

# Персональная гибридная вычислительная система. Тестирование производительности

Д.Ахмедов<sup>1</sup>, С.Елубаев<sup>1</sup>, Ф.Абдолдина<sup>2</sup>, Т.Бопеев<sup>1</sup>,  
Д.Муратов<sup>1</sup>, Р.Поветкин<sup>1</sup>, А.Каратаев<sup>1</sup>

<sup>1</sup>Лаборатория имитационного моделирования компьютерных систем,  
Институт космической техники и технологий, Алматы, Республика Казахстан

<sup>2</sup>Кафедра Программного обеспечения систем и сетей,

Институт информационных и телекоммуникационных технологий,

Казахский национальный технический университет им. К.И. Сатпаева, Алматы, Республика Казахстан

akhmedov.d@istt.kz, elubaev.s@istt.kz, farida\_mail@mail.ru

**Аннотация.** В статье рассматриваются вопросы разработки персональных гибридных вычислительных систем на базе графических процессоров для высокопроизводительных вычислений. Изложены результаты тестирования реальной производительности системы в тесте Linpack при различных моделях графических карт последнего поколения (Nvidia Tesla M2090, Nvidia Tesla C2075) и разных объемах оперативной памяти (от 4,2 Гб до 107,5 Гб).

## Ключевые слова

Параллельные вычисления, высокопроизводительные вычисления, персональная гибридная вычислительная система, графический процессор, CUDA технология.

## 1 Введение

На сегодняшний день одним из главных векторов современного развития информационных технологий является развитие суперкомпьютеров. Во многих странах мира суперкомпьютерные технологии отнесены к числу важнейших направлений развития техники и технологий. Их активное и эффективное применение не только способствует глубокой модернизации промышленности, но и обеспечивает лидерство в глобальной экономической конкуренции [1].

Суперкомпьютеры, как отмечают эксперты, позволяют решить большое количество задач, которые невозможно проверить натурным экспериментом, и в настоящее время практически не осталось отраслей, где не были бы задействованы и суперкомпьютерное моделирование, и сами суперкомпьютеры.

Появившиеся возможности современной базы комплектующих для параллельных вычислений, и особенно резкое увеличение производительности графических видеокарт нового поколения, делают реальным самостоятельное создание и развитие суперкомпьютерных систем в Казахстане. Одним из перспективных направлений развития суперкомпьютерных технологий является использование вычислительных возможностей графических плат Nvidia Tesla для создания персональных гибридных вычислительных систем.

В Институте космической техники и технологий разработана персональная гибридная вычислительная система на базе графических карт с использованием CUDA – технологии ISTT HPC 2000, внешний вид которой приведен на рисунке 1, основные характеристики приведены в таблице 1. В центре сертификации продукции и услуг был получен Сертификат соответствия (рисунок 1).

Персональная гибридная вычислительная система в целом представляет собой персональную вычислительную машину, располагаемую непосредственно на рабочем месте пользователя. Состав персональной гибридной вычислительной системы в целом совпадает с составом персонального компьютера, за исключением того, что в его составе присутствуют специализированные высокопроизводительные графические процессоры Nvidia Tesla, собственно и осуществляющие быстрые вычисления. Наличие мощного центрального процессора и нескольких графических процессоров в совокупности образуют высокопроизводительную вычислительную систему, где ресурсоемкие задачи могут быть разделены между данными процессорами, тем самым обеспечивая высокую параллельность и скорость вычислений.



Рис. 1. Персональная гибридная вычислительная система ISTT HPC 2000. Сертификат соответствия

Таб.1. Основные характеристики персональной гибридной вычислительной системы ISTT HPC2000

<i>Компонент</i>	<i>Характеристики</i>
Пиковая производительность	2 и 4 Тфлопс двойной и одинарной точности
Процессор	2 x 2,9ГГц
Графический процессор	4 x Nvidia Tesla C2075
Оперативная память	8 x 16 Гб
Материнская плата	Asus Z9PE-D8 WS

При проектировании персональной вычислительной системы основную сложность составляет расчет схемы оптимального соотношения мощного центрального процессора и графических процессоров, системы воздушного охлаждения, оптимальной схемы электропитания. Для обеспечения необходимого уровня охлаждения при интенсивной работе графических процессоров применяется эффективная система воздушного охлаждения, построенная таким образом, что бы воздушный поток свободно циркулировал между вычислительными узлами персональной вычислительной системы, не допуская их перегрева. Для этого применяются наиболее эффективные механические элементы системы охлаждения – вентиляторы с низким уровнем выделяемых шумов, а так же специальный тип корпуса системного блока, позволяющий в некоторых пределах произвольно изменять расположение устройств охлаждения.

Наличие нескольких графических процессоров, мощных центральных процессоров, жестких дисков и других элементов требует обеспечения качественного электропитания. Для этого в составе системного блока вычислительной системы присутствует качественный блок питания с рассчитанной мощностью, достаточной для обеспечения продолжительной бесперебойной работы системного блока.

## 2 Предшествующие работы

Результаты исследований, представленные в статье, являются частью работ, проводимых в рамках научного проекта «Разработать персональную гибридную вычислительную систему на базе GPU-процессоров», входящего в бюджетную программу «Разработать технологию создания суперкомпьютерного гибридного кластера с применением GPU-процессоров» Министерства образования и науки Республики Казахстан.

Целью данного проекта является выполнение опытно-конструкторских работ по созданию персональной гибридной вычислительной системы (ПГВС) на базе GPU-процессоров.

В рамках проекта создан опытный образец персональной гибридной вычислительной системы, предназначенный для использования в научно-исследовательских работах, для задач с повышенной точностью. На предварительном этапе создания опытного образца ПГВС создан экспериментальный образец ПГВС, на базе которого проведены исследования зависимости реальной производительности от характеристик отдельных компонент аппаратной составляющей системы. Выполненные работы в рамках данного этапа позволили оценить степень влияния отдельных комплектующих ПГВС на производительность системы, позволили удостовериться в правильности направления создания ПГВС и легли в основу процесса создания опытного образца ПГВС.

Решаемая задача позволила формировать состав аппаратной части ПГВС таким образом, чтобы производительность системы удовлетворяла требованиям поставленных задач, и при этом достигалось наилучшее соотношение цена/производительность создаваемой системы.

Экспериментальный образец ПГВС имеет пиковую производительность около 3 ТФлопс одинарной и 1,5 ТФлопс двойной точности. Разработанный экспериментальный образец ПГВС является хорошей альтернативой дорогостоящим суперкомпьютерам, демонстрируя существенно лучшие показатели производительности на Ватт потребляемой мощности. Мощность энергопотребления экспериментального образца ПГВС составила около 1200Вт. Соотношение производительности на Ватт потребляемой мощности составило 552,5 МФлопс/Вт.

Производительность экспериментального образца ПГВС в расчете на стоимость владения также чрезвычайно высока. Соотношение цена производительность составило примерно \$20 за 1 ГФлопс/сек на тесте Linpack.

Тестирование производительности экспериментального образца ПГВС на базе трех графических процессоров проводилось при разных тактовых частотах (1066 и 1333МГц) и разных объемах оперативной памяти (от 2 до 24 Гб с шагом 2Гб), так же изучено влияние на производительность пропускной способности шины PCI-Express.

На момент тестирования производительности экспериментального образца ПГВС конфигурация системы отображена в таблице 2.

**Таб.2.** Характеристики ЭО ПГВС с тремя графическими процессорами

<i>Компонент</i>	<i>Описание / характеристики</i>	<i>Количество</i>
Процессор	Intel core i7-960 3.2Ghz	1
Графический процессор	Nvidia Tesla C2050	3
Видеокарта	GeForce 9800GTX+	1
Оперативная память	4Gb DDR3 1066/1333MHz	6
Материнская плата	Asus Rampage III	1

Проведенное тестирование экспериментального образца ПГВС позволило определить реальную производительность системы в тесте Linpack, которая составила 663 ГФлопс, что составляет 44,2% от пиковой производительности.

### 3 Результаты тестирования производительности ПГВС в форм-факторе Full Tower

В данной статье представлены результаты оценки производительности опытного образца ПГВС ISTT НРС 2000 в форм-факторе FullTower с 4-мя графическими процессорами.

**Таб. 3 -** Результаты тестирования 1-го варианта ПГВС на базе четырех графических процессоров Nvidia Tesla M2090

<i>№</i>	<i>Объем оперативной памяти, Гб</i>	<i>Размерность задачи</i>	<i>Размерность разбиения задачи</i>	<i>Время, сек</i>	<i>Реальная производительность (Linpack), Гфлопс</i>	<i>Эффективность, %</i>	<i>Удельная стоимость 1 ГФлопс/сек, тг</i>
1	2	14 273	512	4,6	422,3	16,0%	5 022,50
2	4	20 822	512	9,7	621,1	23,5%	3 414,91
3	6	25 502	768	18,2	606,7	23,0%	3 504,20
4	8	29 447	512	21,7	785,1	29,7%	2 707,94
5	10	32 923	512	28,2	843,2	31,9%	2 527,28
6	12	36 066	1 024	35,8	873,5	33,1%	2 439,61
7	14	38 956	768	42,3	930,9	35,3%	2 294,55
8	16	41 645	512	46,9	1 027,0	38,9%	2 079,84
9	18	44 171	512	57,4	1 001,0	37,9%	2 138,86
10	20	46 561	512	63,8	1 055,0	40,0%	2 029,38
11	22	48 833	768	73,0	1 063,0	40,3%	2 018,81
12	24	52 224	512	89,3	1 064,0	40,3%	2 016,92
13	26	56 853	512	111,9	1 095,0	41,5%	1 964,38
14	28	57 344	1 024	114,6	1 097,0	41,6%	1 960,80
15	30	58 368	512	118,7	1 117,0	42,3%	1 930,17
16	32	59 396	768	124,7	1 120,0	42,4%	1 925,00

Для обоснования выбора элементной базы ПГВС были построены 2 варианта персональной гибридной вычислительной системы на базе графических карт с использованием CUDA-технологии.

В составе 1-го варианта ПГВС использовались два центральных процессора с тактовой частотой 2ГГц, четыре графических процессора Nvidia Tesla M2090, в составе 2-го варианта ПГВС - два центральных процессора с тактовой частотой 2,9ГГц, четыре графических процессора Nvidia Tesla C2075.

Для обоих вариантов ПГВС проведены тесты по определению реальной производительности в тесте Linpack. Использовался Cuda Accelerated Linpack 2.0 для массивно-параллельных гибридных систем [2]. Данный тест позволяет задать все значимые параметры алгоритма, подбирая их для определения наиболее точного значения реальной производительности.

Тестирование производительности ПГВС на базе четырех графических процессоров проводилось при разных объемах оперативной памяти. Для 1-го варианта ПГВС объем менялся от 2 Гб до 32 Гб с шагом 2Гб (таблица 3), для 2-го варианта от 4,2 Гб до 107,5 Гб (таблица 4).

Для 1-го варианта ПГВС при объеме оперативной памяти в 2 Гб реальная производительность составила 422,3 ГФлопс а при 32 Гб достигла предела, равного 1 120,0 ГФлопс. Таким образом, реальная производительность возросла в 2,7 раза.

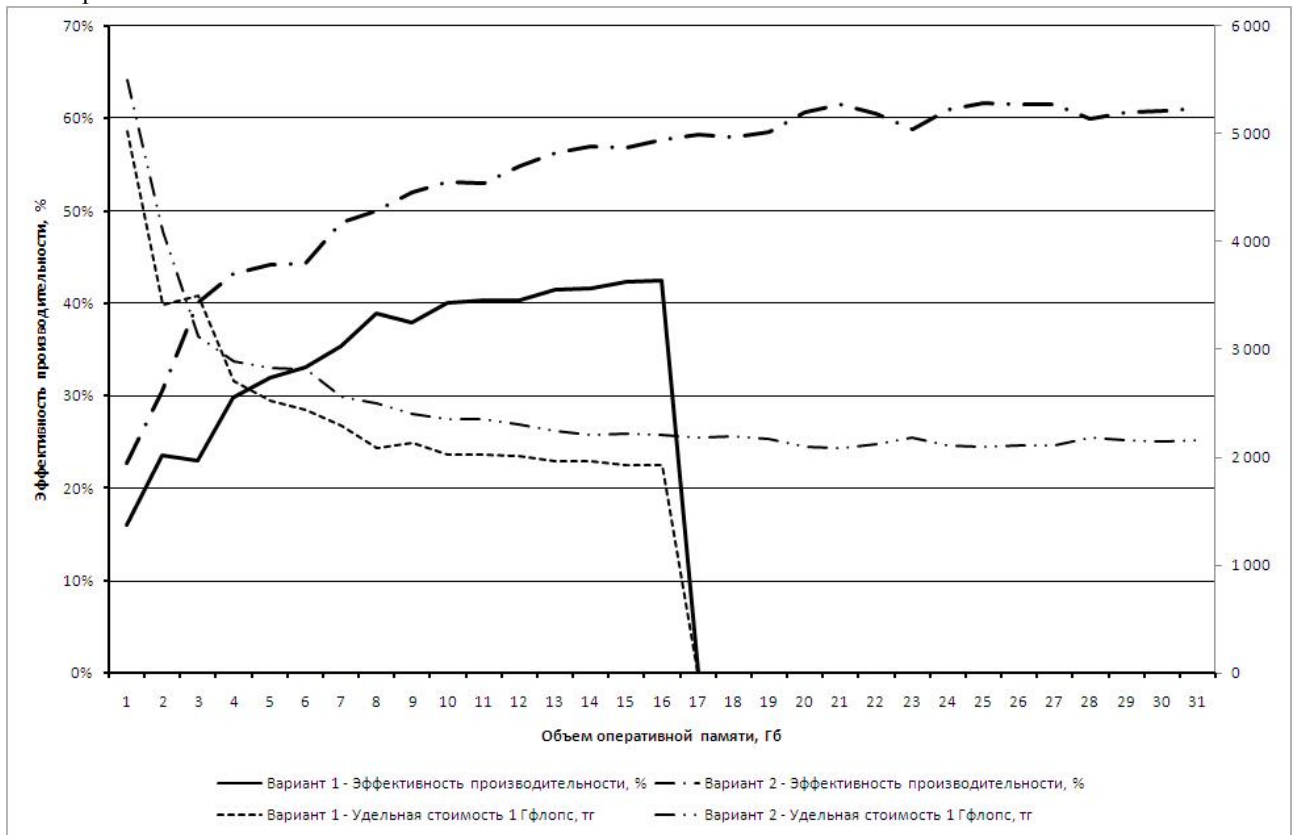
**Таб. 4** - Результаты тестирования 2-го варианта ПГВС на базе четырех графических процессоров Nvidia Tesla C2075

<i>№</i>	<i>Объем оперативной памяти, Гб</i>	<i>Размерность задачи</i>	<i>Размерность разбиения задачи</i>	<i>Время, сек</i>	<i>Реальная производительность (Linpack), Гфлопс</i>	<i>Эффективность, %</i>	<i>Удельная стоимость 1 ГФлопс/сек, мг</i>
1	4,2	14 273	512	4,1	468,8	22,8%	5 501,28
2	6,1	20 822	512	9,6	629,0	30,5%	4 100,16
3	7,8	25 502	768	13,4	824,4	40,0%	3 128,34
4	8,5	29 447	768	19,1	891,3	43,3%	2 893,53
5	10,2	32 923	512	26,1	911,7	44,3%	2 828,78
6	13,7	36 066	1 024	34,3	912,4	44,3%	2 826,61
7	15,9	38 956	768	39,3	1 004,0	48,7%	2 568,73
8	17,6	41 645	1 024	46,8	1 029,0	50,0%	2 506,32
9	19	44 171	768	53,7	1 071,0	52,0%	2 408,03
10	21,6	46 561	512	61,5	1 094,0	53,1%	2 357,40
11	23,4	48 833	1 024	71,0	1 093,0	53,1%	2 359,56
12	25,7	52 224	1 024	84,1	1 129,0	54,8%	2 305,58
13	29,3	56 853	1 024	105,6	1 160,0	56,3%	2 243,97
14	33,54	60 068	1 024	123,1	1 174,0	57,0%	2 217,21
15	35,1	63 284	1 024	144,2	1 172,0	56,9%	2 220,99
16	40,1	66 499	768	165,3	1 188,0	57,7%	2 211,28
17	42,7	69 715	1 024	188,2	1 200,0	58,3%	2 189,17
18	46,3	72 930	1 024	216,4	1 195,0	58,0%	2 198,33
19	50,1	76 146	768	244,0	1 206,0	58,5%	2 178,28
20	54,7	79 361	1 024	266,6	1 250,0	60,7%	2 101,60
21	58,8	82 577	1 024	296,2	1 268,0	61,6%	2 090,69
22	62,4	85 793	1 024	337,5	1 246,0	60,5%	2 127,61
23	66,8	89 008	1 280	388,2	1 211,0	58,8%	2 189,10
24	71,3	92 224	1 024	416,3	1 256,0	61,0%	2 110,67
25	75,3	95 439	1 024	456,3	1 270,0	61,7%	2 106,30
26	80,9	98 655	1 024	505,4	1 267,0	61,5%	2 111,29
27	85,1	101 870	1 024	555,9	1 268,0	61,6%	2 109,62
28	91,1	105 086	1 024	626,9	1 234,0	59,9%	2 187,20
29	96,4	108 301	1 024	677,6	1 250,0	60,7%	2 159,20
30	103	111 517	1 024	737,8	1 253,0	60,8%	2 154,03
31	107,5	114 733	1 024	801,0	1 257,0	61,0%	2 166,27

Для 2-го варианта ПГВС при объеме оперативной памяти в 4,2 Гб реальная производительность составила 468,8 ГФлопс а при 107,5 Гб достигла предела, равного 1 257,0 ГФлопс. Таким образом, реальная производительность для 2-го варианта возросла так же в 2,7 раза.

2-ой вариант комплектации ПГВС позволил получить 61,7% значение эффективности на тесте Linpack, в то время как для 1-го варианта было достигнуто только 42,4%. Эффективность производительности 2-го варианта превышает производительность 1-го варианта в среднем на 13% (рисунок 2).

Однако же удельная стоимость 1 ГФлопс/сек для 1-го варианта ПГВС оказалась в среднем на 11% дешевле 2-го варианта.



**Рис. 2** – Зависимость эффективности производительности и стоимости 1 ГФлопс/сек от объема оперативной памяти для 1-го и 2-го варианта ПГВС

## 4 Заключение

Проведенные работы позволили оценить степень влияния производительности центрального и графических процессоров на эффективность системы в целом, что легло в основу выбора электронной компонентной базы для опытного образца ПГВС.

## Список литературы

- [1] Top500 Supercomputer sites: [Электронный ресурс]. 2000-2012. URL: <http://www.top500.org/>
- [2] Техническая поддержка Nvidia: [Электронный ресурс]. URL: <http://developer.nvidia.com>.