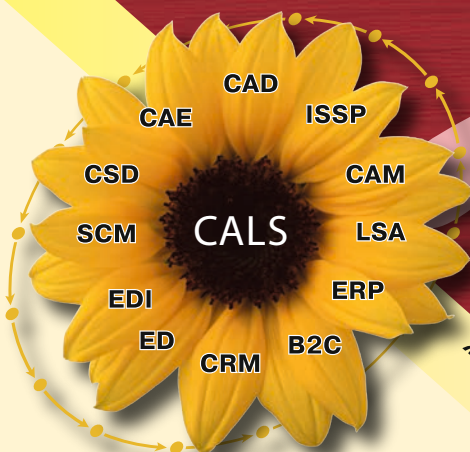


# КАЧЕСТВО

# ИННОВАЦИИ

# ОБРАЗОВАНИЕ

№ 11  
2015



журнал в журнале

КАЧЕСТВО и ИПИ (CALS)-технологии

[www.quality-journal.ru](http://www.quality-journal.ru)

КАЧЕСТВО  
ИННОВАЦИИ  
ОБРАЗОВАНИЕ

№11 (126)  
ноябрь 2015

СОДЕРЖАНИЕ

МЕНЕДЖМЕНТ КАЧЕСТВА И ИННОВАЦИОННЫЙ МЕНЕДЖМЕНТ

Ю.С. ШРАЙНЕР  
Формирование признаков модели зрелости  
и критериев их оценки на основе системного подхода ..... 3

О.А. ПЛУГАРЬ, И.А. АМИНОВА  
Эффективность и системы менеджмента ..... 10

ПРОБЛЕМЫ ПОДГОТОВКИ СПЕЦИАЛИСТОВ

Д.Р. ЮСИПОВА  
Форматирование научных публикаций в соответствии  
с профессиональной специализацией и с учетом мировой практики ..... 13

КАЧЕСТВО И ИПИ (CALS)-ТЕХНОЛОГИИ

ПРИБОРЫ, МЕТОДЫ И ТЕХНОЛОГИИ

В.В. ЖАДНОВ  
Методы повышения достоверности оценки  
межповерочных интервалов электронных измерительных приборов ..... 20

А.В. БЕЛЯКОВ, Ю.Э. ВАСИЛЬЕВ, Ш.Н. ВАЛИЕВ  
Совершенствование концепции автоматизации  
и методов применения передвижных дорожных диагностических машин ..... 28

А.Д. КАЛУЖСКИЙ  
Об оценке эффективности функционирования систем:  
метод дискретной эффектометрии ..... 38

СЕТЕВЫЕ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

Ю.Л. ЛЕОХИН, А.М. ЖИГУНОВ, Г.В. МОРОЗОВ  
Исследование эффективности решения прикладных задач  
с параллельными вычислениями на ПЛИС, CPU, GPU ..... 47

КАЧЕСТВО ЖИЗНИ

Р.С. ГОЛЬДБЕРГ  
Уточненная оценка экологических рисков нефтепереработки  
на основе байесовского статистического вывода ..... 59

ЭКОНОМИКА И УПРАВЛЕНИЕ

И.В. ДЕГТЯРЕВА, Г.Ф. ТОКАРЕВА, И.В. БАГАУТДИНОВА, О.И. ШАЛИНА  
Воспроизводство оборотного капитала российской экономики:  
современные проблемы и пути их решения ..... 64

ГЛАВНЫЙ РЕДАКТОР ОБЪЕДИНЕННОЙ  
РЕДАКЦИИ  
Азаров В.Н.

РЕДАКЦИОННАЯ КОЛЛЕГИЯ  
Алешин Н.П. (Москва), Батыров У.Д.  
(Нальчик), Бойцов Б.В. (Москва),  
Васильев В.А. (Москва), Васильев В.Н.  
(Санкт-Петербург), Гудков Ю.И.  
(Москва), Домрачев В.Г. (Москва),  
Карабасов Ю.С. (Москва), Кортюв С.В.  
(Екатеринбург), Лонцих П.А. (Иркутск),  
Львов Б.Г. (Москва), Мищенко С.В.  
(Тамбов), Олейник А.В. (Москва),  
Сергеев А.Г. (Владимир), Смакотина  
Н.Л. (Москва), Старых В.А. (Москва),  
Стриханов М.Н. (Москва), Тихонов А.Н.  
(Москва), Фирстов В.Г. (Москва), Харин  
А.А. (Москва), Червяков Л.М. (Курск),  
Шленов Ю.В. (Москва)

ЗАРУБЕЖНЫЕ ЧЛЕНЫ РЕДКОЛЛЕГИИ  
Андрес Каррион Гарсиа, Вагстафф  
П., Диккенсон П., Зайчек В., Иняц Н.,  
Кемпбелл Д., Папис Л., Пупиус М.,  
Роджерсон Д., Фарделф Д.

АДРЕС РЕДАКЦИИ И ИЗДАТЕЛЯ  
105118, Москва, ул. Буракова, д. 8  
Тел.: +7 (915) 001-31-49

E-mail: [quality@eqc.org.ru](mailto:quality@eqc.org.ru)  
[www.quality-journal.ru](http://www.quality-journal.ru)  
[www.quality21.ru](http://www.quality21.ru)

ИЗДАТЕЛЬ  
Европейский центр по качеству

НАУЧНЫЙ РЕДАКТОР  
Гудков Ю.И.  
[yugudkov@hse.ru](mailto:yugudkov@hse.ru)

ХУДОЖЕСТВЕННЫЙ РЕДАКТОР  
Кудрявцева А.И.

ЛИТЕРАТУРНЫЙ РЕДАКТОР  
Савин Е.С.

ОТВЕТСТВЕННЫЙ СЕКРЕТАРЬ  
Мартюкова Е.С.

ЖУРНАЛ ЗАРЕГИСТРИРОВАН  
в Министерстве РФ по делам печати,  
телерадиовещания и средств массовых  
коммуникаций. Свидетельство  
о регистрации ПИ №77-9092

ПОДПИСНОЙ ИНДЕКС  
в каталоге агентства «Роспечать» 80620,  
80621  
в каталоге агентства «Урал-Пресс» 14490  
на сайте НЭБ eLIBRARY.RU 80620

ОТПЕЧАТАНО  
ФГУП Издательство «Известия» УД ПРФ  
127254, г. Москва, ул. Добролюбова, д. 6

© «Европейский центр по качеству», 2014

Журнал входит в перечень ВАК РФ

Статьи рецензируются

Сведения о членах редколлегии и об авторах статей можно найти на сайте [www.quality-journal.ru](http://www.quality-journal.ru)

Ю.Л. Леохин, А.М. Жигунов, Г.В. Морозов

**ИССЛЕДОВАНИЕ ЭФФЕКТИВНОСТИ РЕШЕНИЯ  
ПРИКЛАДНЫХ ЗАДАЧ С ПАРАЛЛЕЛЬНЫМИ ВЫЧИСЛЕНИЯМИ  
НА ПЛИС, CPU, GPU**

В статье описана структура испытательного стенда для оценки эффективности решения прикладных задач с параллельными вычислениями на ПЛИС, CPU, GPU; представлены результаты оценки эффективности решения трех задач: поиск в потоке по шаблону, операции с большими матрицами, умножение матриц.

*Ключевые слова: гетерогенные серверы, параллельные вычисления, центральные процессоры, графические процессоры, программируемые логические интегральные схемы, испытательный стенд*

**В** настоящее время наблюдается переход на гибридные, гетерогенные архитектуры серверов, вычислительные модули которых состоят из универсальных процессоров (CPU), например, семейства x86, программируемых логических интегральных схем (ПЛИС) и графических процессоров (GPU). Использование гетерогенной архитектуры позволяет повысить эффективность решения задач, реализуемых в центрах обработки данных: web-приложения, поиск и индексация больших массивов данных, хостинг физических и виртуальных серверов, маршрутизация, параллельные и «облачные» вычисления.

Каждый из вычислительных модулей на базе CPU, GPU и ПЛИС имеет свои достоинства. Так, например, ПЛИС обладают большим вычислительным потенциалом, который в полной мере может быть реализован для задач вычислительного устройства. Перспективными являются архитектуры вычислительных модулей, в которых ПЛИС тесно связана с интерконнектом и используется и как маршрутизатор интерконнекта, и как средство поддержки базовых операций, и как ускоритель вычислений [1, 2, 3].

Вычислительные модули на базе CPU лучше выполняют последовательные задачи, но при большом объеме обрабатываемой информации, с которой можно работать параллельно, очевидное преимущество имеют вычислительные модули на базе GPU. Это связано с особенностями архитектуры CPU и GPU – архитектура CPU предполагает последовательную обработку информации, а GPU исторически предназначался для обработки компьютерной графики, поэтому рассчитан на массивно параллельные вычисления. Как показывают многочисленные испытания, пиковая производительность современного GPU в разы выше, чем производительность современного CPU. В скорости доступа к видеопамяти GPU также имеет значительное превосходство над CPU. Эффективная организация подсистемы памяти повышает общую эффективность графического процессора при работе с неграфическими задачами [4, 5, 6].

В рамках проекта «Разработка и организация высокотехнологичного производства энергоэффективных многопроцессорных аппаратно-программных серверных комплексов для государственных и корпоративных информационных систем и центров обработки данных» был разработан испытательный стенд для исследования эффективности решения прикладных задач с параллельными вычислениями на основе программируемых логических интегральных схем (ПЛИС), на центральных (CPU) и графических (GPU) процессорах (далее по тексту – испытательный стенд).

Испытательный стенд предназначен для проведения испытаний и отладки аппаратно-программных компонентов многопроцессорного аппаратно-программного серверного комплекса (МАПК), использующих ПЛИС, GPU и CPU.

Стенд включает в себя следующие аппаратные (таблица 1) и программные компоненты:

Таблица 1. Аппаратные компоненты испытательного стенда и их характеристики

№ п/п	Наименование и комплектация (сайт производителя)	Кол-во	Основные характеристики
1.	Процессор Intel® Core™ i7-5820K Haswell-E (3300MHz, LGA2011-3, L3 15360Kb)  <a href="http://www.intel.ru/buy/ru/ru/product">http://www.intel.ru/buy/ru/ru/product</a>	1	Socket LGA2011-3 Количество ядер – 6 Техпроцесс – 22 нм Тактовая частота – 3300 МГц Системная шина – DMI Коэффициент умножения – 35 Встроенный контроллер памяти, полоса 68 Гб/с Тепловыделение – 140 Вт Максимальная рабочая температура – 66.8 °C Объем кэша L2/L3: 1536 Кб/15360 Кб
2	Материнская плата GIGABYTE GA-X99-UD4 (rev. 1.0)  <a href="http://www.gigabyte.ru/products/page/mb/ga-x99-ud4rev_10">http://www.gigabyte.ru/products/page/mb/ga-x99-ud4rev_10</a>	1	Форм-фактор – ATX Socket – LGA2011-3 Чипсет – Intel X99 Память – DDR4 DIMM, включая RDIMM, 1333 – 3000 МГц Количество слотов памяти – 8 Количество разъемов SATA 6Gb/s: 10, RAID: 0, 1, 5, 10 на основе Intel X99 Слоты расширения – 4xPCI-E x16, 3xPCI-E x1 Интерфейсы: 16 USB, из них 8 USB 3.0 (6 на задней панели), выход S/PDIF, Ethernet, PS/2 (клавиатура), PS/2 (мышь) Ethernet – 1000 Мбит/с
3	Оперативная память 64ГБ (модуль – Corsair CMK8GX4M1A2400C14)  <a href="http://www.corsair.com/ru-ru/vengeance-lpx-8gb-1x8gb-ddr4-dram-2400mhz-c14-memory-kit-red-cmk8gx4m1a2400c14r">http://www.corsair.com/ru-ru/vengeance-lpx-8gb-1x8gb-ddr4-dram-2400mhz-c14-memory-kit-red-cmk8gx4m1a2400c14r</a>	8	Тип памяти – DDR4 Форм-фактор – DIMM 288 Тактовая частота – 2400 МГц Пропускная способность – 19200 Мб/с Объем – 1 модуль 8 Гб Напряжение питания – 1.2 В
4	Графический процессор (видеокарта ASUS GeForce GTX 750 Ti 1124Mhz PCI-E 3.0 2048Mb 5400Mhz 128 bit DVI HDMI HDCP)  <a href="http://www.asus.com/ua/Graphics_Cards/STRIXGTX750TI0C2GD5/specifications/">http://www.asus.com/ua/Graphics_Cards/STRIXGTX750TI0C2GD5/specifications/</a>	1	Графический процессор – NVIDIA GeForce GTX 750 Ti Интерфейс – PCI-E 16x 3.0 Техпроцесс – 28 нм Максимальное разрешение – 4096x2160 Частота графического процессора – 1124 МГц Объем видеопамяти – 2048 Мб Тип видеопамяти – GDDR5 Частота видеопамяти – 5400 МГц Разрядность шины видеопамяти – 128 бит Частота RAMDAC – 400 МГц Число универсальных процессоров – 640 Версия шейдеров – 5.0 Число текстурных блоков – 40 Число блоков растеризации – 16 Максимальная степень анизотропной фильтрации – 16x Максимальная степень FSAA – 32x Поддержка стандартов DirectX 11.2, OpenGL 4.4 Размеры 196x122 мм

5	SSD жесткий диск Samsung MZ-75E500BW (SSD жесткий диск 500ГБ – всего 1ТБ) <a href="http://www.samsung.com/in/business/business-products/ssd/ssd-850-evo/MZ-75E500BW">http://www.samsung.com/in/business/business-products/ssd/ssd-850-evo/MZ-75E500BW</a>	2	Объем – 250 Гб Скорость записи/Скорость чтения – 520/540 Мб/с Скорость случайной записи (блоки по 4Кб) 88000 IOPS Объем буферной памяти – 512 Мб Подключение – SATA 6Gbit/s Внешняя скорость передачи данных – 600 Мб/с
6	Компьютерный корпус Full-Desktop <a href="http://ru.fractal-design.com/home/product/cases/node-series/node-605">http://ru.fractal-design.com/home/product/cases/node-series/node-605</a>	1	Форм-фактор ATX, mATX, Mini-ITX Максимальная высота процессорного кулера 125 мм Встроенные вентиляторы 2×120×120 мм Максимальная длина видеокарты 290 мм Габариты (Ш×В×Г) – 445×164×349 мм Вес 6 кг Разъемы на лицевой панели: USB 3.0 x2, FireWire, наушники, микрофон
7	Блок питания Chieftec BPS-850C2 850W <a href="http://www.oldi.ru/catalog/element/0186815/">http://www.oldi.ru/catalog/element/0186815/</a>	1	Мощность 850 Вт Стандарт ATX12V 2.3 Система охлаждения – 1 вентилятор Ток по линии +3.3 В = 25 А Ток по линии +5 В = 25 А Ток по линии +12 В = 70 А Ток по линии -12 В = 0.8 А Ток по линии +5 В Standby = 3 А Размеры (Д×Ш×В) – 160×150×87 мм
8	Система охлаждения для центрального процессора Thermalright TRUE Spirit 120M <a href="http://www.thermalright.com/html/products/cpu_cooler/true_spirit_120i.html">http://www.thermalright.com/html/products/cpu_cooler/true_spirit_120i.html</a>	1	Socket: S775, S1150/1155/S1156, S1356/S1366, S2011, AM2, AM2+, AM3/AM3+/FM1, FM2/FM2+ 1 вентилятор 120 мм скорость 700-1500 об/мин радиатор из алюминия и меди уровень шума 17-28.5 дБ Размеры вентилятора (Д×Ш×В) – 120×120×25 мм Размеры кулера (Ш×В×Г) – 132×145×79 мм Скорость вращения 700 – 1500 об/мин
9.	Система бесперебойного питания (ИБП APC Smart-UPS SUA1500I) <a href="http://www.apc.com/products/">http://www.apc.com/products/</a>	1	Максимальная выходная мощность 980 Ватт / 1500 ВА Максимальное задаваемое значение мощности 980 Ватт / 1500 ВА Номинальное выходное напряжение 230V Диапазон входного напряжения при работе от сети 160 – 286 В Диапазон регулировки входного напряжения при работе от сети 151 – 302 В Типовая продолжительность работы в автономном режиме половинной нагрузки – 23.9 минуты (490 Ватт) Типовая продолжительность работы в автономном режиме полной нагрузки – 6.7 минуты (980 Ватт)

10.	<p>Отладочная система для ГПУ Tegra K1 (NVIDIA Jetson TK1)</p> <p><a href="https://developer.nvidia.com/jetson-tk1">https://developer.nvidia.com/jetson-tk1</a></p> <p><a href="http://labs.beatcraft.com/en/index.php?CUDA6%2FJetson">http://labs.beatcraft.com/en/index.php?CUDA6%2FJetson</a></p>	1	<p>GPU с 192 ядрами CUDA</p> <p>4-х ядерный Cortex A15 процессор</p> <p>2 Гб памяти DDR3L</p> <p>4.51 16 Гб eMMC память</p> <p>порты ввода-вывода, SPI, I2C, UART, HDMI, SATA, LVDS</p> <p>Габариты: 120×120 мм.</p>
11.	<p>ПЛИС-модуль (компания PLDA, версия XpressGX5LP)</p> <p><a href="https://www.plda.com/plda-announces-xpressgx5lp#sthash.E9nG2JEL.dpuf">https://www.plda.com/plda-announces-xpressgx5lp#sthash.E9nG2JEL.dpuf</a></p>	1	<p>Поддержка PCIe 3.0 at Gen3, Gen2, Gen1 speeds, in x1, x2, x4, x8</p> <p>QSFP+ enabling 40Gb Ethernet 4 x 10GbE</p> <p>PLDA's QuickPCIe IP core</p> <p>Не требуется внешний источник питания</p>
12.	<p>Система на кристалле Terasic DE1-SoC REV.C (с блоком питания и MicroSD картой объемом 8Гб)</p> <p><a href="http://www.terasic.com.tw/cgi-bin/page/archive.pl?CategoryNo=167&amp;Language=English&amp;No=836">http://www.terasic.com.tw/cgi-bin/page/archive.pl?CategoryNo=167&amp;Language=English&amp;No=836</a></p>	1	<p>Микросхема Cyclone V SoC 5CSEMA5F31C6;</p> <p>Двухъядерный Cortex-A9 (HPS)</p> <p>Память:</p> <p>64MB (32Mx16) SDRAM on FPGA</p> <p>1GB (2x256Mx16) DDR3 SDRAM on HPS</p> <p>Micro SD Card Socket on HPS</p> <p>Два порта USB 2.0 Host (ULPI интерфейс с USB разъемом типа A);</p> <p>USB-UART (разъем microUSB typeB);</p> <p>10/100/1000 Ethernet;</p> <p>Два 40-pin разъема расширения, один 10-pin разъем ADC;</p> <p>Один LTC разъем (один последовательный периферийный интерфейс (SPI) Master, один I2C и один GPIO интерфейс);</p> <p>24-bit CODEC, линейный вход, линейный выход, вход микрофона;</p> <p>TV Decoder (NTSC/PAL/SECAM) и входной TV разъем;</p> <p>АЦП: разрядность 12 бит, 8 каналов, номинальная скорость выборки 1 MSPS;</p> <p>11 светодиодов пользователя (FPGAx10; HPSx1);</p> <p>шестиразрядный семисегментный дисплей;</p> <p>входной диапазон аналогового сигнала 0 ~ 2.5 V или 0 ~ 5V;</p> <p>12 V DC разъем питания</p>

Программная часть стенда собрана на современном ядре Linux и включает следующее:

- 1) ОС CentOS 7 в конфигурации "Developers", либо другие версии Linux;
- 2) ОС Windows 7 pro 64;
- 3) кросс-компилятор arm-linux-gnueabihf для сборки приложений под Terasic DE1-SoC REV.C;
- 4) инструментарий разработчика Altera SDK for OpenCL в составе инструментария Altera Quartus версий 13.1 для PLDA xpressgx5lp и версии 14.0 для сборки под Terasic DE1-SoC REV.C;
- 5) лицензионный сервер для ПО quartus;
- 6) пакет поддержки модуля BSP для PLDA xpressgx5lp;
- 7) пакет поддержки модуля BSP для Terasic DE1-SoC REV.C;
- 8) образ Linux с BSP для Terasic DE1-SoC REV.C;
- 9) OpenSSH сервер для удаленной сборки ПО.

**Общий вид стенда**

На рисунке 1 представлен стенд для исследования эффективности решения прикладных задач с параллельными вычислениями на ПЛИС, CPU, GPU.



Рис. 1. Стенд для исследования эффективности решения прикладных задач с параллельными вычислениями на ПЛИС, CPU, GPU

Элементы стенда:

- 1) терминал доступа к плате DE1-SoC;
- 2) терминал по работе с платой PLDA;
- 3) PCI-E плата PLDA xpressgx5lp;
- 4) система на кристалле Terasic DE1-SoC REV.C

Отладочная система для ГПУ Tegra K1 показана на рисунке 2.



Рис. 2. Отладочная плата ГПУ Tegra K1





В результате работы программ получены следующие гистограммы:

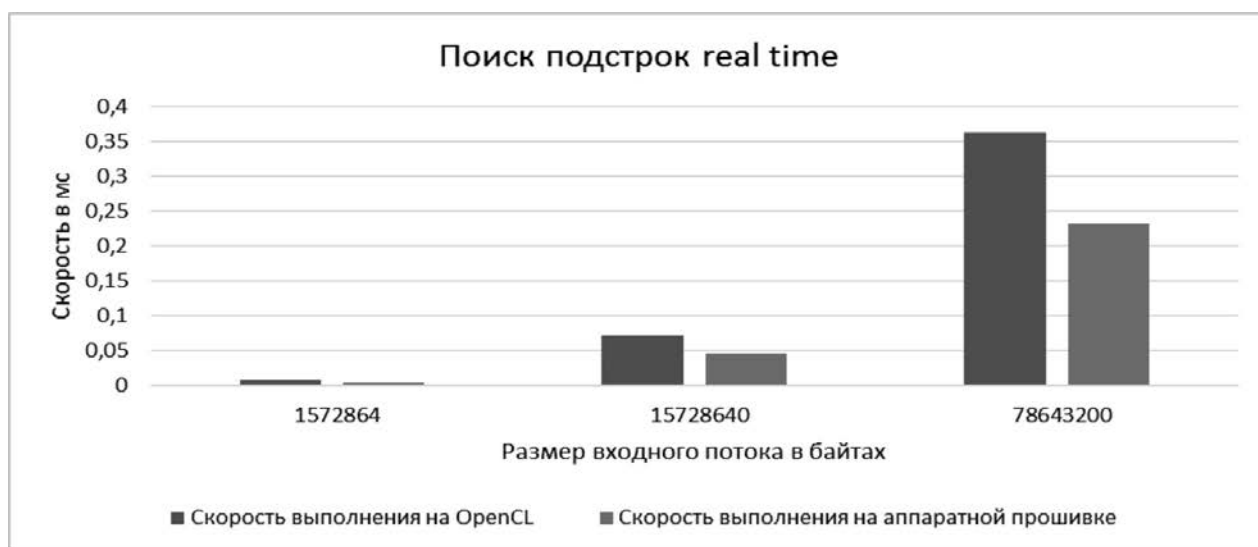


Рис. 4. Поиск по шаблону, реализованный на разных платформах

Анализ результатов показывает, что скорость решения задачи при входных потоках больших объемов в 1,5 раза выше для аппаратной реализации алгоритма, чем для реализации алгоритма на языке высокого уровня OpenCL. При этом, реализация алгоритма решения задачи на OpenCL показала также хорошую производительность на входных потоках данных меньшего объема, что говорит об эффективной оптимизации OpenCL.

**Операции с большими матрицами**

В данном тесте используется одна из оптимизированных под ПЛИС функций задачи по нахождению оптимальных поверхностей в многомерном векторном пространстве, описанных на третьем этапе, а именно, функция вычисления корреляции двух временных рядов.

```
rifle@GPU-station:~/projects/inventum/build$ bin/fpga_compute --
kernelfile=cov_pipeline.aocx --datasize=1048576
```

Листинг 2. Запуск программы для расчета корреляции временных рядов, размерностью 1 млн. элементов.

```
Platform: Altera SDK for OpenCL
Using 1 device(s)
xpressgx5lp_a7 : XpressGX5LP
Using AOCX: cov_pipeline.aocx
datasize 1048576
Time: 3.48306 ms
Kernel time: 0.032057 ms
CPU time: 4.41551 ms
FPGA result: [-0.933629]
CPU result: [0.0314513]
```

Листинг 3. Вывод в консоль

```
rifle@GPU-station:~/projects/inventum/build$ bin/fpga_compute --kernelfile=cov_pipeline.
aocx --datasize=2048
```

Листинг 4. Запуск программы для расчета ряда размерностью 2048 элементов

```
Platform: Altera SDK for OpenCL
Using 1 device(s)
xpressgx5lp_a7 : XpressGX5LP
Using AOCX: cov_pipeline.aocx
datasize 2048
Time: 0.355005 ms
Kernel time: 0.045746 ms
CPU time: 0.0245571 ms
FPGA result: [-0.933629]
CPU result: [0.929363]
```

Листинг 5. Вывод в консоль

```
rifle@GPU-station:~/projects/inventum/build$ bin/fpga_compute --
kernelfile=cov_pipeline.aocx
```

Листинг 6. Запуск программы для расчета ряда размерностью 256 элементов

```
Platform: Altera SDK for OpenCL
Using 1 device(s)
xpressgx5lp_a7 : XpressGX5LP
Using AOCX: cov_pipeline.aocx
datasize 256
Time: 0.343561 ms
Kernel time: 0.046793 ms
CPU time: 0.00405312 ms
FPGA result: [-0.933629]
CPU result: [-0.93363]
```

Листинг 7. Вывод в консоль

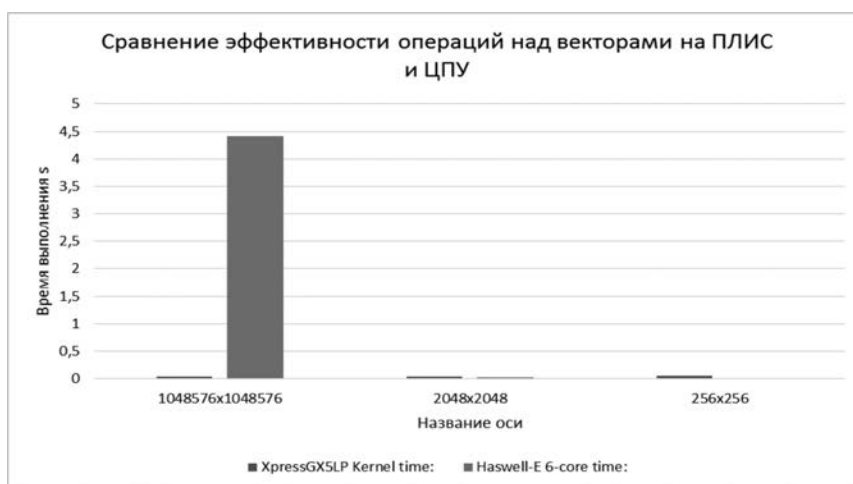


Рис. 4. Сравнение результатов векторных вычислений на ПЛИС и CPU

В данном примере продемонстрировано значительное преимущество параллельных вычислений на ПЛИС относительно CPU. Это преимущество особенно заметно при работе с очень большими векторами, размерностью, примерно, в 1 миллион элементов. При средних размерах векторов результаты отличаются на порядок, при малых размерах – соизмеримы. Объясняется это тем, что при небольших размерах данных главные накладные расходы приходятся не на вычисления, а на поблочную передачу данных.

### Умножение матриц

На примере умножения матриц, замерим производительность различных аппаратных ускорителей.

```
rifle@GPU-station:~/NVIDIA_CUDA-7.0_Projects/bin/x86_64/linux/release$ ./p2pBandwidthLatencyTest
[P2P (Peer-to-Peer) GPU Bandwidth Latency Test]
Device: 0, GeForce GTX 750 Ti, pciBusID: 4, pciDeviceID: 0, pciDomainID:0
***NOTE: In case a device doesn't have P2P access to other one, it falls back to normal memcopy
procedure.
So you can see lesser Bandwidth (GB/s) in those cases.
P2P Cliques:
[0]
Unidirectional P2P=Disabled Bandwidth Matrix (GB/s)
D\D  0
    0 38.34
Unidirectional P2P=Enabled Bandwidth Matrix (GB/s)
D\D  0
    0 37.25
Bidirectional P2P=Disabled Bandwidth Matrix (GB/s)
D\D  0
    0 38.39
Bidirectional P2P=Enabled Bandwidth Matrix (GB/s)
D\D  0
    0 37.78
P2P=Disabled Latency Matrix (us)
D\D  0
    0 3.62
P2P=Enabled Latency Matrix (us)
D\D  0
3.62
```

Листинг 8. Запуск умножения матриц на графических процессорах

```
[Matrix Multiply Using CUDA] – Starting...
GPU Device 0: "GeForce GTX 750 Ti" with compute capability 5.0
MatrixA(320,320), MatrixB(640,320)
Computing result using CUDA Kernel...
done
Performance= 198.36 GFlop/s, Time= 0.661 msec, Size= 131072000 Ops, WorkgroupSize= 1024
threads/block
Checking computed result for correctness: Result = PASS
-----
На терпе
[Matrix Multiply Using CUDA] – Starting...
GPU Device 0: "GK20A" with compute capability 3.2
MatrixA(320,320), MatrixB(640,320)
Computing result using CUDA Kernel...
done
Performance= 10.41 GFlop/s, Time= 12.588 msec, Size= 131072000 Ops, WorkgroupSize= 1024
threads/block
Checking computed result for correctness: Result = PASS
```

Листинг 9. Результат умножения матриц на графических процессорах

```

rifle@GPU-station:~/altera/mmul/matrix_mult$ aoc device/matrix_mult.cl -o bin/matrix_mult.aocx
-fp-relaxed=true -fpc=true --no-interleaving default --board xpressgx5lp_a7
Compiler Warning: Vectorized kernel contains loads/stores that cannot be vectorized. This might reduce
performance.
rifle@GPU-station:~/altera/mmul/matrix_mult$ ls
bin device host Makefile matrix_mult matrix_mult.aoco README.html
rifle@GPU-station:~/altera/mmul/matrix_mult$ cd bin/
rifle@GPU-station:~/altera/mmul/matrix_mult/bin$ ls
matrix_mult matrix_mult.aocx
rifle@GPU-station:~/altera/mmul/matrix_mult/bin$ ./matrix_mult
Matrix sizes:
A: 2048 x 1024
B: 1024 x 1024
C: 2048 x 1024
Initializing OpenCL
Platform: Altera SDK for OpenCL
Using 1 device(s)
  xpressgx5lp_a7 : XpressGX5LP
Using AOCX: matrix_mult.aocx
Generating input matrices
Launching for device 0 (global size: 1024, 2048)
Time: 38.648 ms
Kernel time (device 0): 38.580 ms
Throughput: 111.13 GFLOPS
Computing reference output
Verifying
Verification: PASS

```

Листинг 10. Запуск и результат умножения матриц на ПЛИС

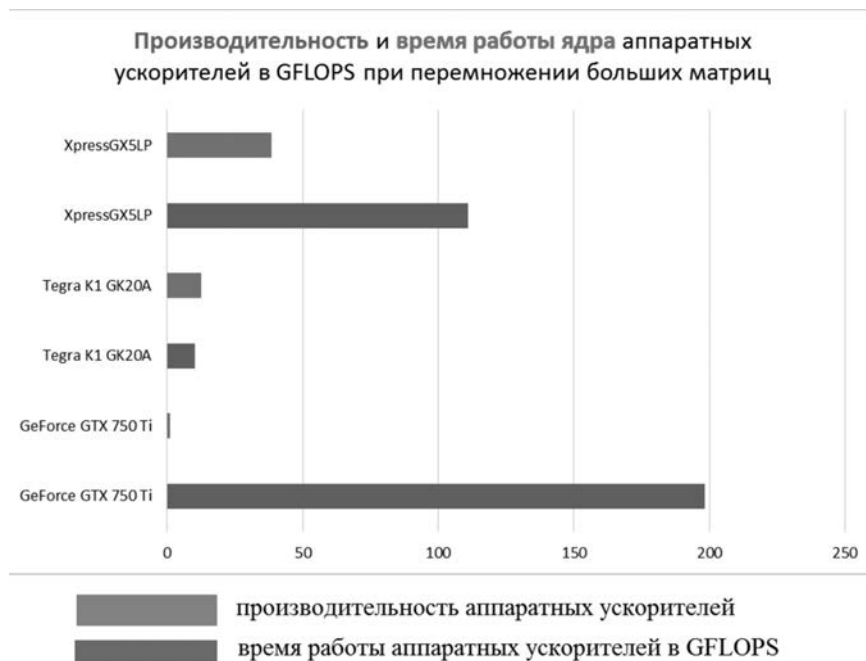


Рис. 5. Замеры производительности на операции умножения больших матриц

Исходный код программы был написан на Cuda для графических процессоров и на OpenCL – для ПЛИС. Результаты анализа замеров производительности показали (рис. 5), что наиболее

производительным ускорителем является GeForce GTX 750 Ti с 640 ядрами cuda. Менее мощный Soc Tegra K1 с 192 ядрами показал наименьшую производительность. ПЛИС XpressGX5LP показал средний результат ввиду того, что оптимизация под конкретный ПЛИС не проводилась. Причем компоненты матриц представляют собой числа с плавающей точкой, что снижает эффективность ПЛИС. В отличие от ПЛИС, графические ускорители очень эффективно работают с таким типом данных. Учитывая, что производительность ПЛИС XpressGX5LP и GeForce GTX 750 Ti одного порядка и вычисления на ПЛИС можно оптимизировать по нескольким параметрам, на ПЛИС можно достичь заметно большей эффективности работы с целочисленными данными.

**Заключение**

Результаты испытаний, проведенных с использованием разработанного стенда, показали эффективность использования гетерогенных аппаратных вычислительных сред для решения различных прикладных задач с параллельными вычислениями.

Написанные библиотеки для прикладных расчётов могут быть оптимизированы под вычислительные модули на базе различных типов процессоров, которыми будет укомплектован разработанный в рамках проекта многопроцессорный аппаратно-программный комплекс.

Статья написана по результатам выполнения проекта «Разработка и организация высокотехнологического производства энергоэффективных многопроцессорных аппаратно-программных серверных комплексов для государственных и корпоративных информационных систем и центров обработки данных», выигранного в рамках Открытого конкурса по отбору организаций на право получения субсидий на реализацию комплексных проектов по созданию высокотехнологического производства (Постановление Правительства РФ от 09 апреля 2010 г. №218).

**Литература:**

1. Панкин А.В. Метод параллельной обработки данных вычислительного устройства на основе ПЛИС FPGA и технологии «система на кристалле» ([http://www.rusnauka.com/4\\_SND\\_2013/Informatica/2\\_127968.doc.htm](http://www.rusnauka.com/4_SND_2013/Informatica/2_127968.doc.htm)).
2. Шагурин И., Шалтырев В., Волон А. «Большие» FPGA как элементная база для реализации систем на кристалле // Электронные компоненты. 2006. №5. С.83-88.
3. <http://www.botik.ru/PSI/RCMS/publications/publ-texts-2010/divn-abram-i-dr-2010.pdf>.
4. Джораев А.Р. Гибридные вычислительные системы на основе GPU для задач биоинформатики // Компьютерные исследования и моделирование. 2010. Т.2, № 2. С. 163–167.
5. Боресков А.В., Харламов А.А. Основы работы с технологией CUDA. – М.: ДМК Пресс, 2010. – 232 с.: ил.
6. NVIDIA's Next Generation CUDATM Compute Architecture: Fermi (Whitepaper) ([http://www.nvidia.com/content/PDF/fermi\\_white\\_papers/NVIDIA\\_Fermi\\_Compute\\_Architecture\\_Whitepaper.pdf](http://www.nvidia.com/content/PDF/fermi_white_papers/NVIDIA_Fermi_Compute_Architecture_Whitepaper.pdf)).

**Леохин Юрий Львович,**  
*Национальный исследовательский университет  
 «Высшая школа экономики»,  
 123458, Москва, ул. Таллинская, д. 34.  
 Зав. Учебно-исслед. лаб. высокопроизводит.  
 аппаратно-прогр. комплексов  
 и локальных вычисл. сетей МИЭМ НИУ ВШЭ.  
 Д-р техн. наук, профессор.  
 Тел.: 7 (495) 772-95-90.  
 e-mail: yleokhin@hse.ru*

**Жигунов Александр Михайлович,**  
*Национальный исследовательский университет  
 «Высшая школа экономики»,  
 101000, Москва, ул. Мясницкая, д. 20.  
 Научный сотрудник.  
 e-mail: gmorozov@hse.ru*

**Морозов Глеб Владимирович,**  
*Национальный исследовательский университет  
 «Высшая школа экономики»,  
 101000, Москва, ул. Мясницкая, д. 20.  
 Научный сотрудник.  
 e-mail: azhigunov@hse.ru*

Y.L. Leokhin, A.M. Jigunov, G.V. Morozov

### RESEARCH OF EFFICIENCY OF APPLIED TASKS WITH PARALLEL PROCESSING ON FPGA, CPU, GPU

The article describes the structure of the test bed for evaluating the effectiveness of applied tasks with parallel processing on FPGA, CPU, GPU, the results of evaluating the effectiveness of three objectives: the search for a pattern in the flow of operations with large matrices, matrix multiplication.

**Keywords:** *heterogeneous servers, parallel computing, CPUs, graphics processors, programmable logic integrated circuits, test bench*

#### References:

1. Pankin A.V. Metod parallelnoj obrabotki dannyh vychislitelnogo ustrojstva na osnove PLIS FPGA i tekhnologii sistema na kristalle ([http://www.rusnauka.com/4\\_SND\\_2013/Informatica/2\\_127968.doc.htm](http://www.rusnauka.com/4_SND_2013/Informatica/2_127968.doc.htm)).
2. Shagurin I., Shaltyrev V., Volov A. «Bolshie» FPGA kak elementnaya baza dlya realizacii sistem na kristalle//Elektronnye komponenty, 2006, №5, s.83—88.
3. <http://www.botik.ru/PSI/RCMS/publications/publ-texts-2010/divn-abram-i-dr-2010.pdf>.
4. A.R. Dzhoraev. Gibridnye vychislitelnye sistemy na osnove GPU dlya zadach bioinformatiki//Kompyuternye issledovaniya i modelirovanie, 2010. T. 2, № 2. S. 163–167.
5. Boreskov A.V., Harlamov A.A. Osnovy raboty s tekhnologiej CUDA. – M.: DMK Press, 2010. – 232 s.: il.
6. NVIDIA's Next Generation CUDATM Compute Architecture: Fermi (Whitepaper) ([http://www.nvidia.com/content/PDF/fermi\\_white\\_papers/NVIDIA\\_Fermi\\_Compute\\_Architecture\\_Whitepaper.pdf](http://www.nvidia.com/content/PDF/fermi_white_papers/NVIDIA_Fermi_Compute_Architecture_Whitepaper.pdf)).

**Leokhin Yuri L.**

*National Research University Higher School of Economics,  
Professor,  
yerokhin@hse.ru*

**Jigunov Alexander M.**

*National Research University Higher School of Economics,  
researcher,  
azhigunov@hse.ru*

**Morozov Gleb V.**

*National Research University Higher School of Economics,  
researcher,  
gmorozov@hse.ru*