

スーパーコンピュータ「京」の システム実装技術

System Packaging Technologies for the K computer

● 前田秀樹 ● 久保秀雄 ● 島森 浩 ● 田村 亮 ● 魏 杰

あらまし

スーパーコンピュータ「京」は2011年6月のTOP500リストで世界一を獲得し、2011年11月のTOP500リストで再び世界一を獲得した。また同時に、単位電力あたりの計算性能のランキングであるGreen500で2011年6月のランキングで6位を獲得した。この高性能・省電力の達成は、高性能・低消費電力のCPUだけでなく、CPUを高密度に実装する筐体技術、CPU間的高速データ伝送を実現する接続技術、信頼性を高める冷却技術、電力ロスを低減する電源技術のシステム実装技術によるところも大きい。

本稿では、「京」に適用したシステム実装技術について解説する。

Abstract

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.

まえがき

スーパーコンピュータ「京」^(注)は、高性能CPUやTofuインターコネクタアーキテクチャによる高い計算性能と実行効率が注目されている。しかし、CPUを8万個以上使った「京」がシステムとして高い計算性能を発揮できるのは、そのシステムを小型・高信頼・低消費電力に実現したシステム実装技術によるところも大きい。

今回「京」の大規模なシステムを実現するため、今まで培った実装技術をベースに個々の要素技術から全体システムに至るまで詳細に技術のチューニングを行い、世界最速を支えるシステム実装技術を開発した。

本稿では、「京」のシステム実装技術として、筐体・接続・冷却・電源の各技術について紹介する。

筐体技術

本章では、「京」で採用された筐体技術を紹介する。

● 筐体全体

「京」は図-1のようにCPUを4個搭載したシステムボード（以下、SB）24枚と、CPUを1個搭載したIOSB 6枚の合計102CPUを搭載している。これは当社の従来装置の3倍強の搭載数となっている。このような高密度実装はCPUなどの高発熱部品には水冷、メモリなどには空冷のハイブリッド冷却を採用して実現した。これらの冷却については後述の冷却技術の章で紹介する。

(注) 理化学研究所が2010年7月に決定したスーパーコンピュータの愛称。

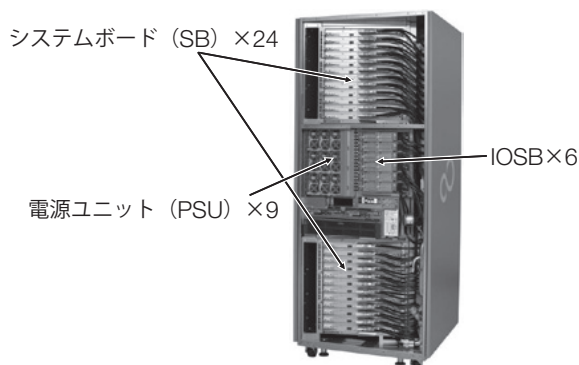


図-1 筐体構成

高密度実装により筐体質量は筐体間を接続するインターコネクタケーブル（Tofuケーブル）を除いても当社の従来装置の約1.5倍となる1000 kgとなった。そのため、キャスターや筐体ベースを新規に開発して対応した。また、質量の大きなSBを高所に搭載しているため、質量が大きいだけでなく重心が高くなり、振動対策が非常に難しくなったが、シミュレーションを駆使して対策を実施し、従来装置と同様に1000 Gal（震度6強レベル）の地震に耐えられる構造となっている。

● SB斜め実装

SBは、CPUなどの高発熱部品には水冷を採用したが、DIMMなどには空冷を採用した。そのため、冷却風を確保する吸排気口、ファンの場所を確保し、更に水冷配管を配置する必要があった。今回、SBを斜めに実装する構造を開発し、高密度実装を実現した（図-2）。

● システム接続とケーブル実装

CPU間の最短距離接続を実現するために、TofuケーブルはSBを搭載したバックパネル（以下、BP）に直接実装する構造を採用した。BPでは、表側のSB接続に使用されるコネクタと、背面のケーブル用コネクタをスルーホールで共有する構造になっている（図-3）。この構造により、伝送ロスや信号反射が最小に抑えられている。

また、BPは電波シールドを兼ねたケーブル保持シェルフで覆われており、Tofuケーブル装着時はワンタッチでロックされる方式で、静電対策と強度的な保持を両立させた構造を採っている（図-4）。

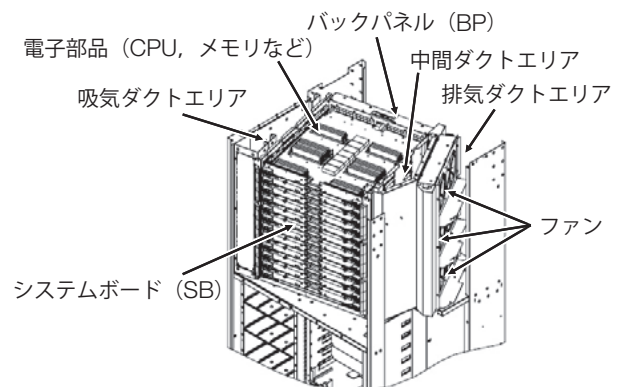


図-2 SB斜め実装

接続技術

● Tofuケーブル

Tofuケーブルは、敷設時の外力による特性変化を受けない堅ろう性と高速性の両立、更にはケーブル量産製造時における長手方向での特性安定性を考慮して新規に開発を行った。

線材の被覆には低誘電率の絶縁体ePTFEを採用し、ドレイン線を差動信号の横に配置しつつ、電気的な対称性を満足するよう特殊なシールド構造を採用した。これにより、高速伝送特性を維持した上で、芯線の高密度化と、優れた端末加工性を達成した。また、専用ケーブル製造設備での製造パラメータ管理や、一部の端末加工工程を自動化するなどにより、伝送ロスとスキューのばらつきが非常に小さい製品の大量生産を可能とした。特に、筐体間を接続するインターコネクトケーブル（Tofuケーブル）として差動信号のポジネガ間スキューは30 ps/10 m以下と世界最高性能を誇り、「京」の高い伝送品質に貢献している。

TofuケーブルAssyの外観を図-5に示す。コネクタの最大外形は34 mm×14 mmと小型で、作業性

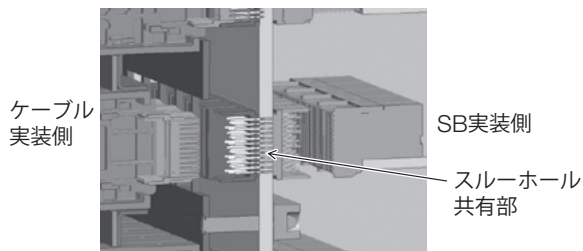


図-3 BPでのコネクタ接続

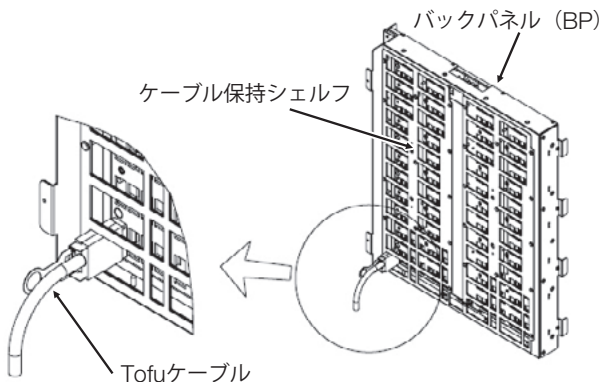


図-4 BPへのケーブル接続

を犠牲にしない高密度搭載を可能とするワンタッチロック機構を備えている。ケーブルAssyの搭載ピッチは14.45 mmと高密度である。

一般に、高速ケーブルは検査パラメータが多く、全信号数をスキャンするためには非常に多くの時間を要する。「京」は、大量のケーブルを短期間で生産する必要があり、試験時間の短縮が大きな課題であった。そのため、タイムドメインと周波数ドメインの特性間のコリレーションを高精度に行い、タイムドメイン検査項目の一部を周波数ドメイン特性からのシミュレーションに置き換えることにより、従来比、約1/10に試験時間を短縮した。

● Tofuのケーブルマネジメント

「京」では、大量のケーブルを最短ルートで高密度に接続する必要があり、そのためのフォーミング技術や敷設方法を、あらかじめ詳細に検討し、収容体積効率や余長を考慮した最適なケーブルルートとスペースを確保した。

1枚のBPに接続されるケーブル本数は約200本で、上下2枚のBPで構成される1筐体あたりのケーブル本数は400本を超える。ケーブルは、筐体の上下で配線が分かれ、天井と床下のスペースを効率的に活用して、筐体間の最短配線を実現した。システム全体で使用するケーブルは約20万本で、総敷設長は約1000 kmに及ぶ。

ケーブル配線例を図-6に示す。

冷却技術

「京」では高効率・高信頼を実現した水冷システムと、高効率・低コストを実現した空冷システムを組み合わせたハイブリッド冷却システムを採用

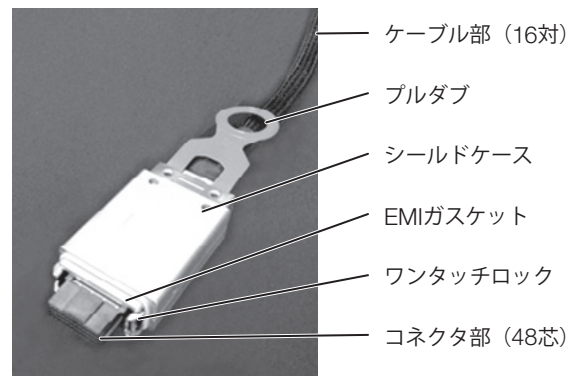
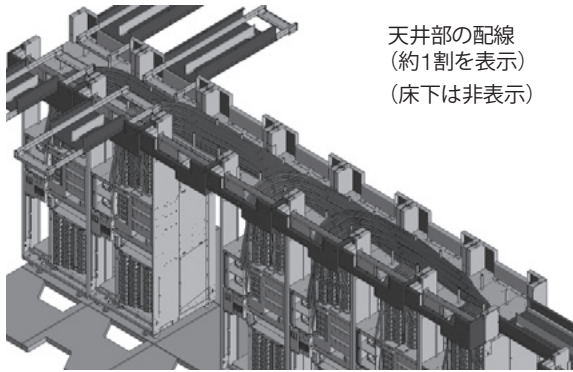


図-5 TofuケーブルAssy

した⁽¹⁾以下、この冷却システムについて述べる。

● 水冷システム

本節では、筐体側とSBにおける水冷システムの構成について述べる。



天井部の配線
(約1割を表示)
(床下は非表示)

図-6 システム間のTofu接続

(1) 筐体の水冷システム

筐体側の水冷システムの構造を図-7に示す。前面右側に2本の配管（給水側、排水側）を縦に実装しており、その配管から分岐した給排水用のホース（SB給排水ホース各24本、IOSB給排水ホース各6本）が出ている。そのほか、この給排水管には、センサ（水温センサ、圧力センサ、結露センサ）や、給排水調整部品（空気抜き弁、電動弁、逆止弁、フィルタ）を保有している。これらを全て2本の配管系に収め、最終的には給水配管と排水配管の二つの配管系部品で構成される水冷システムとした。

給排水において、給水側と排水側をバランスさせるため、配管長さ方向に対し中央から給排水を行い、流路系統内の配管径をコントロールし、更に給排水のホースも損失をバランスする範囲内で均等長さに設計した。その結果、給排水の損失差を±5%以内に抑えた。また配管の実装に関しては、

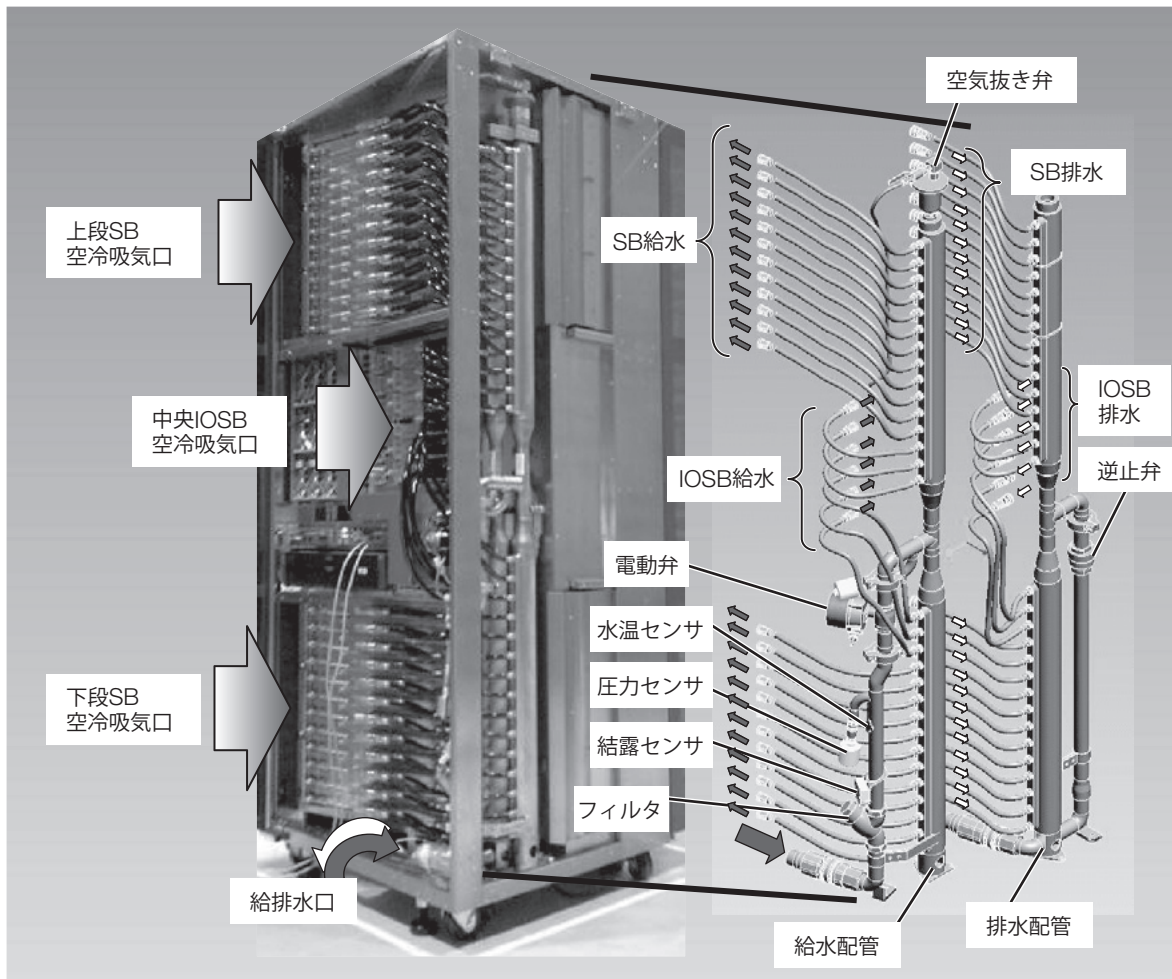


図-7 筐体水冷システムの構造

徹底的な省スペース化を行い、筐体前方右コーナーの極小スペースに給水配管・排水配管を収めた。

このように流量調整弁などの特別な部品を用いることなく流量をコントロールするシンプルな配管システムを設計したことにより、配管単体での徹底した品質確保と高効率・高信頼なシステムを実現した。

(2) SBの水冷システム

SBに実装されたCPUとICC (Interconnect Controller) パッケージ、および一部の電源変換素子 (DC/DCコンバータ) は水冷ユニット (Liquid Cooling Unit : LCU) で冷却している。水冷ユニットはCPUおよびICC冷却用の複数のクーリングプレート (Cooling Plate : CP) により構成され、それぞれを並列2系統の配管経路で連結している (図-8)。水冷ユニットへの冷却水の給排水は、保守性を考慮し水冷ユニットのSB前方に設置している液体コネクタ (Liquid Coupler) を介して行う。

「京」は2万枚以上のSBで構成されるシステムのため、循環水量を抑え効率良く冷却することが課題であった。このため、水冷ユニットは八つのCPのうち四つずつを並列2系統の配管経路で接続し、均等流量の冷却水を各CPに流す設計とした。さらにCP内部は冷却水循環流量を最小限に抑えるため微細チャンネル流路構造を採用した (図-9)。これらの工夫により冷却水の流動損失を抑えると同時

に、CP内部流路構造を調整することで筐体内全てのCPにおいて±3%以内に冷却水流量ばらつきを抑えることができた。また配管で連結、一体化した水冷ユニットをSBに取り付けるためには、それぞれの寸法ばらつきを吸収する構造が必要になる。CPを連結する銅配管の寸法や形状を工夫し、取り付け時の寸法誤差や応力を緩和する構造とした。この一体化構造により、コスト低減とともに水冷ユニットの強度、信頼性も大幅に強化された。さらに、薄型軽量化により、高密度実装も実現した。

● 空冷システム

「京」の空冷システムは特徴的な筐体構造となっている。筐体技術の章でも紹介したとおり、筐体内にシェルフを斜めに搭載し、シェルフに搭載された12枚のSBを6台 (5+1冗長) のファンで冷却している。水冷配管やTofuケーブルのエリアを確保し、BPやファンの保守性、SBへの給排水のための液体コネクタの接続操作性を確保しつつ、SB冷却風の損失を極力抑えるために採用した実装形態である。この斜め実装により、保守性を確保しつつ冷却風の損失を90度の実装シェルフに比較して1/2に削減している。また、ファン近傍に搭載した冷却風の逆流防止機構と合わせることで、ファン台数を1/2に削減できた。

図-10にSBの空冷構造、図-11に冷却風の流れを示す。SBには空冷対象部品としてDIMMおよび電源部品の一部が搭載されている。これらの部品の冷却風はSB斜め前方より入り、対象部品を冷却後、斜め後方へ排出される。このねじれた冷却風の流れを使用して、冷却風の流れを阻害する水冷ユニット

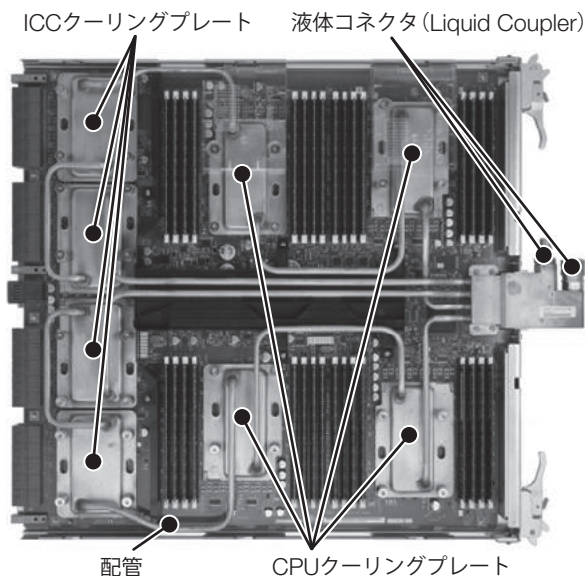


図-8 SB搭載水冷ユニット (LCU)

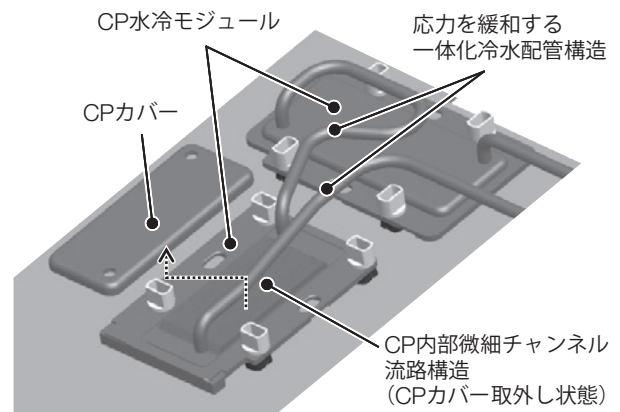


図-9 クーリングプレート (CP) の構造

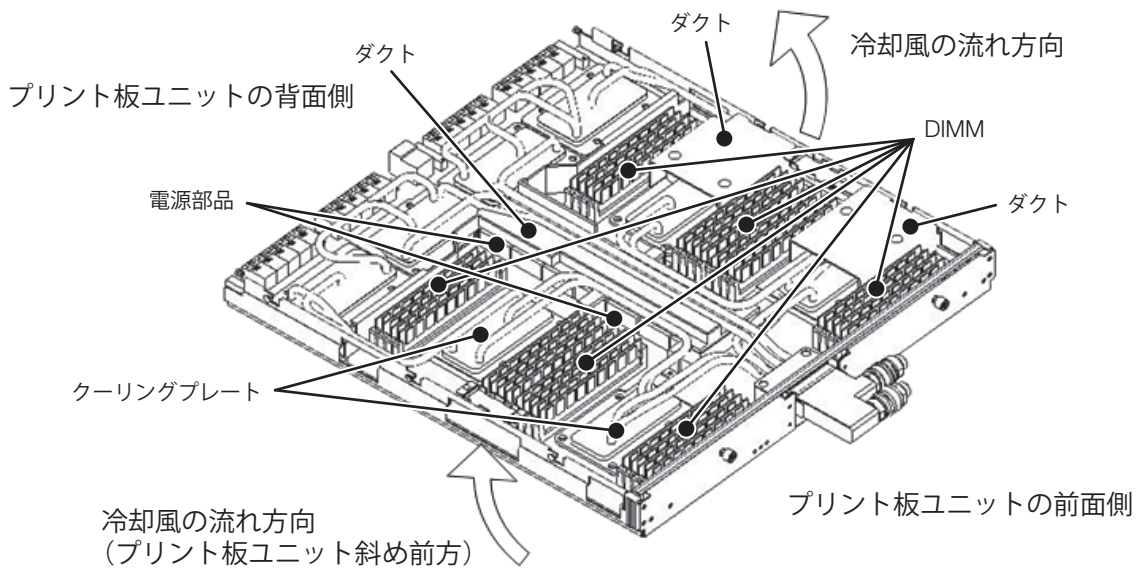


図-10 SBの空冷構造

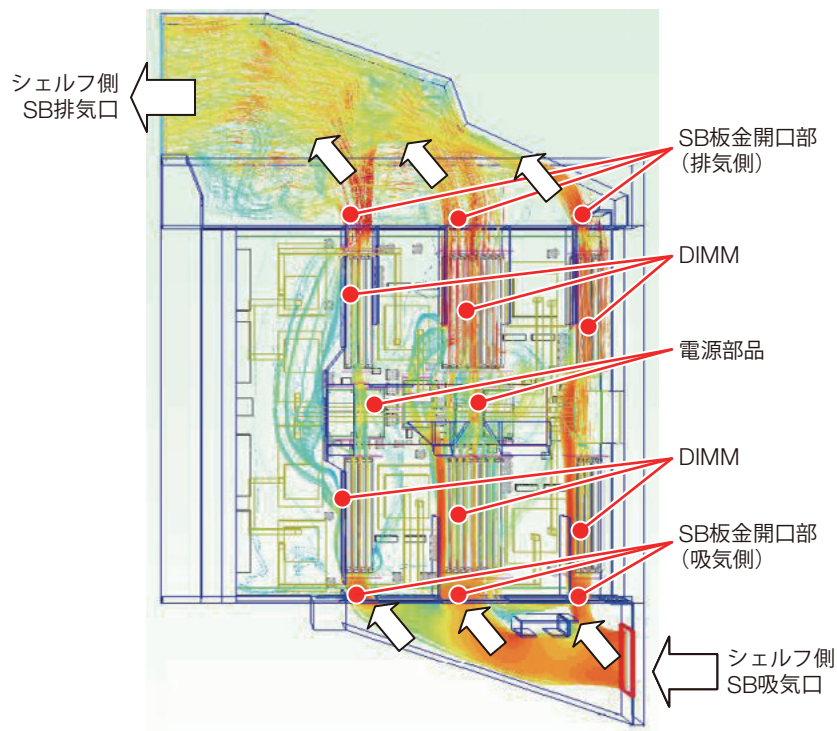


図-11 SB冷却風の流れ(上面視図)

トを考慮し、最適な冷却風を供給する必要があった。熱流体解析を駆使し、SBの構造部品やダクト形状を決定した。冷却に必要な風量を確保・制御するために、冷却風供給側と排気側の板金の開口形状と位置の最適化や、SB内ダクトの形状や配置を最適化し、対象部品にバランス良く冷却風を配分した。これらの工夫によりDIMMなどの温度ば

らつきを極力抑え、効率的な冷却システムを実現した。

電源技術

「京」で採用した電源システムの構成、電力変換効率および信頼性について述べる。

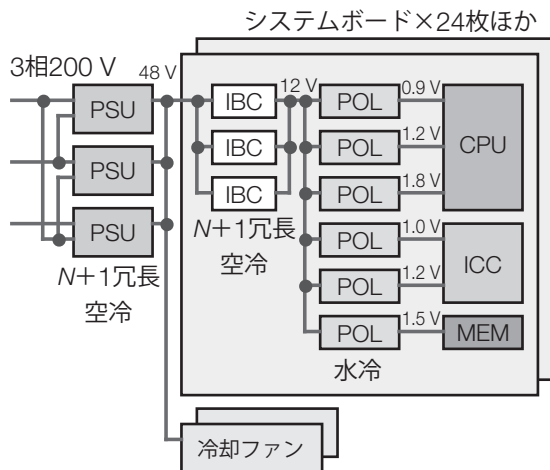


図-12 電源システムの構成

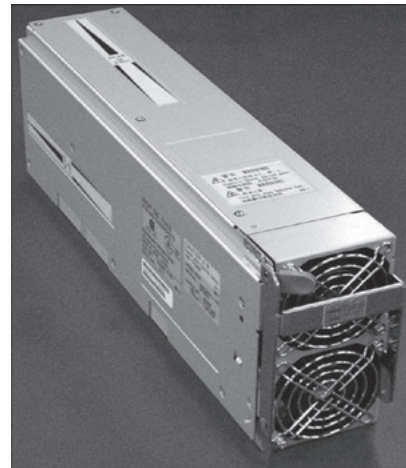


図-13 PSUの外観

● 電源システムの構成

電源システム構成の概略を図-12に示す。

装置は3相200Vを受電し、単相200V入力タイプの電源ユニット（PSU）に分配する（図-13）。各PSUは出力電流が均等になるように制御しながら、SB、冷却ファンなどに48Vを供給する。SBでは各中間バスコンバータ（IBC）が電流バランス制御を行いながら48Vを12Vに変換し、ディスクリットタイプの各非絶縁DC/DCコンバータ（POL）に供給する。そして各POLは12Vを変換し、最終段となるCPUほかの負荷に0.9Vなどの低電圧、大電流を供給する。電源システムの冷却は、PSUおよびIBCは空冷、POLは水冷を採用している。

● 電源システムの電力変換効率

各電源の仕様を表-1に示す。

PSUの電力変換効率を従来の88%から91%に向上させた。これは新電力デバイスの採用、電力変換回路の工夫、実装技術の進歩によるものである。特にスイッチングスピードが速く低ON抵抗の新しい電力デバイスであるシリコンカーバイドタイプのダイオードとスーパージャンクションFETを採用してリカバリー電流などの主要なスイッチング損失を低減した効果が大きい。その結果、従来2000W出力であった同一の筐体サイズ内に1.5倍の3000W出力を高密度実装することができた。

SB内の電源システムの特徴は、従来の絶縁型DC/DCコンバータをIBCとPOLに分離し、POLをCPUなどの負荷の直近に配置したことである。これによりSB内のプリント板配線パターンの電力損

表-1 各電源の仕様

	PSU	IBC	POL
入力電圧 (V)	200-240	48	12
出力電圧 (V)	48	12	0.9-3.3
出力電流 (A)	60	20	4-120
形状 (mm)	70×123×325	35×58	35×50

失を大きく削減できた。IBCとPOL（0.9V）を合わせた電力変換効率は、従来の絶縁型DC/DCコンバータが80%程度であったものを84%に向上させている。また、POLの高速応答化と負荷直近配置により負荷のデカップリングコンデンサ容量を従来の1/10程度に削減できたことも大きな特徴である。

以上の技術開発の結果、電源システム全体の電力変換効率は従来の約70%から76%にアップさせることができた。

● 電源システムの信頼性

「京」のような大規模システムでは、膨大な数値計算を長時間にわたって継続するため、電源システムの信頼性が重要な要素となる。本電源システムでは、PSUおよびIBCにN+1冗長方式を採用し、信頼性を飛躍的に向上させた。POLは冗長構成を採ってはいないが、水冷のCPでCPUと一緒に冷却する方式を採用し、ジャンクション温度を40℃程度の低温にすることにより十分な信頼性を確保することができた。

む す び

本稿では、スーパーコンピュータ「京」の高密度・高効率を実現したシステム実装技術として筐体・接続・冷却・電源の各技術について具体例を交えて紹介した。

スーパーコンピュータだけでなくサーバ装置全体で高密度・高効率システムの要求が高まって

おり、今後もその要求に応える技術開発を行っていく。

参考文献

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

著者紹介



前田秀樹 (まえだ ひでき)

エンタプライズサーバ事業本部実装技術開発統括部 所属
現在、サーバ装置の実装構造技術開発に従事。



田村 亮 (たむら あきら)

富士通アドバンステクノロジー(株) 実装技術開発統括部 所属
現在、サーバ装置の実装要素技術開発に従事。



久保秀雄 (くぼ ひでお)

エンタプライズサーバ事業本部実装技術開発統括部 所属
現在、サーバ装置の冷却技術開発に従事。



魏 杰 (うえい じえ)

富士通アドバンステクノロジー(株) HPC適用推進統括部 所属
現在、サーバ装置の冷却要素技術開発に従事。



島森 浩 (しまもり ひろし)

エンタプライズサーバ事業本部実装技術開発統括部 所属
現在、サーバ装置の電源技術開発に従事。