

Семинар “Технологии разработки и анализа программ”

**Состояние и перспективы развития
информационных технологий,
актуальные задачи**

**Л.К.Эйсымонт
(16 октября 2014 года)**

Общая картина в области СКТ

- Внедрение результатов программы DARPA HPCS (2002-2010), коммерческие образцы и военные 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

Ожидаемые результаты по суперЭВМ экза- и более уровня

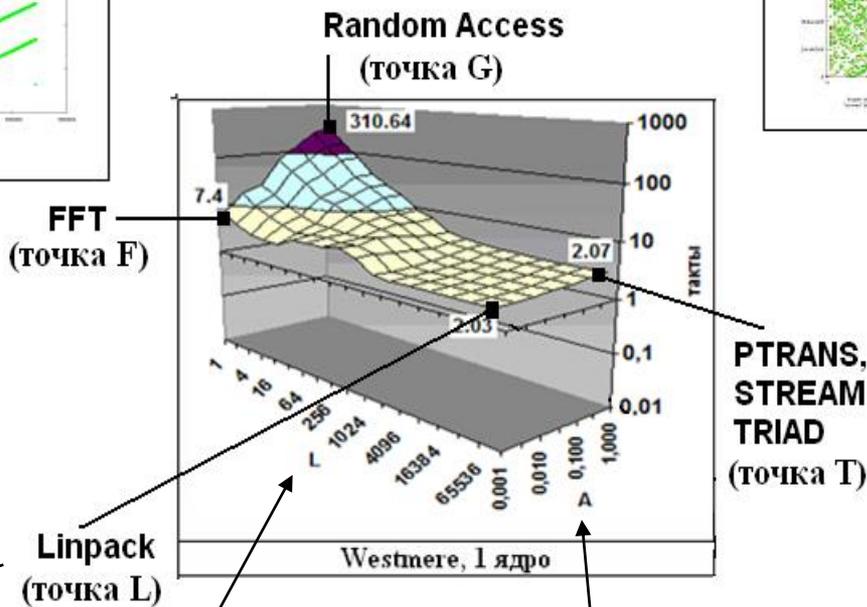
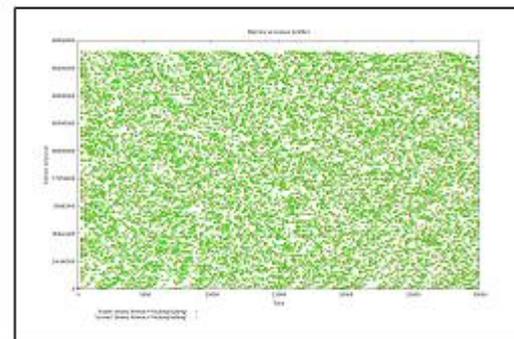
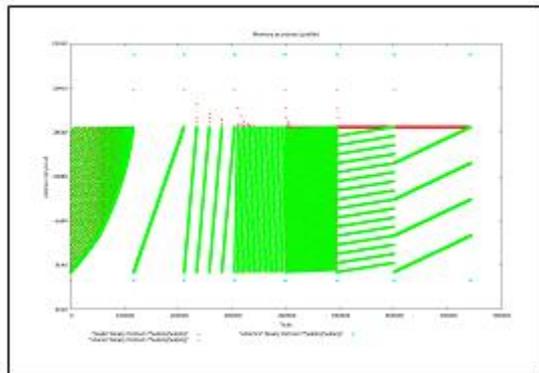
-2015-2017 – военные суперкомпьютеры (ВКСН) экза-уровня, CF- и DIS-задачи

-2018-2020 – эволюционная суперЭВМ экзафлопсного уровня NNSA DoE

- после 2022 – инновационная суперЭВМ экзафлопсного уровня OS/ASCR DoE

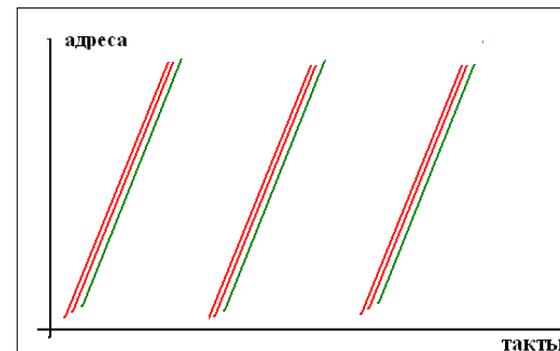
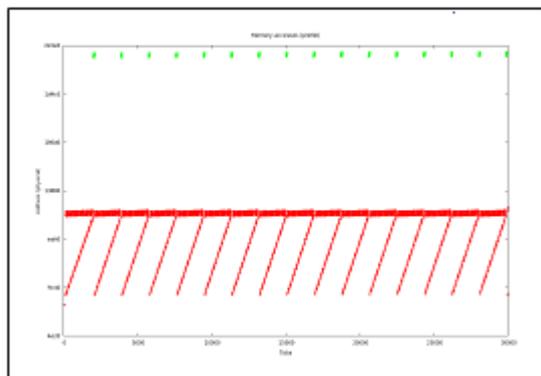
- после 2020 - военные суперкомпьютеры (ВКСН) зетта-уровня (~ 2020) и йотта-уровня (~ 2024), технологии RSFQ, QCA и квантовые аналогово-спиновые (~D-Wave)

Разные режимы пространственно-временной локализации и эффективность работы памяти



Пространственная локализация

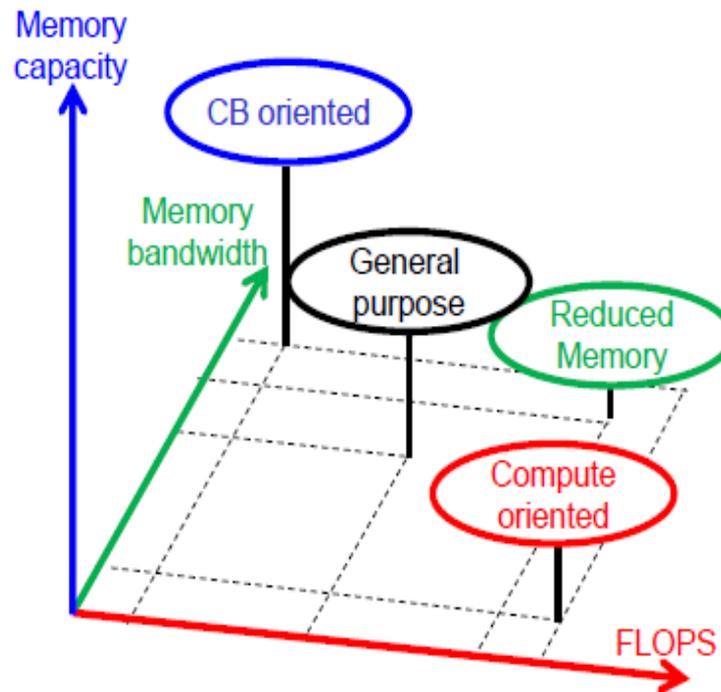
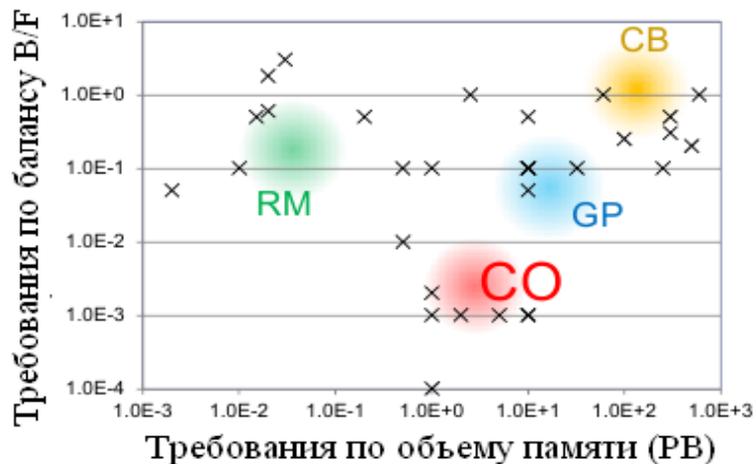
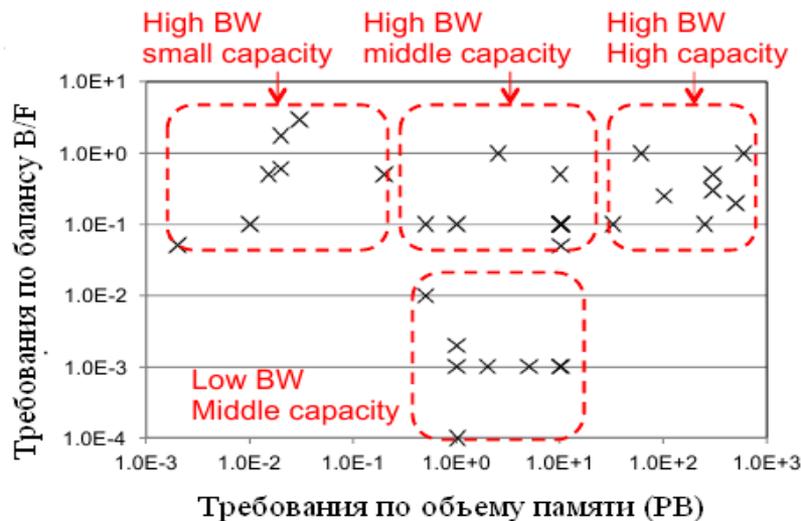
Временная локализация



Рейтинг Top500 на вычислительных задачах - тесты HP Linpack и HP CG

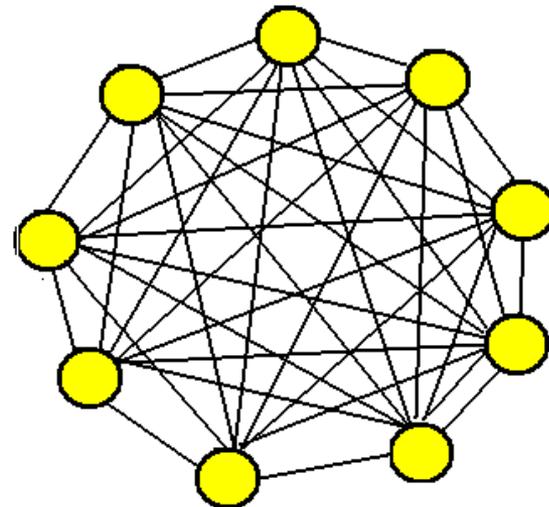
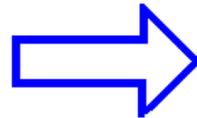
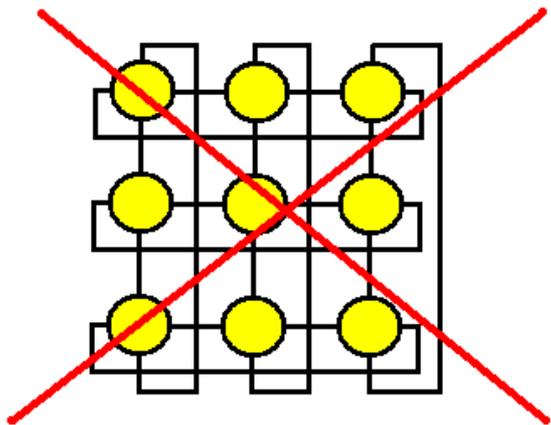
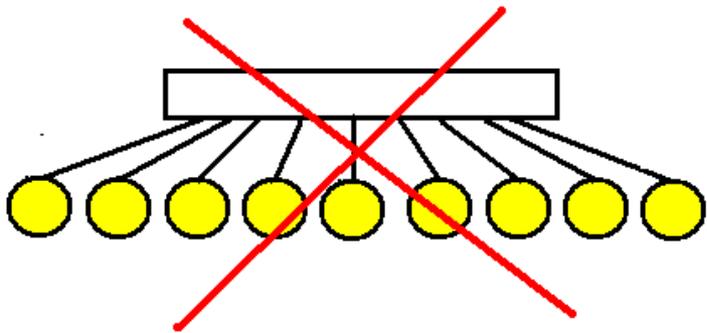
Site	Computer	Cores	HPL Rmax (Pflops)	HPL Rank	HPCG (Pflops)	HPCG/HPL
NSCC / Guangzhou	Tianhe-2 NUDT, Xeon 12C 2.2GHz + Intel Xeon Phi 57C + Custom	3,120,000	33.9	1	.580	1.7%
RIKEN Advanced Inst for Comp Sci	K computer Fujitsu SPARC64 VIIIfx 8C + Custom	705,024	10.5	4	.427	4.1%
DOE/OS Oak Ridge Nat Lab	Titan, Cray XK7 AMD 16C + Nvidia Kepler GPU 14C + Custom	560,640	17.6	2	.322	1.8%
DOE/OS Argonne Nat Lab	Mira BlueGene/Q, Power BQC 16C 1.60GHz + Custom	786,432	8.59	5	.101*	1.2%
Swiss CSCS	Piz Daint, Cray XC30, Xeon 8C + Nvidia Kepler 14C + Custom	115,984	6.27	6	.099	1.6%
Leibniz Rechenzentrum	SuperMUC, Intel 8C + IB	147,456	2.90	12	.0833	2.9%
CEA/TGCC-GENCI	Curie tine nodes Bullx B510 Intel Xeon 8C 2.7 GHz + IB	79,504	1.36	26	.0491	3.6%
Exploration and Production Eni S.p.A.	HPC2, Intel Xeon 10C 2.8 GHz + Nvidia Kepler 14C + IB	62,640	3.00	11	.0489	1.6%
DOE/OS L Berkeley Nat Lab	Edison Cray XC30, Intel Xeon 12C 2.4GHz + Custom	132,840	1.65	18	.0439 #	2.7%
Texas Advanced Computing Center	Stampede, Dell Intel (8c) + Intel Xeon Phi (61c) + IB	78,848	.881*	7	.0161	1.8%
Meteo France	Beaufix Bullx B710 Intel Xeon 12C 2.7 GHz + IB	24,192	.469 (.467*)	79	.0110	2.4%
Meteo France	Prolix Bullx B710 Intel Xeon 2.7 GHz 12C + IB	23,760	.464 (.415*)	80	.00998	2.4%
U of Toulouse	CALMIP Bullx DLC Intel Xeon 10C 2.8 GHz + IB	12,240	.255	184	.00725	2.8%
Cambridge U	Wilkes, Intel Xeon 6C 2.6 GHz + Nvidia Kepler 14C + IB	3584	.240	201	.00385	1.6%
TiTech	TUSBAME-KFC Intel Xeon 6C 2.1 GHz + IB	2720	.150	436	.00370	2.5%

Классификация суперкомпьютеров



Петафлопс

Общая тенденция – многосвязность и иерархичность



**Нужна эффективная
глобально адресуемая
память огромного объема !**

Стойка суперкомпьютера Power 775

Data Center In a Rack

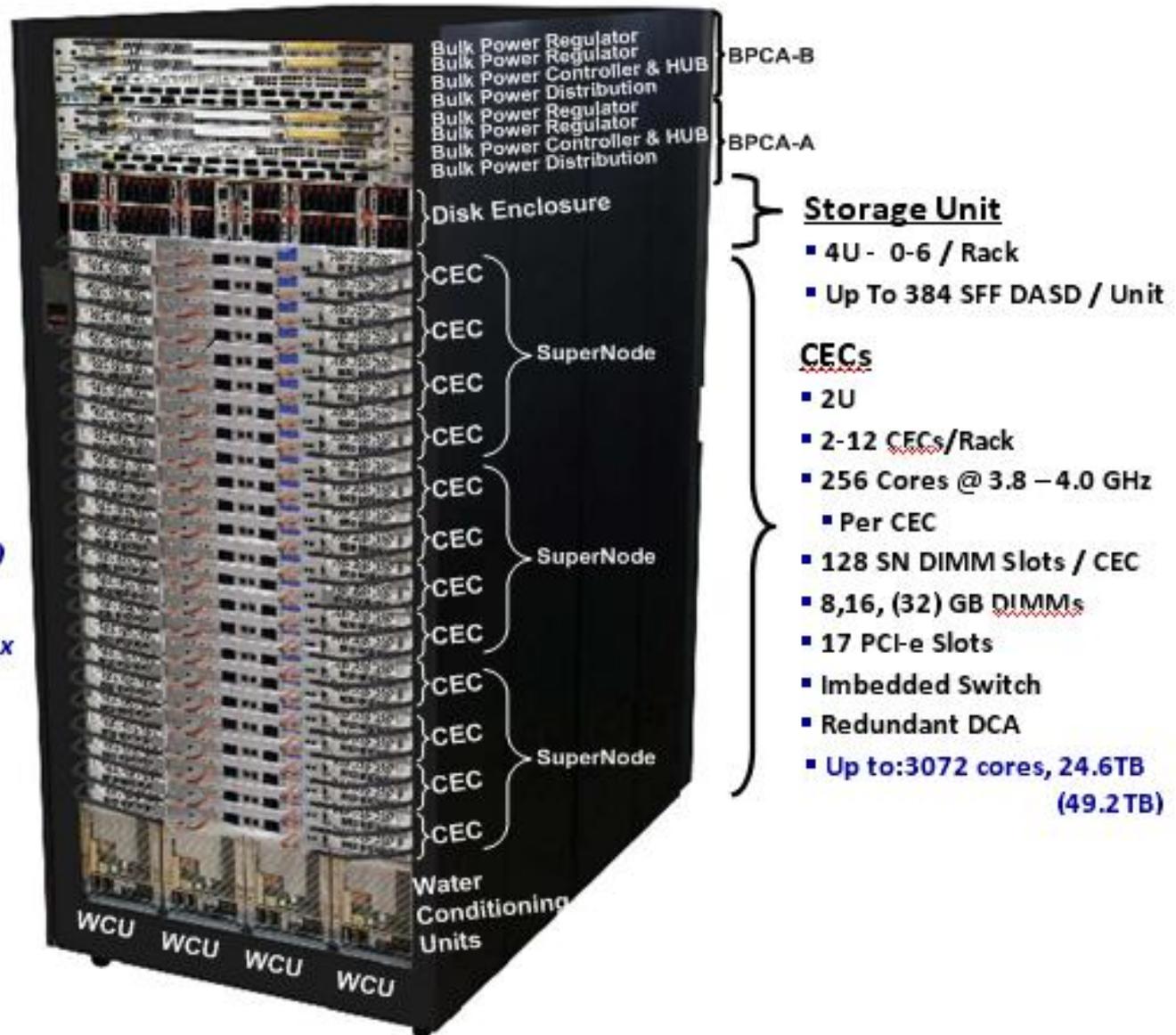
Compute
Storage
Switch
100% Cooling

Input:

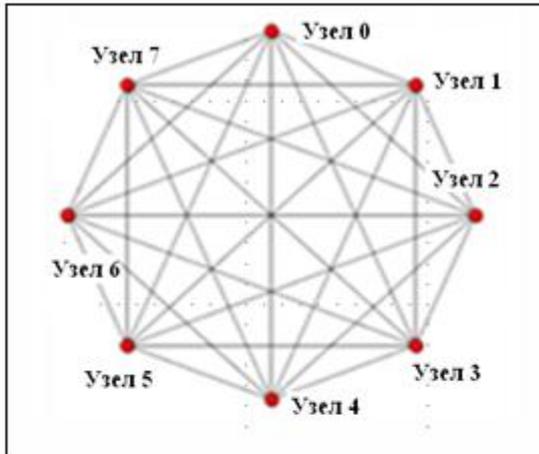
- 8 Water Lines
- 4 Power Cords

Out:

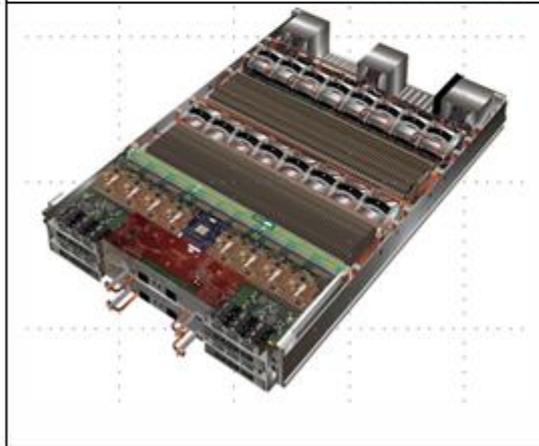
- ~100TFLOPs / 24.6TB (492)
- 153.5TB Storage
- 192 PCIe 16x / 12 PCIe 8x



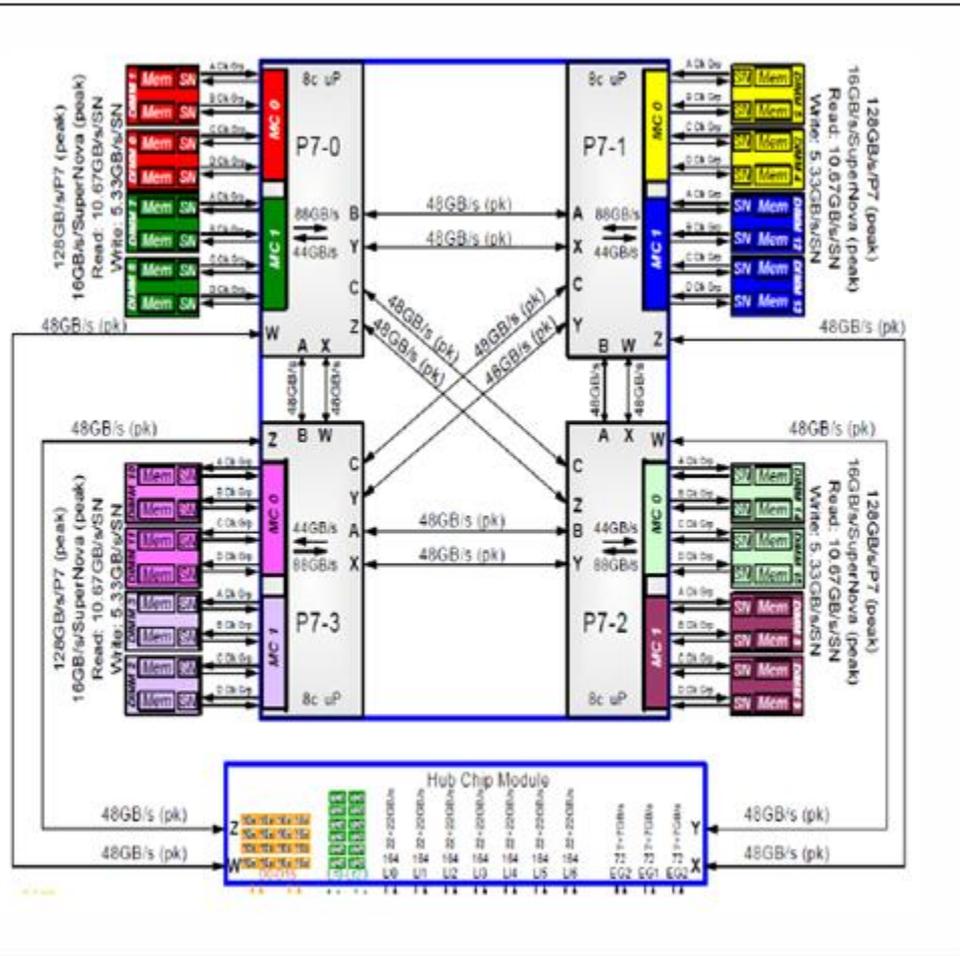
Узел и макроузел IBM Power 775



Внутренняя сеть макроузла

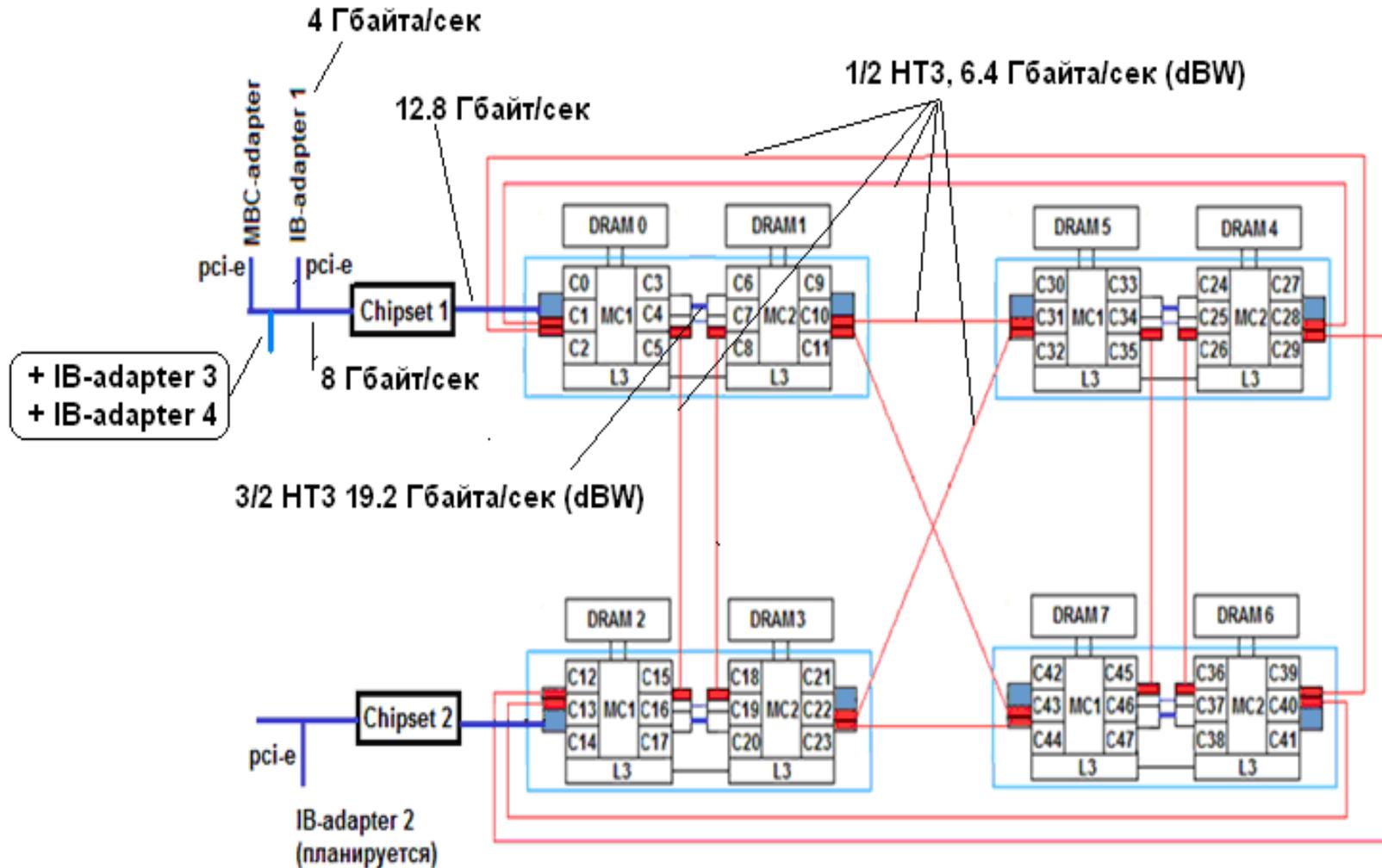


Внешний вид макроузла

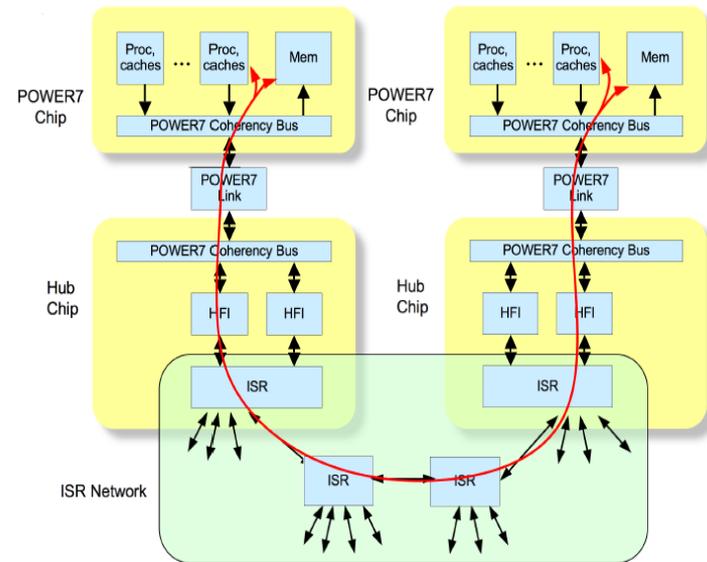
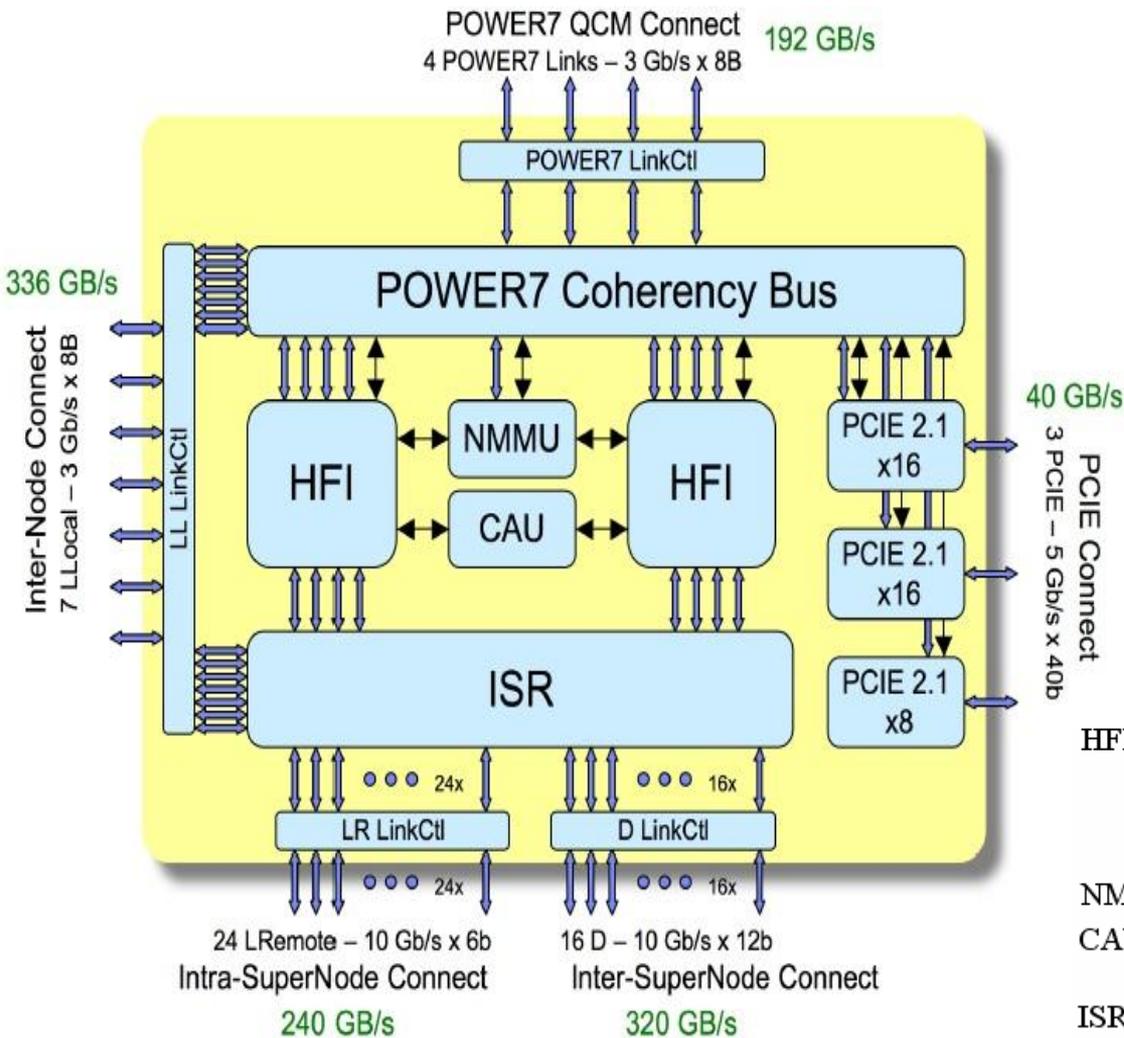


4-х процессорный узел (4 Power7 + HUB)

Серверная плата ПТК (СПБГПУ)



Многопортовый HUB-маршрутизатор сети PERCS



HFI – интерфейс процессоров с сетью, управление через “окна в виртуальной памяти”, поддержка доменов защиты, пакетов MPI, GAS (глобальной памяти), IP (интернет);

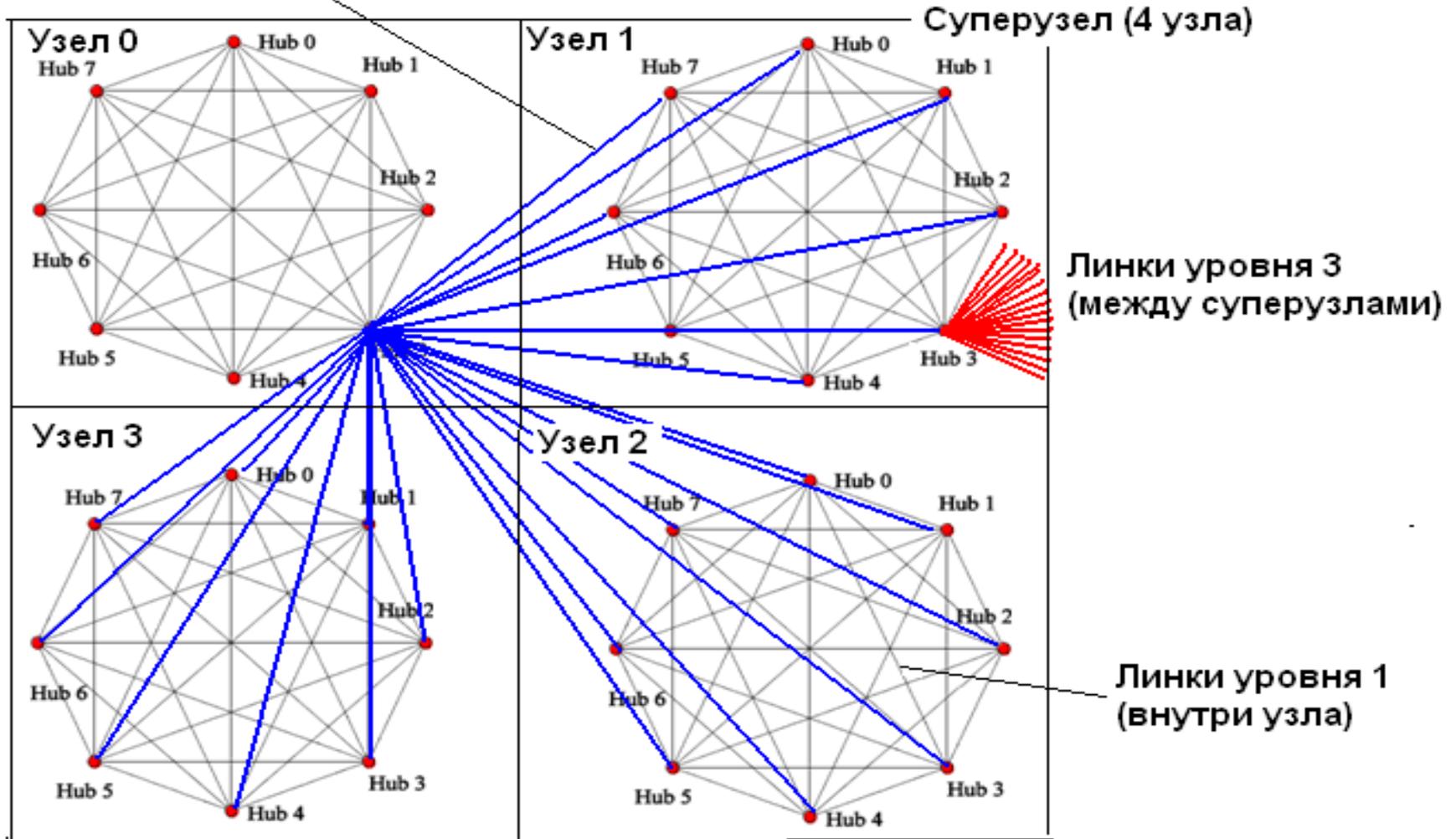
NMMU – блок трансляции виртуальных адресов;

CAU – блок ускорения выполнения коллективных операций;

ISR – интегрированный коммутатор-маршрутизатор (56x56 портов, таблицы маршрутизации, виртуальные каналы, защита от дедлоков, адаптивная маршрутизация, отказоустойчивость на уровне линков и сети)

Многоуровневая сеть PERCS суперкомпьютера Power 775

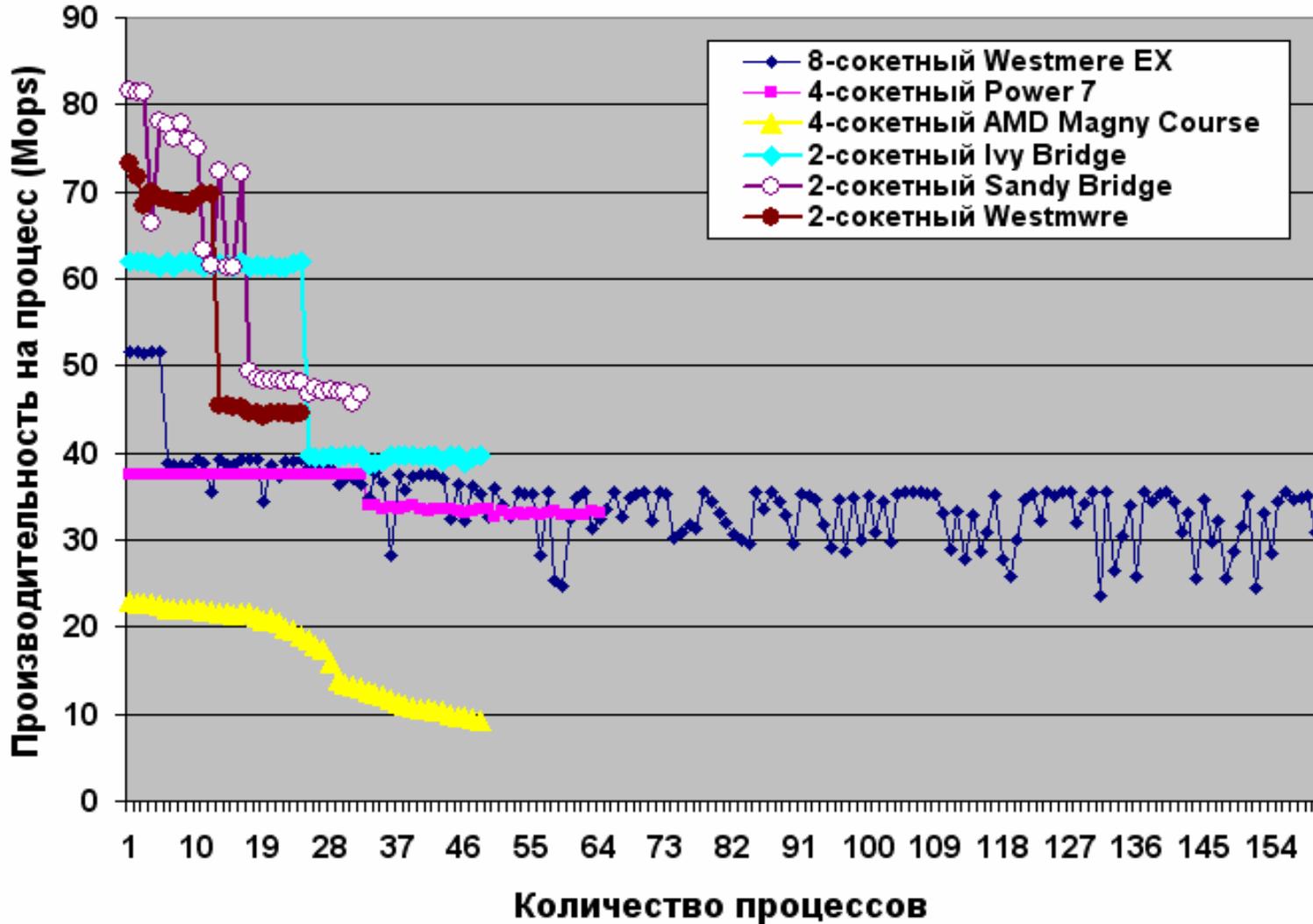
Линки уровня 2
(внутри суперузла)



Один QCM

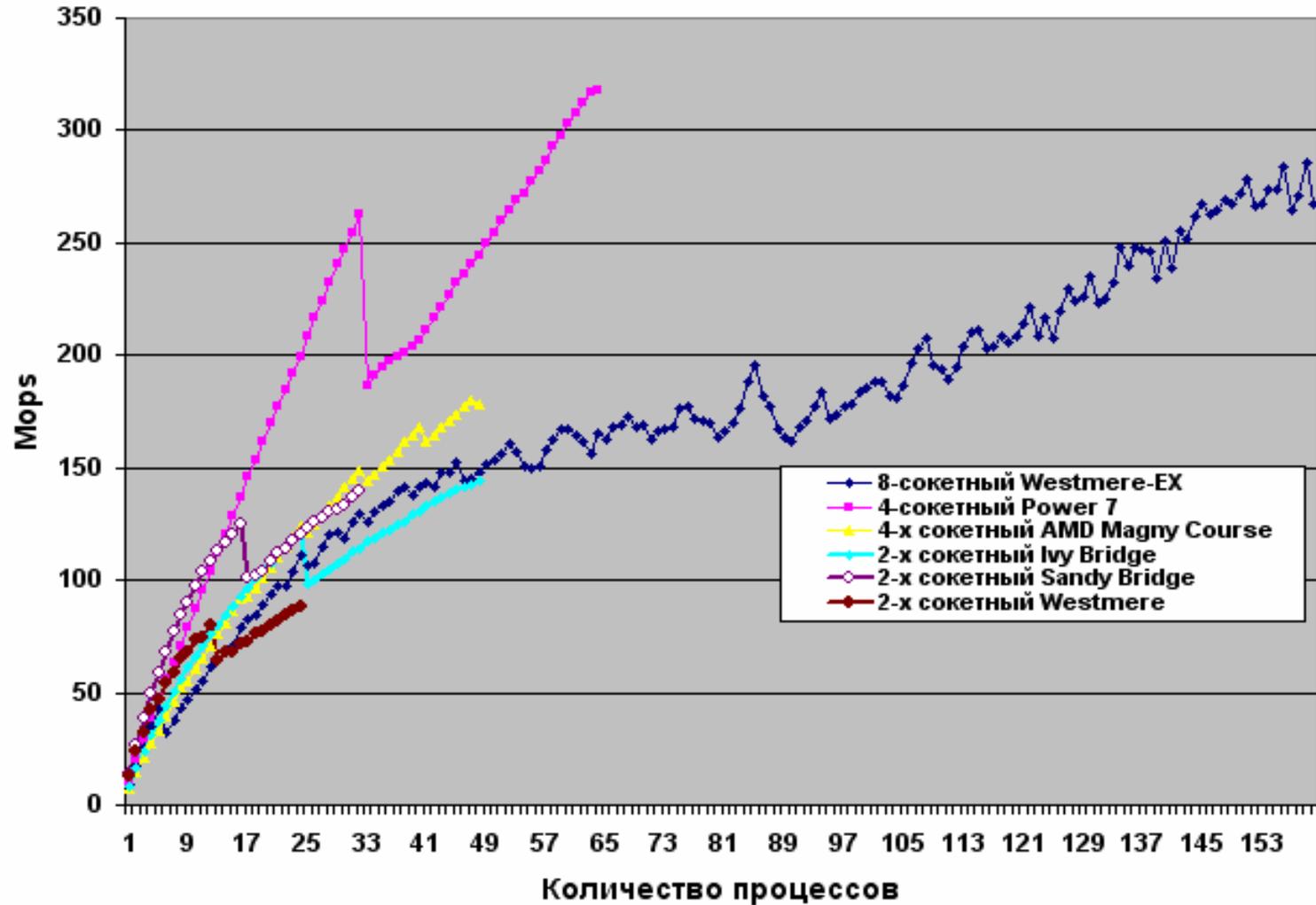
IBM Power 775 - тест EP (класс C)

Тест EP, класс C, OpenMP,
производительность на один процесс



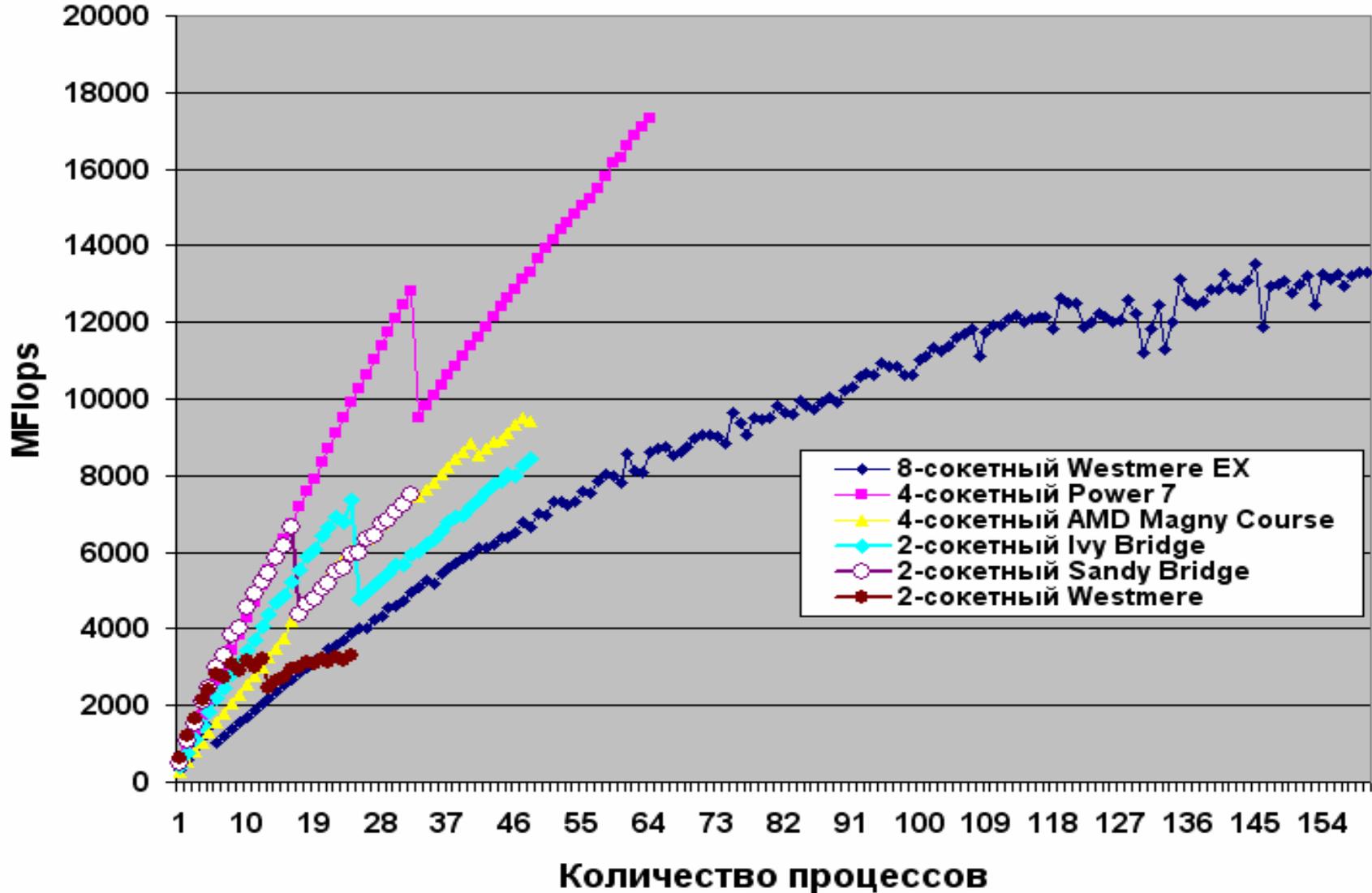
IBM Power 775 - тест UA (класс C)

Тест UA, класс C, OpenMP



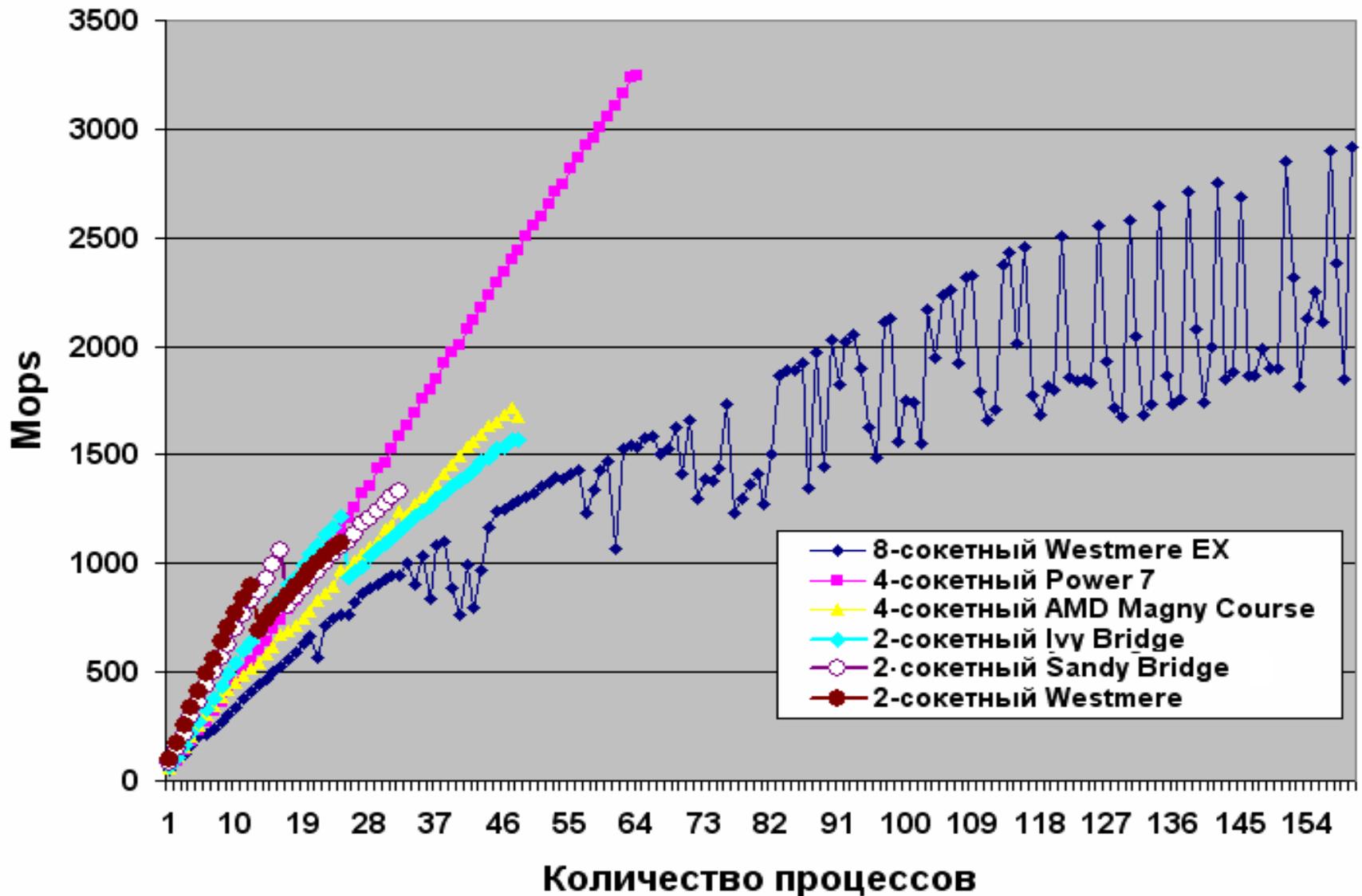
IBM Power 775 - тест CG (класс C)

Тест CG, класс C, OpenMP



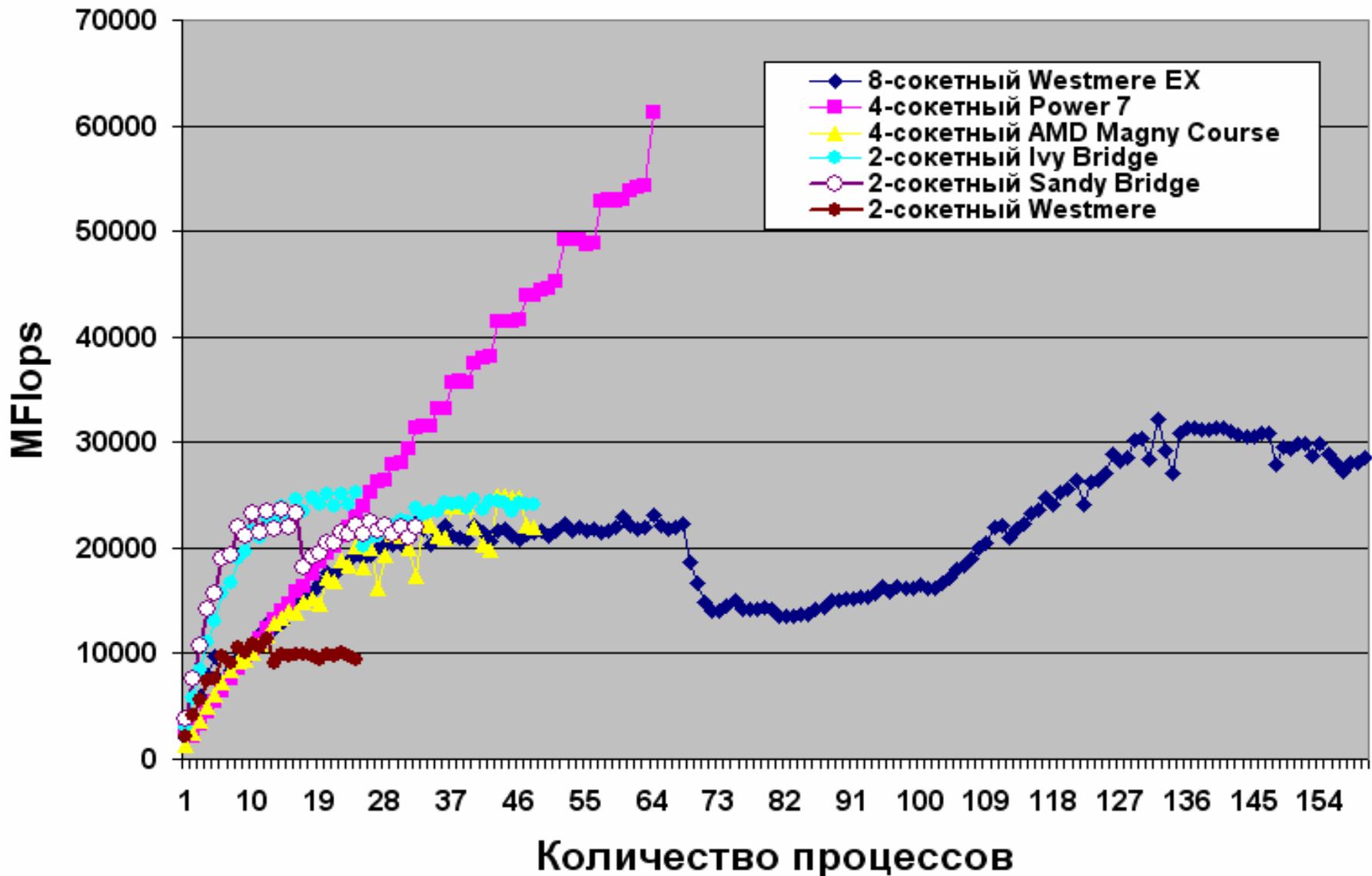
IBM Power 775 - тест IS (класс C)

Тест IS, класс C, OpenMP



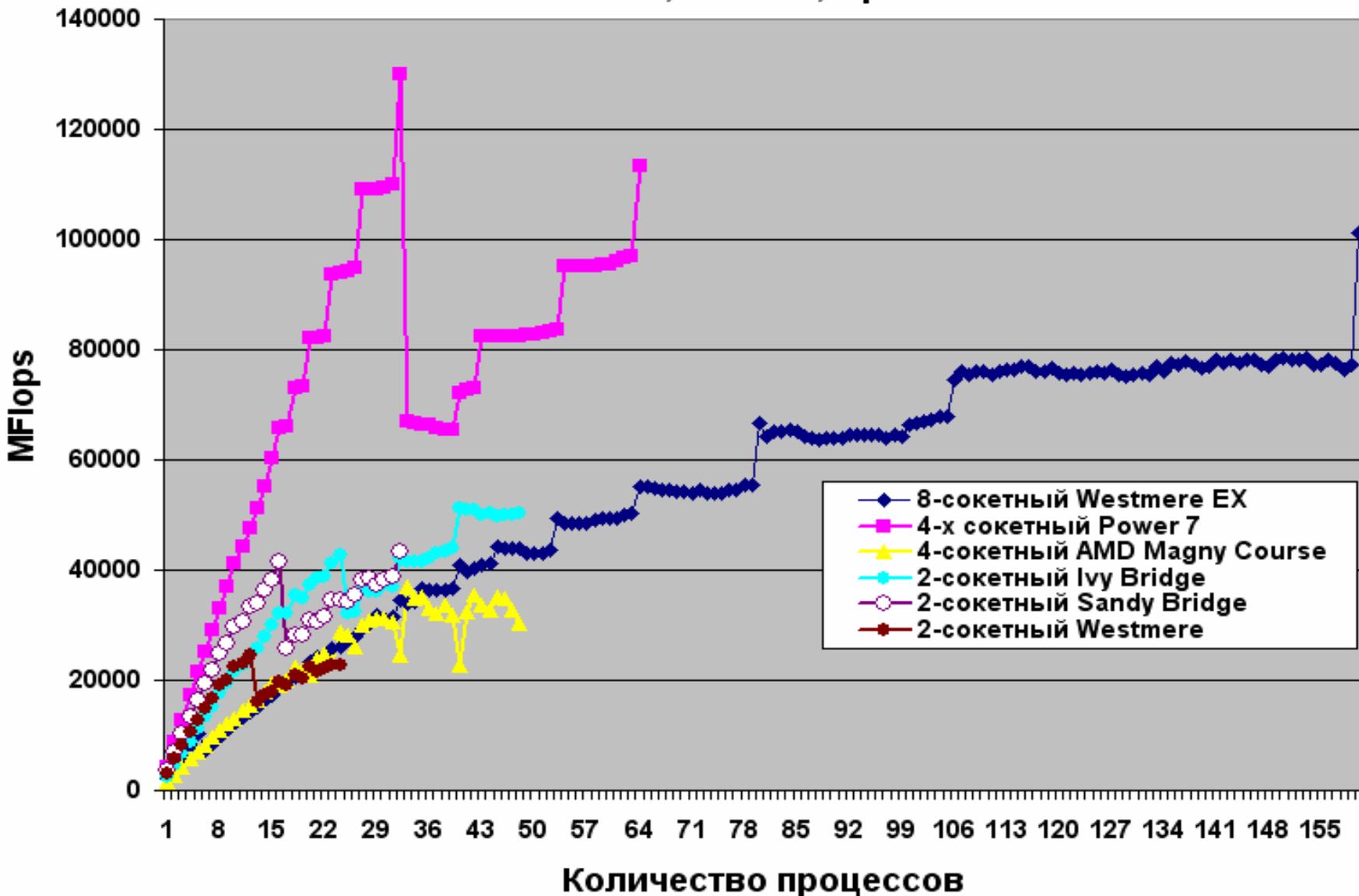
IBM Power 775 - тест MG (класс C)

Тест MG, класс C, OpenMP



IBM Power 775 - тест VT (класс C)

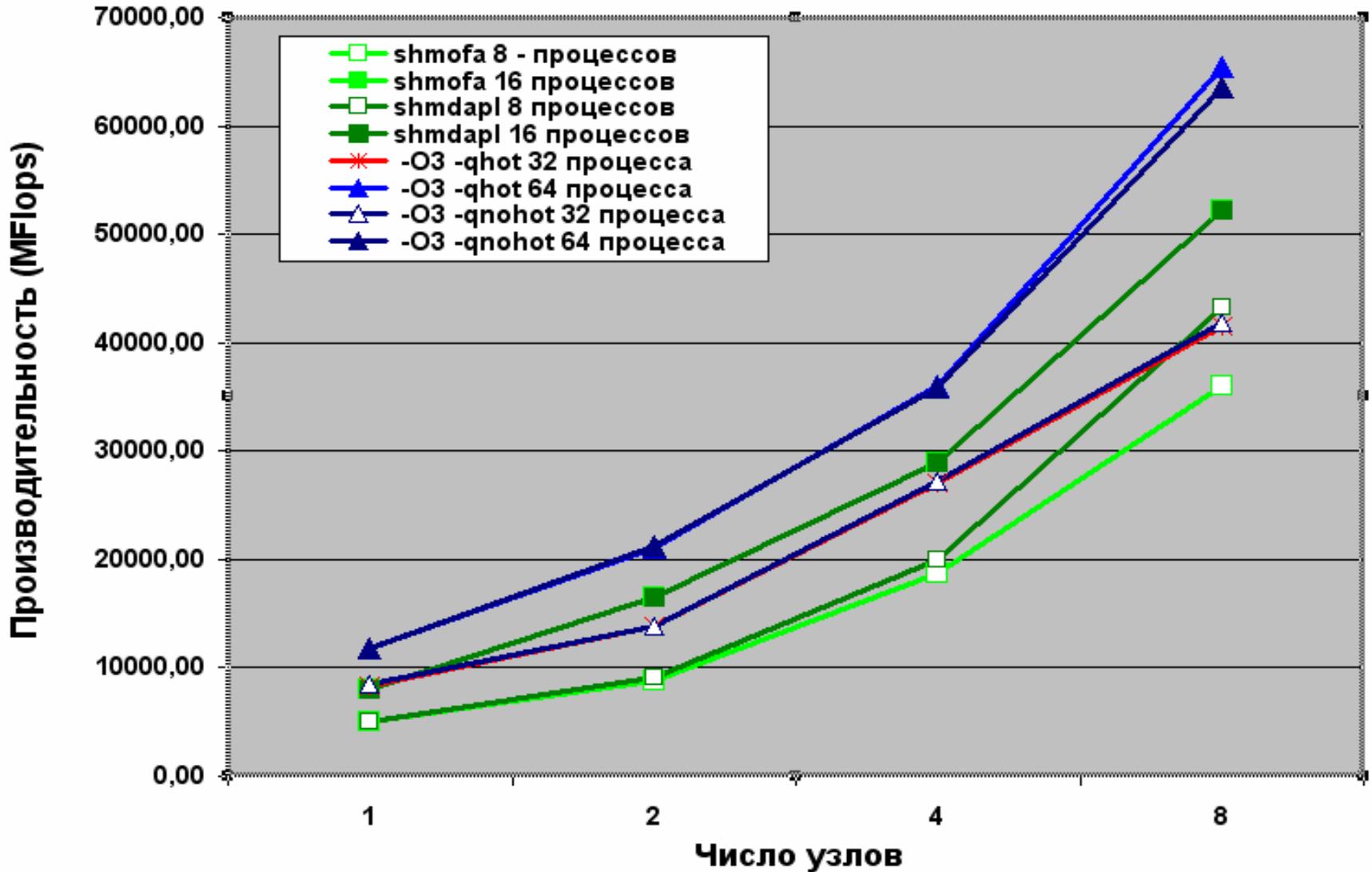
Тест VT, класс C, OpenMP



**Много QCM одной
серверной платы
(на ней до 8 QCM)**

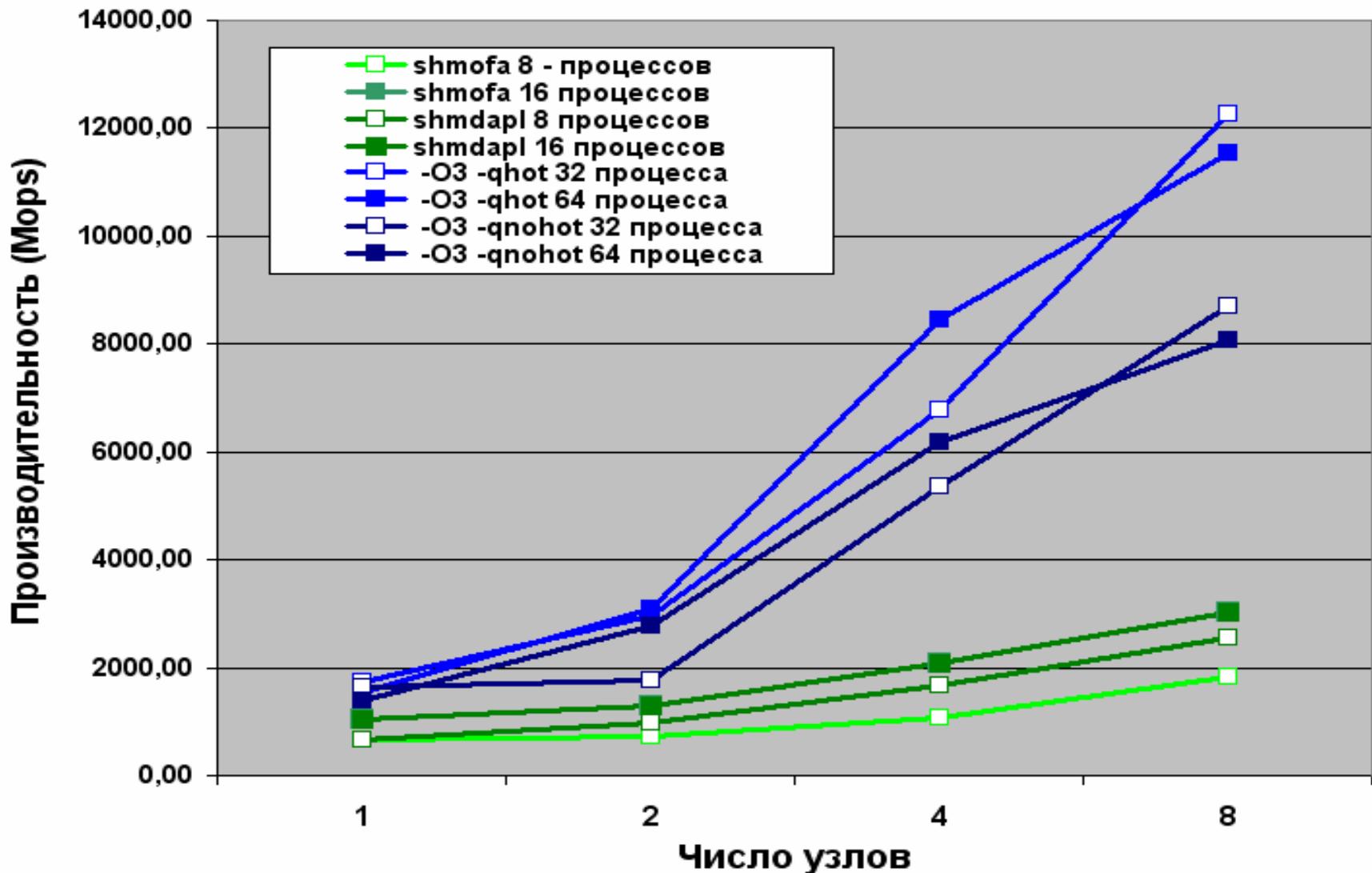
IBM Power 775 – тест CG (класс C)

Тест CG, класс C, MPI



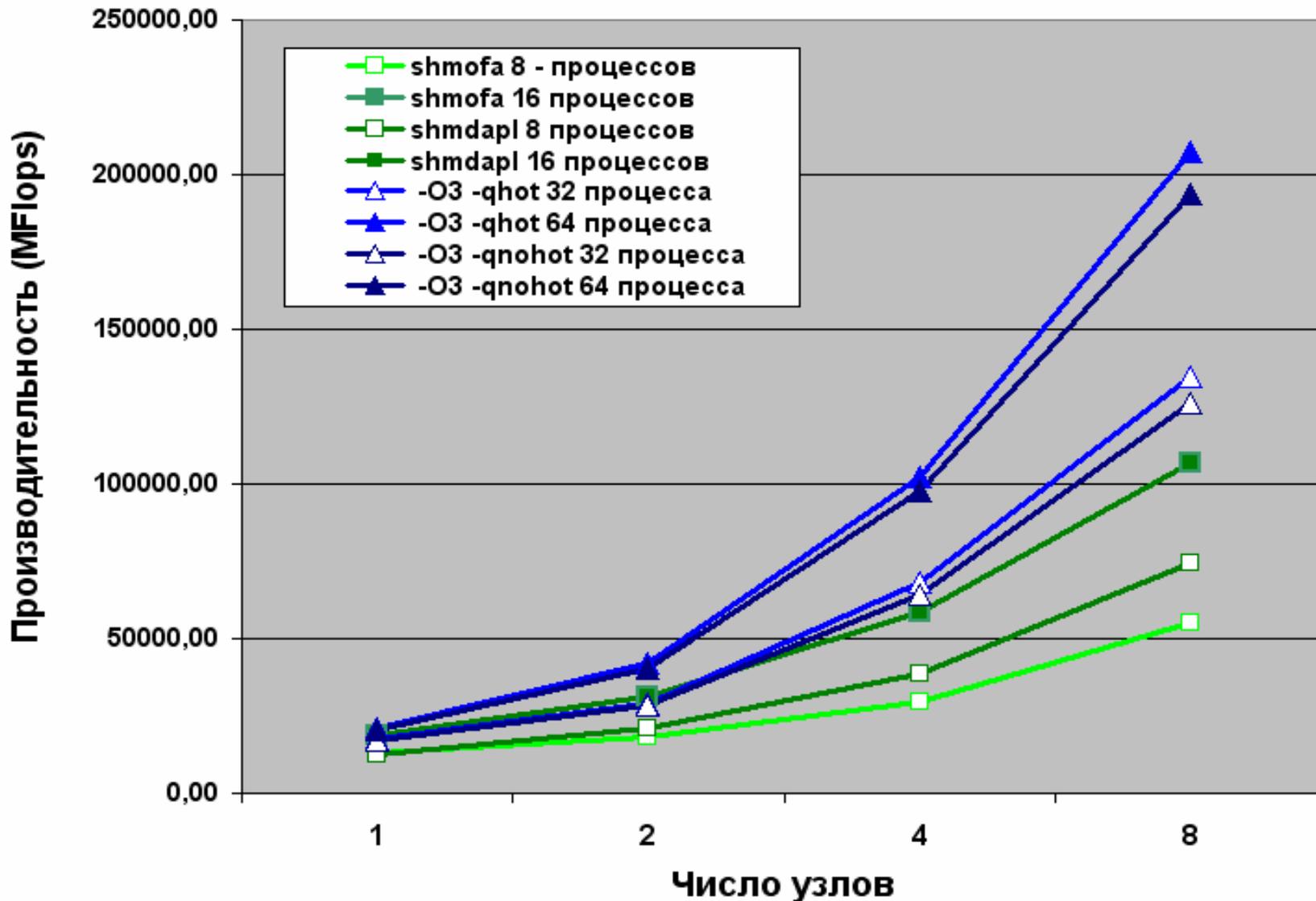
IBM Power 775 - тест IS (класс C)

Тест IS, класс C, MPI



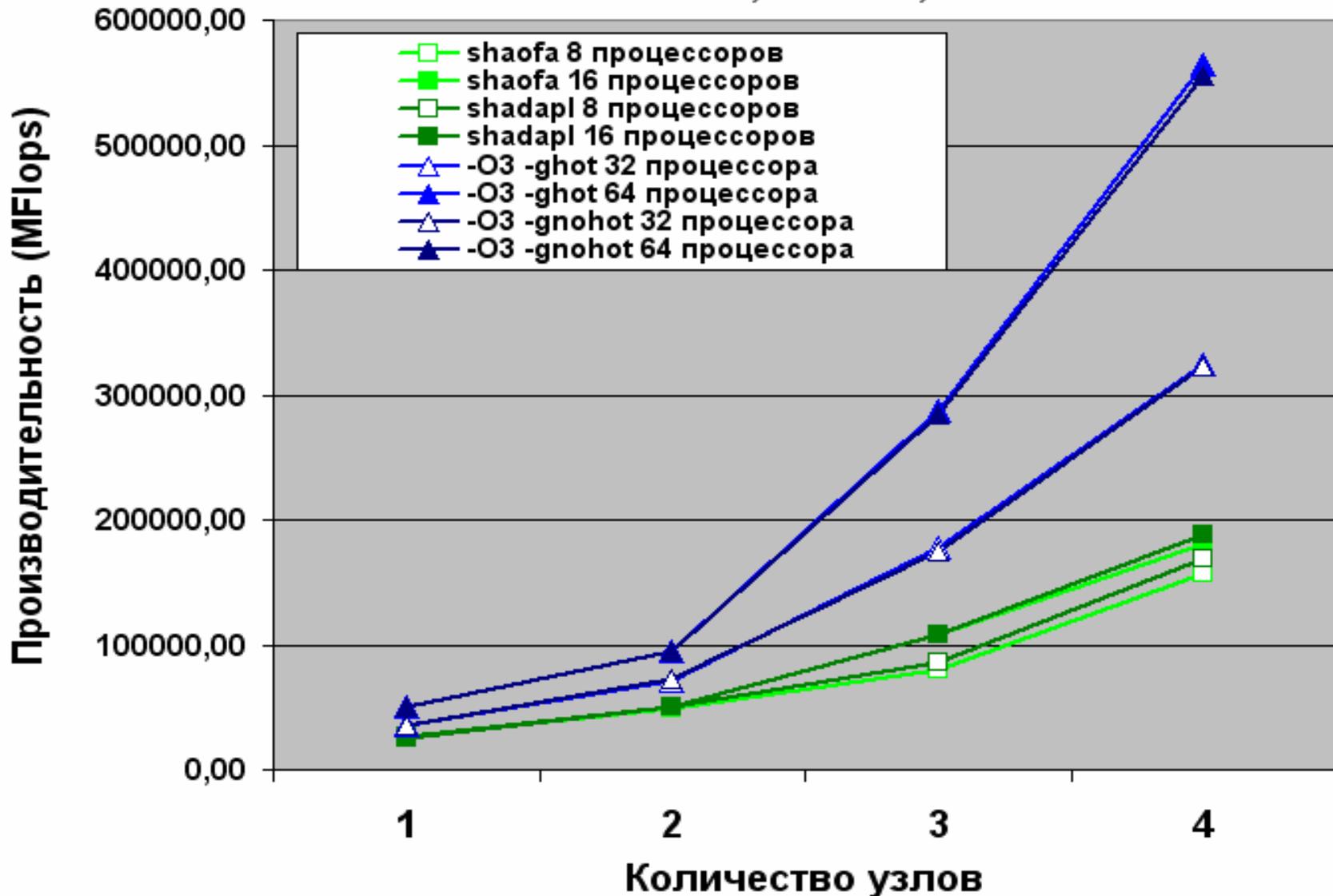
IBM Power 775 - тест FT (класс C)

Тест FT, класс C, MPI



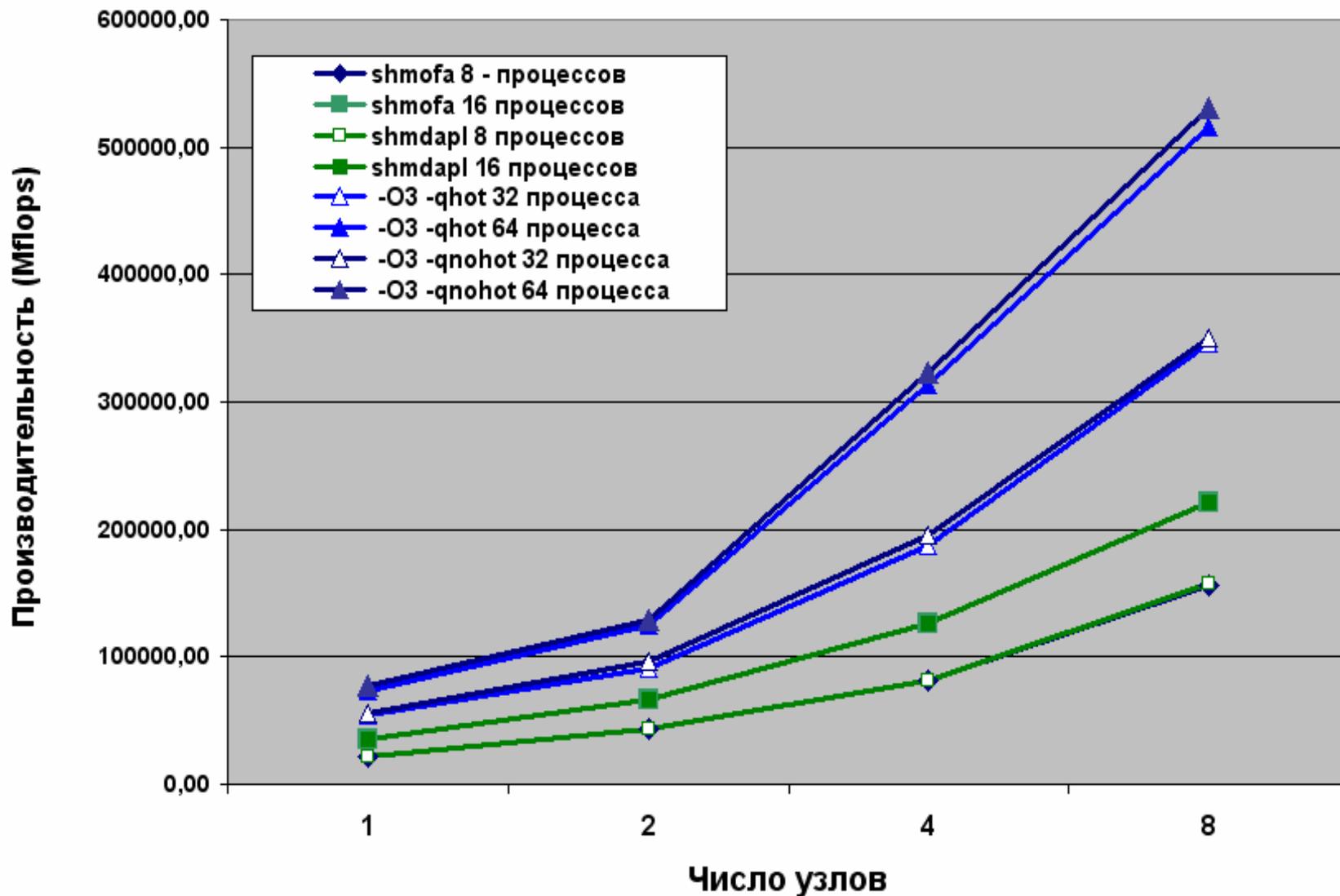
IBM Power 775 - тест FT (класс C)

Тест MG, класс C, MPI



IBM Power 775 - тест LU (класс C)

Тест LU, класс C, MPI

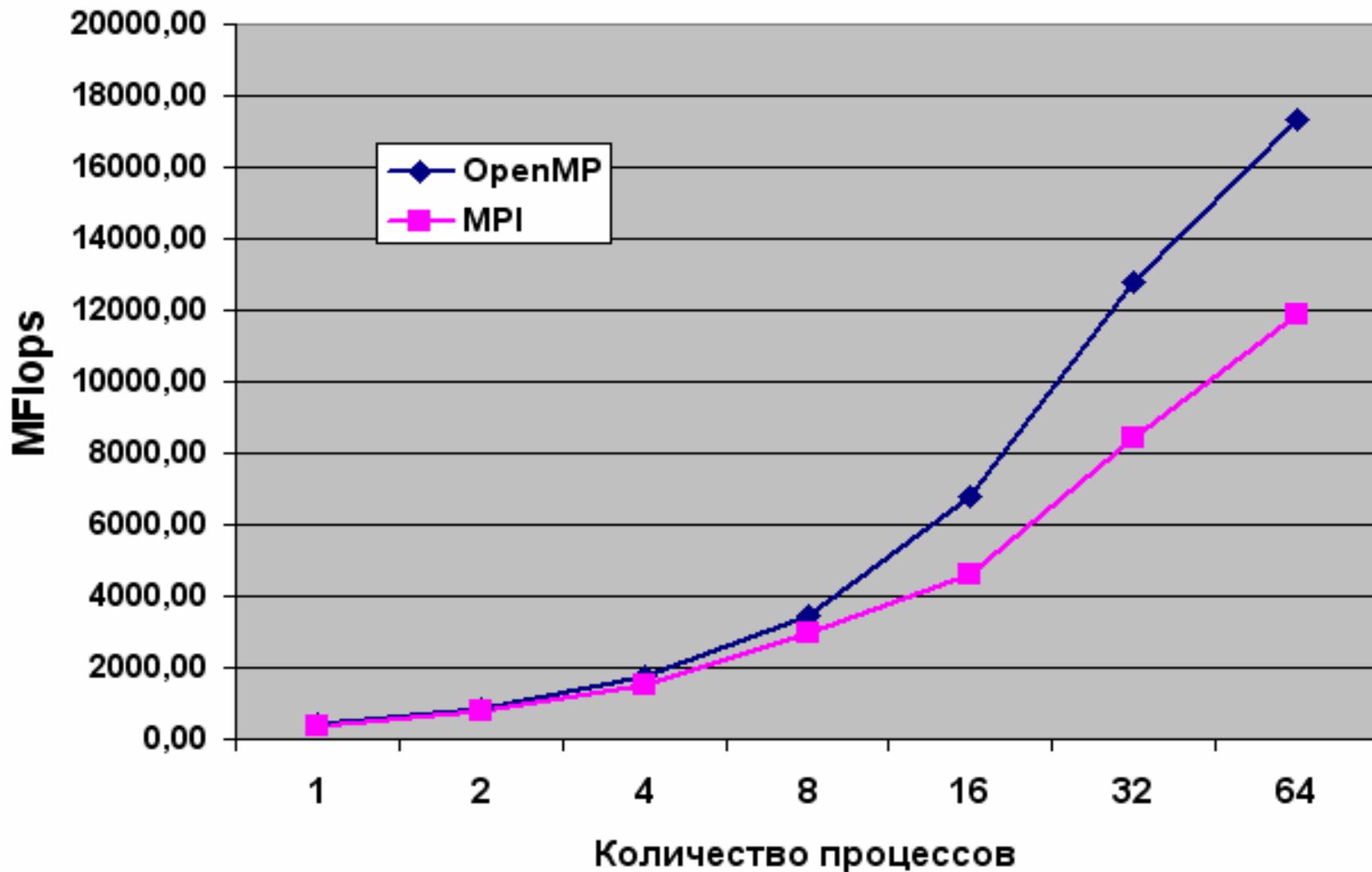


Один QCM

OpenMP vs MPI

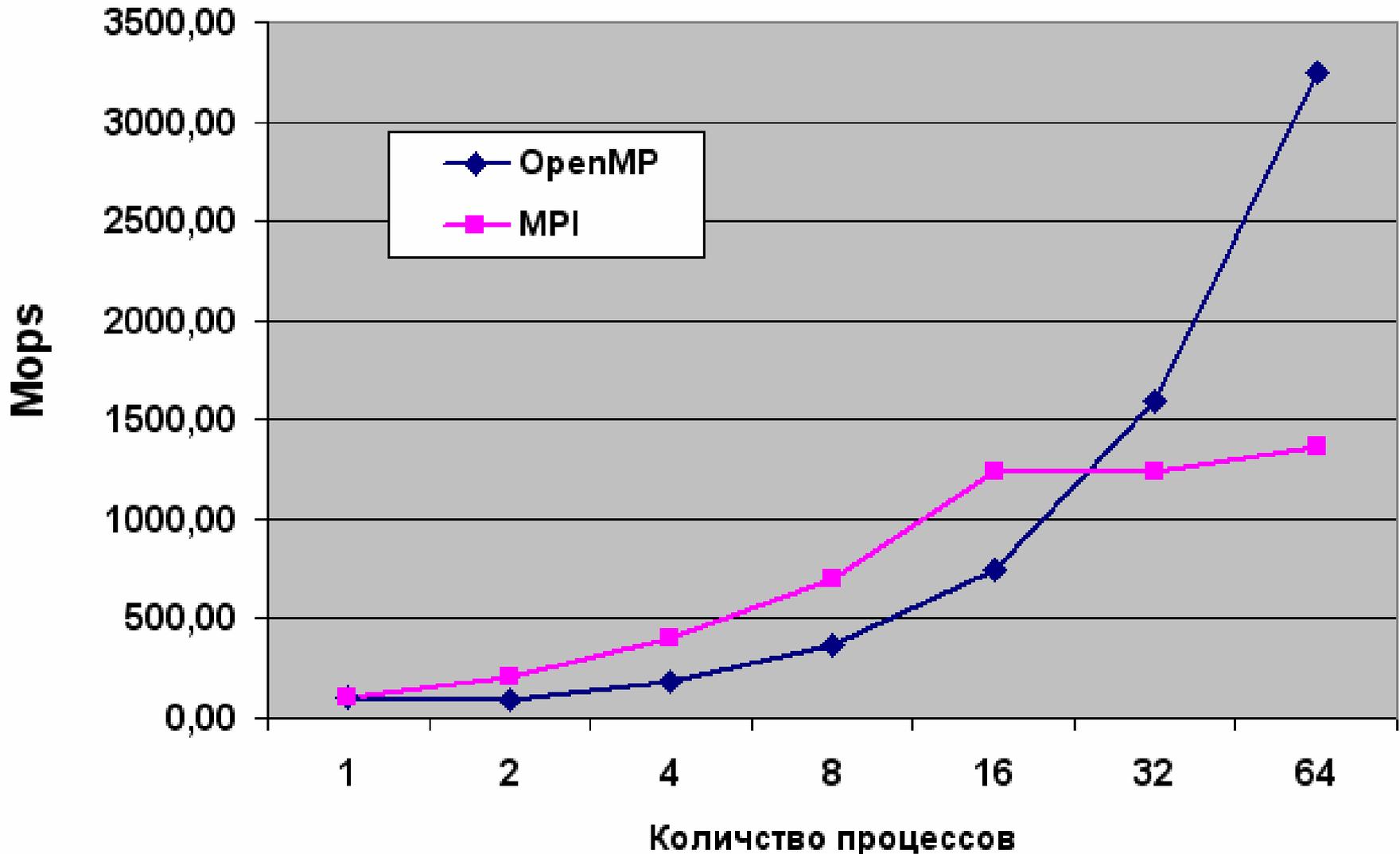
IBM Power 775 - тест CG (класс C)

Тест CG, класс C, OpenMP и MPI, один узел p775



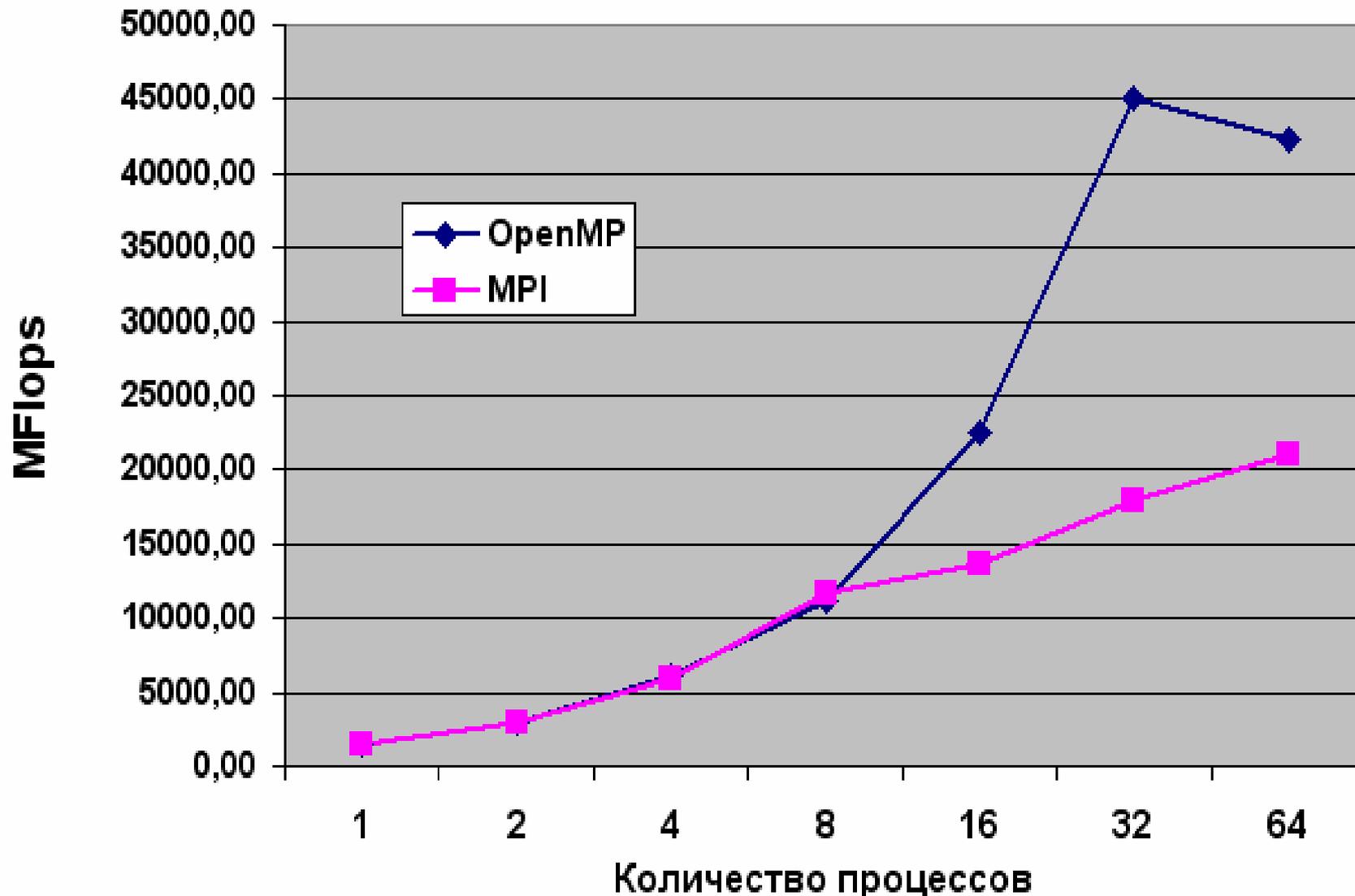
IBM Power 775 - тест IS (класс C)

Тест IS, класс C, OpenMP и MPI, один узел p775



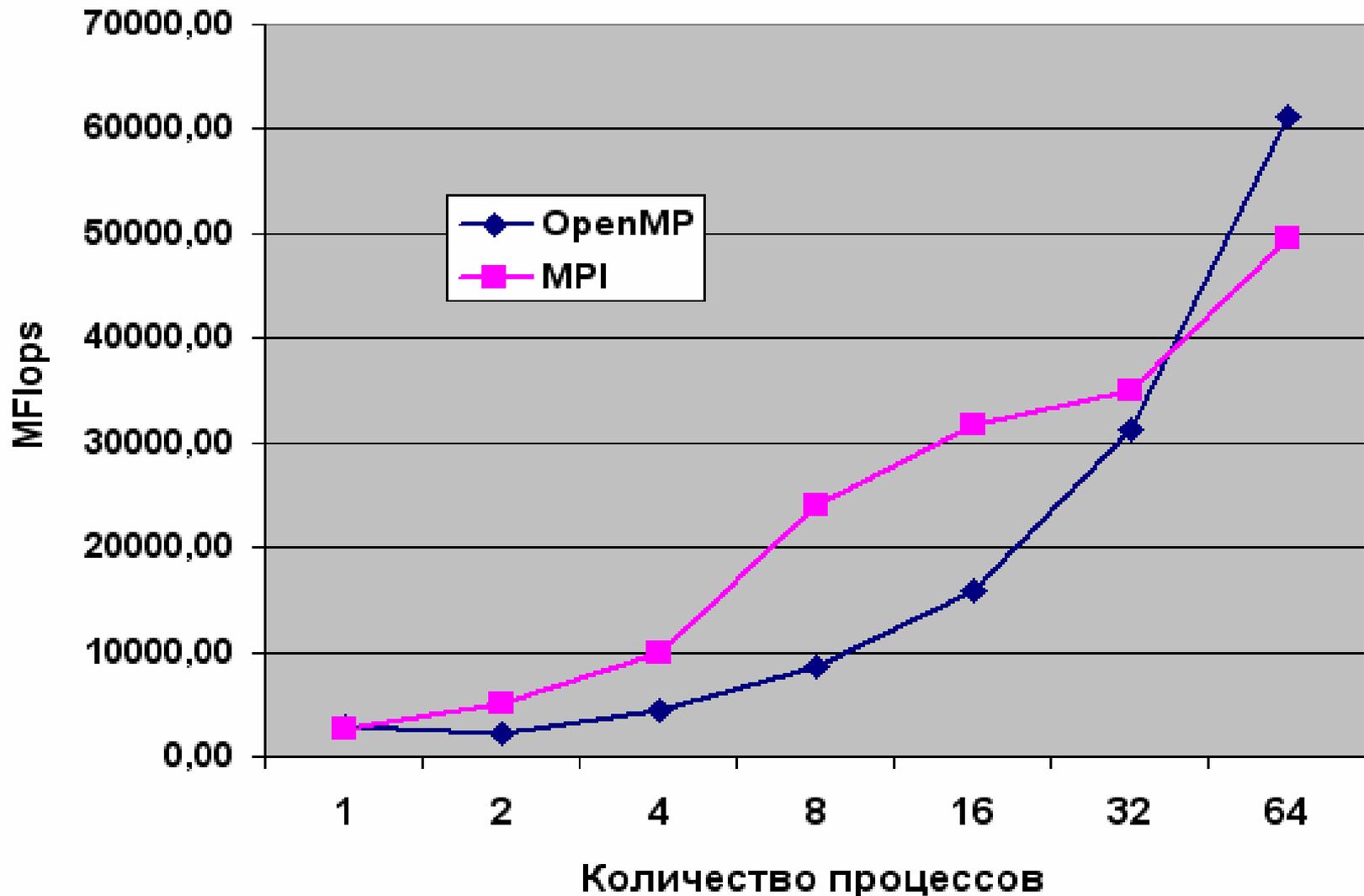
IBM Power 775 - тест FT (класс C)

Тест FT, класс C, OpenMP и MPI, один узел р775



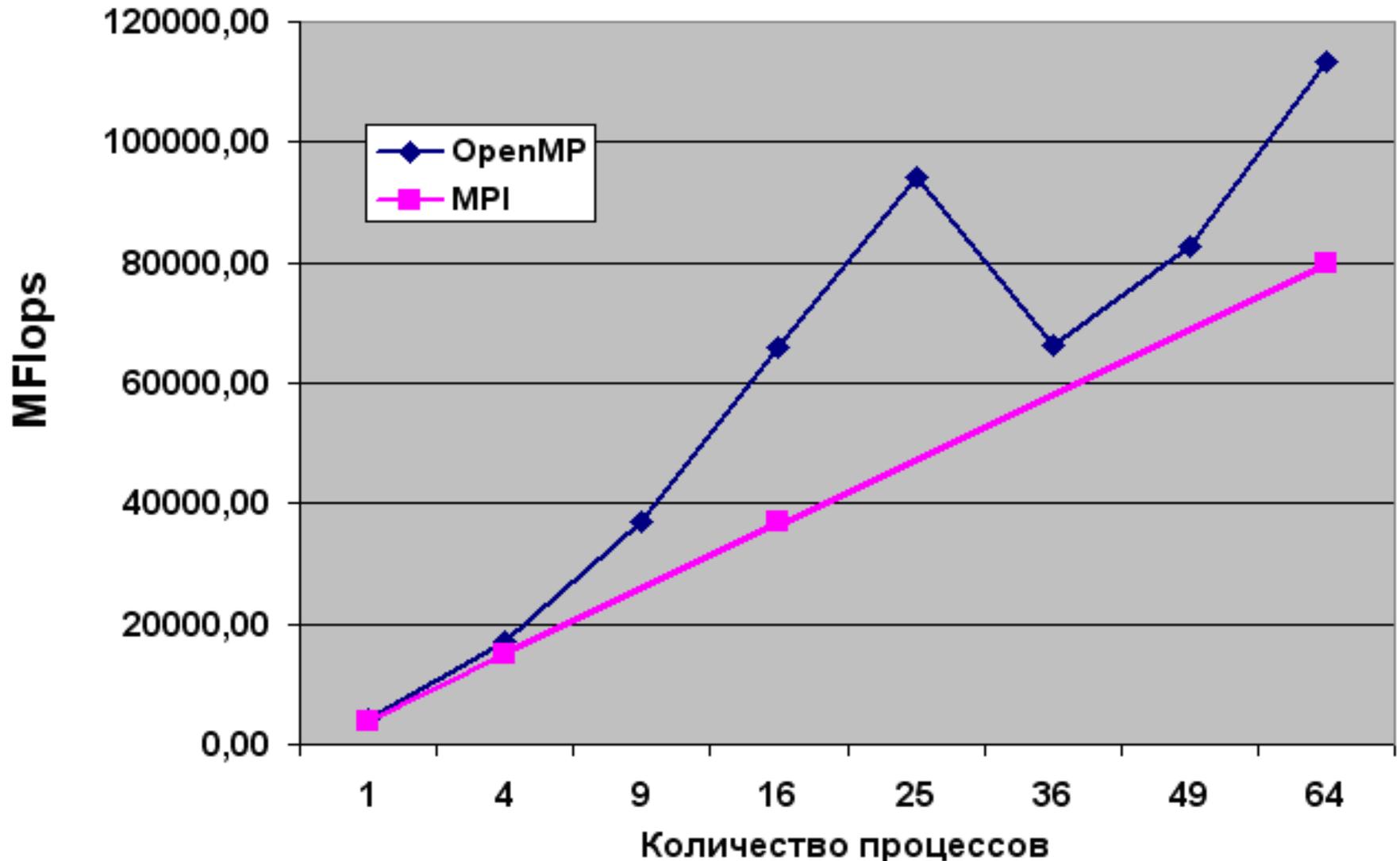
IBM Power 775 - тест MG (класс C)

Тест MG, класс C, OpenMP и MPI, один узел p775

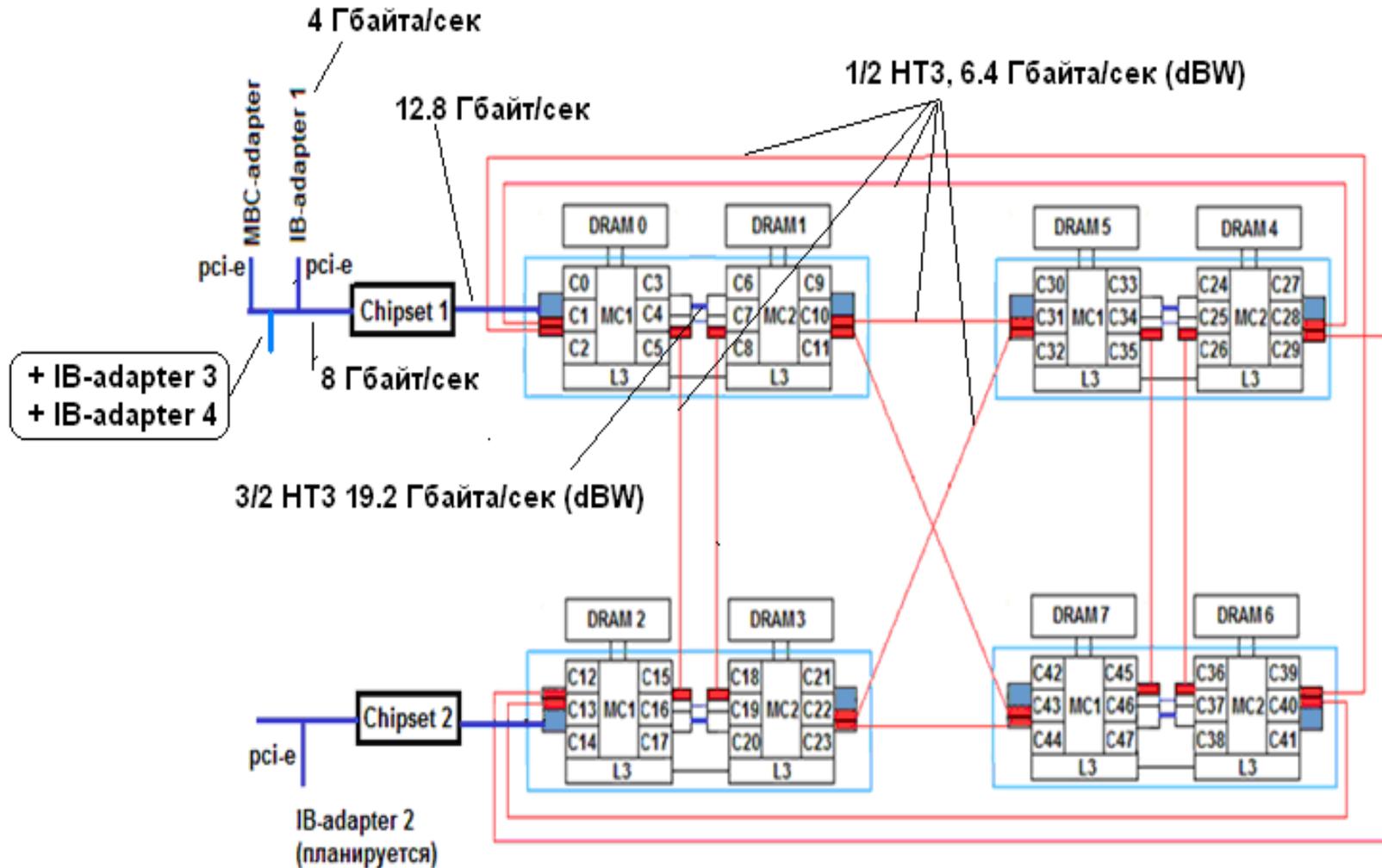


IBM Power 775 - тест VT (класс C)

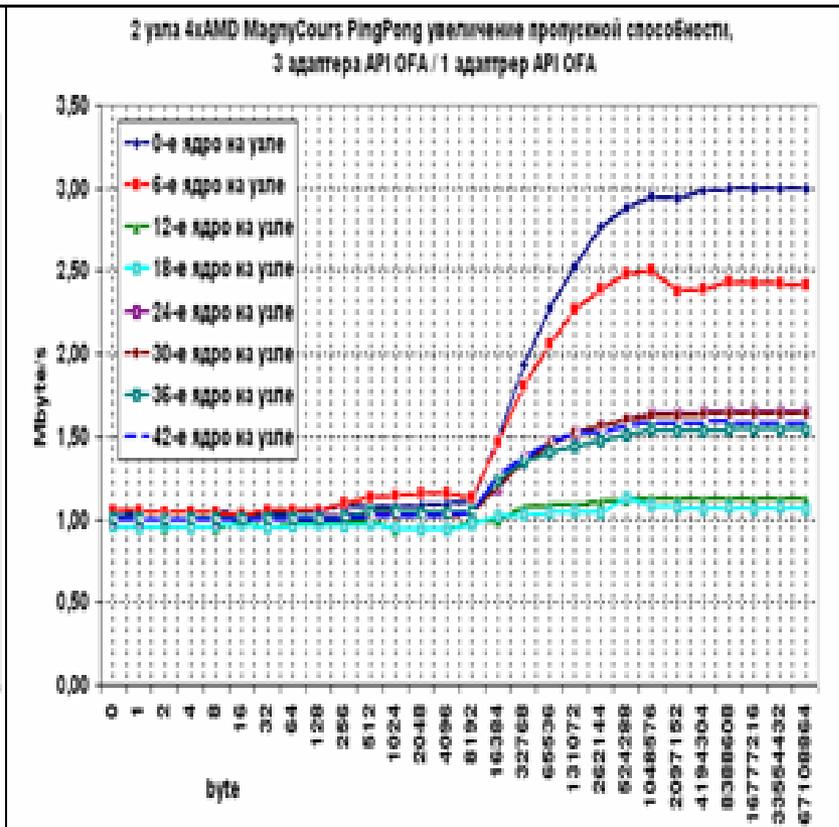
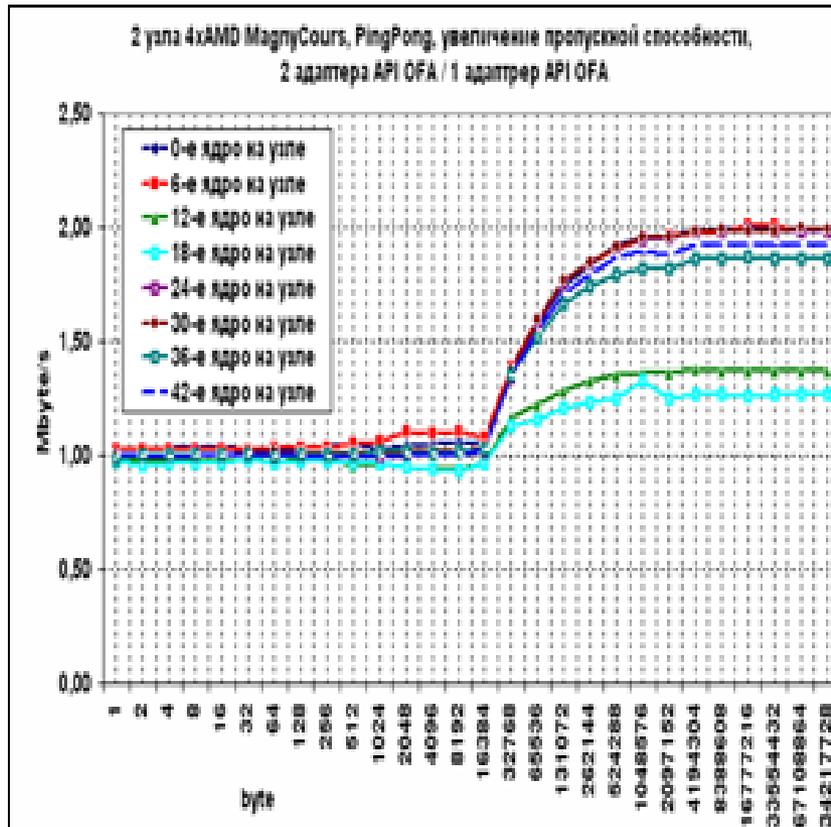
Тест VT, класс C, OpenMP и MPI, один узел p775



Серверная плата ПТК (СПБГПУ)



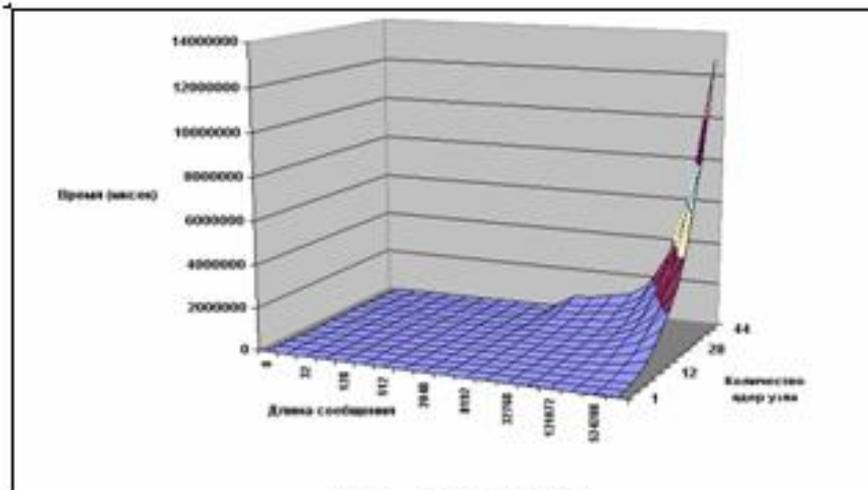
Ускорение операций “точка-точка” на 2-х и 3-х адаптерной установке



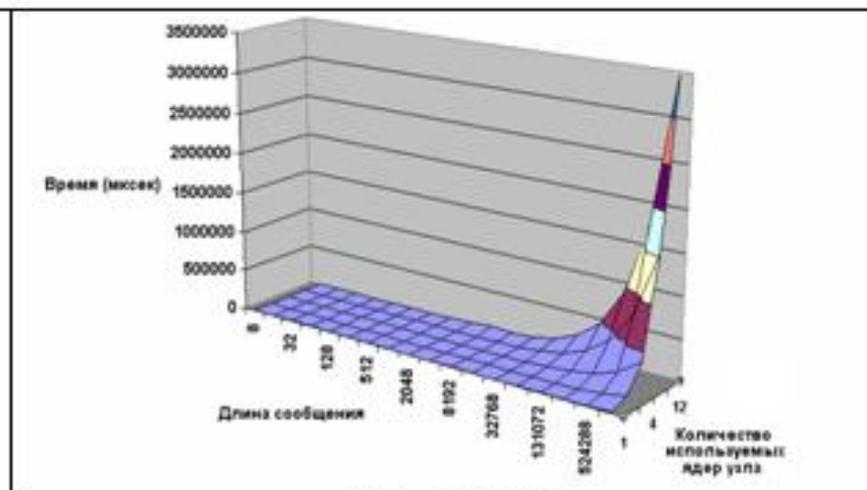
(A) Межузловые взаимодействия, Mellanox, PMPI+ofa, 1 и 2 адаптера на узле.

(B) Межузловые взаимодействия, Mellanox, PMPI+ofa, 1 и 3 адаптера на узле.

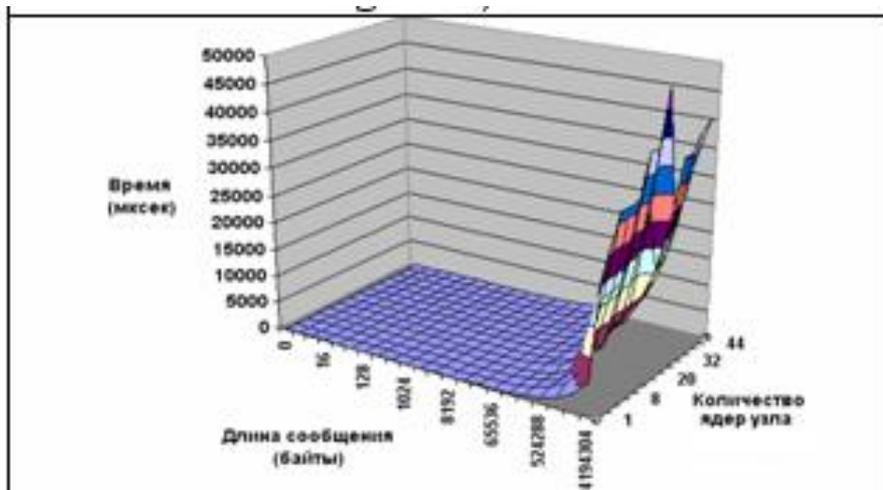
Времена коллективных операций AlltoAll и Allreduce (линейная шкала)



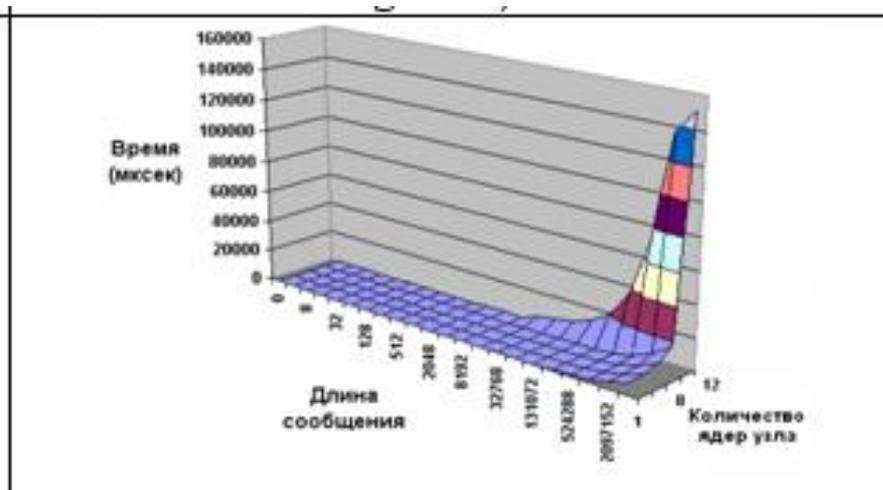
Alltoall, ПТК



Alltoall, БК

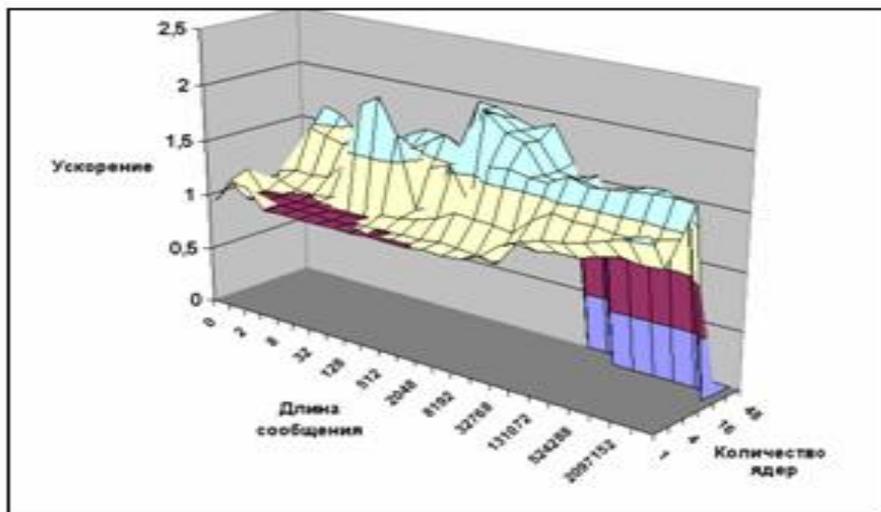


Allreduce, ПТК

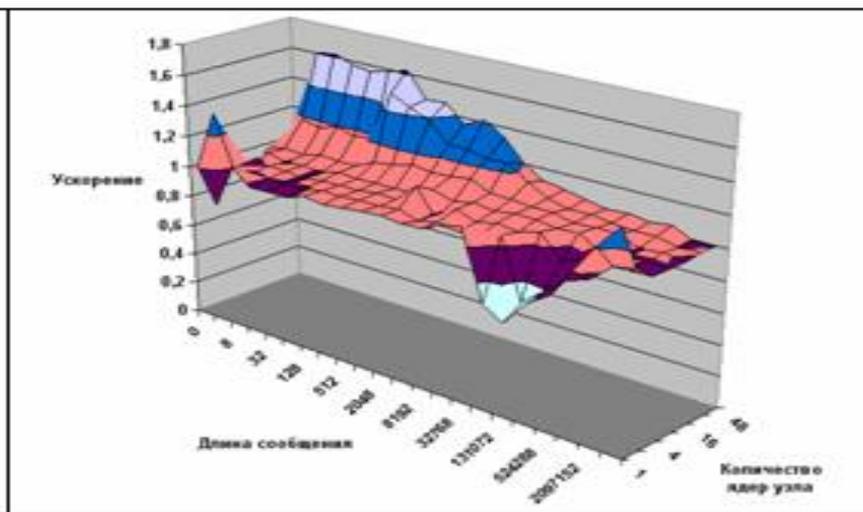


Allreduce, БК

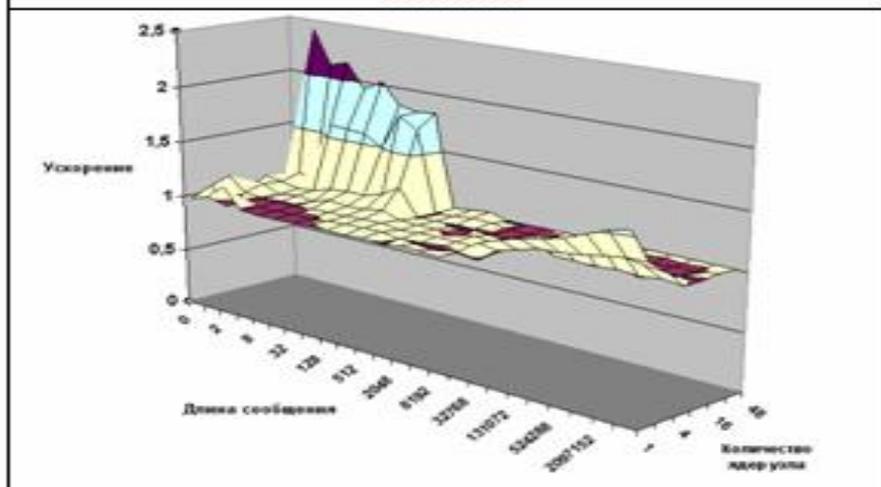
Ускорение коллективных операций на 2-х адаптерной установке



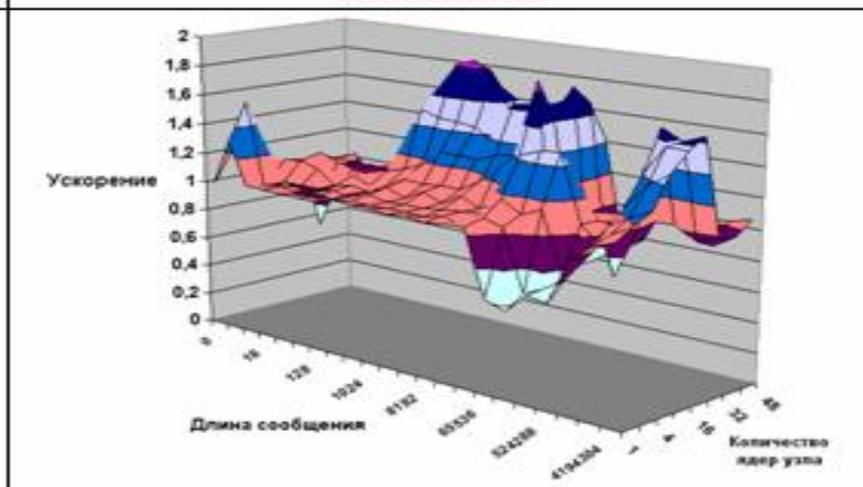
Alltoall



Allreduce

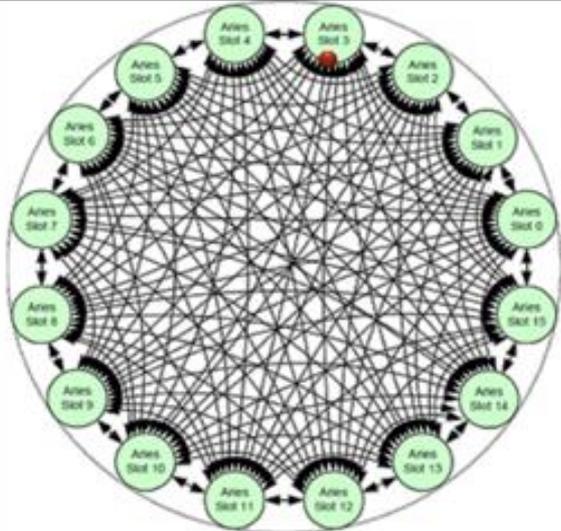


Allgather

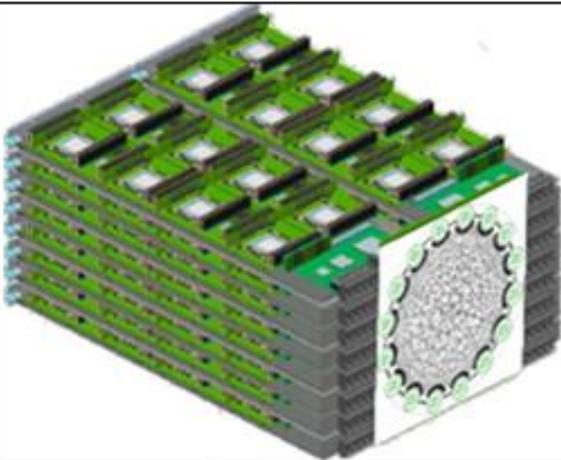


ReduceScatter

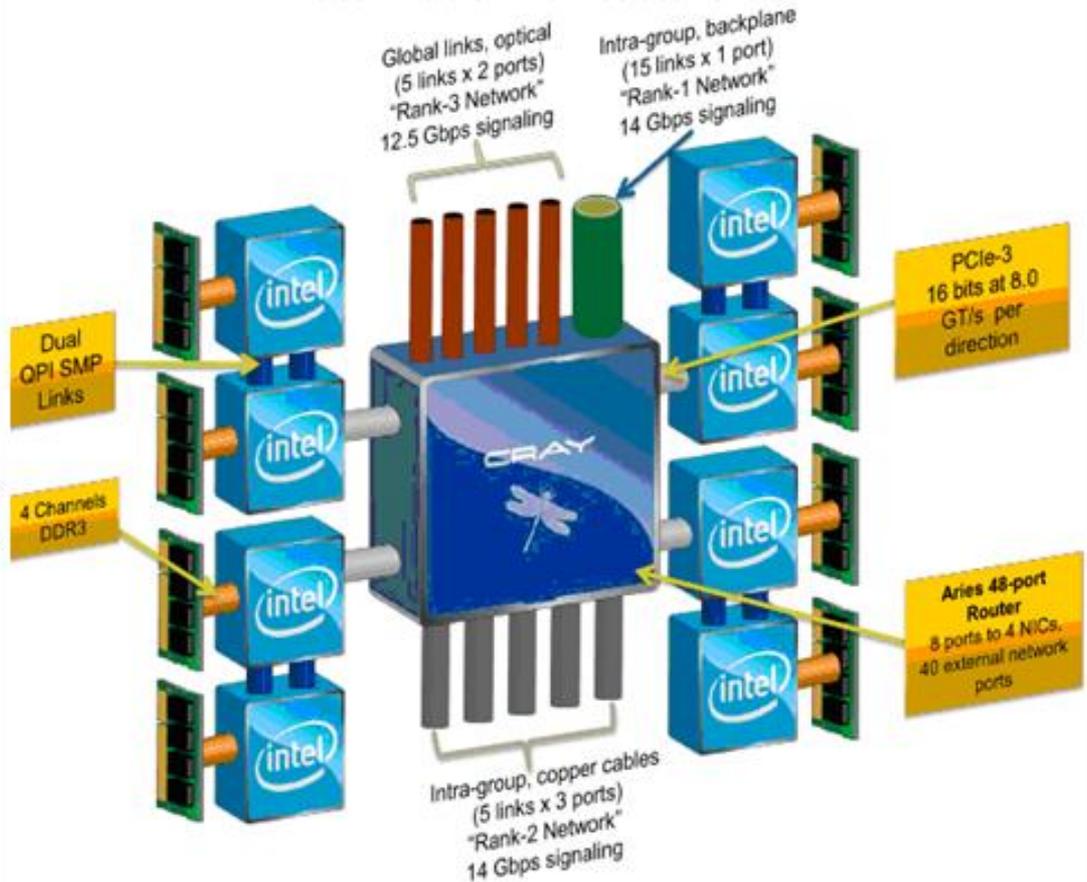
Узел и макроузел Cray XC30



Внутренняя сеть макроузла

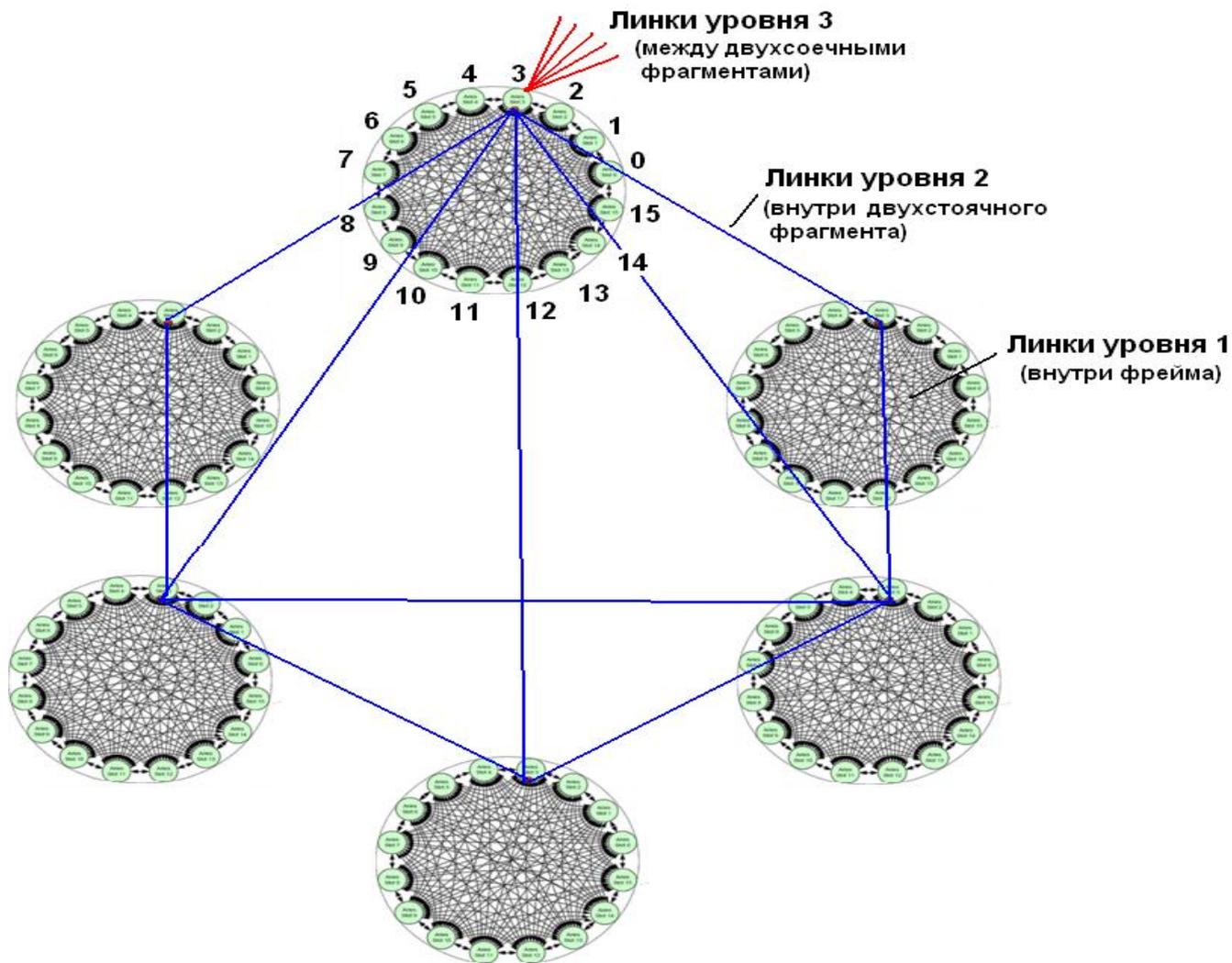


Внешний вид макроузла

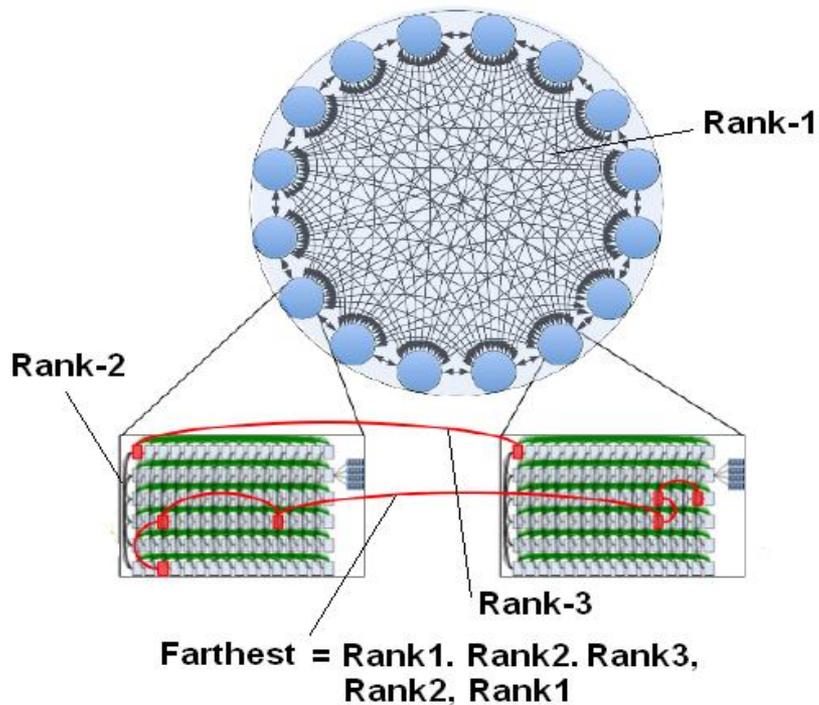
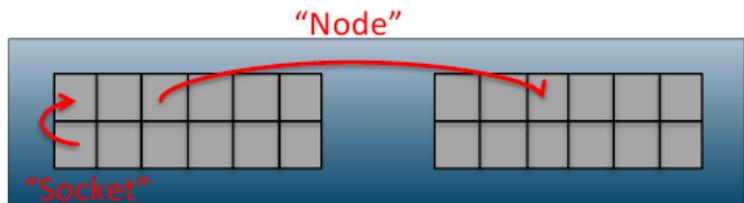


8-ми процессорный узел (8 Intel + YARC)

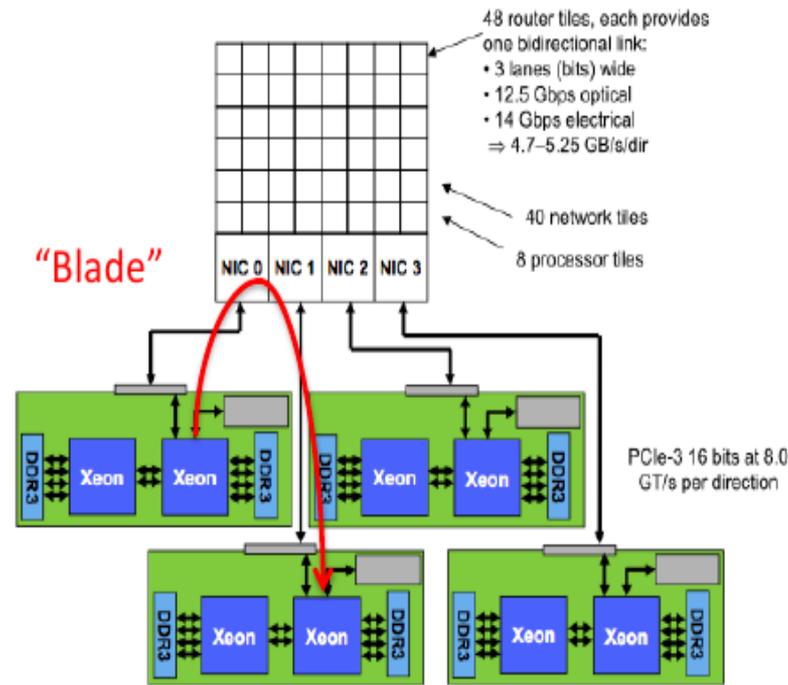
Многоуровневая сеть суперкомпьютера Cray XC30



Характеристики по уровням на тесте ping-pong



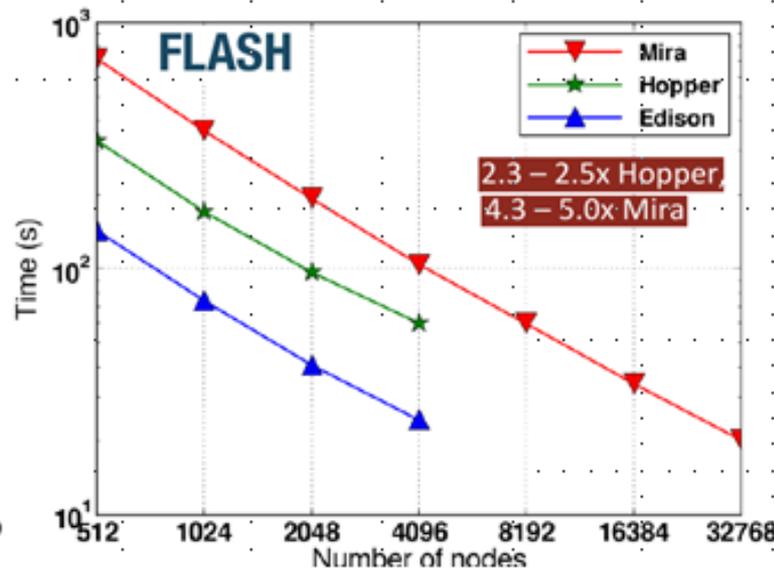
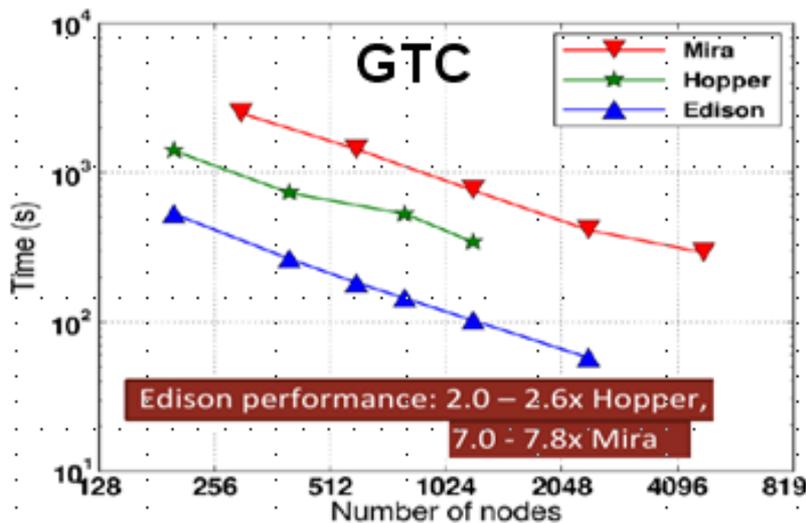
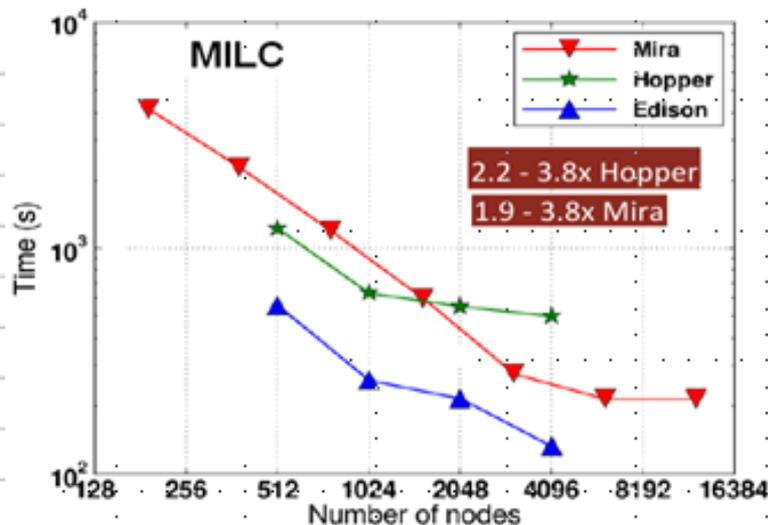
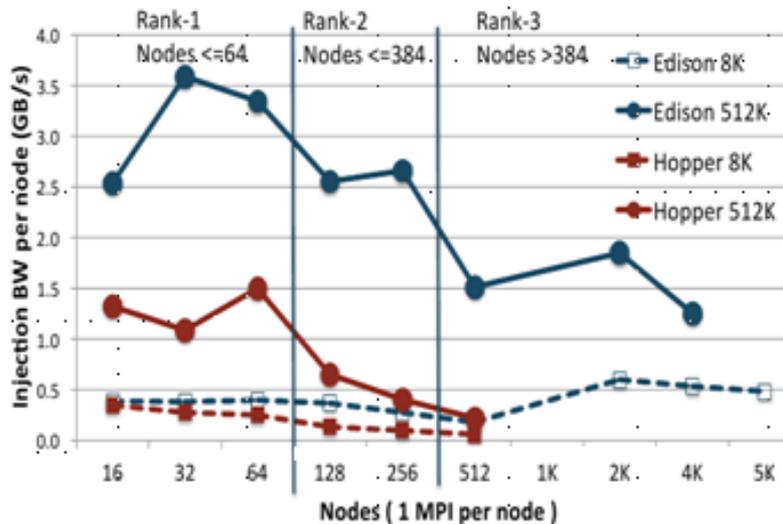
	Latency (us)	Bandwidth (GB/s)	Multi-BW (GB/s)
Rank-1	1.5	15.4	27.0
Rank-2	1.5	15.4	16.2
Rank-3	2.2	15.3	10.0
Farthest	2.3	15.3	5.1



	Latency (us)	Bandwidth (GB/s)
Socket	0.3	
Node	0.7	
Blade	1.3	14.9
Rank-1	1.5	15.4
Rank-2	1.5	15.4
Rank-3	2.2	15.3
Farthest	2.3	15.3

Характеристики на задачах

Тест AlltoAll



Экзафлопс

Перспективные архитектурно-программные принципы создания экзамасштабных суперЭВМ

- массовая однородная и неоднородная мультиредовость;
- неоднородность мультиредовости по объему (зернистости) и типу выполняемых в них операций;
- выделение в приложениях отдельных тесно взаимодействующих асинхронных процессов доступа к данным и вычислительных процессов;
- статическое и динамическое распараллеливание приложений;
- адаптивная локализация данных и вычислений;
- иерархическая организация памяти;
- специализация с ориентацией на приложения определенного типа и даже на отдельные задачи задач;
- применение функциональных, непроектурных и потоковых моделей организации программ для представления на уровне пользователя параллелизма выполнения программ в суперкомпьютере.

Перспективные направления развития элементно-конструкторской базы экзамасштабных суперЭВМ

- мультитядерность (1000-кратно);
- повышенную пропускную способность кристаллов по вводу-выводу (3D-компоновка, TSV-технологии);
- оптические соединения между платами (Holley-технологии матриц миниатюрных лазеров и линз передачи информации через оптоволоконные кабели, WDM-технологии передачи по каналам информации одновременно на разных длинах волн) и внутри кристаллов (нанофотонные волноводы);
- новая технология памяти (HMC и NVRAM с 3D-компоновкой, WDM-интерфейсами, встроенными процессорами).

10 актуальных проблем разработки экзафлопсных суперкомпьютеров в США

1. Энергоэффективность - создание энергоэффективных схем процессоров, маршрутизаторов коммуникационных сетей, технологий обеспечения питанием и технологий охлаждения.

2. Технологии соединений как внутри вычислительных узлов, так и между вычислительными узлами – увеличение производительности и сокращение задержек передачи данных, достижение энергоэффективности в линиях связи коммуникационных сетей и интерфейсах, наиболее важный показатель – снижение энергопотребления и повышения быстродействия обращений к памяти удаленных узлов (RDMA).

3. Технологии оперативной памяти – интеграция новых улучшенных технологий памяти для повышения емкости при повышении плотности размещения запоминающих элементов и сокращения их стоимости, снижение задержек выполнения операций с памятью, повышение их сложности и локализации, повышение пропускной способности даже для обращений с большой мелкозернистостью, т.е. когда обращения происходят к небольшим участкам памяти, а не к блокам большого объема.

10 актуальных проблем разработки экзафлопсных суперкомпьютеров в США

4. Создание масштабируемого (при увеличении параллелизма) системного программного обеспечения в виде операционных систем нового типа и систем поддержки выполнения программ (run-time систем), обеспечивающего высокий параллелизм уровня 10^9 (основной прирост, до трех порядков, ожидается непосредственно внутри вычислительного узла, а межузловой – на порядок), энергоэффективность за счет глубокого проникновения в управление работой оборудования и отказоустойчивость.

5. Системы программирования для пользователей – создание новых систем программирования, которые обеспечивают: создание эффективных с массовым параллелизмом программ, прозрачную для пользователя работу с иерархической глобально адресуемой памятью с обеспечением как эффективных удаленных обращений к памяти, так и локализацию данных при вычислениях и вычислений при данных; высокую многоуровневую отказоустойчивость.

6. Управление данными – создание хранилищ данных, программного обеспечения, которые бы справились с объемами и интенсивностью поступающих данных, их обработкой и хранением, ожидаемым разнообразием типов данных

10 актуальных проблем разработки экзафлопсных суперкомпьютеров в США

7. Создание экзамасштабных алгоритмов – переформулирование научных проблем и реконструирование или переработка алгоритмов их решения с целью эффективного выполнения на создаваемых суперкомпьютерах экзафлопсного класса.

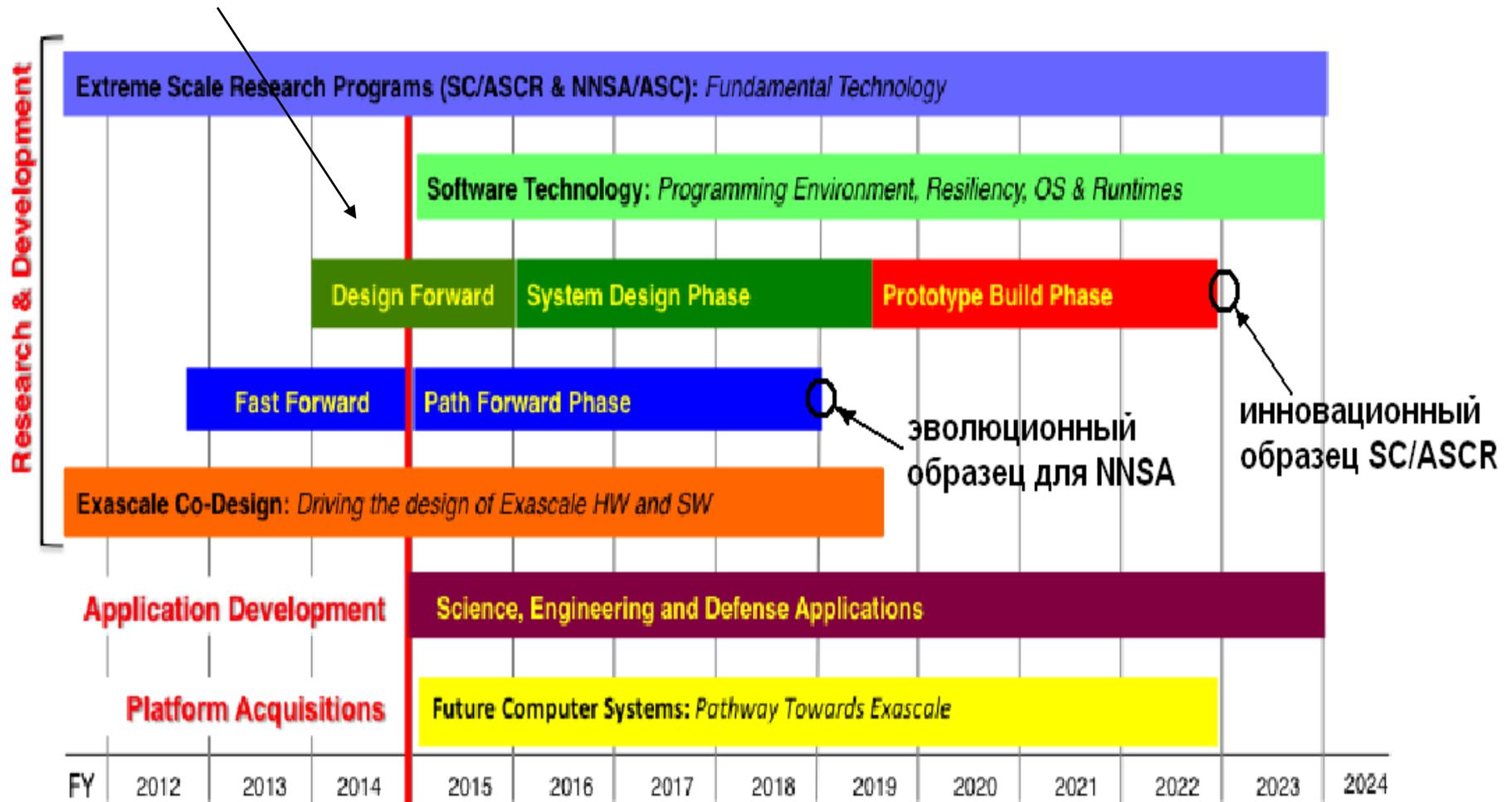
8. Создания алгоритмов автоматизации принятия решений по получаемым на экзафлопсных суперкомпьютерах результатам, что связано с оптимизацией принятия инженерных решений при создании сложных технических изделий, извлечением знаний из результатов научных расчетов и поступающих данных от сенсоров и физических установок.

9. Обеспечение как отказоустойчивости, так и достоверности вычислений в условиях наличия сбоя и отказов оборудования, ошибок программ и информационных не повторяемости результатов вычислений.

10. Повышение продуктивности разработки прикладных программ.

Дорожная карта создания экзафлопсных суперкомпьютеров DoE США

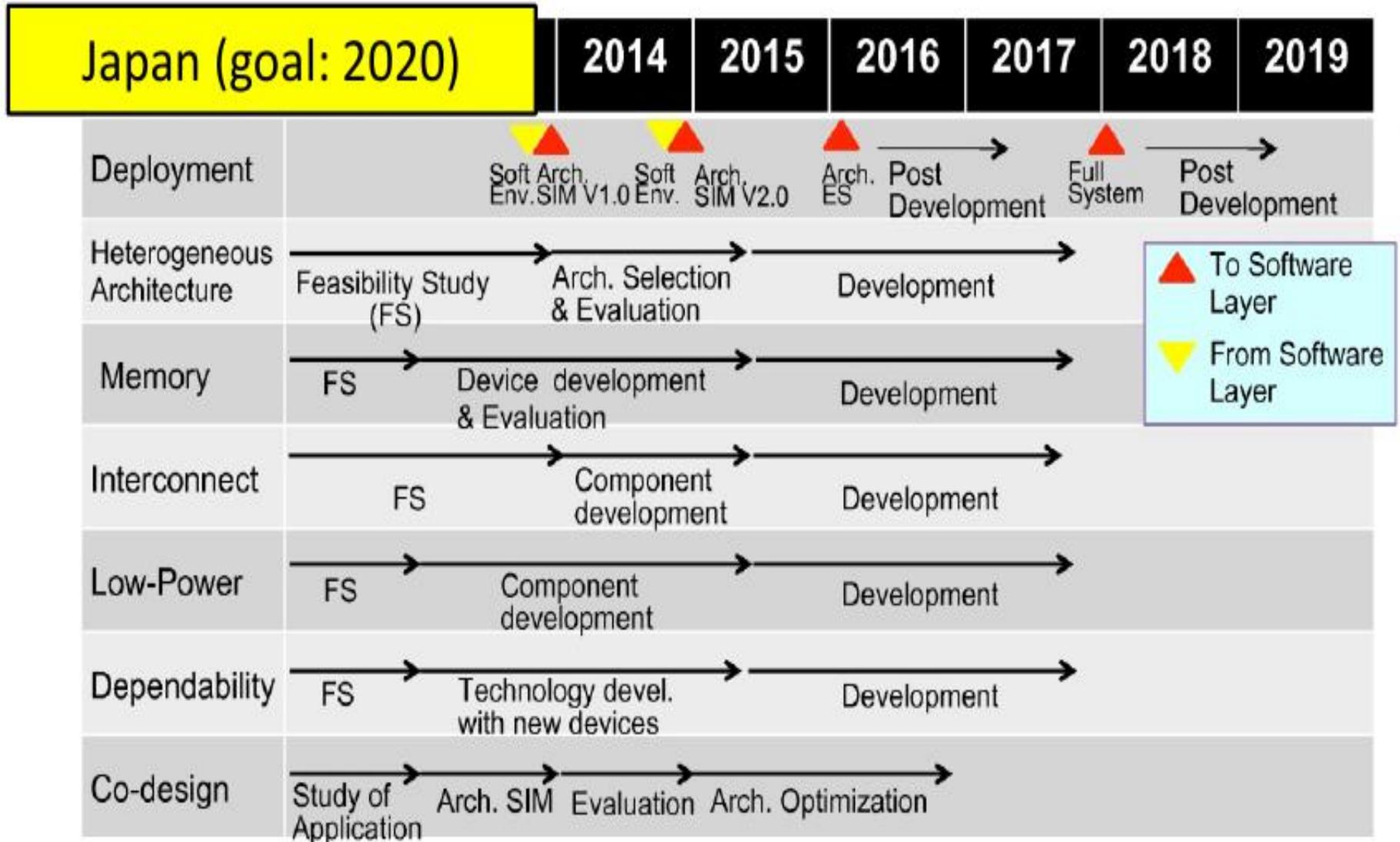
- ORNL Titan, Cray XK7, 27 PF
- LLNL Sequoia, IBM BG/Q 20 PF
- ANL Mira, IBM BG/Q, 10 PF
- LBNL Edison, Cray XC30, 2 PF
- LANL Cielo, Cray XE6, 1.1 PF



Реальная история и перспектива суперкомпьютеров NERSC (LBNL)

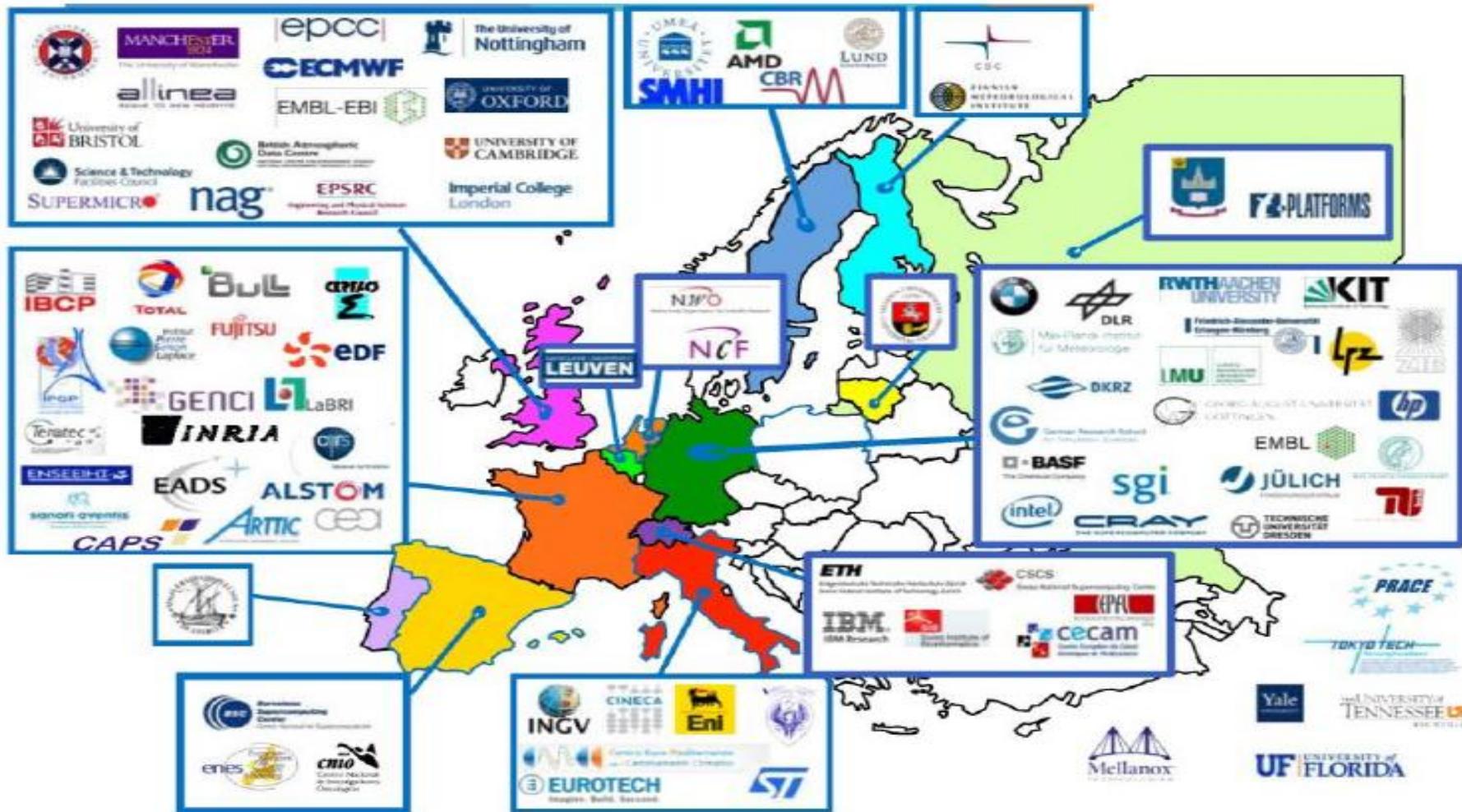
System attributes	NERSC-6	NERSC-7	NERSC-8 (proposed)	NERSC-9 (Proposed)
	Hopper	Edison	Cori 2015-2016	? 2019-2020
System peak	1.3 PF	2.6PF	20-40PF	200-300 PF
Power	2.9 MW (Peak) 2.2MW (Typical)	2.3 MW (Peak) 1.6 MW (Typical)	<5 MW (Peak)	< 15 MW (peak)
System memory	0.21 PB	0.35 PB	1-2 PB	~10 PB (128 GB on package, 512-1024 GB DRAM)
Node performance	202GF	460 GF	2-3.5TF	~10 TF
Node memory BW	50 GB/s	90 GB/s	100-500 GB/s	~200 GB/s ? 2-4 TB/s on package
Node concurrency	24 AMD Magnycours cores	24 Intel Ivy Bridge Cores	up to 300 Knight Landing Haswell	Up to 2048
System size (nodes)	6,384 nodes	5,576 nodes	8,000-12,000 nodes	O(10,000)
MPI Node Interconnect BW	~3 GB/s	~9GB/s	~9 GB/s	Up to 50 GB/s

Дорожная карта создания экзафлопсных суперкомпьютеров Японии



Участники работ по экзафлопсной тематике в Европе

European Union (goal: 2020)



Российские инновационные СКТ

Первоочередные задачи для страны в области инновационных СКТ

- **Разработка отечественного микропроцессора с массово-мультитредовой архитектурой и многопортового коммутатора**
- **Разработка проблемно-специализированных микропроцессоров на базе перспективных архитектурных принципов**
- **Проведение исследований и разработок по новым моделям вычислений и организации памяти, а также эмуляции массово-мультитредовых суперкомпьютеров с глобально адресуемой памятью**
- **Организация работы экспертного сообщества для планирования и оценки работ, формирование нескольких десятков исследовательских групп для работы в области инновационных суперкомпьютерных технологий**
- **Срочное формирование коллективов и организация работ по сверхпроводниковой электронике и нанофотонике, для целенаправленной работы в области ЭКБ перспективных суперЭВМ**

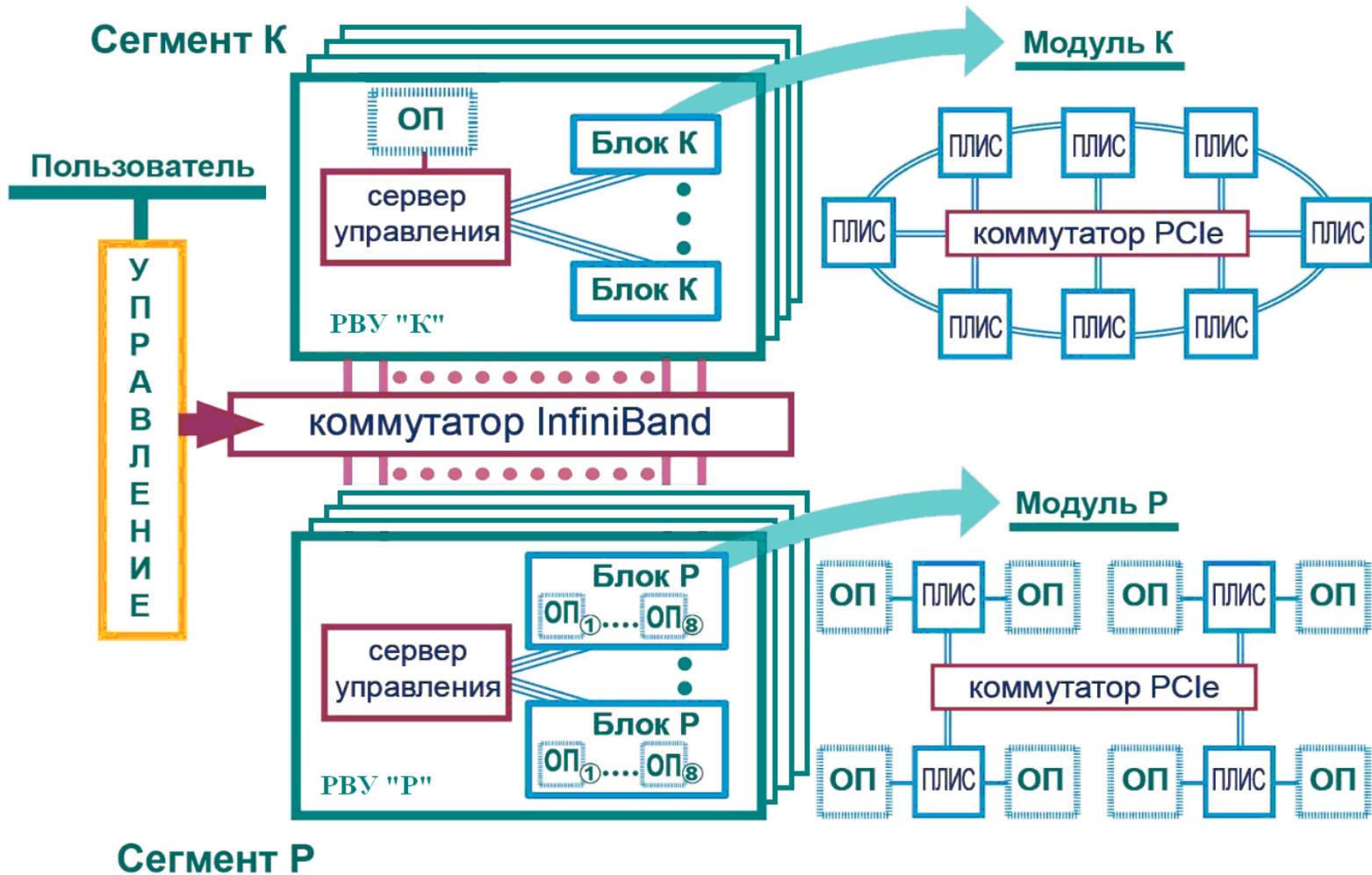
Некоторые актуальные задачи ФГУП “НИИ “Квант”

- 1. Оптимизация параллельных программ реализации интересующего класса алгоритмов на кластерных суперкомпьютерах с ускорителями на GPU.**
- 2. Разработка реализаций алгоритмов на ускорительных платах с множеством ПЛИС**
- 3. Оптимизация настроек операционных систем и разработка программ мониторинга суперкомпьютеров, планировщиков выполнения программ**
- 4. Аппаратно-программные комплексы для работы с большими данными**
- 5. Перспективные исследования по аппаратно-программным принципам и ЭКБ суперкомпьютеров экза-, зетта- и йотта-уровня (в частности - проект МГВС РАН)**

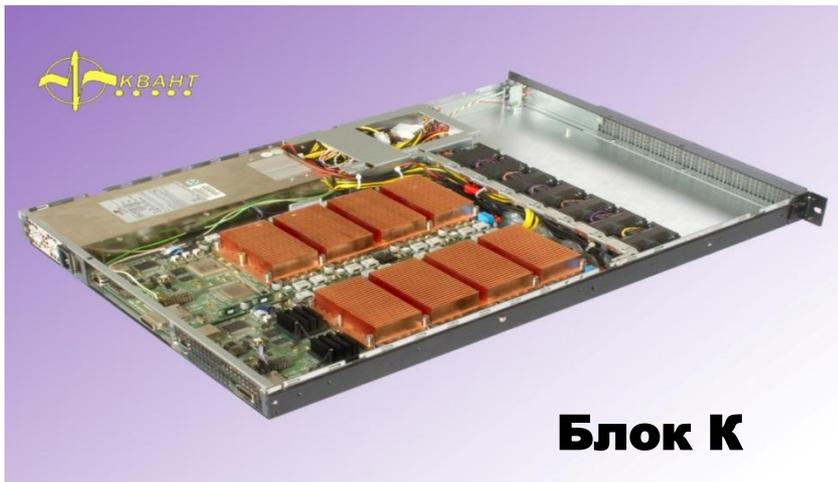
**МГВС - Моделирующая
Гибридная Вычислительная
Система**

keldysh.ru/exaflops.pdf

Структура МГВС



Реконфигурируемые вычислительные узлы сегментов К и Р



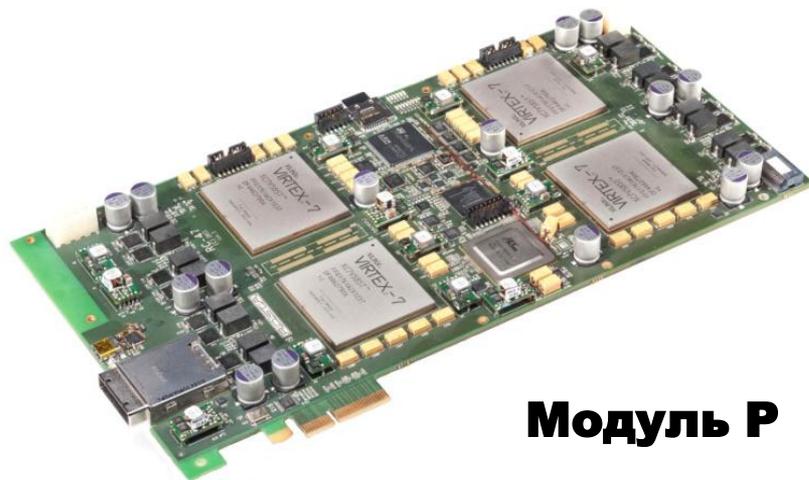
Блок К



Модуль К



Блок Р



Модуль Р

Направления исследований на по проекту МГВС по экзамасштабной тематике

Восток-1: легкие однородные узлы, массовая многоядерность, разделенная глобально адресуемая память

Восток-2: тяжелые гибридные узлы с массовой мультитредовостью, многоядерностью, потоковостью и иерархической разделенной глобально адресуемой памятью

Восток-2: основные направления исследований

HMT (Hybrid MultiThreaded) - экспериментальный гибридный многоядерный мультитредово-поточковый микропроцессор

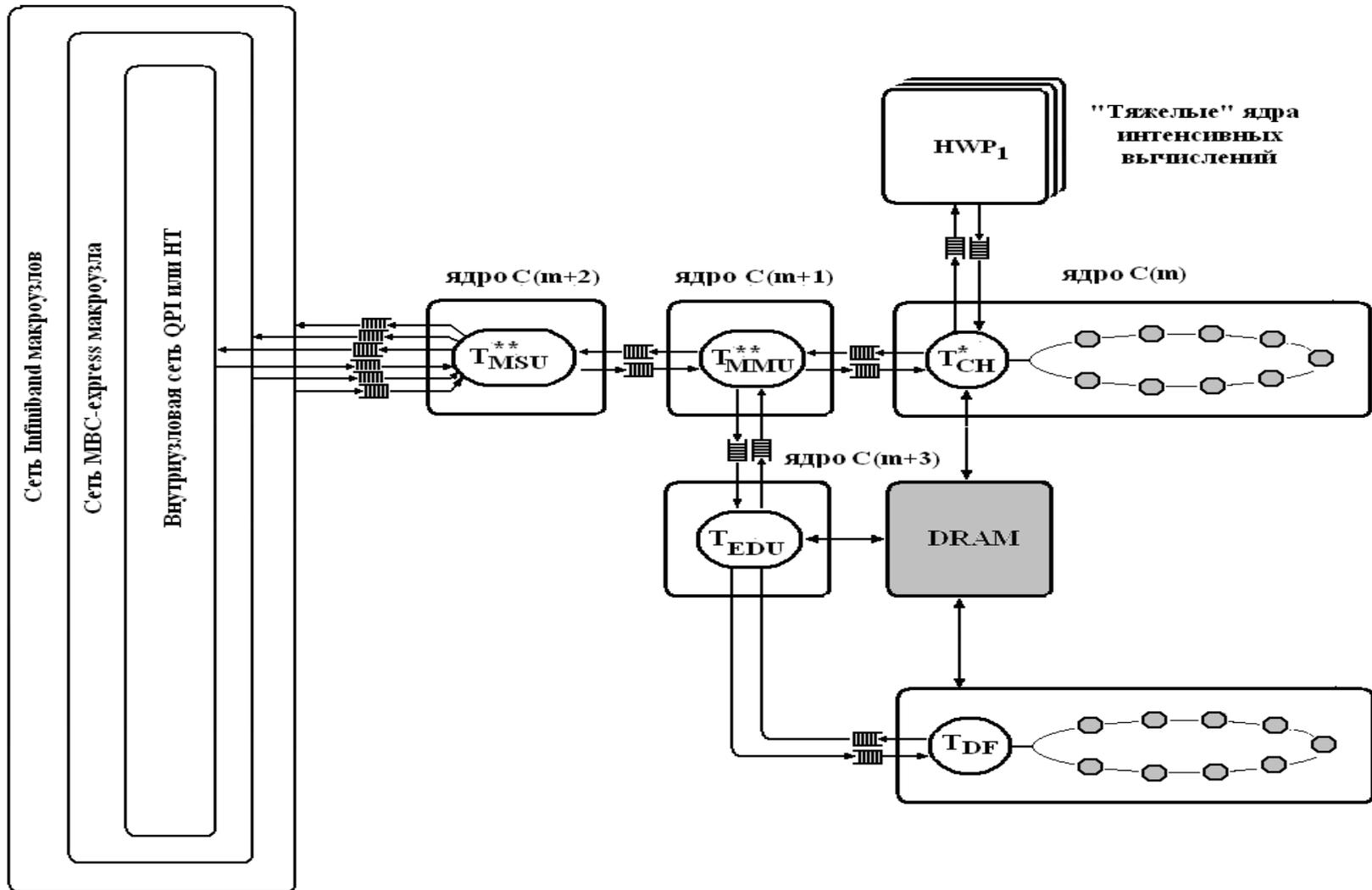
MTE (MultiThreaded supercomputer Emulation) - эмуляция суперкомпьютеров с мультитредовой архитектурой и виртуальной глобально адресуемой памятью

FSC (Functional Super Computer) - реализация суперкомпьютера с внутренним языком высокого уровня на базе непроцедурного функционального высокого уровня (Рефал)

SDF (Static Data Flow) - процессор в виде настраиваемой неоднородной среды вычислительных элементов и элементов памяти

DDF (Dynamic Data Flow) - суперкомпьютер на базе динамических потоковых графов и средств ассоциативной обработки информации

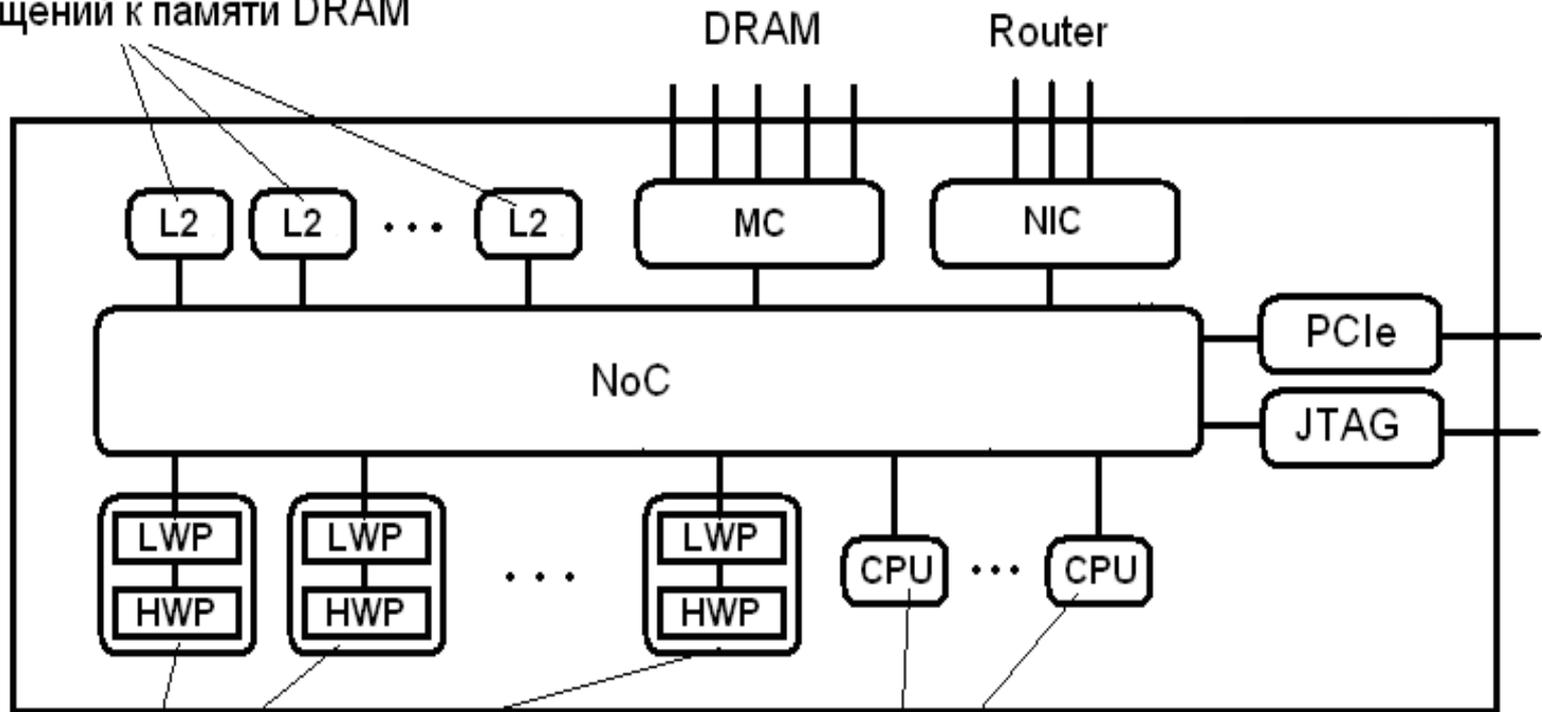
Базовая реализация модели вычислительного узла суперкомпьютера *Восток-2*



**Разработка специализированной
микропроцессорной СБИС с
гибридной массово -
мультитредовой архитектурой.**

Структура специализированной микропроцессорной СБИС

Кэш-памяти для локализации обращений к памяти DRAM



Гибридные мультитредово-потокковые HW/LW-макроядра с ядрами поддержки легких (LWP) и тяжелых (HWP) тредов

Суперскалярные ядра (типа ARM)

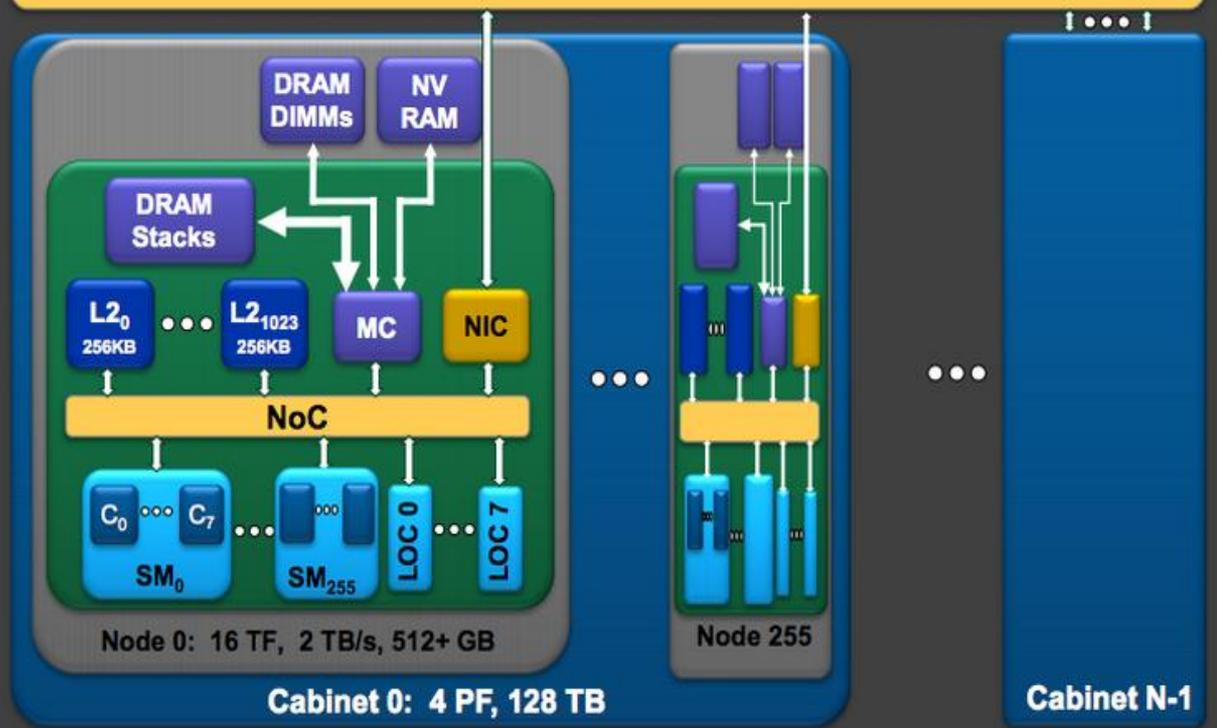
Проект Echelon

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

2018 Vision: Echelon Compute Node & System



System Interconnect



Echelon System (up to 1 EF)

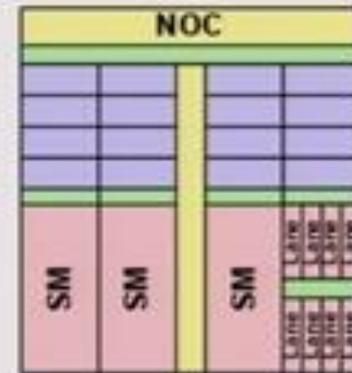
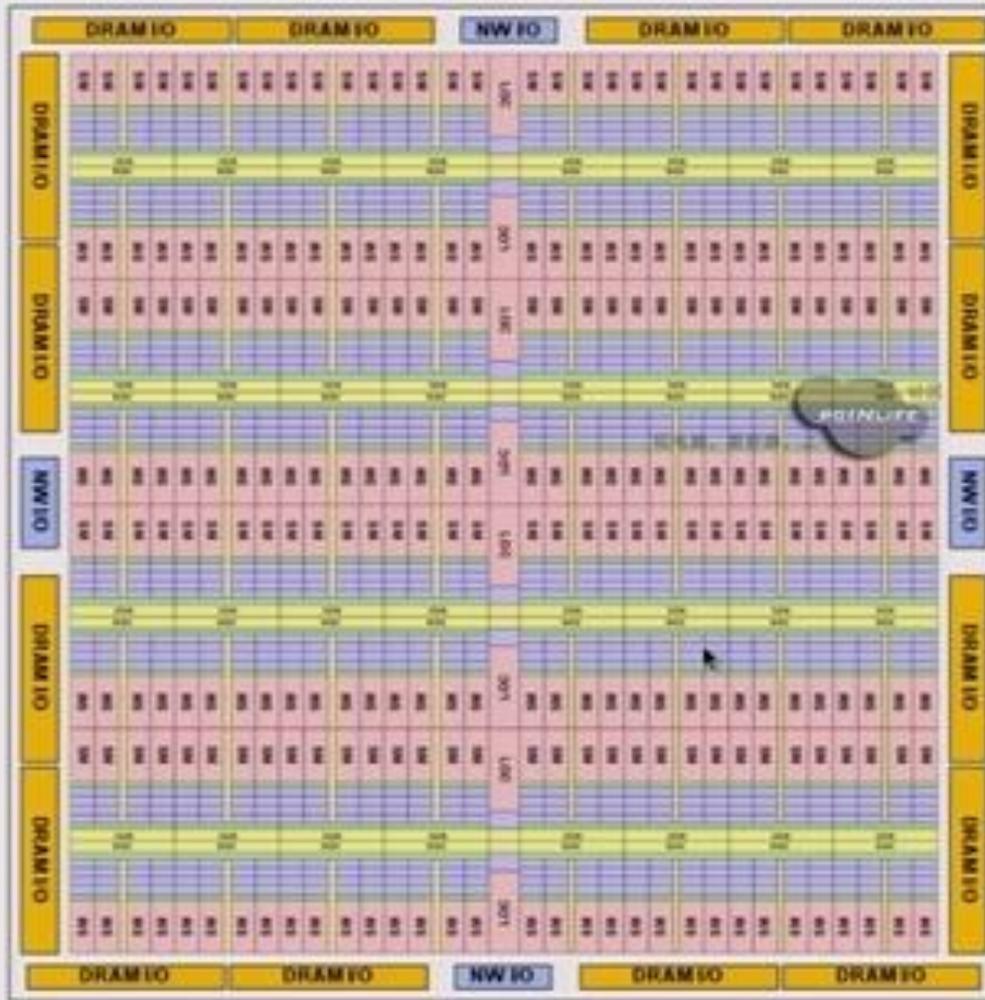
Key architectural features:

- Malleable memory hierarchy
- Hierarchical register files
- Hierarchical thread scheduling
- Place coherency/consistency
- Temporal SIMT & scalarization
- PGAS memory
- HW accelerated queues
- Active messages
- AMOs everywhere
- Collective engines
- Streamlined LOC/TOC interaction

Структура микропроцессора



Echelon Chip Floorplan



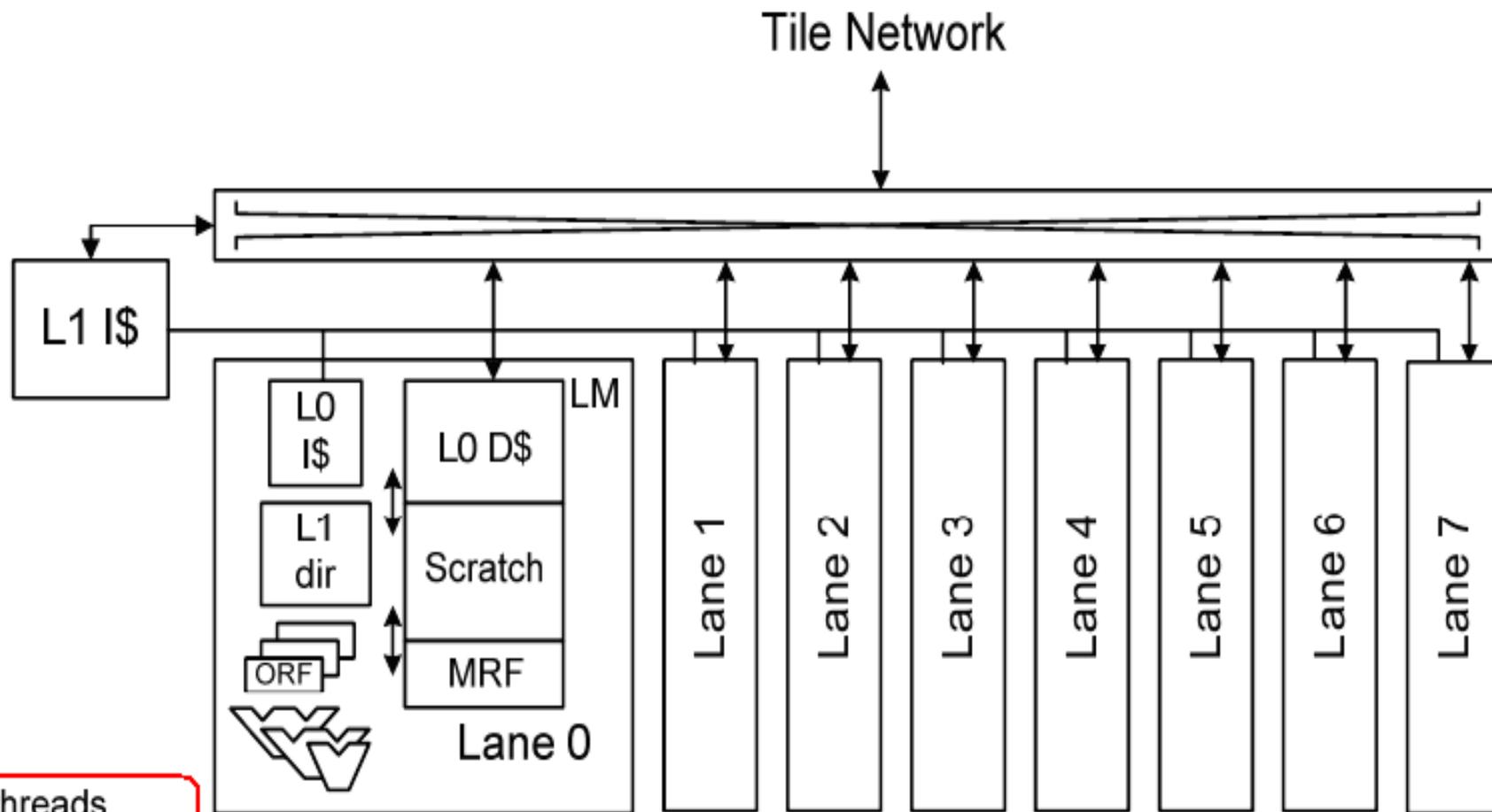
17mm

10nm process
290mm²



中国科学院

Проект Echelon: структура SM-ядра



512 threads

32 active threads

16 DFMAs (32 FLOPs/clock)

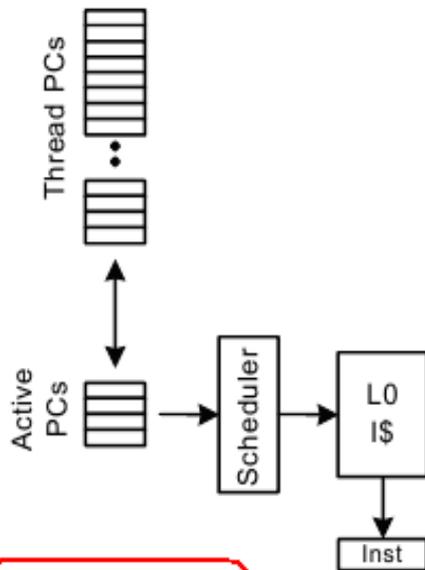
L1 I\$: 2K instructions (32KB)

RF/Scratch/D\$: 256KB

L0 caches in other lanes form L1 cache

Проект Echelon: полоса обработки (Lane) SM-ядра

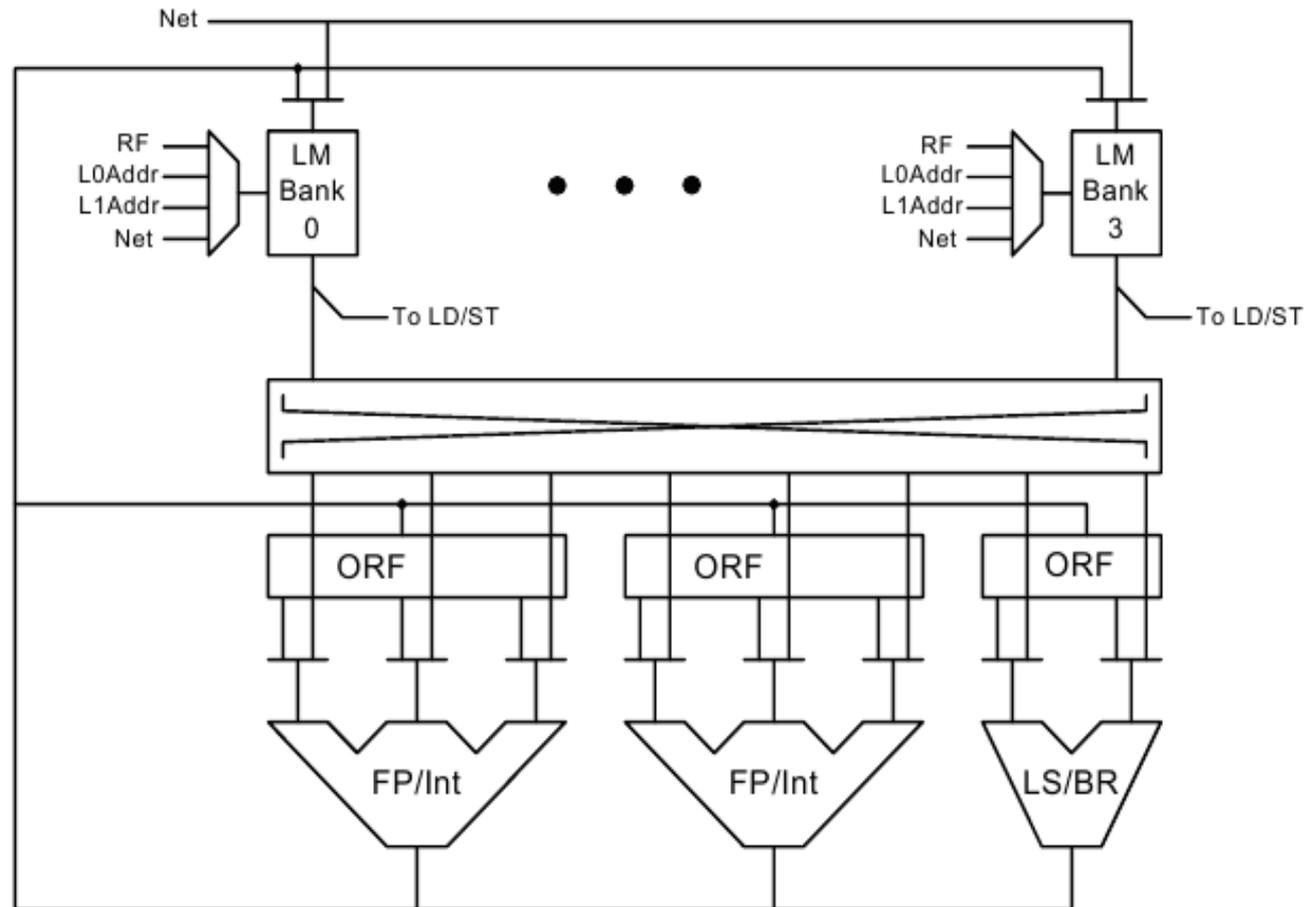
Control Path



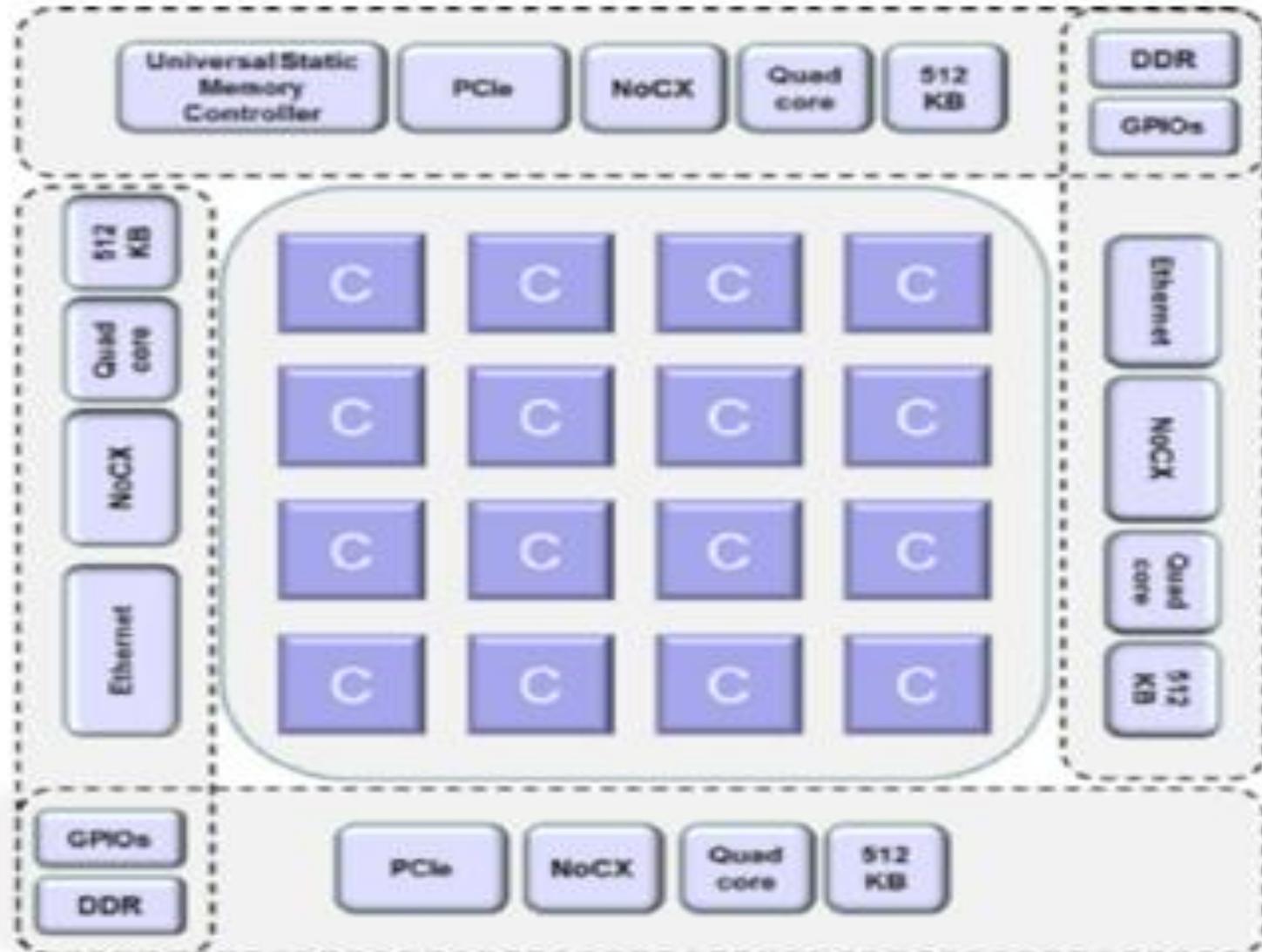
64 threads
4 active threads

2 DFMAs (4 FLOPS/clock)
ORF bank: 16 entries (128 Bytes)
L0 I\$: 64 instructions (1KByte)
LM Bank: 8KB (32KB total)

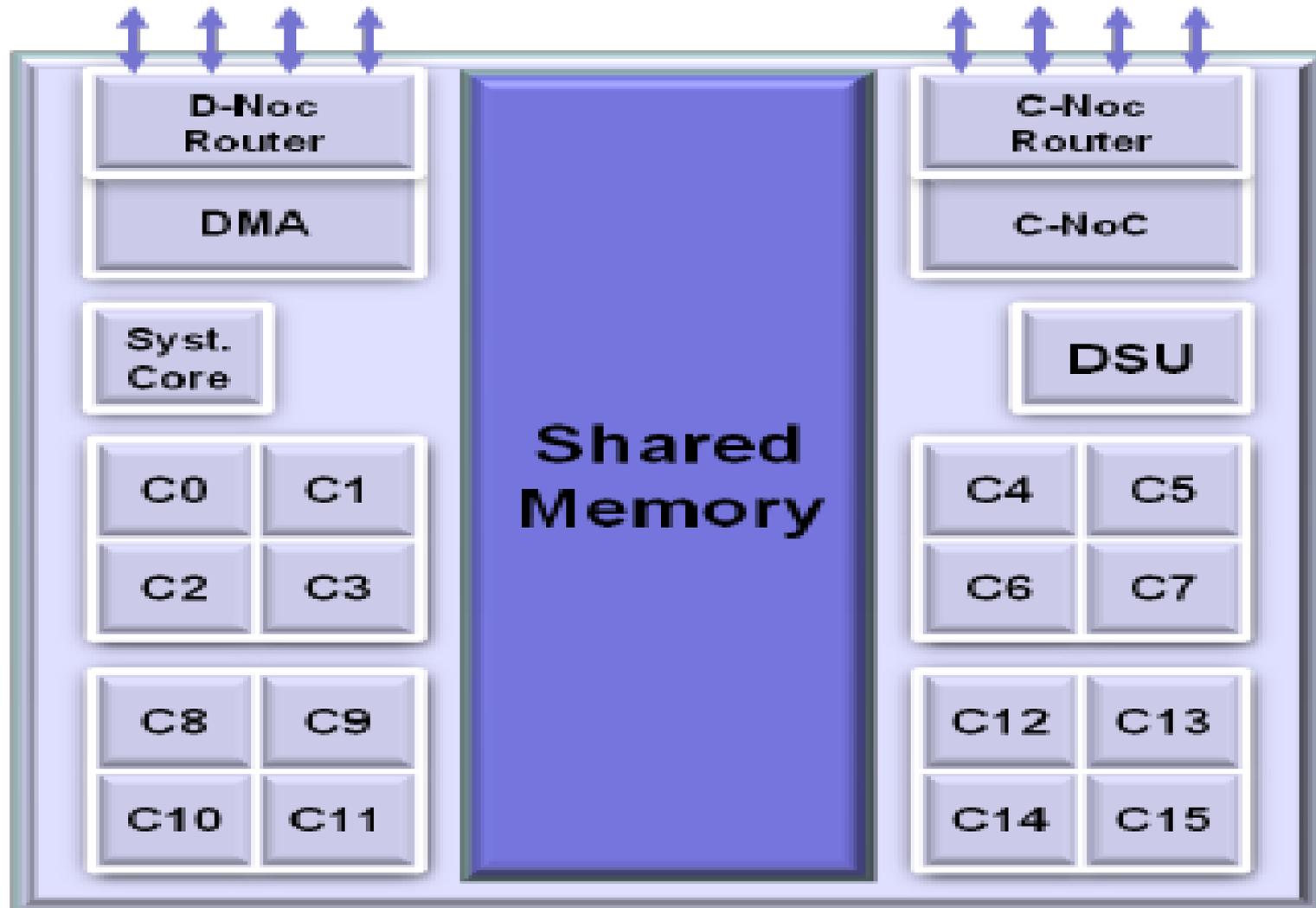
Data Path



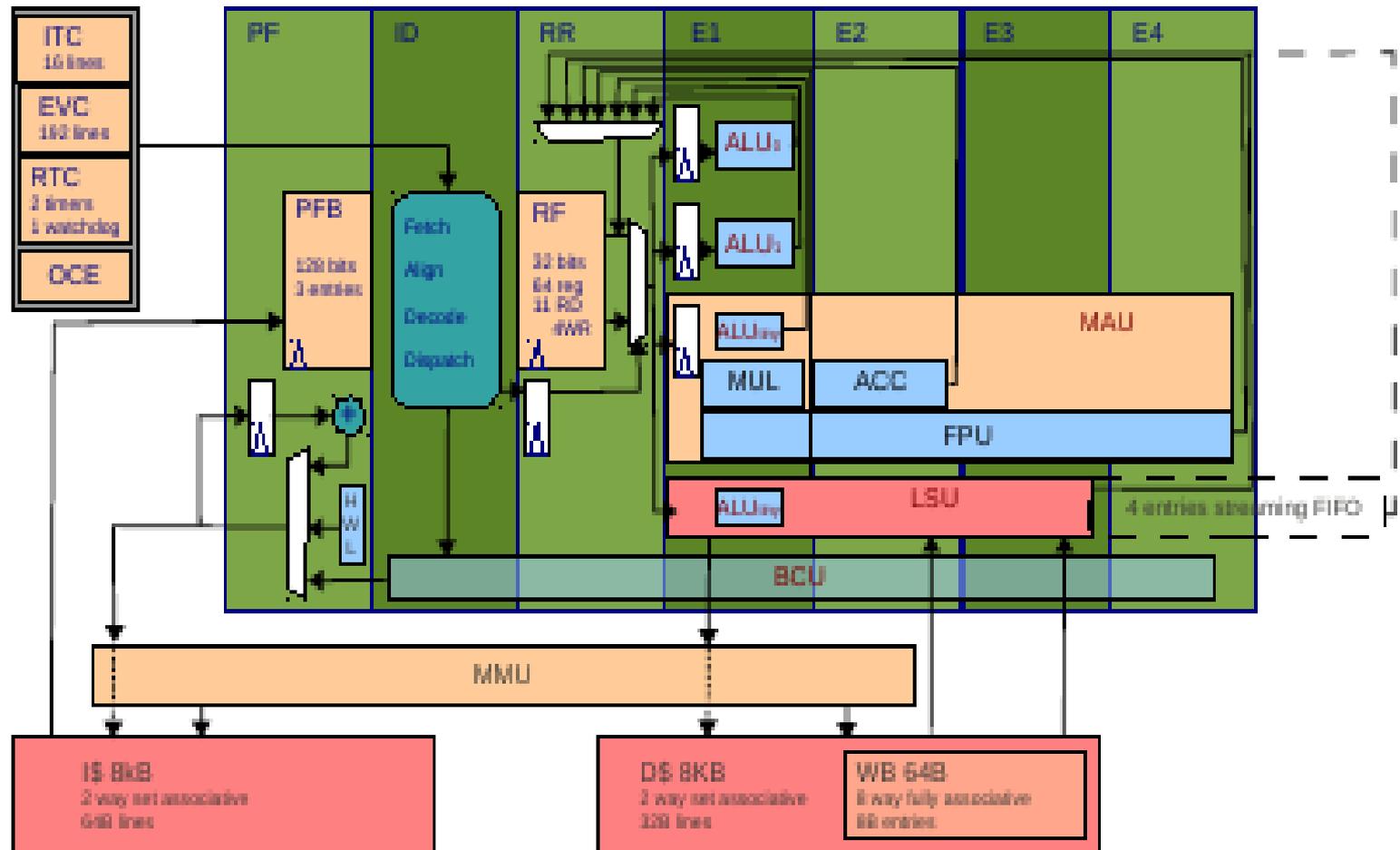
Микропроцессор MPPA-256, фирма Kalray (Франция)



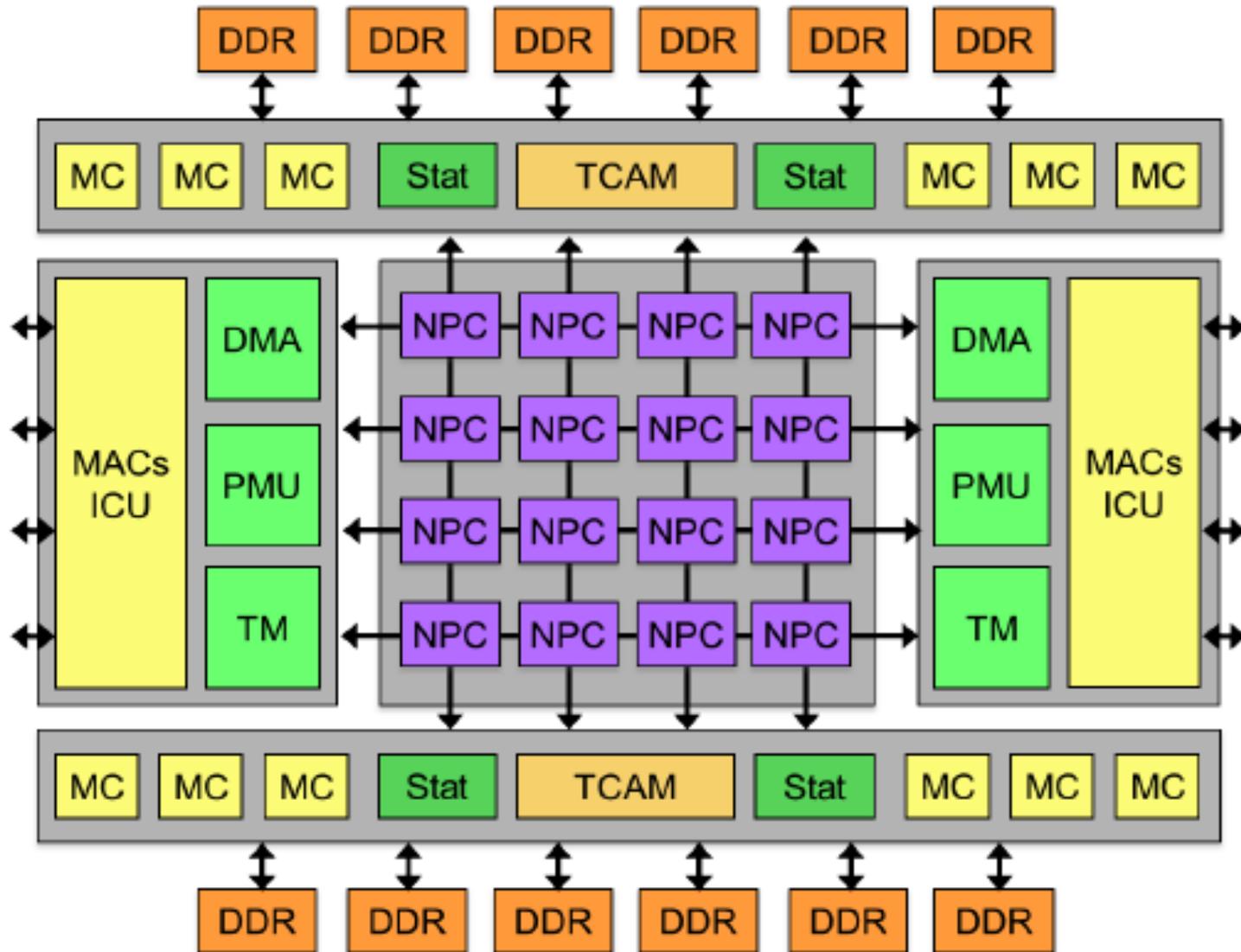
Кластер микропроцессора MPRA-256 (16 VLIW-ядер + ядро управления)



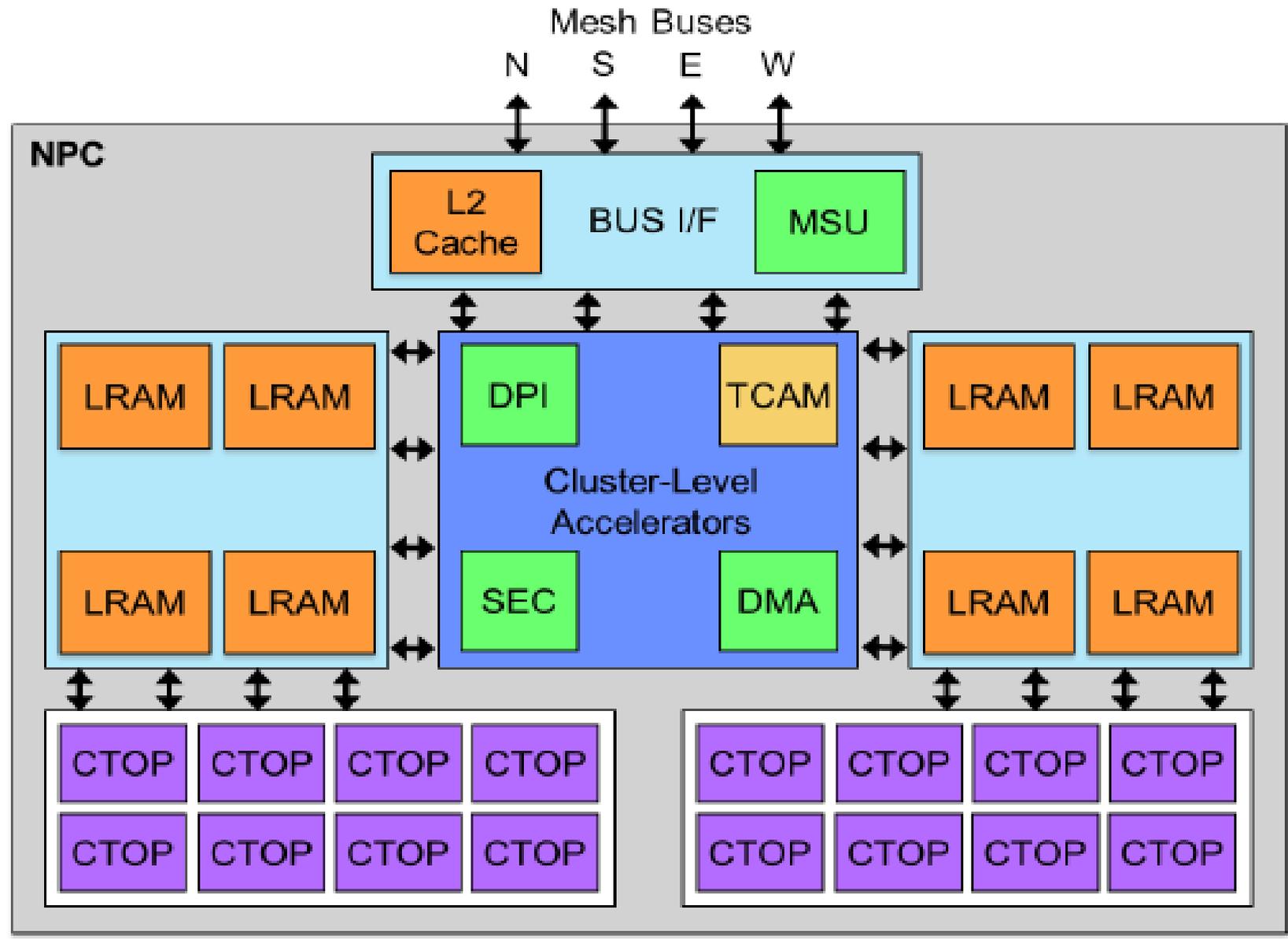
VLIW-ядро кластера микропроцессора MPRA-256



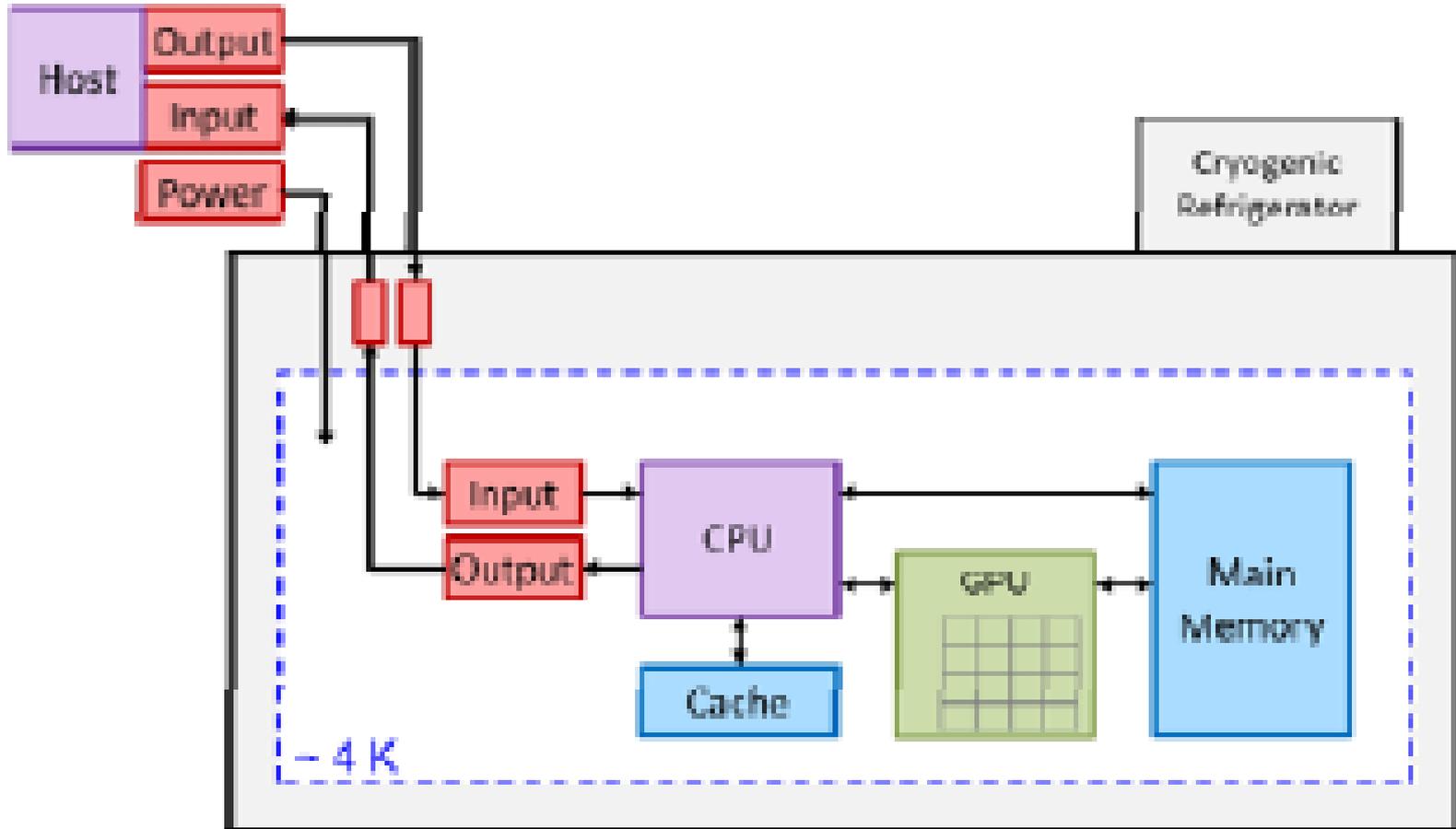
Сетевой микропроцессор NPS-400 фирмы EZChip (Израиль)



Кластер NPC микропроцессора NPS-400



Вычислительный узел криогенного суперкомпьютера, создаваемого по программе IARPA C3 (5 лет)



Характеристики модуля и компонентов

#	Parameter	Goal
1	Throughput (bit-op/s)	10^{13}
1	Efficiency @ 4 K (bit-op/J)	10^{15}
2	Main memory, total (B)	2^{27}
2	Cache memory, total (B)	2^{15}
2	Input/Output (bit/s)	10^8

Общие

CPU

GPU

#	Parameter	Goal
1	Word size (bit)	64
2	Efficiency @ 4 K (bit-op/J)	5×10^{15}
2	Processor class	ARM™ or Intel Atom™
2	Instruction set	ARM™ or simple x86
2	ALUs	1 integer
2	Throughput (bit-op/s)	10^{12}

#	Parameter	Goal
2	Efficiency @ 4 K (bit-op/J)	10^{16}
2	Instruction size (bit)	64
2	Data size (bit)	64
2	PU count	8
2	PU register size (bit)	64×128
2	PU ALUs	1 integer
2	PU throughput (bit-op/s)	2×10^{12}
3	PU area (mm ²)	10

Cache

Memory

#	Parameter	Goal
1	Access energy @ 4 K, ave. (J/bit)	$5e-18$
1	Power, static (W/bit)	$< 5e-19$
2	Write time (ps)	200
2	Read time (ps)	100
2	Read rate, burst mode (Gbit/s)	300
2	Capacity per memory chip (bit)	$\geq 2^{16}$
2	Density (bit/cm ²)	$1e+7$
3	Read/Write error rate	$1e-12$
3	Read/Write disturb rate	$1e-12$

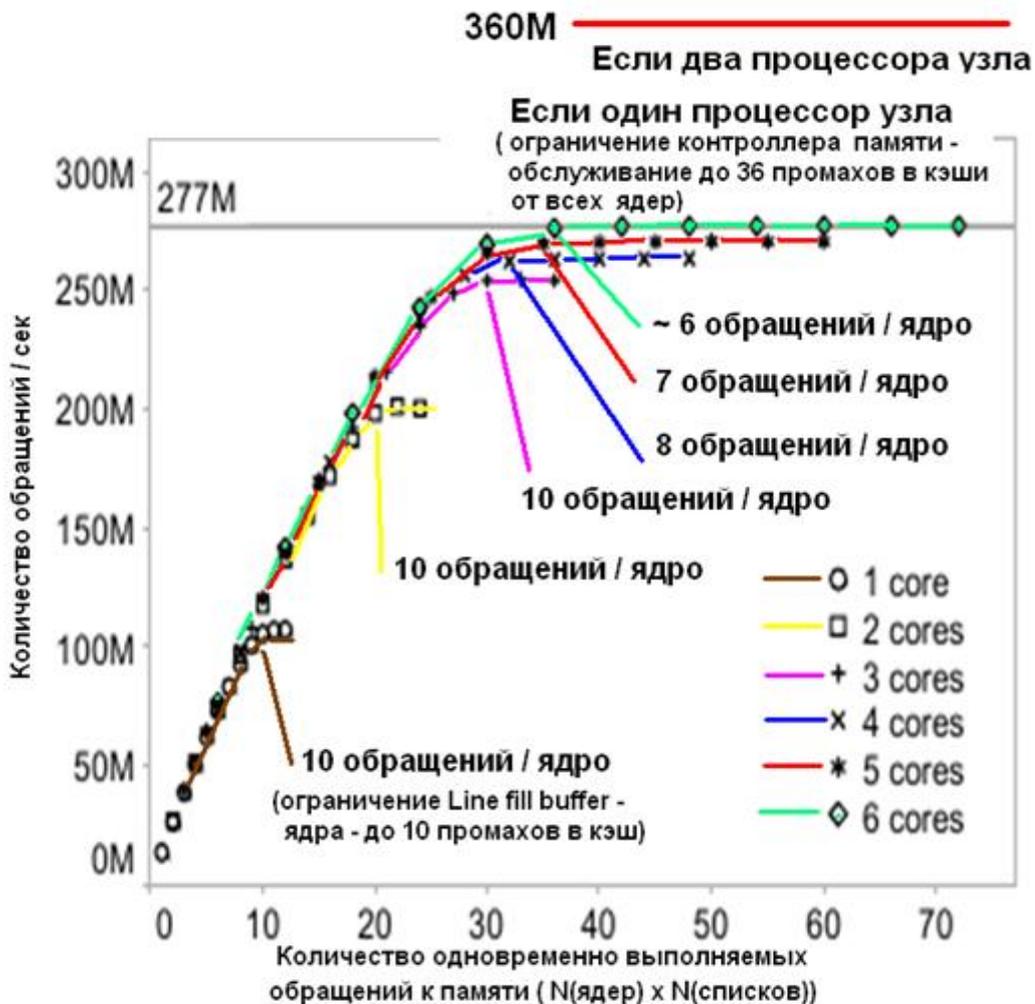
#	Parameter	Goal
1	Access energy @ 4 K, ave. (J/bit)	$5e-17$
1	Power, static (W/bit)	$< 5e-19$
2	Write time (ps)	2,000
2	Read time (ps)	500
2	Read rate, burst mode (Gbit/s)	50
2	Capacity per memory chip (bit)	$\geq 2^{26}$
2	Density (bit/cm ²)	$1e+8$
3	Read/Write error rate	$1e-12$
3	Read/Write disturb rate	$1e-12$

**Эмуляция HPGAS и массовой
мультитредовости на кластерных
суперкомпьютерах.
Модели параллельных программ
для экзафлопсных машин.
Функциональный суперкомпьютер**

Аппаратная поддержка одновременного выполнения обращений к памяти

```
while (count-- > 0) {  
    list1 = list1->next;  
    list2 = list2->next;  
    ...  
    listn = listn->next;  
}
```

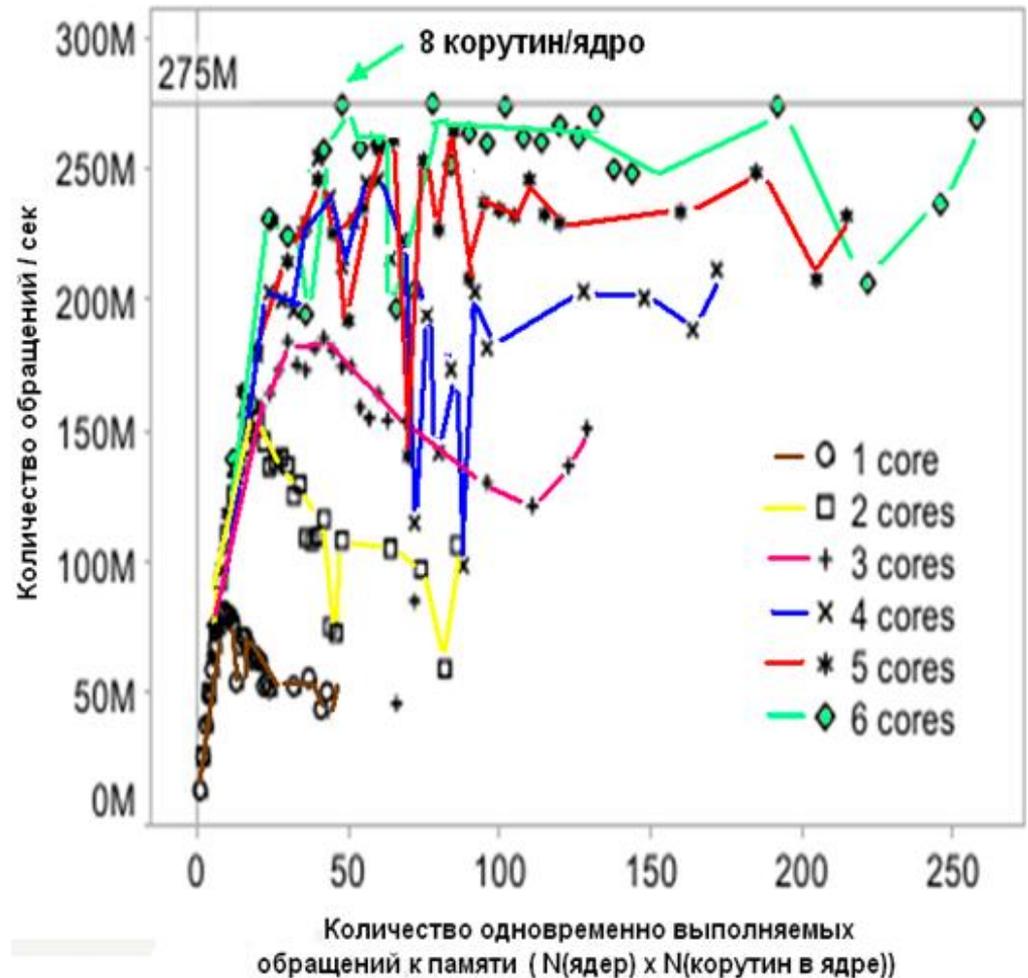
(a)



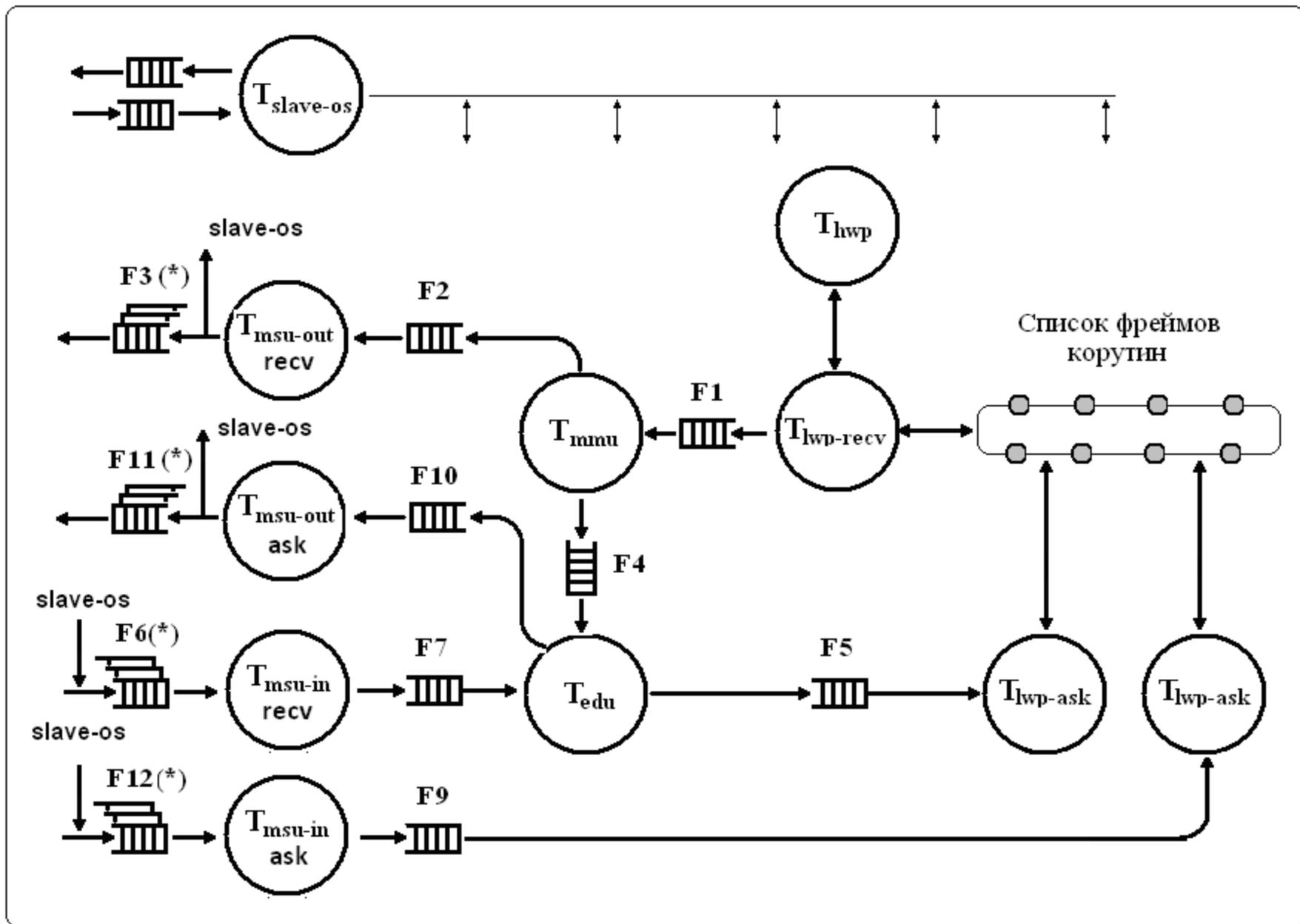
Программная поддержка одновременного выполнения обращений к памяти через корутины.

```
while (count-- > 0) {  
    readIssue(&(list->next));  
    yield();  
    list = readComplete(&(list->next));  
}
```

(b)

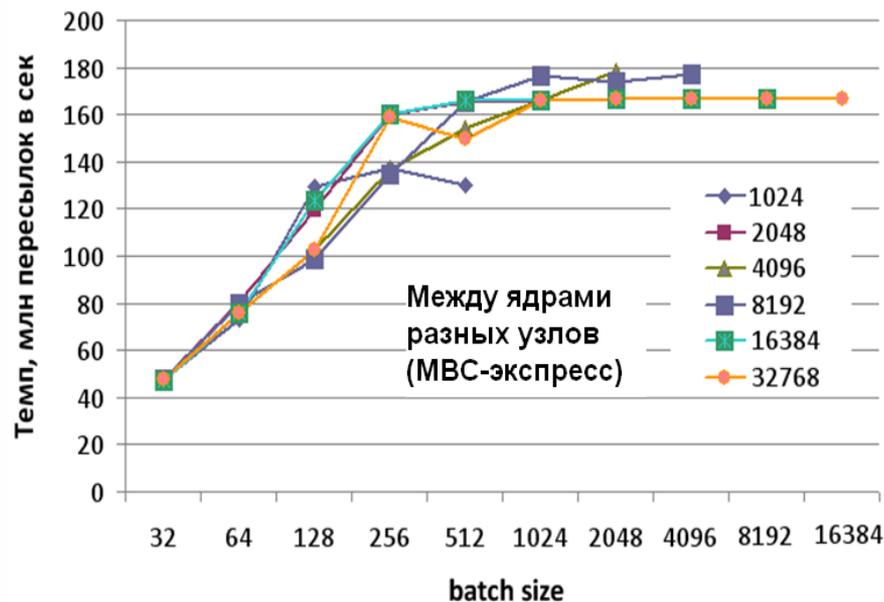
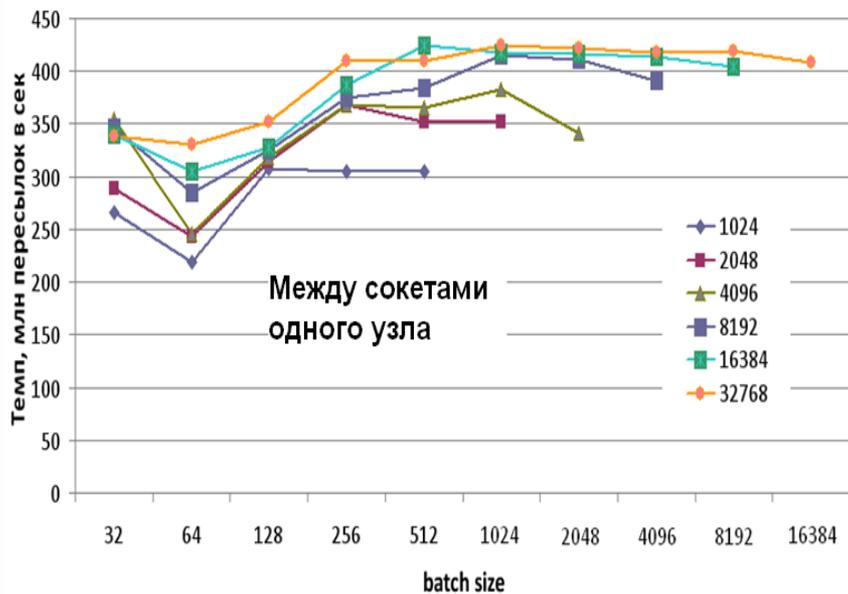
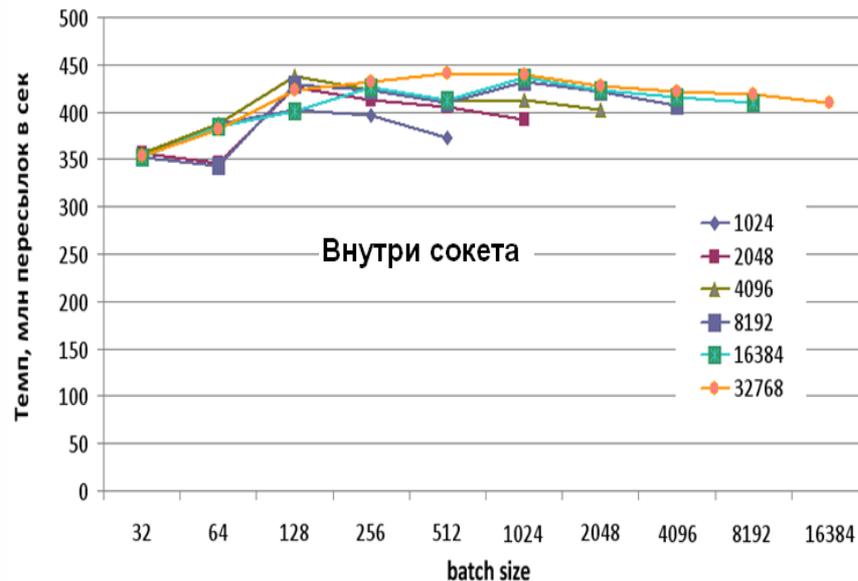
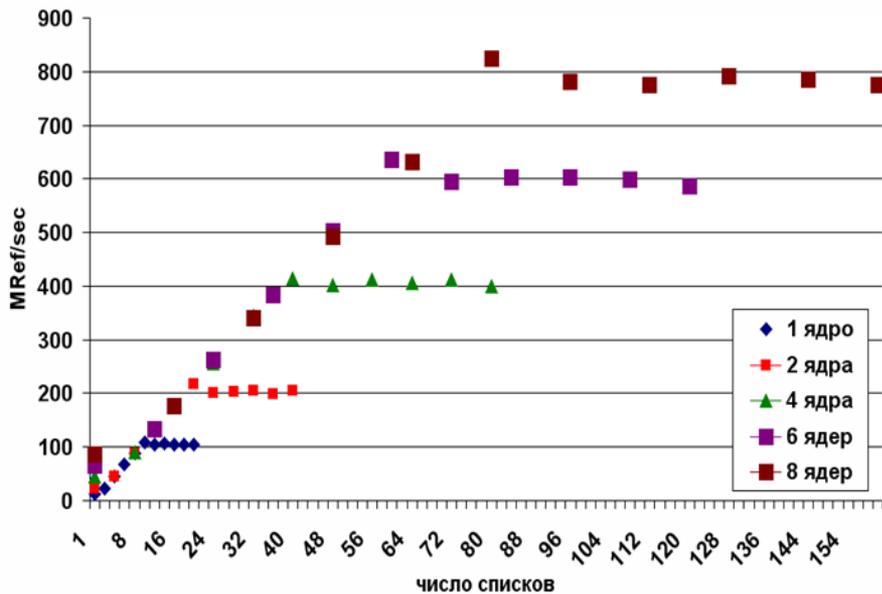


Уточненная схема логического узла

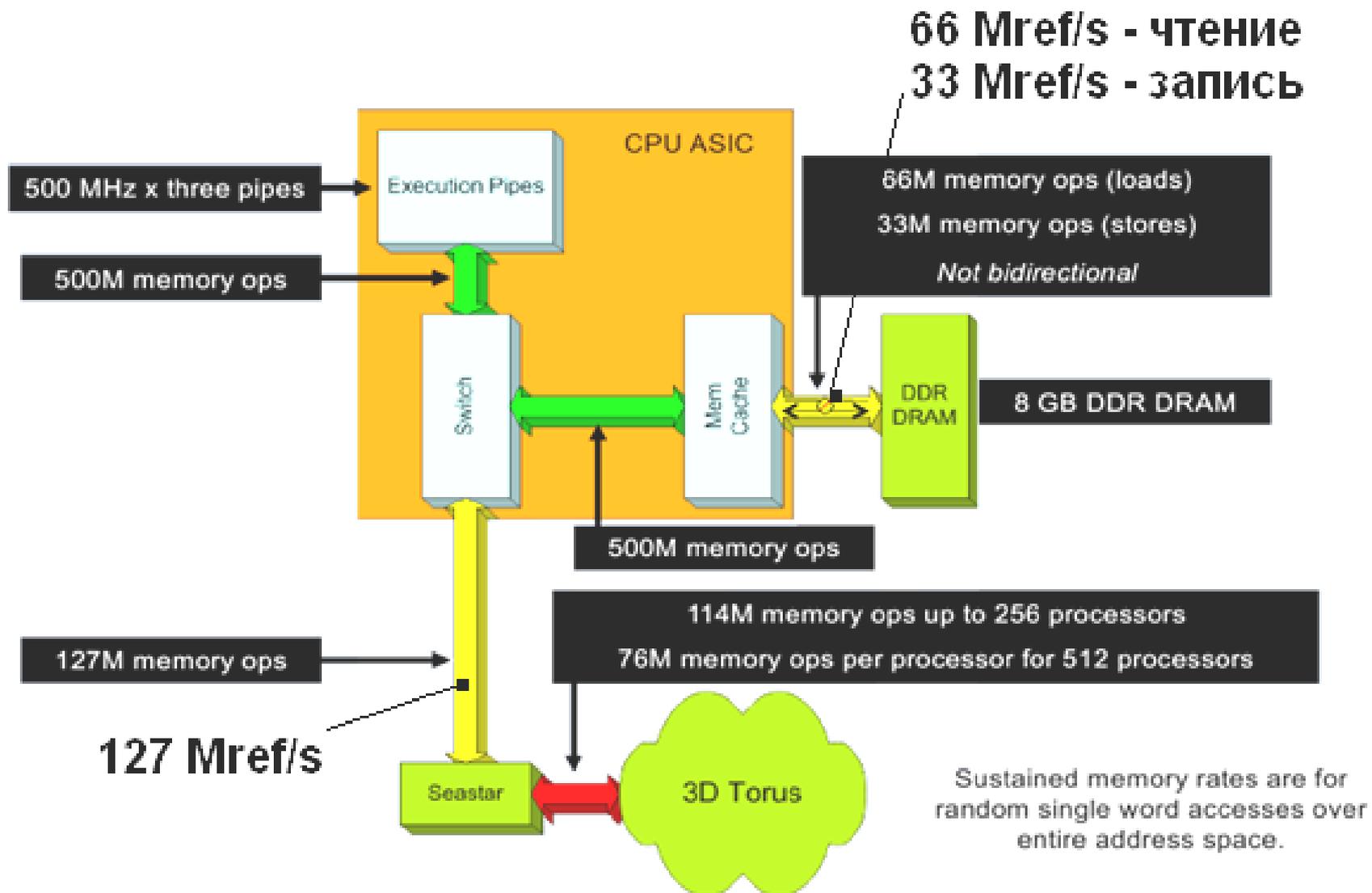


Результаты экспериментов

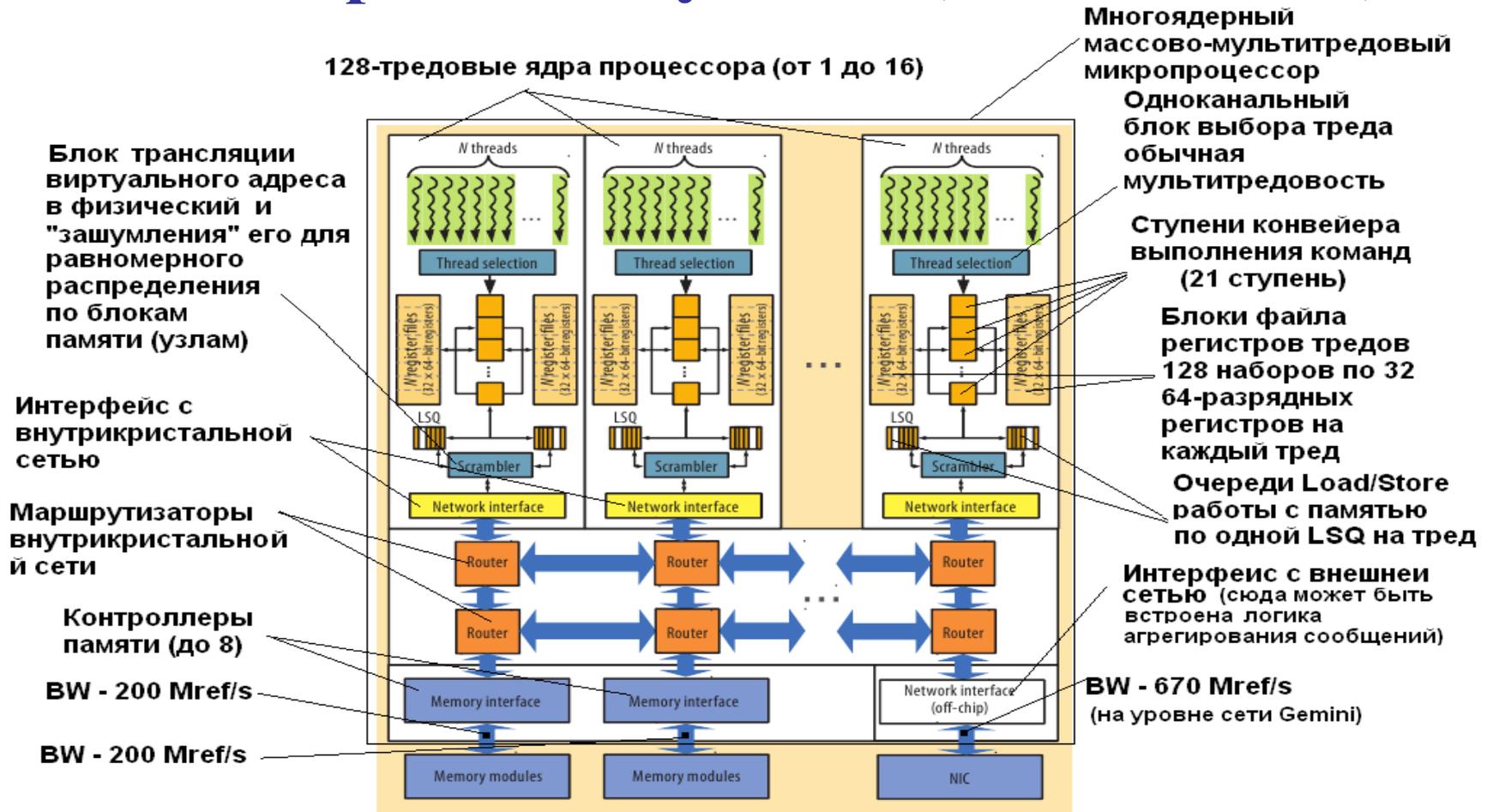
Простой обход списков, 2 x E5 2660, icss (165150720 элементов на поток)



Текущий уровень интерфейсов Threadstorm



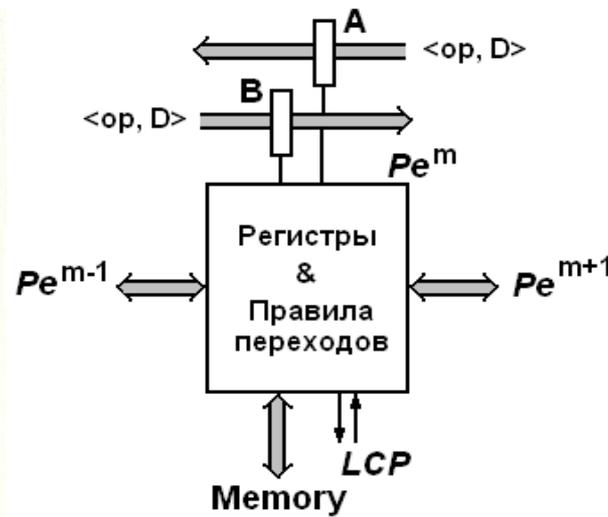
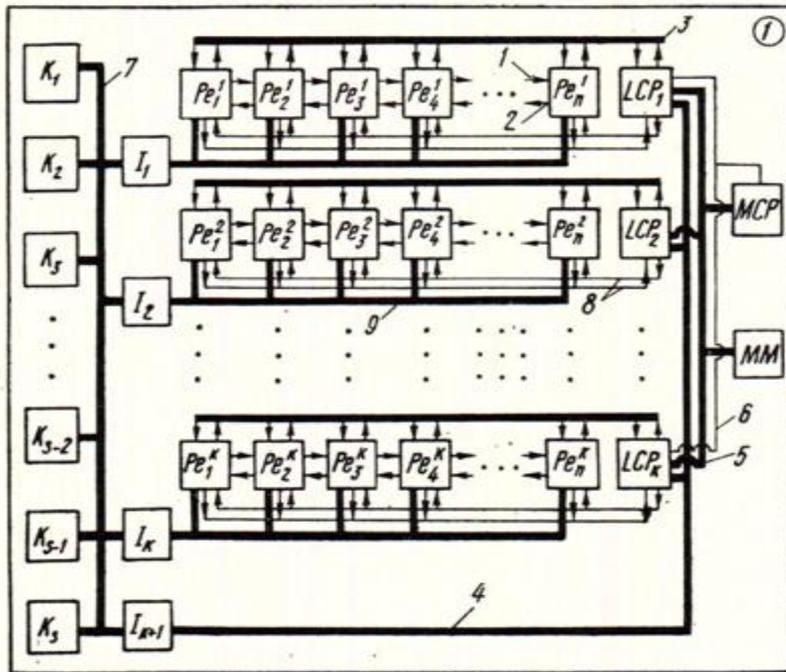
Имитационное моделирование многоядерного Cray XMT (Threadstorm)



Tumeo S. *et al.* Designing Next-Generation Massively Multithreaded Architectures for Irregular Applications. *COMPUTER*, August 2012, pp.53-61.

Villa O., Tumeo A., Secchi S., Manzano J.B., Fast and Accurate Simulation of the Cray XMT Multithreaded Supercomputer, *IEEE Transactions on Parallel and Distributed Systems*, 13 Feb. 2012. 9 pp.

Восток-2: FSC, функциональный суперкомпьютер



Регистры Pe

	Биты
Γ_g	64
Γ_b	$3 \times 64 + 8$
Γ_{fi}	$64 + 8$
Γ_a	64
Γ_1	64
Γ_2	64
n_0	8
n_1	8
n_2	8
q	6
s_m	4
W	16
G	3

Операции Pe

$K^m - 4$ $K^{RRI} - 17$ $K^{PL} - 3$
 $K^W - 15$ $K^{RRO} - 22$ $K^{AB} - 14$
 $K^{ZTR} - 8$

Правила переходов Pe

$$\Psi (S_{i-1}(t), S_i(t), S_{i+1}(t), W, G) \rightarrow [(K_{i1}(t))(K_{i2}(t))] S_i(t+1)$$

Правила проектирования – 206

Правила замены - 88



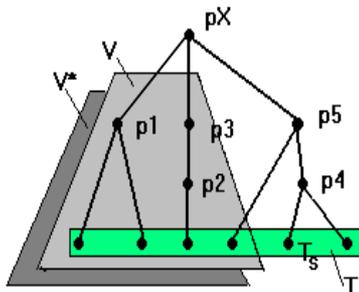
Крупнозернистое распараллеливание программ – одновременное выполнение функций.

Программа

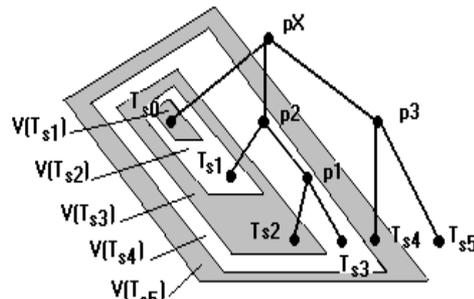
$$\begin{aligned}
 & \S 1 kfe_1 + e_2 > kfe_{1 _}, kfe_{2 _}, + \\
 & \S 2 kfe_1 * e_2 > kfe_{1 _}, kfe_{2 _}, * \\
 & \S 3 kfe_1 > e_1
 \end{aligned}
 \tag{1}$$

Последовательное выполнение

- шаг 1. $kfA * B + C * D _$
- шаг 2. $kfA * B _, kfC * D _, +$
- шаг 3. $kfA _, kfB _, *, kfC * D _, +$
- шаг 4. $A, kfB _, *, kfC * D _, +$ (2)
- шаг 5. $A, B, *, kfC * D _, +$
- шаг 6. $A, B, *, kfC _, kfd _, *, +$
- шаг 7. $A, B, *, C, kfd _, *, +$
- шаг 8. $A, B, *, C, D, *, +$



Дерево вызовов функций (конкретизаций).



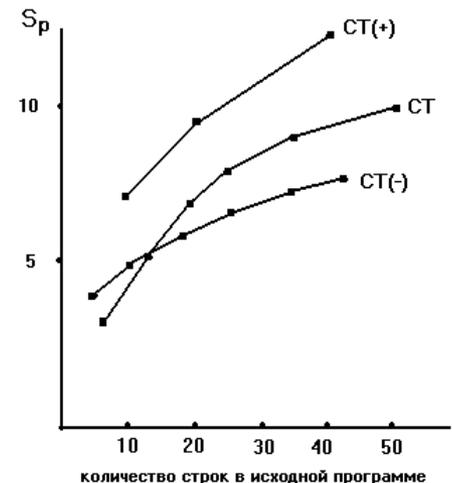
Информационные зависимости на дереве конкретизаций.

Параллельное выполнение

- шаг 1. $kfA * B + C * D _$
- шаг 2. $kfA * B _, kfC * D _, +$ (3)
- шаг 3. $kfA _, kfB _, *, kfC _, kfd _, *, +$
- шаг 4. $A, B, *, C, D, *, +$

```

Тест CT(-)  real A,B,C
            B = C+A
            ...
            B = C+A
            end
Тест CT      real A,B,C
            A = B + C/(C+2.3-B)
            B = C+A
            ...
            A = B + C/(C+2.3-B)
            B = C+A
            end
Тест CT(+)  real A,B,C
            A = B + C/(C+2.3-B)
            ...
            A = B + C/(C+2.3-B)
            end
    
```



Мелкозернистое распараллеливание проектирования и замены.

Пример левой части

$$\S 1kfe_1 + e_2(w_3A (**) e_4BC) > . . .$$

Левая часть с расставленными номерами шагов проектирования

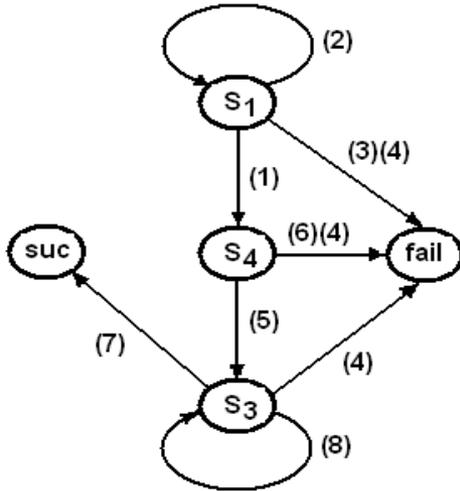
$$\S 1 \overset{1}{k} \overset{3}{f} \overset{15}{e_1} + e_2 (\overset{17}{w_3} \overset{5}{A} (\overset{6}{*} \overset{7}{*}) \overset{8}{e_4} \overset{10}{B} \overset{11}{C}) > . . .$$

Возможное совмещение проектирования

такт 1.	k	$>$
такт 2.	kf	$) >$
такт 3.	kfe_1	$(C) >$
такт 4.	$kfe_1 + (w_3$	$BC) >$
такт 5.	$kfe_1 + e_2(w_3A$	$BC) >$
такт 6.	$kfe_1 + e_2(w_3A ($	$BC) >$
такт 7.	$kfe_1 + e_2(w_3A (*)$	$BC) >$
такт 8.	$kfe_1 + e_2(w_3A (**)e_4$	$BC) >$

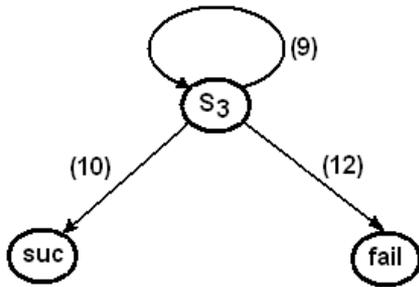
На следующем слайде – процесс реального проектирования для обращения с аргументом $BC+M(())A(**)MCPBC$). Получаем: $e1 = BC$, $e2=M$, $w3=()$, $e4=MCP$

δ1 - Правила проектирования для $Pe(k)$



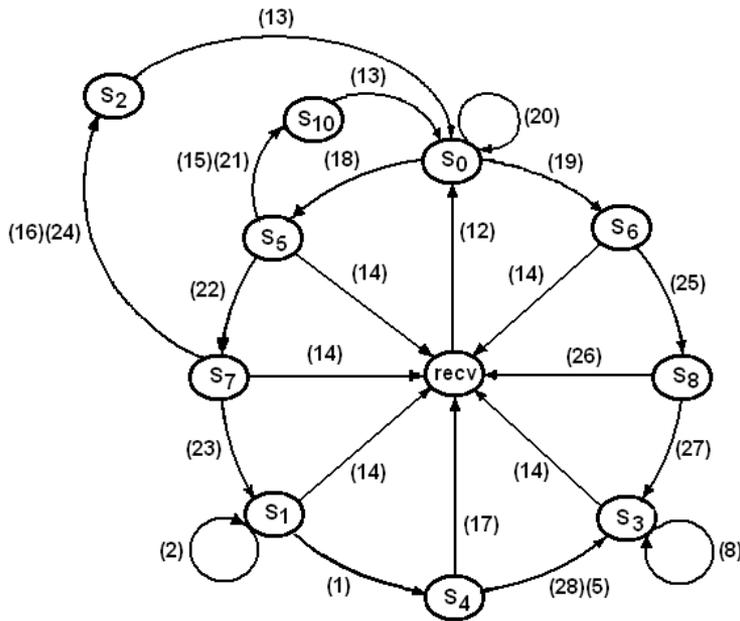
- (1) $(S^{m-1} \in S) \ \& \ (S^m = s_1) \ \& \ (S^{m+1} \in \{s_1, s_3, s_4\}) \ \& \ (\sim w_{10}) \rightarrow [(K_1^{RRO}) (K_1^W)] s_4$
- (2) $(S^{m-1} \in S) \ \& \ (S^m = s_1) \ \& \ (S^{m+1} \in \{s_1, s_2, s_3, s_4\}) \ \& \ (\sim w_{10}) \rightarrow$
- (3) $(S^{m-1} \in S) \ \& \ (S^m = s_1) \ \& \ (S^{m+1} = s_2) \ \& \ (\sim w_{10}) \rightarrow [() (K_1^{PL})]$
- (4) $(S^{m-1} \in S) \ \& \ (S^m = S) \ \& \ (S^{m+1} = S) \ \& \ w_{10} \rightarrow [() (K_1^{PL})]$
- (5) $(S^{m-1} \in S) \ \& \ (S^m = s_4) \ \& \ (S^{m+1} = S) \ \& \ (\sim w_{10} \ \& \ w_1) \rightarrow [(K_1^{ZTR}) ()] s_3$
- (6) $(S^{m-1} \in S) \ \& \ (S^m = s_4) \ \& \ (S^{m+1} \in S) \ \& \ (\sim w_{10} \ \& \ \sim w_1) \rightarrow [() (K_1^{PL})]$
- (7) $(S^{m-1} \in S) \ \& \ (S^m = s_3) \ \& \ (S^{m+1} \in S) \ \& \ (\sim w_{10} \ \& \ w_{11}) \rightarrow [() (K_2^{PL})]$
- (8) $(S^{m-1} \in S) \ \& \ (S^m = s_3) \ \& \ (S^{m+1} \in S) \ \& \ (\sim w_{10} \ \& \ \sim w_{11}) \rightarrow$

δ2 - Правила проектирования для $Pe(\sim)$



- (9) $(S^{m-1} \in S \setminus \{s_3\}) \ \& \ (S^m = s_3) \ \& \ (S^{m+1} \in S) \ \& \ (\sim w_{10}) \rightarrow$
- (10) $(S^{m-1} = s_3) \ \& \ (S^m = s_3) \ \& \ (S^{m+1} \in S) \ \& \ (\sim w_{10}) \rightarrow [() (K_1^{AB})]$
- (12) $(S^{m-1} \in S) \ \& \ (S^m \in S \setminus \{s_3, s_1\}) \ \& \ (S^{m+1} \in S \setminus \{s_0, s_1\}) \ \& \ (w_{10}) \rightarrow [(K_2^{ZTR}) ()] s_0$

б3 - Правила проектирования для $Pe(\varphi)$



$$(S^{m-1} \in S) \ \& \ (S^m = s_1) \ \& \ (S^{m+1} \in \{s_1, s_3, s_4\}) \ \& \ (\sim w_{10}) \ \rightarrow \ [(K_1^{RRO}) (K_1^W)] \ s_4 \quad (1)$$

$$(S^{m-1} \in S) \ \& \ (S^m = s_1) \ \& \ (S^{m+1} \in \{s_1, s_2, s_3, s_4\}) \ \& \ (\sim w_{10}) \ \rightarrow \quad (2)$$

$$(S^{m-1} \in S) \ \& \ (S^m = s_4) \ \& \ (S^{m+1} = S) \ \& \ (\sim w_{10} \ \& \ w_1) \ \rightarrow \ [(K_1^{ZTR}) \ 0] \ s_3 \quad (5)$$

$$(S^{m-1} \in S) \ \& \ (S^m = s_3) \ \& \ (S^{m+1} \in S) \ \& \ (\sim w_{10} \ \& \ \sim w_{11}) \ \rightarrow \quad (8)$$

$$(S^{m-1} \in S) \ \& \ (S^m \in S) \ \& \ (S^{m+1} \in S) \ \& \ (w_{10}) \ \rightarrow \ [(K_2^{ZTR}) \ 0] \ s_0 \quad (12)$$

$$(S^{m-1} \in S) \ \& \ (S^m \in \{s_2, s_{10}\}) \ \& \ (S^{m+1} \in S \setminus \{s_2, s_{10}\}) \ \& \ (\sim w_{10}) \ \rightarrow \ [(K_2^{ZTR}) \ 0] \ s_0 \quad (13)$$

$$(S^{m-1} \in S) \ \& \ (S^m \in S \setminus \{s_0, s_2, s_{10}\}) \ \& \ (S^{m+1} = s_2) \ \& \ (\sim w_{10}) \ \rightarrow \ [0 \ (K_2^{AB})] \quad (14)$$

$$(S^{m-1} = s_{10}) \ \& \ (S^m = s_5) \ \& \ (S^{m+1} \in S \setminus \{s_2, s_{10}\}) \ \& \ (\sim w_{10}) \ \rightarrow \ s_{10} \quad (15)$$

$$(S^{m-1} = s_2) \ \& \ (S^m = s_7) \ \& \ (S^{m+1} \in S \setminus \{s_2, s_{10}\}) \ \& \ (\sim w_{10}) \ \rightarrow \ s_2 \quad (16)$$

$$(S^{m-1} \in S) \ \& \ (S^m = s_4) \ \& \ (S^{m+1} \in S) \ \& \ (\sim w_{10} \ \& \ \sim w_1) \ \rightarrow \ [0 \ (K_2^{AB})] \quad (17)$$

$$(S^{m-1} \in S \setminus \{s_0, s_2, s_9, s_{10}\}) \ \& \ (S^m = s_0) \ \& \ (S^{m+1} \in S) \ \& \ (\sim w_{10}) \ \rightarrow \ [(K_2^{RRO}, K_3^{RRO}) (K_1^M, K_4^W)] \ s_5 \quad (18)$$

$$(S^{m-1} \in \{s_0, s_9\}) \ \& \ (S^m = s_0) \ \& \ (S^{m+1} \in S \setminus \{s_0, s_2, s_9, s_{10}\}) \ \& \ (\sim w_{10}) \ \rightarrow \ [(K_4^{RRO}) (K_1^M)] \ s_6 \quad (19)$$

$$(S^{m-1} \in \{s_0, s_9\}) \ \& \ (S^m = s_0) \ \& \ (S^{m+1} \in \{s_0, s_2, s_9, s_{10}\}) \ \& \ (\sim w_{10}) \ \rightarrow \quad (20)$$

$$(S^{m-1} \in S \setminus \{s_2, s_{10}\}) \ \& \ (S^m = s_5) \ \& \ (S^{m+1} \in S \setminus \{s_2, s_{10}\}) \ \& \ (\sim w_{10} \ \& \ \sim w_4) \ \rightarrow \ s_{10} \quad (21)$$

$$(S^{m-1} \in S \setminus \{s_2, s_{10}\}) \ \& \ (S^m = s_5) \ \& \ (S^{m+1} \in S \setminus \{s_2, s_{10}\}) \ \& \ (\sim w_{10} \ \& \ w_4) \ \rightarrow \ [(K_2^W, K_3^W) \ 0] \ s_7 \quad (22)$$

$$(S^{m-1} \in S \setminus \{s_2, s_{10}\}) \ \& \ (S^m = s_7) \ \& \ (S^{m+1} \in S \setminus \{s_2, s_{10}\}) \ \& \ (\sim w_{10} \ \& \ \sim w_3 \ \& \ w_2) \ \rightarrow \ [(K_1^{ZTR}) \ 0] \ s_1 \quad (23)$$

$$(S^{m-1} \in S \setminus \{s_2, s_{10}\}) \ \& \ (S^m = s_7) \ \& \ (S^{m+1} \in S \setminus \{s_2, s_{10}\}) \ \& \ (\sim w_{10} \ \& \ (w_3 \vee \sim w_2)) \ \rightarrow \ s_2 \quad (24)$$

$$(S^{m-1} \in S) \ \& \ (S^m = s_6) \ \& \ (S^{m+1} \in S \setminus \{s_2, s_{10}\}) \ \& \ (\sim w_{10}) \ \rightarrow \ [(K_2^W, K_3^W) \ 0] \ s_8 \quad (25)$$

$$(S^{m-1} \in S) \ \& \ (S^m = s_8) \ \& \ (S^{m+1} \in S \setminus \{s_2, s_{10}\}) \ \& \ (\sim w_{10} \ \& \ (\sim w_2 \vee w_3)) \ \rightarrow \ [0 \ (K_2^{AB})] \quad (26)$$

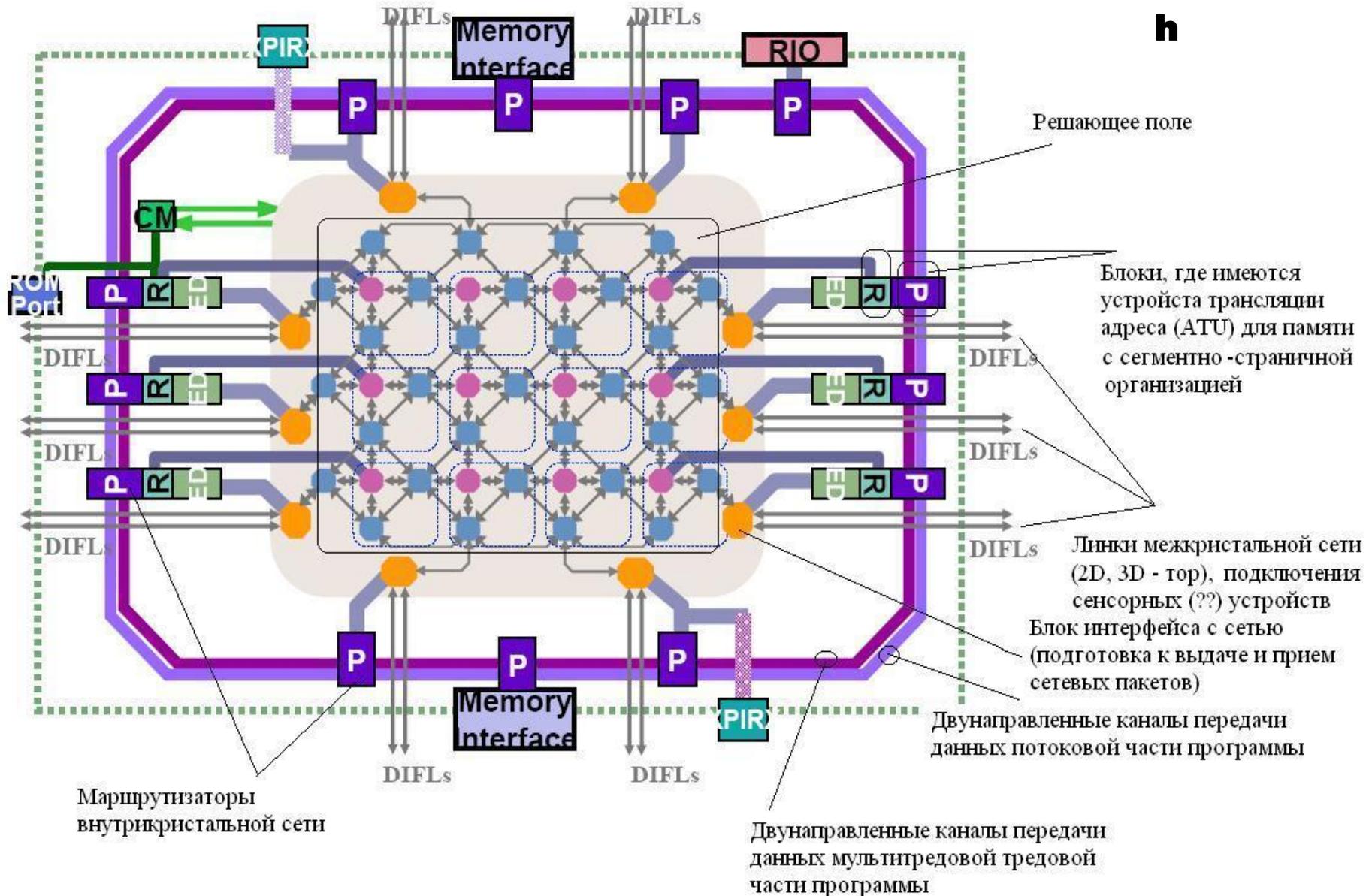
$$(S^{m-1} \in S) \ \& \ (S^m = s_8) \ \& \ (S^{m+1} \in S \setminus \{s_2, s_{10}\}) \ \& \ (\sim w_{10} \ \& \ (w_2 \ \& \ \sim w_3)) \ \rightarrow \ [(K_1^{ZTR}) \ 0] \ s_3 \quad (27)$$

$$(S^{m-1} \in S) \ \& \ (S^m = s_4) \ \& \ (S^{m+1} = s_0) \ \& \ (\sim w_{10}) \ \rightarrow \ [(K_1^{ZTR}) \ 0] \ s_3 \quad (28)$$

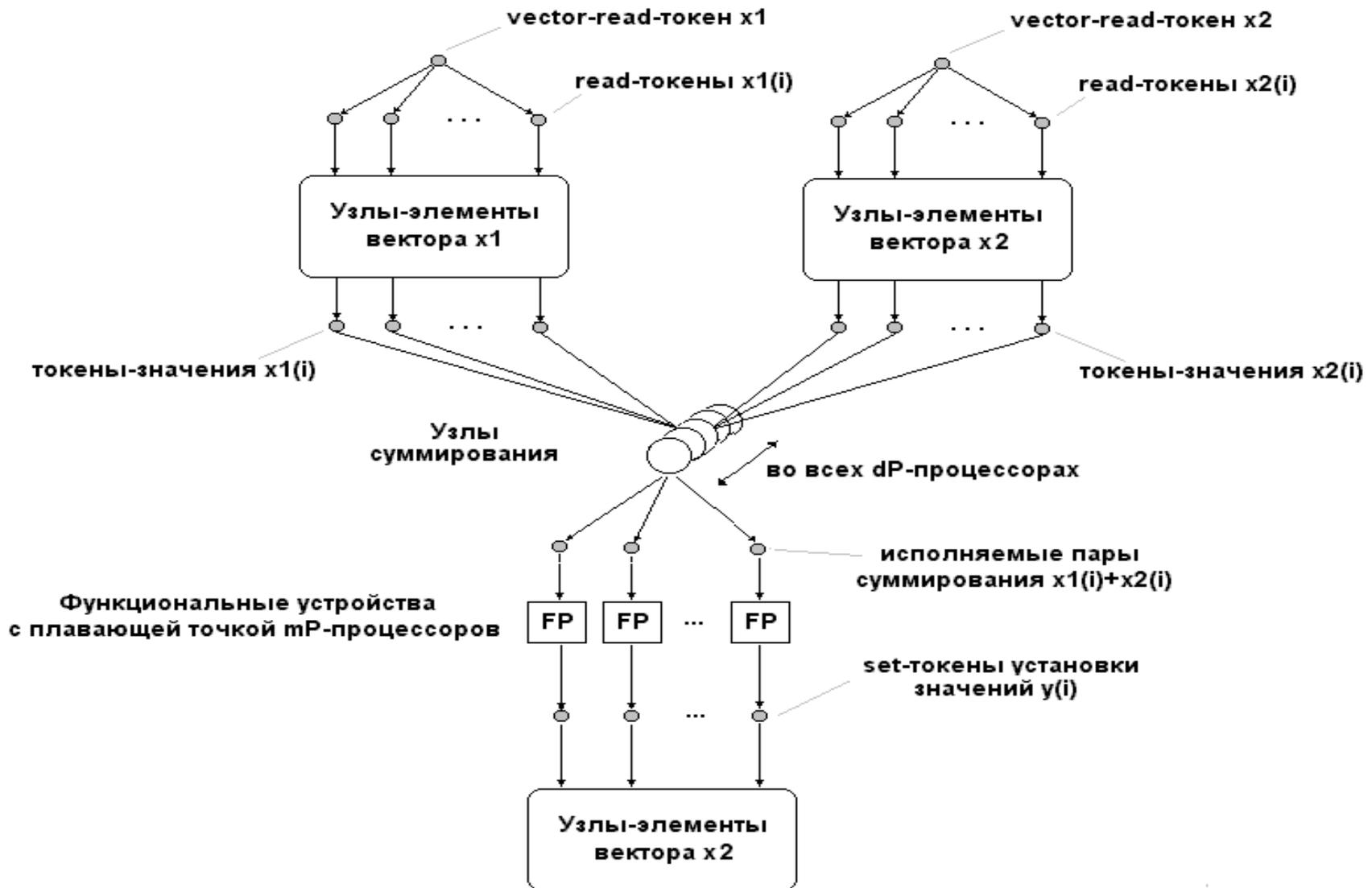
Разработка суперкомпьютерных блоков с потоковой архитектурой (SDF и DDF)

Восток-2 SDF, близкие работы

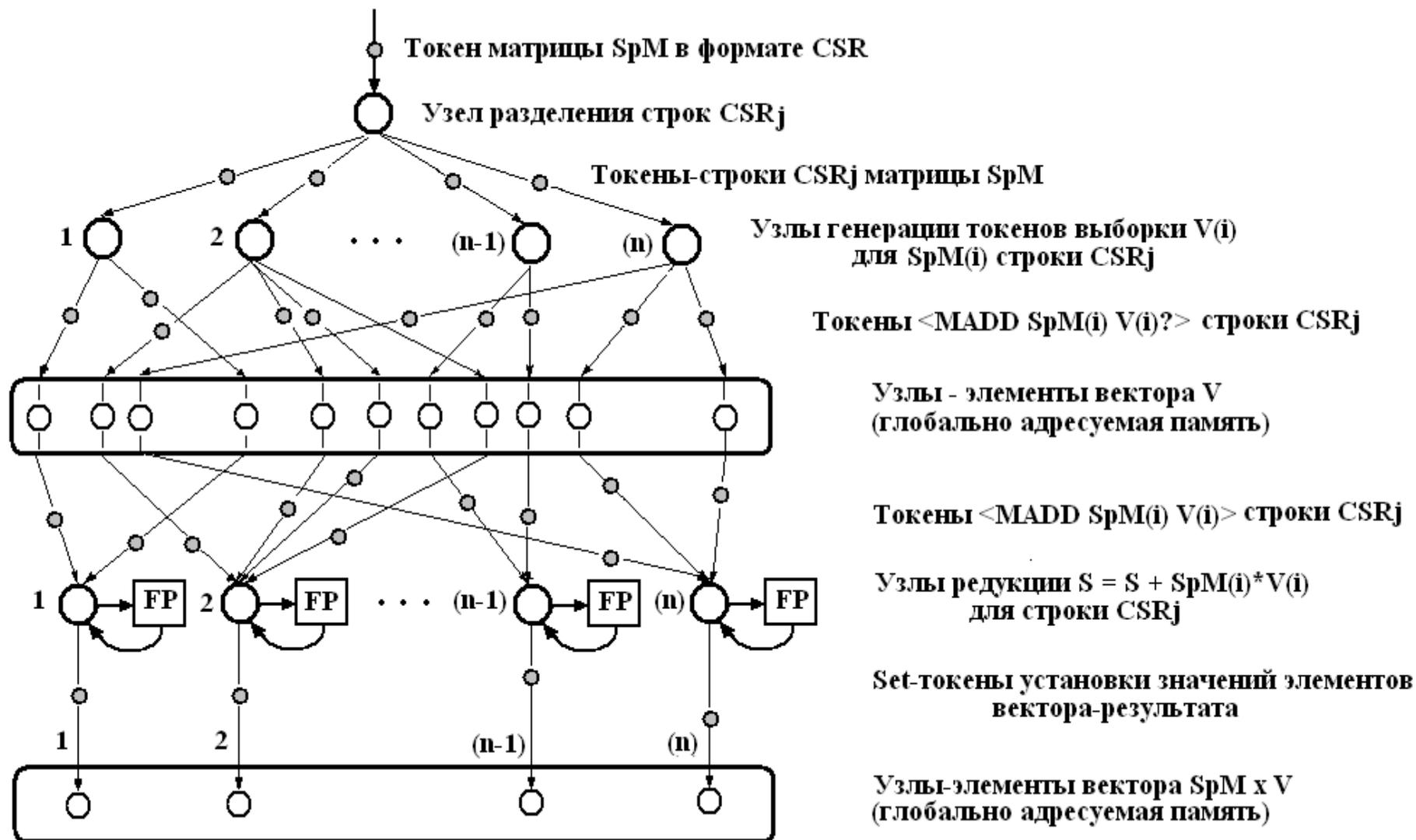
~ 2006 Monarc h



Восток-2: DDF, Схема выполнения векторного сложения $y(i) = x1(i) + x2(i)$



Восток-2: DDF - умножение разреженной матрицы на вектор



Вопросы ?

Эйсымонт Леонид Константинович

(ФГУП "НИИ"Квант", verger-lk@yandex.ru, eisymont@rdi-kvant.ru)