



# КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ COMPUTER SIMULATION HISTORY

УДК 004.275.56

## МЕТОДЫ ОЦЕНКИ ПРОИЗВОДИТЕЛЬНОСТИ ВЫЧИСЛИТЕЛЬНЫХ КОМПЛЕКСОВ

## METHODS FOR EVALUATING THE PERFORMANCE OF COMPUTER SYSTEMS

С.А. Сорокин<sup>1</sup>, С.М. Чудинов<sup>1</sup>, А.П. Сорокин<sup>1</sup>, Е.В. Болгова<sup>2</sup>  
S.A. Sorokin<sup>1</sup>, S.M. Chudinov<sup>1</sup>, A.P. Sorokin<sup>1</sup>, E.V. Bolgova<sup>2</sup>

<sup>1</sup>) АО «НИИВК им. М.А. Карцева», 117437, Москва, Профсоюзная ул., 108

<sup>1</sup>) «M.A. Kartsev Scientific and Research Institute of Computing Systems», Russia, 117437

<sup>2</sup>) Белгородский государственный национальный исследовательский университет,  
Россия, 308015, Белгород, ул. Победы, 85

<sup>2</sup>) Belgorod State National Research University, 85 Pobeda St, Belgorod, 308015, Russia

e-mail: Sergey@prosoft.ru, Chud35@yandex.ru, Bolgova\_e@bsu.edu.ru

### Аннотация

Представлены методы оценки производительности вычислительных комплексов (платформ) и возможности их использования в автоматизированных системах управления для обработки информации.

### Abstract

Methods for evaluating the performance of computer systems (platforms) and the possibility of their use in automated control systems for information processing are presented.

**Ключевые слова:** вычислительная платформа, оценка производительности вычислительных комплексов, управление параллелизмом операций, характеристики вычислительных комплексов.

**Keywords:** group computing platform, estimation of computer systems performance, control of operation parallelism, characteristics of computer systems.

### Введение

В последние годы широко и все более быстрыми темпами осуществляется разработка, выпуск и внедрение в автоматизированные системы обработки информации и управления, вычислительные центры и вычислительные сети разнообразных многопроцессорных и многомашинных вычислительных систем. Такие вычислительные системы отличаются повышенной гибкостью и характеризуются высокими, а в ряде случаев – рекордными показателями производительности, объемов внутренней памяти, надежности и живучести и называются вычислительными платформами.

Программирование для многопроцессорных и многомашинных систем связано с распараллеливанием и синхронизацией вычислений и с организацией выполнения параллельных вычислительных процессов. Это выдвигает целый ряд сложных задач, среди которых весьма важными являются расчет характеристик времени и количества операций, требующихся для выполнения параллельных программ, и построение расписаний (планов)



выполнения параллельных программ на многопроцессорных и многомашинных вычислительных системах.

В настоящее время в отечественных вычислительных системах типа «Эльбрус» (разработчик и производитель АО «МЦСТ») и вычислительных платформ типа «ГРИФОН» (разработчик и производитель ЗАО «Доломант») широко используются параллельные методы обработки, в связи с этим представляет интерес оценка производительности вычислительных комплексов [Бычков И.Н. и др., 2016] и их возможность использования в автоматизированных системах управления.

### Управление параллелизмом операций

Важнейшей особенностью архитектуры «Эльбрус» и «ГРИФОН» является явное указание процессорному ядру параллельно исполняемых операций, при этом анализ независимости и их планирование выполняет компилятор. Это позволяет отказаться от сложной и энергоемкой аппаратуры обеспечения внеочередного исполнения команд, используемой во всех современных универсальных процессорах, и делает осмысленным расширение парка исполнительных устройств и повышение предельной производительности на такт до уровней, превосходящих возможности конкурирующих решений. В первых четырех поколениях процессоров с архитектурой «Эльбрус», включая процессоры Эльбрус-2С+ и Эльбрус-4С, с помощью широкой команды на одном процессорном ядре в каждом такте можно запускать на исполнение: на ациклических участках кода до 16 операций, при выполнении циклов – до 23 операций, 8 64-разрядных вещественных операций. В процессоре Эльбрус-8С уже можно запускать на исполнение до 25 операций и 12 64-разрядных вещественных операций. Для снижения потерь от доступа за данными в память реализован аппаратно-программный механизм асинхронной предварительной подкачки данных в специальные буферные памяти. Специальные операции подготовки переходов обеспечивают возможность эффективно осуществлять переходы без использования предсказателя переходов. Аппаратура поддерживает средства (исполнение под управлением предикатов) программно-управляемого переупорядочивания операций и одновременного исполнения нескольких условных ветвей программы. Наряду с параллелизмом на уровне операций в архитектуре «Эльбрус» и «ГРИФОН» реализован векторный параллелизм за счет операций над упакованными данными. Благодаря этому возможности выполнения 32-разрядных вещественных операций на одном ядре возрастают в 2 раза и позволяют исполнять: на «Эльбрус-2С+» и «Эльбрус-4С» – до 16, а на процессоре «Эльбрус-8С» – до 24 упакованных операций (Флопс). Упакованные операции реализованы также для целочисленных операций. Они находят применение в таких областях, как обработка сигналов, обработка изображений, графика и др. Во вновь разрабатываемых процессорах с архитектурой «Эльбрус» возможности выполнения упакованных операций удваиваются. Процессоры с архитектурой «Эльбрус» поддерживают и расширяют при переходе к новым поколениям параллелизм потоков управления на основе многоядерности. Кроме этого поддерживается параллелизм многопроцессорности на общей памяти с когерентными многоуровневыми кэш-памятями на базе неодинакового времени обращения к общей памяти (NUMA). Все описанные механизмы параллельного исполнения эффективно поддерживаются оптимизирующим компилятором, что позволяет автоматически получать коды, использующие мощное процессорное ядро и другие параллельные возможности архитектуры [Бычков И.Н. и др., 2016; Волконский В.Ю. и др., 2012]. Технологии параллельного исполнения операций с использованием для этого оптимизирующего компилятора находят применение в развитии известных архитектур. Хотя в большинстве универсальных процессоров используется внеочередное аппаратное планирование потока операций, некоторые элементы планирования вычислений с помощью компилятора заложены в архитектуру Sparc64, начиная со Sparc64 V8fx – специального расширения архитектуры Sparc v9 для суперкомпьютеров. Как и в архитектуре «Эльбрус», в этой архитектуре используются большие регистровые файлы, а также возможность закодировать одновременную выборку



сразу двух операций, с помощью которых оптимизирующему компилятору удастся гораздо лучше распараллеливать вычисления, особенно в циклах. Для использования вычислительных платформ в автоматизированных системах необходима разработка оценки их производительности.

### Оценка производительности вычислительных комплексов на микропроцессорах с архитектурой «Эльбрус»

Для определения показателей производительности вычислительных модулей и комплексов существуют несколько различных подходов. Оптимальный вариант – применение эталонных тестов для сравнения различных вычислительных модулей и комплексов, входящих в вычислительную платформу.

Примером эталонных тестов являются тестовые пакеты организации *SPEC* (Консорциум по оценке производительности систем) [*SPEC CPU<sup>TM</sup> 2006*], которые измеряют производительность вычислительного комплекса, зависящую от таких факторов, как объем оперативной памяти, конфигурация кэша, количество ядер, и в качестве основной нагрузки используют математические операции, нечувствительные к качеству оптимизаций компилятора.

Другой пример – направленные тесты пакета *C++ Performance Benchmarksuite* компании *AdobeSystemsInc.*, предназначенные для тестирования архитектурной связки «компилятор – процессорное ядро» [*C++ Performance Benchmark suite*]. Результаты тестирования позволяют обнаружить недостатки реализации оптимизаций в компиляторе, а также процессорного ядра для архитектур с оптимизацией в режиме исполнения кода, например *x86*. Специфика тестовых программ данного пакета заключается в том, что на них не влияют объемы оперативной памяти, кэша и количество ядер.

Поскольку оригинальность вычислительных модулей на основе микропроцессоров с архитектурой «Эльбрус» заключается именно в тесной связи процессорного ядра и компилятора, то сравнение производительности таких комплексов с зарубежными аналогами проводилось посредством тестирования с помощью пакета *C++ PerformanceBenchmarksuite*.

Для сравнения были выбраны вычислительные комплексы на основе микропроцессоров *Intel* архитектурами *x86* и *IA-64 (Itanium)*. Архитектура процессорных ядер *IA-64* относится к типу *EPIC (explicitly parallel instruction computing)* и наряду с архитектурой ядра процессора «Эльбрус» входит в класс архитектур *VLIW (very large instruction word)*. Таким образом, стала возможной оценка особенностей реализации компилятора и процессора.

Для исключения преимущества в производительности из-за более высокой тактовой частоты процессоров *Intel* были введены поправочные коэффициенты. Также для сопоставимости результатов различных тестов была проведена нормализация по времени.

Характеристики тестируемых вычислительных комплексов на основе микропроцессоров разных архитектур приведены в табл. 1-3.

Таблица 1  
Table 1

#### Характеристики вычислительного комплекса с архитектурой *x86* Characteristics of a computational complex with *x86* architecture

Характеристики	1	2	3
Обозначение			
Процессор	Intel Core 2 Quad Q9400		Intel Xeon 5160
Микроархитектура	Yorkfield		Woodcrest
Количество ядер, шт.	4		2
Тактовая частота, Гц	2,66		3,00
Объем кэша L1, Мб	6		4
Объем ОЗУ, Гб	4	4	12
Операционная система	Linux 2.6.33	Linux 2.6.27	Linux 2.6.27
Компилятор	GGG 5.2.0	GGG 4.1.2	GGG 4.1.2



Проведенное тестирование показало, что вычислительный комплекс на основе микропроцессоров с архитектурой «Эльбрус» продемонстрировал более высокую производительность (об этом свидетельствует большинство тестов – 24 из 39) по сравнению с зарубежными аналогами, причем iuv тесты пакета C++ PerformanceBenchmarksuite он исполнял быстрее, чем вычислительный комплекс на основе платформы Itanium.

Таким образом, результаты тестовых испытаний доказали, что лучшую реализацию системы «компилятор – процессное ядро» в классе архитектуры VLIW имеет вычислительный комплекс на основе платформы «Эльбрус».

Таблица 2  
Table 2

**Характеристики вычислительного комплекса с архитектурой IA-64**  
**Characteristics of a computing complex with IA-64 architecture**

Характеристики	4	5
Обозначение		
Процессор	IntelItanium 2	
Микроархитектура	Madison	
Количество ядер, шт.	2	
Тактовая частота, ГГц	1,60	
Частота FSB, ГГц	533	
Объем кэша L2, Мб	9	
Объем ОЗУ, Гб	8	
Операционная система	HP – UX S688816 B.11.31 U	
Компилятор	GGG 4.2.3	HP Acc (HP C/Ac++ b3910b a.06.28)

Таблица 3  
Table 3

**Характеристики вычислительного комплекса с архитектурой «Эльбрус»**  
**Characteristics of the computer complex with the Elbrus architecture**

Характеристики	6	7
Обозначение		
Процессор	Эльбрус-4С	Эльбрус-8С
Микроархитектура	E2K	
Количество ядер, шт.	4	8
Тактовая частота, ГГц	0,75	1,20
Объем кэша L2, Мб	8	4
Объем кэша L3, Мб	-	16
Объем ОЗУ, Гб	12	16
Операционная система	Linux 2.6.33	Linux 3.14.53 (314.2.0)
Компилятор	LGG 1.20.09, режим LP64	LGG 1.20.10, совместимый с GGG 4.4.0, режим LP64

Приведенные данные в сравнении позволяют пользователям автоматизированных систем определить необходимый тип вычислительного комплекса для обработки информации. Другим важным направлением использования отечественных платформ для обработки информации является конкретный пример по обработке изображений.

**Высокопроизводительная гетерогенная вычислительная платформа**

В статье представлены материалы по разработанному отечественному многоцелевому комплексу аппаратно-программных средств (платформа), предназначенном для построения встраиваемых систем и организации высокопроизводительных вычислений, в том числе обработки изображений.

Специально для успешного решения указанных выше ответственных задач в ЗАО «НПФ «ДОЛОМАНТ» совместно с АО «НИИВК им. М.А. Карцева» была разработана новая отечественная высокопроизводительная гетерогенная вычислительная платформа (ВГВП), рис. 1 [Галаган П.В., Тумакин Д.А., 2016].



Рис. 1. Внешний вид базового вычислительного блока  
 Fig. 1. Appearance of the base computing unit

ВГВП представляет собой универсальную суперкомпьютерную вычислительную платформу для решения задач высокопроизводительных вычислений в жёстких условиях эксплуатации – при критических массогабаритных ограничениях аппаратной части – для решения задач встраиваемого класса.

Большинство задач встраиваемого класса, для которых необходимо применение платформы на мировом рынке отнесены к отдельному сегменту – НРЕС (HighPerformanceEmbeddedComputing), т. е. встроенные высокопроизводительные вычисления (в русской аббревиатуре ВВВ). Сегодня это особый рынок, объединяющий сразу несколько сегментов в один уникальный сегмент со своими потребителями, игроками, прикладными задачами и, конечно, со своими подходами к созданию таких вычислительных систем.

Укрупнённо можно выделить следующие основные принципы создания ВГВП.

1. Открытая архитектура.
2. Массогабаритные ограничения и энергоэффективность.
3. Возможность применения в жёстких условиях эксплуатации.
4. Возможности по созданию многомашинных конфигураций.
5. Низкие задержки и высокоинтенсивный обмен данными между вычислителями.
6. Высокая вычислительная плотность (производительность на кубический сантиметр).

7. Богатые ресурсы ввода-вывода данных, как в цифровом, так и в аналоговом виде.

8. Неоднородность (гетерогенность) вычислительной среды.

ВГВП – изделие отечественного производства, построенное с учётом всех перечисленных подходов и реализующее целостную концепцию многоцелевой высокопроизводительной вычислительной платформы с гетерогенной вычислительной средой для решения задач высокопроизводительных вычислений в жёстких условиях эксплуатации [Иванов М.И., Сорокин С.А., 2017].

Высокопроизводительная гетерогенная вычислительная платформа (ВГВП) позволяет строить и эффективно применять гетерогенные конфигурации. Выбор конкретной гетерогенной конфигурации обусловлен комплексом исходных технических требований, типом данных и режимов их обработки. На базе ВГВП представляется возможным осуществлять конвейерную обработку данных с применением гетерогенной архитектуры. Идея использования гетерогенных вычислительных конвейеров заключается в том, чтобы на каждом этапе последовательной обработки (участке конвейера) обработчик на базе оптимальной для работы с конкретным типом данных архитектурой, выполнив свою работу, передавал бы результат для дальнейшей обработки на следующий участок конвейера для обработки вычислителем – обработчиком другой архитектуры, одновременно принимая новый объем входных данных для следующей итерации цикла конвейерной обработки. При этом большинство задач машинного зрения хорошо поддаются распараллеливанию при обработке данных. Например, каждая видеочасть передает один видеопоток, если таких камер несколько, то для повышения общего быстродействия весьма эффективно разделить конвейер на участки параллельной обработки, где это возможно, получив прирост



производительности. Механизм параллельно-конвейерной обработки является признанным классическим методом повышения быстродействия систем обработки данных, и если структура данных и алгоритм позволяют распараллеливать задачу, то это почти всегда повышает эффективность такой обработки. Так, гетерогенность, архитектурные решения и программные механизмы взаимодействия модулей различной архитектуры позволяют эффективно применять ВГВП для гетерогенной параллельно-конвейерной обработки данных. Рассмотрим возможности ВГВП для организации параллельно-конвейерной обработки данных на примере системы обработки видео высокого разрешения в режиме реального времени. Постановку задачи можно кратко сформулировать следующим образом: требуется в режиме реального времени принять данные от четырех камер высокого разрешения, провести предварительную обработку, передать данные на отдельный обработчик для отработки прикладных алгоритмов компьютерного зрения с дальнейшей передачей результата для принятия решения центральным процессором. Исходя из постановки данной задачи был сконфигурирован аппаратный состав базового вычислительного блока – гетерогенного вычислителя на базе ВГВП – табл. 4, а дополнительные аппаратные средства представлены в табл. 5.

Таблица 4  
Table 4

**Аппаратный состав гетерогенного вычислителя обработки видео высокого разрешения на базе ВГВП**  
**Hardware composition of heterogeneous high-resolution video processing on the basis of VGVP**





Наименование	Описание	Внешний вид	Количество
<b>CPC512</b>	Модуль центрального процессора (может использоваться в микропроцессорах с архитектурой Эльбрус)		1 шт.
<b>FPU500</b>	Модуль ПЛИС		1 шт.
<b>VIM556</b>	Модуль графического процессора		4 шт.
<b>KIC551</b>	Модуль коммутации PCIe		1 шт.
<b>KIC550</b>	Модуль-носитель HDD-накопителя		1 шт.
<b>MIC2003</b>	Мезонинный модуль ввода		1 шт.

Таблица 5  
Table 5

**Дополнительные аппаратные средства**  
**Additional hardware**

Наименование	Количество
Камеры full-hd	4 шт.
3G-SDI-коннекторы	4 шт.
Мониторы	4 шт.



На рис. 2 представлена схема параллельно-конвейерной обработки данных на базе ВГВП.



Рис. 2. Параллельно-конвейерная обработка данных на примере системы обработки видео высокого разрешения в режиме реального времени, построенной на базе ВГВП

Fig. 2. Parallel-pipeline processing of data on the example of a high-resolution video processing system in real time, built on the basis of VGVP

В статье рассмотрены этапы работы конвейера на конкретном примере. Для ввода данных в вычислительный контур сразу от нескольких камер по стандарту 3G-SDI используется мезонинный submodule МПС, смонтированный на вычислительный модуль FPU500, что позволяет, во-первых осуществить ввод данных через нестандартные интерфейсы, а во-вторых осуществить ввод «напрямую» (без транзита по общей транспортной шине PCIe) на модуль FPU500 для дальнейшей обработки.

Поступающие на модуль FPU500 кадры видеоизображения разрешением 1920x1080 в формате 3G-SDI, далее декодируются и сохраняются в памяти модуля в формате YUV420, организованной в виде кольцевого буфера емкостью 16 кадров для каждой камеры. При очередной записи кадра модуль генерирует прерывание на шине PCIe, по которому управляющая программа на модуле центрального процессора CPC512 выдает команду на копирование кадра из памяти FPU500 в память модуля графического процессора VIM556 по линиям шины PCIe. Один модуль FPU500 может одновременно обслуживать видеопотоки от 4-х видеокамер. На модуле графического процессора VIM556 в режиме реального времени средствами CUDA и компонентами библиотеки OpenCV отрабатываются нужные прикладные алгоритмы: поиск изображения, детектирование движения изображения лавинно-образного типа, дополнительная фильтрация. Далее средствами библиотек OpenGL и XLib прошедший обработку на VIM556 кадр без передачи по PCIe в режиме реального времени отображается на подключенном к модулю VIM556 мониторе. Процесс такой обработки идет по 4 параллельным гетерогенным конвейерам по количеству входных потоков данных – в данном примере задействованы 4 камеры. При этом основная нагрузка делегируется для выполнения средствами модуля на базе ПЛИС FPU500 и модулей графического процессора VIM556. Модуль центрального процессора CPC512 не задействован непосредственно в обработке данных, а выдает только управляющие команды, что существенно снижает его загрузку, высвобождая ресурсы для выполнения другого функционала.

Авторы разработали руководство пользователя модулей FPU500 v2.0. Высокая производительность обработки данных в модуле FPU500 обеспечивается FPGA Virtex-6 фирмы Xilinx и двумя параллельными банками памяти DDR3 SDRAM общим объемом 4 Гигабайт. В качестве загрузчика битстримов для основного FPGA Virtex-6 применяется CPLD CoolRunner II. Действительно, следует особо отметить, что одним из важных преимуществ ВГВП является поддержка режима «каждый с каждым» (peer-to-peer/P2P) при



межмодульном взаимодействии по высокоскоростной шине PCIe. Это позволяет осуществлять пересылку данных от одного вычислительного модуля другому без участия центрального процессора. В данном примере механизмы прямого межмодульного взаимодействия в режиме «каждый с каждым» позволяют высвободить ресурсы центрального процессора и снизить нагрузку на основной транспортный интерконнект по шине PCIe, что на практике позволяет минимизировать время обработки кадра по конвейеру. Важным параметром ВГВП при разработке является производительность. К основным характеристикам производительности конвейера можно отнести следующие параметры:

- конвейерная задержка
- пропускная способность
- уровень загрузки ЦП.

### Заключение

Приведенные в статье научные подходы по оценке производительности вычислительных платформ показывают их возможность использования в автоматизированных системах по обработке информации. Приведенные данные обеспечения вычислительных средств показывают их конкурентоспособность по сравнению с иностранными образцами и отражают возможность работ по импортозамещению.

### Список литературы

#### References

1. Бычков И.Н., Молчанов И.А., Фельдман В.М., Юрлин С.В., 2016. Вычислительные комплексы на микропроцессорах с архитектурами «Эльбрус» и SPARC для построения автоматизированных систем управления. Научно-производственный и культурно-образовательный журнал «Качество и Жизнь». Специальный выпуск: 74-80.

Buchkov I.N., Molchanov I.A., Feldman V.M., Yurlin S.V., 2016. Computing systems on microprocessors with "Elbrus" and SPARC architectures for constructing automated control systems. Scientific-production and cultural-educational magazine "Quality and Life". Special issue: 74-80.

2. Волконский В.Ю., Брегер А.В., Бучнев А.Ю., Грабежной А.В., Ермолицкий А.В., Муханов Л.Е., Нейман-заде М.И., Степанов П.А., Четверина О.А., 2012. Методы распараллеливания программ в оптимизирующем компиляторе. Вопросы радиоэлектроники серия ЭВТ, 3: 63-88.

Volkonsky V.Yu., Breger A.V., Buchnev A.Yu., Grabeznoy A.V., Ermolitsky A.V., Mukhanov L.E., Neiman-zade M.I., Stepanov P.A., Chetverina O.A., 2012. Methods of parallelizing programs in an optimizing compiler. Questions of radio electronics series EVT, 3: 63-88.

3. Галаган П.В., Тумакин Д.А., 2016. Высокопроизводительная гетерогенная вычислительная платформа для построения встраиваемых систем. Вопросы радиоэлектроники, 10: 21-31.

Galagan P.V., Tumakin D.A., 2016. High-performance heterogeneous computing platform for building embedded systems. Voprosy radioelektroniki [Questions of radio electronics] 10: 21-31.

4. SPEC CPU™ 2006 [Электронный ресурс]: Standard Performance Evaluation Corporation. URL: <http://www.spes.org/benchmarks.html#cpu> (дата обращения: 29.06.2016).

SPEC CPU 2006 [Electronic resource]: Standard Performance Evaluation Corporation. URL: <http://www.spes.org/benchmarks.html#cpu> (date of circulation: June 29, 2016).

5. C++ Performance Benchmark suite [Электронный ресурс]: Adobe Systems. URL: <https://stlab.adobe.com/performance> [дата обращения: 01.04.2016].

C ++ Performance Benchmark suite [Electronic resource]: Adobe Systems. URL: <https://stlab.adobe.com/performance> [date of circulation: 01/04/2016].

6. Иванов М.И., Сорокин С.А., 2017. Обработка изображений в системе технического зрения с использованием высокопроизводительных вычислительных платформ. Научные ведомости БелГУ. Сер. Экономика. Информатика. 2(251): 153-160.

Ivanov M.I., Sorokin S.A., 2017. Processing images in the system of technical vision using high-performance computing platforms. Nauchnye vedomosti BelGU. Jekonomika. Informatika [Belgorod State University Scientific Bulletin. Economics Information technologies] 2(251): 153-160.