Федеральное государственное бюджетное образовательное учреждение высшего образования

«Сибирский государственный университет телекоммуникаций и информатики»

(СибГУТИ)

Кафедра вычислительных систем

Расчетно-графическое задание по дисциплине

«Архитектура вычислительных систем»

Вариант 4

Выполнил:

Студент гр. ИП-813

Гаврик В. А.

Проверил:

к.т.н доцент кафедры ВС:

Ефимов А.В.

Новосибирск 2020 г.

СОДЕРЖАНИЕ

[**1 ЗАДАНИЕ 3**](#_Toc59428183)

[**1.1 Задание 3**](#_Toc59428184)

[**1.2 Общие сведения 3**](#_Toc59428185)

[**1.3 Характеристики суперкомпьютера Sunway TaihuLight 5**](#_Toc59428186)

[**1.4 Процессор 6**](#_Toc59428187)

[**1.5 Вычислительная система 10**](#_Toc59428188)

[**1.6 Сетевая система 13**](#_Toc59428189)

[**1.7 Периферийные системы 13**](#_Toc59428190)

[**1.8 Система технического обслуживания и диагностики 14**](#_Toc59428191)

[**1.9 Система электроснабжения и система охлаждения 14**](#_Toc59428192)

[**1.10 Программный стек 15**](#_Toc59428193)

[**1.11 Область применения Sunway TaihuLight 16**](#_Toc59428194)

[**2 ЗАДАНИЕ 18**](#_Toc59428195)

[**2.1 Расчет диаметра структуры 18**](#_Toc59428196)

[**2.2 Расчет бисекционной пропускной способности 19**](#_Toc59428197)

[**2.3 Вычисление среднего диаметра 21**](#_Toc59428198)

[**3 СПИСОК ЛИТЕРАТУРЫ 22**](#_Toc59428199)

# **ЗАДАНИЕ**

## **Задание**

Выполнить анализ архитектуры супер ВС из списка Top 500. В соответствии с моделью коллектива вычислителей выделить и описать уровни мультиархитектуры супер ВС. В том числе для каждого уровня показать функциональную структуру, сущность вычислителя, топологию сети связей, доступные технологии программирования и область эффективного применения, а также структурные характеристики.

Расчет структурных характеристик (диаметр, средний диаметр, бисекционная пропускная способность) выполнить для одного из уровней мультиархитектуры.

## **Общие сведения**

Китайская Суперэвм (Sunway TaihuLight) удерживала лидирующую позицию в рейтинге TOP500 с 2016 до 2018 года. В соответствии с тестами LINPACK ее считали самым производительным суперкомпьютером, минимум в полтора раза превосходящим ближайшего конкурента и втрое опережающим самую производительную американскую модель Titan. Разработка и строительство вычислительной системы обошлось в 1,8 млрд. юаней или 270 млн долларов.

Является разработкой специалистов Национального исследовательского центра параллельных компьютерных вычислений и технологий (National Research Center of Parallel Computer Engineering & Technology, NRCPC), а установлен он в помещении Национального суперкомпьютерного центра (National Supercomputing Center) в Уси.

Инвесторами проекта были правительство Китая, администрация китайской провинции Цзянсу и города Уси.

Суперкомпьютер занимает площадь 605 кв. м. Название модели дали в честь расположенного рядом озера Тайху, третьего по величине пресноводного водоема Китая.

Пиковый расход энергии под максимальной нагрузкой у суперкомпьютера Sunway TaihuLight составляет 15.37 МВт, что в пересчете на производительность равно 6 гигафлопс на ватт. Такой показатель позволяет системе Sunway TaihuLight занять одну из первых строчек рейтинга Green500, в котором суперкомпьютеры ранжируются по показателю эффективности.

В состав системы Sunway TaihuLight входит 10 649 600 вычислительных ядер, сгруппированные в 40 960 узлов. Этот суперкомпьютер более чем в два раза быстр, и в три раза более эффективен, нежели суперкомпьютер Tianhe-2, который на тесте LINPACK показывал производительность в 33.86 петафлопс.



**Рисунок 1 - Sunway TaihuLight supercomputer**

Модель Sunway TaihuLight применяется для выполнения сложных вычислений в области медицины, горнодобывающей промышленности и производстве. С помощью вычислительной машины прогнозируют погоду, исследуют новые лекарства и анализируют «большие данные» – массивы информации, обработать которые не получится даже у самого мощного серийного компьютера.

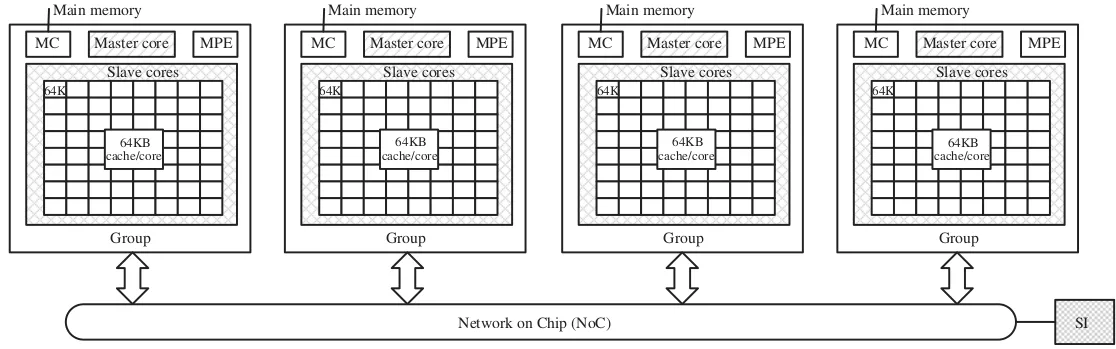
## **Характеристики суперкомпьютера Sunway TaihuLight**

|  |  |
| --- | --- |
| Процессор (CPU) | SW26010 (ShenWei) |
| Разработчик (Developer) | Национальный исследовательский центр параллельной вычислительной техники и технологий (ориг. National Research Center of Parallel Computer Engineering & Technology) |
| Изготовитель чипов (Chip Fab) | Поставщик процессоров-Шанхайский Центр проектирования высокопроизводительных микросхем |
| Набор инструкций (Instruction set) | Набор инструкций Shenwei-64 |
| Процессорные ядра узлов (Node processor cores) | 256 вычислительных элементов (Computing Processing Elements, CPE),  4 элемента управления (Management Processing Element, MPE) |
| Частота работы процессора (Clock frequency CPU) | 1.45 GHz |
| Технологический процесс (Process technology) | N/A (предположительно 28 нм) |
| Потребляемая мощность при выполнении теста Linpack (Power consumption when running the Linpack test) | 15.371 МВатт |
| Пиковая производительность узла (Node Peak Performance) | 3.06 терафлопс |
| Производительность Linpack (Linpack performance) | 93 Пфлопс |
| Пиковая производительность системы (Peak Performance of system) | 125.4 Пфлопс |
| Целевое применение (Target application) | высокопроизводительные вычисления (HPC) |
| Объем ОЗУ (Total RAM) | 1.31 петабайта |
| Общая пропускная способность памяти (Total memory bandwidth) | 5591.5 терабайт/с |
| Общий объем памяти (Total storage) | 20 петабайт |
| Пропускная способность сетевого канала (Network link bandwidth) | 16 Гб/с |
| Общая пропускная способность ввода-вывода (Total I/O bandwidth) | 288 Гб/с |
| Шкафы (Cabinets) | 40 |
| Ядер на узел (Cores per node) | 260 cores |
| Узлов на шкаф (Nodes per cabinet) | 1024 |
| Узлы (Nodes) | 40,960 |
| Общее количество ядер системы (Total system core count) | 10,649,600 |
| Система охлаждения (Cooling system) | Представляет собой замкнутое водяное охлаждение с индивидуальным блоком жидкостного водяного охлаждения |

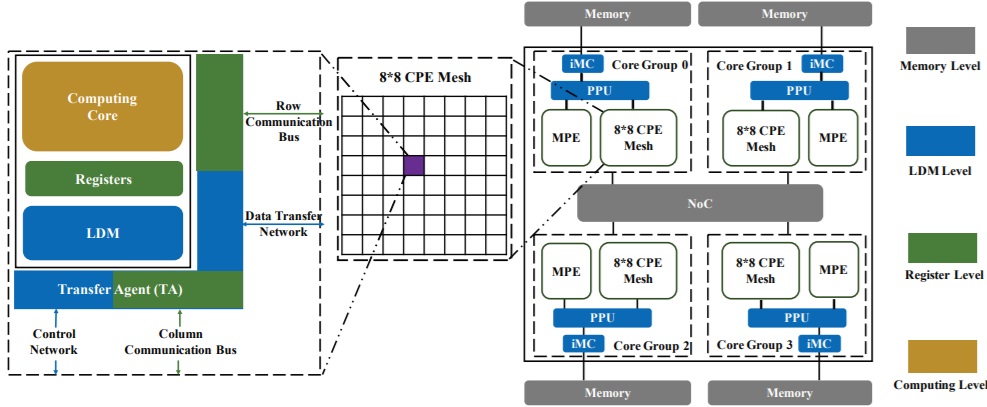
## **Процессор**

Одним из основных технологических инноваций в TaihuLight суперкомпьютер Sunway-это отечественный SW26010 многоядерный процессор. Общая архитектура процессора SW26010. Процессор включает в себя четыре группы ядер (CGs). Каждый CG включает в себя один элемент обработки управления(MPE), один кластер вычислительных элементов обработки (CPE) восемь на восемь CPE и один memory controller (MC). Эти четыре группы ядер связаны по сети (кольцу) на кристалле (NOC). Каждый CG имеет свое собственное пространство памяти, которое соединено с MPS и кластером CPS через MC. Процессор подключается к другим внешним устройствам через системный интерфейс (SI).

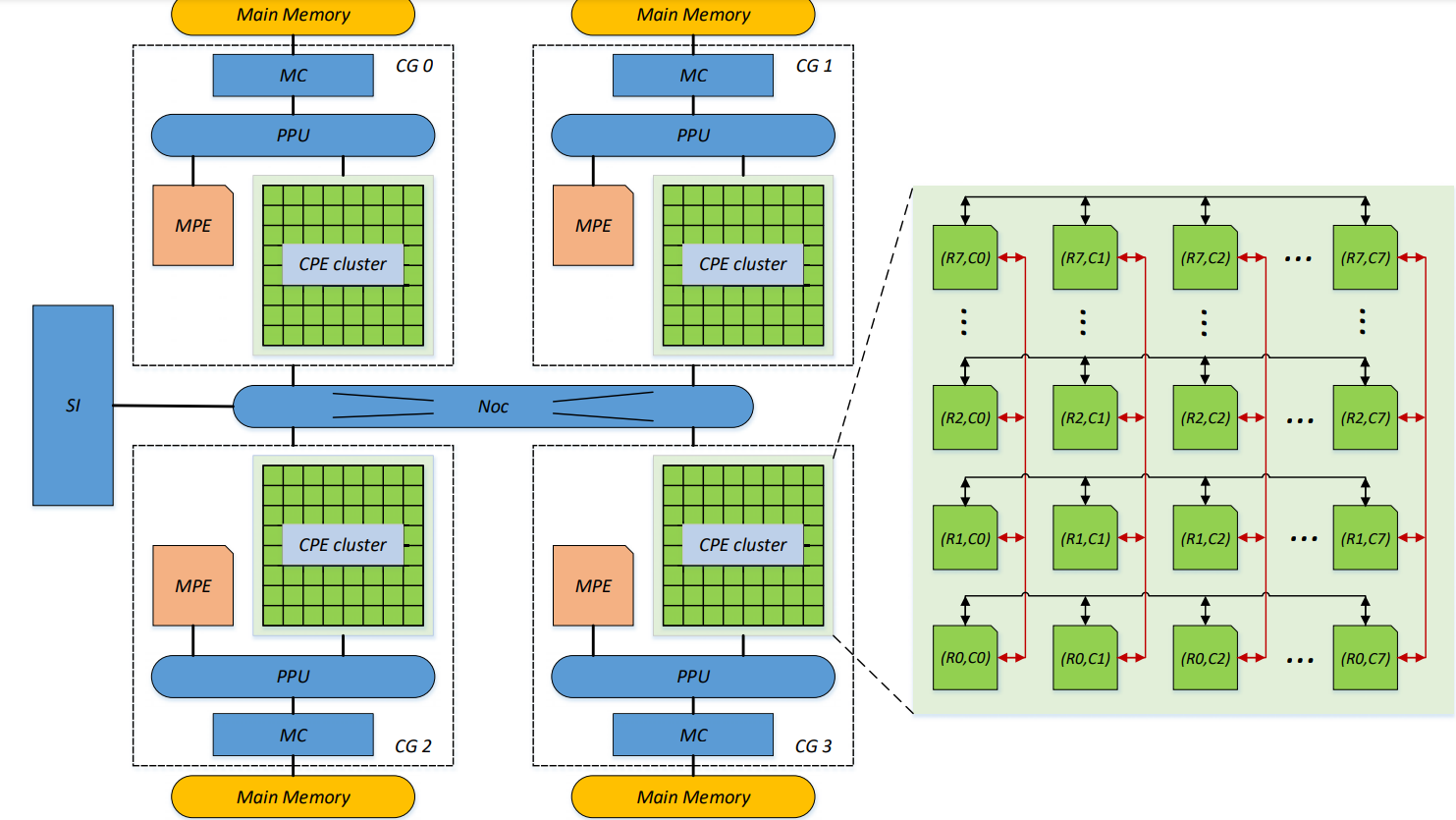
MPE и CPE основаны на архитектуре RISC, 64-битной, SIMD, неупорядоченной микроструктуре. И MPE, и CPE участвуют в приложении пользователя. Управление производительностью MPE, связь и вычисления в то время как CPE в основном выполняют MPE-это полное 64-битное RISC-ядро, которое может работать как в пользовательском, так и в системном режимах. MPE полностью поддерживает функции прерывания, управления памятью, суперскалярной обработки и выполнения вне порядка. Таким образом, MPE является идеальным ядром для обработки функций управления и связи. В отличие от этого, CPE также является 64-битным RISC-ядром, но с ограниченными функциями. CPE может только работать в пользовательском режиме и не поддерживает функции прерывания. Целью проектирования этого элемента является достижение максимальной агрегированной вычислительной мощности при минимизации сложности микроархитектуры. Кластер CPE организован как сетка восемь на восемь, с сетчатой сетью для достижения передачи данных с низкой задержкой регистрации между восемь на восемь CPE. Сетка также включает в себя контроллер сетки, который обрабатывает элементы управления прерываниями и синхронизацией. И MPE, и CPE поддерживают 256-битные векторные структуры. С точки зрения иерархии памяти, каждый MPE имеет 32 КБ кэша команд L1 и 32 КБ кэша данных L1, а также 256 КБ кэша L2 как для команд, так и для данных. Каждый CPE имеет свой собственный 16-килобайтный кэш L1 для инструкций и управляемую пользователем память scratch pad (SPM). SPM может быть сконфигурирован либо как быстрый буфер, поддерживающий точное управление пользователем, либо как программно эмулируемый кэш, обеспечивающий автоматическое кэширование данных. Однако, поскольку производительность программно-эмулируемого кэша невелика, в большинстве случаев для достижения хорошей производительности нам нужена схема буферизации, контролируемая пользователем. Объединяя четыре CGs кластеров MPE и CPE, каждый процессор Sunway обеспечивает пиковую производительность 3,06 Тфлопс при соотношении производительности к мощности более 10 Гфлопс/Ватт. В то время как производительность вычислений и энергоэффективность являются одними из лучших по сравнению с существующими графическими процессорами и микросхемами, размер встроенного буфера и пропускная способность памяти относительно ограничены. Процессор SW26010 использует технологию SoC и интегрирует четыре контроллера памяти DDR3, интерфейсы pcie3.0, Gigabit Ethernet и JTAG. Его пропускная способность памяти составляет 136,51 Гбит/с, а пропускная способность двунаправленного сетевого интерфейса-16 Гбит / с. Частота процессора составляет 1,45 ГГц, и он использует пакет fcbga3832.



**Рисунок 2 - общая архитектура нового процессора Sunway (SW26010)**



**Рисунок 3 - процессор SW26010**



**Рисунок 4 процессор SW26010 (Подробнее)**



**Рисунок 5 - Система Sunway TaihuLight**

## **Вычислительная система**

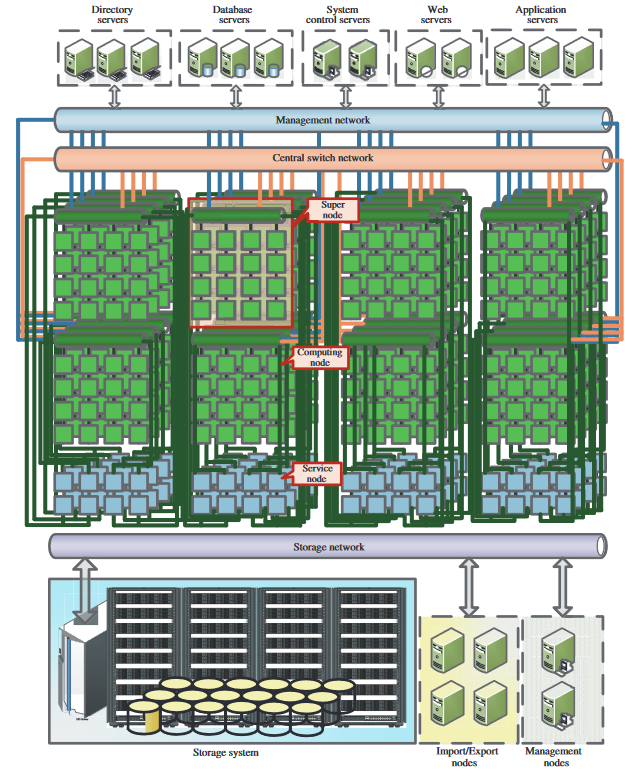
Стремясь к пиковой производительности 125 Пфлопс, вычислительная система Sunway TaihuLight построена с использованием полностью настраиваемого интеграционного подхода с несколькими различными уровнями:

* Вычислительный узел (один процессор на вычислительный узел);
* Суперузел (256 вычислительных узлов на суперузел);
* Шкаф (4 суперузла на шкаф);
* Вся вычислительная система (40 шкафов).

Вычислительные узлы являются базовыми узлами вычислительной системы и включают в себя:

* Один процессор SW26010;
* 32 ГБ памяти;
* Контроллер управления узлом;
* Блок питания;
* Интерфейсные схемы.

Группы из 256 вычислительных узлов интегрируются в тесно связанный суперузел с использованием полностью подключенного перекрестного коммутатора, чтобы поддерживать вычислительно-интенсивные, коммуникационно-интенсивные и интенсивные вычислительные задания ввода-вывода.



**Рисунок 6 - общая архитектура Sunway TaihuLight системы**

## **Сетевая система**

Сетевая система состоит из трех различных уровней, с центральной коммутационной сетью наверху, суперузловой сетью посередине и сетью совместного использования ресурсов внизу. Пропускная способность сети биссекции составляет 70 ТБ/ с, а диаметр сети-7.

Каждый суперузел включает в себя 256 процессоров Sunway, которые полностью соединены сетью суперузлов, что обеспечивает как высокую пропускную способность, так и низкую задержку для связи "все-ко-всем" между всеми 65536 элементами обработки.

Центральная коммутационная сеть отвечает за построение соединений и обеспечение обмена данными между различными суперузлами.

Сеть совместного использования ресурсов соединяет общие ресурсы с суперузлами и предоставляет услуги для связи ввода-вывода и отказоустойчивости вычислительных узлов.

## **Периферийные системы**

Периферийная система состоит из сетевой системы хранения данных и периферийной системы управления. Сетевая система хранения данных включает в себя как сеть хранения данных, так и дисковый массив хранения данных, обеспечивающий общее хранение 20 ПБ и высокоскоростную и надежную службу хранения данных для вычислительных узлов. Периферийная система управления включает в себя системную консоль, сервер управления и сеть управления, которые обеспечивают управление системой и обслуживание

## **Система технического обслуживания и диагностики**

Система технического обслуживания и диагностики обеспечивает комплексное оперативное управление техническим обслуживанием, мониторинг состояния и окружающей среды, обнаружение и регистрацию неисправностей, а также услуги безопасности для всей системы.

Приняв многоуровневую архитектуру управления, система технического обслуживания и диагностики интегрирует встроенные модули для тестирования и обслуживания узлов, суперузлов и всей вычислительной системы. Используя распределенный динамический подход к сбору данных, система обеспечивает мониторинг времени выполнения и визуализацию состояния вычислительной системы.

## **Система электроснабжения и система охлаждения**

Суперкомпьютер TaihuLight использует взаимно резервную входную мощность 2=35 кВ. В шкафах системы используется трехуровневый (300 в-12 в-0,9 в) режим питания постоянным током. Выходы внешнего источника питания 300 в, который напрямую подключен к шкафу. Основной источник питания шкафа преобразует 300В постоянного тока в 12В постоянного тока, а источник питания процессора преобразует 12В в напряжение, необходимое процессору.

Шкафы вычислительных и сетевых систем используют косвенное водяное охлаждение, в то время как периферийные устройства используют воздушное и водяное охлаждение, а энергетическая система использует принудительное воздушное охлаждение. Шкафы используют замкнутый контур, статическое гидравлическое давление для cavum, косвенную технологию водяного охлаждения параллельного потока, которая обеспечивает эффективное охлаждение для полномасштабного запуска Linpack

## **Программный стек**

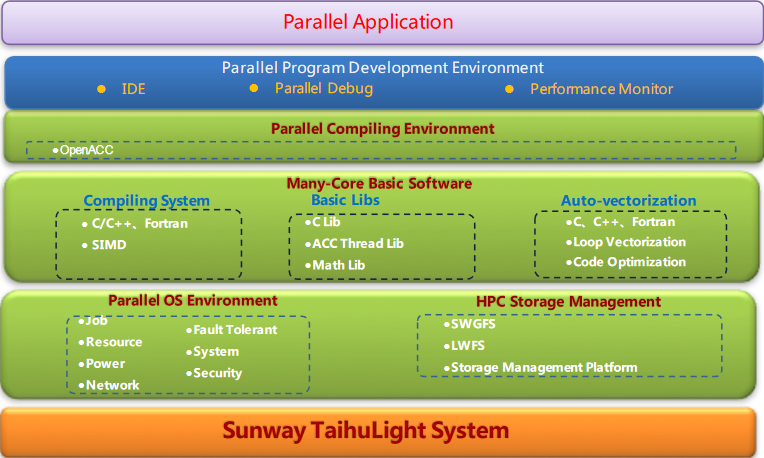
Программная система Sunway TaihuLight обеспечивает поддержку приложений в различных научных областях и отраслях промышленности. Основные компоненты включают базовое программное обеспечение для отечественного многоядерного процессора, среду параллельной операционной системы, высокопроизводительную систему управления хранилищем, язык параллельного программирования и среду компиляции, а также среду параллельной разработки.

Базовое программное обеспечение для многоядерного процессора включает базовые компоненты компилятора, такие как компиляторы C/C++ и Fortran, инструмент автоматической векторизации и базовые математические библиотеки. Среда параллельной операционной системы, которая включает в себя параллельную операционную систему, систему сетевого управления и системы управления доступностью и питанием, предоставляет пользователям вычислительные, прикладные и управленческие услуги в дополнение к другим системным услугам.

Высокопроизводительная система управления хранилищем включает в себя параллельную файловую систему, облегченную файловую систему и платформу управления хранилищем, которые обеспечивают поддержку хранения для запуска системного программного обеспечения и крупномасштабных параллельных приложений ввода-вывода. В дополнение к базовому программному обеспечению система TaihuLight также имеет язык параллельного программирования и среду компиляции для поддержки распараллеливания на различных уровнях.

Для распараллеливания на узловом уровне обычно применяется MPI. Для четырех единиц в пределах одного процессора, мы можем использовать либо через MPI или OpenMP. Для распараллеливания внутри CG используется Sunway OpenACC, настроенный инструмент параллельной компиляции, который поддерживает синтаксис OpenACC 2.0 и нацелен на кластеры CPE. Настроенный Sunway OpenACCtool поддерживает параллельное управление задачами, неоднородное извлечение кода и описания передачи данных.

Кроме того, основываясь на специфических особенностях архитектуры процессора Sunway, Sunway OpenACCtool также сделал ряд синтаксических расширений из оригинального стандарта OpenACC 2.0, таких как управление над многомерной буферизацией массива и упаковка распределенных переменных для передачи данных.



**Рисунок 7 - программный стек Sunway TaihuLight**

## **Область применения Sunway TaihuLight**

В настоящее время существует четыре ключевых области применения системы TaihuLight Sunway:

* Передовое производство: CAD, CAE приложений.
* Моделирование земной системы и прогнозирование погоды.
* Естественные науки.
* Аналитика больших данных.

Есть три заявки, которые являются финалистами премии Гордона Белла в SC16, которые основаны на новой системе Sunway TaihuLight. Эти три приложения:

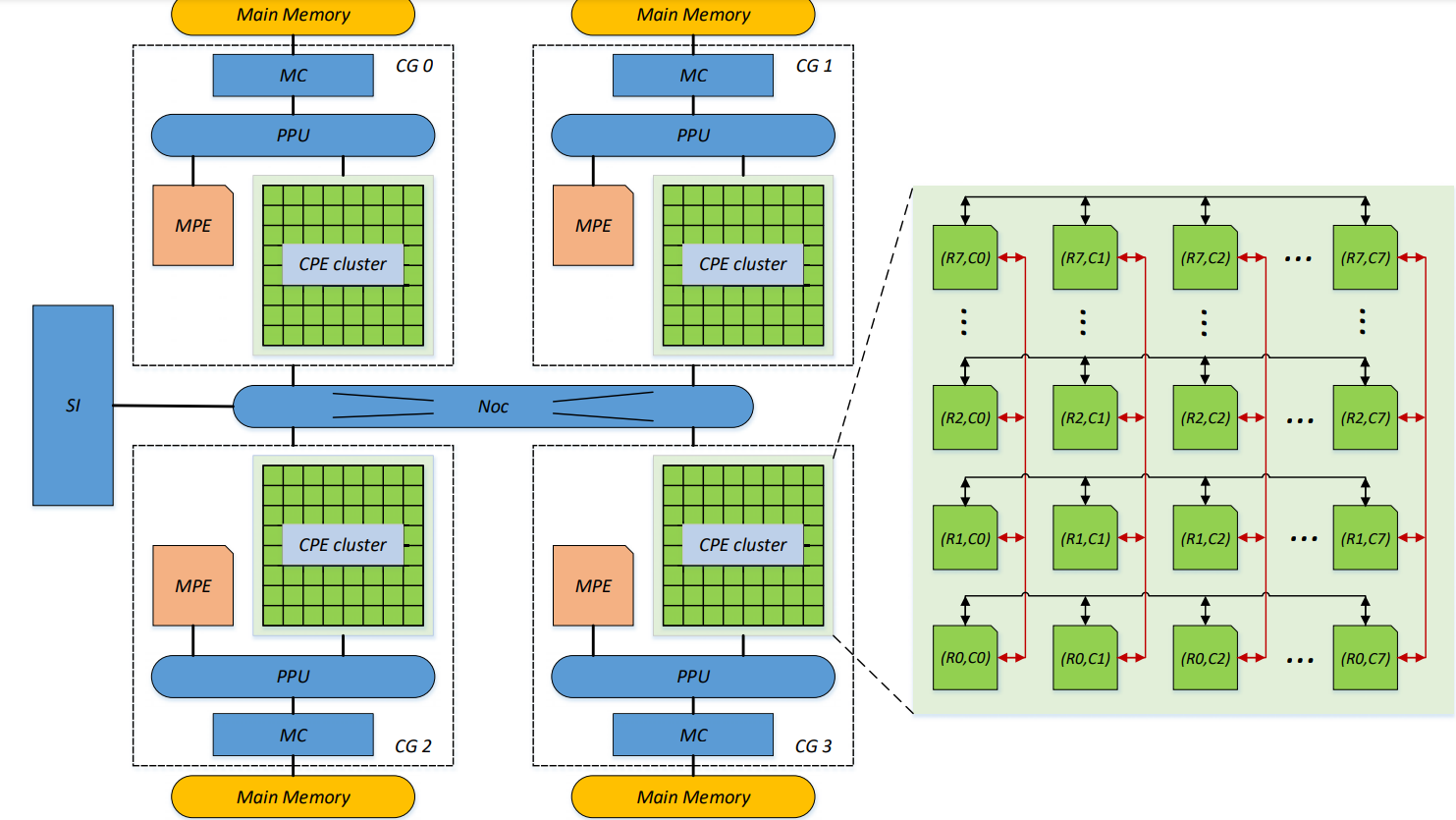
* полностью неявный негидростатический динамический решатель для моделирования атмосферы с разрешением облаков;
* высокоэффективное численное моделирование глобальных поверхностных волн со сверхвысоким разрешением;
* крупномасштабное моделирование фазового поля для динамики укрупнения на основе уравнения Кана-Хиллиарда с вырожденной подвижностью.

Все эти три приложения масштабируются примерно до 8 миллионов ядер (близко к полному системному масштабу). Приложения, которые поставляются с явным методом (например, моделирование волн и моделирование фазового поля), достигли устойчивой производительности от 30 до 40 петафлопс. В отличие от этого, неявный решатель достигает устойчивой производительности около 1,5 петафлопс с хорошей скоростью сходимости для крупномасштабных задач.

# **ЗАДАНИЕ**

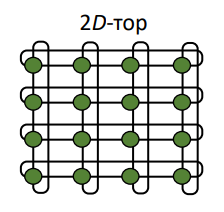
## **Расчет диаметра структуры**

Расчеты будем производить для уровня вычислительного узла (SW26010). Диаметр графа - длина максимального из кратчайших путей в графе (характеристика числа транзитных передач между эл. машиной)

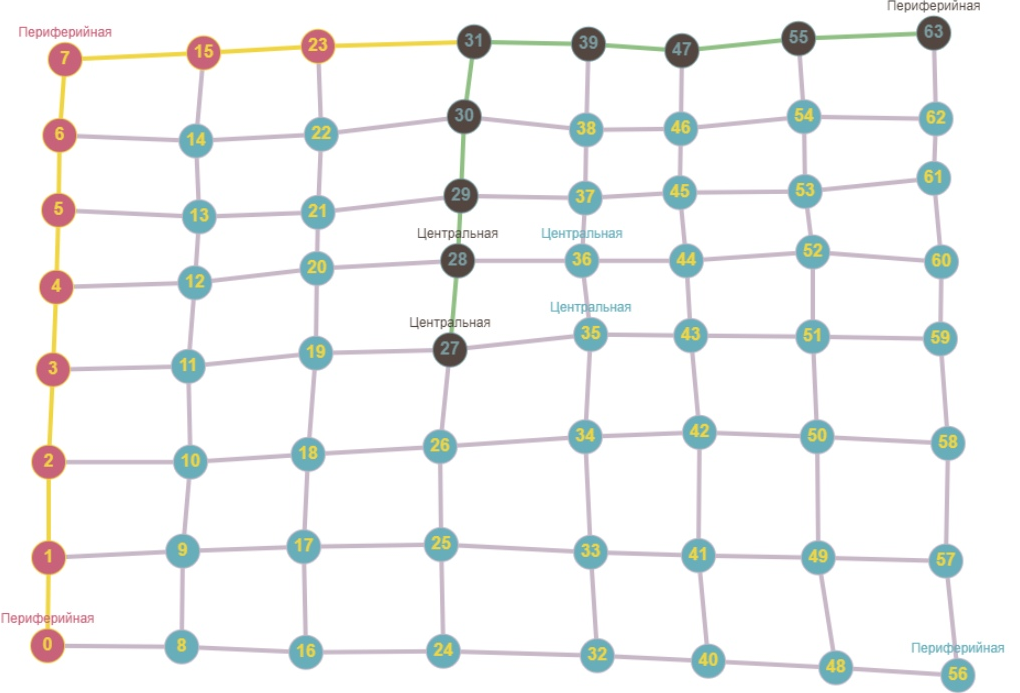


**Рисунок 8 - архитектура SW 26010**

Core group процессора SW 26010 соответствуют тороидальной решетке



**Рисунок 9 - тороидальная решетка**

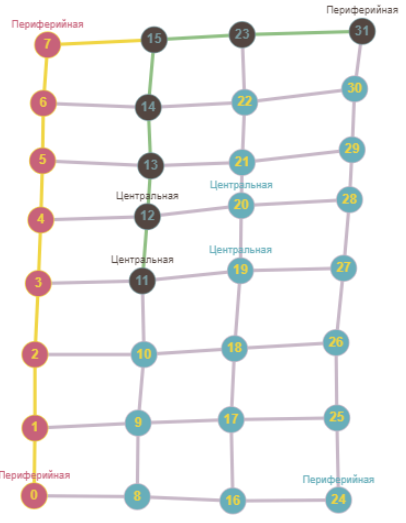


**Рисунок 10 - представление core group в виде графа**

**Диаметр графа равен:** 14 (0 > 1 > 2 > 3 > 4 > 5 > 6 > 7 > 15 > 27 > 31 > 39 > 47 > 55 > 63)

## **Расчет бисекционной пропускной способности**

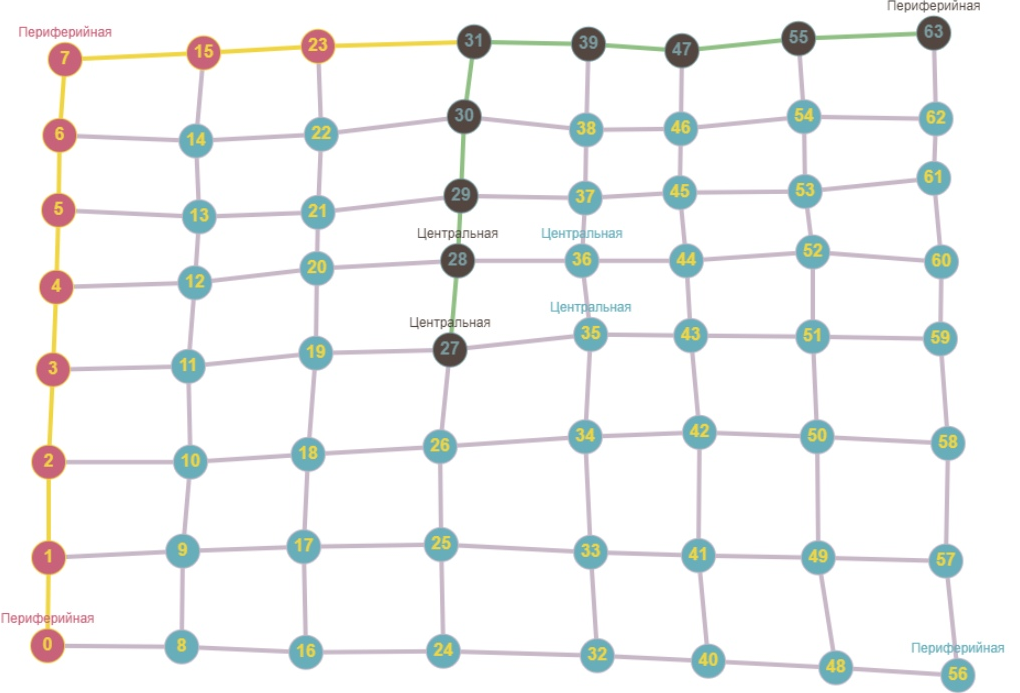
Бисекционная пропускная способность (bisection bandwidth) – суммарная пропускная способность каналов связи между двумя непересекающимися подмножествами машин системы (для худшего разбиения, минимальное значение)

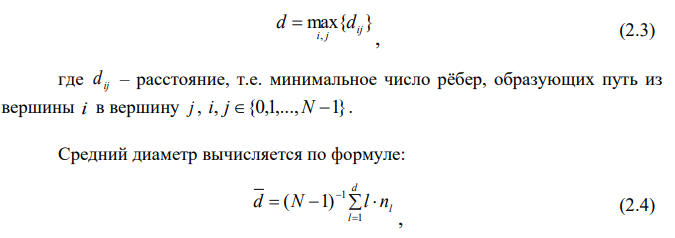


**Рисунок 11 расчет бисекционной пропускной способности**

Бисекционная пропускная способность равна: 10 (0 > 1 > 2 > 3 > 4 > 5 > 6 > 7 > 15 > 23 > 31)

## **Вычисление среднего диаметра**





Средний диаметр равен d = 1/63 \* (2 \* 1 + 3 \* 2 + 4 \* 3 + 5 \* 4 + 6 \* 5 + 7 \* 6 + 8 \* 7 + 7 \* 8 + 6 \* 9 + 5 \* 10 + 4 \* 11 + 3 \* 12 + 2 \* 13 + 1 \* 14) = 7.1111

# **СПИСОК ЛИТЕРАТУРЫ**

1. Дмитрий Волков. Стратегические ИТ: китайский сюрприз 863 // Открытые системы.СУБД. — 2010. — № 3. — С. 32–37. URL: [**www.osp.ru/os/2010/03/13001879**](http://www.osp.ru/os/2010/03/13001879).
2. The Sunway TaihuLight supercomputer: system and applications // Science China Information Sciences. — 2016. — Т. 59. — № 7. — С. 072001.
3. Dongarra J. Report on the Sunway TaihuLight System // PDF). www.netlib.org. Retrieved June. — 2016. — Т. 20.
4. <https://www.researchgate.net/publication/317303891_26_PFLOPS_Stencil_Computations_for_Atmospheric_Modeling_on_Sunway_TaihuLight>
5. <https://www.mkurnosov.net/teaching/docs/pct-lecture1.pdf>
6. <https://www.researchgate.net/publication/322871432_An_automatic_performance_model-based_scheduling_tool_for_coupled_climate_system_models>
7. <https://www.researchgate.net/publication/334685337_swATOP_Automatically_Optimizing_Deep_Learning_Operators_on_SW26010_Many-Core_Processor>