УДК 004.2

Исследование характеристик гибридного вычислительного кластера

Борисов В. В., Зернов М. М., Федулов А. С., Якушевский К. А.

Постановка задачи: *в условиях возрастающего объема обрабатываемой информации, увеличения интенсивности и сложности решаемых поисковых и аналитических задач все более насущным для различных организаций и отдельных исследователей является создание и применение гибридных вычислительных систем, имеющих в своем составе как центральные, так и графические процессоры. Эти системы обладают потенциальной производительностью, позволяющей отнести их к классу суперкомпьютеров, при сохранении приемлемых для пользователей ценовых, массогабаритных, эргономических характеристик. Вместе с тем, следует отметить дефицит публикаций, посвященных вопросам исследования характеристик таких систем в зависимости от особенностей их архитектуры и используемого информационного обеспечения, а также выработке рекомендаций по оптимизации их структурных решений и параметров.* ***Целью работы*** *является предложение инструментальных средств и исследование с их использованием производительности гибридных вычислительных систем (на примере кластера филиала ФГБОУ ВО «НИУ «МЭИ» в г. Смоленске – СФМЭИ) для выработки конкретных рекомендаций по оптимизации структурных решений и параметров.* ***Используемые методы****: методы анализа и проектирования вычислительных систем.* ***Новизна:*** *представленные результаты могут послужить основанием для обобщения подхода к исследованию гибридных вычислительных систем, ориентированных на решение разных нечисловых и вычислительных задач и различающихся архитектурой и используемым информационным обеспечением.* ***Результат:*** *представлена архитектура гибридного вычислительного кластера СФМЭИ. Проведена оценка максимальной теоретической производительности. Выполнены оценки производительности на основе тестов.* ***Практическая значимость:*** *результаты исследований могут быть использованы для анализа характеристик гибридных вычислительных систем, ориентированных на решение различных нечисловых и вычислительных задач.*

Ключевые слова: *гибридная вычислительная система, центральный процессор, графический процессор.*

Актуальность

В условиях возрастающего объема обрабатываемой информации, увеличения интенсивности и сложности решаемых поисковых и аналитических задач все более насущным для различных организаций и отдельных исследователей является создание и применение гибридных вычислительных систем, имеющих в своем составе как центральные (CPU), так и графические (GPU) процессоры. Эти системы обладают потенциальной производительностью, позволяющей отнести их к классу суперкомпьютеров, при сохранении приемлемых для пользователей ценовых, массогабаритных, эргономических характеристик. Вместе с тем, следует отметить дефицит публикаций, посвященных вопросам исследования характеристик таких систем в зависимости от особенностей их архитектуры и используемого информационного обеспечения, а также выработке рекомендаций по оптимизации их структурных решений и параметров.

В работе исследуются характеристики гибридного вычислительного кластера (ГВК) на примере кластера филиала ФГБОУ ВО «НИУ «МЭИ» в

г. Смоленске (Смоленского филиала МЭИ – СФМЭИ). Полученные результаты могут послужить основанием для обобщения подхода к исследованию гибридных вычислительных систем, ориентированных на решение нечисловых и вычислительных задач и различающихся архитектурой и используемым информационным обеспечением.

1. Постановка задачи. Состав и взаимосвязь компонентов гибридного вычислительного кластера

Общая схема ГВК СФМЭИ представлена на рис. 1.

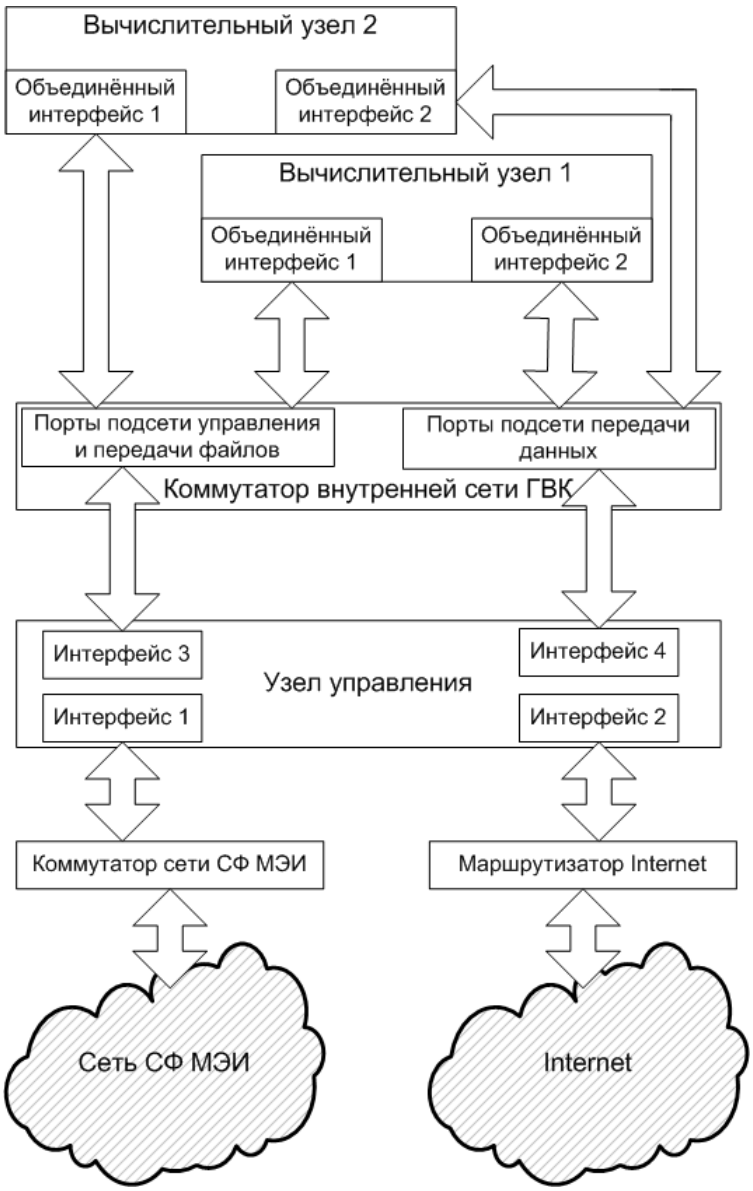

Рис. 1. Взаимосвязь основных компонентов кластера

Как видно из рис. 1, кластер состоит из двух вычислительных узлов, узла управления, необходимых интерфейсных (связных) компонентов.

Каждый вычислительный узел (ВУ) включает в себя два 10-ядерных процессора, оперативную память объемом 64 Гбайт, два твердотельных накопителя SSD с интерфейсом SATA III, организованных в массиве RAID 1,

объемом 240 Гбайт, а также по два интерфейсных канала, обеспечивающих подключение вычислительных узлов к внутренней сети кластера.

Кроме того, ВУ 1 содержит две видеокарты на базе GPU NVIDIA, а также два дисковых накопителя HDD объёмом 1 Тбайт с интерфейсом SATA III, организованных в виде массива RAID 1. Вычислительный узел 2 содержит две видеокарты на базе GPU AMD.

Дисковые накопители вычислительных узлов имеют следующее назначение. Основные разделы операционной системы узлов размещаются на SSD-накопителях. Файловая система вычислительных узлов для хранения общих данных, исходных текстов и бинарных файлов программ, исходных данных и результатов работы программ, запускаемых пользователями параллельно, размещается на HDD-накопителе вычислительного узла 1. Таким образом, в случае, когда вычислительным заданиям достаточно ресурсов одного узла, их эффективнее запускать на ВУ 1.

Узел управления содержит два 8-ядерных процессора Intel Xeon E5-2620 v4, оперативную память объемом 64 Гбайт стандарта DDR4, девять дисковых накопителей: SSD-накопитель объёмом 256 Гбайт с интерфейсом PCIe 3.0×4 в стандарте M.2; два SSD-накопителя объёмом 250 Гбайт с интерфейсом SATA III в массиве RAID 1; два HDD-накопителя объёмом 2 Тбайт с интерфейсом SATA III в массиве RAID 1; четыре HDD-накопителя объёмом 3 Тбайт с интерфейсом SATA III, размещенных в массиве RAID 10 с объёмом 6 Тбайт. Наличие нескольких дисковых накопителей позволяет, в том числе, создавать на управляющем узле требуемую систему виртуальных машин.

Внутренняя сеть ГВК построена на основе 16-портового коммутатора HP V1910-16G Switch и разделена на две подсети: для передачи файлов (используется сетевой файловой системой и общесистемными инструментами администрирования); для передачи данных (используется менеджером ресурсов для управления вычислительной нагрузкой и обеспечивает передачу сообщений по протоколу MPI). Каждая из подсетей дополнительно разделена на две параллельно работающие части, что позволяет, используя функции связывания каналов, получить соединения с удвоенной пропускной способностью.

Коммутатор сети СФМЭИ и маршрутизатор сети Интернет обеспечивают связь кластера с внешним миром. Доступ к кластеру осуществляется как из внутренней сети филиала, так и по внешнему Internet-каналу.

Вычислительные узлы находятся под управлением дистрибутива Centos 7.2 операционной системы Linux, управление вычислительными ресурсами осуществляет менеджер ресурсов SLURM версии 15.08. Обработка данных может осуществляться посредством как центральных процессоров на основе технологий MPI и Open MP, так и графических на основе технологий OpenCL и CUDA.

2. Архитектура вычислительных узлов кластера

В данном разделе представлена архитектура вычислительных узлов ГВК с точки зрения оценки производительности (рис. 2 и 3). Вычислительные узлы построены по однотипной схеме, хотя их состав отличается друг от друга.

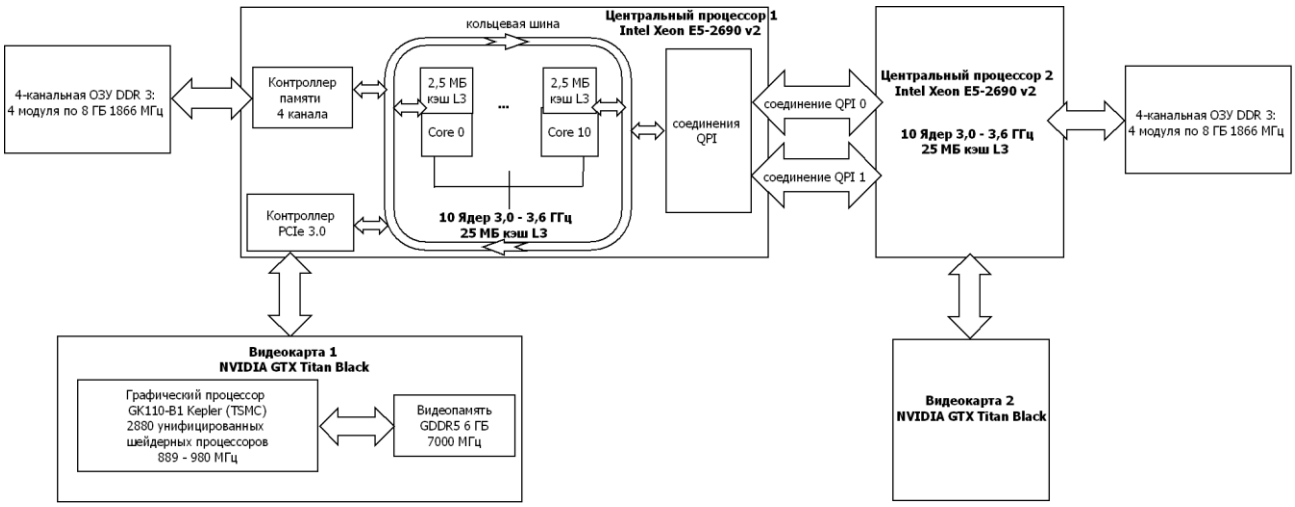

Рис. 2. Архитектура вычислительного узла 1

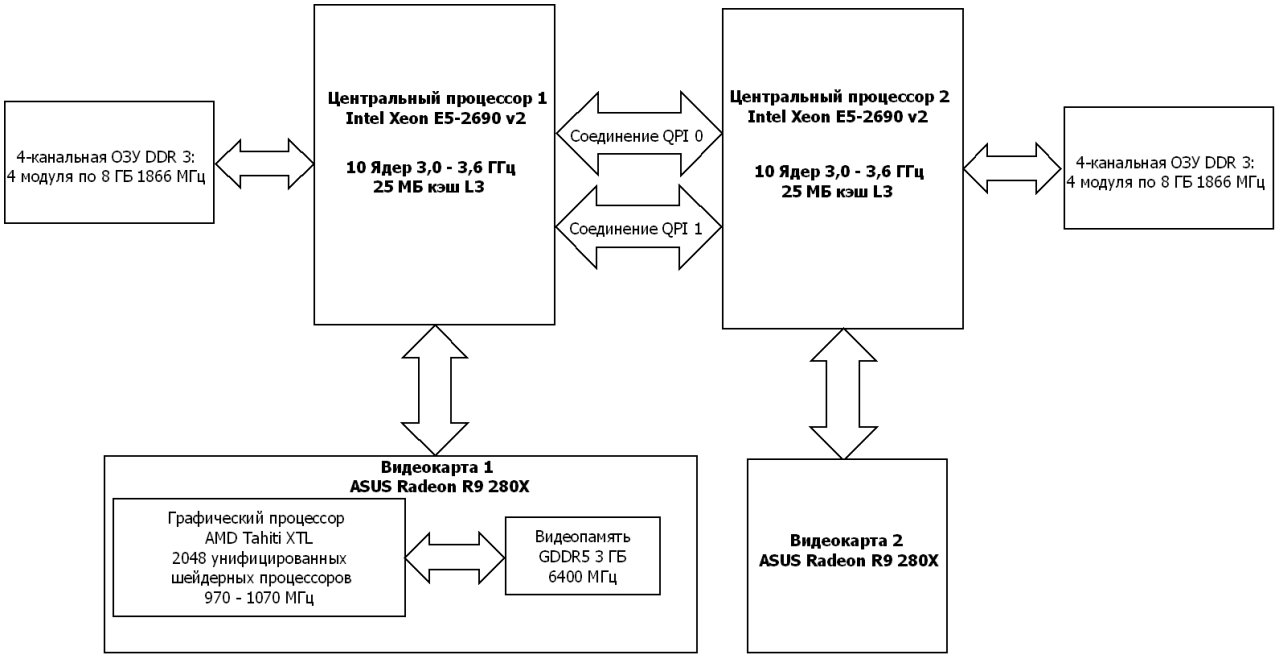

Рис. 3. Архитектура вычислительного узла 2

Каждый вычислительный узел содержит два 10-ядерных центральных процессора Intel Xeon E5-2690V2, связанных между собой шиной QPI. Каждое ядро имеет кэш-память 3-го уровня емкостью 2,5 Мбайт [1]. Соответственно, общий объем кэш-памяти одного процессора составляет 25 Мбайт. Каждый процессор имеет собственную оперативную память емкостью 32 Гбайт в формате DDR3, состоящую из четырех модулей емкостью по 8 Гбайт. Каждый модуль памяти подключен к отдельному каналу связи с процессором.

В процессорах поддерживается расширенный набор инструкций AVX [1] и технология Hyper Threading [2, 3], позволяющая выполнять на одном ядре сразу два потока вычислений и представляющая одно физическое ядро как два логических на уровне операционной системы.

Вычислительный узел 1 включает в себя две видеокарты NVIDIA GeForce GTX Titan Black на базе графического процессора GK110-B1 Kepler [4-6]. Каждая видеокарта оснащена 6 ГБайт графической памяти формата GDDR5.

Вычислительный узел 2, в отличие от первого, использует видеокарты ASUS Radeon R9280X, построенные на GPU компании AMD с кодовым наименованием Tahiti XTL [7, 8], использующими архитектуру вычислительных блоков GCN. Обе видеокарты ВУ 2 включают по 3 Гбайт графической памяти формата GDDR5.

Связь центральных процессоров с графическими процессорами осуществляется по шине PCIe 3.0 (с поддержкой 16-ти линий) под управлением встроенного в центральный процессор контроллера. Каждый процессор непосредственно связан с одной видеокартой.

По способу доступа центральных процессоров к оперативной памяти в вычислительных узлах реализована архитектура с неравномерным доступом (Non-Uniform Memory Access – NUMA), в которой каждый процессор обладает собственной локальной памятью с быстрым доступом, но при этом имеет более медленный доступ к памяти другого процессора [9].

3. Оценка производительности гибридного вычислительного кластера

3.1. Оценка максимальной теоретической производительности

Максимальная (пиковая) производительность кластера может быть оценена с учетом его архитектуры и заявленных в спецификациях характеристик производительности отдельных компонентов.

Обобщенные сведения и характеристики теоретической пиковой производительности основных компонентов вычислительных узлов ГВК представлены в таблице 1.

Номинальная частота каждого центрального процессора вычислительного узла составляет 3,0 ГГц, в режиме Turbo повышается до 3,6 ГГц. В соответствии с расчётами по методике, представленной в [10], максимальная (теоретическая) производительность центральных процессоров при использовании AVX-инструкций составляет для вычислений с плавающей точкой одинарной точности 480 Гфлопс, для вычислений с плавающей точкой двойной точности 240 Гфлопс.

Модули оперативной памяти вычислительных узлов работают на частоте 1866 МГц. Пропускная способность одного канала процессор-память составляет 14933 Мбайт/с. С учетом возможности параллельной работы четырех каналов, максимальная скорость обмена процессоров с памятью составляет 59,7 Гбайт/с [1].

Обмен данными между двумя процессорами одного вычислительного узла осуществляется посредством 2-х соединений шины QPI. Каждое соединение имеет пропускную способность 8 Гигатранзакций/с в каждую сторону, что означает реальную пропускную способность 16 Гбит/с в каждую сторону на одно соединение, 32 Гбит/с в каждую сторону для двух соединений, 64 Гбит/с при одновременной передаче в обе стороны [11].

Интерфейс связи CPU и GPU поддерживает 16 линий шины PCIe 3.0, которая имеет пропускную способность одной линии 1 Гбит/с. Следовательно, пропускная способность интерфейса CPU↔GPU составляет 16 Гбит/с [12].

Таблица 1 – Сводные данные по основным компонентам вычислительных узлов ГВК

Вычислительный узел 1			
CPU			
Количество и название	Суммарное число ядер/ потоков	Частота процессора базовая/ турбо	Суммарная теоретическая пиковая производительность вычислений с плавающей точкой одинарной точности/двойной точности
Два процессора Intel Xeon E5-2690V2 Ivy Bridge-EP	20/ 40	3000 МГц/ 3600 МГц	960 Гфлопс/ 480 Гфлопс
GPU			
Количество и название	Суммарное число процессоров одинарной точности	Суммарный объём видеопамяти	Суммарная теоретическая пиковая производительность вычислений с плавающей точкой одинарной точности/двойной точности
Две видеокарты NVidia GTX Titan BE	5760	12 Гбайт	11 290 Гфлопс/ 3 762 Гфлопс
ОЗУ			
Ёмкость	Эффективная частота	Тип	Суммарная пропускная способность
64 Гбайт (8 модулей по 8 Гбайт)	1866 МГц	4-канальная DDR3 ECC Reg	119,4 Гбайт/с (2×59,7 Гбайт/с)
Вычислительный узел 2			
CPU			
Количество и название	Суммарное число ядер/ потоков	Частота процессора базовая/ турбо	Суммарная теоретическая пиковая производительность вычислений с плавающей точкой одинарной точности/двойной точности
Два процессора Intel Xeon E5-2690V2 Ivy Bridge-EP	20/ 40	3000 МГц/ 3600 МГц	960 Гфлопс/ 480 Гфлопс
GPU			
Количество и название	Суммарное число процессоров одинарной точности	Суммарный объём видеопамяти	Суммарная теоретическая пиковая производительность вычислений с плавающей точкой одинарной точности/двойной точности
Две видеокарты ASUS Radeon R9 280X	4096	6 Гбайт	8 768 Гфлопс/ 2 192 Гфлопс
ОЗУ			
Ёмкость	Эффективная частота	Тип	Суммарная пропускная способность
64 Гбайт (8 модулей по 8 Гбайт)	1866 МГц	4-канальная DDR3 ECC Reg	119,4 Гбайт/с (2×59,7 Гбайт/с)

Процессоры GK110-B1 Kepler видеокарт NVIDIA, установленных на ВУ 1, работают на базовой частоте 889 МГц (в режиме разгона, в зависимости от нагрузки, частота возрастает до 980 МГц). Графическая память видеокарты работает на эффективной частоте 7000 МГц (реальная частота 1750 МГц). Ширина шины GPU↔видеопамять составляет 384 бит, а ее пропускная способность составляет 336 Гбит/с. Графический процессор GK-110-Bl содержит 2880 ядер CUDA одинарной точности и втрое меньше ядер для обработки операций с плавающей точкой двойной точности [6, 13].

Анализ показывает, что пиковая теоретическая производительность видеокарты NVIDIA GeForce GTX Titan Black в операциях с плавающей точкой одинарной точности составляет 5645 Гфлопс, в операциях двойной точности – 1881 Гфлопс [6, 13, 14].

Второй вычислительный узел, в отличие от первого, использует видеокарты ASUS Radeon R9280X, построенные на GPU компании AMD с кодовым наименованием Tahiti XTL, использующими архитектуру вычислительных блоков GCN. Базовая частота GPU равна 970 МГц, частота автоматического разгона – 1070 МГц. Графическая память работает на эффективной частоте 6400 МГц (реальная частота 1600 МГц). Ширина шины памяти 384 бит, максимальная пропускная способность интерфейса GPU↔видеопамять составляет 307,2 Гбит/с. Графический процессор Tahiti XTL содержит 2048 потоковых процессора, обеспечивающих пиковую производительность операций с плавающей точкой одинарной точности 4384 Гфлопс, операции двойной точности используют те же самые потоковые процессоры и выполняются в темпе ¼ от одинарной – 1096 Гфлопс [8, 14, 15].

3.2. Тестовая оценка производительности

3.2.1. Состав применяемого программного обеспечения

Тестирование производительности кластера проводилось с использованием программного обеспечения (ПО), представленного в таблице 2.

Таблица 2 – Версии системного и тестового ПО

№	Тип ПО	Наименование и версия ПО
1	Операционная система	CentOS 7.2.1511
2	Ядро операционной системы	Linux 3.10.0-327
3	Командная строка (shell) пользователя	Bash 4.2.46
4	Компилятор языка C	GCC (GNU Compiler Collection) 4.8.5 20150623, модель многопоточности – POSIX
5	Библиотека C	Glibc (GNU libc) 2.17
6	Платформы GPGPU	
	Узел 1	NVIDIA UNIX x86-64 Kernel Module 352.79 OpenCL 1.2 CUDA 7.5 NVCC NVidia Cuda compiler driver 7.5.27
	Узел 2	AMD fglrx 14.50.2 driver OpenCL 1.2
7	Реализация MPI	OpenMPI 1.10.0
8	Реализация OpenMP	OpenMP 3.1
9	Тесты LINPACK	HPL 2.2 Intel Optimized MP LINPACK Benchmark for Clusters l_mklb_p_2017.1.013
10	Инструментарий гетерогенного тестирования	The Scalable HeterOgeneous Computing (SHOC) Benchmark Suite 1.1.5

3.2.2. Тестирование производительности центральных процессоров вычислительного кластера на основе тестов LINPACK

Тестирование производительности ГВК с вычислениями, проводимыми на CPU, выполнялось на основе тестов LINPACK открытой версии HPL 2.2 с использованием библиотеки BLAS [16,17] и той же версии, в которой была использована специальная библиотека математических функций (math kernel library – MKL), оптимизирующая вычисления для архитектуры процессоров Intel и кластеров, построенных на их основе [18].

Особенностью LINPACK HPL 2.2 является возможность изменения параметров алгоритма решения системы линейных алгебраических уравнений (СЛАУ), лежащего в основе теста. Для параметрической настройки используется специальный файл конфигурации HPL.dat. При тестировании ГВК СФМЭИ часть параметров теста была выбрана по умолчанию из руководства по настройке, а часть – изменялась для поиска вариантов, обеспечивающих наилучшую производительность.

При работе алгоритма матрица коэффициентов СЛАУ с числом уравнений N разбивается на подматрицы размером $NB \times NB$, а каждая из полученных подматриц разбивается на $P \times Q$ блоков, где P – число блоков по строкам, Q – по столбцам. Каждый из полученных блоков ассоциируется с процессом, закрепляемым за отдельным физическим ядром процессора.

Тестирование ГВК СФМЭИ тестом LINPACK без оптимизации под архитектуру Intel было выполнено для двух совместно работающих вычислительных узлов.

Существенное влияние на производительность оказывают размер матрицы N и размер подматрицы NB. На рис. 4 представлены результаты серии экспериментов с различными сочетаниями параметров N и NB при фиксированном значении ряда параметров конфигурации теста:

```
PMAP   : Column-major process mapping
P      :       4
Q      :      10
PFACT  :   Right
NBMIN  :       4
NDIV   :       2
RFACT  :   Right
BCAST  :  1ringM
DEPTH  :       0
SWAP   : Mix (threshold = 64)
L1     : transposed form
U      : transposed form
EQUIL  : yes
ALIGN  : 8 double precision words
```

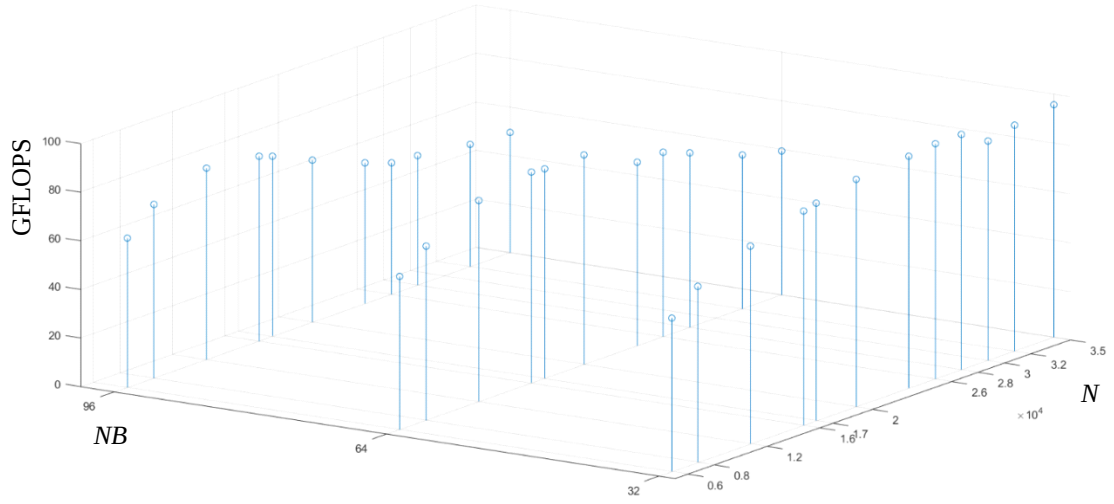

Рис. 4. Зависимость производительности ГВК СФМЭИ от *N* и *NB*

Наилучшие результаты при любых размерах матрицы *N* были получены для размера подматрицы *NB* = 32 элемента.

В результате дальнейшего тестирования было выявлено, что выбор вариантов реализаций матричных операций алгоритма (параметры конфигурации PFACT и RFACT) не оказывает существенного влияния на производительность.

Среди различных алгоритмов передачи данных наилучшие результаты показал алгоритм Increasing-ring [17] (параметр конфигурации BCAST=1ring).

В результате серии экспериментов наилучшая оценка производительности кластера, которая была достигнута с помощью теста LINPACK без оптимизации с использованием MKL, составила 101,7 Гфлопс (при *N* = 35000, *NB* = 32) с параметрами теста:

```
N      :    35000
NB     :   32
PMAP   : Column-major process mapping
P      :       5
Q      :      8
PFACT  :   Right
NBMIN  :       4
NDIV   :       2
RFACT  :   Right
BCAST  :  1ring
DEPTH  :       0
SWAP   : Mix (threshold = 64)
L1     : transposed form
U      : transposed form
EQUIL  : yes
ALIGN  : 8 double precision words
```

Для тестирования узлов кластера с помощью оптимизированной версии LINPACK, использующей математическое ядро MKL, были применены настройки по умолчанию с различными вариантами размера матрицы и подматрицы:

```
PMAP      : Column-major process mapping
P         :        1
Q         :        1
PFACT     :    Crout
NBMIN     :        4
NDIV      :        2
RFACT     :    Crout
BCAST     :    Bpush
DEPTH     :        0
SWAP      : Binary-exchange
L1        : no-transposed form
U         : no-transposed form
EQUIL     : no
ALIGN     :     8 double precision words
```

Тесты были проведены для каждого из вычислительных узлов отдельно при размере задачи *N*, изменяющимся от 10000 до 38000 с шагом 2000, и размере подматрицы *NB*, изменяющимся от 32 до 256 с шагом 32. При размерах матрицы от 30000 и больше для размера подматрицы 256 элементов оценки производительности приближаются к теоретическому максимуму для вычислений с плавающей точкой двойной точности: 480 Гфлопс для 20 ядер. Результаты тестов представлены на графиках (рис. 5 и 6).

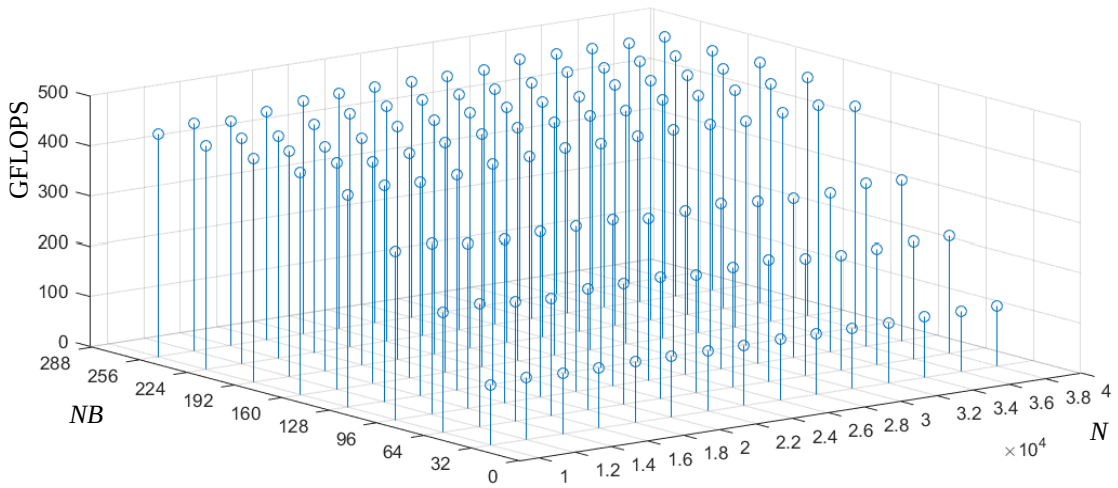

Рис. 5. Результаты оценки производительности в оптимизированном LINPACK для вычислительного узла 1

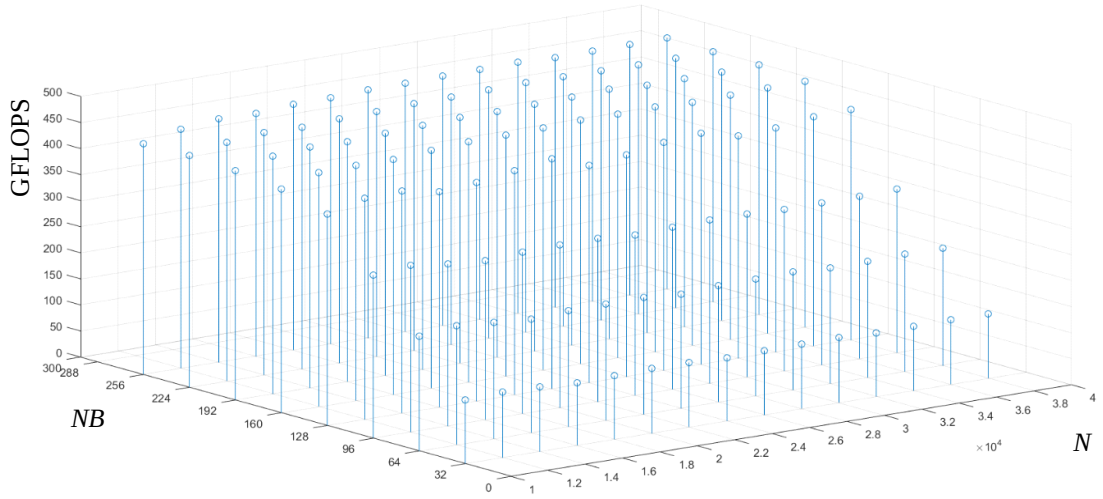

Рис. 6. Результаты оценки производительности в оптимизированном LINPACK для вычислительного узла 2

3.2.3. Тестирование производительности видеокарт гибридного вычислительного кластера на основе тестов SHOC

Тестирование возможностей ГВК с использованием видеокарт проводилось на основе тестового инструментария SHOC – The Scalable HeterOgeneous Computing [19, 20]. В данной работе представлены результаты тестирования ГВК с помощью программы MaxFlops, входящей в состав пакета SHOC. Программа MaxFlops служит для измерения максимально достижимой производительности при выполнении синтетических программных кодов, сформированных из различных комбинаций операций с плавающей точкой.

Наилучшие результаты среди видеокарт вычислительного узла 1 были получены для видеокарты 1 (рис. 7). Максимальная зафиксированная производительность в вычислениях одинарной точности составила 5202.83 Гфлопс, что довольно близко к теоретической (пиковой) производительности. В вычислениях двойной точности процессоры GPU GK110-Bl показали максимальный результат 255,55 Гфлопс. На рис. 7 приведены результаты для частных тестов, входящих в пакет MaxFlops. Использованы следующие обозначения:

SP – тесты вычислений с плавающей точкой одинарной точности;

DP – тесты вычислений с плавающей точкой двойной точности;

AddN – тесты, использующие *N* независимых потоков операций сложения;

MAddN – тесты, использующие *N* независимых потоков операций умножения-сложения;

MAddU – тест на основе вручную настроенного ядра, вычисляющего множественные операции умножения-сложения;

MulN – тесты, использующие *N* независимых потоков операций умножения;

MulMaddN – тесты, использующие N независимых потоков операций умножения и умножения-сложения;

MulMaddU – тест на основе вручную настроенного ядра, использующего смесь операций умножения и умножения-сложения.

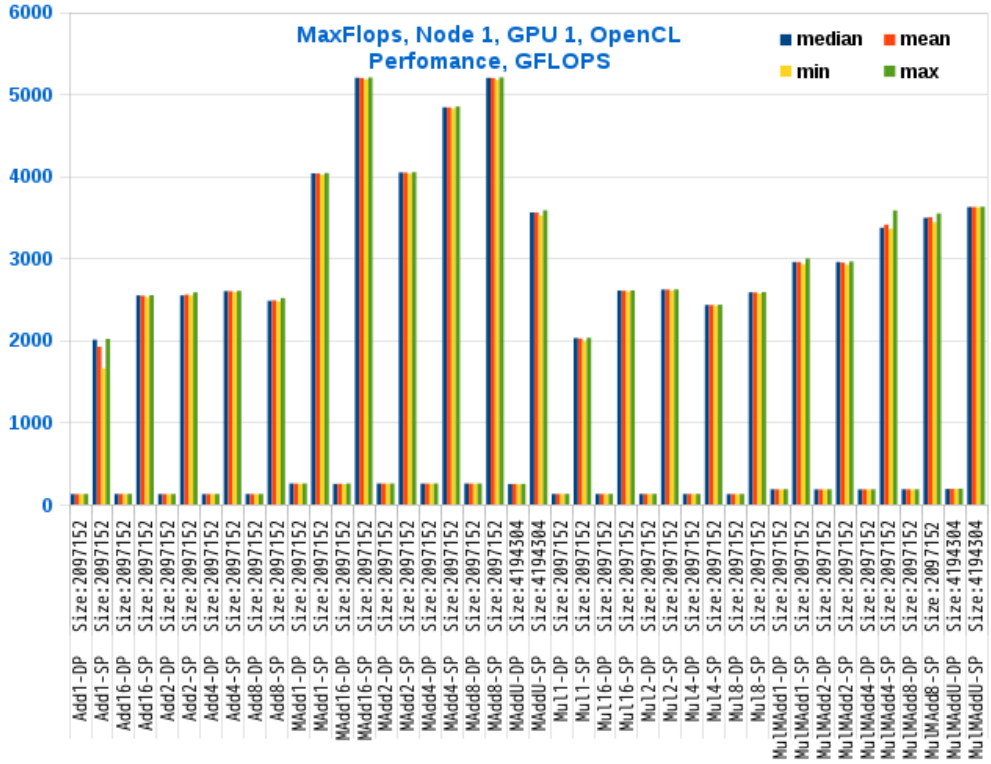

Рис. 7. Результаты теста MaxFlops для видеокарты 1 вычислительного узла 1

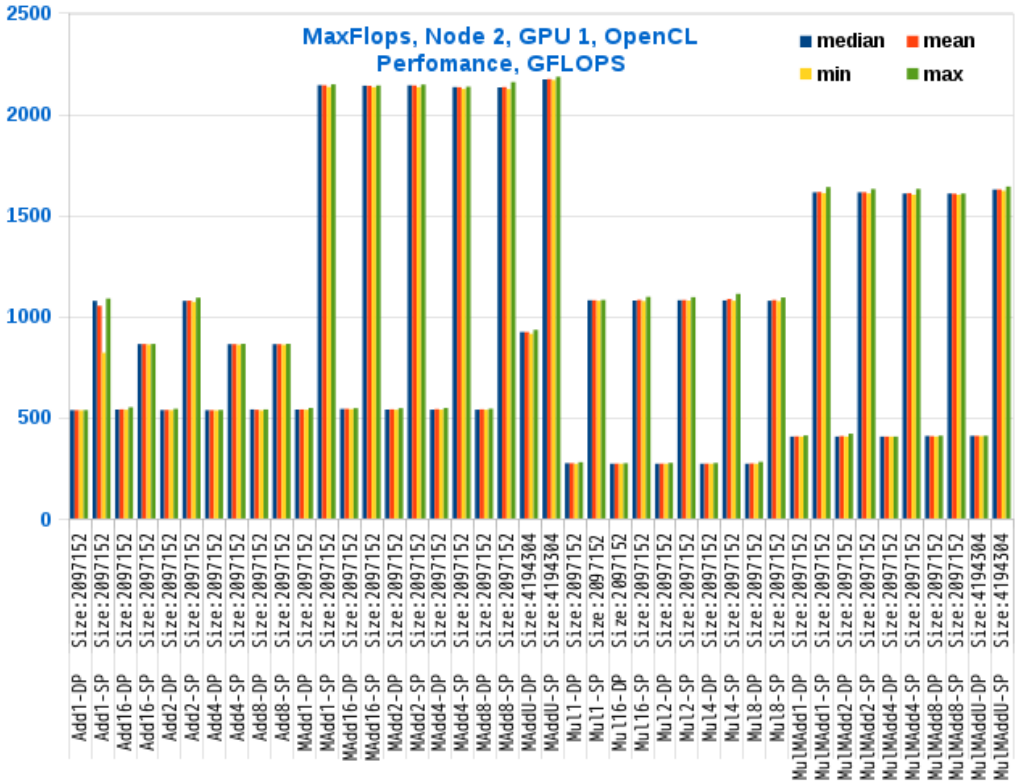

Рис. 8. Результаты теста MaxFlops для видеокарты 2 вычислительного узла 2

Максимальная зафиксированная производительность для видеокарт вычислительного узла 2 в вычислениях одинарной точности составила 4277,66 Гфлопс. Полученный результат близок к теоретической производительности. Так же близкий к теоретическому максимуму результат был получен и для вычислений двойной точности: 1072,39 Гфлопс. Подробно результаты теста представлены на рис. 8. Обозначения на рис. 8 аналогичны приведенным выше.

Заключение

В работе приведена архитектура гибридного вычислительного кластера филиала ФГБОУ ВО «НИУ «МЭИ» в г. Смоленске. Проведена оценка максимальной теоретической производительности. Выполнены оценки производительности на основе тестов.

Суммарная теоретическая производительность центральных процессоров вычислительных узлов ГВК в вычислениях с плавающей точкой одинарной точности составляет: 960 Гфлопс для каждого узла и 1920 Гфлопс в сумме.

Для вычислений двойной точности оценки теоретической производительности составляют: 480 Гфлопс для каждого узла и 960 Гфлопс в сумме.

В тесте LINPACK HPL 2.2 без оптимизации с использованием MKL, получена наилучшая оценка 101,7 Гфлопс.

В тесте LINPACK HPL 2.2 оптимизированном для процессоров Intel с использованием библиотеки MKL, оценки производительности отдельных вычислительных узлов подтверждают рассчитанный теоретический максимум для вычислений с плавающей точкой двойной точности.

Теоретическая производительность графических процессоров вычислительных узлов ГВК в вычислениях с плавающей точкой одинарной точности составляет: 2×5645 Гфлопс для видеокарт узла 1 (графический процессор NVIDIA GK-110-Bl) и 2×4384 Гфлопс для видеокарт узла 2 (графический процессор Tahiti XTL), 20058 Гфлопс в сумме для всего ГВК.

Для вычислений двойной точности оценки теоретической производительности составляют: 2×1881 Гфлопс для видеокарт узла 1 (графический процессор NVIDIA GK-110-Bl) и 2×1096 Гфлопс для видеокарт узла 2 (графический процессор Tahiti XTL), 5954 Гфлопс в сумме для всего ГВК.

В тесте SHOC 1.1.5 MaxFlops были получены следующие лучшие результаты:

- 5202,83 Гфлопс в вычислениях с плавающей точкой одинарной точности и 255,55 Гфлопс в вычислениях двойной точности среди видеокарт узла 1;
- 4277,66 Гфлопс в вычислениях с плавающей точкой одинарной точности и 1072.39 Гфлопс в вычислениях двойной точности среди видеокарт узла 2.

Представленные результаты планируется использовать для теоретического обобщения подхода к исследованию гибридных

вычислительных систем, ориентированных на решение разных нечисловых и вычислительных задач и различающихся архитектурой и используемым информационным обеспечением.

Литература

1. Intel Xeon Processor E5-2690 v2 // Intel [Электронный ресурс]. 20.11.2016. – URL: http://ark.intel.com (дата обращения 20.11.2016).

2. Intel Hyper-Threading Technology // Intel [Электронный ресурс]. 20.11.2016. – URL: http://www.intel.com (дата обращения 20.11.2016).

3. Чеканов Д. Технология Hyper-Threading от Intel // 3Dnews [Электронный ресурс]. 20.11.2016. – URL: http://www.3dnews.ru (дата обращения 20.11.2016).

4. Bradley T. Inside Kepler // NVIDIA [Электронный ресурс]. 20.11.2016. – URL: http://www.nvidia.ru (дата обращения 20.11.2016).

5. NVIDIA Kepler GK110 Next-Generation CUDA Compute Architecture // NVIDIA [Электронный ресурс]. 20.11.2016. – URL: http://www.nvidia.ru (дата обращения 20.11.2016).

6. Видеокарта GeForce GTX TITAN Black // NVIDIA [Электронный ресурс]. 20.11.2016. – URL: http://www.nvidia.ru (дата обращения 20.11.2016).

7. Семейства видеокарт AMD Radeon. Справочная информация // IXBT.com [Электронный ресурс]. 20.11.2016. – URL: http://www.ixbt.com (дата обращения 20.11.2016).

8. Тест и обзор: ASUS Radeon R9 280X DirectCU II TOP (R9280X-DC2T-3GD5) // Hardwareluxx.ru [Электронный ресурс]. 20.11.2016. – URL: http://www.hardwareluxx.ru (дата обращения 20.11.2016).

9. NUMизматика, NUMерология и просто о NUMA // Habrahabr.ru [Электронный ресурс]. 20.11.2016. – URL: https://habrahabr.ru (дата обращения 20.11.2016).

10. Optimizing Performance with Intel Advanced Vector Extensions // Intel [Электронный ресурс]. 20.11.2016. – URL: http://www.intel.ru (дата обращения 20.11.2016).

11. Intel Xeon E5-2600 v2: больше ядер, кэша и лучшая эффективность // Tom's Hardware [Электронный ресурс]. 20.11.2016. – URL: http://www.thg.ru (дата обращения 20.11.2016).

12. PCI Express 3.0: новый стандарт производительности и функциональности // Tom's Hardware [Электронный ресурс]. 20.11.2016. – URL: http://www.thg.ru (дата обращения 20.11.2016).

13. Обзор видеокарты Gigabyte GeForce GTX TITAN Black GHz Edition: «Титаново-чёрный» гигагерц // 3Dnews [Электронный ресурс]. 20.11.2016. – URL: http://www.3dnews.ru (дата обращения 20.11.2016).

14. AMD Radeon and NVIDIA GeForce FP32/FP64 GFLOPS Table // Geeks3D [Электронный ресурс]. 20.11.2016. – URL: http://www.geeks3d.com (дата обращения 20.11.2016).

15. ASUS R9 280X DirectCU II TOP // Techpowerup.com [Электронный ресурс]. 20.11.2016. – URL: https://www.techpowerup.com (дата обращения 20.11.2016).

16. Petitet A., Whaley R. C., Dongarra J., Cleary A. HPL – A Portable Implementation of the High-Performance Linpack Benchmark for Distributed-Memory Computers // Netlib.org [Электронный ресурс]. 20.11.2016. – URL: http://www.netlib.org (дата обращения 20.11.2016).

17. Dongarra J., Luszczek P., Petitet A. The linpack benchmark: past, present and future // Concurrency and Computation: Practice and Experience, 2003, vol. 15, no. 9, pp. 803–820.

18. Vipin Kumar E. K. HPL application note// Intel [Электронный ресурс]. 20.11.2016. – URL: https://software.intel.com (дата обращения 20.11.2016).

19. Vetter J.S. The Scalable HeterOgeneous Computing (SHOC) Benchmark Suite// Github.com [Электронный ресурс]. 20.11.2016. – URL: https://github.com (дата обращения 20.11.2016).

20. Danalis A., Marin G., McCurdy C., Meredith J., Roth P., Spafford K., Tipparaju V., Vette J. The Scalable HeterOgeneous Computing (SHOC) Benchmark Suite // In Proc. of the Third Workshop on General-Purpose Computation on Graphics Processors (GPGPU 2010), March 2010.

References

1. Intel Xeon Processor E5-2690 v2. *Intel*, 20 November 2016. Available at: http://ark.intel.com (accessed 20 November 2016).

2. Intel Hyper-Threading Technology. *Intel*, 20 November 2016. Available at: http://www.intel.com (accessed 20 November 2016).

3. Chekanov D. Tehnologija Hyper-Threading ot Intel [Chekanov D. Intel Hyper-Threading Technology]. *3Dnews*, 20 November 2016. Available at: http://www.3dnews.ru (accessed 20 November 2016) (in Russian).

4. Bradley T. Inside Kepler. *NVIDIA*, 20 November 2016. Available at: http://www.nvidia.ru (accessed 20 November 2016).

5. NVIDIA Kepler GK110 Next-Generation CUDA Compute Architecture. *NVIDIA*, 20 November 2016. Available at: http://www.nvidia.ru (accessed 20 November 2016).

6. Videokarta GeForce GTX TITAN Black [Graphics Card GeForce GTX TITAN Black]. *NVIDIA*, 20 November 2016. Available at: http://www.nvidia.ru (accessed 20 November 2016) (in Russian).

7. Semeistva Videokart AMD Radeon. Spravochnaja Informatsija [AMD Radeon Graphics Card Family. Reference Information]. *IXBT.com*, 20 November 2016. Available at: http://www.ixbt.com/video2/spravka-r2013.shtml (accessed 20 November 2016) (in Russian).

8. Test i Obzor: ASUS Radeon R9 280X DirectCU II TOP (R9280X-DC2T-3GD5) [Test and Review: ASUS Radeon R9 280X DirectCU II TOP (R9280X-DC2T-3GD5)]. *Hardwareluxx.ru*, 20 November 2016. Available at: http://www.hardwareluxx.ru (accessed 20 November 2016) (in Russian).

9. NUMizmatika, NUMerologija i Prosto o NUMA [NUMismatics, NUMerology and Just About NUMA]. *Habrahabr.ru*, 20 November 2016. Available at: https://habrahabr.ru (accessed 20 November 2016) (in Russian).

10. Optimizing Performance with Intel Advanced Vector Extensions. *Intel*, 20 November 2016. Available at: http://www.intel.ru (accessed 20 November 2016).

11. Intel Xeon E5-2600 v2: Bolshe Jader, Kjesha i Luchshaja Effectivnost [Intel Xeon E5-2600 v2: More Coreser, More Cache and Better Efficiency]. *Tom's Hardware*, 20 November 2016. Available at: http://www.thg.ru (accessed 20 November 2016) (in Russian).

12. PCI Express 3.0: Noviy Standart Proizvoditelnostji i Funkcionalnosti [PCI Express 3.0: New Standard of Performance and Functionality]. *Tom's Hardware*, 20 November 2016. Available at: http://www.thg.ru (accessed 20 November 2016) (in Russian).

13. Obzor Videokarti Gigabyte GeForce GTX TITAN Black GHz Edition: "Titanovo-Chjornyj" Gigagerc [Review of Graphic Card Gigabyte GeForce GTX TITAN Black GHz Edition: "Titan Black" GHz]. *3Dnews*, 20 November 2016. Available at: http://www.3dnews.ru (accessed 20 November 2016) (in Russian).

14. AMD Radeon and NVIDIA GeForce FP32/FP64 GFLOPS Table. *Geeks3D*, 20 November 2016. Available at: http://www.geeks3d.com (accessed 20 November 2016).

15. ASUS R9 280X DirectCU II TOP. *Techpowerup.com*, 20 November 2016. Available at: https://www.techpowerup.com (accessed 20 November 2016).

16. Petitet A., Whaley R. C., Dongarra J., Cleary A. HPL – A Portable Implementation of the High-Performance Linpack Benchmark for Distributed-Memory Computers. *Netlib.org*, 20 November 2016. Available at: http://www.netlib.org (accessed 20 November 2016).

17. Dongarra J., Luszczek P., Petitet A. The linpack benchmark: past, present and future // *Concurrency and Computation: Practice and Experience*, 2003, vol. 15, no. 9, pp. 803-820.

18. Vipin Kumar E. K. HPL application note. *Intel*, 20 November 2016. Available at: https://software.intel.com (accessed 20 November 2016).

19. Vetter J. S. The Scalable HeterOgeneous Computing (SHOC) Benchmark Suite. *Github.com*, 20 November 2016. Available at: https://github.com (accessed 20 November 2016).

20. Danalis A., Marin G., McCurdy C., Meredith J., Roth P., Spafford K., Tipparaju V., Vette J. The Scalable HeterOgeneous Computing (SHOC) Benchmark Suite. *In Proc. of the Third Workshop on General-Purpose Computation on Graphics Processors (GPGPU 2010)*, March 2010.

Статья поступила 05 декабря 2016 г.

Информация об авторах

Борисов Вадим Владимирович – доктор технических наук, профессор. Профессор кафедры вычислительной техники. Филиал «Национального исследовательского университета «Московский энергетический институт» в

г. Смоленске. Старший научный сотрудник научно-исследовательского центра. Военная академия войсковой противовоздушной обороны Вооруженных сил Российской Федерации им. А.М. Василевского. Области научных интересов: нечеткий и нейро-нечеткий анализ, моделирование сложных систем и процессов; интеллектуальная поддержка принятия решений; ассоциативные системы хранения и обработки информации. Тел.: +7 4812 65 14 61. E-mail: vadim.v.borisov@mail.ru

Зернов Михаил Михайлович – кандидат технических наук. Доцент кафедры вычислительной техники. Филиал «Национального исследовательского университета «Московский энергетический институт» в г. Смоленске. Области научных интересов: интеллектуальная поддержка принятия решений; интеллектуальный анализ данных; высокопроизводительная обработка данных. Тел.: +7 4812 65 14 61. E-mail: zmmioml@yandex.ru

Федулов Александр Сергеевич – доктор технических наук, профессор. Заведующий кафедрой вычислительной техники. Филиал «Национального исследовательского университета «Московский энергетический институт» в г. Смоленске. Области научных интересов: нечеткий когнитивный анализ и моделирование сложных систем и процессов; цифровая обработка сигналов; интеллектуальная поддержка принятия решений. Тел.: +7 4812 65 14 61. E-mail: fedulov_a@mail.ru

Якушевский Константин Андреевич – студент кафедры вычислительной техники. Филиал «Национального исследовательского университета «Московский энергетический институт» в г. Смоленске. Области научных интересов: высокопроизводительная обработка данных; параллельные вычисления. Тел.: +7 4812 65 14 61. E-mail: jakushevskij@gmail.com

Адрес: 214013, Россия, г. Смоленск, Энергетический проезд, д. 1.

Analysis of the Characteristics of Hybrid Computing Cluster

V. V. Borisov, M. M. Zernov, A. S. Fedulov, K. A. Yakushevsky

Purpose. *The development and application of hybrid computing systems is relevant in the context of increasing information, increasing the intensity and complexity of search and analytical tasks. These systems have the capacity to include them to the class supercomputers, while maintaining acceptable to users of price, size, ergonomic features. However, it should be noted shortage of articles devoted to the study of the characteristics of such systems depending on the nature of their architecture and software. The purpose of this work is the analysis the performance of hybrid computing cluster (in the example of the hybrid computing cluster of the branch of National Research University "Moscow Power Engineering Institute" in Smolensk).* ***Methods***. *Methods of analysis and designing of computing systems.* ***Novelty***. *The presented results can serve as a basis for the generalization of the approach to the study of hybrid computing systems, focused on solving various non-numerical and computational problems.* ***Practical relevance***. *The results can be used to analyze the characteristics of various hybrid computing systems, focused on solving different non-numerical and computational problems.*

Keywords: *hybrid computer system, CPU, GPU.*

Information about Authors

Vadim Vladimirovich Borisov – Dr. habil. of Engineering Sciences, Professor. Professor of the Dept of Computer Engineering. The Branch of National Research University "Moscow Power Engineering Institute" in Smolensk. Senior researcher. Military Academy of Army Air Defence A.M. Vasilevsky. Fields of research: fuzzy and fuzzy neural models and networks, intellectual decision-making support, associative memory, associative systems of storage and processing of the information and knowledge. Tel.: +7 4812 65 14 61. E-mail: vadim.v.borisov@mail.ru

Mikhail Mikhailovich Zernov – Ph.D. of Engineering Sciences. Associate Professor of the Dept of Computer Engineering. The Branch of National Research University "Moscow Power Engineering Institute" in Smolensk. Fields of research: intellectual decision-making support, intellectual decision-making, high-performance data processing. Tel.: +7 4812 65 14 61. E-mail: zmmioml@yandex.ru

Alexander Sergeevich Fedulov – Dr. habil. of Engineering Sciences, Professor. Head of the Dept of Computer Engineering. The Branch of National Research University "Moscow Power Engineering Institute" in Smolensk. Fields of research: fuzzy cognitive analysis and modeling of complex systems and processes; digital signal processing; intellectual decision-making. Tel.: +7 4812 65 14 61. E-mail: fedulov_a@mail.ru

Konstantin Andreevich Yakushevsky – Student of the Dept of Computer Engineering. The Branch of National Research University "Moscow Power Engineering Institute" in Smolensk. Fields of research: high-performance data processing, parallel computing. Tel.: +7 4812 65 14 61. E-mail: jakushevskij@gmail.com

Address: Russia, 214013, Smolensk, Energeticheskiy proezd, 1.