

УДК 681.3

## ИССЛЕДОВАТЕЛЬСКИЙ ПРОГРАММНО-АППАРАТНЫЙ КОМПЛЕКС ВЫСОКОПРОИЗВОДИТЕЛЬНОЙ ОБРАБОТКИ И ВИЗУАЛИЗАЦИИ МЕДИЦИНСКИХ ДАННЫХ

<sup>1</sup>Талалаев А.А., <sup>1</sup>Фраленко В.П., <sup>2</sup>Хачумов М.В.

<sup>1</sup>ИПС им. А.К. Айламазяна РАН, Вельково, e-mail: arts@arts.botik.ru;

<sup>2</sup>Институт системного анализа Федерального исследовательского центра  
«Информатика и управление» Российской академии наук, Москва, e-mail: khmike@mail.ru

Предметом исследования является построение программно-инструментального комплекса высокопроизводительной обработки данных МРТ и изображений микроскопической съемки головного мозга животных для решения сложных научных задач биомедицинского назначения. Комплекс коллективного пользования строится по клиент-серверной технологии гибридных вычислительных систем, имеющих высокую производительность, относительно невысокую цену и гибкость, снабжен интеллектуальным графическим интерфейсом для быстрого формирования схемы исследования и когнитивной визуализации результатов. Программное обеспечение строится по методологии, реализующей модульный принцип, за счет чего обеспечивается гибкость и наращиваемость функциональных возможностей комплекса. В результате исследования разработана общая архитектура интегрированной высокопроизводительной системы медицинского назначения, выработаны требования и определен состав прикладного программного обеспечения. Областью применения полученных результатов являются информационные территориально распределенные системы коллективного пользования, предназначенные для выполнения совместных исследований в рамках крупных научных проектов.

**Ключевые слова:** программно-инструментальный комплекс, высокопроизводительная обработка, визуализация данных, МРТ, биомедицинские исследования, графический интерфейс

## RESEARCH SOFTWARE AND HARDWARE COMPLEX OF HIGH-PERFORMANCE PROCESSING AND VISUALIZATION OF MEDICAL DATA

<sup>1</sup>Talalaev A.A., <sup>1</sup>Fralenko V.P., <sup>2</sup>Khachumov M.V.

<sup>1</sup>Ailamazyan Program Systems Institute of the Russian Academy of Sciences, Veskovo, e-mail: arts@arts.botik.ru;

<sup>2</sup>Institute for Systems Analysis, Federal Research Center «Computer Science and Control» of Russian Academy of Sciences, Moscow, e-mail: khmike@mail.ru

The subject of the research is the construction of a program-instrumental complex of high-performance data processing of MRI images and microscopic photography of an animal brain to solve complex scientific problems for biomedical applications. The complex of collective use is based on client-server technology of hybrid computing systems with high performance, relatively low price and flexibility, equipped with an intelligent graphical user interface for quick generating of the research diagram and cognitive visualization of results. The software is based on the methodology implementing the modular principle, providing flexibility and scalability of the complex functionality. The general architecture of the integrated high-performance system for medical use, the structure of the application software and requirements it should meet are results of the research. The results obtained during the research are applicable in geographically distributed information systems of collective use, designed to perform joint research in large scientific projects.

**Keywords:** program-instrumental complex, high-performance processing, data visualization, MRI, biomedical research, graphical interface

В России и за рубежом ведутся интенсивные разработки программных и аппаратных комплексов для решения задач в области обработки снимков медицинского и промышленного назначения. Созданы как универсальные программные и программно-аппаратные комплексы, на основе которых могут быть построены прикладные системы, так и специализированные комплексы, предназначенные для анализа медицинских данных. Рассмотрим особенности построения подобных систем и приведем их краткое описание.

В работе [4] изложено видение проблемы внедрения актуальных технологических решений как для лечебных и профилак-

ческих учреждений, так и для профильных ведомств и министерств. Показаны пути внедрения современных информационных технологий и систем в медицину для существенного повышения качества диагностики, лечения и реабилитации пациентов, расширения спектра решаемых медициной задач. В частности, говорится о необходимости объединения действующих медицинских систем, их модификации, интеграции и адаптации к новейшим задачам и технологиям с целью удовлетворения потребностей врачей-исследователей в медицинских сведениях.

Архитектура системы обработки больших объемов изображений на осно-

ве модели распределенных вычислений «MapReduce» [14], используемая для параллельных вычислений над большими наборами данных в компьютерных кластерах с автоматическим распараллеливанием, представлена в работе [7]. Особенностью этой разработки в области анализа изображений на кластерных установках является освобождение прикладного программиста, занимающегося обработкой изображений, от знания деталей внутреннего устройства, связанных с организацией параллельной работы и загрузкой изображений. Программист разрабатывает функцию, обрабатывающую только одно изображение, запуск которой для большого объема изображений в параллельно-конвейерном режиме выполняется системой автоматически.

Компанией «Открытые Технологии» [13] разработаны конфигурации компактных гибридных вычислительных систем на графических процессорах (GPU, Graphics Processing Unit). Основные особенности этих решений:

- наличие стандартных компонентов, достаточных для большинства задач, заказчику необходимо указать только номер конфигурации (или пиковую производительность), используемый тип GPU-вычислителей, необходимость высокоскоростной сети Infiniband, тип операционной системы (Linux либо Windows);

- система поставляется в виде готовой к работе платформы, решение вопросов интеграции системы в инфраструктуру сводится к минимуму;

- возможность использования серверного варианта GPU-вычислителя Nvidia Tesla для обеспечения повышенной надежности, удобства обслуживания и гибкости в подключении к серверам кластера;

- сопоставимости со стоимостью нескольких профессиональных рабочих станций при несоизмеримо более высокой совокупной производительности и возможностях.

Оценка современного состояния предметной области указывает на наличие следующих тенденций: повышенный интерес к проблеме построения универсальных инструментальных программных средств обработки медицинской информации при одновременном отсутствии, несмотря на имеющийся прогресс, эффективных методов построения сложных прикладных информационно-аналитических систем. Наблюдается активное патентование в США и других странах методов и систем обработки и визуализации, в которых все большую роль играют интеллектуальные методы анализа многомерных данных, включая методы

классификации и когнитивного отображения информации.

Целью настоящего исследования является разработка архитектуры нового высокопроизводительного программно-аппаратного комплекса коллективного пользования (ПАК) для поддержки проведения научных исследований, связанных с изучением миграции клеток, трансплантированных в головной мозг в норме и при экспериментальном ишемическом повреждении. Настоящее исследование предполагает построение ПАК для коллективного использования результатов исследований, проводимых рядом институтов на уникальном оборудовании. Часть данных получается на томографе для малых лабораторных животных «ClinScan» компании «Bruker BioSpin» и с помощью ряда микроскопов. Постобработка данных томографии ведется на рабочей станции «HP Z400», которая характеризуется высокой производительностью, функциональностью, удобством обслуживания. Источником изображений микроскопической съемки выступает следующее оборудование: конфокальный микроскоп «Nikon A1R MP» (проприетарный формат получаемых данных), конфокальный микроскоп «Zeiss AxioPlan 2» (формат Tiff) и флуоресцентный микроскоп «Keyence BZ-9000E» (формат Tiff).

Создание подобного программно-инструментального комплекса призвано существенно усилить отдачу дорогостоящего оборудования, обеспечить накопление и доступ к базам данных и знаний, дать возможность большему числу исследователей глубже анализировать информацию и принимать обоснованные решения. В настоящей работе дается в виде научной концепции обоснование выбора архитектуры и приводится описание функциональных модулей программно-аппаратной системы высокопроизводительной обработки и визуализации данных МРТ и изображений микроскопической съемки для выполнения сопутствующих расчетов и интеллектуального анализа ситуации в реальном времени.

#### **Требования к основным характеристикам исследовательского комплекса**

Проект выполняется на базе имеющегося аппаратного обеспечения. Предполагается, что оно будет дополнено новым программным обеспечением и, по необходимости, улучшено путем расширения функциональных возможностей.

Томограф «ClinScan» работает с программным обеспечением «syngo® MR B15». К основным положительным характеристикам системы можно отнести:

– интегрированную технологию параллельного сбора данных «iPAT» (параллельная визуализация во всех направлениях), предназначенную для увеличения пропускной способности томографа;

– опцию «Multi-core», которая эффективна при использовании программ, предназначенных для использования многоядерных процессоров;

– высокую чувствительность за счет применения системы «CryoProbe TM Autorac», позволяющей повысить пространственное или временное разрешение и сократить время сканирования изображений микроскопических структур в живом организме мелких животных;

– возможность одновременного получения нескольких серий снимков;

– широкие возможности графического интерфейса, который, в частности, позволяет запускать несколько независимых друг от друга операций, например, запустить подгрузку новых изображений с томографа и в то же время делать некоторую обработку уже ранее загруженных изображений.

Имеется ряд обстоятельств, требующих совершенствования лабораторной базы как в части математического, так и аппаратного обеспечения:

– необходимость расширения функциональных возможностей и существенного усложнения исследовательских задач, требующих обучения, создания классов, тематической сегментации и выполнения измерений; задача осложняется такими факторами, как отсутствие априорной информации и универсального математического аппарата, решающего эту задачу с требуемой точностью;

– необходимость обработки и научной визуализации в интерактивном режиме больших объемов биомедицинских данных, поступающих от томографа и микроскопов, для чего целесообразно привлечение современных высокопроизводительных вычислительных средств и создание библиотек алгоритмов и программ, реализуемых в том числе на графических процессорах; уровень современных аппаратных платформ позволяет достаточно эффективно использовать ресурсы графических ускорителей и других имеющихся ускорителей для автоматизации процессов обработки биомедицинских данных и облегчить их анализ;

– необходимость в удобном графическом интерфейсе, позволяющем строить схемы решения задач на основе модульного принципа, а также реализующего интерактивную работу исследователя с данными с помощью встроенных инструментальных средств; задача формируется пользователем

в виде иерархического орграфа, вершинами которого служат унифицированные программные модули, связываемые дугами передачи информации;

– возможность параллельной обработки информации одновременно несколькими исследователями, в том числе удаленными, на разнотипных вычислительных платформах, что требует реализации кроссплатформенности.

Построение ПАК требует выработки научной концепции на основе проработки особенностей существующих медицинских информационно-аналитических систем, включая используемые форматы и стандарты, методы обработки и визуализации данных, задачи выбора алгоритмов и построения библиотек, а также анализа современных программно-аппаратных решений. Наличие серьезных ограничений приводит к целесообразности построения нового исследовательского комплекса коллективного пользования на базе имеющегося лабораторного оборудования за счет совершенствования программно-аппаратной базы.

Требования, отвечающие научной концепции построения комплекса, можно сформулировать следующим образом:

– интерактивность – возможность врача проводить исследования в интерактивном режиме, т.е. режиме диалога с медицинским оборудованием и программно реализованными методами исследования;

– функциональная полнота – наличие методов, алгоритмов и методик, обеспечивающих достаточность функционального обеспечения комплекса при решении задач обработки и когнитивной визуализации многомерных биомедицинских данных;

– возможность параллельной обработки за счет распараллеливания программ применительно к многоядерным/многопроцессорным системам и графическим ускорителям;

– кроссплатформенность;

– интеграция территориально распределенного уникального оборудования, программного обеспечения и данных исследований в единую информационно-аналитическую систему на базе компьютерной вычислительной сети, томографа и микроскопов.

#### **Требования по составу алгоритмов**

На основе анализа научных источников выявлены следующие необходимые программные компоненты, опирающиеся как на продукты с открытыми кодами, так и на оригинальные разработки:

– алгоритмы конвертации данных из DICOM;

- стандартные и оригинальные алгоритмы предварительной обработки данных [2];
- методы выделения и тематической сегментации области мозга: подпрограмма «BET» (Brain Extraction Tool) из пакета программ «FSL», подпрограмма «seg\_EM» программы «NiftySeg», а также специальные разработки [1];

- методы глубокой обработки данных с распознаванием и измерением областей поражения [5];

- программные средства визуализации данных («ImageJ», «MRICron», «RadiAnt DICOM Viewer», «xjView», «Novo Spark», «3D Slicer» и др.).

### Требования к графическому интерфейсу

Для повышения эффективности моделирования и оперативного отображения информации в состав ПАК должны входить соответствующие программные средства, позволяющие повысить оперативность построения схемы исследований, анализа и принятия решений. Программные средства отображения результатов должны обеспечивать оператору возможность выбора на схеме задачи точек вывода информации для отображения в удобном когнитивно-графическом виде. Должен быть определен формат передаваемых данных от модулей обработки данных, используемых в ходе решения прикладной задачи, в модули отображения интерфейса пользователя. Должна быть обеспечена поддержка следующих возможностей:

- построение визуальных образов для обрабатываемых данных и результатов обработки;

- передача информации о зонах или областях, представляющих особый интерес для пользователя (например, таких как выделенные аномалии);

- отображение нескольких потоков обрабатываемых данных в одном модуле визуализации;

- поддержка бесконфликтной работы нескольких удаленных пользователей;

- возможность настройки стилиевой и визуальной составляющей отображаемых сущностей (в частности – модулей обработки, каналов связи);

- сохранение описаний задач (схем) в виде текста в формате XML;

- окно визуализации схем задач должно обеспечивать стандартные интерфейсы управления, такие как перемещение по рабочему пространству с помощью манипулятора типа «мышь»; кнопки уменьшения/увеличения масштаба; горизонтальные и вертикальные полосы прокрутки;

- возможность подключения модулей для отображения специальных когнитив-

ных образов, в том числе с помощью современных средств визуализации, например, платформонезависимых программных интерфейсов «OpenGL» и/или «Vulkan»;

- кроссплатформенность и единство внешнего вида на разных операционных системах (Windows, Linux и пр.).

### Концепция построения высокопроизводительного исследовательского комплекса коллективного пользования

Предлагаемая концепция программно-инструментального комплекса учитывает современные направления развития медицинских информационных технологий в области обработки и визуализации изображений [3], в том числе принципы:

- унификации и интеллектуализации медицинских и промышленных интерфейсов;

- стандартизации форматов представления данных и создания библиотек методов их обработки;

- стандартизации способов передачи данных в синхронном и асинхронном режимах;

- интеллектуализации автоматизированной и автоматической обработки слабо-структурированных данных;

- визуализации многомерной и разнородной информации (в том числе когнитивной);

- поддержки принятия решений пользователя;

- модульности и наращиваемости программного обеспечения;

- параллельной обработки на много-процессорных вычислительных системах и графических процессорах.

Схема интегрированного программно-аппаратного комплекса представлена на рис. 1.

Клиентская часть включает в себя графический интерфейс пользователя, обеспечивающий формирование схем решения прикладных задач в парадигме визуально-блочного программирования и отображение результатов обработки биомедицинских данных, полученных с использованием исследовательского оборудования.

Серверная часть программно-аппаратного комплекса обеспечивает хранение и многопользовательский доступ к данным проводимых исследований (посредством БД) и возможность их высокопроизводительной конвейерно-параллельной обработки в соответствии со сформированными пользователями задачами с использованием ядра программной системы, обеспечивающего вычислительную среду, и модулей обработки, реализующих требуемые алгоритмы.

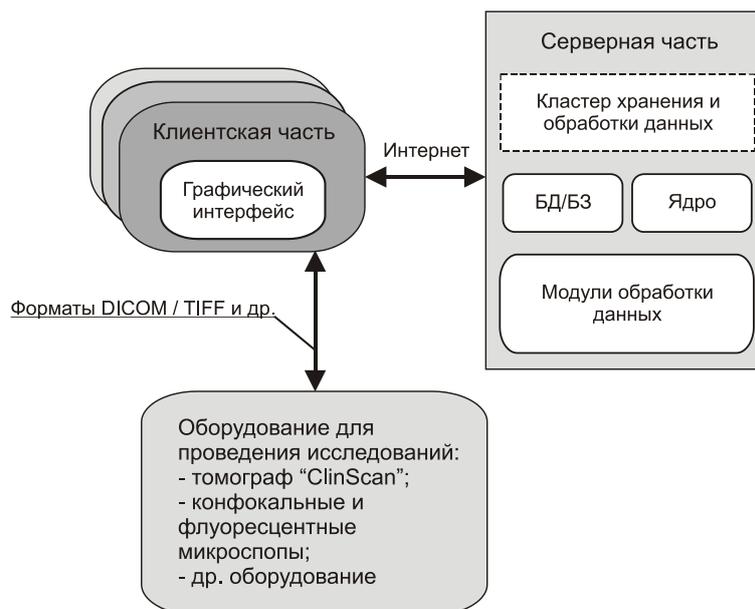


Рис. 1. Схема интегрированного комплекса

Концепция построения высокопроизводительной системы обработки и научной визуализации данных МРТ и изображений микроскопической съемки опирается на имеющиеся в ИПС им. А.К. Айламазяна РАН разработки в области конвейерно-параллельных вычислений. Как правило, вычислительные ресурсы существующих медицинских комплексов не обладают высокой производительностью, что препятствует обработке больших объемов анализируемых данных и реализации многопользовательского режима. Применение графических процессоров для решения задач обработки медицинских снимков является перспективным направлением, способным значительно сократить временные затраты на обработку данных. Процессоры общего назначения (CPU) и графические процессоры дополняют друг друга. CPU показывает лучшую производительность при работе с последовательными задачами, но при большом объеме обрабатываемой информации, с которой можно работать параллельно, очевидное преимущество имеет GPU.

Предлагаемые авторами программно-аппаратные решения опираются на имеющийся научный задел в области высокопроизводительной конвейерно-параллельной обработки данных [10]. Наиболее перспективным подходом видится использование концепции GPGPU, которая позволит достичь многократного повышения производительности за счет архитектурных особенностей графических ускорителей [9].

Ускорение автоматического анализа потоков медицинских данных выполняется за счет учета конвейерности процессов обработки как особого вида параллелизма. Программно-инструментальный комплекс для организации конвейерно-параллельных вычислений имеет в качестве прототипа архитектуру системы, описанную в работах [3, 8, 12]. Архитектура ПАК предоставляет пользователю возможности формирования различных стратегий параллельной обработки данных на мультипроцессорных системах. Основу составляют ядро, графический интерфейс, модули параллельных реализаций искусственных нейронных сетей, фильтров и других алгоритмов. Универсальное программное ядро обеспечивает связь модулей в процессе выполнения задачи. С помощью графического интерфейса формируется образ задачи, который записывается в виде XML-файла и может быть сохранен и/или выполнен на вычислителе.

Прототипом аппаратной части системы может служить высокопроизводительный вычислительный комплекс «MIX-1» [6]. Общая архитектура высокопроизводительного гетерогенного вычислительного комплекса для параллельной обработки представлена на рис. 2. Архитектура комплекса удовлетворяет современным требованиям к структуре, способам обработки и передачи информации и соответствует мировым тенденциям в данной области. Комплекс представляет собой гетерогенную вычислительную си-

стему с распределенной памятью. Вычислительные узлы имеют одинаковые аппаратные характеристики, что предполагает легкое горизонтальное масштабирование системы с целью увеличения общей производительности комплекса при решении требуемых задач.

В таблице далее представлены основные аппаратные характеристики вычислительных узлов системы.

На рис. 3 представлена архитектура программного обеспечения высокопроизводительного гетерогенного вычислительного комплекса для обработки данных МРТ и изображений микроскопической съемки.

Для выполнения исследований комплекс оснащен прикладным программным обеспечением, которое позволяет эффективно решать как вычислительные задачи, допускающие внутренний (поточный) параллелизм в рамках одного

процесса, так и задачи, требующие распределения по данным между несколькими узлами кластерного вычислительного устройства.

В настоящее время ПАК может комплектоваться библиотекой «XRay» алгоритмов обработки изображений медицинского и промышленного назначения с использованием GPU [11]. Библиотека содержит модули программ-обработчиков, ориентированные на гетерогенную вычислительную среду, и позволяет решать спектр задач обработки как данных МРТ, так и снимков медицинского назначения (в том числе изображений микроскопической съемки). В состав библиотеки входят модули для интегральных преобразований, обработки данных в частотной области; обработки с помощью фильтров (масок); сегментации и маркировки; работы с гистограммами; поэлементной обработки изображений.

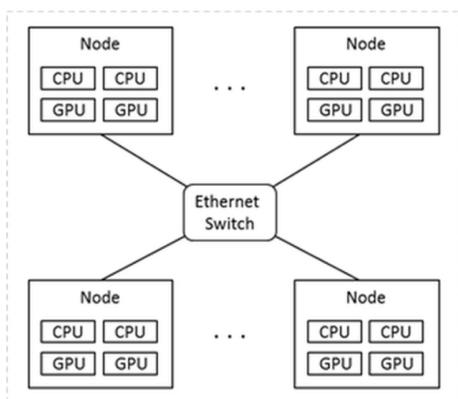


Рис. 2. Архитектура и общий вид вычислительной системы «MIX-1»



Рис. 3. Архитектура программной части комплекса

Аппаратные характеристики  
вычислительного узла системы «MIX-1»

Наименование	Значение
Конструктив узла	1U RackMount
Тип процессора	Intel®Xeon® X5650 2,66 ГГц
Тактовая частота внешней шины	1333 МГц
Количество процессоров	2
Тип коммуникационной сети	Gigabit Ethernet
Оперативная память	32 Гб ECC
Блок питания	2 x 1400 Вт
Дисковая память	2 x 300 Гб
Система охлаждения	Воздушная
Графические ускорители	2 x Nvidia Tesla M2050

**Заключение**

Разработана концепция универсального программно-аппаратного комплекса коллективного пользования для конвейерно-параллельной обработки данных МРТ и изображений микроскопической съемки в гетерогенной вычислительной среде. Можно выделить следующие основные свойства и особенности комплекса:

- наличие средств высокопроизводительной обработки потоков данных на многоядерных и графических процессорах, в том числе методов и технологий фильтрации, выделения и текстурного анализа;

- наличие унифицированных прикладных библиотек алгоритмов базовой обработки данных для обеспечения работы врачей-исследователей;

- возможность удаленной коллективной работы над накопленными данными.

В дальнейшем планируется уточнить состав программных средств для решения различных прикладных задач.

Авторы выражают благодарность Губскому Илье Леонидовичу (ГБОУ ВПО РНИМУ им. Н.И. Пирогова Минздрава России) и Вахрушеву Игорю Викторовичу (НИИ биомедицинской химии имени В.Н. Ореховича) за ценные консультации и экспертную оценку выдвинутых предложений.

*Работа выполнена при частичной финансовой поддержке проекта РФФИ № 16-29-07116-офи м «Информационно-аналитические модели и методы когнитивной визуализации процессов миграции и хоуминга трансплантированных мезенхимальных стволовых клеток для поддержки*

*проведения научных исследований и разработки подходов к лечению заболеваний методами клеточной терапии».*

**Список литературы**

1. Абдулракеб А.Р.А., Сушкова Л.Т., Лозовская Н.А. Обзор методов сегментации опухолей на МРТ-изображениях головного мозга // Прикаспийский журнал: управление и высокие технологии. – 2015. – № 1. – С. 122–138.

2. Гуляев М.В. Управление контрастом в магнитно-резонансной томографии в полях 0,5 и 7 тесла: дисс.... канд. физико-математических наук. – М., 2013. – 114 с.

3. Заднепровский В.Ф. и др. Программно-инструментальный комплекс высокопроизводительной обработки изображений медицинского и промышленного назначения // Информационные технологии и вычислительные системы. – 2014. – № 1. – С. 61–72.

4. Лищук В.А. и др. Стратегия информатизации медицины – основные положения, принципы и предложения. – Ейск: ЮгПолиграф, 2011. – 237 с.

5. Магонов Е.П. Комплексная магнитно-резонансная морфометрия очаговых и атрофических изменений головного мозга (на примере рассеянного склероза и ранних стадий ВИЧ-инфекции): дисс.... канд. мед. наук. – СПб., 2015. – 167 с.

6. Попков А.Ю., Соченков И.В., Хачумов В.М. Гетерогенный вычислительный комплекс для решения инженерных и научных задач // III Национальный Суперкомпьютерный Форум: тезисы докл. (с. Вельсково, 24–27 ноября 2014 г.). – с. Вельсково, 2014. URL: [http://2014.nscf.ru/TesisAll/8\\_Stendovaya/02\\_074\\_KhachumovVM.pdf](http://2014.nscf.ru/TesisAll/8_Stendovaya/02_074_KhachumovVM.pdf) (дата обращения: 06.06.2016).

7. Созыкин А.В., Гольдштейн М.Л. Архитектура системы обработки больших объемов изображений с автоматическим распараллеливанием // Научный сервис в сети Интернет: поиск новых решений: Труды Международной суперкомпьютерной конференции (Новороссийск, 17–22 сентября 2012 года). – Новороссийск, 2012. – С. 58–62.

8. Талалаев А.А. Организация конвейерно-параллельных вычислений для обработки потоков данных // Информационные технологии и вычислительные системы. – 2011. – № 1. – С. 8–13.

9. Талалаев А.А. и др. Концепция высокопроизводительного комплекса визуализации и обработки медицинских и промышленных снимков // Научно-технический вестник Поволжья. – 2013. – № 2. – С. 233–238.

10. Талалаев А.А., Фраленко В.П. Архитектура комплекса конвейерно-параллельной обработки данных в гетерогенной вычислительной среде // Вестник Российского университета дружбы народов. Серия «Математика. Информатика. Физика». – 2013. – № 3. – С. 113–117.

11. Хачумов В.М. и др. Библиотека алгоритмов обработки изображений медицинского и промышленного назначения с использованием графических вычислительных устройств (Библиотека «XRay») // Свидетельство о государственной регистрации программы для ЭВМ № 2013618678, дата приоритета: 18.07.2013, дата регистрации: 13.09.2013.

12. Хачумов В.М., Талалаев А.А. Технические характеристики кластерных вычислителей и анализ эффективности параллельных программных средств обработки потоков данных // Авиакосмическое приборостроение. – 2011. – № 12. – С. 3–17.

13. Open Technologies [Электронный ресурс]. URL: <http://www.ot.ru/en/> (дата обращения: 06.06.2016).

14. Stonebraker M. and etc. MapReduce and Parallel DBMSs: Friends or Foes? // Communications of the ACM, 2010. – Vol. 53, No. 1. – P. 64–71.