

**Второй Национальный Суперкомпьютерный Форум
(26 ноября 2013)**

Эксафлопсные суперкомпьютеры: достижения и перспективы

**В.С.Горбунов, Л.К.Эйсымонт
(ФГУП «НИИ «Квант»)**



Общая картина в области СКТ (на примере США)

- Внедрение результатов программы DARPA HPCS (2002-2010), коммерческие образцы и заказные суперЭВМ (2013-2017)**
- Выполнение программы DARPA UNPC (2010-2020) и программ DoE по экзамасштабным технологиям и суперЭВМ экза-уровня**
- Выполнение программы DARPA STARNet (с 2013 года) по оптимизации использования КМОП-технологий и разработки технологий пост-Муровской эры, зетта- и йотта-уровень**

Экзафлопсные проекты США

DARPA UNPC – проекты Runnemedede (Intel), Echelon (NVIDIA, Cray), X-Calibr(Sandia), Angstrom (MIT,Tilera)

DoE – проекты FastForward, C— Design Centers, X-Stack, OS/R, проекты крупных лабораторий и центров

Специальные проекты, выполняемые крупными фирмами – Intel, Cray, IBM, HP, Google...

Ожидаемые суперЭВМ экза-и выше уровня в США

- **2015-2017 – заказные суперкомпьютеры экза-уровня, оптимизация под CF- и DIS-задачи**
- **2018-2020 – эволюционная суперЭВМ экзафлопсного уровня NNSA DoE**
- **после 2022 – инновационная суперЭВМ экзафлопсного уровня OS/ASCR DoE**
- **после 2020 - заказные специализированные суперкомпьютеры зетта-уровня (~ 2020) и йотта-уровня (~ 2024), технологии RSFQ, QCA и квантовые аналогово-спиновые (типа D-Wave)**

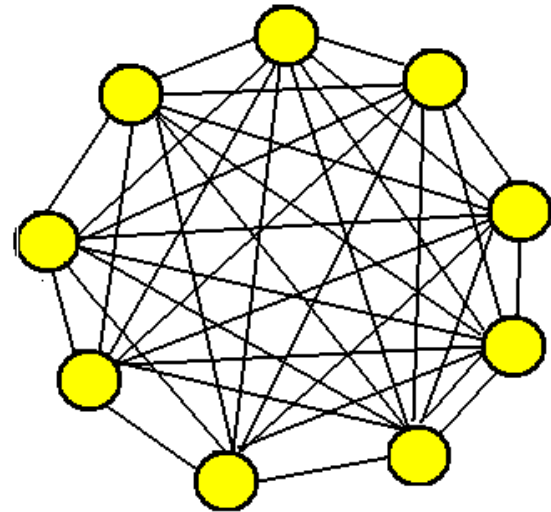
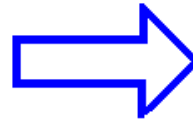
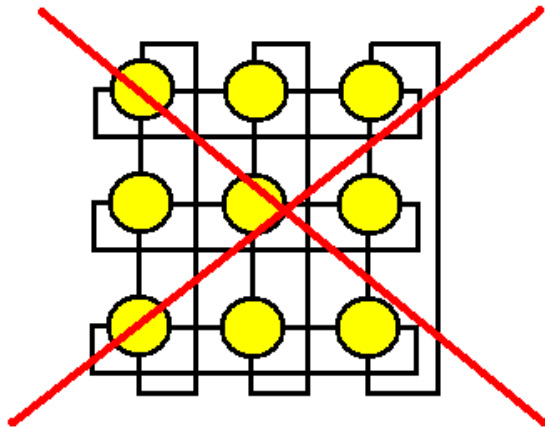
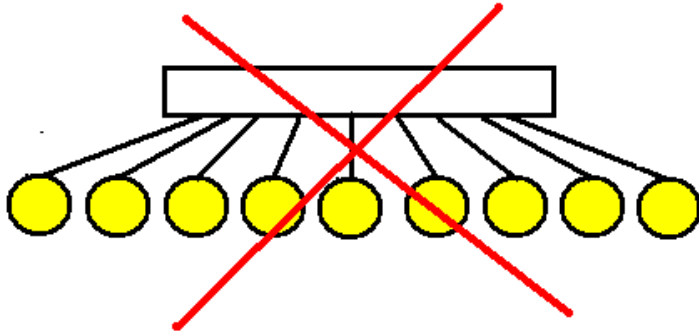
Вопрос для обсуждения N1

Какая инфраструктура исследований и разработок по экзамасштабной тематике возможна в России?

**Характерные черты будущих
эксамасштабных систем, проявившиеся уже
сейчас в лучших рейтинговых и
коммерческих суперкомпьютерах –
Tianhe-2, IBM Power 775, Cray XC30**

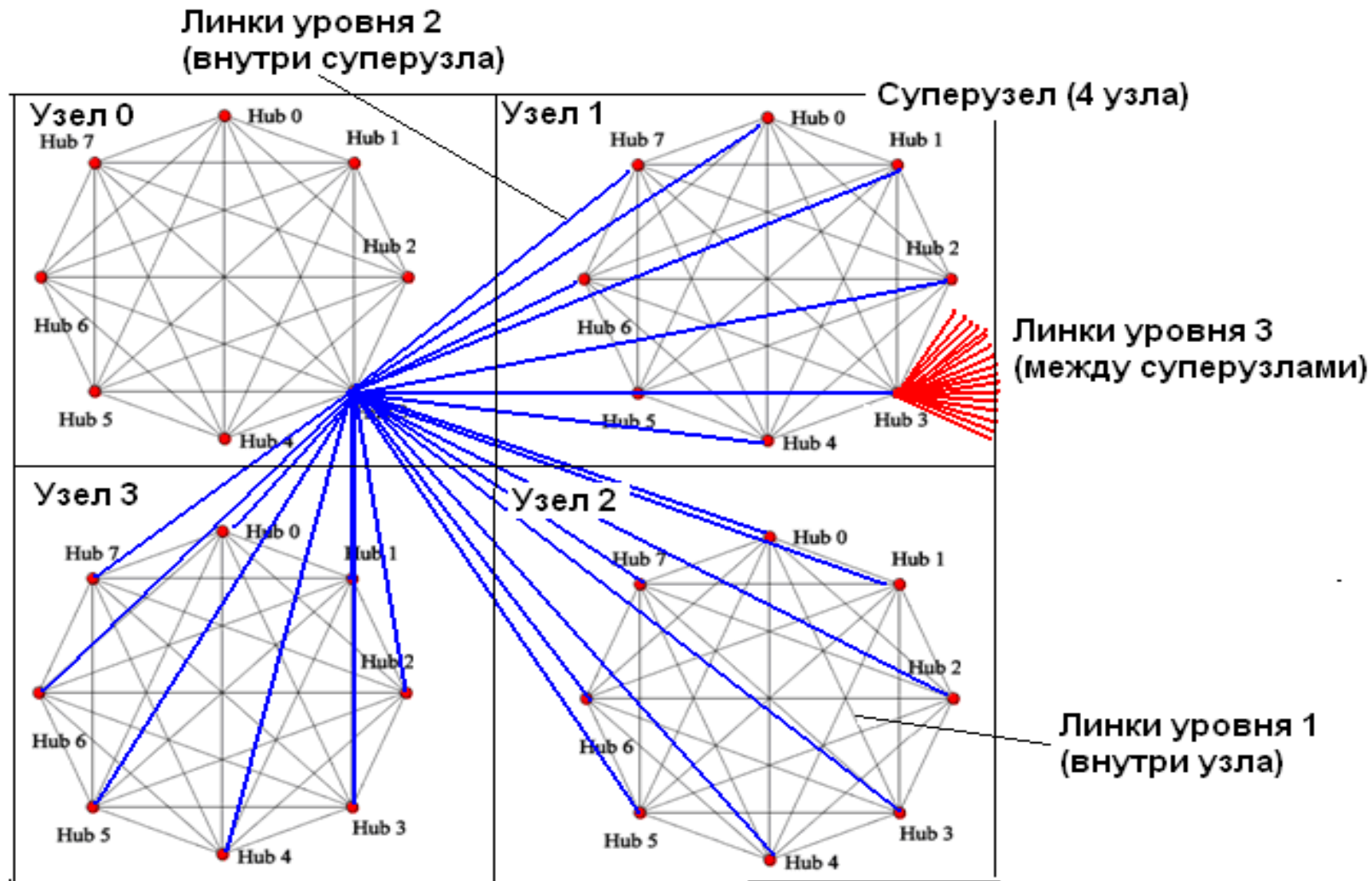
- Иерархичность**
- Гетерогенность**
- Гибридность**

ИЕРАРХИЧНОСТЬ



**Потому, что нужна
эффективная
глобально адресуемая
память огромного
объема !**

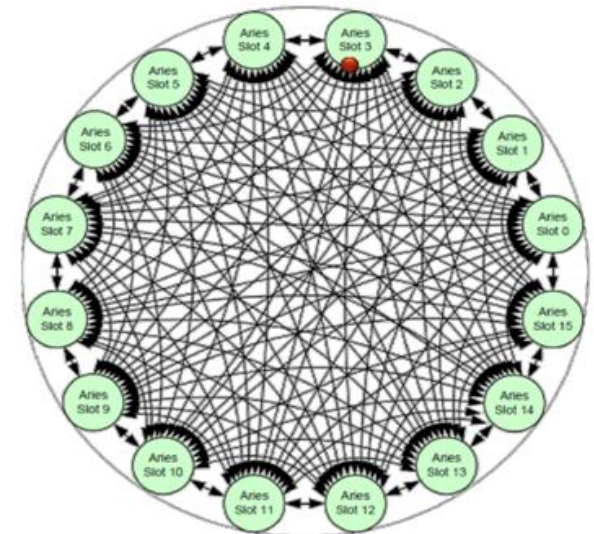
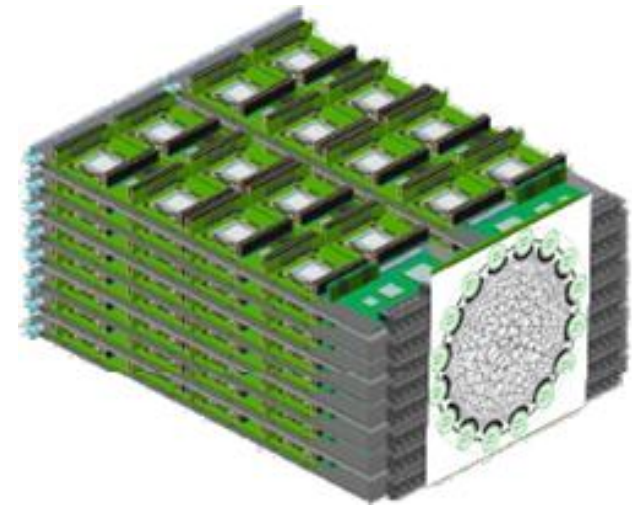
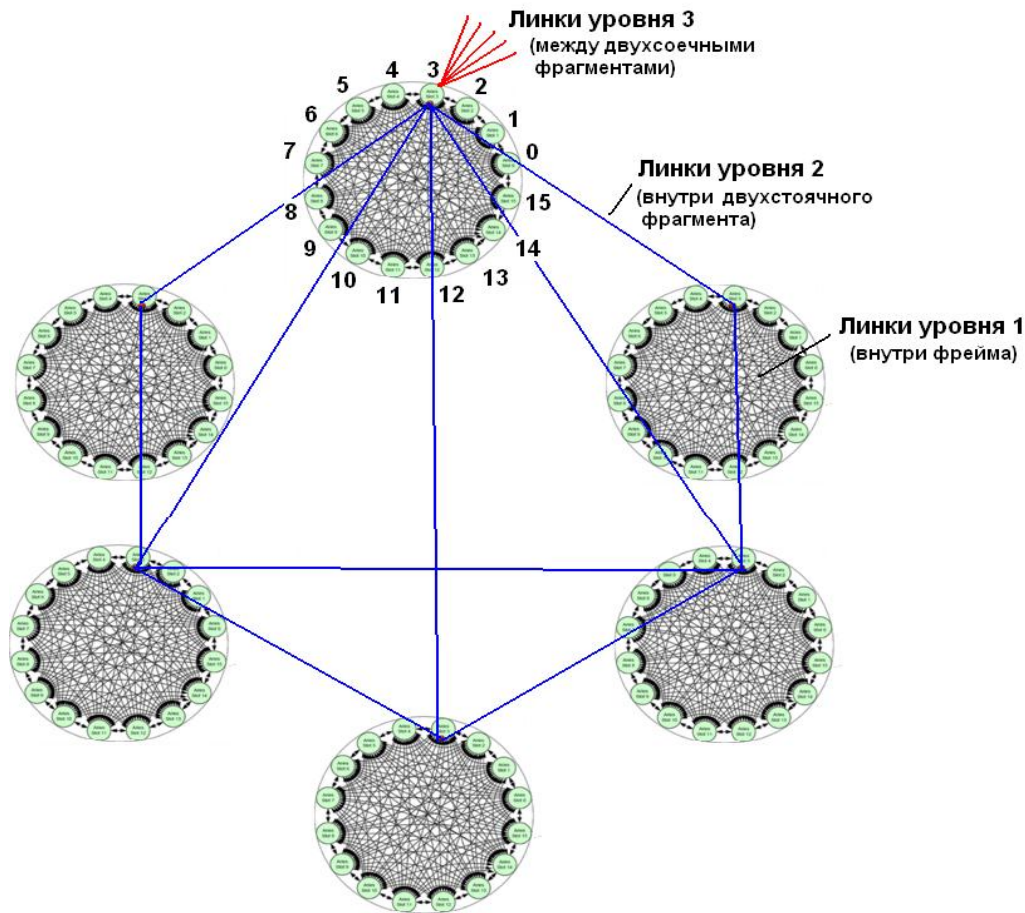
Многоуровневая сеть PERCS суперкомпьютера Power 775



Многоуровневая сеть суперкомпьютера Cray XC30

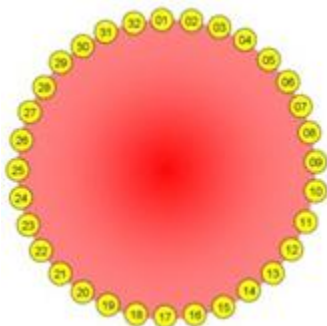
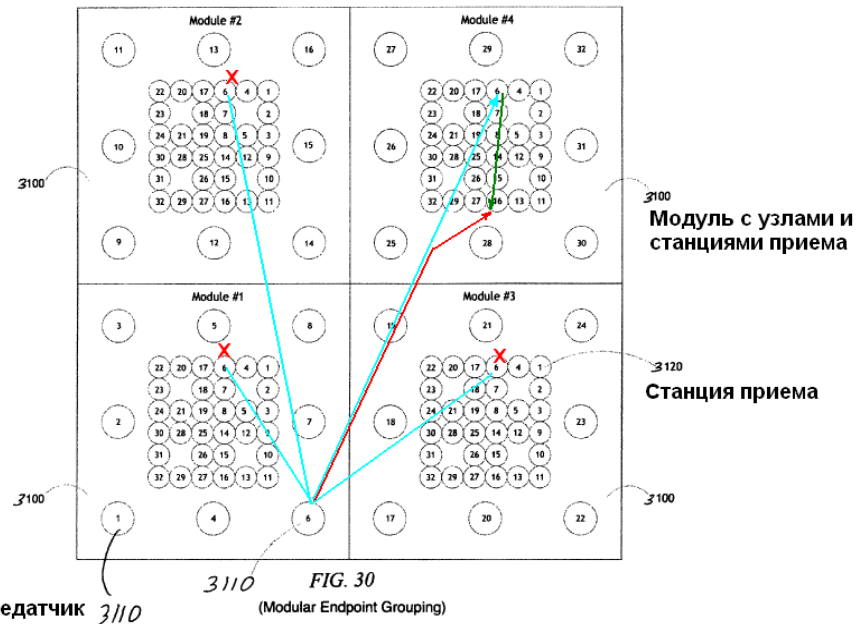
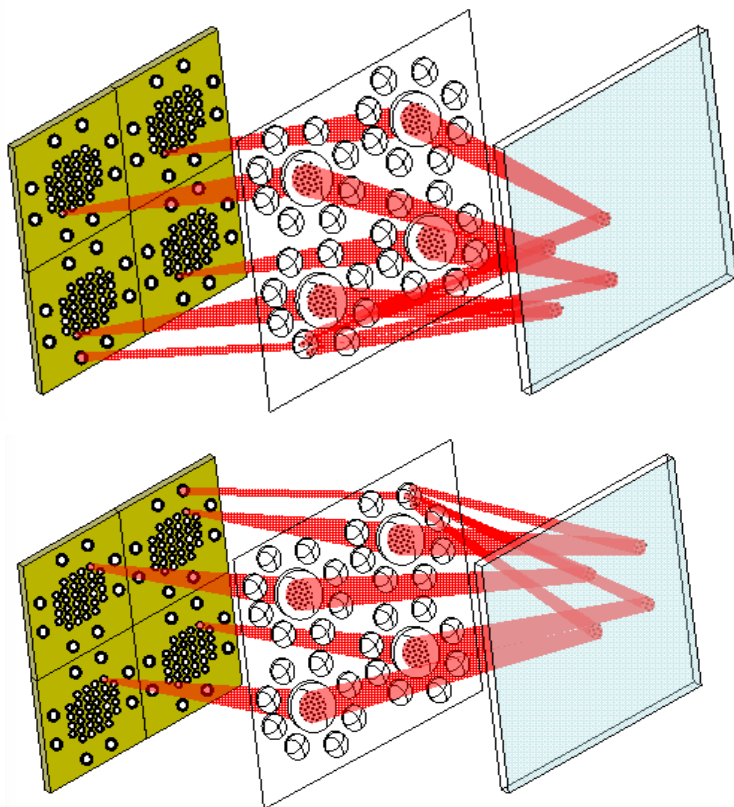
Фрейм

Двухстоечный
фрагмент



Оптические сети – фирма Lightfleet

без проводов типа broadcast)



Вопрос для обсуждения N2

Нужен ли России суперкомпьютер с аппаратной поддержкой глобально адресуемой памяти ?

ГЕТЕРОГЕННОСТЬ

**(наличие специализированных
фрагментов в суперкомпьютерах и ЦОД-ах)**

Tianhe-1A - 2048 FT-1000

(8 ядер, 64 треда, 1 GHz)

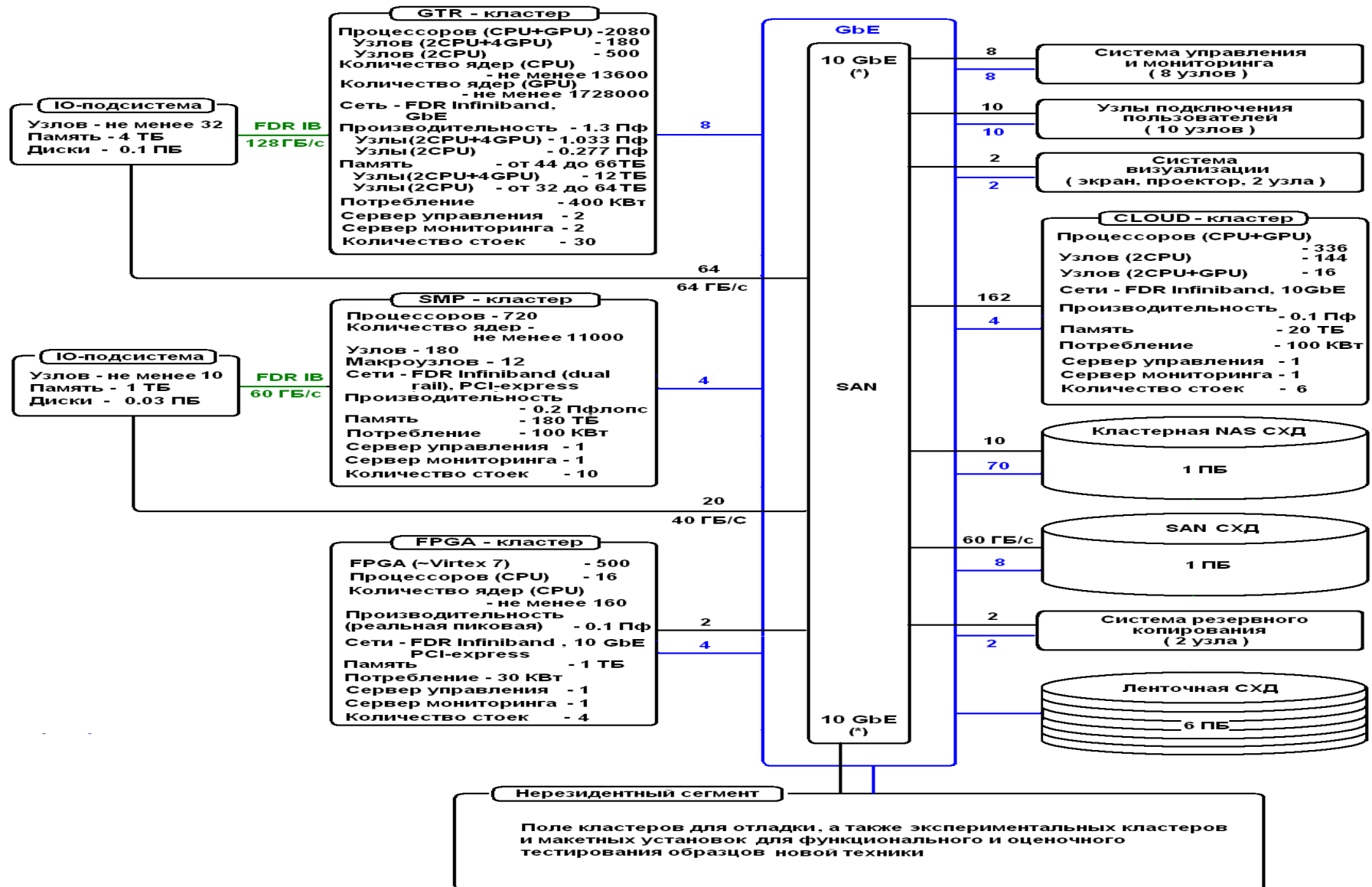
Tianhe-2 - 4096 FT-1500

**(16 ядер, 256 тредов, 1.8GHz,
144 GFlops, 60 W)**

**Специализированные блоки,
например, аналоговые – например,
D-Wave**

Гетерогенный ЦОД СПбГПУ

(2012 год – рекомендовано, а что будет на самом деле?)



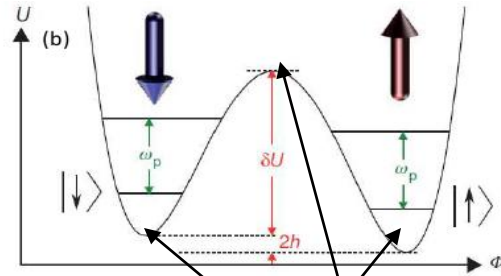
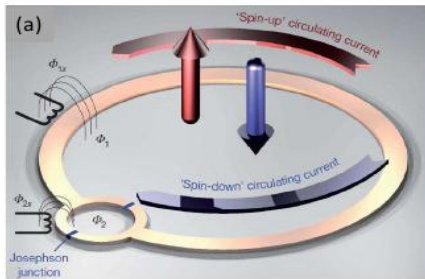
Важный пример необычного специализированного фрагмента – 128-кубитовый квантовый аналогово-спиновый суперкомпьютер D-Wave – 1 (сейчас уже выпущен 512-кубитовый D-Wave-2)

Умеет только минимизировать

$$E_{\text{Ising}}(s_1, \dots, s_N) = - \sum_{i=1}^N h_i s_i + \sum_{(i,j) \in E} J_{i,j} s_i s_j$$

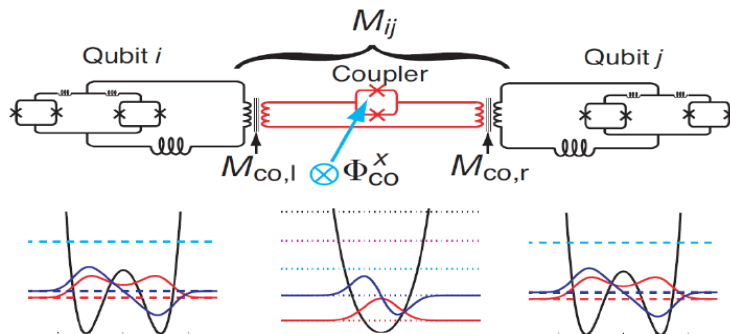
При заданных h_i и $J_{i,j}$ (настроечные коэффициенты) находит такие s_i – спины, +1 или -1, что достигается глобальный минимум минимизации такого

функционала сводятся многие задачи, в т.ч. SAT-проблема!

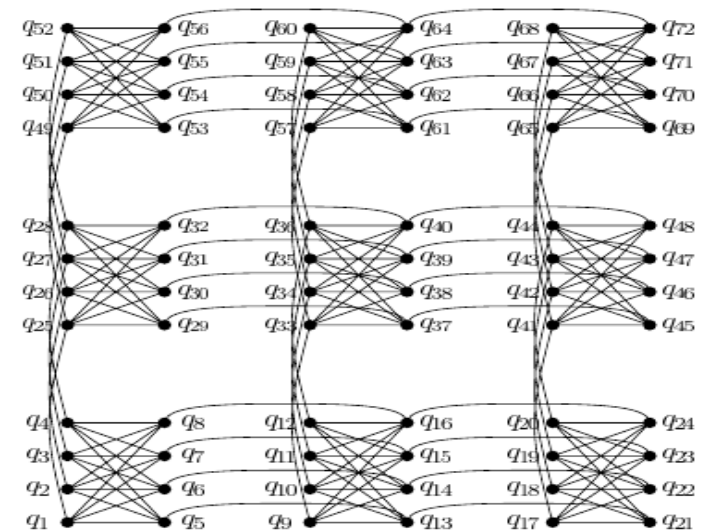


Одно или другое, после минимизации

Перед минимизацией

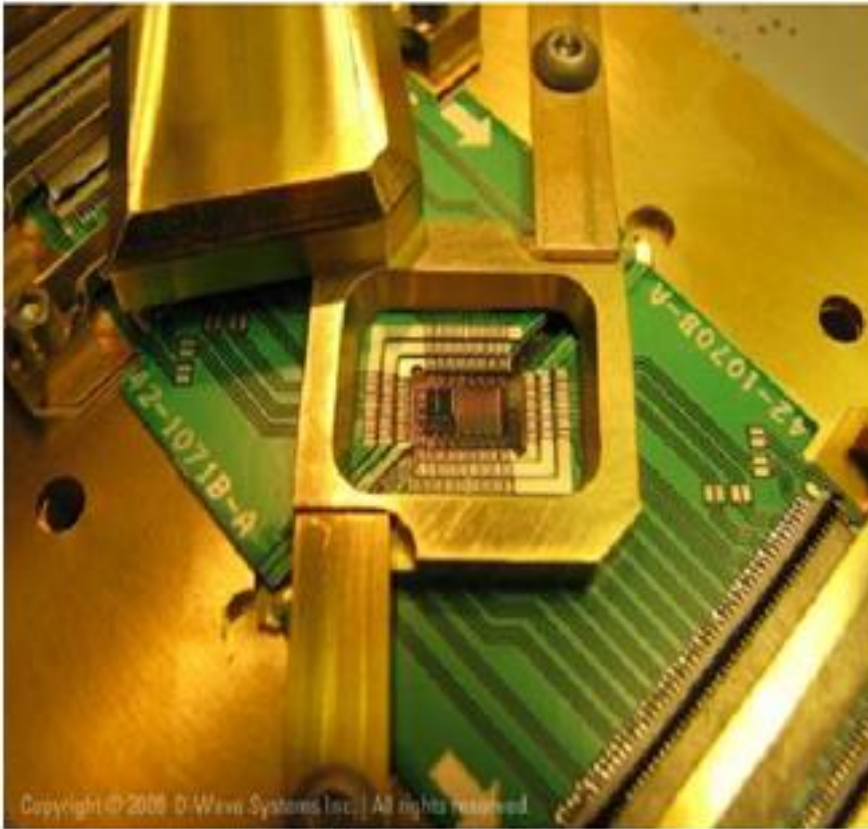


Соединение двух q-битов

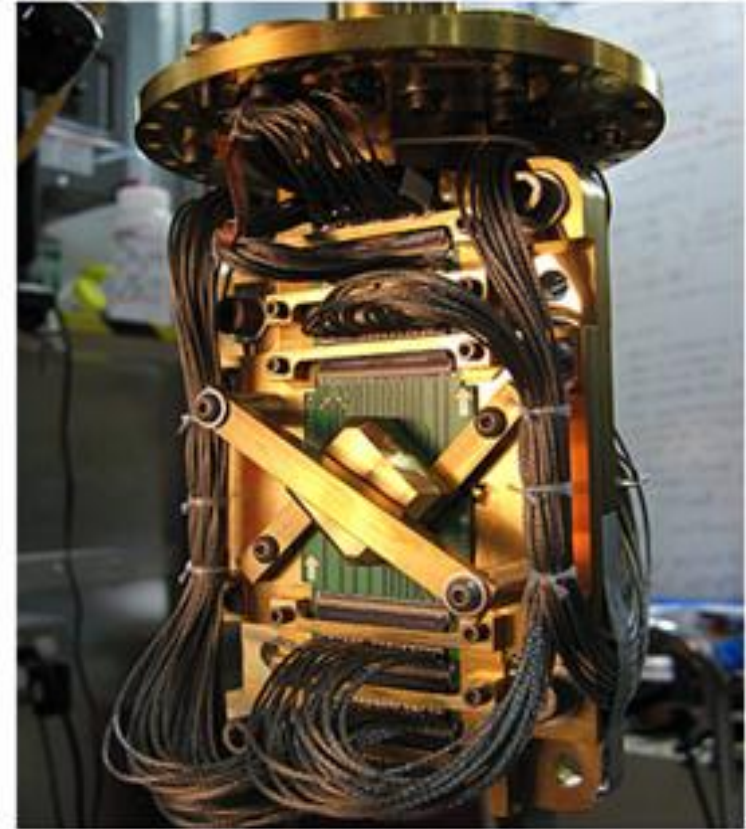


Логическое соединение 128 q-битов

Квантовый аналогово-спиновый суперкомпьютер D-Wave - 2



А. Процессор в камере непосредственного жидкостного охлаждения



Б. Процессор в конструктиве охлаждения и защиты

Рабочая температура ~ 20 мК, но работает в 10000 раз быстрее одного микропроцессора Xeon !

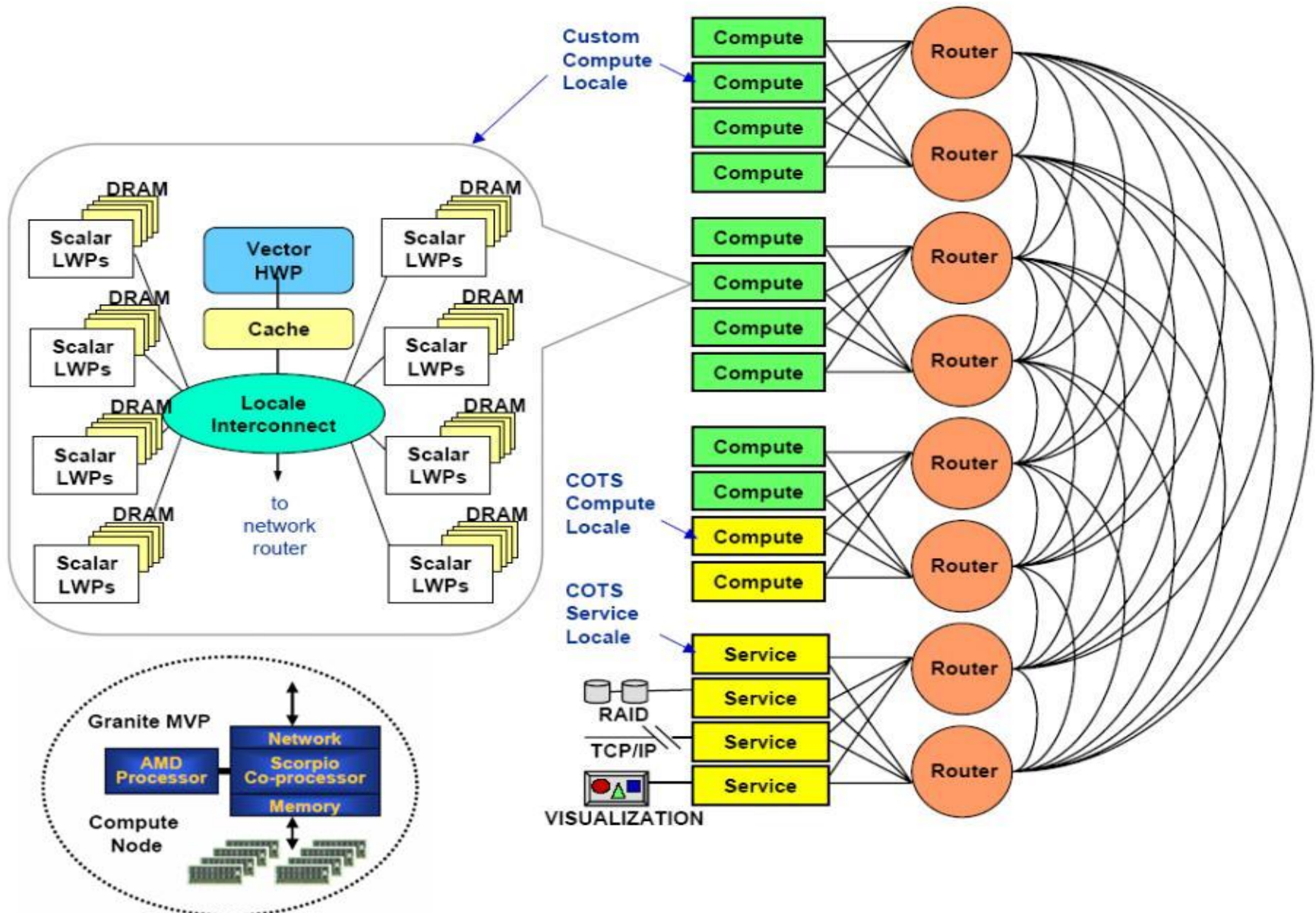
Вопрос для обсуждения N3

Нужны ли России суперкомпьютеры и центры обработки данных с гетерогенной структурой ?

ГИБРИДНОСТЬ

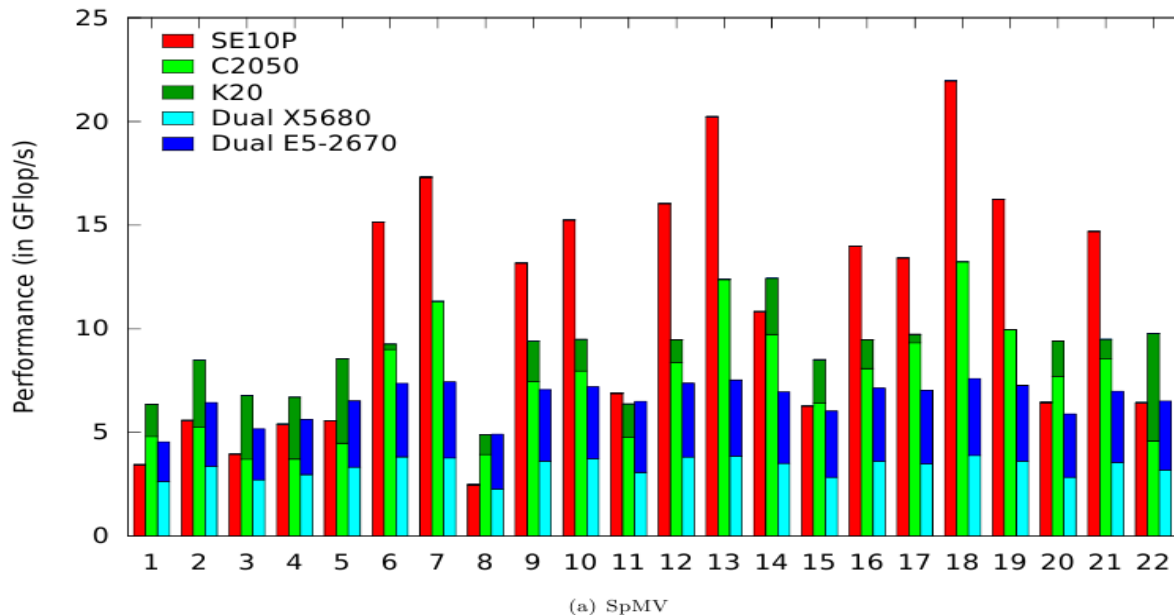
**Массово-многоядерные
мультитредовые
микропроцессоры,
однородные и гибридные**

Фирма Cray, проект Cascade программы DARPA HPSCS (~ 2003 год)



HPCG (SpMV) vs HPL(Top500)

Установка	Реальная производительность на SpMV (% от пиковой)
2 x Intel Xeon X5680 (Westmere)	0,78 - 1,09
2 x Intel Xeon E5-2670 (Sandy Bridge)	1,36 - 2,12
NVIDIA Tesla C2050 (Fermi).	0,68 - 2,52
Tesla K20 (Kepler)	0,5 - 1,3
Intel Xeon Phi	0,25 - 2,25



2xW - 320 GF

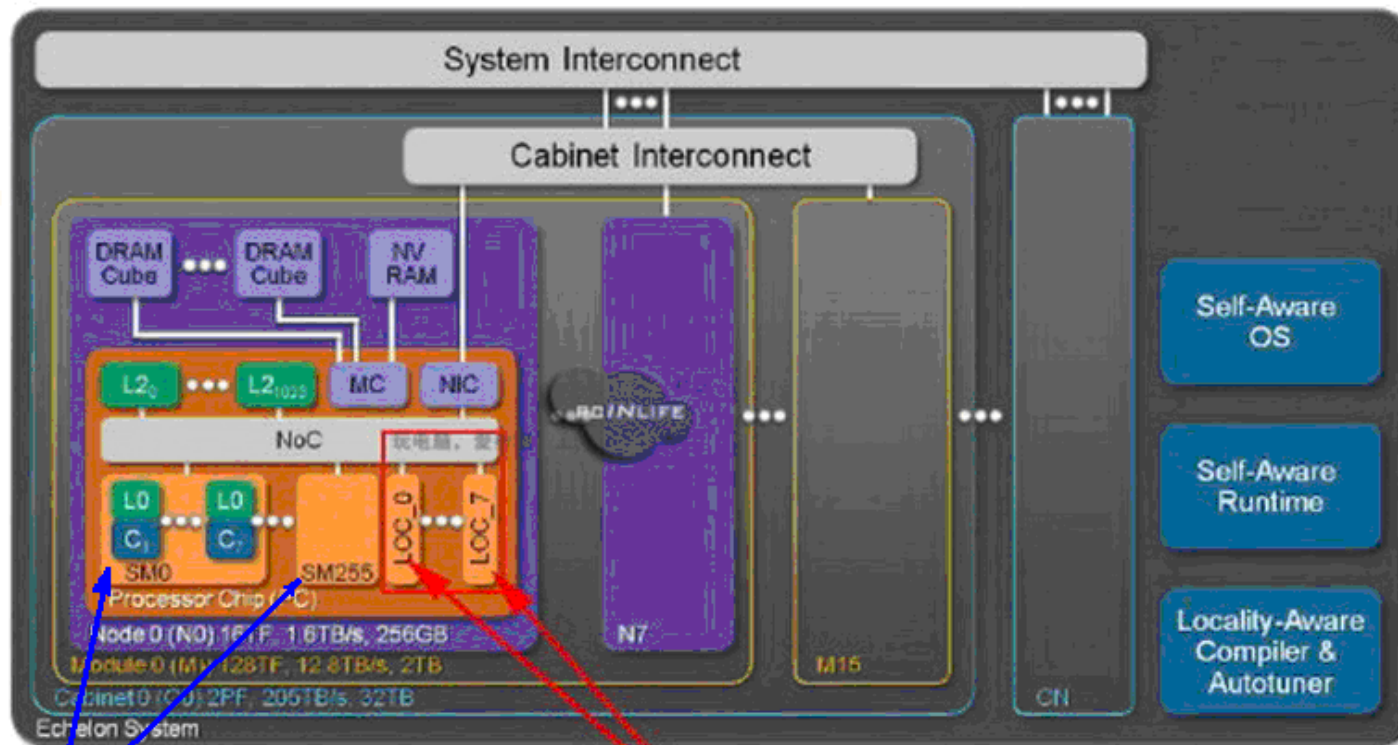
2xSB - 330 GF

F - 512 GF

K - 1 TF

Phi - 1 TF

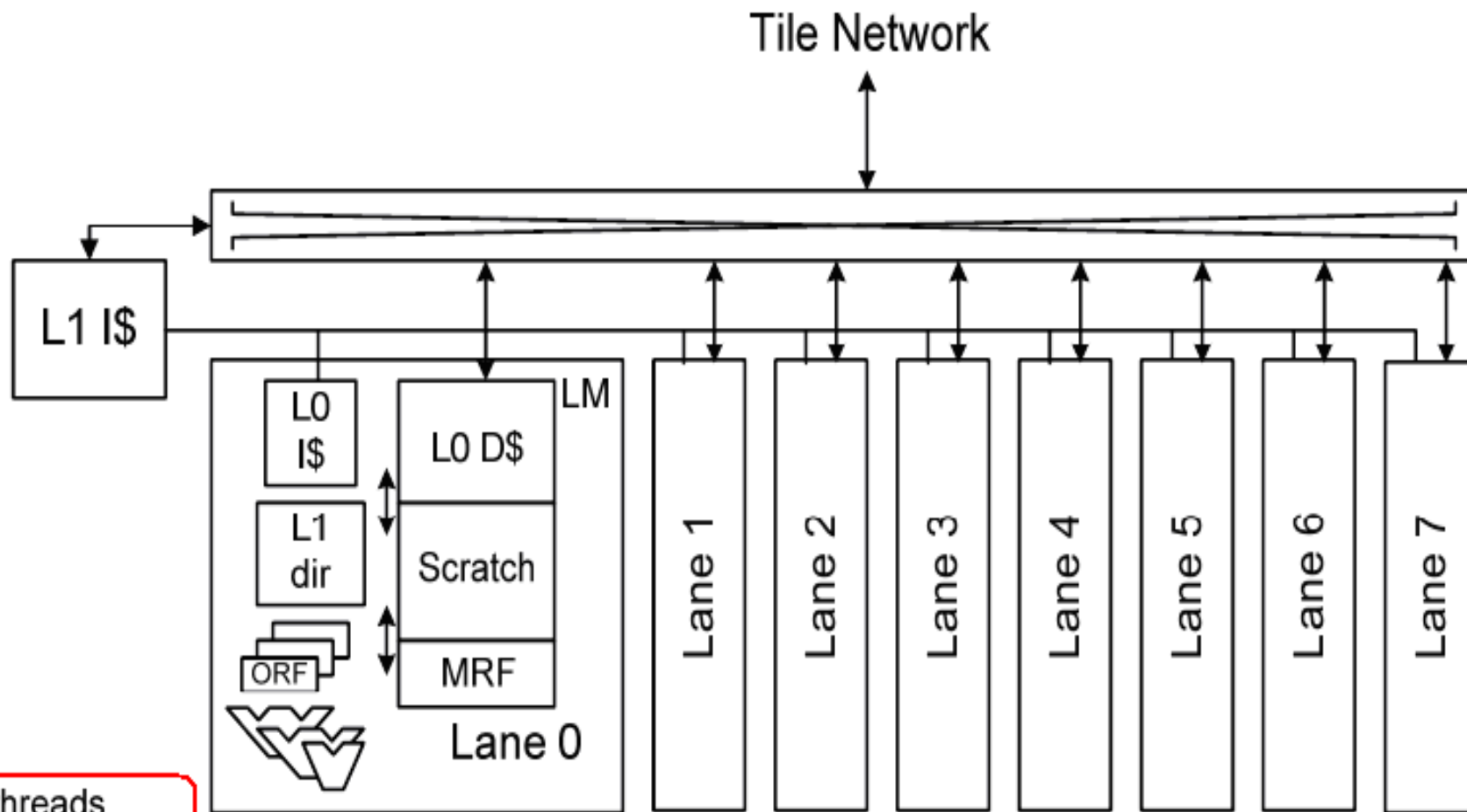
Проект Echelon (NVIDIA, Cray, ORNL, Lockhead Martin, 8 университетов)



Массово-мультитредовые
ядра

Суперскалярные ядра

Структура SM-ядра



512 threads

32 active threads

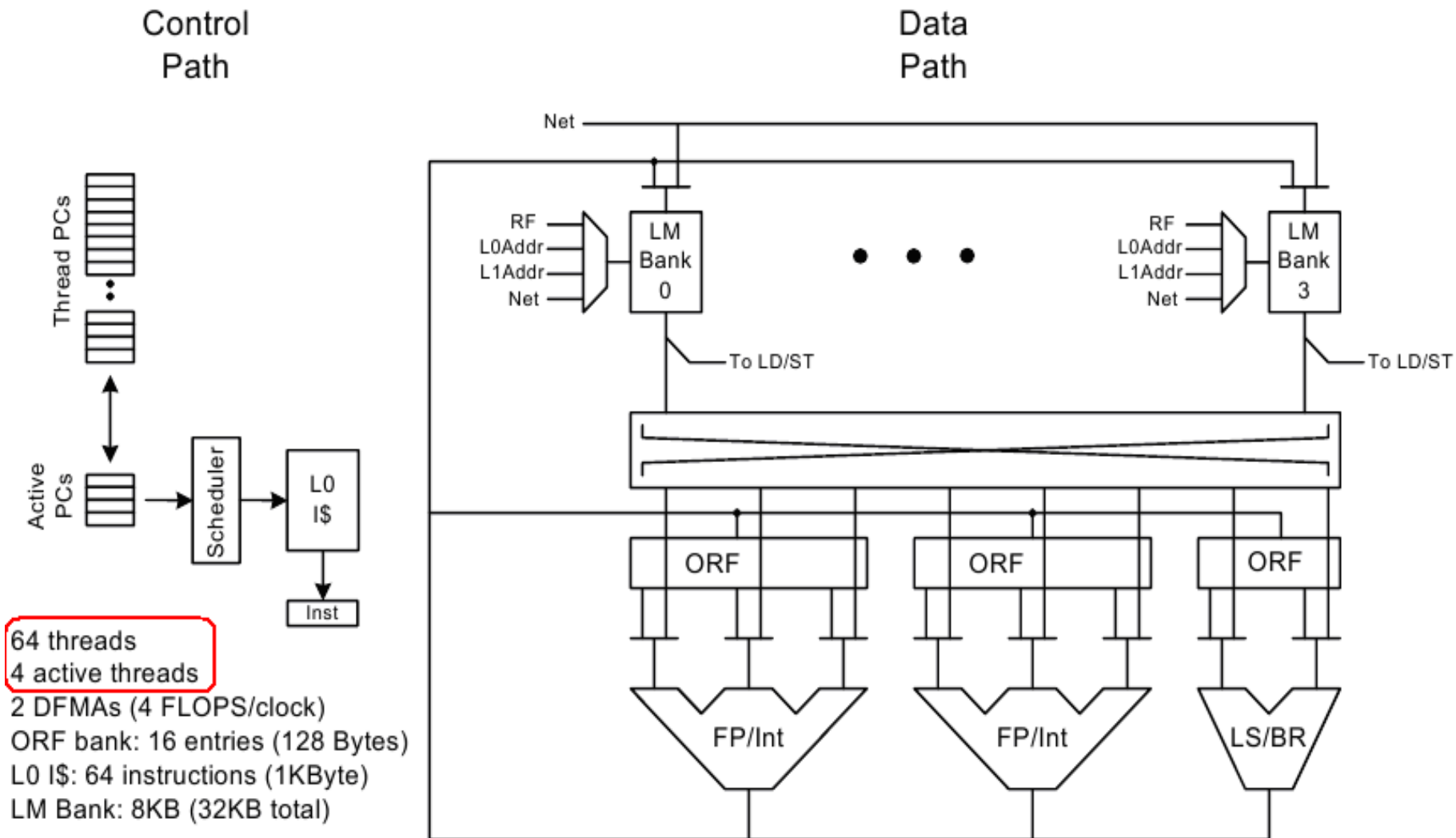
16 DFMAs (32 FLOPs/clock)

L1 I\$: 2K instructions (32KB)

RF/Scratch/D\$: 256KB

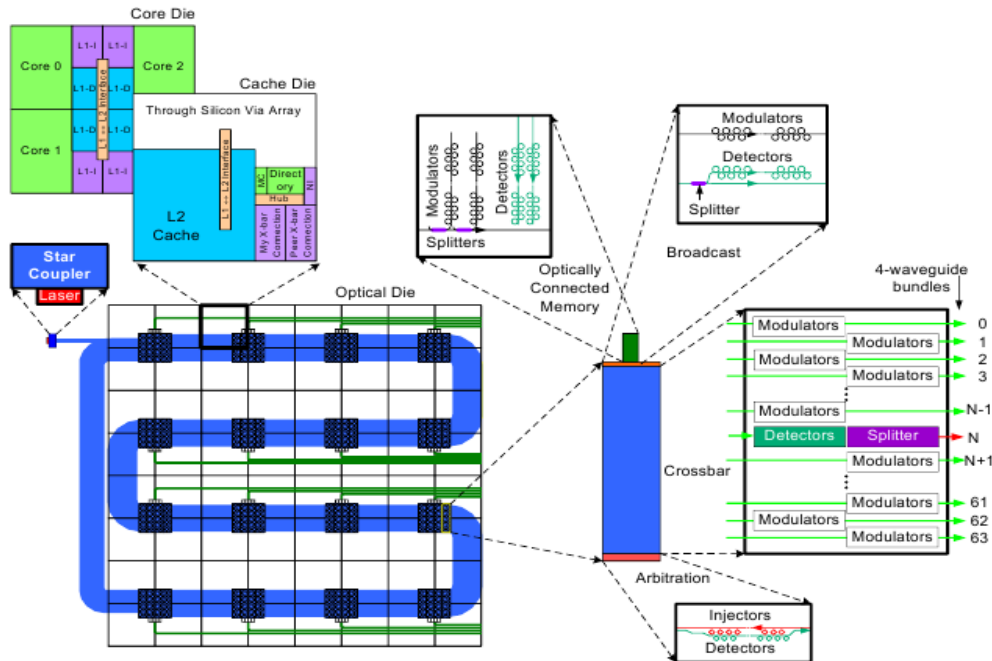
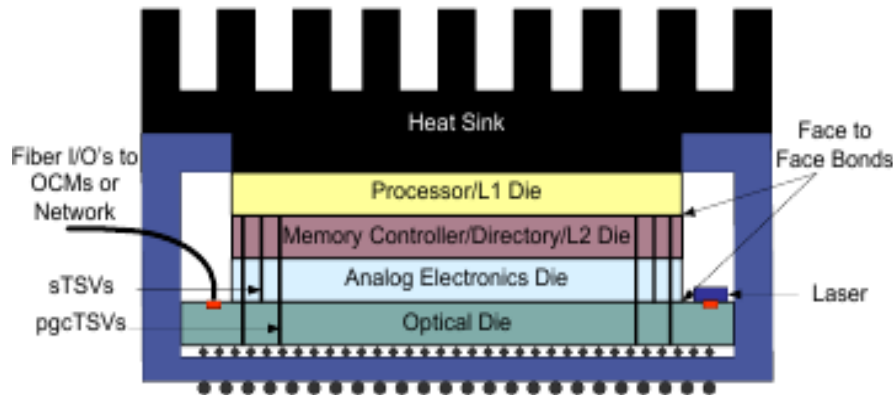
L0 caches in other lanes form L1 cache

Полоса обработки (Lane) SM-ядра



Проект Corona (Hewlett-Packard)

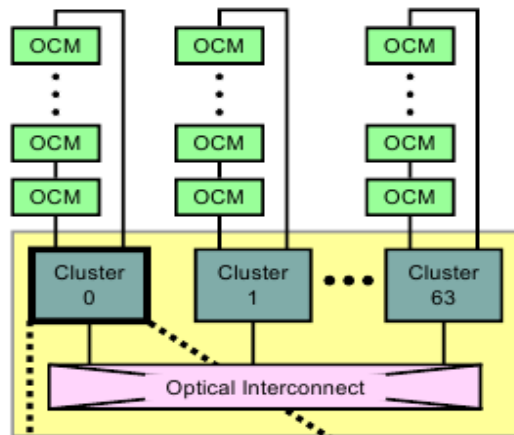
3D-процессорный модуль



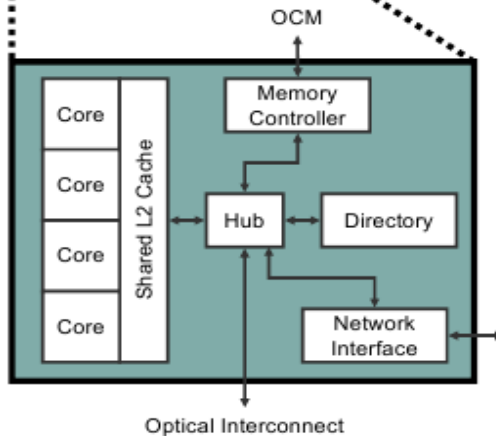
Resource	Value
Number of clusters	64
Per-Cluster:	
L2 cache size/assoc	4 MB/16-way
L2 cache line size	64 B
L2 coherence	MOESI
Memory controllers	1
Cores	4
Per-Core:	
L1 ICache size/assoc	16 KB/4-way
L1 DCache size/assoc	32 KB/4-way
L1 I & D cache line size	64 B
Frequency	5 GHz
Threads	4
Issue policy	In-order
Issue width	2
64 b floating point SIMD width	4
Fused floating point operations	Multiply-Add

Photonic Subsystem	Waveguides	Ring Resonators
Memory	128	16 K
Crossbar	256	1024 K
Broadcast	1	8 K
Arbitration	2	8 K
Clock	1	64
Total	388	≈ 1056 K

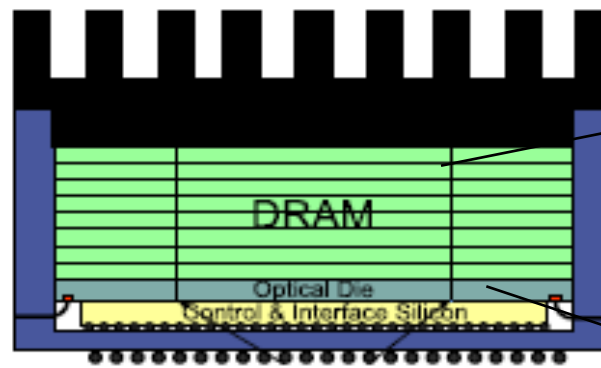
Проект Corona – 3D-модуль памяти



(a)



(b)

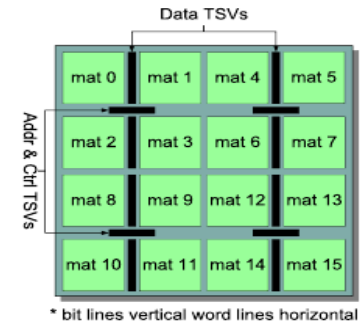


Optically Connected Memory

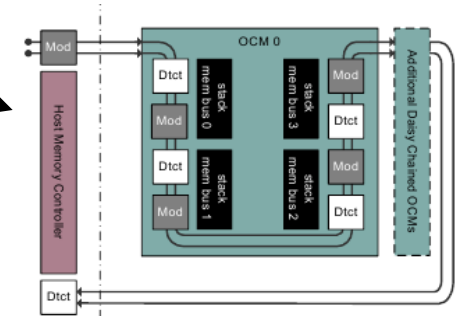
$$BW_{\text{mem}} = 10 \text{ TB/s}$$

ECM – 160W (2 mW/Gb/s)

OCM – 6W (0.078 mW/Gb/s)



* bit lines vertical word lines horizontal



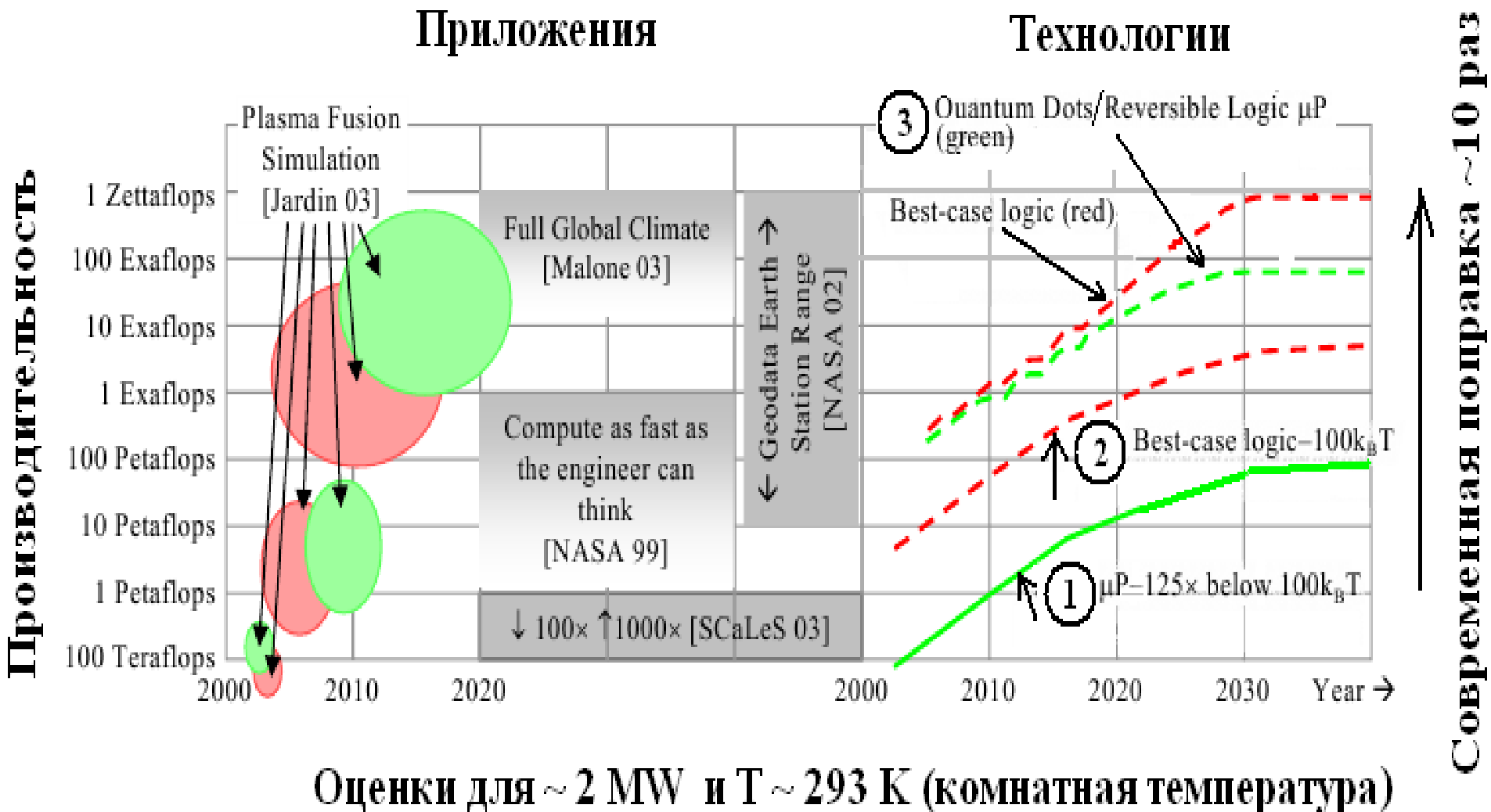
Resource	OCM	ECM
Memory controllers	64	64
External connectivity	256 fibers	1536 pins
Channel width	128 b half duplex	12 b full duplex
Channel data rate	10 Gb/s	10 Gb/s
Memory bandwidth	10.24 TB/s	0.96 TB/s
Memory latency	20 ns	20 ns

Вопрос для обсуждения N4

**Нужно ли России расширение типов выпускаемых микропроцессорных ядер?
Например, не пора ли освоить выпуск массово-мультитредовых ядер?**

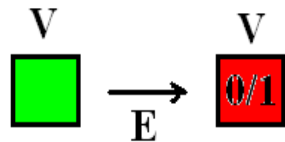
Физические ограничения и пост-Муровская эра

Прогноз (2004 год) роста потребностей производительности и возможностей создаваемых суперкомпьютеров



Ограничение Лэндауэра, принцип Неймана-Лэндауэра и реверсивная логика

(А) Минимальная энергия преодоления шумов электронов в микроскопическом объеме независимо от технологий



$dL \sim 1.5 \text{ нм}$
 $dT \sim 0.04 \text{ ps (40 fs)}$

при $T = 293 \text{ K}$

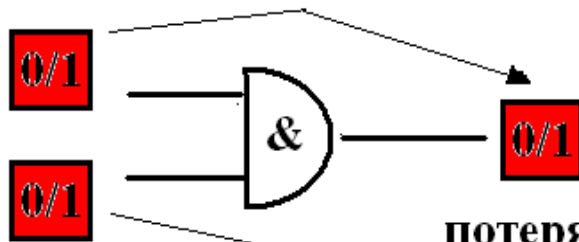
$$E = kT \ln 2 = 2.85 \times 10^{-21} \text{ J} = 2.85 \text{ zJ (зептоджоулей)}$$

$$2.85 \times 10^{-9} \text{ pJ (обозначают кТ)}$$

при $T = 4 \text{ K}$

$$E_{\text{bit}}\{\text{RSFQ}\} \sim 10^{-19} \text{ J} = 10^{-7} \text{ pJ} \sim 100 \text{ кТ}$$

(Б) Ограничение Лэндауэра (1961)



потеря в виде тепла $E_b > kT \ln 2$

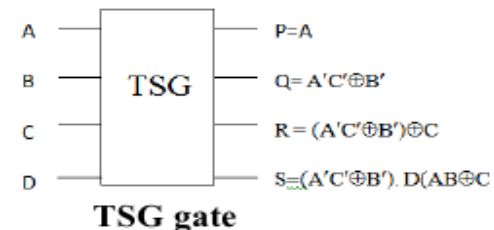
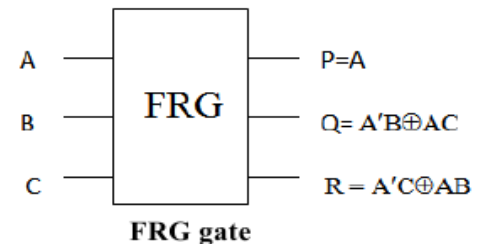
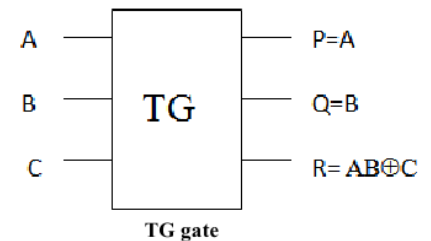
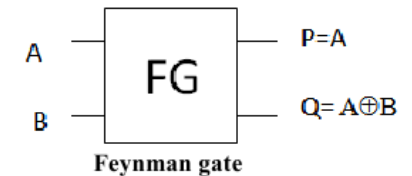
КМОП : 100000 - 1000000 кТ

ERSFQ: схема сумматора, 1000 кТ/bit

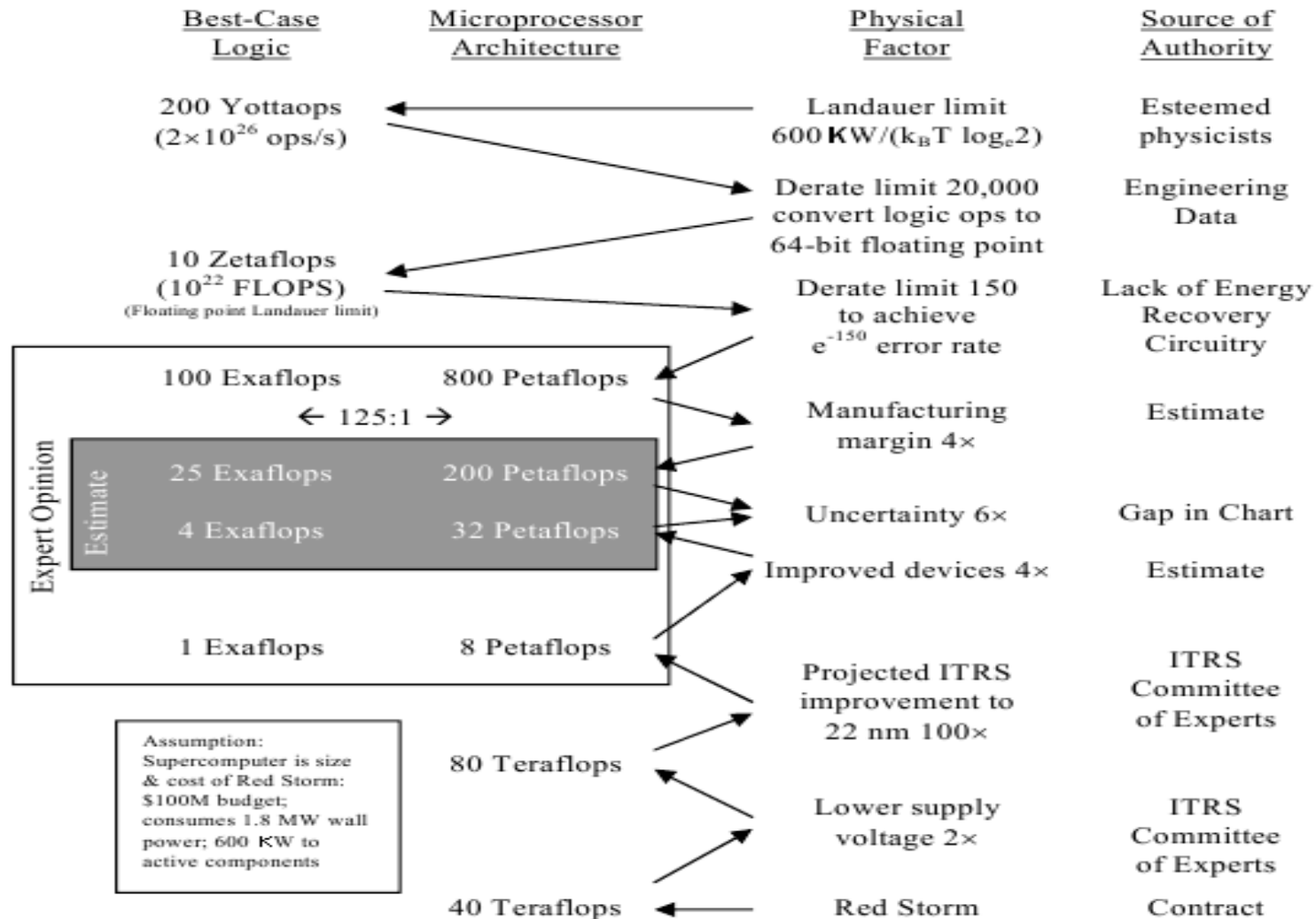
: схема сдвигового регистра, 3 кТ/bit

(Г) Принцип Неймана-Лэндауэра (Беннет, 1973)
 Для реверсивных машин возможно $E_b \ll kT \ln 2$

(В) Реверсивные логические элементы ($N_{\text{ВХ}} = N_{\text{ВЫХ}}$)

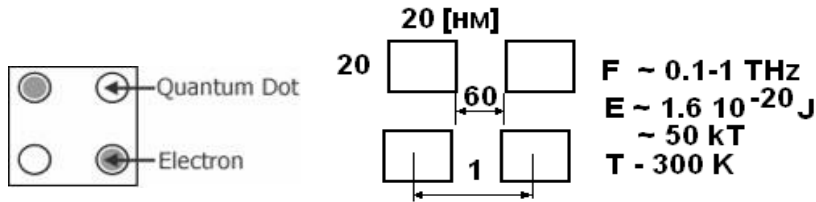


Как вычисляется предел производительности нереверсивных суперкомпьютеров, “точка Стерлинга”, для мощности 600 KW. (DeBenidics, 2004)

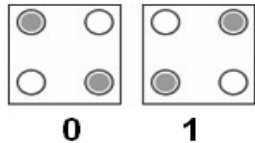


Элементы QCA-логики

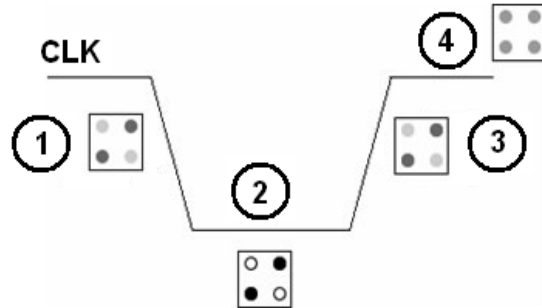
4-х точечная QCA-ячейка



QCA-ячейки в состоянии 0 и 1



Управление QCA-ячейкой



Сигнал CLK влияет на уровень тунnelирования:

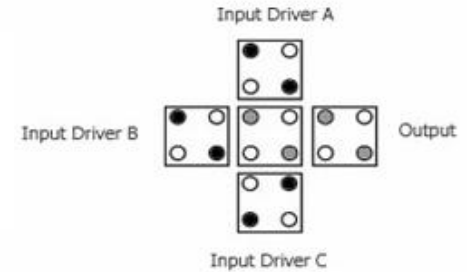
- 1 - переходное состояние (меняется уровень)
- 2 - стабильное состояние (уровень высокий)
хранение информации
- 3 - переходное состояние (меняется уровень)
- 4 - рабочее состояние (уровень низкий)
обработка информации, логика клеточного автомата

QCA majority gate

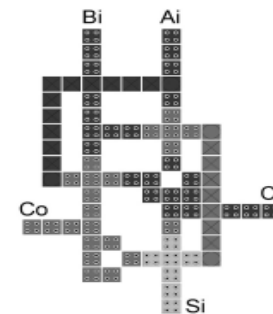
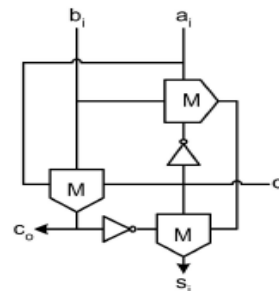
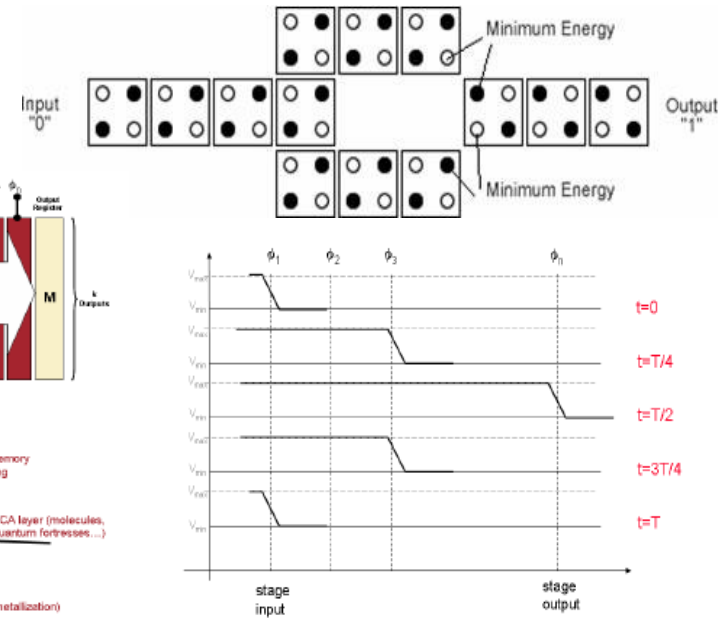
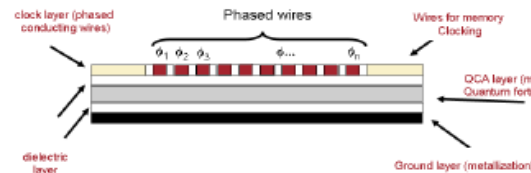
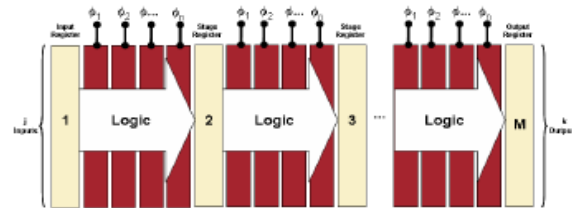
$$M(a,b,c) = ab + bc + ca$$

$$a.b = M(a,b,0)$$

$$a+b = M(a,b,1)$$



QCA inverter



Разряд сумматора

Вопрос для обсуждения N5

Нужна ли России вертикальная интеграция архитектурно-программных разработок и разработок по элементно-конструкторской базе в рамках крупных проектов экза- и зетта-суперкомпьютеров?

Выводы

- 1. За рубежом (США, Китай, Япония) началось внедрение результатов проведенных там НИР прошлого десятилетия по петафлопсной тематике**
- 2. Усилился объем и глубина исследований по экзамасштабной тематике и новой элементной базе пост-Муровской эры. Начало выполнения крупной экзафлопсной программы в DoE ожидается не ранее 2014 года**
- 3. Важнейшие направления системных исследований и разработок: иерархические сети, гибридные массово-мультитредовые многоядерные микропроцессоры, гетерогенность (включая специализированные фрагменты, в том числе фрагменты аналогового типа)**

4. Важнейшие направления по элементно - конструкторской базе: нанофотоника, сверхпроводниковая электроника, квантовая электроника (спиновая, клеточные автоматы, реверсивная логика), TSV-соединения, оптоволоконные соединения через матрицы линз и лазеров.

5. Новая элементно-конструкторская база позволит преодолеть зетта- и йотта- уровень производительности, это намечено в первую очередь в заказных специализированных зарубежных суперкомпьютерах

Публикации, в которых изложены основные положения доклада

1. В.Горбунов, Г.Елизаров, Л.Эйсымонт. Экзафлопсные суперкомпьютеры: достижения и перспективы. Открытые системы, N7, 2013.
2. Д.Андрюшин, В.Горбунов, Л.Эйсымонт. Перспективные особенности Tianhe-2. Открытые системы, N8, 2013.
3. В.Горбунов. НРС. Оценить, измерить, оптимизировать. Суперкомпьютеры, 3(15) осень 2013.
4. В.Горбунов, Л.Эйсымонт. Комплексная методика тестирования производительности суперкомпьютеров, профессиональный подход. Вычислительные методы и программирование. 2013 (в печати).
5. А.Речинский, В.Горбунов, Л.Эйсымонт. Суперкластер с глобально адресуемой памятью. Открытые системы, №7, 2011.

Вопросы ?

Горбунов Виктор Станиславович (gorbunov@rdi-kvant.ru)

Эйсымонт Леонид Константинович (eisymont@rdi-kvant.ru)