

Комплексная методика тестирования производительности суперкомпьютеров. Профессиональный подход.

В.С. Горбунов, Л.К. Эйсымонт

Комплексная методика тестирования производительности суперкомпьютеров. Профессиональный подход.

В. Горбунов, Л. Эйсымонт

HPEC Challenge Green-500 ЦТСС Graph-500 HPC Challenge TOP-500

Абрау-Дюрсо, 26 сентября 2013

Динамика роста Top500

Тор500		Место в рейтинге (Тflops)				
год	месяц	1	2	3	4	5
2010	июнь	1759.0	1271.0	1042.0	831.7	825.5
		ХТ5	Nebulae	Roadrunner	ХТ5	BG/P
	ноябрь	2566.0	1759.0	1271.0	1192.0	1054.0
		Tianhe-1A	ХТ5	Nebulae	Tsubame	ХЕ6
2011	июнь	8162.0	2566.0	1759.0	1271.0	1192.0
		K-comp	Tianhe-1A	ХТ5	Nebulae	Tsubame
	ноябрь	10510.0	2566.0	1759.0	1271.0	1192.0
		K-comp	Tianhe-1A	ХТ5	Nebulae	Tsubame
2012	июнь	16324.8	10510.0	8162.0	2897.0	2566.0
		BG/Q	K-comp	BG/Q	IBM-cluster	Tianhe-1A
	ноябрь	17590.0	16324.8	10510.0	8162.4	4141.2
		ХК7	BG/Q	K-comp	BG/Q	BG/Q
Динамика 2012/2005 (Nov, разы)		62.69	178.8	165.78	133.81	108.13
Динамика 2012/2010 (Nov, разы)		6.86	9.28	8.27	6.85	3.93

Динамика роста Graph500

Graph500		Место в рейтинге (GTEPS(scale))				
год	месяц	1	2	3	4	5
2010	ноябрь	6.6 (36)	5.22 (32)	1.22 (29)	1.17 (29)	0.533 (29)
		BG/P	XT4	XMT	XMT	X5670+IB
2011	июнь	43.47 (37)	25.08 (37)	19.96 (36)	18.51 (38)	18.42 (38)
		Lomonosov	XE6	XT4	BG/P	BG/P
	ноябрь	253.4 (32)	112.7 (37)	103.1 (37)	99.86 (36)	92.34 (36)
		BG/Q	XE6	Lomonosov	Tsubame	BG/P
2012	июнь	3541 (38)	3541 (38)	508.1 (35)	358.1 (38)	317.09 (35)
		BG/Q	BG/Q	p7-775	K-comp	HP-cluster
	ноябрь	15383 (40)	10461 (39)	5848 (38)	5524 (40)	2567 (37)
		BG/Q	BG/Q	BG/Q	K-comp	BG/Q
Динамика 2012/2010 (разы)		2330	2004	4793	4721	4816

Динамика роста HPC Challenge

HPCC Awards (Nov)	G-HPL (TFlops)		G-RandomAccess (GUPS)		G-FFT (TFlops)		EP-Stream-Triad (Tbyte/s)	
	1	2	1	2	1	2	1	2
2005	259	-	35	-	2.3	-	160	-
	BG/L	-	BG/L	-	BG/L	-	BG/L	-
2006	259	67	35	17	2.311	1.122	160	55
	BG/L	BG/L	BG/L	BG/L	BG/L	XT3	BG/L	p5-575
2007	259	94	35.5	33.6	2.870	2.311	160	77
	BG/L	XT3	BG/L	XT3	XT3	BG/L	BG/L	XT3
2008	902	259	103	35.5	5.080	2.870	330	160
	XT5	BG/L	BG/P	BG/L	BG/P	XT3	XT5	BG/L
2009	1533	736	117	103	11.0	8.0	398	267
	XT5	XT5	BG/P	BG/P	BG/P	BG/P	XT5	BG/P
2010	1533	736	117	103	11.88	11.7	398	267
	XT5	XT5	BG/P	BG/P	SX-9	XT5	XT5	BG/P
2011	2118	1533	121	117	34.7	11.88	812	398
	K-comp	XT5	K-comp	BG/P	K-comp	SX-9	K-comp	XT5
2012	9796	1533	2021	472	205.9	132.7	3857	525
	K-comp	XT5	p7-775	K-comp	K-comp	p7-775	K-comp	p7-775
Динамика 2012/2006 (разы)	37.8	22.88	57.75	7.62	89.52	118.27	24.01	4.27
Динамика 2012/2010 (разы)	6.4	2.08	17.27	4.58	17.33	11.34	9.69	1.97

Зачем нужны тесты?

Задачи: $A = \{A_1, A_2 \dots A_N\}$

Компьютер S:

Hw – аппаратная часть

SW_{OS} - операционная система

$SW_{RT/LIB}$ - run-time системы и библиотеки

S_{WC} - компиляторы

Схема 1: Узнать как работает существующий компьютер на задачах **Const**(A , Hw , SW_{OS} , $SW_{RT/LIB}$, S_{WC})

Схема 2: Улучшить существующий компьютер на задачах **Const**(A , Hw), **Var**(SW_{OS} , $SW_{RT/LIB}$, S_{WC})

Схема 3: Задачи не сформулированы, узнать какие характеристики и возможности компьютера **Const**(Hw , SW_{OS} , $SW_{RT/LIB}$, S_{WC}), **Var**(A)

Схема 4: Разработка компьютера **Var**(Hw , SW_{OS} , $SW_{RT/LIB}$, S_{WC})



Многоуровневая методика оценочного тестирования

1 тест APEX

Общий тест подсистемы памяти
APEX-map

2 тесты граничных точек
арех-поверхности

G-RandomAccess, G-HPL (High Performance Linpack), G-StreamTriad, G-FFT (Fast Fourier Transform)

3 sysmark - тесты

Специальные системные тесты
компонентов (Euroben, Pallas MPI Benchmark)

4 stressmark - тесты

Тесты базовых алгоритмов для выявления
влияния системных характеристик
DGEMM, SpMV, QR, SVD

5 kernel benchmark
тесты

Тесты – ядра приложений NAS Parallel benchmark (CG, EP, FT, IS, MG)

6 benchmarks
(compact application)

Тесты – компактные приложения
NAS Parallel benchmark (BT, LU, SP, UA)

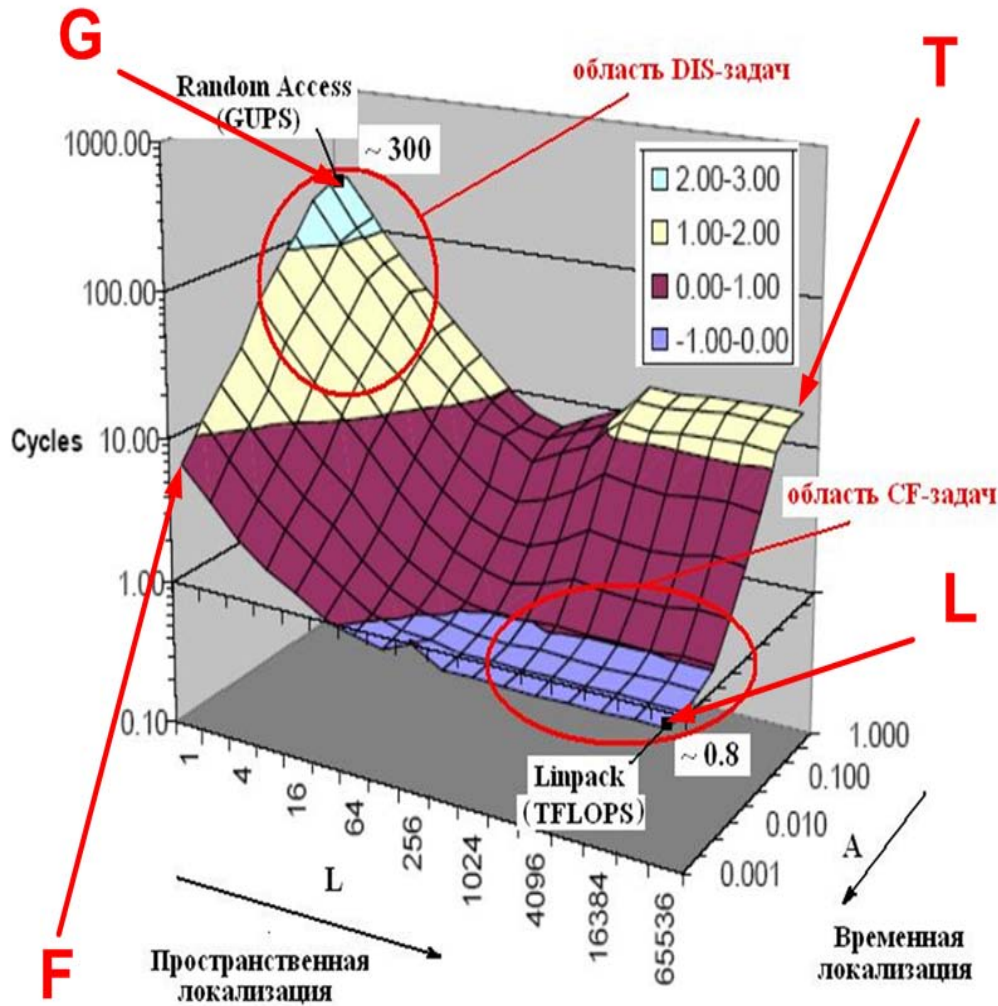
7 application

Тесты приложения

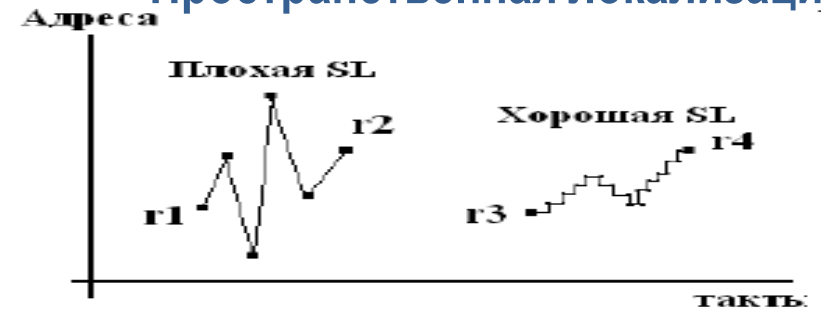
Тест APEX – интегральная оценка эффективности подсистемы памяти

1

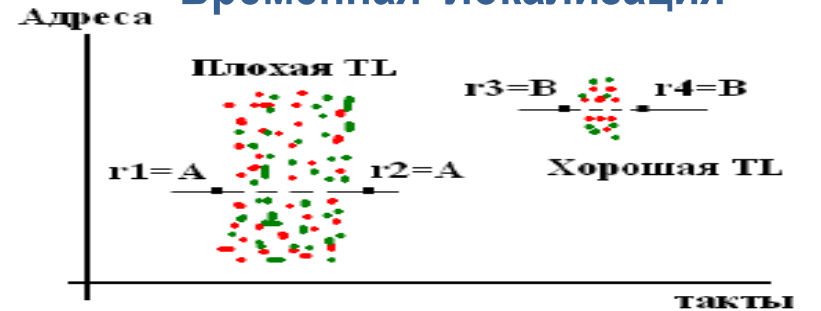
Цель: оценка времени доступа к элементу памяти в зависимости от пространственно-временной локализации



Пространственная локализация



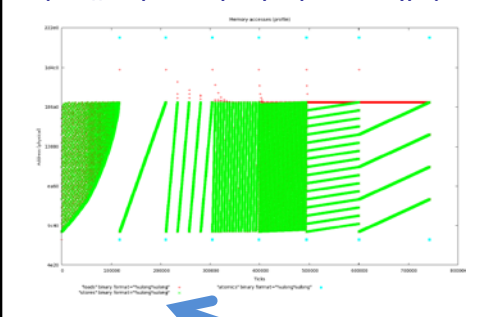
Временная локализация



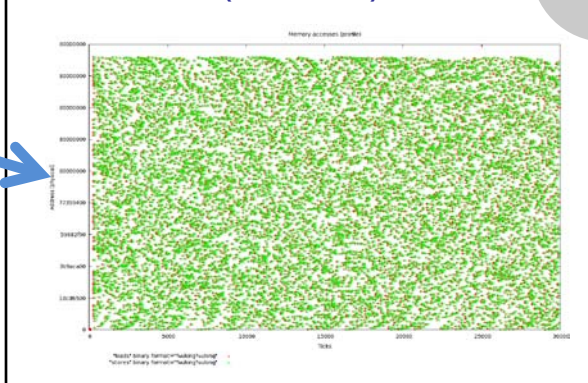
	Временная локализация	Пространственная локализация
L	хорошая	хорошая
F	хорошая	плохая
T	плохая	хорошая
G	плохая	плохая

HPC Challenge

Пример профиля работы с памятью для задачи DIS-класса
(FFT-одномерное быстрое преобразование Фурье)



Пример профиля работы с памятью для задачи DIS-класса
(RandomAccess)

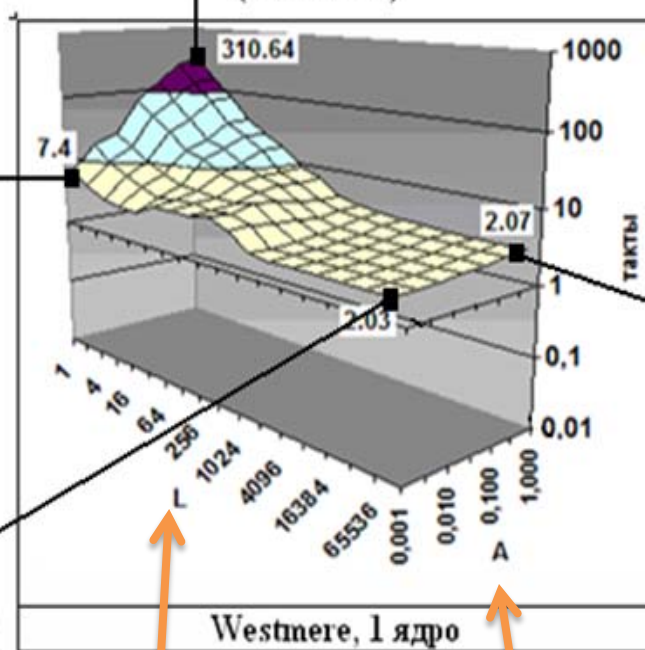


FFT
(точка F)

Random Access
(точка G)

Linpack
(точка L)

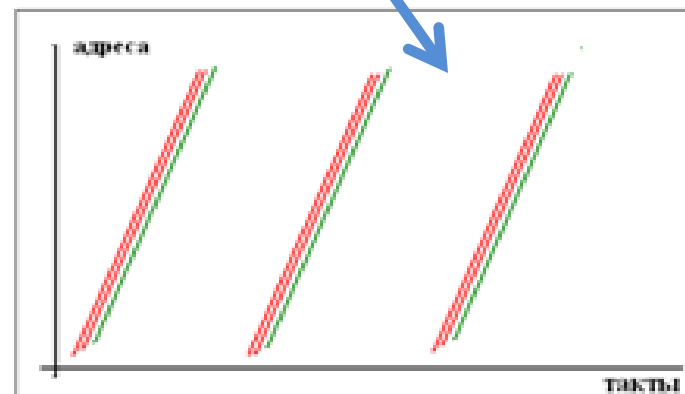
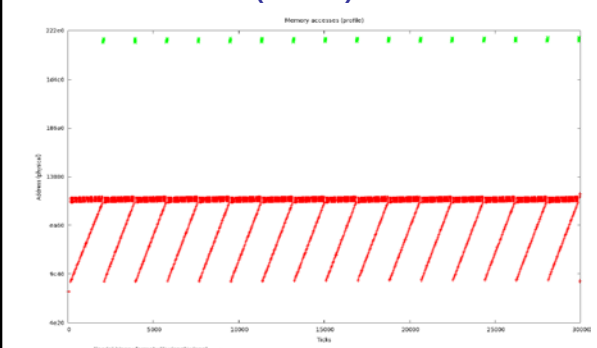
PTRANS,
STREAM
TRIAD
(точка T)



Пространственная
локализация

Временная
локализация

Пример профиля работы с памятью задачи CF-класса
(DGEMM)





Пакет тестов Euroben, группы операций

3

тест Euroben (модуль mod1a)

31 тест простых операций с векторами

Цели: Оценка производительности вычислительных элементов в зависимости от схемы доступа и соотношения числа вычислительных операций и операций доступа к памяти

Сравнение компиляторов и опций

Группа 1 (усложнение доступа к памяти)	Группа 2 (повышение интенсивности вычислений)
$y(i) = x1(i) * x2(i), i = 1, n$	$y(i) = x1(i) + x2(i), i = 1, n$
$y(i) = x1(i) * x2(i), i = 1, 3*n, 3$	$y(i) = x1(i)*x2(i) + x3(i)*x4(i), i = 1, n$
$y(i) = x1(i) * x2(i), i = 1, 4*n, 4$	$y(i) = x1(i+1) - 2*x1(i) + x1(i-1), i = 1, n$
$y(i) = x1(ind(i)) * x2(ind(i)), i = 1, n$	9th DegrPolyn

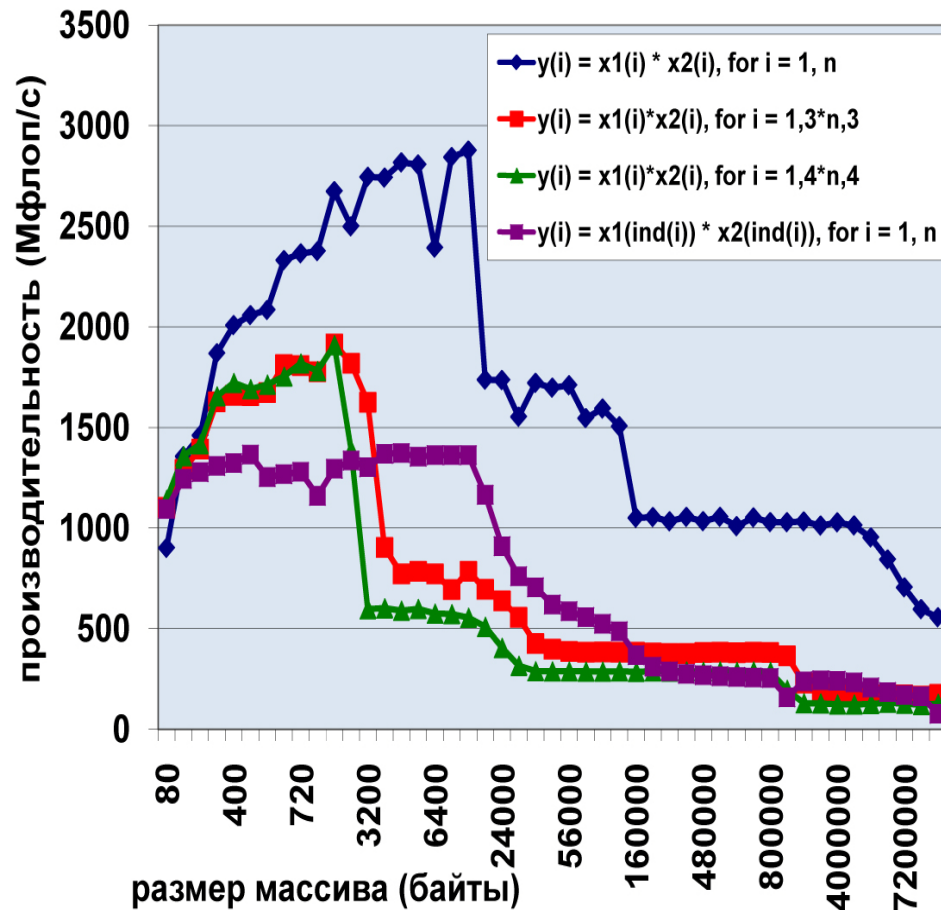
Пакет тестов Euroben

Intel E5-2660 Sandy Bridge

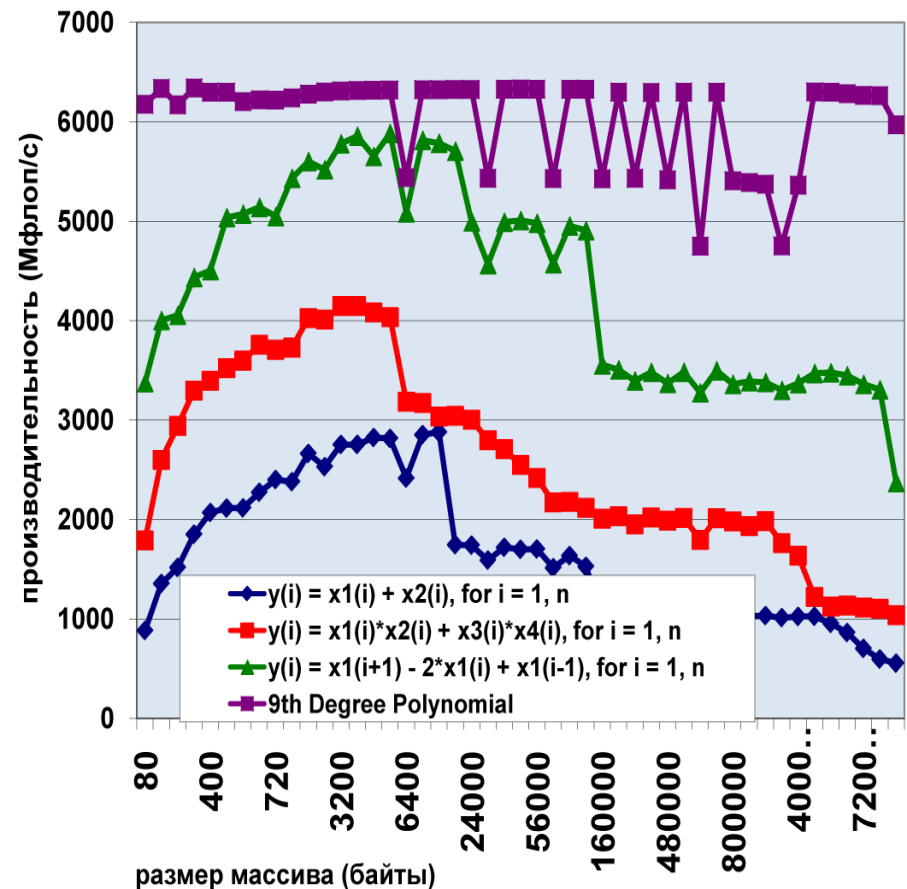
(Rpeak 17,6 Гфлоп/с)

3

Группа 1



Группа 2



Коммуникационная сеть

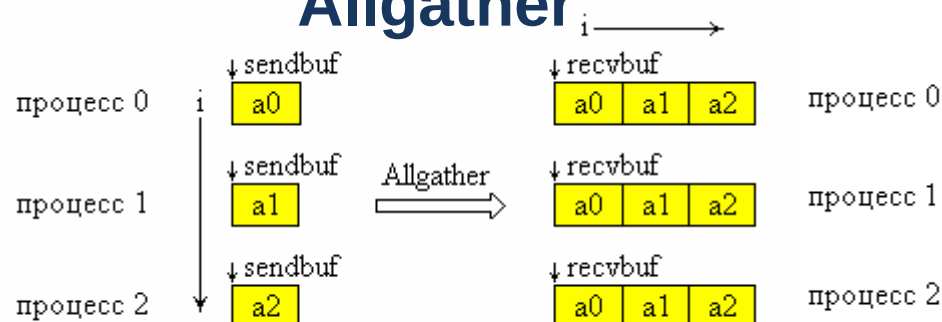
«коллективные операции»

3

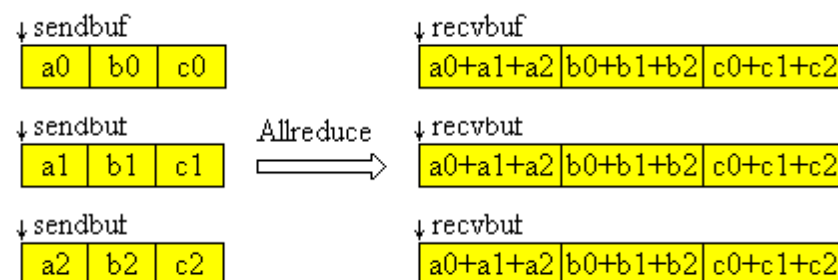
Цели: Оценка эффективности выполнения операций на уровне пользовательской библиотеки MPI

Сравнение библиотек MPI и алгоритмов выполнения коллективных операций

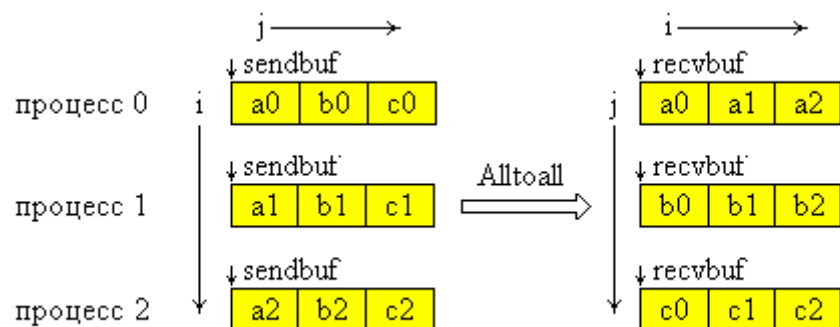
Allgather



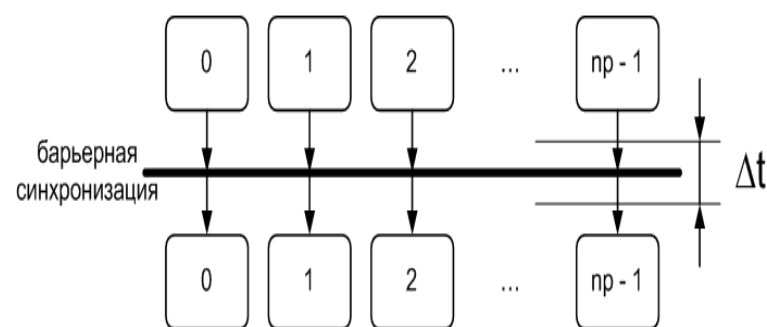
Allreduce



Alltoall



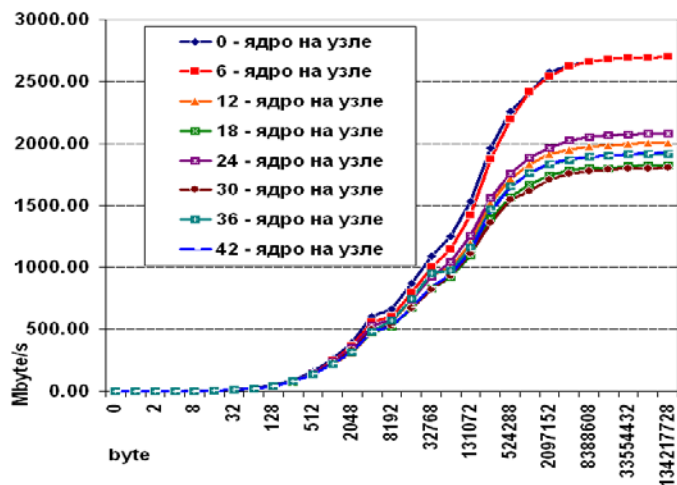
Barrier



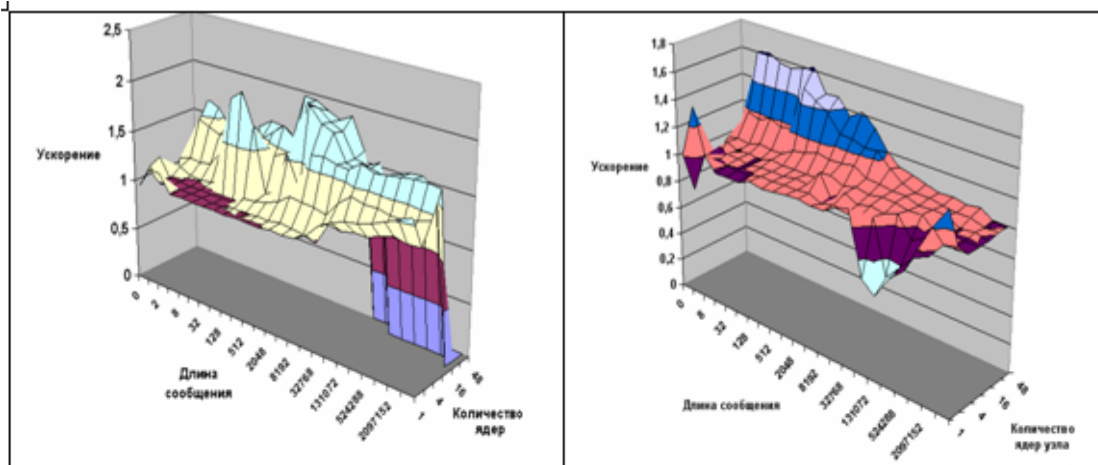
Коммуникационная сеть

3

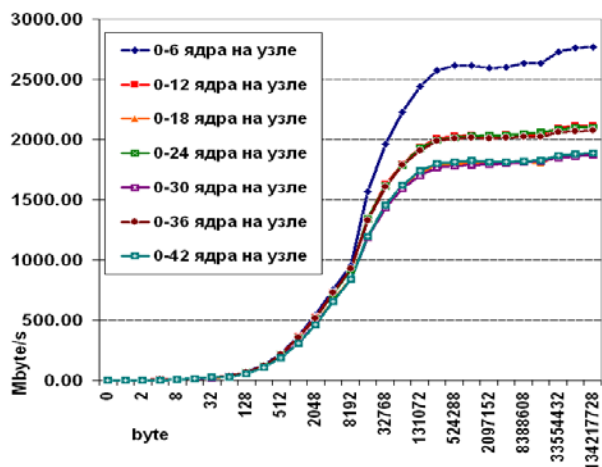
PingPong, межузловой



Ускорение коллективных операций для 2-х адаптеров на узле

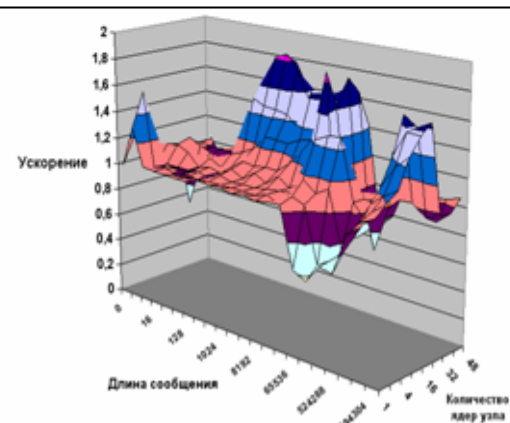
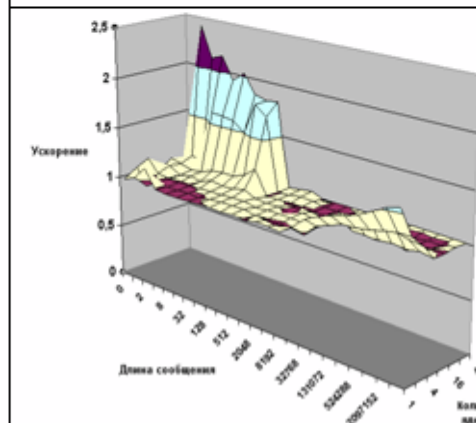


PingPong, внутриузловой



Alltoall

Allreduce



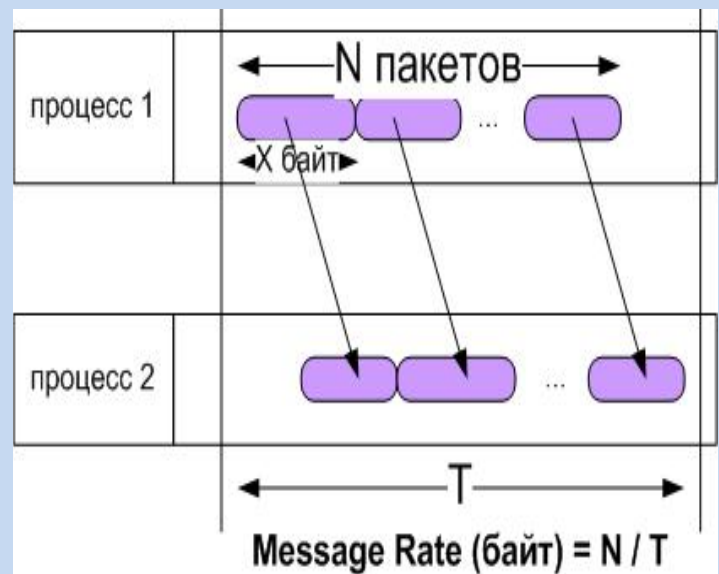
Allgather

ReduceScatter

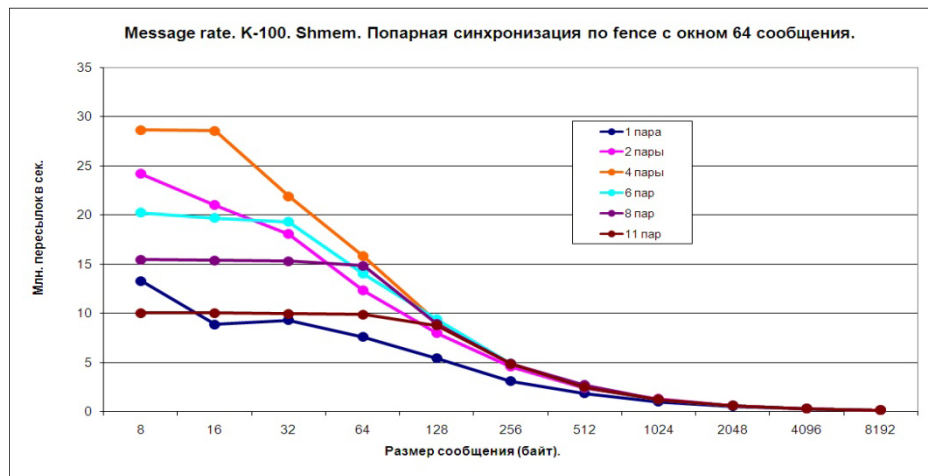
Коммуникационная сеть Message Rate

3

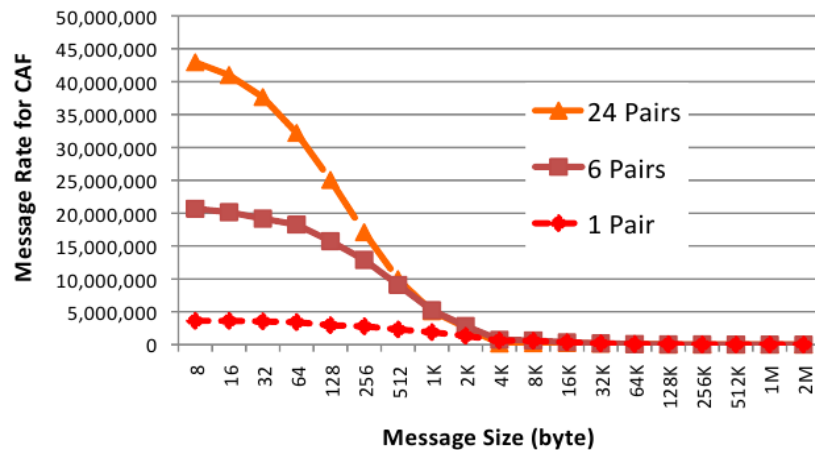
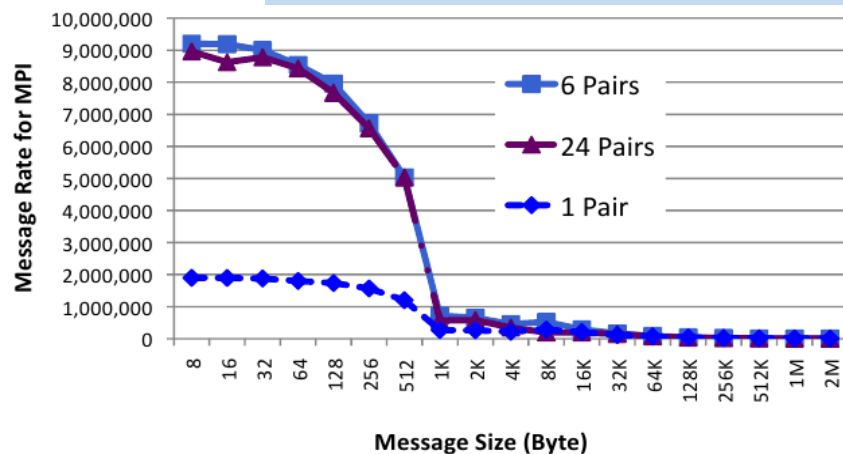
Цель: оценка темпа выдачи сообщений в сеть



Сеть «МВС-Экспресс», gen2



Сеть Cray Gemini

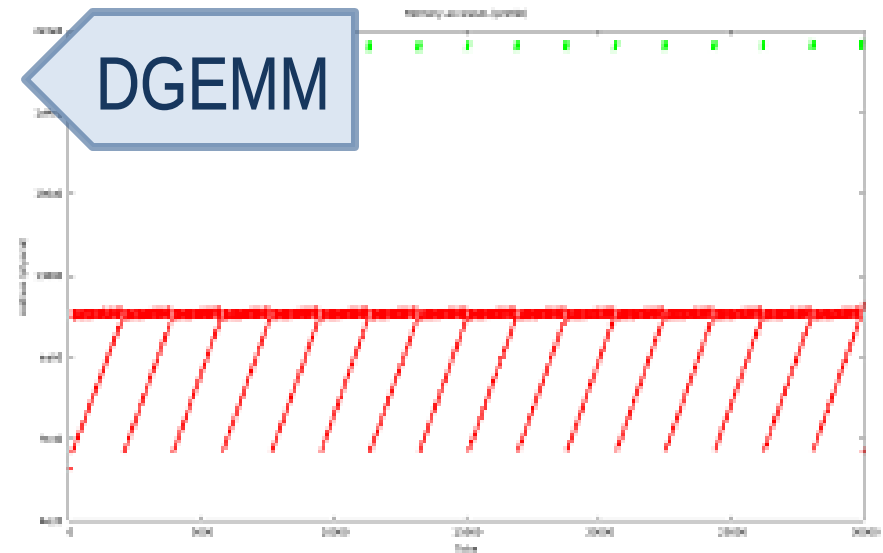
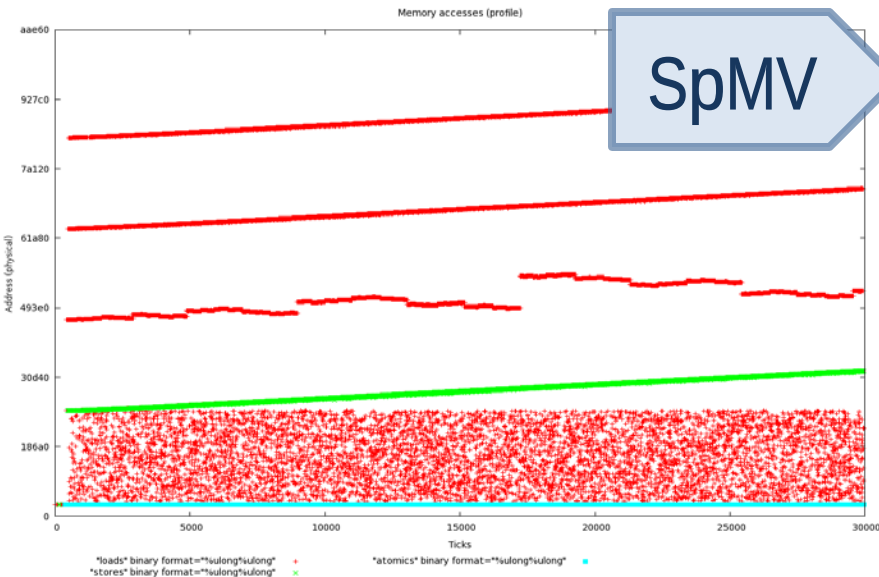


HPCG (SpMV)

vs

HPL

4



установка	реальная производительность (% от пиковой) SpMV
2 x Intel Xeon X5680 (Westmere)	0,7 - 1,3
2 x Intel Xeon E5-2670 (Sandy Bridge)	1,2 - 2,3
NVIDIA Tesla C2050 (Fermi).	1,2 - 5,3
Tesla K20 (Kepler)	0,4 - 1,2
Intel Xeon Phi	0,5 - 4,5

NAS Parallel Benchmarks (NPB)

5

Цели: Оценка производительности на приложениях
Сравнение компиляторов и библиотек MPI
Выбор оптимальных параметров запуска

EP

Embarrassing Parallel

Накопление
двумерной статистики
псевдослучайных
чисел
с распределением
Гаусса.

Позволяет оценить
наибольшую
производительность
на задачах
с интенсивными
вычислениями с
плавающей точкой.

CG

Conjugate Gradient

Метод сопряженных
градиентов нахождения
приближения
наименьшего
собственного значения
разреженной матрицы.

Находится в области
плохой временной
и пространственной
локализации обращения
в память.

FT

Fast-Fourier Transform

Быстрое
преобразование
Фурье для решения
дифференциального
уравнения с частными
производными.

Тяжелый тест на
производительность
при обменах между
процессорами.

IS

Parallel Sort

Сортировка
целых чисел.

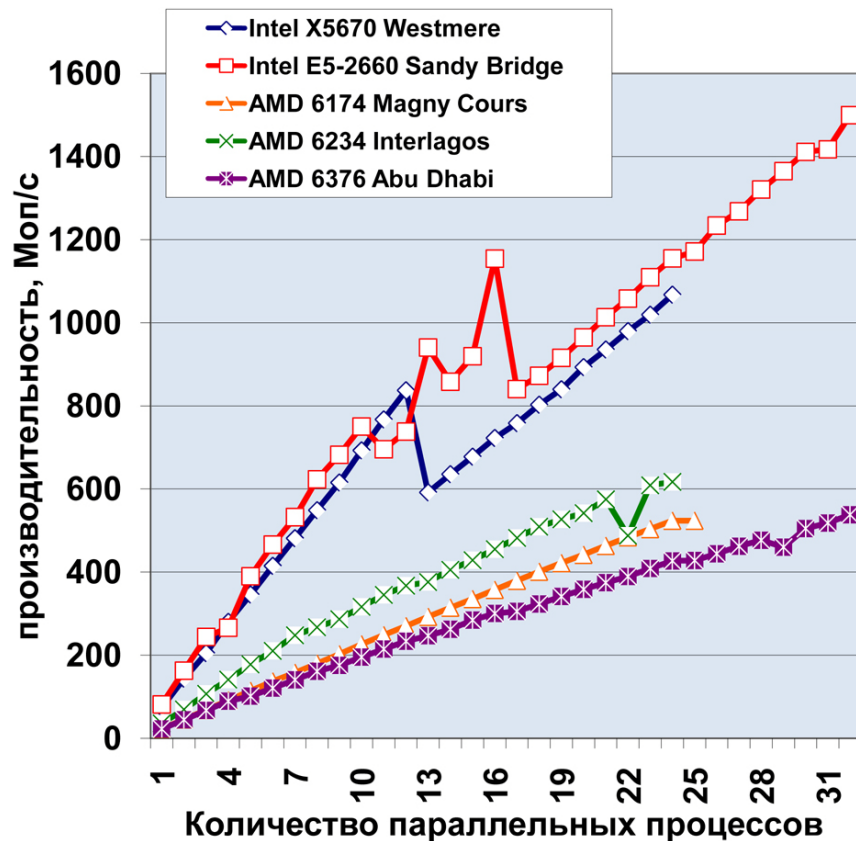
Затраты на
коммуникации в
разы превосходят
затраты на
вычисления.

NPB EP и CG, OpenMP, класс C

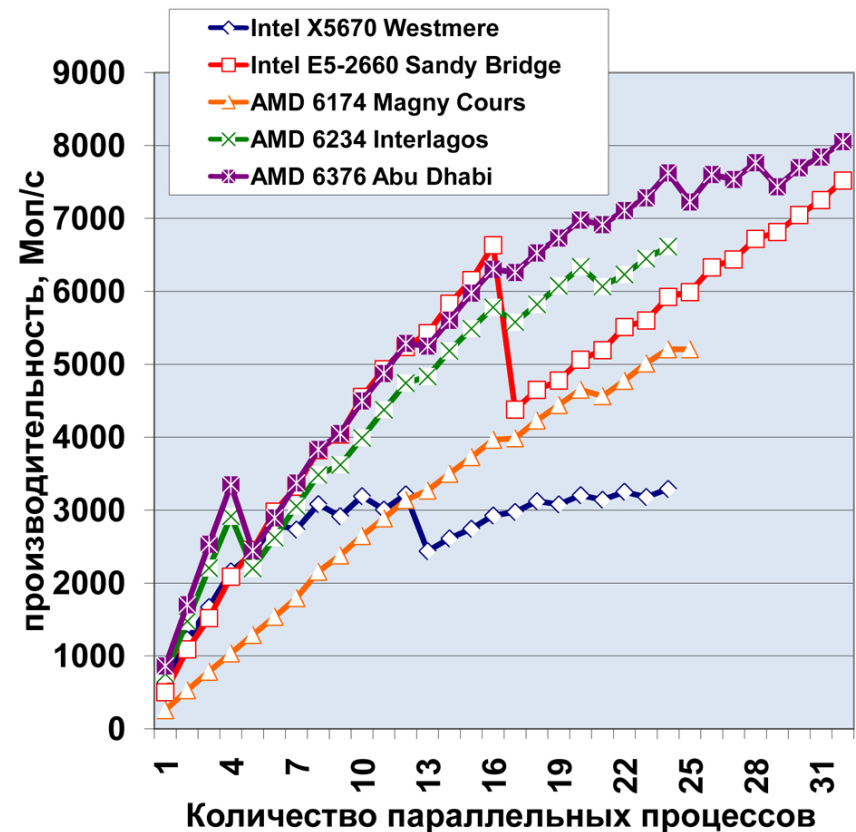
двухсокетные узлы

производительность

NPB EP



NPB CG



NPB UA и EP, OpenMP, класс C, результаты

5

NPB UA

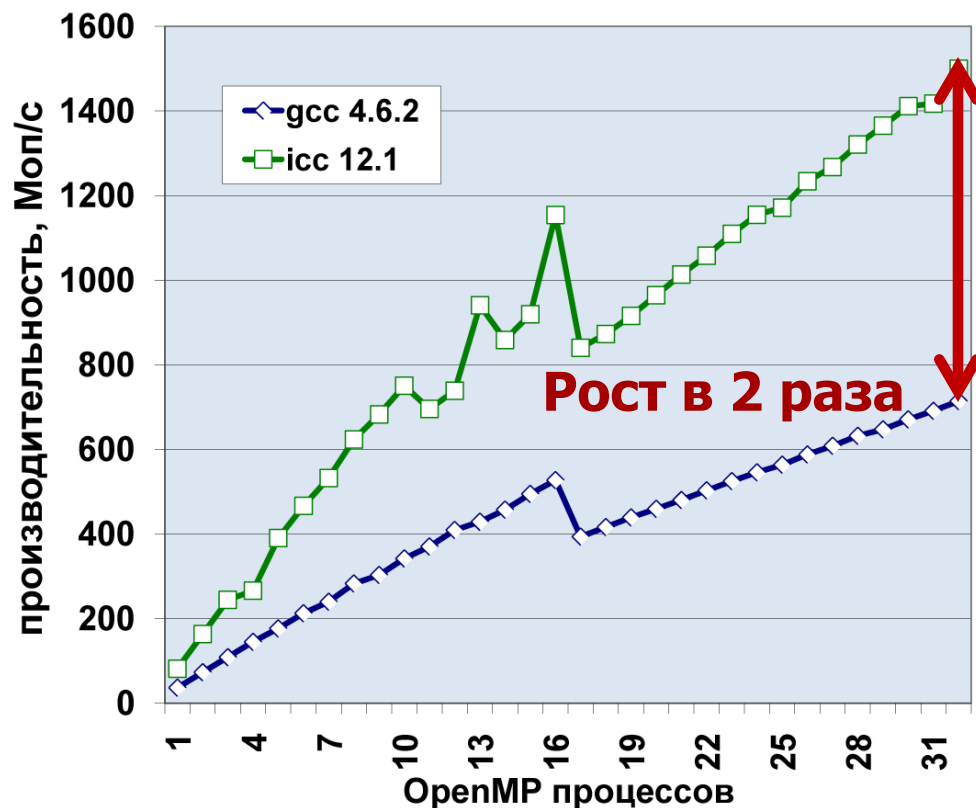
узел 8 x Intel Westmere-EX



привязка и оптимальное
размещение процессов
по ядрам узла

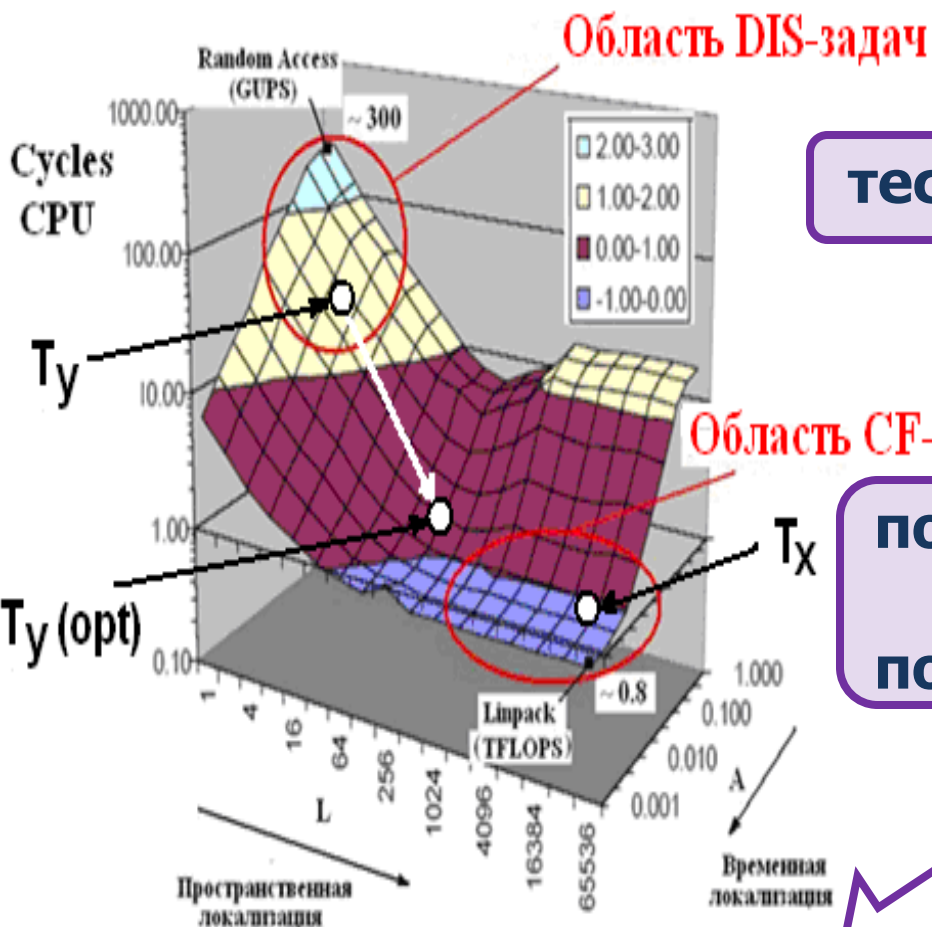
NPB EP

узел 2 x Intel E5-2660 Sandy Bridge



выбор оптимального
компилятора

Технология оптимизации на основе профилирования



тестирование

профилирование,
вычисление
локализаций
(SL, TL)

построение
APEX-
поверхности

отображение
на APEX -
поверхность

определение резервов
оптимизации

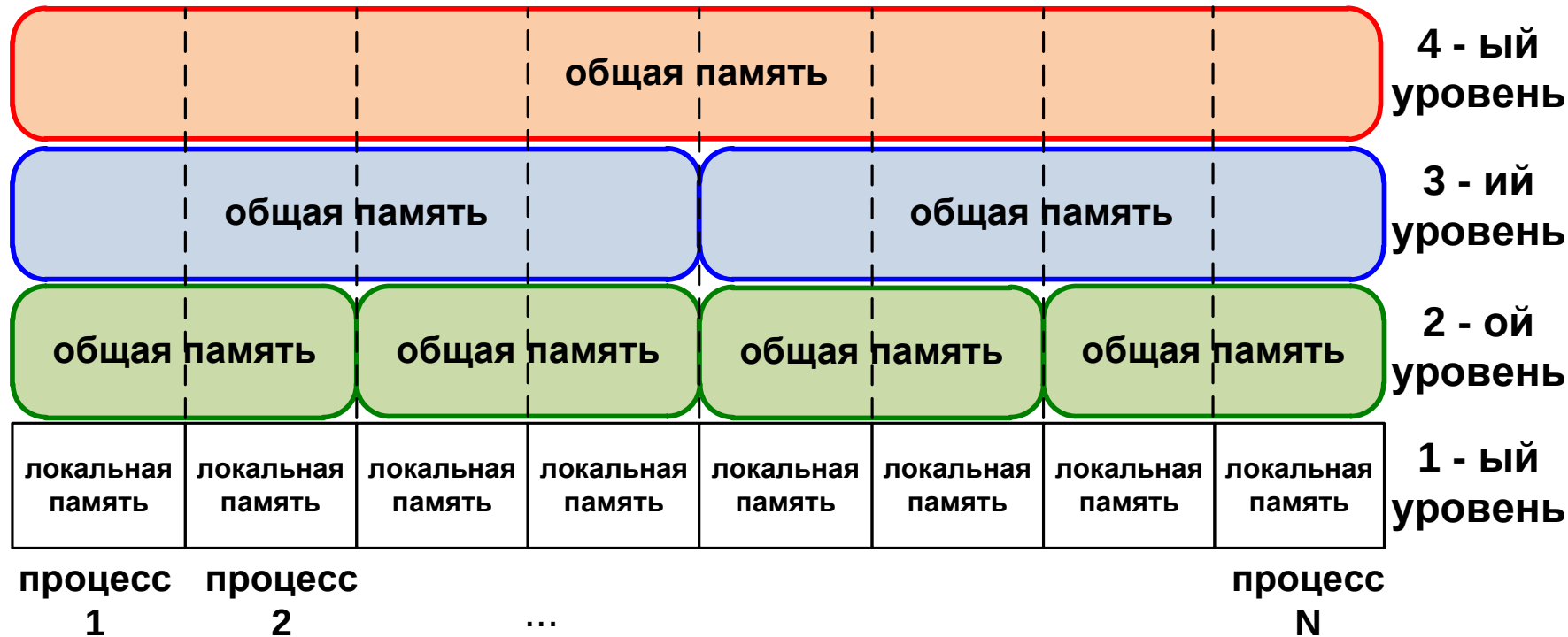
Моделирование
будущих
архитектур

повышение
толерантности

улучшение
локализации

Заключение

HPGAS (Hierarchical PGAS)



HPGAS – результаты тестирования используются при составлении и использовании модели

**Спасибо
за внимание**