BladeCenter system overview

D. M. Desai
T. M. Bradicich
D. Champion
W. G. Holland
B. M. Kreuz

The IBM eServer™ BladeCenter® system was conceived in the fall of 1999 by members of the IBM xSeries® server brand technical team (at the time, called the Netfinity® brand). It was a new idea, and for some, too revolutionary to consider. After considerable end-user feedback, technical refinement, and internal discussion and debate, the concept was developed into a leadership server product. This paper provides a brief historic overview of the development of the architecture and blade servers, and a technical description of the BladeCenter system and technology.

Introduction

The summer of 1999 was a dynamic time for server technology. Intel processors were on the verge of transitioning from Pentium** III to Pentium 4 technology [1]. Pentium 4 technology promised higher clock frequencies than the Pentium III technology could achieve. One-gigabit Ethernet had started rapid market growth with the ratification of the 1000Base-T copper Gigabit Ethernet standard in June (IEEE 802.3ab) [2]. At this time, Fibre Channel was introduced in the IBM X-Architecture*, the guiding architecture of the xSeries* [3]. The yearly xSeries strategic vision study had reported on a number of opportunities emerging from customer and technology trends. These perceived opportunities included the convergence of local area networks (LANs), storage area networks (SANs), and wide area networks (WANs) with clusters, and one intriguing item that as yet had no solution proposal—*cableless servers*.

In the networking industry, Transmission Control Protocol/Internet Protocol (TCP/IP) [4] Ethernet had grown from being just one option out of a dozen or so common protocol or network combinations (or both) that a server had to support to being the only network needed anywhere. IP storage, which would eventually be standardized as Internet Small Computer System Interface (iSCSI), was a growing topic of research and development. InfiniBand** input/output (I/O) [5] was being designed for high-speed, low-latency interconnections, and its specification included modular physical form factors which would later serve as the basis for the BladeCenter* blade and module form factors. The typical server grew to offer many Peripheral Computer Interconnect (PCI) [6] I/O slots to meet the need to use a server flexibly on any type of network or on multiple networks with sufficient cumulative I/O performance. This flexibility was to be preserved in the BladeCenter design. The ubiquity and high performance of Gigabit Ethernet would enable servers to connect through just one or two cables, making adapter slots less necessary.

These various technological innovations and changes were occurring in the time period during which the *dot-com* industry developed. That was typified by a demand for server systems that scaled out, supported high-speed networking fabrics, and above all, were dense. The emerging segments of the information technology industry held server rack space at a premium.

Four-way servers (with four processors) that in 1995 had been 11U (19 in.) high were, by 1999, being designed to fit into 4U (7 in.) of vertical rack space. Two-way servers (with two processors) were also being compressed, from 5U (8.75 in.) down to the smallest rack-mount package, 1U (1.75 in.).

At this density, with 42 servers in a 42U-high rack, installation and servicing problems arose. Not least of these was the sheer number of cables required. Forty-two servers in a rack, each with a keyboard, video, mouse (KVM), power, and two network cables, resulted in more than 240 cables that had to be managed. Moreover, these early 1U servers lacked redundant power and redundant cooling, and had to be lifted out of the rack for servicing. They required an infrastructure of switches for switching the KVM from server to server, and they required LAN and SAN networking switches to carry data to and from end users and on and off storage. To simply install the rails for 42 servers was a two-hour task. Innovative ways of reducing the number of KVM cables, such as cable-chaining technology, were introduced, as were tool-less rails, but these were only patches. The need for even

0018-8646//$5.00 © 2005 IBM

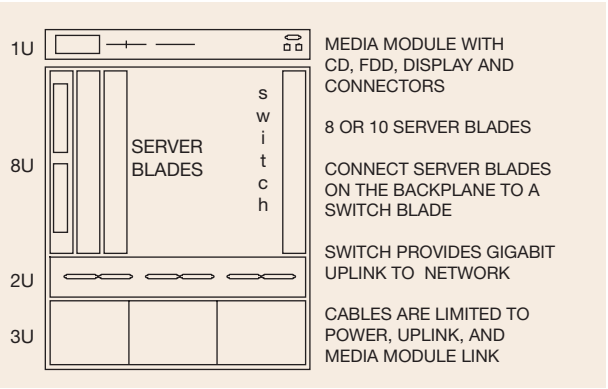


Figure 1

The earliest concept drawing of the BladeCenter architecture, created on October 20, 1999.

greater density and more compute power per rack continued growing. Ideas such as placing two servers side by side in a 1U space were explored, but this would have compounded the existing serviceability problems. A breakthrough was needed, and this breakthrough was to be the BladeCenter system.

The concept of computing devices in the blade form factor is not new. They have existed in the telecommunications industry, and have been offered by a limited number of server vendors [7–9]. Router, switch, concentrator, and packet processing blades existed in large-chassis network switches and hubs, such as the IBM 8260 and Cisco Systems 7500. The xSeries development personnel included engineers and architects who previously had designed this class of product while part of the IBM Networking Hardware Division. Because of this, it seemed that extending the blade concept to server blades would have been an obvious next step—but it was not.

High-volume xSeries servers had grown out of the IBM personal computer (PC) tradition and hence were principally tower models. Using industry-standard components, the development focus was on cost and time to market. The pressure was on to ship new processors and features in systems before customers could buy them from the competition. Customers and decision makers, including server and network administrators, saw servers as end nodes on a network. They were traditionally maintained and managed separately from the network itself, by people in different parts of the organization. The only place the server and the network intersected was the network interface card selection, and the only place the server intersected with the SAN was the host bus adapter and the selection of the multipath failover driver.

Networking equipment design was foreign to the xSeries server engineers, who had only recently (1995) begun to migrate from towers to true rack-mounted servers.

Thus, when the question was finally asked, *Why not a server blade?*, it was not a concept universally greeted with understanding or enthusiasm. That being said, senior technical people did take it seriously. An unofficial technical team began regular meetings in September of 1999. The areas of expertise represented on the team covered many disciplines: system architecture, power, cooling, logic, software, mechanical packaging, human factors, reliability, switching architecture, and serviceability. The work was initially unfunded and without mandate; it was a skunk works, but one of the benefits of this was that a regular development team would not have had such a range of talent available to it.

The very first concept drew heavily on the telecommunications hub model. It comprised a chassis with a backplane that mounted into a two-post telecommunications rack. The server blades would plug to the backplane, lined up like books on a shelf. The infrastructure—power, cooling, and media—could be shared over a number of servers and be redundant. Network switches would also plug into the backplane. Connectivity for media, power, and data would be over the backplane, eliminating the need for individual cables for each function. Since everything plugged into the front, service would be from the front only; if the components were made hot-swappable, servicing it would be quick and easy. The final product looked much different from this, although many of the ideas were maintained. The earliest concept drawing, created on October 20, 1999, is shown in **Figure 1**.

When the first designs were presented, they were rejected as too revolutionary, and emphasis was then placed on designing for a standard cabinet rack. As a result, a model was built that had a midplane rather than a backplane, was designed with front-to-rear cooling, and allowed components to be installed from both the front and the rear of the chassis. It held eight server blades in a 7U package, resulting in a density of 56 servers per rack rather than 42. This model was discussed with the marketing organization and shown to customers. The feedback was positive, but marketing wanted yet higher density and asked for 84 servers per rack. Here was the true challenge: to fit into a 7U chassis 14 servers with redundant power, cooling, management, network switch modules, shared KVM and media, and, of course, redundant interfaces on the midplane.

The development effort was finally chartered as a formal development project in 2000. Many development challenges and problems arose during development and during the period when early systems were delivered to

customers. Ultimately, the development team had to quickly pull together imaginative solutions to address areas that had not been anticipated. For example, there was a clear need to continue providing the classic serial port interface. For system management, there were problems in arranging failover between the redundant management modules.

Looking back, the major achievements were doubling the density of servers in a rack, blending the server and telecommunications functions within a single chassis, devising a way to manage the system, and providing a high level of installability and serviceability when compared with 1U systems. While we were solving problems of density and cabling, we never foresaw the impact we would have on the industry or on data center infrastructure. As it turned out, the BladeCenter architecture represented a new way of looking at storage, servers, and networking—as a nexus of all three, at the same time simplifying the way some things are done and making other things more complicated. What is clear is that it provided a higher level of integration going forward than had been seen before.

Blade servers

Traditionally, industry-standard servers (servers utilizing Intel processors and capable of running such operating systems as Microsoft Windows**, Linux**, and Novell NetWare**) are thought of as standalone enclosures that are deployed in rack-mounted or tower configurations. These servers contain a system board with processors, memory, and I/O devices, and a power supply, hard disks, a CD-ROM (compact disk–read-only memory), a diskette drive, a network interface, and PCI I/O expansion adapter slots. Using this traditional packaging approach, the industry has shrunk the size of a rack-mounted server down to what appears at first glance to be a logical limit on the minimum size that a server can attain: a height of 1U (1.75 in.). However, using a new approach with blade servers, the industry is breaking through the 1U barrier and achieving higher density.

A blade server is a subset of an industry-standard server that is implemented as a thin, pluggable board with a top and bottom protective enclosure. It slides into a *chassis*, or enclosure, designed specifically to house multiple units. Each blade connects to the midplane, from which it shares common resources such as power, cooling, network connectivity, management functions, and access to other shared resources (such as a front-panel, CD-ROM drive, or diskette drive). In most cases, a blade is implemented as a single unit. However, in some cases the covers can enclose more than one board, or the blades might snap together to form multiblade units that occupy multiple slots.

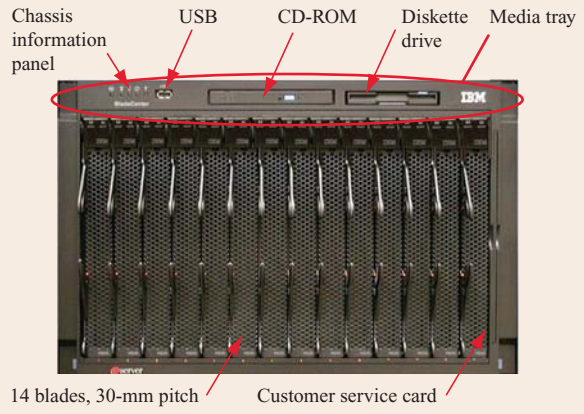
BladeCenter system overview

Power, packaging, and cooling

The IBM eServer* BladeCenter system is a high-density, rack-mounted packaging architecture for servers. It can support virtually any processor family as long as it is designed to be compatible with the I/O, systems management, power, and packaging architecture of the BladeCenter system. A blade is not necessarily limited to being a server or compute node. Moreover, it integrates networking and I/O with servers in a fundamental way with the convergence of physical, administrative, and operational functions. That makes it very well suited for a data center infrastructure. All blades and modules (with the exception of the midplane) are *hot-pluggable*, which means that if one fails, it can be replaced without shutting down the system power. The BladeCenter system is designed to provide complete redundancy of power, cooling, and a midplane to support a very high level of availability. This redundancy allows continuous operation of a BladeCenter system in the event one of these subsystems fails. However, it cannot tolerate a double fault within the same subsystem.

A BladeCenter system consists of a 7U by 28-in. (711-mm)-deep chassis [10] with a support structure for up to 14 blade slots, a midplane, a media tray containing a CD-ROM and floppy disk drive, a universal serial bus (USB) port, a chassis information panel, and a customer service card in the front [**Figure 2(a)**]. In the rear of the chassis, it supports the shared infrastructure for four switch module bays, two management module bays, four power module bays, a rear information panel, and two blowers [**Figure 2(b)**]. These subsystems all interface with one another via the midplane. The BladeCenter system shares common power, packaging, cooling, systems management, and I/O infrastructure among all subsystems, which allows doubling the server-node density compared with a standard 1U server. Each blade server functions as an independent server with its own operating system and application. Blades, switch modules, management modules, power modules, and shared resources interconnect via a redundant midplane. Airflow is from the front to the rear, controlled by the blower modules at the rear. All service access is from either the front or the rear of the unit; the BladeCenter chassis does not have to be removed from the rack for service. This eliminates the need for retractable chassis slides and for an articulated cable-management arm.

The BladeCenter system power is divided into two independent power boundaries: A and B. Each domain can be equipped with a pair of redundant ac input power modules and provides +12-VDC power to the system loads.



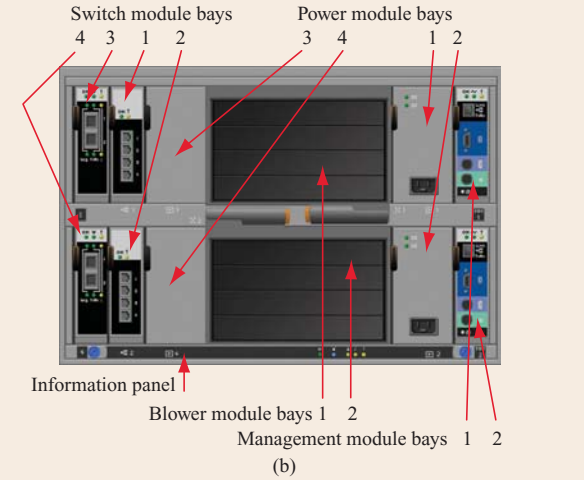
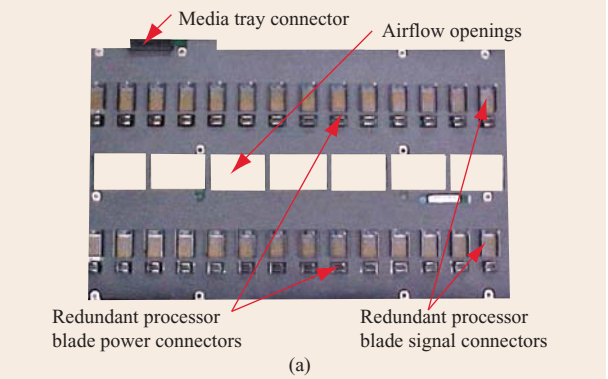


Figure 2

BladeCenter (a) front and (b) rear views.



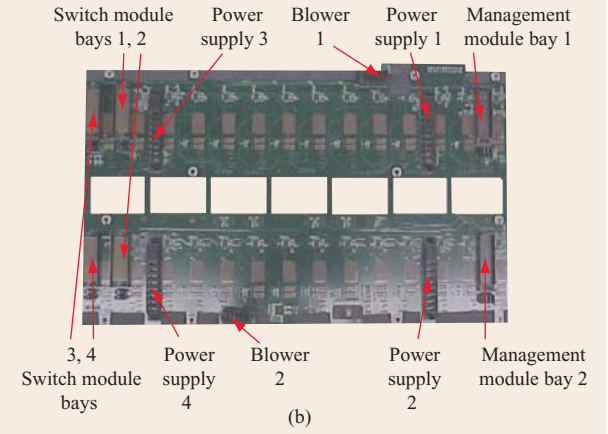


Figure 3

(a) Front view of midplane. (b) Rear view of midplane.

Midplane

The midplane is mounted on the switch, power, and cooling (SPC) subchassis. The SPC subchassis slides into the rear of the BladeCenter chassis, providing the bays in which the switch, power, management, and blower modules are installed. It provides all of the interconnections among the blades, modules, media tray, and dc power distribution throughout the chassis (**Figure 3**). Air flows above and below the midplane as well as through the seven openings in its center.

The midplane is logically divided into an upper and a lower half, which, with few exceptions, are identical. All of the connectors and interconnections on the upper half are duplicated on the lower half [11]. There is no single point of failure, and this provides the redundancy required for nonstop operation, which enables the BladeCenter system to achieve our reliability, availability, and serviceability objectives. All blades and modules and the media tray plug into the midplane and are hot-pluggable. Nonvolatile data specific to the chassis is stored in an electrically erasable programmable read-only memory (EEPROM) on the midplane.

Processor blade

A typical processor blade is implemented on a board with a protective top and bottom enclosure and plugs into a midplane in a BladeCenter chassis [12]. **Figure 4** shows the BladeCenter components discussed here and below. The processor blade shown in Figure 4(a) is the single-wide HS20 processor blade. The blade enclosure is based on the InfiniBand specification. It is an InfiniBand I/O double-high (6U), single-width (29 mm) enclosure. However, it has a custom depth of approximately 446 mm. Airflow is from front to rear. A typical blade server has a processor, memory, I/O subsystem, and other logic necessary to interface with the midplane. The blade interacts with other subsystems within a BladeCenter chassis via the midplane using two 60-pin Teradyne

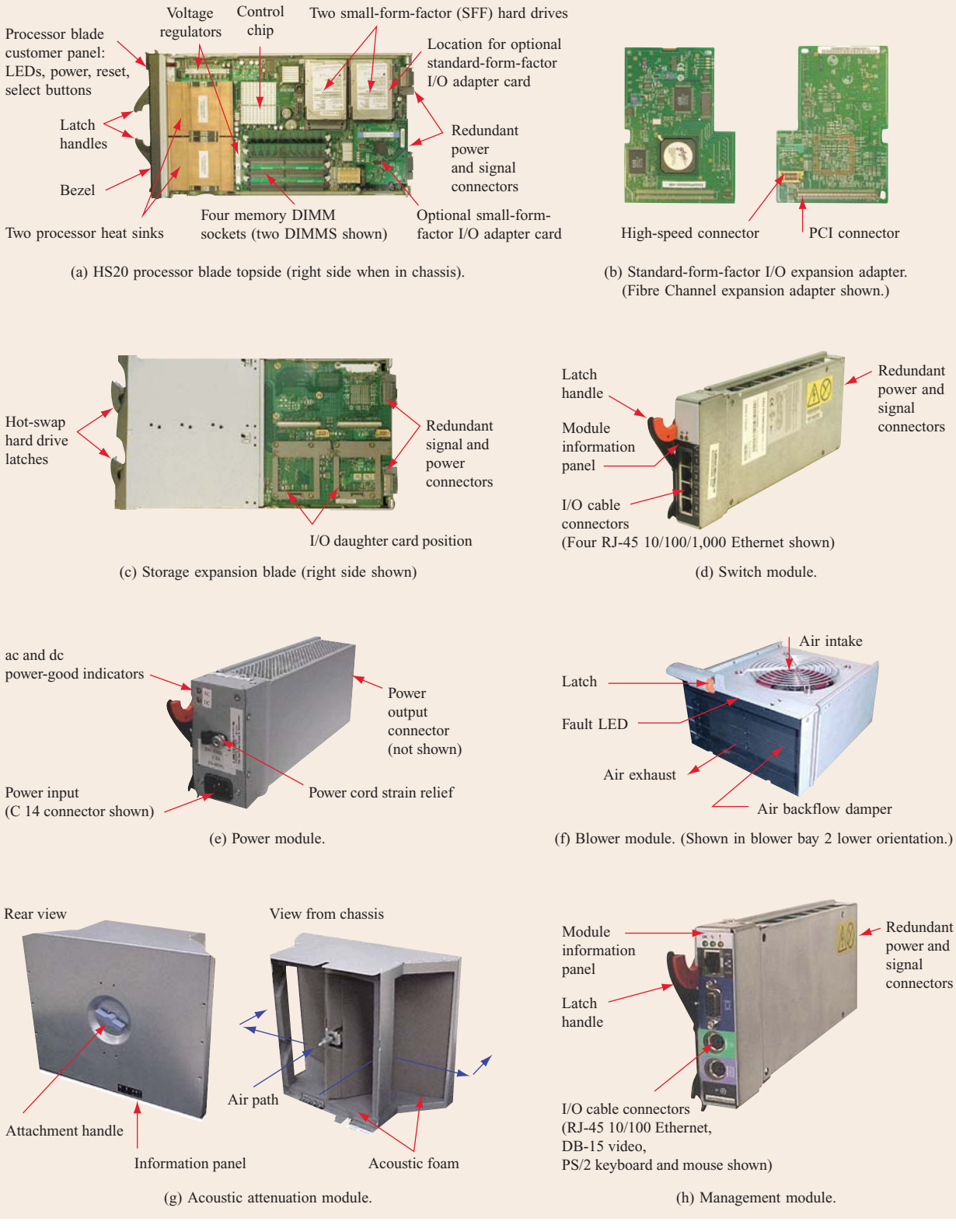


Figure 4

BladeCenter components.

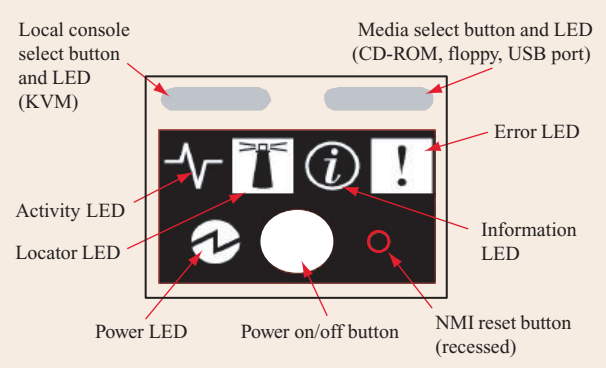


Figure 5

Processor blade information panel.

VHDM** [13] connectors. To avoid a single point of failure, these two connectors provide identical interfaces with the midplane for redundancy, ensuring continuous blade operation in case a power module, management module, or switch module fails.

In most cases, a blade is typically implemented as a single board 15.6 in. × 8.9 in. (396.8 mm × 226.9 mm), which is single width (29 mm) and takes up one slot. However, in some cases one blade might take up more than one slot, depending upon cooling requirements or as a result of a requirement to add I/O expansion adapters (see below), hard disk drives, or both. Blades interface with other subsystems in the chassis via the midplane. The midplane offers the following interfaces [13]:

- RS-485.
- Analog video bus—red, green, blue (RGB).
- USB.
- High-speed serializer/deserializer (SerDes) (ports 1 and 2 are typically Ethernet).
- Control signals (slot ID, blade detect, select A or B connector, etc.).
- 12-V continuous power.

These interfaces provide blades with the ability to communicate with other subsystems in the BladeCenter chassis. Subsystems can include components such as management modules, a media tray, or switch modules, such as Ethernet, Fibre Channel, or InfiniBand. These interfaces are duplicated on the midplane to provide redundancy. All blades must have an Intelligent Platform Management Interface (IPMI)-compliant baseboard management controller to support the RS-485 interface so that the management module can manage the blade. Each blade also has a blade customer panel with five information light-emitting diodes (LEDs) and four push buttons. The LEDs display the current status of a blade, and the push buttons provide control for power on/off, selection of a CD-ROM or floppy disk, selection of the blade for the KVM, and nonmaskable interrupt (NMI) reset for core dumps (**Figure 5**).

I/O expansion adapter

A blade is designed to support two different-form-factor I/O expansion adapters that are unique to the BladeCenter system [Figure 4(b)]. Both types of I/O expansion adapters use same physical connector interface; however, only one I/O expansion adapter can be plugged into a blade. An I/O expansion adapter provides an expansion area for additional I/O functions [12]. It supports a 64-bit PCI-X 1.0 electrical interface at 133 MHz via a 200-pin board-to-board stack connector to the host [13]. It also provides an interface with high-speed ports from blade to switch module bays 3 and 4.

Numerous functions can be deployed on an I/O expansion adapter or switch module, or both. The blade can gain access to a network, such as Fibre Channel SAN, Ethernet, or InfiniBand, via an I/O adapter and switch module combination. Once an I/O expansion adapter is installed in one blade, all other blades in a given chassis must use an adapter with the same protocol, or use no adapter at all. The management module checks that an I/O expansion adapter and switch module are compatible and support the same protocol before turning on the blade. If there is a mismatch, the management module logs an error, sends an alert, and does not turn on the blade.

The blade supports two types of I/O adapter form factors: *standard* and *small*. Since both use the same physical interface, they are mutually exclusive. The small-form-factor I/O expansion adapter, supported by 8843 and beyond, allows two hard disk drives (HDDs) and an I/O adapter and also supports a standard I/O adapter. Currently, only Fibre Channel and Ethernet I/O expansion adapters are available in a small form factor. The legacy I/O expansion adapter slot is positioned near the rear of the board adjacent to the VHDM connector. The I/O expansion adapter slot overlays the rear HDD bay on the blade so that only the rear HDD or I/O expansion adapter can be installed.

Expansion blade

The expansion blade is a mezzanine cartridge that can be added to the base blade to expand its function [Figure 4(c)]. The expansion blade is an add-on blade assembly that communicates with the base blade via a PCI-Express** bus, to which it is attached. It can be added on by removing the top cover of the processor blade and plugging the expansion blade into the expansion connector on the processor blade board. The

retention of the expansion blade is by the same tool-less method used to retain the single-wide blade cover. In all cases, the power for the expansion blade is obtained from the midplane. Expansion options include 1) a blade storage expansion (BSE) unit that can support two 3.5-in. hot-swappable SCSI hard drives, and 2) a PCI I/O expansion unit (PEU) that can support two full-size PCI adapters [14]. For this option to be viable, a blade must be designed to be compatible with the electrical and mechanical interface that supports the capability of expanding a blade attachment via an expansion blade. The add-on structure increases the width of the blade assembly by 30 mm to 60 mm, depending on the expansion unit option. In some cases, expansion options can provide additional features, such as additional I/O daughter cards.

Modules

A module is a device that is installed in the rear of the BladeCenter chassis. There are four basic types of modules: *switch modules* provide network connectivity between the blade I/O and the network; *power modules* provide power to the system; *blower modules* cool it (an optional acoustic module is offered to reduce noise); and *management modules* manage the chassis, blades, and other modules. Switch and management modules are 29 mm wide and are referred to as *single-wide* modules. Power modules are 59 mm wide and are referred to as *double-wide* modules.

Switch module

The switch module provides networking or switch functions, or both, to the blade servers. Each blade is capable of accessing up to four switch modules via point-to-point links across the midplane. Duplicate switches can be added for the purpose of providing redundancy, speed, or additional function. The mechanical enclosure of the switch module is based on the standard 3U single-wide, single-high InfiniBand module form factor [Figure 4(d)]. It uses a VHDM signal connector to interface with the midplane. Airflow is from top to bottom or bottom to top, depending on whether the switch module is installed in a top or bottom slot of the chassis. Four switch modules are supported in the chassis. The management module provides a single point of control for the BladeCenter chassis. An Inter-Integrated Circuit (I^2C) Serial Bus Interface is required between the management module and switch modules. The I^2C bus is used to collect information and perform initial configuration, and to provide monitoring and control of the switch modules [13]. Each management module also has a 100-Mb/s Ethernet interface with each of the switch modules. This Ethernet interface can be used once IP traffic has been established between the management module and the switch module. The internal IP connection is used to configure and monitor the switches.

Switch module bays 1 and 2 [Figure 2(b)] are typically Ethernet modules, and switch module bays 3 and 4 support another two-network port based on an installed I/O expansion adapter, such as a Fibre Channel, Ethernet, or InfiniBand adapter [15].

Power modules

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

Power modules 1 and 2 provide redundant power for power domain A, and power modules 3 and 4 provide redundant power for power domain B. Each power module provides dc power to all entities in its power domain. If one power module in a domain fails, the other power module in the domain continues to handle the load. Within domain A, the following subsystems are powered: processor blades 1 through 6, blowers 1 and 2, management modules 1 and 2, switch modules 1 through 4, and front and rear operator panels.

In domain B, power is delivered to processor blades 7 through 14. The power modules are designed for hot-swap removal, insertion, or replacement. Each power module communicates with the management module through the I^2C bus protocol using a microcontroller or I/O expander for power-module control and status information [10]. The power modules are designed in a double-wide, single-high, custom-length InfiniBand-based mechanical enclosure [Figure 4(e)]. Airflow through power modules 1 and 3 is from top to bottom, while airflow through power modules 2 and 4 is from bottom to top. The design supports up to 2,000 watts per power module.

Blower modules

The BladeCenter system has two blower module bays. The reverse-impeller variable-speed blowers draw air from the front of the BladeCenter system and exhaust it out the rear. Each of the two blowers has a backflow damper [Figure 4(f)]. This damper prevents air from recirculating into a failed blower, which would deprive the blades and modules within the chassis of airflow. Each bay also has a backflow damper to prevent air recirculation if a blower is removed.

The blowers are hot-swappable and redundant; a single blower will cool all electronics (within a set operating range). Air passes through the 14 blades and through or around the midplane. More than 50% of the air passes through openings in the center of the midplane and

directly into the two blowers. Approximately 20% of the air passes over and 20% passes under the midplane and then is directed into the modules before being exhausted by the blowers.

The blowers are powered by a constant 12-VDC source. The blower speed is controlled by the management module, which provides a low-voltage, low-current control signal to each blower. A single blower will produce approximately 265 ft^3/min (CFM) at 0.9 in. H_2O. A thermistor, mounted in the media tray, indicates inlet air temperature, which is monitored by the management module. Blades and modules have thermal sensors that are also monitored by the management module. The blowers pull fresh air across the blades to cool them. Blade processors are cooled using low-profile vapor chamber heat sinks. Memory, hard drives, and blade board electronics are cooled, for the most part, by air that has been preheated by first passing over the processors, and then continues to pass to the rear of the blade, where it cools these downstream components.

The blowers are responsible for most of the noise generated by the system. Acoustic requirements are achieved by limiting blower revolutions per minute up to 25°C operating temperature. Above 25°C, blower speed is allowed to increase. Consequently, acoustic requirements are not met above 25°C. Up to six BladeCenter servers can be installed in a standard 42U rack. However, even with the optional acoustic module installed, the acoustic requirement cannot be met with six BladeCenter systems. The maximum number of BladeCenter systems in a rack for which acoustic requirements can be met is four.

Air enters the blades and passes through the processor heat sinks, over blade components such as memory, control logic, and hard drives. It then passes to the rear of the chassis via three paths [10]. Air passes above, through the center of, and below the midplane. The air at the bottom passes up through the modules and into a common plenum. The air at the top of the chassis passes down through the upper modules and into the common plenum. Air also passes through the midplane and into the common plenum. Air then passes into the two blowers and is exhausted out the rear of the server.

Acoustic attenuation module (optional)

The optional acoustic attenuation module is shown in Figure 4(g). This assembly can be attached to the rear of the chassis if the user desires to reduce noise emissions. The acoustic module reduces the noise by approximately 5 dB. The addition of the acoustic attenuation module increases the air impedance through the system by a small amount but does not adversely reduce the cooling of the system. However, for optimal cooling, it is recommended that the module not be used.

Management module

The main task of the management module is to manage the BladeCenter chassis, blades, modules, and shared resources. It also provides functionality that allows a data center management application, such as IBM Director, to be used to manage the BladeCenter system. The management module consists of a processor and KVM switch function [Figure 4(h)]. It has an Ethernet point-to-point connection with each of the four switch modules and with all other major components in the BladeCenter chassis via I^2C, USB, and RS-485 buses [16]. It uses a Teradyne HDM** signal connector to interface with the midplane. Airflow is from top to bottom or from bottom to top, depending on the location of the management module in the chassis. The mechanical enclosure is a standard, single-wide, single-high InfiniBand module. The depth of the management module is greater than that of the InfiniBand standard. There are two management module bays for redundant operation, and hot-standby redundant operation is supported in case of failure. The management module provides the following features:

- System-management processor functions for the BladeCenter system.
- Ethernet connection to a management network.
- Video port (local and remote console).
- IBM PS/2* keyboard port.
- PS/2 mouse port.
- 10/100-Mb/s Ethernet connection.

System management

The resources shared among blades in a BladeCenter chassis enable new management paradigms. The system is designed to reduce downtime by providing redundancy for all subsystems. The chassis management architecture is based on a multi-tiered management concept involving a management module that provides support for all chassis components. The baseboard management controller (BMC), described more fully below, is located on a blade and works in conjunction with the management module to manage the blade.

Chassis-level components include power modules, customer-interface panels, CD-ROM or floppy drives, and switch modules (Ethernet, Fibre Channel, or other switches). They communicate with the management modules over I^2C buses to provide various levels of control and status. The blowers, which have unique signals, are connected directly to the management modules. In addition, the management modules support an Ethernet connection to each switch slot for the purpose of fabric configuration and management [16]. In addition to supporting normal processor management function, the baseboard management controller on each blade provides control and status by communicating with

the management module over the RS-485 bus using an IPMI protocol. An external Ethernet link on the management module provides connectivity for *remote* management, including full console capability (KVM) with keystroke selection of the target processor blade. PS/2 and video ports on the management module enable *local* console (KVM) access to individual blades with keystroke selection [16]. Functions provided by the management module include but are not limited to the following:

- Chassis configuration.
- Chassis cooling (blower control and temperature sensing).
- Power module control.
- Blade initialization.
- Switch module initialization.
- Media selection and control (CD-ROM or floppy disk drive).
- Remote and local console control.
- Customer interface panel.
- Chassis-level power management.
- Power on/off control.
- Chassis thermal sensing (monitor thermal status and post alerts).
- Serial-over-LAN (SOL) session control and terminal server.

Functions provided by the blade BMC include but are not limited to the following:

- Power on/off control.
- Media control (request and enable or disable CD-ROM or floppy disk drive access).
- Keyboard and mouse control (access to USB bus on the midplane).
- Video control (access to video bus on the midplane).
- Thermal sensing (monitor thermal status and post alerts).
- Management module interface (communications with management module).
- Blade power management.
- SOL session.

The management modules support hot-plug capability so that if one becomes disabled, either as a result of a component failure on the management module or by its removal from the chassis, there will be no disruption of ongoing normal operation of blades and modules. Also, chassis configuration cannot be changed, monitored, or managed when a management module is being replaced because of failure. Additionally, upon becoming reenabled, the management module does not disrupt the then-current operation or configuration of the chassis components without policy intervention.

Redundant and failover operation provides two distinct capabilities: redundant bus support and management module failover. Redundant bus capability allows the active management module to switch the bus being used to communicate with the chassis components when an I^2C or RS-485 bus communications problem has been detected. Failover operation is provided when a chassis is configured with two management modules. When operating in this mode, one management module operates as the active manager, with the other in standby mode. Each management module monitors the state of the other management module within the chassis and, upon detecting a failure of the active module, the standby module takes over as the active management module. If the active management module detects a failure of the standby management module, an alert is posted indicating that failover operation is inactive. In the event of a failover of the active management module, there will be no disruption of the then-current operation or configuration of the chassis components.

Chassis management

Effectively managing hardware is important, especially for users who want to get the most effective use from their equipment. However, since it can be a time-consuming task, there are two methods of managing BladeCenter units. The first method makes use of the Web graphical user interface (GUI) that is integrated into the management module, allowing remote connection from any terminal connected to the same network as that of the management module. The second method of managing units is with IBM Director.

Web GUI

The Web GUI of a management module can be accessed by opening the Web browser on a terminal that has access to the BladeCenter system via the network. One would then type the IP address of the management module into the address or Uniform Resource Locator (URL) field and proceed to log in to the management module. The management module is divided into four main sections for easy navigation [16, 17]: monitors, blade tasks, switch tasks, and management module control.

Monitors

The monitors section of the management module enables one to view the status, settings, and other information for each of the key components within a system. Listed are the monitor components and their function:

- System status, which allows viewing of the overall status of all chassis components.

- Event log, which allows the user to view any entries that are currently stored in the management module event and error log.
- LEDs: The information LED is turned on when a noncritical event occurs and is logged in the error log; the location LED identifies the chassis and rack location of the blade server where the event has occurred. The LEDs are turned off by the user after appropriate actions are taken.
- BladeCenter system unit and an individual blade by controlling the identity LED.
- Vital product data (VPD), which allows viewing of the VPD data of all blades and modules. The VPD includes the slot number, serial number, machine type, and firmware of each component, along with other important information.

Blade tasks

The blade task section is used primarily to control and change the settings or configurations of blades. The choices under this section are the following:

- Power/restart (power on/off, restart blade, enable/disable local power control).
- Enable/disable wake on LAN, reset blade BMC.
- Remote control (remote control status, redirect console).
- Disable local KVM and local media tray switching.
- Firmware update.
- Configuration (blade information such as name and policy settings).

Switch tasks

The choices in the switch tasks section are used to manage and change the settings or configurations of the switch modules in the BladeCenter system. There are two choices in this section: power/restart and management. Power/restart allows resetting the switch modules and turning them on or off. Management allows viewing or changing the IP configuration of the switch modules.

Management module control

The management module control section is used to change the settings or configurations of the management module. This includes the following choices:

- General setting, which allows setting the name, date, and time.
- Creating login profiles, which allows setting authority, access rights, and passwords.
- Alerts management, which allows management of various types of alert settings and status.
- Network interfaces, which allows configuring the management module Ethernet connections for a remote console and for communicating with the switch modules.

Real-time diagnostics

Real-time diagnostics (RTD) is a manageability tool that runs on a BladeCenter system to help prevent or minimize downtime, thereby increasing system availability. Before RTD, it was necessary to shut down a server to run diagnostics on it. This could greatly disrupt availability and adversely affect the client's business productivity. RTD gives administrators the ability to run diagnostic utilities on servers while they are in operation, thus allowing administrators to proactively maintain their servers.

IBM Director

IBM Director is a comprehensive workgroup manager designed for use with IBM eServer servers, PCs, notebooks, and now BladeCenter systems. IBM Director includes support for BladeCenter systems, enabling users to manage, deploy, and monitor systems much more efficiently [17, 18]. It also includes features such as self-management and proactive and predictive tools, which provide higher levels of availability and reliability than does the Web GUI-based management capability. The IBM Director software consists of IBM Director Server, IBM Director Agent, and IBM Director Console.

A different combination of these components is required for each of the hardware groups in an IBM Director environment. The management server must contain all three of these components. IBM Director Console must be installed on the management console or any system from which a system administrator will remotely access the management server. IBM Director Agent must be installed on each system an administrator intends to manage.

IBM Director Server

IBM Director Server is the main component of IBM Director. The server component contains the management data, the server engine, and the application logic. IBM Director Server provides basic functions such as discovery of managed systems, storage of configuration and management data, inventory database, event listening, security and authentication, management console support, and administrative tasks.

IBM Director comes with the Microsoft Jet database engine; however, other database applications can be used in larger IBM Director management solutions. IBM Director Console and IBM Director Agent are automatically installed when IBM Director Server is installed. Every IBM eServer BladeCenter system comes

with an IBM Director Server license. An important new feature of IBM Director is its support for the BladeCenter system and its features. This includes new options under Groups and Tasks of IBM Director Console. In the Groups pane, users can detect BladeCenter systems and their configuration, which are then displayed under the Group Contents pane in the center of the console. In the Tasks pane is a new section headed *BladeCenter*, which has options for configuring, managing, and deploying IBM eServer BladeCenter systems [15].

IBM Director Agent

IBM Director Agent allows Director Server to communicate with systems on which it is installed. Director Agent provides the server with management data, which can be transferred using the TCP/IP, NetBIOS, and Internet Packet eXchange (IPX) protocols. IBM Director Agent Web-based access can be enabled only on Microsoft Windows** operating systems. All IBM BladeCenter processor blades come with an IBM Director Agent license. Additional licenses can be purchased for non-IBM systems.

IBM Director Console

IBM Director Console enables systems administrators to manage all systems that have an IBM Director Agent installed. This is done easily via the GUI by either a drop-and-drag action or a single click. Unlike Director Agent, Console and Server communicate and transfer data using TCP/IP. IBM Director Console does not require Director Agent to be installed unless an administrator wants to manage this system as well, in which case Director Agent must be installed separately. IBM Director Console does not require a license and can be installed on as many systems as needed.

Conclusion

The IBM eServer BladeCenter system provides IBM and its clients with many advantages over conventional server form factors. Blade server architecture has been available for decades, primarily in the telecommunications equipment industry. However, the application of blade servers in high-volume, mainstream, and enterprise server segments is a relatively new development that is profoundly relevant to IBM because these segments make up the overwhelming majority of IBM server, software, and services businesses. The BladeCenter system integration of networking and systems management also supports a critical role in data center infrastructure to deploy scale-out applications, such as clustering and Web server. It provides capabilities that were not previously available in scale-out environments. These are among the many reasons why the BladeCenter system has become the fastest-growing server product in IBM history.

*Trademark or registered trademark of International Business Machines Corporation.

**Trademark or registered trademark of Intel Corporation, Microsoft Corporation, Linus Torvalds, Novell, Inc., InfiniBand Trade Association, Teradyne, Inc., or PCI-SIG in the United States, other countries, or both.

References

1. Intel Pentium III and Pentium 4 Processor Family, Intel Corporation; see *http://www.intel.com/products/processor*.
2. *IEEE 802 Specifications,* Institute of Electrical and Electronics Engineers; see *http://www.ieee802.org/3/ap/*.
3. M. T. Chapman, "Introducing IBM Enterprise X-Architecture Technology," IBM Corporation white paper, August 2001; see *ftp://ftp.software.ibm.com/pc/pccbbs/pc_servers_pdf/exawhitepaper.pdf*.
4. *RFC 793* and *RFC 894 Specifications;* see *http://www.rfc-archive.org*.
5. *InfiniBand I/O Specification*, volumes 1 and 2, InfiniBand Trade Association; see *http://www.infinibandta.org/home*.
6. *PCI Bus Specification*, PCI Special Interest Group; see *http://www.pcisig.com*.
7. "RLX Server Blade Scalable Compute Clusters: A Technical White Paper," RLX Technologies, September 2002; see *http://whitepapers.zdnet.co.uk/0,39025945,60072680p-39000690q,00.htm*.
8. Egenera BladeFrame System, Gartner, Inc., ID number DPRO-103356, Gartner Group, April 2004; see *http://gartner.com/DisplayDocument?doc_cd=103356*.
9. "ProLiant BL p-Class System Overview and Planning White Paper," Hewlett-Packard Company, May 2003; see *http://whitepapers.zdnet.co.uk/0,39025945,60042483p-39000482q,00.htm*.
10. 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, and B. A. Trumbo, "BladeCenter Packaging, Power, and Cooling," *IBM J. Res. & Dev.* **49**, No. 6, 887–904 (2005, this issue).
11. 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).
12. 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).
13. IBM Corporation, BladeCenter Platform Design Specifications; see *http://www-1.ibm.com/servers/eserver/bladecenter/open_specs.html*.
14. W. G. Holland, P. L. Caporale, D. S. Keener, A. B. McNeill, and T. B. Vojnovich, "BladeCenter Storage," *IBM J. Res. & Dev.* **49**, No. 6, 921–939 (2005, this issue).
15. 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).
16. 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).
17. R. Credle, D. Brown, L. Davis, D. Robertson, and T. Ternau, "IBM eServer BladeCenter Systems Management," *IBM*

Redpaper, first edition, IBM Corporation, November 8, 2002; see *http://www.redbooks.ibm.com/abstracts/redp3582.html*.
18. 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).

Received January 2, 2005; accepted for publication March 14, 2005; Internet publication October 7, 2005

Dhruv M. Desai *IBM Systems and Technology Group, 3039 Cornwallis Road, Research Triangle Park, North Carolina 27709 (ddesai@us.ibm.com)*. Mr. Desai is a Distinguished Engineer in IBM eServer xSeries development, working as a BladeCenter system chief architect and strategist. He holds an M.S. degree in computer engineering from Nova Southwestern University and an M.S.E.E. degree from Texas A and M University. Mr. Desai has 24 years of experience in systems design and architecture of Microsoft Windows**/Intel-based systems. He holds 33 patents.

Thomas M. Bradicich *IBM Systems and Technology Group, 3039 Cornwallis Road, Research Triangle Park, North Carolina 27709 (bradic@us.ibm.com)*. Dr. Bradicich is the Chief Technology Officer, IBM xSeries/BladeCenter servers. Since 1998 he has directed the xSeries Architecture and Technology organization, where he is responsible for the architecture and technology of the IBM Intel-based server line. He received B.S. (1982) and M.S. (1983) degrees in electrical engineering and a Ph.D. degree (1996), all from the University of Florida. Since 1983, Dr. Bradicich has worked for the IBM Corporation in many engineering and management capacities. Prior to his current assignment, he directed the Technology Development organization in the Office of the CTO of the Personal Computer Division, and managed the strategic technology investment portfolio for the Intel-based desktop, mobile, workstation, home PC, and server lines. He co-founded and co-directed the IBM Personal Systems Institute, a technology management system for accelerating leading-edge developments from worldwide IBM Research Laboratories into IBM Intel-based product lines. He was named an IBM Distinguished Engineer in 2001 and was elected to the IBM Academy of Technology in 2004. Dr. Bradicich has served on the faculty at several universities; he is an adjunct professor in the College of Management and the College of Computer and Electrical Engineering at North Carolina State University. He is a member of the IEEE, the Computer Science Accreditation Board, and the National Science Foundation Center for Technology Commercialization Executive Board.

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.

William G. Holland *IBM Systems and Technology Group, 3039 Cornwallis Road, Research Triangle Park, North Carolina 27709 (wholland@us.ibm.com)*. Mr. Holland is a Senior Technical Staff Member working in BladeCenter development. Having provided technical guidance for the iSCSI, InfiniBand, and Fibre Channel BladeCenter options, he has ongoing responsibility for the BladeCenter I/O architecture and overall storage solution strategy. He received a B.S. degree in electrical engineering from Worcester Polytechnic Institute in 1984. Mr. Holland has worked in a number of roles at IBM, including circuit board tools development, S/390* processor logic design, worldwide product engineering manager for

S/390, PCI network adapter design, network router architecture and performance, and xSeries performance analysis. With this diverse experience base, he was an original member of the team that created and refined the BladeCenter design from 1999 until it first shipped in 2002. Mr. Holland has been awarded 11 patents.

Benjamin M. Kreuz *IBM Systems and Technology Group, 11400 Burnet Road, Austin, Texas 78758 (bmkreuz@us.ibm.com).*
Mr. Kreuz is currently the Power, Packaging, and Cooling (PPP&C) team leader for the JS20 family and LS20 blade products. He previously worked on other xSeries, pSeries*, and storage eServer products within the IBM Systems and Technology Group. He joined IBM as an engineering co-op student in Austin in 1997, and became a regular employee after graduating from the University of Texas at Austin in 2000 with a B.S. degree in mechanical engineering. Mr. Kreuz received an M.S. degree in engineering management from the National Technological University in 2004. He holds two patents and has one patent pending.