BladeCenter packaging, power, and cooling

M. J. Crippen
R. K. Alo
D. Champion
R. M. Clemo
C. M. Grosser
N. J. Gruendler
M. S. Mansuria
J. A. Matteson
M. S. Miller
B. A. Trumbo

This paper addresses the packaging, power, and cooling of the IBM eServer™ BladeCenter® compact server infrastructure consisting of 14 servers in a 7U (1U = 44.45 mm) vertical space for installation in an industry-standard rack. A typical rack is 42U tall. Therefore, six BladeCenter systems will fit in a rack, for a total of 84 servers. The density of a BladeCenter system (servers/U) is double that of previous 1U rack-optimized servers. To build such a dense server system required overcoming a multitude of challenges in packaging, power, and cooling design. Our approach to these challenges is described, but in the broader context of not only increasing the density, but setting a new server standard for a highly available redundant infrastructure with integrated systems management and network switching that uses less power and is simple to maintain on site or remotely, even by nonspecialized personnel. Server processors, memory, storage, and input/output devices were combined in a single compact server unit called a processor blade, while the support infrastructure, such as systems management, network connectivity, optical media, power, and cooling, was consolidated in a single structure and shared among many servers. The result is a package architecture that lends itself well to standardization. Custom server blades and input/output devices may be designed in accordance with BladeCenter base specifications and be effectively integrated into the blade server system.

Introduction

Innovation in system-level packaging, power, and cooling technology over the past several years has proven to be one of the most significant competitive differentiators for IBM Intel-processor-based servers. The IBM eServer* BladeCenter* system [1] is among the leaders in the industry in volumetric density, power density, cooling capability, power management, cooling management, and ease of use. Blade servers make it possible to increase server packaging density and eliminate some of the problems inherent in 1U servers. These problems include excessive cables, difficulty in servicing, complex connections for input/output (I/O) devices and systems management, and, when many 1U servers are assembled in a rack, the repetitive use of hardware components such as optical media, fans, and power supplies. In the 7U, 14-processor-blade BladeCenter system, the density (servers/U) is double that of previous 1U rack-optimized servers and provides a highly available, redundant infrastructure with integrated systems management and network switching, all while using less power [2].

When the power, packaging, and cooling design requirements for the IBM blade system were initially defined, they were focused primarily on improving the deficiencies in the 1U server packaging architecture. However, it was quickly determined that with a processor blade form factor, much more could be realized than simply reducing hardware complexity and content, and the requirements list expanded to reflect a server blade system with improved reliability, availability, serviceability (RAS), network connectivity, ease of use, and efficient systems management.

Many key challenges were faced during the design of the BladeCenter architecture, and this resulted in innovative solutions using state-of-the-art technologies. Some of the challenges were doubling the volumetric packaging density (servers/U of rack space used) while supporting server-level Intel processors, designing a highly available server support infrastructure to support multiple servers reliably with various operating systems, and devising a system that could keep up with rapid

0018-8646/05/$5.00 © 2005 IBM

advances in processor, memory, and server I/O technologies through the planned life of the product.

A fundamental requirement from the onset was the need to support Intel's enterprise-level processors so that the BladeCenter system could deliver the performance customers demanded. Low-powered processor families were briefly considered but were discounted because they would not provide the performance IBM customers expect. It was also a requirement to build an infrastructure that would last through several processor life cycles. Critical to meeting this challenge was the selection of a blade pitch and size (the distance from one blade to another and the size of the main printed circuit board) that would result in a high concentration of servers that offered all necessary function, but still left sufficient space within the blade to package, power, and cool the components. The architecture had to be scalable with respect to the number of processors, memory, and I/O devices in the blade [3]. While the primary focus was on dual-processor blades, single-processor blades and the ability to expand to four-way servers was an absolute requirement. This was accomplished through the use of multiwidth blades. If more function is needed in a blade (such as adding processors), the blade width can be doubled or tripled to provide more room.

Developing a product with high RAS was of utmost importance. It was imperative that the BladeCenter system be highly reliable and available. Minimizing single points of failure that would cause a system to shut down was essential. Therefore, redundant subsystems and very reliable components, such as hard drives, were selected. For those components for which redundancy was not practical, "hot-swapping" without affecting primary system function was employed. This led to the use of redundant power supplies, blowers, networking switches, and management modules. The ability to support redundant system power feeds was also a requirement, leading to the use of $N + N$ power: the use of an even number of power supplies with the option of half being powered by one building power source and half by a second, redundant source. The power supplies must be arranged such that one of the system power sources can fail without interrupting server function. The midplane, which is used for communications among virtually all system components, had to have redundant signals and connectors. The optical media devices are not redundant; however, they can be removed and replaced while the system is operating. In all cases, innovative techniques were used so that quick, intuitive removal and replacement of devices was possible.

The system had to be compact, self-contained in a single chassis, easy to use, and quick to service. This led to using only front and rear service access and easy-swap devices that did not require the use of tools. A goal was established of a 30-second service time limit for removal and replacement of a field-replaceable unit (FRU) and 30 minutes for a midplane. In addition, the ability to expand processor blade function, for example by adding additional hard drives or personal computer interconnect (PCI) [4] cards without the use of tools and with minimal parts on the floor, was required. It was also essential to design the system such that no damage would occur if a processor blade or module were misplugged. Customer information lights on the front and rear of the chassis were required. Since BladeCenter electronic components are accessible to the operator while the system is powered on, precautions were needed to ensure that the product was safe. This included, but was not limited to, ensuring that the operator is never exposed to a hazardous energy of greater than 240 volt-amps. This was achieved by using an innovative interlock device with retractable plungers to prevent installation of the blade in the chassis without the proper blade enclosure cover in place.

Cooling and power subsystem architecture met the challenges through many innovations: multipath serial cooling of components, the use of vapor-chamber processor heat sinks with top and bottom fins, and high-performance backward-curved impeller blowers. It resulted in the use of one of the highest-density power supplies (watts per unit volume) used in IBM products to date. Many innovations in power and cooling system controls have greatly helped to minimize the cost and space occupied while providing very high-performance subsystems. Power system *oversubscription* uses the available capacity of the power system during normal operation while still providing redundant power in a power-supply failure mode. System power demand reduction is used during power-supply and blower failure modes. Blower speed control is used to provide quiet-mode operation when needed.

Satisfying the requirements defined for BladeCenter systems proved to be very challenging. The resulting technical solutions have made IBM the leader in the server blade market [5].

Packaging

Architecture

The development of the packaging architecture was the result of taking 1U-rack packages to the next level. For example, to increase server density, it was proposed that two servers that shared both cooling and power resources be placed in a single 1U package. However, the result was not a significant improvement over the existing 1U server. In a full rack of 42 1U servers, more than 300 cables would have to be routed in the rear portion of a rack (**Table 1**). Because of space limitations within a 1U server, two servers would share a nonredundant infrastructure,

Table 1 Hardware comparison of 14 1U systems with a single 14-processor-blade BladeCenter system.

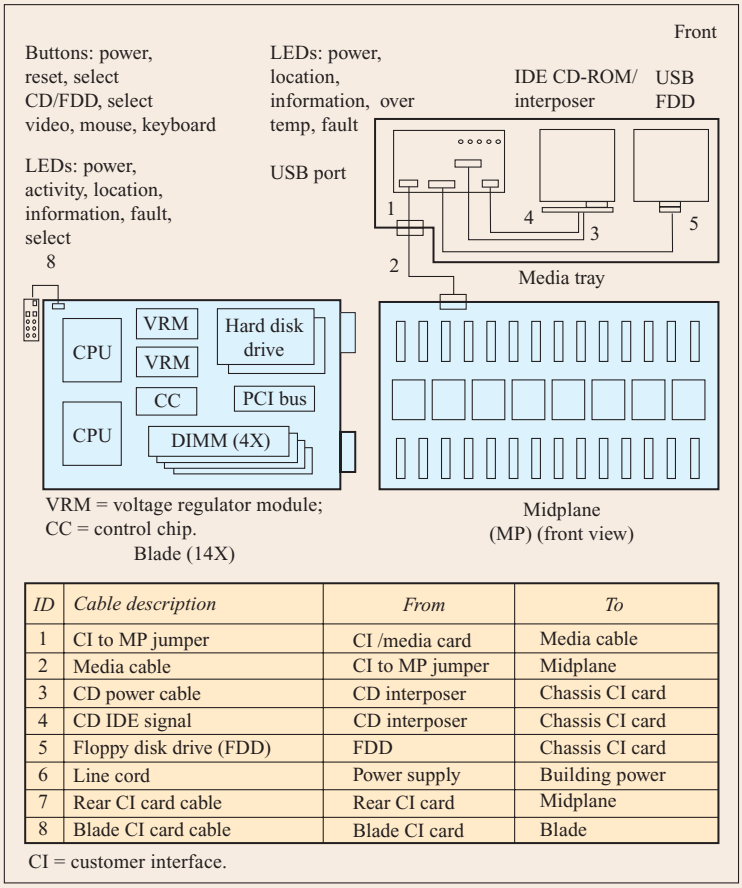
1U server (quantity)	*Functional description*	*BladeCenter systems (quantity)*
14	Diskette drive	1
14	CD-ROM	1
112	Fans	2
28	Power supplies	4
28	Line cords	4
14	Keyboard, video, and mouse (KVM) cables	1
28/56	Ethernet cables (two module/four module)	8/16
28/0	Fibre Channel cables (two module/zero module)	4/0
13	Systems management cables (one or two BladeCenter systems)	0/1
394 lb	Weight	240 lb

resulting in poor server availability if a component failed. Service would not be significantly improved because of all the cables connected to the rear of the chassis.

The obvious requirement was to reduce the cabling complexity. The best solution was to create an enclosure containing a midplane [6] that could handle the vast majority of the interconnect function. The next obstacle was to define the size of the processor blade and chassis. Some system developers took the quickest time-to-market design point by stripping the power supplies out of the 1U server, making it narrower, and sliding it into an enclosure like a book onto a shelf. This provided a time-to-market advantage and low development cost, but it did not go far enough to simplify the setup and maintenance of the system. For instance, one competitor's blade product contained many fans and a clumsy connection to a separate external bulk power source, both resulting in excess internal and external cables.

In parallel with BladeCenter development, many IBM engineers were involved in the creation of the InfiniBand** industry standard [7], and in particular the InfiniBand blade packaging standard that was being developed. Since the concept for the IBM blade server and the InfiniBand packaging philosophy had much in common, the decision was made to expand on the InfiniBand design to create a blade server and associated *modules*—subcomponents within the BladeCenter chassis. The resulting design is the 7U IBM BladeCenter chassis, which supports 14 processor blade servers in the front and associated modules in the rear. (Detailed system illustrations and the many components discussed in this paper can be found in a companion paper by D. M. Desai et al. [1].) Both blades and modules plug into a common printed circuit board, called a *midplane*. The processor with its blade server functions must be configured and treated as an autonomous server. Typical blade types include processor blades, hard drive blade storage expansion (BSE), and the PCI expansion unit (PEU). Modules make up the infrastructure that is shared with processor blades and are installed in the rear of the chassis. Modules are arranged so that those with the most cables are aligned with the sides of the chassis, minimizing trapping of modules behind cables. This is especially important for large modules, such as the blowers. Module functions include systems management, Ethernet network switches, Fibre Channel switches, I/O optical passthrough, blowers, power supplies, and acoustic attenuation. A media tray located in the top front of the chassis contains customer interface indicator light-emitting diodes (LEDs), a Universal Serial Bus (USB) port, a compact disk read-only memory (CD-ROM) or digital video disk (DVD), and a diskette drive.

Sharing resources and establishing common interfaces to simplify the customer's experience is the cornerstone of the packaging architecture. In Table 1, it can be seen that the complexity of a multiserver system is greatly reduced by using a BladeCenter system. As startling as this may seem, it becomes much more dramatic once viewed at the rack level. The introduction of networking hardware into the system infrastructure and sharing it among the many servers in the chassis is very helpful in reducing complexity. Another important architectural accomplishment is the association of volume with server scaling. The system allows for multiwidth processor blades, such as the single-wide (30-mm pitch) and double-wide (60-mm pitch) processor blades that can piggyback with expansion blades. Depending on the number of blade widths a server occupies, the amounts of cooling, power, and I/O are scaled proportionately. A double-wide processor blade can make use of double the power, cooling, I/O, and so on.



ID	Cable description	From	To
1	CI to MP jumper	CI /media card	Media cable
2	Media cable	CI to MP jumper	Midplane
3	CD power cable	CD interposer	Chassis CI card
4	CD IDE signal	CD interposer	Chassis CI card
5	Floppy disk drive (FDD)	FDD	Chassis CI card
6	Line cord	Power supply	Building power
7	Rear CI card cable	Rear CI card	Midplane
8	Blade CI card cable	Blade CI card	Blade

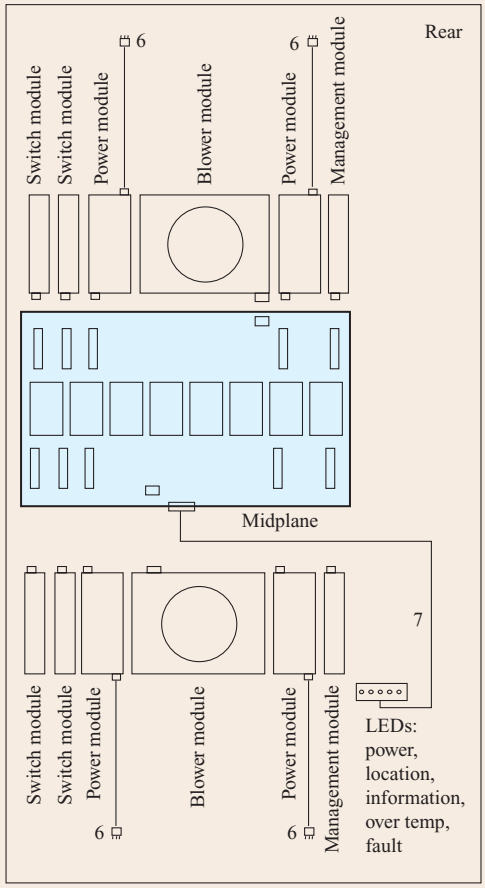


Figure 1

BladeCenter chassis system.

As a result of combining technical requirements with the need to simplify, the chassis has become physically symmetrical right to left and top to bottom. The midplane interconnect is basically two symmetrical midplanes that are physically interconnected, but, for purposes of redundancy, can essentially operate independently. The modules that plug into the top portion of the midplane provide redundancy for the modules that plug into the bottom. The processor blades provide connections to both the top and bottom modules. One of the most consistent customer comments is this: *It puts all of my eggs in one basket; prove to me that if an egg or two is broken, the whole basket isn't taken with it*. The physical partitioning created by symmetry ensures that one failed area cannot propagate to its redundant twin. Symmetry also makes it possible to cool all of the components. The symmetrical air paths provide even and balanced cooling to the processor blades and the modules behind them. Finally, the symmetry provides the customer and service personnel with an intuitive map that helps them understand what goes on and where.

Figure 1 shows the layout of a typical BladeCenter system, illustrating all major components and associated interconnects. The system symmetry, as discussed above, can be seen clearly. Although the diagram shows what appear to be two midplanes, it is actually a front and a rear view of the single midplane.

Human factors

System availability and serviceability is paramount in a multiserver system. Many innovative features were incorporated to make the system easy to use. These features enhance the user experience and reduce assembly, installation service time, and cost.

Ease of service starts with front and rear service access of the system, rather than front, top, and rear, which is normal for rack-mounted servers. Not requiring top

service access greatly reduces the complexity of the chassis rack mounting hardware. It also removes the need to slide the server out of the rack during hot-swap service operations, resulting in a more reliable system and the elimination of an articulated cable management arm to control cables during service [8].

The modular design allows server blades, switch modules, power modules, blower modules, management modules, and the media tray to be removed independently and replaced quickly without the use of tools. Since all major components plug to a single midplane, there are no buried components, as there are in many standard servers, and no internal cabling that can be daunting to an untrained customer.

Information LEDs are used on the chassis, processor blades, and modules in conjunction with IBM light-path diagnostics to provide system information and to guide the user during service operations. A primary fault indicator, intended to be visible from up to 20 feet, is located in a bank of five information LEDs on the media tray at the front of the chassis. This lights up when any fault occurs in the hardware. Secondary fault indicators are located on every processor blade and module. A unique design challenge associated with this was how to light LEDs located inside the processor blade after it has been removed and power has been removed. This was accomplished by using a charged capacitor. Once the blade is removed and opened, a button can be pressed that lights the LED of the failed component through the use of the capacitor charge.

Another innovative feature is a removable system service card located in a bay at the front of the chassis. This card explains the meaning of all LEDs and has instructions on it showing users how to remove and replace all modules and processor blades. The processor blades have a label on them that shows how to remove and replace the FRUs inside the blade.

Industrial design

The BladeCenter industrial design is critical to the function, value, and appearance of the product and significantly enhances the user experience. The product design is honest, open, and reliable, which reveals rather than conceals its purpose. It remains true to the character of the metal material and visually emphasizes functional elements such as the processor blade handles, the customer panel, and the media, aiding in ease of service. The adjective *reliable* means simple, well made, and durable. The design uses simple geometric forms to consolidate, formalize, and propagate the design. Simplifying complexity through simple forms instills a common-sense logic and consistency in every part. The raven black color with a satin finish was chosen to reflect lasting quality and support the product physical characteristics and design. The sculpted blade and module handles draw attention and are enhanced by curved details that communicate their purpose. The blinking icons, visible through the smoked-glass customer panel door of the processor blade, form a strong contrast with the solid black.

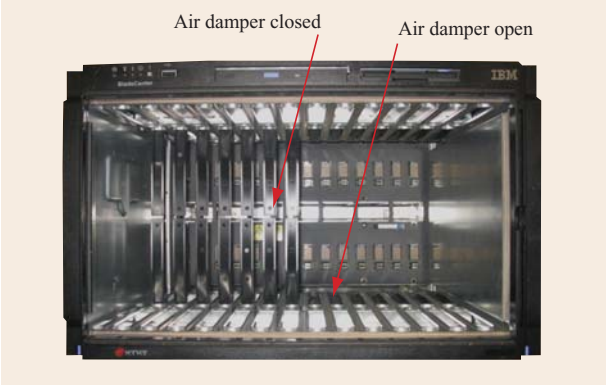


Figure 2

BladeCenter chassis showing blade bays and air dampers.

Chassis

The BladeCenter chassis is a 7U-tall, 28-in.-deep rack with optimized monocoque construction. It is designed so that virtually all components can be removed from the chassis within a few minutes.

The BladeCenter chassis has 14 hot-swap processor blade bays (**Figure 2**). Each processor blade bay contains a set of air dampers, one on the bottom and one on the top of the 30-mm-wide processor blade bay. The dampers fold out of the way as the processor blade is installed, and they close when the processor blade is removed, thus preventing air from short-circuiting, which would reduce airflow to other processor blades and cause them to overheat. Because the dampers do not provide electromagnetic interference (EMI) protection, a processor blade bay must not be left open without a processor blade or blade filler installed.

At the rear, there are 12 module bays: four switch, two management, four power, and two blower module bays. Switch bays 1 and 2 are for Ethernet or I/O optical passthrough modules. Bays 3 and 4 are for modules based on the function of the processor blade I/O expansion adapters. The two blower module bays have air dampers that close when a blower is removed, thus preventing air recirculation during blower service and maintaining proper system cooling.

The media tray (**Figure 3**) is a sheet metal structure that supports a customer interface (CI) card, CD-ROM or

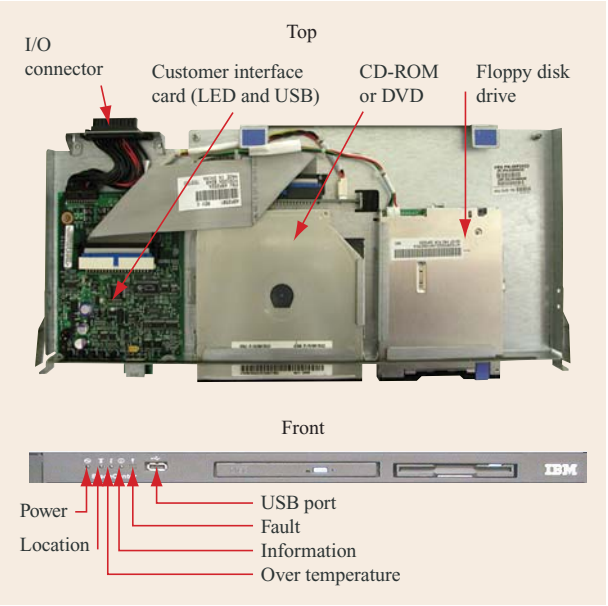


Figure 3

Media tray.

DVD, and diskette drive. The optical devices are shared among the 14 servers. A four-port USB hub on the CI card converts the integrated drive electronics (IDE) CD-ROM to USB and provides USB communication throughout the chassis. The I/O connector on the media tray hot-plugs to a connector on the media cable, located in the media cable tray, which in turn connects to the midplane.

The switch, power, cooling (SPC) chassis (**Figure 4**) supports the midplane, into which all processor blades, modules, and the media tray plug. It installs into the rear of the chassis with the use of cam levers and captive screws. Captive screws that require a tool to tighten or loosen them are used in order to ensure that the chassis is not inadvertently removed, which would cause up to 14 servers to shut down. All processor blades must be unplugged from the midplane prior to removing the SPC chassis. During removal of the SPC chassis, release locks automatically lock the SPC chassis in place halfway out of the BladeCenter chassis as a safety feature. The release locks must be deactivated by the operator prior to final removal of the SPC chassis, ensuring that the operator has a firm grip on the chassis prior to removal.

Air flows above and below the midplane and through the seven openings in the center, as discussed below in the airflow path section. Midplane 240 volt-amps hazardous energy protection is provided by an insulator and mechanical chassis shielding.

Processor blade

A server is a subset of an industry-standard server that is implemented as a thin, pluggable electronics unit with a protective cover. A processor blade is approximately 245 mm tall, 446 mm deep, and 29 mm wide. A double-wide processor blade is 59 mm wide, and a triple-wide is 89 mm wide.

The processor blade (**Figure 5**) consists of one or more printed circuit boards, depending upon its function. A typical processor blade contains two processors, four dual inline memory modules (DIMMs), two small-form-factor (SFF) hard drives, and an SFF-size I/O expansion adapter.

The processors are located at the front of the processor blade so that they receive fresh air for maximum cooling. The memory connectors are arranged to allow the memory DIMMs to be positioned 25 degrees from the plane of the board so that industry-standard DIMMs may be used in this compact processor blade. The

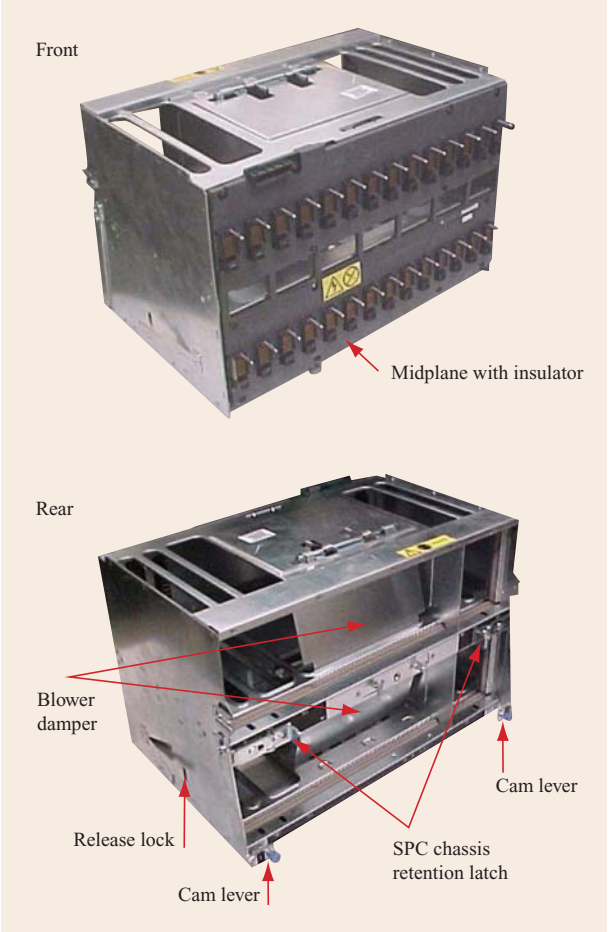


Figure 4

Switch, power, and cooling (SPC) chassis.

SFF hard drives may be IDE or Small Computer System Interface (SCSI) devices, depending on processor blade function, and they can be installed in or removed from the hard drive trays without the use of tools. Some processor blades support two SFF hard drives or one drive and one standard I/O expansion adapter, which is mounted in the area where the rear hard drive is located. Others simultaneously support two SFF hard drives and an SFF I/O expansion adapter, or one hard drive and one standard-size I/O expansion adapter. The SFF I/O expansion adapter plugs into the same connectors used by a standard-size card, but it extends into the area behind the DIMMs, allowing a second drive to be present. Both form-factor cards mount without tools into a tray that supports and retains the card during shipping.

The processor blade enclosure includes a tool-less removable cover and front bezel, both of which may be removed by activating release latches identified by blue touch points on the side of the blade. The bezel contains two latch handles on the front, used to retain the processor blade in the chassis. There are two different processor blade cam lever designs, one for the BladeCenter chassis and a low-profile handle used in the BladeCenter T system, intended for the telecommunications market [9].

The customer panel is located in the top section of the processor blade bezel. It provides considerable function in a very small space owing to the innovative use of buttons and LEDs mounted on a small flexible circuit board. The panel comprises various functions depending on the processor blade. The main interface features include the processor blade power button, nonmaskable interrupt reset button, and one or two select buttons. LEDs provide function and error information. One select button is used to access control of the keyboard, video, and mouse (KVM) devices that can be attached to the management module, located at the rear of the chassis. The other select button is used for gaining control of the diskette drive and CD-ROM or DVD.

A major challenge associated with providing easy service access to the processor blades was to provide tool-less access to components inside the blade while providing 240 volt-amps hazardous energy protection. This was achieved with the use of three primary features. First, the enclosure has two retractable plungers that prevent the processor blade from being installed in the chassis without a top cover or expansion blade attached. (If a processor blade could be installed in a chassis without a cover and the adjacent processor blade was not installed, it would be possible to access 240-volt-amp components.) Second, there are 240-volt-amp keep-out areas defined on the printed circuit board where no 240-volt-amp components are allowed. Third, there is a label on the cover explaining that the cover must be installed before the blade is installed in the chassis.

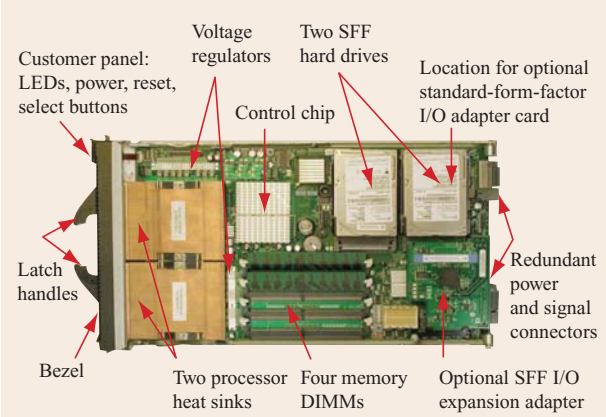


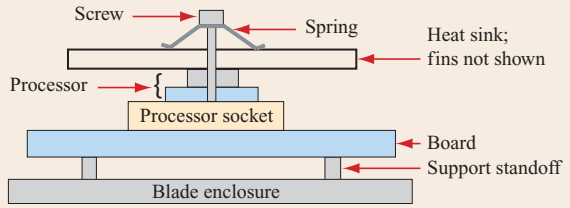
Figure 5

BladeCenter processor blade.

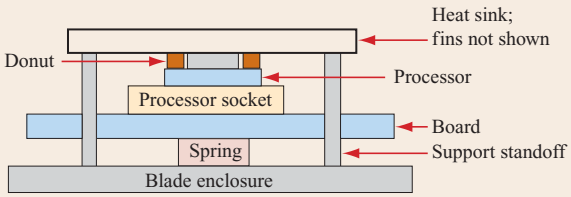
Processor heat-sink technology and mounting techniques are dependent on the processor used in the blade. Typically, for Intel-based processors the heat-sink is a low-profile, folded-fin design that has fins on the top and bottom sides of the heat-sink base plate. On the lowest-powered processors, the base is solid copper. However, most processors require the use of a vapor-chamber-base heat sink to enhance the spreading of heat. In general, the fins are made from aluminum, but higher-powered processors require copper fins. A thermal interface material is used to conduct heat from the processor to the heat sink. This material requires a compression force between the heat sink and processor to ensure proper conduction.

In the IBM HS20 Intel-based processor blade, the compressive static force applied between the heat sink and the processor is approximately 225 N, or 50 lbf (pound force). There is a die-cast retention module that surrounds the processor sockets and is attached to the board with the mounting screws used to retain the board inside the enclosure [**Figure 6(a)**]. For this type of processor, heat sinks were designed with two captive screws that have low-profile leaf springs that compress when tightened. The heat-sink fasteners screw into the retention module to help distribute the force onto the processor.

In the IBM JS20 processor blade, two IBM PowerPC* 970 processors are used. These processors are open die, which means that the silicon chip is exposed rather than being protected by a cover, as in many processors. Open-die processors are fragile and more susceptible to damage than covered processors. When a heat sink is installed on a processor, a compression force must be applied between



(a) Heat sink attaches to retention module.
The spring pushes the heat sink against the processor.



(b) Heat sink attaches to the blade enclosure.
The spring pushes the board, socket, and processor against the heat sink.

Figure 6

Processor blade heat-sink mounting techniques.

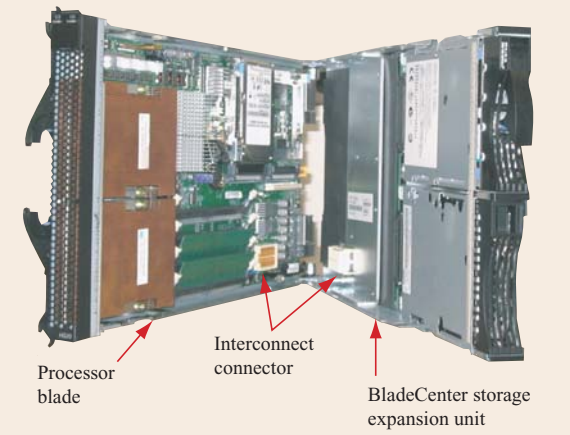


Figure 7

Processor blade with BladeCenter storage expansion unit (BSE).

the heat sink and the processor to ensure that the thermal interface material placed between the two will function properly. This posed the challenge of how to apply enough force to ensure a good thermal interface but not so much that the processor would be damaged. Just the weight of the heat sink could potentially damage the processor during shock loading. To remove the heat-sink mass from the die during attachment and operation, the heat sink was designed to screw directly to the blade enclosure [**Figure 6(b)**]. A square foam pad is located under the processor between the processor blade chassis and the board. It acts as a spring and provides an upward force on the processor. After assembly, the processor package and planar are sandwiched between the heat-sink surface and the foam spring under the planar. Another component of the heat-sink retention is the protective *donut*, which adheres directly to the processor substrate. The donut frame has a hole through which the die protrudes. The frame walls sit slightly under the die top surface to minimize the rocking angle of the heat sink during assembly. This protects the die edges from being damaged.

There are several I/O expansion adapter options for the processor blade, such as Fibre Channel and Ethernet. These cards provide redundant ports from each processor blade to switch module positions 3 and 4. Once a specific I/O expansion adapter is installed in one processor blade, all other processor blades in a given chassis must also use this card or no card at all [10].

The function of a processor blade may be expanded by adding an expansion blade. An expansion blade is an add-on blade assembly that communicates with the base processor blade to which it is attached and, in some cases, with the BladeCenter midplane. An expansion blade can be added by simply removing the top processor blade cover and the processor blade bus terminator card, if present, and then plugging the expansion blade into the base processor blade. The retention of the expansion blade is the same tool-less method used to retain the processor blade cover. In all cases, most of the power for the expansion blade is obtained from the midplane. Expansion options include the BSE, which supports two 3.5-in. hot-swap SCSI hard drives, and a PEU, which supports two full-size PCI cards. The add-on structure increases the width of the processor blade assembly by 30 mm. **Figure 7** shows a processor blade with a BSE in a partially open position.

Modules

The shared BladeCenter infrastructure is provided by modules that plug into the rear of the chassis and connect to the midplane. The modules are sized such that two modules will fit in the vertical height of one processor blade. This vertical scaling allows room for redundant modules within the chassis. Like processor blades, most modules are based on a standard width of 29 mm and are on a pitch of 30 mm; they can be single-wide or double-wide. Cooling for modules is transverse, exiting into the center air plenum of the chassis. The primary modules are the switch module (either Ethernet or Fibre Channel), the passthrough module (either Ethernet or Fibre Channel), the management module, the power module, the blower

module, the acoustic attenuation module, and the filler module.

The switch modules are single-wide form factor. They provide either Ethernet or Fibre Channel network switch functions. On one end of the module are signal, power, and management connections to the chassis midplane; the other end provides customer interfaces such as cable connections, information LEDs, and the extraction lever.

The management module is very similar physically to a switch module. It consists of the service processor and the KVM switch. The management module interfaces with all of the processor blades, the switch modules, the power modules, and the blower modules. The external interfaces of the management module include connections to external KVM, a dedicated management Ethernet port, information LEDs, and the extraction lever.

The power module is double-wide. It provides up to 2,000 W of 12-VDC power to the chassis midplane and associated signals. The external interfaces are the high-line power-cord, ac and dc power-good LEDs, and the extraction lever.

The blower module form factor is defined by the cooling requirements. It consists of an encapsulated backward-curved impeller blower that draws air from a plenum area in the center of the rear portion of the chassis. Covering the air exit of the blower module is a damper that is drawn to a closed position by system suction if the blower fails. A fault LED indicates whether a blower module has failed.

The acoustic attenuation module is an option that reduces noise emission by approximately 5 dB. It fits over the rear of the system centered over the blower modules. The exhaust air is ducted through the acoustic attenuation module, the inside of which is lined with specially shaped acoustic foam that absorbs the noise, directs airflow, and reflects the noise to reduce that which exits the rack enclosure. It can be removed from the chassis by a simple turn of a threaded screw with a T-handle. A light pipe is located on the bottom of the module to transmit light-path function from LEDs located at the rear of the chassis. Clearance is provided so that cables can be routed to either side of the acoustic attenuation module.

Filler modules are installed in module bays in the chassis in the place of absent switch, power, and management modules. Their function is to maintain the EMI integrity of the system and to maintain the airflow balance required to evenly cool the components in the chassis.

Power

The power system architecture provides redundant, hot-swappable power to each of the loads (processor blades and modules). Up to four ac–dc power-supply modules can be installed in the chassis, converting a high-line ac input voltage to a 12-VDC (+12.2 VDC nominal) output, providing up to 164 amps of current (2,000 watts). The 12 VDC is distributed throughout the system and is locally transformed by point-of-load dc–dc converters on the processor blades and modules. A 12-VDC distribution bus rather than a higher voltage, such as 24 VDC or 48 VDC, was chosen to enable higher-efficiency dc–dc converter technologies to be used. A −48-VDC bus was considered but negated owing to several factors: the limited space available on the blade and module cards, the cost of conversion from 48 VDC to low voltage, and the need for 12 VDC on the cards.

The power module is designed to convert power from a single-phase (three-wire) external ac input source to 12 VDC for distribution within the system. The ac input voltage range is 200 VAC_{RMS} to 240 VAC_{RMS} nominal at 50 Hz or 60 Hz and is designed to meet worldwide safety, emissions, and other regulatory requirements. The power module is designed to meet the following functional requirements:

- A near-unity power factor (ac current in phase with ac voltage) with a minimum of 0.98 to maximize customer ac power distribution capacity.
- Parallel operation and, in the event of failure, one power module that continues to power the system loads for a given power domain.
- Hot-swap exchange facilitated using ORing FETS on the other modules. By distributing the ORing throughout the system, I^2R losses typically associated with power supplies are reduced.
- Means to communicate with the systems-management module through Inter-Integrated Circuit (I^2C) Bus Protocol using a microcontroller for power module control and status information.

The power module also features a differential remote voltage sense; specified limits for the 12-VDC output are ±3% measured at the remote sense points. Additionally, it offers active current sharing of the output load between power supplies in redundant operation.

The ac–dc power conversion challenge of the chassis was to package the power-supply modules in approximately half the volume (twice the power density in $W/in.^3$) of previous xSeries* products. The power density constraints of a product are fueled mainly by cooling and airflow limitations in the system chassis. The layout of the system and airflow paths, shown in the cooling and acoustics section below, allows the blower modules to provide a nearly optimal airflow for the power modules. The power modules were developed in an iterative process that kept the power system cost to a minimum. Processor

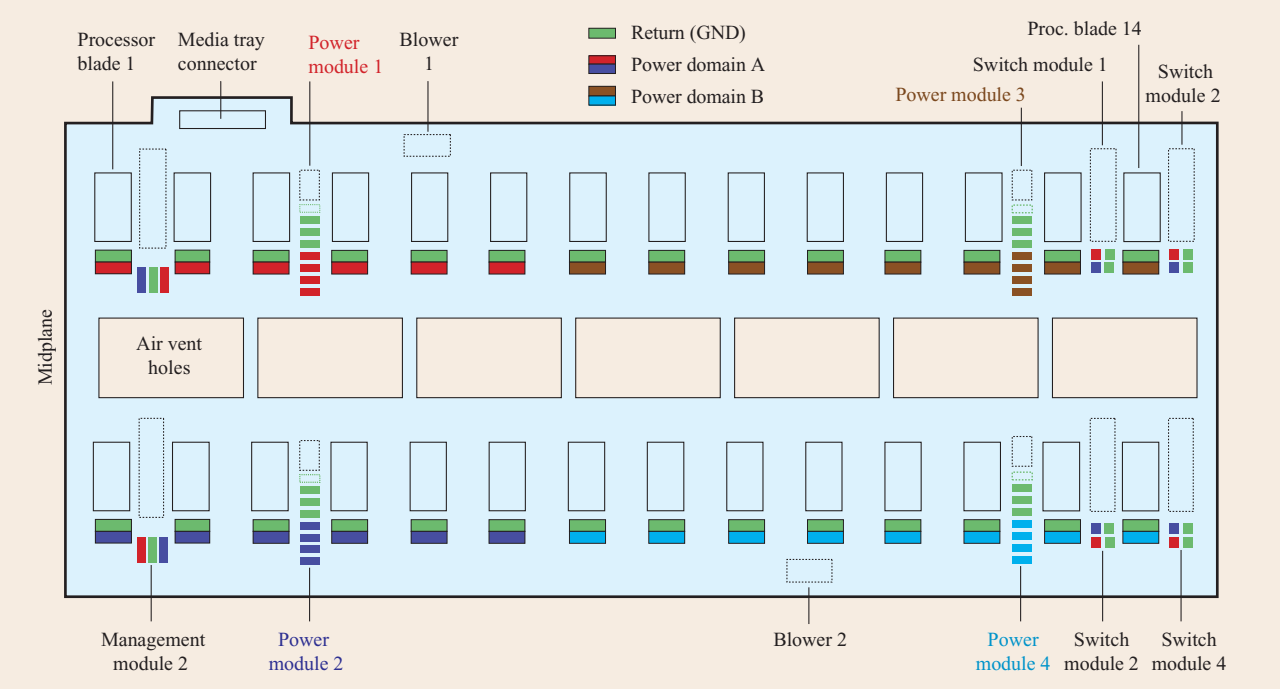


Figure 8

Midplane power connections.

blades released during the first year of production required less power than current processor blades. A 1,200-W power-supply module was initially released. With processor blade power increasing over time, the power supplies were redesigned to achieve the current 2,000-W output at 20 W/in.^{3} A technical feasibility study was conducted and determined that 2,500 W (25 W/in.^{3}) could be achieved in this form factor. However, the airflow coming to the power modules is preheated by the processor blades. The more power delivered to the processor blades, the higher will be the preheating of the air that provides cooling to the power modules. The maximum allowable ambient temperature at the inlet to the power module is currently 52°C with a 2,000-W load. Higher output power would cause the inlet air temperature to exceed the cooling capacity of the power module with the given airflow. The design goals also specified an increase of 5% in ac–dc power module efficiency over power module designs used in other xSeries servers. There were two major benefits in doing so: More output power is delivered to the load for a given input power, and less heat is generated within the supply during power conversion, which helped in achieving the power density.

A further architectural feature was the division of the chassis into two independent power domains, A and B, each powered by one of the redundant pairs of power modules. The power module locations can be seen in **Figure 8**. The power modules installed in bays 1 and 2 comprise power domain A, which powers processor blade bays 1 through 6, blowers 1 and 2, switch modules 1 through 4, management modules 1 and 2, the media tray, and the front and rear operator panels. The power modules installed in bays 3 and 4 comprise power domain B, which powers processor blade bays 7 through 14. Domain B power modules are not installed if fewer than seven processor blades are populated in the chassis.

Figure 8 represents the power distribution on the midplane. Each processor blade or module power connector has been color-coded to the associated power module connector. Two power modules of the same wattage/amperage must be installed in each domain. If a single power module fails, the domain will continue to supply power from the remaining power module. The power distribution buses (amount of copper) on the midplane were designed for acceptable power distribution losses when operating in nonredundant mode (100% output from one power module). Under nominal redundant operating conditions, this allows for one-

quarter the loss (I^2R) with each power supply operating at a maximum of 50% output current.

Processor blade power

Each processor blade has two separate power connectors which accept input power from two isolated 12-VDC inputs on the system midplane. The processor blades provide diode (or MOSFET) ORing which combines two 12-VDC power inputs to create one redundant power source for the blade. This way, if one of the power supplies fails, loses ac input voltage, or is replaced, the remaining source keeps the card powered. Auxiliary (or standby) power is taken from the ORed voltage. Main 12-VDC power to the processor blade is taken from the ORed voltage and passes through a MOSFET switch controlled by the system firmware. The processor blade input power circuit is depicted in **Figure 9**. The input power circuit is designed to provide the following functions:

- Hot-plugging of the card into an energized midplane.
- Auxiliary (continuous) voltage to onboard service processor or card control circuitry.
- ORing of the two 12-VDC input voltages (allows continued power to the card if one of the inputs is removed because of a fault).
- Fast response of the circuit to reverse currents that indicate a shorted 12-VDC input source. This minimizes the effect on the redundant voltage source by not allowing a fault to load the remaining input source, causing a voltage drop to the blade or an overload on the remaining power supply.
- Main power to the card, which is controlled by an *enable* signal. The on/off logic is set by the management module [11] through the onboard baseboard management controller (BMC). In the event that no management module is installed or there is a management module fault, the card can be turned on and off using a power-on push button on the processor blade operator panel.
- Soft start to minimize impact on input when the card is turned on.
- Current sensing and limiting. This circuit reacts quickly to an overload on the card to minimize adverse effects to the rest of the system, such as an overload on the power supplies or voltage drop on adjacent blades. This circuit latches card main power off when detected.

A recent addition to the processor blade power input architecture is embedded power-measurement circuitry. This is done as a means of enabling power-management functions in the processor blade firmware. To manage power effectively, the load power must be known. The most recent Intel central processing units (CPUs) have enabled power-management functions [12] that allow control of the CPU power utilization on the basis of thermal or other external criteria. Previously, only CPU thermal events initiated a reduction in CPU power. The system BMC firmware [13] has been modified to take advantage of this capability by enabling additional power utilization in the chassis while maintaining system redundancy features.

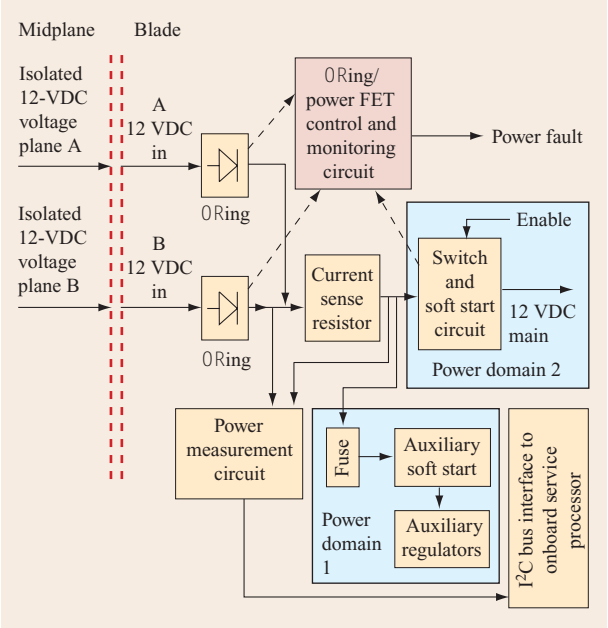


Figure 9

Processor blade power input.

Processor blades that span multiple bays are allowed and may connect to multiple connector sets on the midplane. The total power available for a multislot processor blade is the sum of the individual processor blade slots. Multiwidth processor blades that exceed the power capacity of a single slot must draw power from other occupied slots. The power load and 12-VDC distribution must be segmented to keep each individual slot power connection isolated from the others. This is because the midplane power distribution is divided into two power domains, and it is possible that the multislot processor blade can span this power domain. The processor blade will determine whether all midplane slot connections are powered before allowing main power to be applied to the processor blade.

The 12-VDC main power on the processor blade is distributed on the server system circuit board to the dc–dc converter point-of-load regulators. Since the total power demand of a processor blade can exceed the safety

extra-low-voltage limits of 240 volt-amps, mechanical interlocks and keep-out areas have been designed as previously discussed. The majority of the power consumed on the processor blades is to drive the two processors. Two multiphase buck regulators are used to provide this power (the number of phases is dependent on the CPU power requirement), and each processor blade design is tuned to achieve very efficient high-current dc–dc conversion. Most processor blades also contain six to eight additional single- or two-phase embedded dc–dc regulators to provide the various voltages and currents required by the onboard circuitry.

Module power

The power architecture of the switch and management modules is a simplified version of the 12-VDC distributed processor blade architecture. These modules accept power from the isolated domain A power modules, as shown in Figure 9. Since the total power consumed by an individual module is a fraction of what a processor blade consumes, using a Schottky diode for ORing is sufficient. A fuse is used to provide the over-current and 240 volt-amps protection within a small mechanical keep-in area. Switch and management modules can therefore eliminate ORing FET control and current sensing and power measurement circuitry. High-efficiency dc–dc conversion is still a design criterion, and each switch module typically has five or more dc–dc point-of-load regulators.

The blower modules operate off a single 12-VDC source and therefore are the single exception to the load containing the ORing function. ORing diodes are located on the midplane at each of the blower connectors.

System and rack power

The aggregate power of 14 two-way processor blades, management, switch, and blower modules in a 7U chassis has led to the need for new power requirements at customer data centers. A single chassis can exceed 5 kW of ac input power, and the capability of placing six BladeCenter chassis in a 42U rack means that customers must potentially be able to provide more than 30 kW of ac power. High-density power distribution units with three-phase 60-amp ac power feeds were offered by IBM to provide customers with the ability to power their racks. These high-power levels are still at least 15% less than an equivalent number of IBM 1U servers, and the power savings—based on other 1U rack servers, the types of processor blades, and processor utilization [2]—could be as high as 57%.

Cooling and acoustics

The BladeCenter server package is extremely dense and poses significant thermal challenges beyond those of traditional server packages. A common cooling infrastructure provides the necessary airflow across all processor blades. The air-moving devices consist of two high-performance blowers, which provide redundant cooling. The airflow is carefully balanced across the processor blade and module bays, including times when the bays are not fully populated. As described above, air dampers provide proper air restriction when processor blades are removed for service. The power densities—310 W per processor blade, 5 kW per server, and 30 kW per rack—are among the highest for IBM servers. The processor blade form factor, believed to be the highest volumetric density in the industry using enterprise-level Intel processors, presented unique cooling challenges. These included the high-power density and low-temperature specifications on the processor and memory DIMMs, the use of mobile laptop hard drives, and the requirement to limit the air temperature off the processor blade to ensure proper cooling of downstream modules. The system uses cooling- and power-management algorithms [13] to further achieve a balance between system cooling and performance requirements. Several generations of processor blades and higher-power processors have required improvements to the heat-sink design and more sophisticated power- and cooling-management algorithms. As a result of the cooling capacity required and the resulting high airflow rates, an acoustic attenuation module is used to reduce noise. Finally, BladeCenter deployment and its impact on the data center environment have become one of the major challenges.

Airflow path

The airflow through the BladeCenter chassis is convoluted. All of the air enters the chassis uniformly across the 14 processor blade bays. The airflow passes through the processor blades and exits at the rear via three separate paths (**Figure 10**). By horizontal symmetry, the air exits openings in the rear top and bottom of the processor blade into plenums A and B, and additionally, straight through the midplane into the common blower plenum C. Air in plenums A and B crosses over or under the midplane, turns 90 degrees, proceeds through the switch, management, and power-supply modules, and rejoins the airflow from the rear center of the processor blade in common plenum C. The air is then pulled into the two blowers and exhausted from the chassis.

Chassis airflow rates and impedance

The total airflow targeted was derived as a function of a tolerable temperature rise across the entire chassis, given the expected power. Once the system-level airflow rates were understood, the appropriate subsystem component

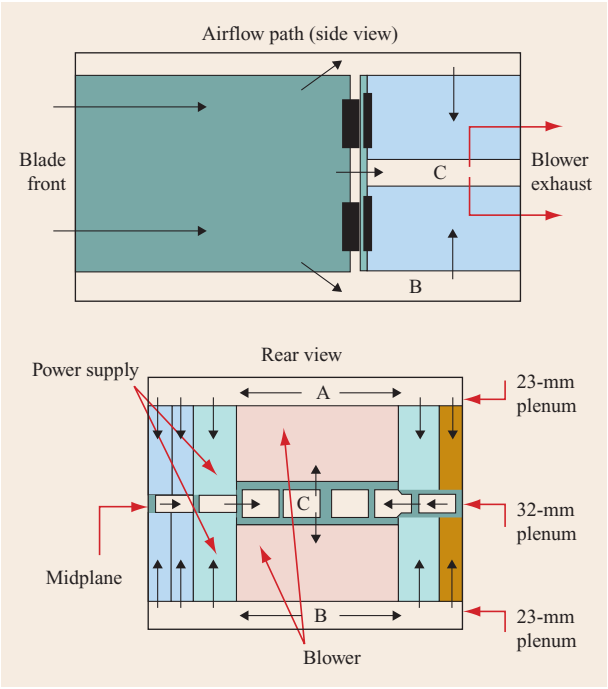


Figure 10

Airflow path.

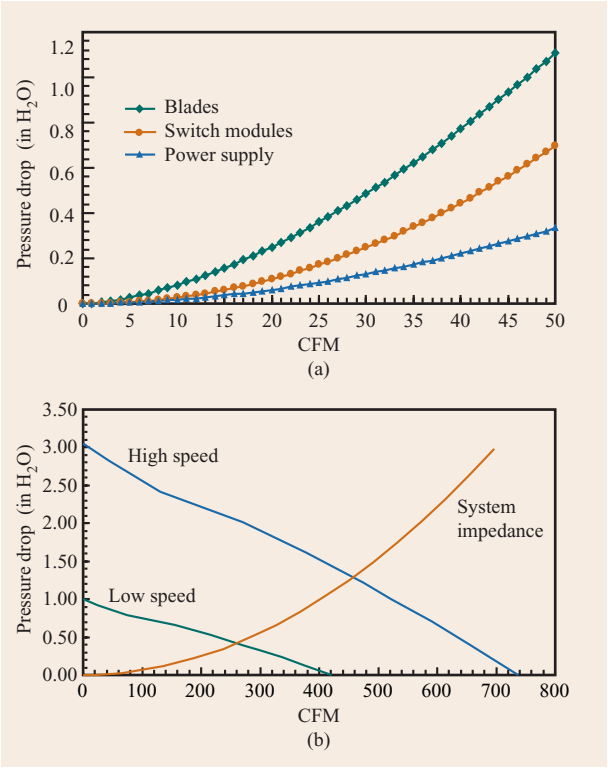


Figure 11

(a) Processor blade and module impedance curves. (b) Blower performance and system curves.

airflow rates and an allowable temperature rise could be determined. The analysis showed that 255 and 455 cubic feet per minute (CFM) of airflow would be required for the system in low-speed and full-speed modes, respectively. This equates to 18 CFM through each processor blade at low blower speed and 32.5 CFM through each processor blade at high speed. The low- and high-blower-speed airflow rates for the modules were determined to be 7 CFM and 14.5 CFM for the switches and 13 CFM and 26 CFM for the power supplies. To achieve the required airflow to the modules, more than 20% of the processor blade air must flow through each of the top and bottom plenums in the rear, with the remainder exiting directly to the common blower plenum. Computational fluid dynamics (CFD) modeling and prototype analysis yielded the pressure drop requirements for the blowers and subsystem components. It is important to note that the ability to cool the power supplies and switch modules was optimized, in that the airflow path was such that it traversed the modules side-to-side rather than lengthwise, which minimized air impedance. Finally, maximum inlet temperature to the switches and power supplies was specified at 52°C. Impedance curves for processor blades and primary modules are shown in **Figure 11(a)**. The blower performance and system curves for low and high blower speed are shown in **Figure 11(b)**.

Air-moving device

To achieve the desired flow rates through the high-impedance and convoluted layout, it was determined that two backward-curve blowers were best suited for the application. The blowers, arguably the most crucial part of the cooling solution, were engineered to a tight set of parameters, which included a maximum of 120 W of dc power draw while being capable of delivering the desired airflow. Several iterations of blower impeller size were proposed prior to the determination that 175 mm would work best for the application. Among other design issues, the close proximity of the two blower inlets resulted in a 15% performance degradation that had to be overcome. One blower at full speed provided 265 CFM, while both blowers at full speed could provide only 455 CFM, compared with the anticipated 530 CFM. The reliability requirement of these blowers was to achieve an L10 of 50,000 hours at 43°C. (The term *L10* means that only 10%

of the total number of blowers will have failed in the stated time period.)

Blades with CPUs

There are many types of blades with CPUs, including processor blades, storage blades, and I/O blades. There are single-wide, double-wide, and triple-wide blades with CPUs. As varying blade solutions become more prominent, the primary requirement is that the power must not exceed a certain limit which would in turn violate the 52°C temperature specification to downstream components. In addition, airflow impedance limits must not be exceeded. It is critical that all blades with CPUs and modules have matched impedance characteristics to ensure that airflow is always balanced.

The IBM HS20 was the first processor blade offered and was based on dual 400-MHz Intel Xeon** PIV processors. Its thermal design power was 89 W with a maximum case temperature of 75°C. Given the best known thermal interface material, a required heat-sink-to-air thermal resistance of 0.29°C/W in a worst-case condition and 0.43°C/W in a typical low-speed acoustic mode condition was needed. Given the spatial constraints of the processor blade form factor, it was determined that using a heat sink with a vapor-chamber base (high-performance heat-sink technology) was the only viable option to obtain these values. Note that initial versions of the Intel Xeon processors did not require vapor-chamber technology. With vapor-chamber technology, a 14% improvement in heat-sink-to-air thermal resistance was realized compared with more conventional technologies in which just aluminum or copper heat spreaders were used.

The aggressive cooling requirements dictated that the processors be located upstream from all other hardware. This resulted in undesirable preheating to all downstream components, namely memory and hard drives. Cooling with preheated air presented extreme cooling challenges for the hard drives. Careful balancing of airflow and life prediction tests at given air temperatures was required. For higher-powered processor blades, the processor blade front allows some fresh air to bypass the processors, providing much-needed local cooling to hard drives. Higher-power processors required further enhancements to the heat sink—copper fins instead of aluminum, increased flatness of the heat sink, and improvements to the Intel processor integrated heat-spreader lid. High-density 2-GB dual-rank memory DIMMs required increased system management intelligence; the blowers would operate at a slightly higher speed upon insertion and presence detection of these DIMMs [11].

In addition to the HS20 processor blade, other processor blades have posed unique cooling challenges. The two most worthy of discussion are the JS20 and HS40 processor blades. The JS20 processor blade, using two IBM PowerPC 970 processors, had to be able to sustain a much higher heat flux (W/mm^2) while maintaining a much lower junction temperature. In addition, a bare-die heat-sink attachment was required, which made the integrity of the thermal interface much more difficult to achieve. The HS40 processor blade is a double-wide four-CPU processor blade. The challenge was cooling the 2×2 processor arrangement, because undesirable preheating from the upstream processors flowed to the two downstream processors. The challenge was met by using a double-wide processor blade package and tall processor heat sinks.

Acoustics

As a result of the high airflow rates through the BladeCenter system, the acoustic requirements were threatened. IBM has a maximum rack-level acoustic declaration of 7.5 bels for a typical configuration operating at or below 25°C inlet air temperature. A BladeCenter chassis is 7U tall, so one 42U rack can support up to six chassis, but a typical rack system configuration is considered to be four chassis. Test results showed that each chassis would generate 7.4 bels. The total acoustic noise due to multiple sources is equal to the logarithmic summation of the individual sources. Therefore, four chassis at 7.4 bels per chassis equated to 8 bels: $\log [(10^{7.4}) \times 4] = 8$ bels.

It was determined that the dominant noise source was the blowers, as opposed to any structural resonance or amplification. By providing a device that eliminated the line of sight to the noise source and also provided acoustic attenuation, a substantial noise reduction could be achieved. As a result, an acoustic attenuation module (a *muffler*) was developed. The muffler has minimal impact on airflow impedance while providing the necessary reduction in system noise. The acoustic attenuation module results in an approximate 5% reduction in cooling capability due to increased overall airflow impedance but provides a 0.5-bel reduction in noise. A single chassis would now generate only 6.9 bels, compared with the original 7.4. Therefore, with the use of a muffler, the 7.5-bel rack requirement was achieved for a typical rack configuration of four BladeCenter chassis: $\log [(10^{6.9}) \times 4] = 7.5$ bels.

Blower-speed control and thermal-management algorithms

The BladeCenter server was defined with a common blower-speed control algorithm, similar to that used in other xSeries servers. This design uses a sensor located in the media tray at the front of the chassis to monitor the inlet air temperature. Feedback to the management module through the I^2C bus sets the blower speed

appropriately via a lookup table. Below 25°C, the blowers operate at low speed and ramp up linearly to full speed at 32°C. Additional sensing of the processor diodes is used as a catastrophic failsafe mechanism. If a processor exceeds a critical threshold temperature, an alert is issued to the user and the blowers are set to full speed. If the temperature continues to increase, the system shuts down in a controlled manner to maintain data integrity.

As processor blade power increased, more sophisticated blower-speed control algorithms were needed, since low-blower-speed, acoustic-mode environments were becoming inadequate to provide cooling at maximum power. To compensate, the blower speed must be managed as it relates to the processor temperature. For these unique solutions, the customer has the choice of *acoustic mode* or *performance mode*. Acoustic mode fixes the blowers at low speed (assuming that the environment is below 25°C) and force-throttles the processors to reduce the power when critical temperature thresholds are breached. If the customer chooses performance mode, in which processor performance is maximized, blower speed may be driven higher at the expense of acoustics. In this case, if the critical processor temperature threshold is breached, the blowers increase in speed to the point at which temperatures are reduced. In this implementation, there is continual interactive communication between the processor and management module to ensure that the critical values are never exceeded.

Finally, because of the increased difficulty in cooling high-powered processor blades, new algorithms were needed for the time when a blower fails. The original HS20 processor blade design had built-in throttling features that scaled back the processor power in the event the cooling system was degraded. The new solutions cannot scale back enough power in fault mode to prevent reaching critical thresholds; therefore, on demand throttling is required, as previously described in the acoustic-mode section. This provides the increased reduction in power, which in turn contains the critical temperatures.

Thermal design tools, techniques, and validation

During the design phase, several analytical tools and techniques were used to optimize the layout and placement of all components. Flomerics Flotherm** CFD software was used as the primary tool for all thermal simulation and modeling. Models were developed for system-level flow balance and for detailed processor blade and heat-sink analysis. This tool was extremely important in understanding the relative differences between mechanical design points in trying to achieve the perfect thermal solution. Without this tool, the team would have had to rely on much more testing, which would have required more engineering resources and a longer development schedule. Although Flotherm was used as the primary CFD software tool, Icepak** from Fluent Inc. and MacroFlow** from Innovative Research, Inc. were also used.

Once up-front analytical modeling was complete, empirical analysis was performed on all levels of hardware, from first prototype through production-level hardware. It is estimated that more than 100 empirical tests were performed at both the chassis and subsystem component levels. System-level tests were conducted in order to understand flow characteristics during low-speed, high-speed, and blower-fail modes, and to analyze the acoustic attenuation module and other rack-level impacts. Detailed subsystem analysis was also performed to ensure not only that the impedance characteristics of all of the functional modules matched one another, but also that the filler modules (used to balance airflow for partial population) matched. Along with this detailed flow analysis, comprehensive component temperature analysis was performed on all processor blades, power supplies, switches, and management modules at all blower speeds. The modules were moved into all possible positions to guarantee that the worst-case test location was understood. In addition to these obvious tests, a more comprehensive suite of tests were performed in order to understand processor effects on downstream components, such as memory and hard drives.

Rack-level deployment and data center impact

The high packaging density of BladeCenter systems results in high power and cooling demands per unit of volume of space occupied. This, in turn, creates additional on-site challenges when they are deployed in data centers. Typical data centers have an average design cooling capacity of 3 kW per rack, with a maximum of 10–15 kW per rack. A full BladeCenter rack could have a load as high as 30 kW. In addition to these extreme heat loads, airflow delivery is a major problem. A full rack of servers typically requires 800–1,500 CFM of chilled air from the perforated tiles located directly in front of the rack. The typical airflow supplied by the data center to a single rack is approximately 200–500 CFM, with 1,000 CFM being an absolute upper bound.

IBM is helping to solve the data center cooling problem by offering the IBM Rear Door Heat eXchanger [14]. The heat exchanger is an effective solution for a data center that is at the limit of its cooling capacity but with usable floor space still available to add racks of systems. The heat exchanger is a more cost-effective solution than adding another air-conditioning unit. The water-cooled heat-exchanger door is designed to dissipate heat from the back of rack-mounted computer systems before it enters the room. It mounts to IBM racks and attaches to

customer-supplied water using industry-standard fittings and couplings. It can extract up to 50,000 British thermal units per hour (BTUH) (or approximately 15 kW) of heat from air exiting the back of a rack of servers. If the heat exchanger is not used and the rack cannot be supplied with sufficient cool air, it may be necessary to reduce the density of heat-generating equipment on the floor. IBM has been working with customers in the deployment of servers by both providing a list of best practices and working hand-in-hand to optimize data center layouts.

Summary

BladeCenter packaging, power, and cooling delivers impressive volumetric density, power density, cooling capability, power and cooling management, and ease of use. A BladeCenter system has twice the volumetric density of 1U servers. The packaging architecture greatly reduces system complexity in areas such as cable management and ease of service. The infrastructure is redundant and hot-swap, and blades can scale through the use of multiwidth blades and expansion blades. The power system is extremely efficient through the use of dual power domains, 12-V distribution, and point-of-load dc–dc regulators. The innovative cooling system design is able to cool high-power Intel processors and support infrastructure. The use of vapor chamber processor heat sinks, series cooling, and backward-curve blowers has helped meet the challenges of the highly dense packaging architecture. A sophisticated power and cooling control system has aided in extending the life of the BladeCenter architecture over many processor generations and technology advances. As the power of today's servers increases, demands on the data center increase. IBM is assisting data center operators to meet these challenges through innovations such as the IBM Rear Door Heat eXchanger and through data center cooling and power optimization. IBM blade products have caused a server packaging paradigm shift in the industry and resulted in an open server blade and switch architecture.

References

1. D. M. Desai, T. M. Bradicich, D. Champion, W. G. Holland, and B. M. Kreuz, "BladeCenter System Overview," *IBM J. Res. & Dev.* **49**, No. 6, 809–821 (2005, this issue).
2. J. Wright, "Electrical Requirements for Blade Servers," Gartner Research, Inc., Document G00120690, November 30, 2004.
3. J. E. Hughes, M. L. Scollard, R. Land, J. Parsonese, C. C. West, V. A. Stankevich, C. L. Purrington, D. Q. Hoang, G. R. Shippy, M. L. Loeb, M. W. Williams, B. A. Smith, and D. M. Desai, "BladeCenter Processor Blades, I/O Expansion Adapters, and Units," *IBM J. Res. & Dev.* **49**, No. 6, 837–859 (2005, this issue).
4. PCI-SIG Group; see *www.pcisig.com/home*.
5. WinterGreen Research, *Blade Server Market Opportunities, Strategies, and Forecasts, 2005 to 2010*, February 2005; see *http://www.mindbranch.com/products/R49-211.html*.
6. J. E. Hughes, P. S. Patel, I. R. Zapata, T. D. Pahel, Jr., J. P. Wong, D. M. Desai, and B. D. Herrman, "BladeCenter Midplane and Media Interface Card," *IBM J. Res. & Dev.* **49**, No. 6, 823–836 (2005, this issue).
7. *InfiniBand Architecture Specification*, Volume 1, Release 1.2, InfiniBand Trade Association; see *www.infinibandta.org*.
8. Innovations First, Inc.; see *http://www.racksolutions.com/ibm/index.html*.
9. S. L. Vanderlinden, B. O. Anthony, G. D. Batalden, B. K. Gorti, J. Lloyd, J. Macon, Jr., G. Pruett, and B. A. Smith, "BladeCenter T System for the Telecommunications Industry," *IBM J. Res. & Dev.* **49**, No. 6, 873–886 (2005, this issue).
10. S. W. Hunter, N. C. Strole, D. W. Cosby, and D. M. Green, "BladeCenter Networking," *IBM J. Res. & Dev.* **49**, No. 6, 905–919 (2005, this issue).
11. T. Brey, B. E. Bigelow, J. E. Bolan, H. Cheselka, Z. Dayar, J. M. Franke, D. E. Johnson, R. N. Kantesaria, E. J. Klodnicki, S. Kochar, S. M. Lardinois, C. M. Morrell, M. S. Rollins, R. R. Wolford, and D. R. Woodham, "BladeCenter Chassis Management," *IBM J. Res. & Dev.* **49**, No. 6, 941–961 (2005, this issue).
12. Intel Corporation, Enhanced Intel SpeedStep Technology; see *http://www.intel.com/ca/pressroom/2004/0628.htm*.
13. G. Pruett, A. Abbondanzio, J. Bielski, T. D. Fadale, A. E. Merkin, Z. Rafalovich, L. A. Riedle, and J. W. Simpson, "BladeCenter Systems Management Software," *IBM J. Res. & Dev.* **49**, No. 6, 963–975 (2005, this issue).
14. IBM Rear Door Heat eXchanger; see *http://www-1.ibm.com/support/docview.wss?uid=psg1MIGR-60049*.

Received December 16, 2004; accepted for publication April 14, 2005; Internet publication October 7, 2005

Martin J. Crippen *IBM Systems and Technology Group, 3039 Cornwallis Road, Research Triangle Park, North Carolina 27709 (crippen@us.ibm.com)*. Mr. Crippen is a Senior Technical Staff Member in IBM eServer Power, Packaging, and Cooling Development. He received a B.S. degree in mechanical engineering from the Rochester Institute of Technology in 1983 and an M.S. degree in mechanical engineering from Binghamton University in 1994. His experience ranges from the development of mainframe computers to open-system VME computers. Mr. Crippen's field of specialty is electronics packaging, including card-on-board and server-level and rack-level packaging. He is well versed in system cooling and power system architecture. He has been instrumental in the development of many IBM Netfinity* and xSeries servers and is the lead mechanical engineer for the IBM eServer BladeCenter system. Mr. Crippen holds numerous patents in electronics packaging and cooling.

Roland K. Alo *IBM Systems and Technology Group, 3039 Cornwallis Road, Research Triangle Park, North Carolina 27709 (aloha@us.ibm.com)*. Mr. Alo is a Senior Industrial Designer in the IBM Systems and Technology Group. He received a B.F.A. degree in industrial design from Brigham Young University in 1987. He has been instrumental in the design development of the PC Server, IBM Netfinity, and xSeries, and he is the industrial designer for the IBM eServer BladeCenter system. He holds many international design awards. In 2003, the BladeCenter system received awards at Industrie Forum in Hanover, Germany, and at SMAU Industrial Design in Italy. Mr. Alo has authored or coauthored several design patents.

David Champion *IBM Systems and Technology Group, 3039 Cornwallis Road, Research Triangle Park, North Carolina 27709 (dchamp@us.ibm.com)*. Dr. Champion is a Senior Technical Staff Member who leads a human factors team that specializes in hardware usability: that is, ease of installation, service, and upgrades. He is one of the originators of light-path diagnostics, and he has been a major influence in enabling customer-replaceable units in products. Dr. Champion received a B.S. degree in economics from London University, an M.S. degree in ergonomics from Loughborough University (U.K.), and a Ph.D. degree in psychology from North Carolina State University. He also has a postgraduate diploma in education from Nottingham University (U.K.). He joined IBM in 1989. Dr. Champion holds several patents.

Raymond M. Clemo *IBM Systems and Technology Group, 3039 Cornwallis Road, Research Triangle Park, North Carolina 27709 (clemo@us.ibm.com)*. Mr. Clemo is a Senior Engineer in IBM eServer Power, Packaging, and Cooling Development. He received a B.S. degree in engineering (magna cum laude) from the University of Central Florida in 1977. He joined the Power Group at IBM Boca Raton and has worked since then in power design, development, and test. He joined xSeries Power Systems in September 1998, becoming department lead engineer. His current responsibilities include authoring and maintaining the power section of the xSeries Hardware Design Guide. He is a member of the xSeries CBB and System Development Councils. Mr. Clemo has received the IBM First Plateau Invention Achievement Award.

Cynthia M. Grosser *IBM Systems and Technology Group, 3039 Cornwallis Road, Research Triangle Park, North Carolina 27709 (grosserc@us.ibm.com)*. Mrs. Grosser is a manager of server mechanical development in IBM eServer Power, Packaging, and Cooling Development. She received a B.S. degree in mechanical engineering with a minor in engineering mechanics, and an M.S. degree in mechanical engineering from Pennsylvania State University in 1994 and 1996, respectively. She was the mechanical engineering leader for the key teams that designed the first IBM 1U and 2U servers. She created the Mechanical Peer Design Review team for the xSeries brand. Mrs. Grosser holds several patents in mechanical packaging, some of which pertain to hot-plug PCI.

Nickolas J. Gruendler *IBM Systems and Technology Group, 3039 Cornwallis Road, Research Triangle Park, North Carolina 27709 (njg@us.ibm.com)*. Mr. Gruendler is a Senior Engineer in IBM eServer Power, Packaging, and Cooling Development. He received a B.S. degree, with honors, in mechanical engineering from the University of Missouri at Columbia in 1977, and an M.S. degree in electrical engineering from the University of Kentucky at Lexington in 1985. He joined IBM at Lexington in 1977. His most recent assignment is in Research Triangle Park working on xSeries power systems, where he is the lead power engineer for the BladeCenter system. He has been a member of the Power Development Council since 1997 and chairman since 2000. Mr. Gruendler has authored or coauthored several patents and technical bulletin disclosure publications.

Mohan S. Mansuria *IBM Systems and Technology Group, 3039 Cornwallis Road, Research Triangle Park, North Carolina 27709 (mansuri@us.ibm.com)*. Mr. Mansuria is a Senior Engineer in Thermal Technology and Development. He received an M.S. degree in mechanical engineering from North Carolina State in 1966 and an M.S. degree in operations research from Union College in 1976. He has more than 30 years of experience in air and water cooling in first-, second-, and third-level packages in IBM systems. Mr. Mansuria holds more than 15 patents in electronic packaging and cooling technologies. He currently works on the thermal design of xSeries DP and MP servers.

Jason A. Matteson *IBM Systems and Technology Group, 3039 Cornwallis Road, Research Triangle Park, North Carolina 27709 (mattesj@us.ibm.com)*. Mr. Matteson received an A.S. degree in engineering science from the State University of New York at Alfred in 1994 and a B.S. degree in mechanical engineering from the Rochester Institute of Technology in 1997. He began his professional career with IBM as a mechanical engineer, supporting the mechanical development team with system and hardware design. He received an IBM Outstanding Technical Achievement Award for the thermal design and support of the first 1U, dual-P4 Intel Xeon** servers. He supports IBM key customers with data-center cooling concerns and issues resulting from ultra-dense server deployments. Mr. Matteson has authored or coauthored several patents and technical bulletin disclosure publications.

Michael S. Miller *IBM Systems and Technology Group, 3039 Cornwallis Road, Research Triangle Park, North Carolina 27709 (msmiller@us.ibm.com)*. Mr. Miller is a Senior Technical Staff Member supporting xSeries power, packaging, and cooling. His current assignment is as an architect defining and developing future server products and building blocks. He graduated from the University of Florida in 1978 with a B.S.M.E. degree. Mr. Miller has been heavily involved in defining industry standards such as ISA, MicroChannel, PCI, PCMCIA, InfiniBand, PCI-Express**, SIOM, and planar, media, power-supply, and drive form factors. For the last nine years he has been a member of what is now xSeries development and was heavily involved in defining the BladeCenter system. Mr. Miller holds numerous patents and has achieved his third-level invention plateau.

Brian A. Trumbo *IBM Systems and Technology Group, 3039 Cornwallis Road, Research Triangle Park, North Carolina 27709 (batrumbo@us.ibm.com)*. Mr. Trumbo is an Advisory Engineer in IBM eServer Power, Packaging, and Cooling Development. He joined IBM in 1978 as an apprentice tool and model maker with the SEDAB organization in Boca Raton, Florida, and in 1982 moved to the IBM Personal Computer Group. He has 26 years of mechanical engineering, design, and tooling experience with many computing systems. Mr. Trumbo participates as an active member of the IBM RTP Patent Review Board and the RTP PP&C Mechanical Peer Design Review Council. He has been recognized with several IBM awards, holds numerous IBM patents, and is recognized as an IBM RTP Master Inventor.