System Packaging Technologies for the K computer

● Hideki Maeda ● Hideo Kubo ● Hiroshi Shimamori ● Akira Tamura
● Jie Wei

The K computer ranked first on the TOP500 List of June 2011 and maintained its position atop the TOP500 List of November 2011. In addition, it took sixth place in the June 2011 edition of the Green500, which provides a ranking of supercomputers in terms of computational performance per unit of power. The achievement of such high performance and energy efficiency is due not only to the high-performance, low-power-consumption CPU but also largely to the system packaging technologies: rack technology to allow high-density mounting of CPUs, connection technology to achieve high-speed data transmission between CPUs, cooling technology for improved reliability and power supply technology to reduce power loss. This paper describes the system packaging technologies applied to the K computer.

1. Introduction

The K computer[note)] is attracting attention because of its high computational performance and execution efficiency achieved by its high-performance CPU and Tofu interconnect architecture. However, the K computer, which uses over 80 000 CPUs, also mainly has an excellent computational performance because of the system packaging technologies that have given it a small size, high reliability, and low power consumption.

To build the K computer as a large-scale system, we have used the packaging technologies that Fujitsu has cultivated up to now as the basis to fine-tune technologies encompassing individual elemental technologies and the entire system. We have developed system packaging technologies that help produce the world's fastest computers.

As system packaging technologies incorporated in the K computer, this paper presents those relating to its rack, connections, cooling and power supply.

2. Rack technology

This section describes the rack technology used in the K computer.

2.1 Entire rack

As shown in **Figure 1**, the K computer integrates a total of 102 CPUs including 24 system boards (SBs), each with four CPUs, and six IO system boards (IOSBs), each with one CPU. This is over three times the number of CPUs installed in Fujitsu's existing systems. This high-density mounting has been achieved by adopting hybrid cooling, which uses liquid cooling for components that generate a lot of heat such as the CPUs, and air cooling for the memory. These cooling methods will be described later in the section on cooling technology.

note) "K computer" is the English name that RIKEN has been using for the supercomputer of this project since July 2010. "K" comes from the Japanese word "Kei," which means ten peta or 10 to the 16th power.

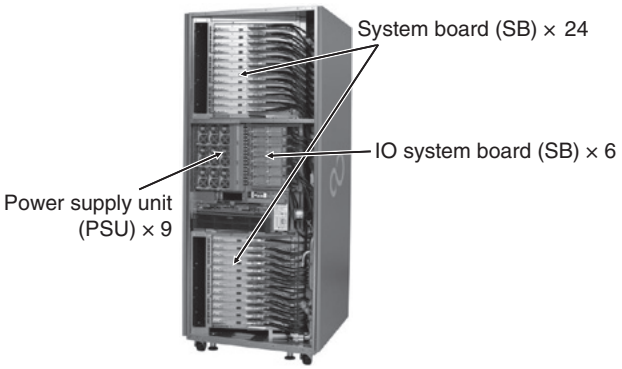

Figure 1
Rack structure.

Due to the high-density mounting, the rack weight has reached 1000 kg even without the interconnect cables (Tofu cables) that connect the racks. This weight is approximately 1.5 times the weight of the existing systems. To deal with this, we have developed new casters and a new rack base. In addition to the weight, heavy SBs are installed in an elevated location and this has the unwanted effect of raising the K computer's center of gravity, making it very difficult to address vibration. We have made full use of simulations to take appropriate measures and realized a structure resistant to earthquakes of 1000 Gal (at a level of 6 upper on the Japan Meteorological Agency [JMA] seismic intensity scale that goes from 0 to 7), which is the same as the existing systems.

2.2 SB slanted mounting

To cool the SBs, we have adopted liquid cooling for components that generate a lot of heat such as the CPUs and air cooling for the DIMMs and other components. For that reason, we needed to secure locations for the air intakes/vents and fans to ensure enough cooling air flow and we also needed to fit the system with manifolds for liquid cooling. We have developed a structure in which SBs are mounted in a slanted manner, and this has helped us to achieve high-density mounting (**Figure 2**).

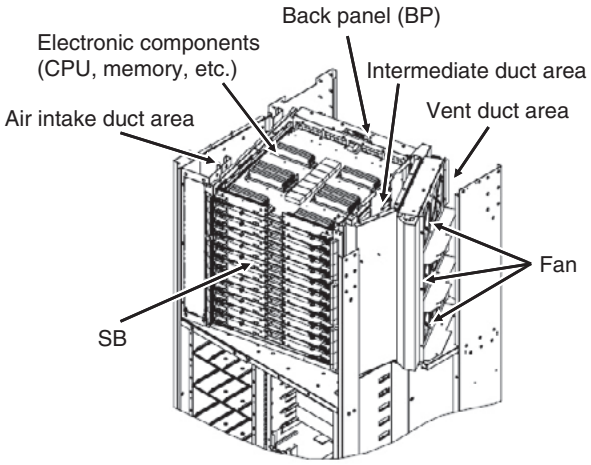

Figure 2
SB slanted mounting.

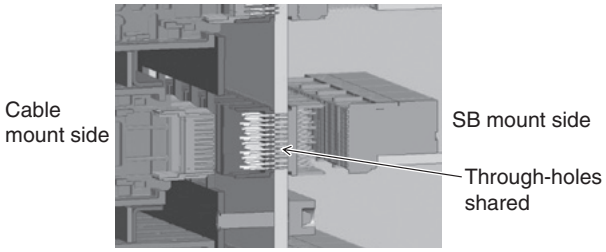

Figure 3
BP connector connection.

2.3 System connection and cable mounting

To minimize the length of the connections between CPUs, a structure has been adopted in which Tofu cables are directly implemented on back panels (BPs) to which SBs are mounted. The BPs have a structure in which connectors used for SB connection on the front side and the connectors for cables on the back side are shared using through-holes (**Figure 3**), minimizing transmission loss and signal reflection.

BPs are covered by a cable-retaining shelf that doubles as a radio shield and Tofu cable mounting is locked with one touch, which provides a structure that prevents electrostatic discharge (ESD) and also securely retains the cables (**Figure 4**).

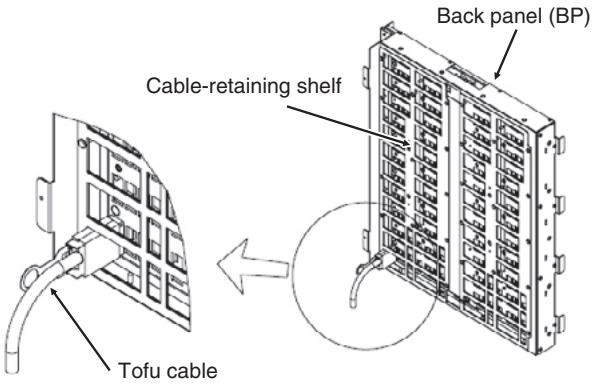

Figure 4
Cable connection to BP.

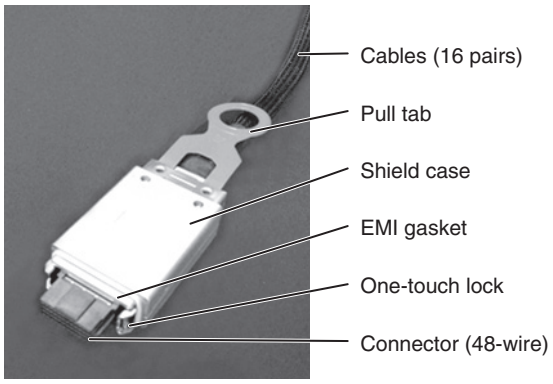

Figure 5
Tofu cable assembly.

3. Connection technology

3.1 Tofu cable

The Tofu cable has been newly developed to be high speed and also robust so that even if it is subject to external forces when being laid, its characteristics do not change. In addition, controlling the consistency of electrical characteristics over the lengthwise direction in manufacturing was carefully considered for the development.

A low-dielectric insulating material expanded polytetrafluoroethylene (ePTFE) has been used as the cladding of wire and a special shielding structure has been adopted that satisfies electric symmetry while arranging the drain and differential signal lines adjacently. This has allowed us to increase the density of the core wires and makes it easier to assemble a high-speed connector to the end of the cable while maintaining high-speed transmission characteristics. In addition, managing manufacturing parameters with dedicated cable manufacturing facilities and automating part of the process for assembling have allowed us to mass-produce products with minimized transmission loss and skew variation. In particular, the Tofu cable has the world's best performance as an interconnect cable connecting racks, with a skew between positive and negative components of differential signal of 30 ps/10 m max., and it contributes to the high transmission quality of the K computer.

Figure 5 shows the appearance of the Tofu cable assembly.

The connector is small with a maximum area of 34 mm × 14 mm and is provided with a one-touch locking mechanism that keeps the high-density mounting from affecting the working efficiency. The mounting pitch of the cable assembly is 14.45 mm, which gives very high dense cabling from the rack.

In general, high-speed cables have many inspection parameters and considerable time is required to scan all signals. For the K computer, we needed to manufacture numerous cables in a short period of time and a major challenge we faced was how to reduce the test time. To that end, we correlated the characteristics of the time and frequency domains with high accuracy in place of conducting simulations from the frequency domain characteristics for some items that need to undergo time domain inspection. In this way, we reduced the test time to approximately one-tenth of the conventional time required.

3.2 Tofu cable management

The K computer needs to have high-density connections via numerous cables over the shortest possible distance. For that purpose, we closely

studied the forming technology and cable-laying method in advance to ensure the optimum cable routing and space in view of the accommodation volume efficiency and extra length.

The number of cables connected to one BP is about 200 and the number of cables per rack consisting of two BPs at the top and the bottom is over 400. Cables are wired separately for the upper and lower parts of a rack and the spaces in the ceiling and under the floor have been efficiently used to achieve the shortest possible wiring between racks. The number of cables used for the entire system is about 200 000 and the total laying length reaches approximately 1000 km.

Figure 6 shows an example of the cable wiring in a ceiling.

4. Cooling technology

For the K computer, we have adopted a hybrid cooling structure, which combines a liquid cooling system featuring high efficiency and high reliability and an air cooling system that offers high efficiency and low cost.[1] The following sections describe this cooling system.

4.1 Liquid cooling system

In this section, the structure of a liquid cooling system for racks and SBs is explained.

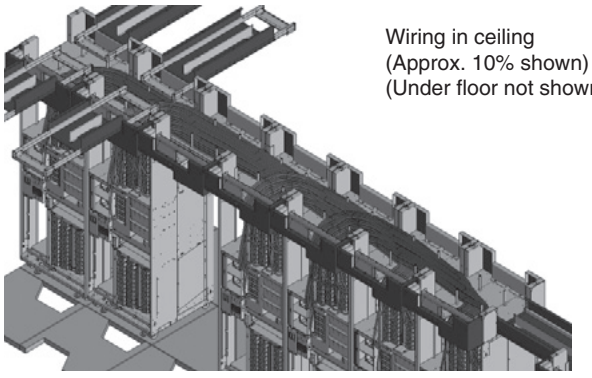

Figure 6
Cable wiring in ceiling.

4.1.1 Liquid cooling system for racks

Figure 7 shows the structure of the liquid cooling system for each rack. Two manifolds (for liquid supply and return) are vertically installed on the right side on the front, from which hoses for liquid supply and return (24 hoses each for SB and six hoses each for IOSB) branch off. In addition, these liquid supply and return manifolds are equipped with sensors (liquid temperature, pressure and dew sensors) and controlling devices (air vent valve, motor valve, check valve and filter). These are all incorporated into two manifolds, which ultimately constitute a liquid cooling system of two manifold sections: liquid supply and return.

To achieve a balance between liquid supply and return, we have designed the system so that liquid is supplied and returned through the center with reference to the length of the manifolds, the diameters of manifold pipes in the flow path system are controlled and the lengths of the liquid supply and return hoses are equal within a range that balances any pressure drops. As a result, the pressure drops between liquid supply and return have been successfully reduced to within ±5%. In installing the manifolds, we have taken thorough space-saving measures to fit the liquid supply and return manifolds into a very small space in the right-hand side corner on the front side of the rack.

In this way, we have designed a simple manifold system capable of controlling the flow rate without using special parts such as a flow regulator, thereby ensuring the utmost quality of the manifolds itself and realizing a high-efficiency, high-reliability system.

4.1.2 Liquid cooling system for SBs

The CPUs and InterConnect Controller (ICC) packages mounted on SBs and some power supply conversion devices (DC-DC converters) are cooled by liquid cooling units (LCUs). An LCU is composed of two or more cooling plates (CPs) to cool the CPUs and ICCs, which are

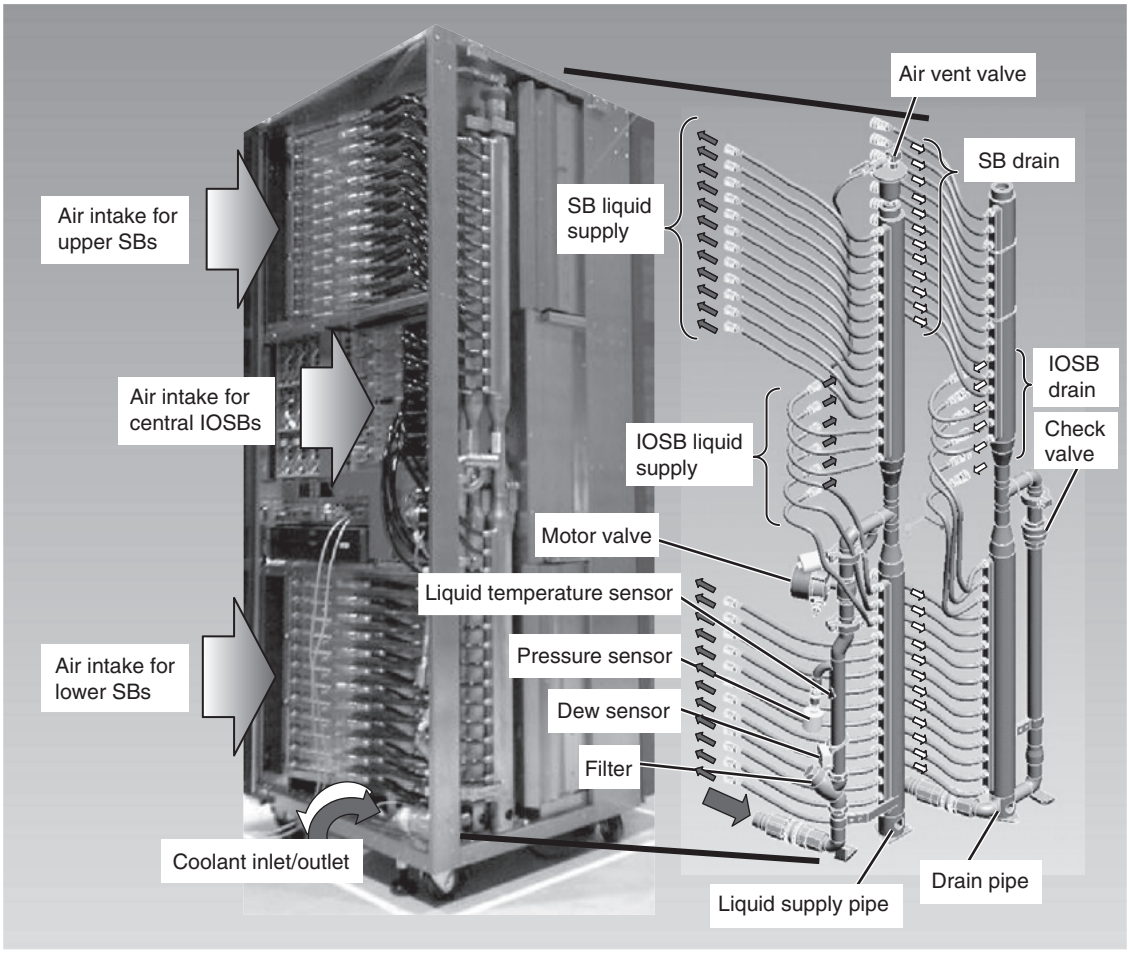

Figure 7
Structure of liquid cooling system for rack.

joined by piping lines in two parallel branches (**Figure 8**). In view of maintainability, coolant is supplied to and returned from the LCU by liquid couplers provided on the front side of an SB.

The K computer is a system composed of over 20 000 SBs and efficient cooling with a small amount of circulating coolant has been a challenge. To that end, the LCU has been designed to have eight CPs connected by piping lines in two parallel branches of four CPs to feed an equal flow rate of coolant to each CP. In addition, a mini-channel flow path structure has been adopted inside the CP to minimize the flow of circulating coolant (**Figure 9**). Flow loss of the coolant has been reduced by taking these measures and the coolant flow variation in all CPs in the rack has been successfully controlled

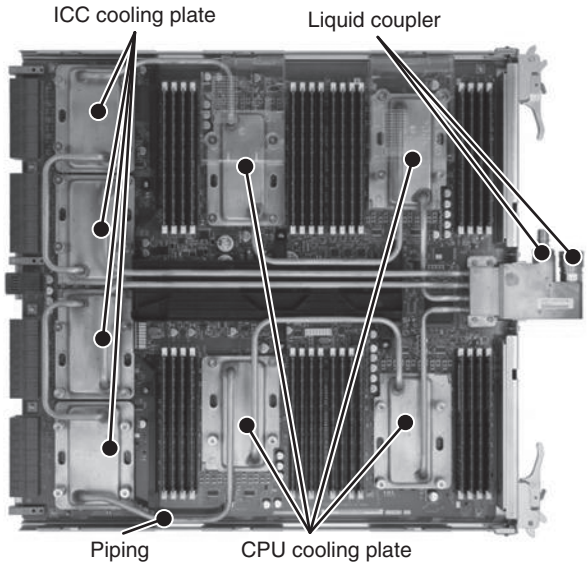

Figure 8
Liquid cooling unit (LCU) on SB.

within ±3% by adjusting the CP internal flow path structure. To mount on an SB an integrated LCU including connections by piping, a structure capable of absorbing the variation in the size is required. We have elaborated the size and shape of copper piping used to join CPs to devise a structure that minimizes the influence of dimension errors and stress in mounting. This integrated structure has significantly helped to improve the strength and reliability of the LCUs in addition to reducing the cost. Furthermore, reduction in height and weight has allowed us to give the system high-density mounting.

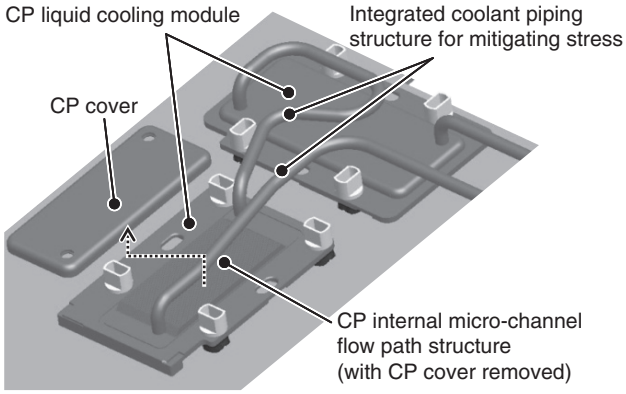

Figure 9
Cooling plate (CP) structure.

4.2 Air cooling system

The air cooling system of the K computer has a characteristic rack structure. As mentioned in the section of "Rack Technology," a shelf is installed aslant in a rack and twelve SBs mounted on the shelf are air cooled by using six (5 + 1 redundancy) fans. This form of installation has been adopted to secure the areas for liquid cooling piping and Tofu cables and minimize loss of SB cooling air while ensuring the maintainability of BPs and fans and connectability of liquid couplers for liquid supply to and discharge from SBs. This slanted mounting has reduced the cooling air loss to one-half that of an upright mounting shelf while ensuring maintainability. The number of fans has also been halved by combining a cooling air backflow prevention mechanism installed near the fans.

Figure 10 shows the SB air cooling structure and **Figure 11** the flow of the cooling air. An SB has dual inline memory modules (DIMMs) and some power supply components to be air-cooled. Cooling air for these components enters diagonally from the front of the SB, cools the components and is discharged diagonally at

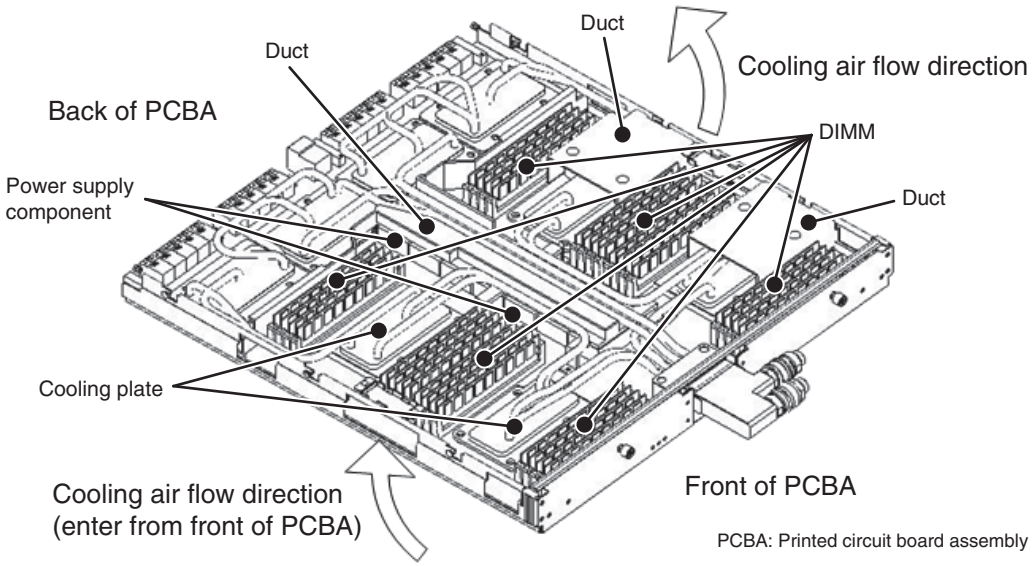

Figure 10
SB air cooling structure.

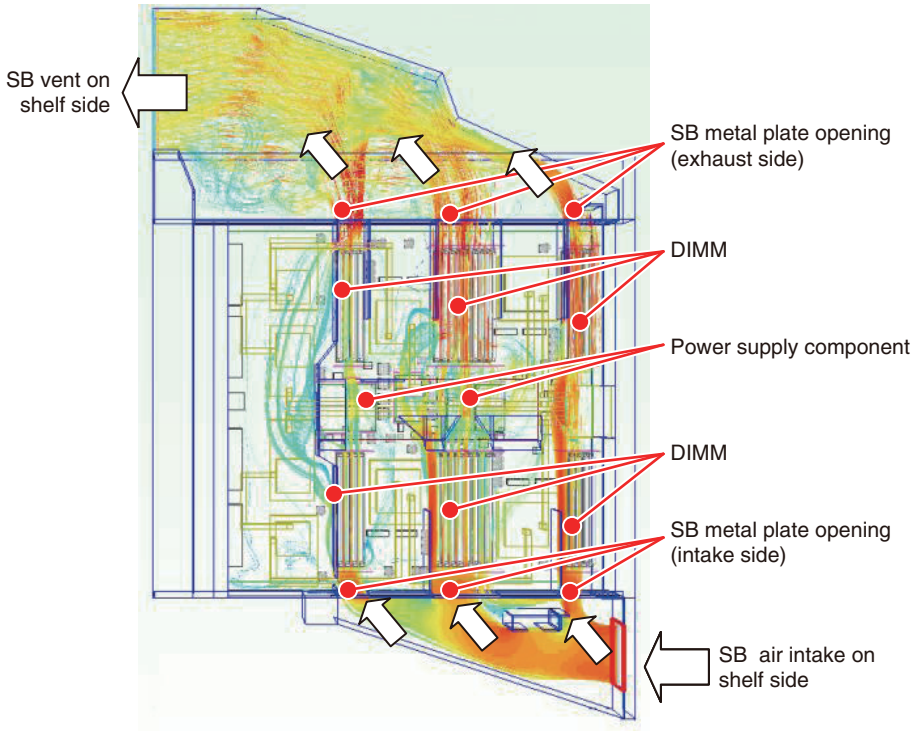

Figure 11
Flow of SB cooling air (top view).

the back. We needed to make use of this twisted flow of cooling air and give consideration to LCUs that block the cooling air flow to supply the optimum cooling air. We have made full use of thermo-fluid analysis to determine the SB structural parts and duct shapes. To ensure and control the air flow required for cooling, we have optimized the shapes and locations of the openings in the metal plates on the cooling air supply and exhaust sides and the shapes and arrangement of ducts in SBs to distribute cooling air among the components to be cooled in a well-balanced manner. By taking these measures, temperature variation of DIMMs and other components has been minimized to realize an efficient cooling system.

5. Power supply technology

This section describes the structure, power conversion efficiency and reliability of the power supply system used for the K computer.

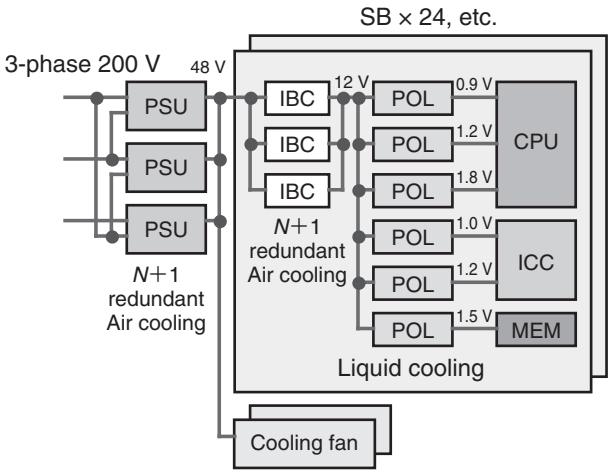

Figure 12
Power supply system configuration.

5.1 Power supply system configuration

Figure 12 shows a schematic diagram of the power supply system configuration.

The system receives three-phase 200 V power and distributes it to single-phase 200 V input power supply units (PSUs) (**Figure 13**).

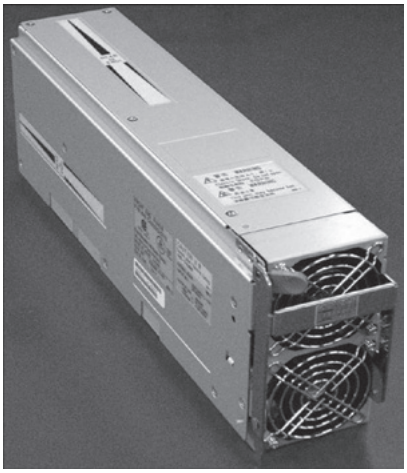

Figure 13
Appearance of PSU.

Each PSU supplies 48 V power to the SBs and cooling fans while controlling the output currents to keep them equal. On the SBs, each intermediate bus converter (IBC) transfers 48 V to 12 V while performing current balancing control and supplies power to each discrete non-isolated DC-DC converter point of loads (POLs). The individual POLs convert the 12 V power to supply 0.9 V and other low-voltage, large-current power to CPUs and other loads in the final stage. To cool the power supply system, air cooling is used for the PSUs and IBCs and liquid cooling is used for the POLs.

5.2 Power conversion efficiency of power supply system

The specifications of each type of power supply are shown in **Table 1**.

The power conversion efficiency of PSUs has been improved to 91% from the conventional 88%. We have achieved this by adopting new power devices, making the power conversion circuit more sophisticated, and improving the packaging technology. In particular, reducing major switching losses such as that from recovery current by using silicon carbide (SiC) diode and super junction (SJ) FET, which are new power devices featuring high switching speed and low ON-resistance, has had a significant effect. As a result, output of 3000 W, which is 1.5 times that of Fujitsu's existing rack with a 2000 W output, has been successfully implemented in high density in a rack of the same size.

Table 1
Power supply specifications.

	PSU	IBC	POL
Input voltage (V)	200–240	48	12
Output voltage (V)	48	12	0.9–3.3
Output current (A)	60	20	4–120
Dimensions (mm)	70 × 123 × 325	35 × 58	35 × 50

One characteristic of the power supply system on each SB is that the conventional isolated DC-DC converters have been separated into IBCs and POLs and the POLs have been located in the immediate vicinity of loads such as CPUs. This led to a significant reduction of power loss in printed circuit board (PCB) patterns in the SBs. The power conversion efficiency resulting from combining IBCs and POLs (0.9 V) has been improved to 84% from 80% with the conventional isolated DC-DC converters. Another major feature is that improved response of POLs and their location in the immediate vicinity of loads have decreased the capacitance of decoupling capacitors for loads to about one-tenth that of existing systems.

Technological development as described above has successfully increased the power conversion efficiency of the entire power supply system from the conventional 70% to 76%.

5.3 Reliability of power supply system

For a large-scale system like the K computer, reliability of the power supply system is an important factor because massive amounts of numerical computations are continued for an extended period of time. With this power supply system, $N + 1$ redundancy has been used for the PSUs and IBCs, which has dramatically

improved the system's reliability. Although POLs do not have a redundant configuration, the system of cooling together with CPUs by using liquid cooling CPs has been adopted to maintain a low junction temperature of around 40°C to ensure sufficient reliability.

6. Conclusion

This paper has given specific examples to present the system packaging technologies including rack, connection, cooling and power supply technologies that have allowed us to give the K computer high density and high efficiency.

Demand for high-density, high-efficiency systems is rising for server devices in general as well as for supercomputers. We intend to continue to work on technological developments that meet this demand.

References

1) J. Wei: Hybrid Cooling Technology for Large-Scale Computing Systems. Proceeding of ASME InterPACK2011, Portland, Oregon, USA (July 2011).

Hideki Maeda
Fujitsu Ltd.
Mr. Maeda is currently engaged in development of structural technology for implementing server devices.

Hideo Kubo
Fujitsu Ltd.
Mr. Kubo is currently engaged in development of cooling technology for server devices.

Hiroshi Shimamori
Fujitsu Ltd.
Mr. Shimamori is currently engaged in development of power supply technology for server devices.

Akira Tamura
Fujitsu Advanced Technologies Ltd.
Mr. Tamura is currently engaged in development of elemental technology for implementing server devices.

Jie Wei
Fujitsu Advanced Technologies Ltd.
Mr. Wei is currently engaged in development of cooling element technology for server technology.