

Технология компоновки системы для К компьютера

•Hideki Maeda •Hideo Kubo •Hiroshi Shimamori •Akira Tamura •Jie Wei
FUJITSU Sci. Tech. J., Vol. 48, No. 3, pp. 286–294 (July 2012)
перевод © ООО «Модуль-Проекты», <http://www.mdl.ru>, ссылки обязательны.

К компьютер занимает первое место в списке TOP500 June 2011 и сохранил свои позиции на вершине списка TOP500 November 2011 (N4 TOP500 November 2013- прим. перев.). Кроме того он занял шестое место в июне 2011 в редакции Green500, которая поддерживает рейтинг суперкомпьютеров с точки зрения вычислительной производительности на единицу мощности. Достижение такой высокой производительности и эффективного энергопотребления обусловлено не только показателями высокой производительности и низкого энергопотребления процессоров, но и в значительной степени технологиями компоновки системы: стоечные технологии позволяют высокую плотность монтажа процессоров, технологии связи достигают высокой скорости передачи данных между процессорами, технологии охлаждения улучшают технологии надежности и энергоснабжения для уменьшения потери мощности. Данная статья описывает технологии компоновки системы, примененные в К компьютере.

1. Введение

К компьютер^{note)} привлекает внимание благодаря своей высокой вычислительной производительности и эффективности работы, достигаемых его высокопроизводительными процессорами и архитектурой интерконнекта Tofu. Однако К компьютер, который использует более 80 000 процессоров, во многом имеет исключительную вычислительную производительность благодаря технологиям компоновки, которые обусловили его небольшой размер, высокую надежность и низкое энергопотребление.

Чтобы построить К компьютер как крупномасштабную систему мы использовали технологии компоновки системы, которые развивались Fujitsu к настоящему времени как основа для технологии тонкой настройки, охватывающей как отдельные элементарные технологии так и всю систему. Мы разработали технологию компоновки систем, которая позволяет производить самые быстрые компьютеры в мире.

Аналогично объединению в К компьютере, технологии компоновки системы в данной статье представлены элементами, относящимися к стойке, связи, охлаждению и питанию.

2. Стоечные технологии

Данный раздел описывает стоечные технологии, используемые в К компьютере.

2.1 Наполненная стойка

Как показано на Рисунке 1, [стойка] К компьютер[a] объединяет в общей сложности 102 процессора, расположенных на 24 системных платах (SB), каждая с четырьмя процессорами и шесть системных плат ввода/вывода (IOSB), каждая с одним процессором. Это более чем в три раза превышает число процессоров, установленных в существующих системах компании Fujitsu.

Такая высокая плотность монтажа была достигнута путем принятия гибридного охлаждения, которое использует жидкостное охлаждение для компонентов, выделяющих большое количество тепла, таких как ЦПУ и воздушное охлаждение для памяти. Эти методы охлаждения будут описаны позже в разделе технологий охлаждения.

^{note)} “К computer” - английское название, которое RIKEN использовал для суперкомпьютера в данном проекте начиная с июля 2010. “К” пришло из японского слова “Kei,” которое обозначает 10пета или 10 в 16й степени.

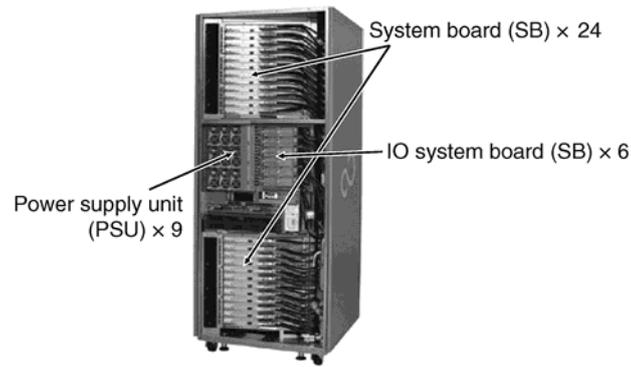


Рисунок 1
Компоновка стойки

Из-за высокой плотности монтажа вес стойки достигает 1000 кг даже без кабелей интерконнекта (кабели Tofu), которые объединяют стойки. Этот вес примерно в 1.5 раза тяжелее существующих систем. Чтобы справиться с этим, мы разработали новые ролики и новую базу стойки. В дополнение к проблеме с весом, тяжелые системные платы устанавливаются на высоких положениях и это создает нежелательный эффект повышения центра тяжести [стойки] К компьютера, что очень усложняет решение проблем с вибрацией. Мы в полной мере использовали моделирование для принятия соответствующих мер и реализовали систему, устойчивую к землетрясениям 100 Gal (на уровне 6, высшем в сейсмической шкале Японского Метеорологического Агентства [JMA], исчисляемой от 0 до 7), что является аналогичным с существующими системами.

2.2 Наклонный монтаж системных плат

Для охлаждения SBs мы применили жидкостное охлаждение для компонентов, выделяющих большое количество тепла, таких как процессоры и воздушное охлаждение для модулей DIMM и других компонентов. По этой причине мы должны были обеспечить место для воздухозаборников/вентиляционных отверстий и вентиляторов для доставки достаточного потока охлаждающего воздуха, а также мы должны были упаковать систему с коллекторами для жидкостного охлаждения. Мы разработали конструкцию, в которой SB монтируются наклонным образом, что помогает достичь высокой плотности монтажа (Рисунок 2).

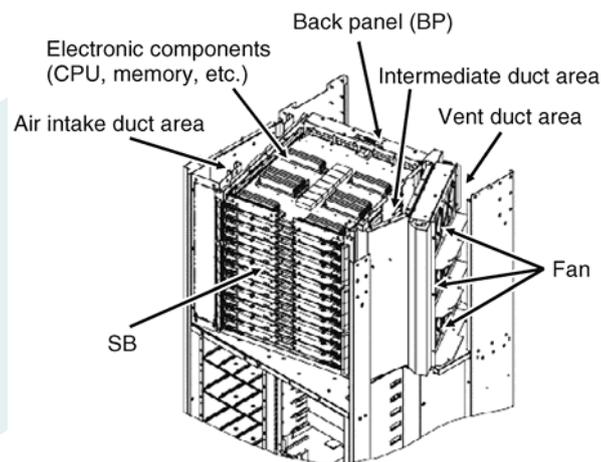


Рисунок 2
Наклонный монтаж SB

2.3 Объединение системы и укладка кабелей

Для минимизации длины соединений между CPU была принята конструкция, в которой кабели Tofu непосредственно реализованы в задних панелях (BPs) к которым крепятся SB. BP имеют конструкцию, в которой разъемы, используемые для соединения с SB на лицевой стороне и разъемы для кабелей на задней стороне являются общими с использованием сквозных отверстий (Рисунок 3), что минимизирует потери при передаче и отражение сигнала.

BP закрыта удерживающим кабели экраном который одновременно радиоэкранирует и позволяет блокировать кабель Tofu одним нажатием при установке, что обеспечивает конструкцию которая предотвращает от электростатических разрядов (ESD), а также надежно сохраняет кабели (Рисунок 4).

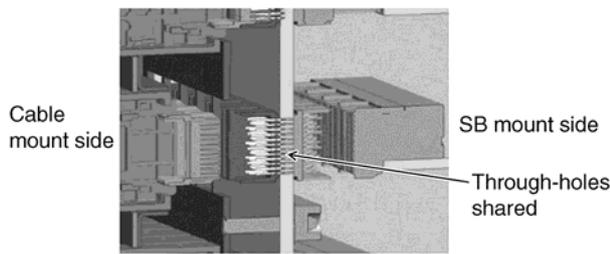


Рисунок 3
Подключение коннектора BP

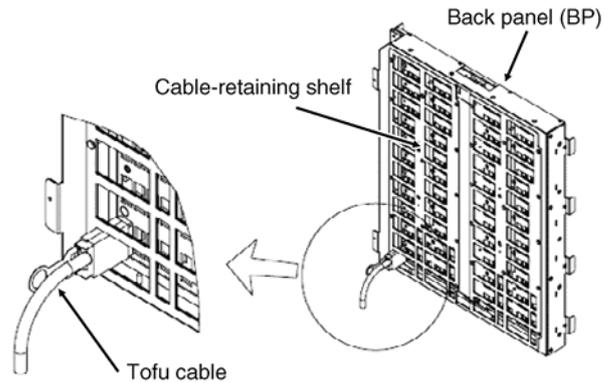


Рисунок 4
Подключение кабеля к BP

3. Технология соединения

3.1 Кабель Tofu

Кабель Tofu был заново разработан для обеспечения высокой скорости помимо надежности, так, что даже если он выступает предметом воздействия внешних сил, то в момент их наличия его характеристики не изменяются. Помимо этого для разработки тщательно учитывался при производстве контроль за согласованностью электрических характеристик вдоль направления длины.

Для оболочки провода был использован вспененный политетрафторэтилен (ePTFE), материал с низкой диэлектрической проницаемостью, а также была применена специальная защитная структура, которая выполняет электрическую симметрию при выравнивании потребления тока и близко расположенных линий дифференциальных сигналов. Это позволило нам увеличить плотность центральных жил и облегчило сборку высокоскоростного разъема на конце кабеля при сохранении высокоскоростных характеристик передачи. Кроме того, управление параметрами производства с заданными характеристиками производства кабеля и автоматизация части процессов при сборке позволило нам осуществлять массовое производство кабелей с минимизированными потерями при передаче и изменениями временных задержек. В частности, кабель Tofu имеет лучшую в мире производительность в качестве кабеля интерконнекта как кабеля, соединяющего стойки, с максимальной временной задержкой между положительной и отрицательной компонентами дифференциального сигнала в 30ps/10m и это способствует высокому качеству передачи в К компьютере.

Рисунок 5 показывает внешний вид кабеля Tofu в сборе.

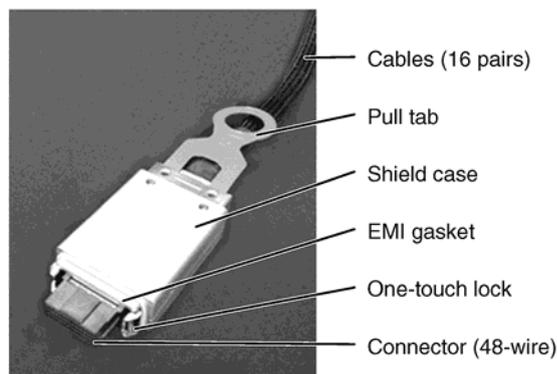


Рисунок 5
Сборка кабеля Tofu

Небольшой разъем с максимальной площадью 34mm x 14mm с запирающим при одном нажатии механизмом, который удерживает высокую плотность монтажа от влияния на эффективность работы. Шаг монтажа кабельной сборки составляет 14.45mm, что дает очень высокую плотность кабелей из стойки.

Обычно высокоскоростные кабели имеют много параметров проверки, и для сканирования всех сигналов требуется значительное время. Для К компьютера мы должны производить большое количество кабелей в короткий промежуток времени и одной из главных задач, с которой мы столкнулись, заключается в том, как сократить время тестирования. С этой целью мы сопоставили характеристики временной и частотной областей с высокой точностью в месте проведения моделирования частотной области характеристик для некоторых элементов, которые должны пройти проверку во временном диапазоне. Таким образом, мы сократили время тестирования приблизительно до одной десятой от обычно необходимого времени.

3.2 Управление кабелями Tofu

К компьютер должен иметь соединения с высокой плотностью с помощью многочисленных кабелей при возможно кратчайших расстояниях. Для этой цели мы предварительно внимательно изучили формообразующие технологии и методы прокладки кабелей, чтобы обеспечить оптимум прокладки кабелей и пространства с учетом эффективности объемного размещения и дополнительной длины.

Количество кабелей, подключенных к одной ВР, составляет около 200, а количество кабелей на стойке, состоящей из двух ВР, в верхней и нижней частях стойке составляет более 400. Кабели подключаются отдельно для верхней и нижней частей стойки, а пространство поверх стойки и под полом было эффективно использовано для достижения кратчайших длин кабелей между стойками. Количество кабелей, используемых для всей системы, составляет около 200 000, а общая длина проложенных кабелей достигает примерно 1000км.

Рисунок 6 показывает пример раскладки кабелей поверх стоек.

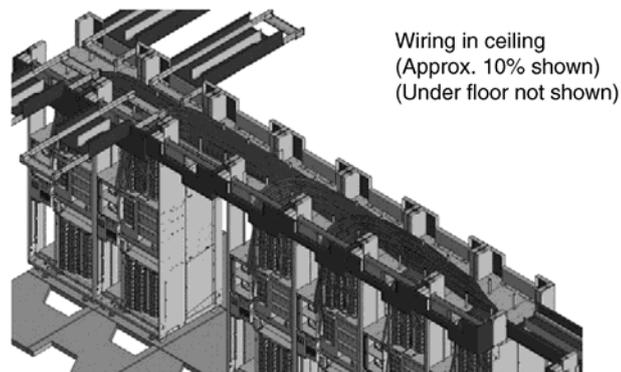


Рисунок 6
Укладка кабелей поверх стоек

4. Технология охлаждения

Для К компьютера мы применили гибридную структуру охлаждения, которая сочетает в себе систему жидкостного охлаждения, характеризующуюся высокой эффективностью и высокой надежностью, и систему охлаждения воздухом, которая обеспечивает высокую эффективность и низкую стоимость.¹⁾ Следующие разделы описывают эти охлаждающие системы.

4.1 Жидкостная охлаждающая система

В данном разделе объясняется структура системы жидкостного охлаждения для стоек и SB.

4.1.1 Система жидкостного охлаждения для стоек

Рисунок 7 показывает структуру системы жидкостного охлаждения для каждой стойки. Два коллектора (для подачи жидкости и ее сбора) установлены вертикально на правой передней стороне, от них ответвляются шланги подачи жидкости и ее возврата (24 шланга на каждую SB и шесть шлангов на каждую IOSB). Кроме того, эти коллекторы подачи и сбора жидкости оснащены датчиками (датчики температуры и давления жидкости, а также датчики росы) и управляющими устройствами (воздушный клапан, клапан двигателя, обратный клапан и фильтр). Все они объединены в две магистрали, которые в конечном итоге составляют собой систему жидкостного охлаждения из двух магистральных систем: подачи жидкости и ее возврата.

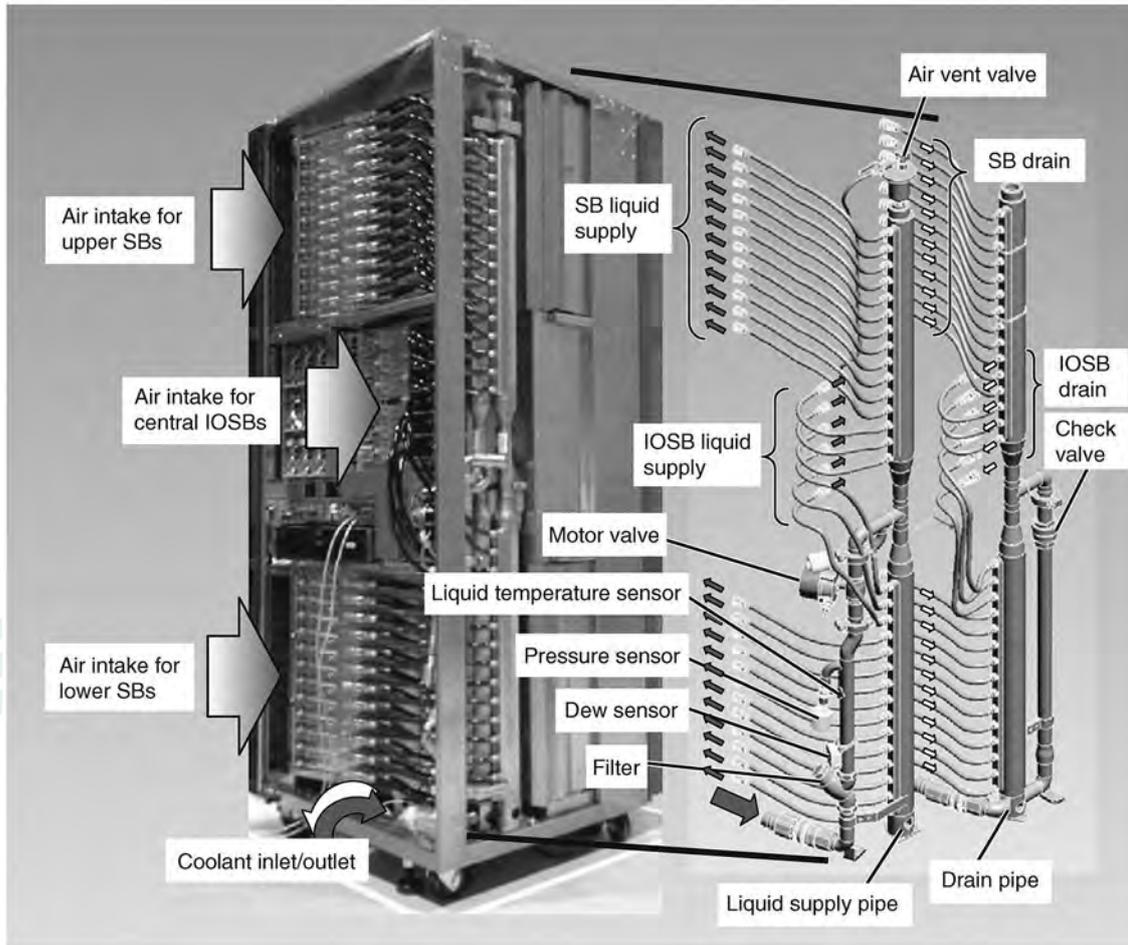


Рисунок 7
Компоновка жидкостной системы охлаждения для стойки

Для достижения баланса между подачей жидкости и ее возвратом, мы разработали систему таким образом, что жидкость подается и возвращается обратно через центр с управляемым эталоном на: длину коллектора, диаметр труб магистралей в пути потока системы, а также длинами шлангов подачи и возврата жидкости в пределах интервала, который уравнивает любые перепады давления. В результате падение давления между подачей и возвратом жидкости было успешно уменьшено до пределов $\pm 5\%$. В установленных коллекторах мы предприняли меры тщательного сохранения пространства, чтобы ограничить коллекторы подачи и возврата жидкости очень малым пространством в углу с правой стороны в передней части стойки.

Таким образом, мы разработали простую систему коллектора, способную контролировать расход без специальных устройств, таких как регулятор потока, тем самым обеспечивая максимальное качество коллекторов ими самими и реализуя высокоэффективную, высоконадежную систему.

4.1.2 Система жидкостного охлаждения SBs

Процессоры (CPU) и пакеты контроллера интерконнекта (ICC) установленные на SB, а также некоторые устройства преобразования электропитания (преобразователи DC-DC) охлаждаются устройствами жидкостного охлаждения (LCUs). LCU состоит из двух или более пластин охлаждения (CP) для охлаждения CPU и ICC, которые соединены между собой через линии трубопроводов в двух параллельных ветвях (Рисунок 8). С целью обеспечения ремонтпригодности охлаждающая жидкость подается и возвращается из LCU через жидкостные соединители, предусмотренные на передней стороне SB.

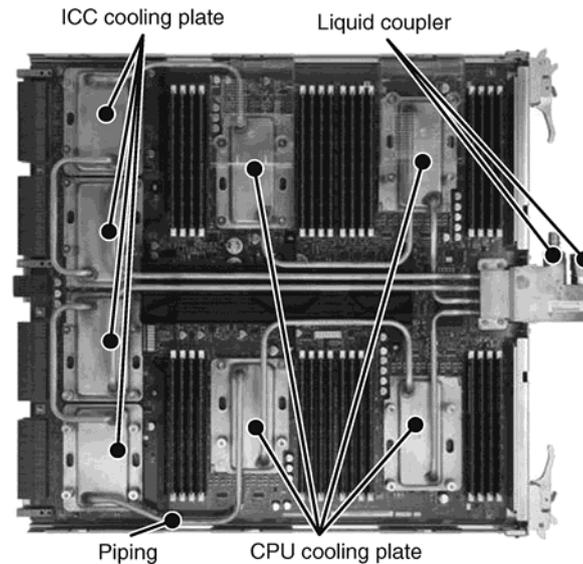


Рисунок 8
Устройство жидкостного охлаждения (LCU, Liquid cooling unit) на SB

К компьютер является системой, состоящей из более чем 20 000 SB и эффективное охлаждение с небольшим количеством циркулирующего хладагента было вызовом. С этой целью LCU была разработана так, чтобы иметь восемь CP, соединенных трубопроводными линиями, параллельными ветвям четырех CP, чтобы снабдить равным расходом теплоносителя каждую CP. Кроме того была принята конструкция микро-канального потока внутри CP, чтобы минимизировать поток циркулирующей охлаждающей жидкости (Рисунок 9). Путем принятия этих мер была снижена потеря расхода теплоносителя и различия потоков теплоносителя во всех CP в стойке успешно контролируется в пределах $\pm 3\%$ выравнивание конструкций внутренних потоков CP. Для монтажа на SB интегрированных LCU, включающих соединения трубопроводами, требуются конструкции, способные нивелировать изменения размеров. Мы разработали размер и форму медных трубочек, используемых для соединения CP в единый механизм, конструкция, которая минимизирует влияние ошибок размеров и напряжений при монтаже. Эта интегрированная структура существенно помогла улучшить прочность и надежность LCU в дополнение к снижению стоимости. Кроме того, уменьшение в размере и весе позволило нам получить систему с высокой плотностью монтажа.

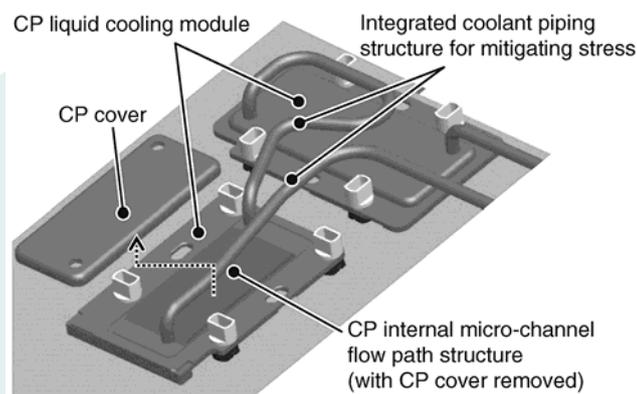


Рисунок 9
Конструктив охлаждающих пластин (CP, Cooling plate)

4.2 Система воздушного охлаждения

Система воздушного охлаждения К компьютера имеет характерную стоечную конструкцию. Как уже упоминалось в разделе "Стойчные технологии", полка устанавливается в стойку под углом и двенадцать SB установлены на полке с воздушным охлаждением при помощи шести (резервирование 5 + 1) вентиляторов. Этот метод установки был принят для обеспечения области для трубопроводов жидкости и кабелей Tofu и минимизации потерь охлаждающего SB воздуха при обеспечении возможности доступа к ВР и вентиляторам и отключения жидкостных соединителей для подачи жидкости и снятия с SB. Такой наклонный монтаж сократил

потери охлаждающего воздуха наполовину по сравнению с вертикальным монтажом на полке при обеспечении ремонтпригодности. Число вентиляторов также было уменьшено наполовину путем соединения с механизмом предотвращения прототока охлаждающего воздуха, установленного рядом с вентиляторами.

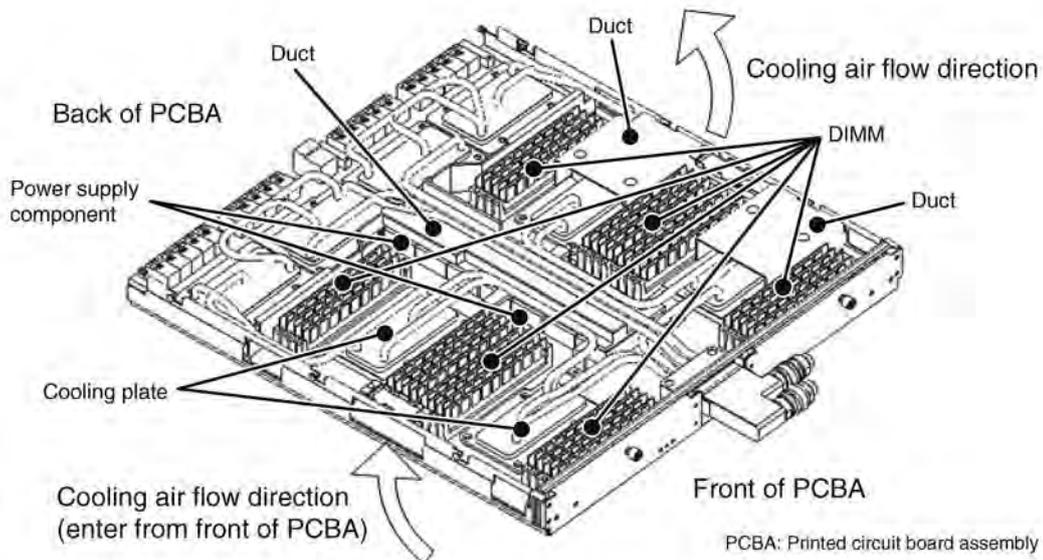


Рисунок 10
Конструктив воздушного охлаждения SB

Рисунок 10 показывает конструкцию воздушного охлаждения SB, а рисунок 11 поток охлаждающего воздуха. SB имеет модули DIMM и некоторые компоненты электропитания, требующие воздушного охлаждения. Охлаждающий воздух для этих компонентов входит по диагонали от передней части SB, охлаждает компоненты и выходит по диагонали в задней части. Нам необходимо было использовать этот искривленный поток охлаждающего воздуха и рассмотреть вопрос о LCU, которые блокируют поток охлаждающего воздуха для поддержания оптимального расхода воздуха. Мы выполнили полный жидкостно-температурный анализ для определения конструктивных частей SB и форм воздуховодов. Для обеспечения и управления потока воздуха требуемого для охлаждения мы оптимизировали форму и расположение отверстий в металлических пластинах со стороны подачи охлаждающего воздуха и со стороны выхлопа, а также форму и расположение каналов в SB для распределения охлаждающего воздуха между компонентами, подлежащими охлаждению на хорошо сбалансированной основе. Путем принятия этих мер изменение температур модулей DIMM и других компонентов было сведено к минимуму для реализации эффективной системы охлаждения.

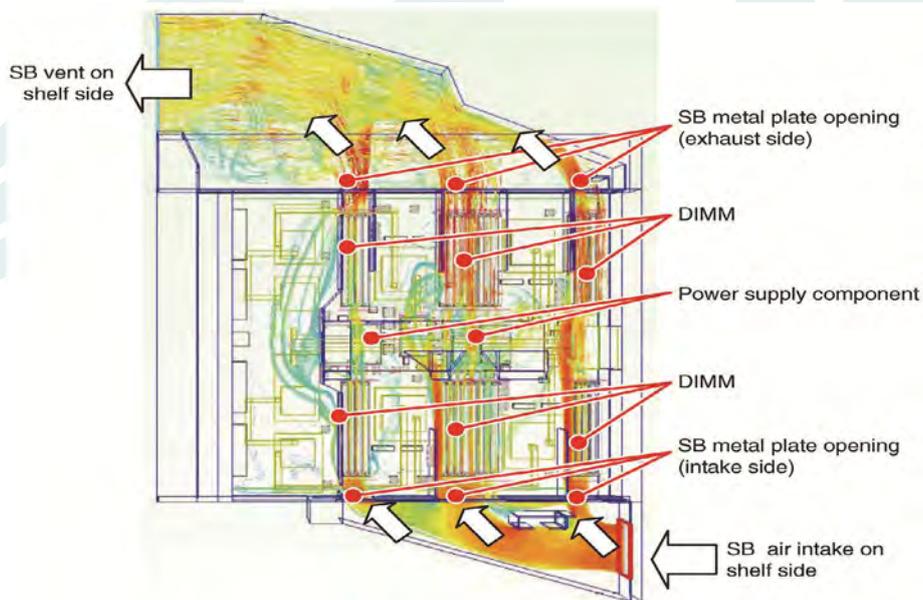


Рисунок 11
Поток охлаждающего SB воздуха (вид сверху)

5. Технология энергоснабжения

В этом разделе описывается структура, эффективность преобразования мощности и надежность системы питания, используемой для К компьютера.

5.1 Конфигурация системы питания

Рисунок 12 показывает схематическую диаграмму конфигурации системы питания.

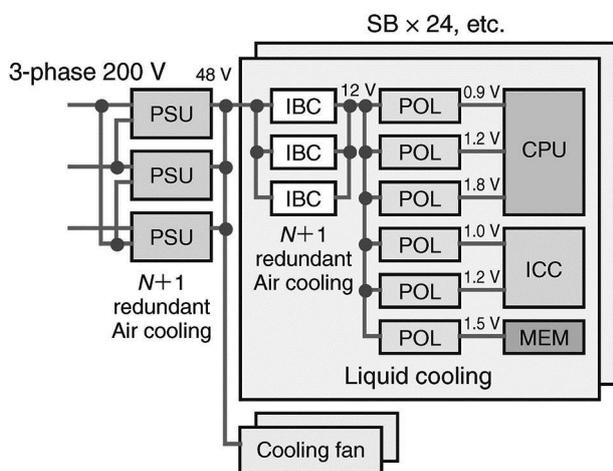


Рисунок 12
Конфигурация системы питания

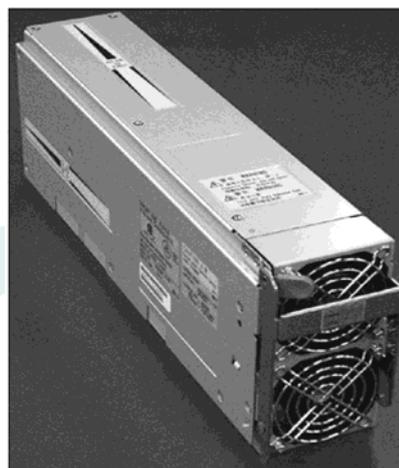


Рисунок 13
Внешний вид PSU

Система получает трехфазное питание 200В и распределяет его по однофазным 200В входным блокам питания (PSUs) (Рисунок 13). Каждый PSU подает питание 48В на SB и вентилятору, управляя при этом выходными токами сохраняя их одинаковыми. На SB каждый конвертер промежуточной шины (IBC, intermediate bus converter) преобразует 48В в 12В при выполнении балансировки токов и подает питание на каждую дискретную неизолированный DC-DC преобразователь точки нагрузки (POL, point of loads). Отдельные POL преобразовывают 12В питание в снабжение 0.9В и другими низковольтными токами большого значения для CPU и других нагрузок конечного этапа. Для охлаждения системы питания используются воздушное охлаждение для PSUs и IBC, в то время как для POL используется жидкостное охлаждение.

5.2 Эффективность преобразования мощностей системы питания

Описание каждого типа энергопитания приводится в Таблице 1.

Таблица 1
Описание системы энергопитания

	PSU	IBC	POL
Входное напряжение (V)	200–240	48	12
Напряжение на выходе (V)	48	12	0.9–3.3
Сила тока на выходе (A)	60	20	4–120
Размеры (mm)	70 × 123 × 325	35 × 58	35 × 50

Эффективность преобразования мощности блоками питания была улучшена до 91% с обычных 88%.

Мы добились этого путем применения новых устройств питания, что делает схемы преобразования питания более передовыми и совершенствуя технологии компоновки. В частности, существенное влияние оказало уменьшение основных коммутационных потерь, таких как токи восстановления, путем использования карбидно-кремниевых диодов (SiC) и полевых транзисторов на основе супер перехода (SJ FET), которые являются новыми устройствами питания, характеризующиеся высокой скоростью переключения и низким сопротивлением во включенном состоянии. В результате в стойке высокой плотности того же размера была реализована выходная мощность 3000Вт, что в 1.5 раза выше существовавших на этот момент стоечных решений Fujitsu с 2000Вт.

Одной из характерных черт системы питания на каждой SB явилось то, что изолированные DC-DC преобразователи были разделены на IBC и POL, что позволило размещать POL-ы в непосредственной близости от основных нагрузок, таких как CPU. Это привело к значительному сокращению потерь мощности на разводке печатных плат (PCB) в SB. Эффективность преобразования мощности в результате комбинирования IBC и POL (0.9В) была улучшена до 84% с 80% для обычных изолированных DC-DC преобразователей. Еще одной важной особенностью является то, что улучшение характеристик POL и их расположение в непосредственной близости от нагрузок снизило емкость конденсаторной развязки нагрузки до одной десятой от значения для существующих систем.

Описанное выше технологическое усовершенствование повысило эффективность преобразования энергии всей системы питания с обычных 70% до 76%.

5.3 Надежность системы питания

Для крупномасштабных систем типа К компьютера надежность системы питания является важным фактором, поскольку огромное количество численных расчетов производится в течение длительного периода времени. В данной системе питания использовалось N+ 1 резервирование для PSU и IBC, что значительно улучшило надежность системы. Хотя POL не используют избыточного резервирования, для обеспечения достаточной надежности была принята совмещенная с CPU жидкостная система охлаждения CP для поддержания низкой температуры перехода в транзисторах, составляющая 40°C.

6. Заключение

Данная статья продемонстрировала конкретные примеры для представления использованных технологий конфигурирования, включая стойку, подключения, охлаждения и энергопитания, которые позволили нам создать К компьютер с высокой плотностью и высокой эффективностью. Спрос на системы с высокой плотностью и высокой эффективностью растет для серверных устройств в целом, наряду с суперкомпьютерами.

Мы намерены продолжать работу по технологическому совершенствованию, отвечающему данному требованию.

Ссылки

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

