-to-chipset (north bridge) interface.
FSC	Fan Speed Control.
FTM	firmware transfer mode
FWH	Firmware Hub. Physical storage hardware circuitry for the system firmware/BIOS.
GB	Gigabyte. 1024 Megabytes.
GCM3	Generic Communication Module (3rd Generation).
GND	Ground
GPIO	General Purpose Input/Output
GUI	Graphical User Interface
GUID	Globally Unique Identifier.
Hard Reset	A hardware driven reset event in the system that initializes all components and invalidates caches.
HBA	host bus adapter
HDD	Hard Disk drive

HDM	High Density Metric
HIS	Integrated Heat Spreader. The metallic surface covering of a processor package intended to encourage the flow of heat away from the processor die and to the heatsink.
HL	Hub-Link
HLT	Halt.
HP	Hot-plug.
HPC	High Pin Count
HPIB	Hot-Plug Indicator Board
HSBP	Hot-swap backplane
HSC	Hot-Swap Controller. The microcontroller that implements the SAF-TE command set and controls the fault lights and drive power on a hot-swap backplane.
HT Technology	Hyper-Threading Technology
I/O	Input / Output
I/O APIC	I/O Advanced Programmable Interrupt Controller.
I^2C or I2C	Inter-Integrated Circuit bus. A multi-master, 2-wire, serial bus used as the basis for the Intelligent Management Bus.
I^2O	Intelligent I/O. An open architecture for the development of device drivers in network system environments
IA	Intel® architecture
IBF	Input buffer
IC	Integrated Circuit
I-cache	Instruction cache. Processor-local cache dedicated for memory locations retrieved through instruction fetch operations.
ICH	I/O controller hub
ICMB	Intelligent Chassis Management Bus
IDE	Integrated Device Electronics
IEC	International Electrotechnical Commission
IERR	internal error
IMB	Intelligent Management Bus
INIT	initialization signal
Intel® I/OAT	Intel® I/O Acceleration Technology
Intel® RMM2	Intel® Remote Management Module 2
Intel® VT	Intel® Virtualization Technology
IOP	I2O-compliant I/O Platforms. These typically contain an I/O processor and I/O subsystem.
IPMB	Intelligent Platform Management Bus. Name for the architecture, protocol, and implementation of a special bus that interconnects the baseboard and chassis electronics and provides a communications media for system platform management information. The bus is built on I^2C and provides a communications path between management controllers such as the BMC and HSC.
IPMI	Intelligent platform Management Interface. An industry standard that defines standardized, abstracted interfaces to platform management hardware.
ISA	Industry Standard Architecture
iSCSI	Internet SCSI.
ISP	In System Programmable
ITE	Information Technology Equipment

ITP	In-Target Probe
JAE	Japan Aviation Electronics
JTAG	Joint Test Action Group. It is a test access port used for testing and debugging. An In-Circuit Emulator uses JTAG as a transport mechanism to access an on-chip debug module which is integrated in to the CPU.
KB	Kilobyte. 1024 bytes.
KCS	Keyboard Controller Style.
KT	keyboard text
KVM	Keyboard, Video, and Mouse.
LAN	Local Area Network.
LBC	Leaky Bucket Counter.
LCD	liquid crystal display
LED	Light Emitting Diode.
LOM	LAN on Motherboard.
LPC	Low Pin Count bus.
LUN	logical unit number. In the context of the Intelligent Management Bus protocol, this is a sub-address that allows messages to be routed to different 'logical units' that reside behind the same I2C slave address.
LVDS	Low Voltage Differential SAS
MAC	Media Access Control
MB	Megabyte. 1024 Kilobytes.
MB/s	Megabytes per second.
MBE	Memory ECC Multi-Bit Error.
MBR	Master Boot Record.
MC	Multi-Core processor technology.
MCH	Memory Controller Hub. A.K.A. Northbridge.
MD5	Message Digest 5. A hashing algorithm that provides higher security than MD2
MIB	Modular information block. A descfripive text translaion of a PET event, contained in a MIB file for use by an SNMP agent hen decoding SEL entries.
MRH-D	Memory Repeater Hub – DDR-II
ms	millisecond
MSI	PCI Message Signaled Interrupt.
MTBF	Mean Time Between Failures
MTRR	Memory Type Range Register.
MUX	multiplexer
NB	Northbridge. The component of the chipset that interfaces the processors to memory and I/O. A.K.A. MCH.
NB or SB	Northbound or Southbound FBD traffic.
NIC	Network Interface Controller/Card.
NMI	Non-Maskable Interrupt.
OBF	output buffer
ODT	On-Die Termination.
OEM	Original Equipment Manufacturer
OLTP	On-line Transaction Processing
OLTT	Open Loop Thermal Throttling for FB-DIMMs

OOB	Out of Band. Fancy acronym for sideband signal.
OPROM	Option ROM. Used to describe the code provided by hardware vendors to provide support for integrated devices or add-in adapters.
OS	Operating System
OTP	Over-Temperature Protection
OVP	Over-Voltage Protection
PAE	Physical Address Extensions.
PAL	Programmable Array Logic
PATA	Parrallel-ATA.
PCI	Peripheral Component Interconnect. Bus with multiplexed address and data lines primarily intended for use as an interconnect system between processor/memory and peripheral components or add-in cards.
PCI-Express*	PCI Express*.
PCR	Platform Configuration Register
PDB	Power distribution board
PECI	Platform Environmental Control Interface.
PEF	Platform Event Filtering.
PEI	Pre EFI Initialization. Component of Intel® Platform Innovation Framework for EFI architecture.
PEP	Platform Event Paging
PET	Platform Event Trap.
PFC	Power Factor Correction
PIA	platform information area
PIC	8259 Programmable Interrupt Controller logic.
PIROM	Processor Information ROM
PLD	Programmable Logic Device.
PME	Power Management Event.
PMI	Platform Management Interrupt.
PnP	Plug and Play. See PnP_BIOS and PnP_ISA.
POST	Power-on Self Test.
PROM	programmable read-only memory
PSMI	Power Supply Management Interface
PSU	Power supply Unit
PTS	Platform Trust Services
PVC	Poly Vinyl Chloride – a plastic
PWM	Pulse Width Modulation. The mechanism used to control the speed of system fans.
PXE	Pre-boot eXecution Environment. This is the system envirionment when EFI firmware is executing during boot and in control of system resources.
RAID	Redundent Array of Independent Disks.
RAM	Random Access Memory
Rank	The set of SDRAM devices on one DDR branch which provides the data packet.
RAS	Reliability, Availability, and Serviceability
RC4	Rivest Cipher 4. A stream cipher designed by Rivest for RSA data security, now RSA security. It is a variable key-size stream cipher with byte-oriented operations. The algorithm is based on a random permutation.

RH	Relative Humidity
RISC	Reduced Instruction Set Computer
RMCP+	Remote Management Control Protocol
ROM	read-only memory
ROMB	RAID on Motherboard.
RPM	Revolutions Per Minute
RT	Runtime. Component of Intel® Platform Innovation Framework for EFI architecture.
RTC	real-time clock
RTM	Root of Trust Measurement
RTR	Root of Trust Reporting
RTS	Root of Trust Storage
SAF-TE	SCSI Accessed Fault-tolerant Enclosure specification. Describes a set of SCSI commands whereby drive fault status can be sent to an enclosure to present that fault information with external indicators, such as fault lights. Other commands are provided so certain management information about the enclosure, such as temperature, voltage, number of drive bays, power status, etc., can be retrieved.
SAS	Serial Attach SCSI
SATA	Serial-ATA.
SBE	Memory ECC Single-Bit Error.
SCA	Single connector Attachment
SCI	System Control Interrupt. A system interrupt used by hardware to notify the operating system of ACPI events.
SCL	Serial Clock
SDA	Serial Data
SDINT	System Diagnostic Interrupt
SDR	Sensor Data Record.
SDRAM	Synchronous DRAM. A type of DRAM that uses an external clock and supports faster memory access than prior DRAM memory types.
SE	Single Ended
SEC	Security. Component of Intel® Platform Innovation Framework for EFI architecture.
SEEPROM	Serial Electrically Erasable Programmable Read Only Memory
SEL	System Event Log.
SEP	SAF-TE Enclosure Processor
SGPIO	Serial general-purpose input / output
SHA1	Secure Hash Algorithm 1
SIO	
SIO3	Server I/O Chip (3rd generation).
SIOH	Server I/O Hub
SMB	Server Management Bus
SMBASE	SMRAM Base Address.
SMBus	System Management Bus. Mastered by a system management controller to read and write configuration registers. Signaling and protocol are loosely based on I2C.
SMI	System Management Interrupt. A special type of interrupt to signal attention from the SMI handler.
SMM	System Management Module
SMP	Symmetric Multiprocessing

SMRAM	System Management RAM
SMS	server management software. Designed to run under the OS.
SNMP	Simple Network Management Protocol
SOL	Serial Over LAN.
SPD	Serial Presence Detect. A 2-signal serial bus used to read and write control registers in SDRAM devices via the SMBus protocol.
SPI	Serial Peripheral Interface
SPT	straight pass-through
SRAM	Serial Random Access Memory
SRK	Storage Root Key
SRTM	Static Root of Trust Measurement
SSI	Server System Infrastructure
TCC	Thermal Control Circuit
TCG	Trusted Computing Group
TDP	Thermal Design Power.
TIM	Thermal Interface Material. Replaces thermal grease between the HIS and the heatsink.
TM	Thermal Monitor. This feature is also known as Automatic Clock Control.
TM2	Thermal Monitor 2
TOE	TCP/IP Offload Engine. A hardware component of Intel Crystal Beach technology.
TPM	Trusted Platform Management or Module
TSS	TCG Software Stack
TTL	Transistor-Transistor Logic
UART	Universal asynchronous receiver and transmitter
UDP	User Datagram Protocol
UE	Uncorrectable Error
UGA	Ultra Graphics Array
UHCI	Universal Host Controller Interface
USB	Universal Serial Bus, a standard serial expansion bus meant for connecting peripherals.
UUID	Universally Unique Identifier. See also GUID.
UV	Under-Voltage
VAC	Alternating current (AC) voltage
VCC	Voltage Controlled Current
VCCI	Voluntary Control Council for Interference by Information Technology Equipment
VGA	Video Graphics Array
VID	Voltage ID
VLAN	Virtual local area network
VMT	Virtual Machine Extensions
VRD	Voltage Regulator Down. Voltage regulation circuitry built into the Baseboard.
VRM	Voltage Regulation Module. Voltage regulation circuitry that can be added to the Baseboard on a plug-in module.
VSB	Voltage StandBy
WfM	Wired For Management
WOL	Wake On LAN.

XD bit	Execute Disable bit. An IA-32 processor that supports the Execute Disable Bit feature can prevent data pages from being used by malicious software to execute code.
XDP	eXtended Debug Port.
ZIF	Zero Insertion Force

This appendix contains important terms used in the preceding chapters. For ease of use, numeric entries are listed first (e.g., "82460GX") with alpha entries following (e.g., "AGP 4x").

Word / Acronym	Definition
82460GX	The chipset used in the server board.
ACPI	TBD
ADC	Analog to Digital Converter.
AGP 4X	High-speed graphics port, a component in the 82460GX chipset.
AP	Application Processor.
API	Application Programming Interface.
APIC	Intel Advanced Programmable Interrupt Controller for Symmetric Multi-processor (SMP) systems.