

# Высокопроизводительная обработка потоков данных на многоядерных и графических процессорах

Кондратьев А.А.<sup>1</sup>, Хачумов В.М.<sup>2</sup>

<sup>1</sup> *Институт программных систем им. А.К.Айламазяна Российской академии наук, ул. Петра I, д. 4 «а», д. Вельхово, Переславский район, Россия*

<sup>2</sup> *Институт системного анализа Российской академии наук, ул. 60-летия Октября, д.9, Москва, Россия*

nonkajitsu@gmail.com, vmh48@mail.ru

**Аннотация.** *Целью работы является исследование методов и инструментальных программных средств обработки интенсивных потоков данных с космических аппаратов на высокопроизводительных гетерогенных вычислительных системах, построенных на основе универсальных многоядерных и графических процессоров. Приведены результаты экспериментальных исследований по обработке космических снимков.*

## Ключевые слова

Гетерогенная среда, многоядерные и графические процессоры, параллельные вычисления, обработка снимков, среда программирования CUDA

## 1 Введение

Вопросы анализа потоков данных с космических аппаратов (служебной телеметрии, данных дистанционного зондирования Земли (ДЗЗ), целевой информации) требуют привлечения значительных вычислительных мощностей. Наиболее эффективным в этом случае являются многопроцессорные вычислительные системы (МВС), включая кластерные установки (КВУ), многоядерные процессоры общего назначения и графические процессорные устройства (ГПУ). Важнейшим условием обеспечения конкурентоспособности в космической области является развитие программно-инструментальных средств и суперкомпьютерных технологий, как основы для технологического перевооружения космической отрасли. Новые решения стали возможны с появлением ускорителей (GPU и Intel MIC), расширивших возможности концепции массового параллелизма, ранее присущие только распределенным вычислительным системам. Сложившаяся ситуация требует разработки и развития инструментальных средств программирования вычислительных систем высокой и сверхвысокой производительности, оснащенных различными процессорами. Такие инструменты должны поддерживать разработку параллельных программ, как для универсальных многоядерных процессоров, так и для графических ускорителей различной архитектуры (Nvidia, AMD и Intel). В настоящей статье проведены некоторые результаты исследований возможности обработки космических снимков на КВУ СКИФ «VEDA», являющихся продолжением работ, начатых в [1]. Исследования инициированы ОАО «Российские космические системы» с целью оснащения научно-производственных процессов компании высокопроизводительными программно-аппаратными комплексами.

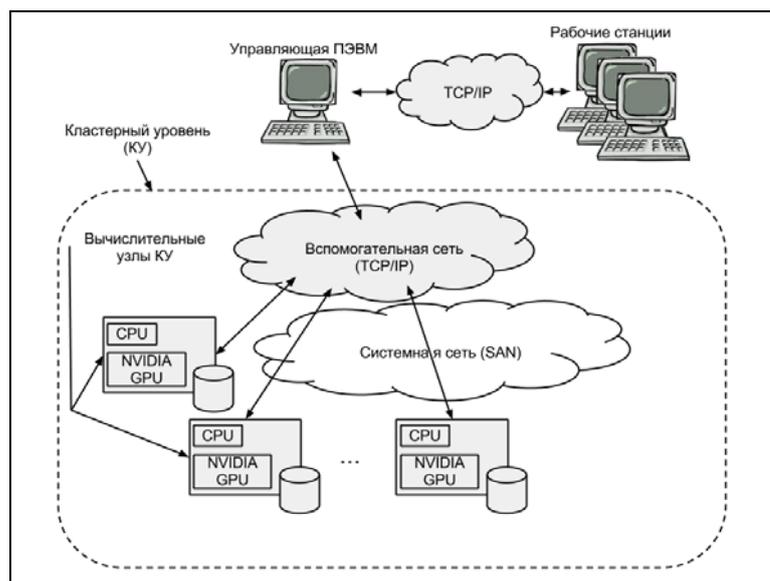
## 2 Базовая гетерогенная кластерная архитектура

В рамках научных программ ИСА РАН, ИПС РАН и РГААТУ проводятся совместные работы по исследованию возможностей гибридных (гетерогенных) вычислительных систем. Примером такой системы служит КВУ СКИФ «VEDA», установленная в ИПС РАН и служащая полигоном для отработки различных программно-инструментальных средств. Аппаратная часть комплекса представляет собой кластерную вычислительную установку, состоящую из вычислительных узлов, объединенных коммуникационной средой. Архитектура комплекса отражена в таблице 1 и на рисунке 1.

**Табл. 1.** Сводные аппаратные характеристики КВУ СКИФ «BEDA»

<i>Наименование характеристики</i>	<i>Значение</i>
Число вычислительных узлов	5
Конструктив узла	Full ATX
Тип процессора	четырёхъядерный Intel®Xeon® E5410 2,33 ГГц, кэш второго уровня 2x6 Мб для каждой пары ядер
Тактовая частота внешней шины	1333 МГц
Пиковая производительность	74.56 Гфлопс x 5 узлов
Тип системной сети	DDR InfiniBand (ConnectXT IB HCA Card)
Максимальная скорость передачи данных сети InfiniBand	20 Гбайт/с
Тип вспомогательной сети	Gigabit Ethernet
Оперативная память	32 Гб
Потребляемая мощность	не более 5 Квт
Дисковая память узлов	25 Тбайт
Система охлаждения	воздушная: 3x90 мм, 2x80, 2x120

В качестве базовой классической системы поддержки параллельных вычислений выбран интерфейс MPI, на базе которого реализована и успешно используется система поддержки конвейерно-параллельных вычислений «ППС ИНС» [2]. В последние годы все более популярной становится идея использования специализированных вычислителей для проведения расчетов общего назначения. К данному направлению можно отнести GPGPU computing, подразумевающий использование аппаратных ресурсов видеокарты для проведения расчетов, или более общий стандарт OpenCL (Open Computing Language), предусматривающий одновременное использование ресурсов CPU и GPU различных архитектур. С целью повышения эффективности комплекса в каждый из узлов устанавливается графический процессор (GPU) TeslaTM C1060, среда программирования: C(CUDA). Имеющаяся гетерогенная среда содержит компоненты, достаточные для решения задач высокопроизводительной обработки потоков данных.



**Рис.1.** Архитектура гибридной вычислительной системы

Среда является гибкой и наращиваемой, что может послужить основой построения программно-инструментальных средств и прикладных систем широкого назначения.

### 3 Экспериментальная обработка потоков космических данных

GPGPU («GPU общего назначения») — техника использования графического процессора видеокарты, который обычно имеет дело с вычислениями только для компьютерной графики, для выполнения расчётов в приложениях для общих вычислений, которые обычно проводит центральный процессор. Существует большое количество различных реализаций библиотек, использующих ресурсы GPU для вычислений: AMD FireStream, CUDA, DirectCompute, OpenCL. По сравнению с традиционным подходом к организации вычислений общего назначения посредством возможностей графических API, у архитектуры CUDA отмечают следующие преимущества в этой области:

- Интерфейс программирования приложений CUDA (CUDA API) основан на стандартном языке программирования Си с некоторыми ограничениями. По мнению разработчиков, это должно упростить и сгладить процесс изучения архитектуры CUDA;
- Разделяемая между потоками память (shared memory) размером в 16 Кб может быть использована под организованный пользователем кэш с более широкой полосой пропускания, чем при выборке из обычных текстур;
- Более эффективные транзакции между памятью центрального процессора и видеопамятью;
- Полная аппаратная поддержка целочисленных и побитовых операций;
- Поддержка компиляции GPU кода средствами открытого LLVM.

Указанные преимущества подтверждаются экспериментальными исследованиями, которые решают задачу оконтуривания (выделения границ) на космических изображениях фильтром Собеля. Алгоритм обработки потоков данных выполняется на модульной программной системе ППС ИНС, функционирующей на КВУ СКИФ «BEDA» и содержит следующую схему:

- 1) Модуль Objects (readpng) считывает предназначенные для обработки фильтром изображения и передает их модулю Sobel.
- 2) Модуль Sobel (edgedetectionsobel) получает изображения, обрабатывает их согласно xml-файлу настроек и передает данные модулю Writer.
- 3) Модуль Writer (savepng) сохраняет полученные от модуля Sobel обработанные изображения в указанную в своих настройках директорию.

Пример обрабатываемого спутникового снимка Земли, представлен на рисунке 2.



**Рис.2.** Исходный космический снимок

Результат фильтрации изображения представлен на рисунке 3.



Рис.3. Результат фильтрации по Собелю

#### 4 Сравнение результатов фильтрации на CPU и GPU

Тестирование программ, реализующих алгоритм фильтрации, проводилось на космических снимках различного размера. Параметры некоторых исходных изображений указаны в таблице 2. Для получения сравнительных характеристик решения задачи фильтрации на GPU и CPU был реализован механизм конвейерной фильтрации изображений. В процессе тестирования измерялось время непосредственно обработки, без учета времени сохранения и передачи данных.

Табл.2. Характеристики обрабатываемых снимков

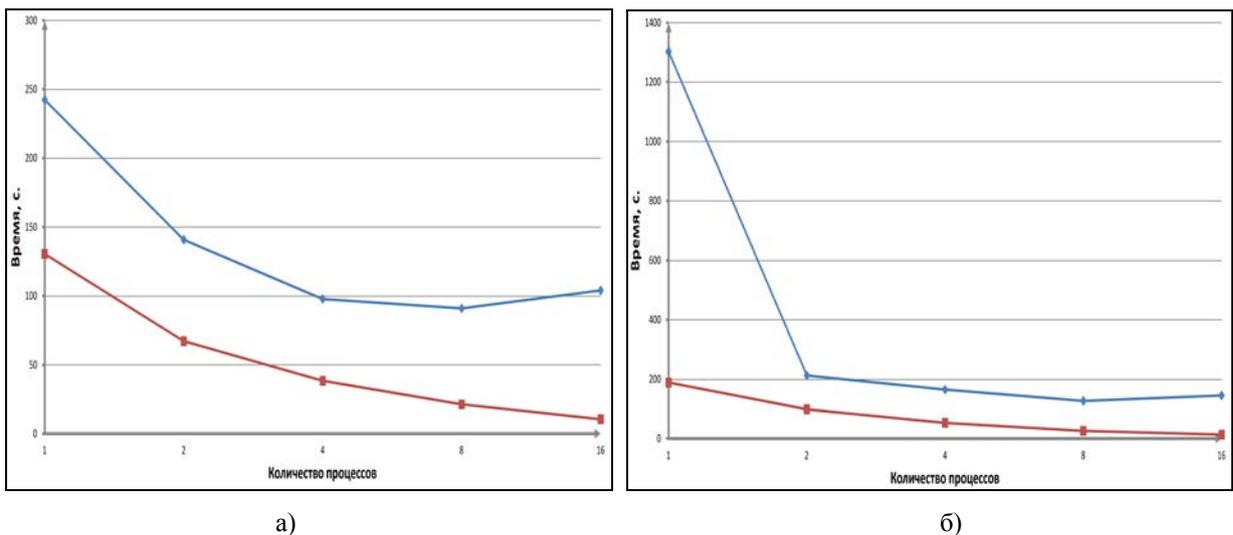
<i>Номер изображения</i>	<i>Ширина, пикс.</i>	<i>Высота, пикс.</i>
1	12150	6075
2	7160	4568
3	5600	7200
4	5134	4947
5	8218	7608
6	5104	4608
7	6750	5246
8	4702	4693
9	3544	5608
10	7381	4147
11	5461	4096
12	9002	3196

Из исходных изображений был сформирован тестовый пакет, в который были включены 10 копий исходных изображений (всего 120 изображений). Тестирование проходило на трех узлах кластерной установки, оснащенных GPU. Для сравнения были проведены тесты производительности подобной реализации на CPU. Результаты тестирования представлены в таблице 3 и на рисунке 4.

**Табл.3.** Характеристики обработки снимков на GPU и CPU

<i>Количество параллельных процессов</i>	<i>Время решения задачи (чтение и обработка), на GPU, сек.</i>	<i>Время обработки одного процесса на GPU (среднее), сек.</i>	<i>Время решения задачи (чтение и обработка) на CPU, сек.</i>	<i>Время обработки одного процесса на CPU (среднее), сек.</i>
1	242.4	130.5	1302.8	189.2
2	140.8	67.2	212.9	98.7
4	97.8	38.4	165	52.9
8	91	21.3	126.7	26.1
16	104	10.4	145.2	13.6

При наличии в системе GPU, поддерживающего OpenCL, его использование заметно сокращает время обработки изображений. Кроме того, освобождается процессор, занятый до этого непосредственно обработкой данных. В частности, при больших данных и длительных расчетах на GPU появляется возможность использовать CPU, что дает заметную прибавку к производительности. Накладные расходы во время тестирования могут быть весьма высокими и в какой-то момент перекрывать ускорение, получаемое за счет GPU. Алгоритм для решения задач на GPU существенно отличается от алгоритма для CPU, что мешает быстрому переносу на них готовых решений. По сравнению с CPU операции имеют более низкую точность, что ограничивает область возможного применения GPU.



**Рис.4** Результат фильтрации по Собелю

(а - обработка снимков на GPU, б) обработка снимков на CPU)

Производительность на GPU падает в программах, где много обращений к памяти (особенно к соседним потокам) и мало операций, которые могут выполняться параллельно.

## 5 Заключение

Исследования направлены на разработку и исследование базового программного обеспечения поддержки параллельных вычислений с применением ускорителей, Решение перечисленных проблем требует проведения фундаментальных поисковых исследований.

К ожидаемым результатам проводимой исследовательской работы можно отнести.

- 1) архитектуру вычислительного ядра с поддержкой конвейерно-параллельной работы гетерогенных систем на основе универсальных и графических процессоров;

- 2) программно-инструментальные средства решения задач мониторинга и имитационного моделирования на высокопроизводительных гетерогенных системах;

Работа выполнена в рамках Программы фундаментальных исследований Президиума РАН №18 «Алгоритмы и математическое обеспечение для вычислительных систем сверхвысокой производительности» (проект «Исследование и разработка инструментальных программных средств мониторинга и обработки потоков данных подсистем космических аппаратов с применением суперкомпьютерных систем, оснащенных многоядерными и графическими процессорами»); Программы ОНИТ РАН «Интеллектуальные информационные технологии, системный анализ и автоматизация» (проект №2.10) и при поддержке РФФИ (проект №13-07-00025 А).

## Литература

- [1] А. А. Кондратьев, И. П. Тищенко: Использование графических вычислителей в процессах обработки и распознавания изображений. – V Всероссийская научно-техническая конференция *«Актуальные проблемы ракетно-космического приборостроения и информационных технологий»* (5-7 июня 2012 г., г. Москва). Тезисы докладов. – М.: Радиотехника, 2012, с.92, ISBN 978-5-88070-025-7.
- [2] А. А. Талалаев: Организация конвейерно-параллельных вычислений для обработки потоков данных. – *Информационные технологии и вычислительные системы*, № 1, 2011, с.8-13.