# Оценка производительности гибридных вычислительных систем на базе современных процессоров IBM POWER*

А. А. Сорокин, С. И. Мальковский

Вычислительный центр Дальневосточного отделения Российской академии наук, г. Хабаровск, Россия

**Аннотация.** Статья посвящена вопросам комплексного изучения аппаратного и программного обеспечения гибридных вычислительных систем на базе современных процессоров IBM семейства POWER и графических сопроцессоров NVIDIA Tesla. Исследована производительность подсистемы памяти и центральных процессоров при проведении параллельных вычислений с применением различных технологий параллельного программирования. Изучена эффективность функционирования математических библиотек, в том числе предусматривающих выгрузку вычислений на сопроцессоры. По результатам проведенной работы даны базовые рекомендации по использованию оборудования подобного класса для решения различных научных задач.

**Ключевые слова:** гибридная вычислительная система, архитектура компьютера, IBM POWER8, IBM POWER9, Intel Xeon Platinum 8160, математическая библиотека, одновременная многопоточность, производительность, тест

**DOI** 10.14357/20718632210303

## Введение

Появление высокопроизводительных алгоритмов и программных средств обработки данных, использующих возможности современных гибридных вычислительных платформ [1], привело к росту востребованности подобных аппаратных систем. При кажущейся ограниченности выбора доступных решений можно выделить несколько производителей оборудования (Intel, IBM и др.), которые развивают относительно самостоятельную экосистему, включающую также наборы различных библиотек и компиляторов. Их производительность может существенно отличаться на разных классах задач. Стоит отметить, что при изучении эффективности новых гибридных вычислительных систем недостаточно рассмотреть производительность использующихся в них ускорителей вычислений. Особого внимания требует оценка влияния на общую производительность центральных процессоров, которые помимо координирующих функций, присущих гибридным платформам, также выполняют операции, связанные непосредственно с численными расчетами, а также шин передачи данных, обеспечивающих взаимодействие всех компонентов системы. В связи с этим, существует необходимость проведения комплексных исследований существующих вычислительных архитектур, направленных на разноплановую

* Работа выполнена при частичной финансовой поддержке РФФИ, грант №18-29-03196

оценку работы оборудования и сопутствующего стека программного обеспечения.

Исследованиям рассматриваемых проблем посвящено большое количество работ, однако большая часть из них связана с анализом функционирования систем с архитектурой x86-64. По этой причине полученные в них выводы нельзя в полной мере применить к альтернативным архитектурам (ARM, POWER, SPARC и др.), в последние годы получившим широкое распространение в российских и международных суперкомпьютерных центрах (по данным рейтинга Top500 самых мощных суперкомпьютеров мира за июнь 2021 года доля таких систем по пиковой производительности составляет 33%). Изучению возможностей одной из таких гибридных архитектур посвящена настоящая работа.

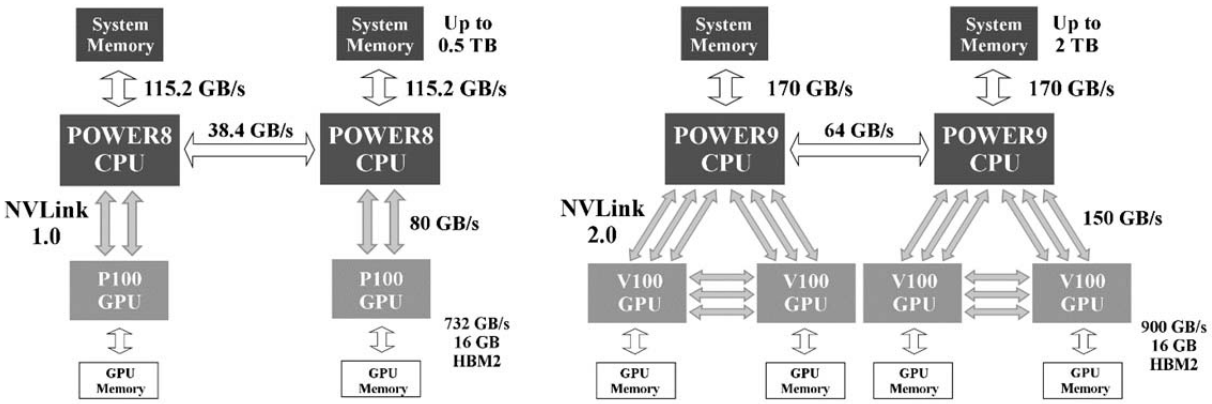
В статье рассматриваются результаты комплексного исследования гибридных вычислительных систем на базе современных процессоров (CPU) IBM POWER и графических сопроцессоров (GPU) NVIDIA Tesla, а также проводится их сравнение с соответствующими оценками производительности гибридной вычислительной системы, построенной с использованием процессоров Intel Xeon. На основе полученных результатов даны рекомендации по использованию систем с архитектурой POWER в решении отдельных научных задач, а также по выбору оптимальных вычислительных платформ для организации работы суперкомпьютерных центров.

## 1. Аппаратное и программное обеспечение вычислительных систем

В работе были исследованы современные гибридные вычислительные системы на архитектуре POWER: IBM Power System S822LC 8335-GTB (далее – система IBM POWER8) и IBM Power System AC922 8335-GTG (далее – система IBM POWER9). Полученные результаты сравнивались с показателями производительности системы Huawei FusionServer G5500 Server G560 V5, построенной на процессорах Intel Xeon (далее – система Intel Xeon Platinum 8160). Рассмотрим ключевые характеристики перечисленных систем подробнее.

Вычислительная система IBM POWER8 (Рис. 1, а) включает два процессора IBM POWER8 [2] с максимальной частотой 4,023 ГГц (пиковая производительность 0,322 ТФлопс), два сопроцессора NVIDIA Tesla P100 GPU, 256 ГБ DDR4 ОЗУ, два жестких диска Seagate 1 ТБ 7200RPM и контроллер EDR InfiniBand.

IBM POWER8 – это десятиядерный суперскалярный процессор, базирующийся на RISC (Reduced Instruction Set Computer - компьютер с сокращенным набором команд) архитектуре POWER [2]. Он поддерживает технологию одновременной многопоточности (SMT) [3], позволяющую запускать до восьми аппаратных потоков на ядро, и внеочередное выполнение команд. Список всех доступных режимов работы технологии SMT приводится в Табл. 1.




(а) (б)

Рис. 1. Архитектура вычислительной системы IBM POWER8 (а) и IBM POWER9 (б)

Табл. 1 Режимы работы процессорных ядер

| Название | Аппаратных потоков на ядро |
|---|---|
| ST | 1 |
| SMT2 | 2 |
| SMT4 | 4 |
| SMT8 | 8 |

Каждое ядро CPU содержит 32 КБ кэша L1 инструкций, 64 КБ кэша L1 данных и 512 КБ кэша L2. Также на каждое ядро приходится 8 МБ eDRAM кэша третьего уровня, разделяемого между всеми вычислительными ядрами (всего 80 МБ на процессор) [4]. Дополнительный разделяемый кэш L4, реализованный по технологии eDRAM, находится вне процессора на микросхемах Centaur, предназначенных для планирования и управления обменами данными с памятью (по микросхеме на канал памяти). Его объем составляет 16 МБ на микросхему, поэтому на процессор IBM POWER8, имеющий четырехканальный контроллер памяти, приходится 64 МБ кэша L4. Теоретическая пропускная способность доступа к памяти для системы IBM POWER8 составляет 230,4 ГБ/с.

Сопроцессоры NVIDIA Tesla P100 GPU (Tesla P100-SXM2-16GB; далее – P100-NVL1), которыми оснащена система IBM POWER8, реализованы на архитектуре Pascal [5]. Графический процессор каждого из них содержит 56 потоковых процессора (SM), включающих 64 CUDA ядра для выполнения операций над числами с плавающей запятой одинарной точности, 32 CUDA ядра для выполнения операций над числами с плавающей запятой двойной точности, 24 КБ разделяемого кэша L1 и 64 КБ разделяемой памяти. Таким образом для выполнения операций над числами двойной точности могут использоваться 1792 CUDA ядра. Помимо кэша L1, доступного в рамках одного SM, этим ядрам доступно 4 МБ кэша L2 и 16 ГБ памяти HBM2 (полоса пропускания 732 ГБ/с). Пиковая производительность одного графического процессора частотой 1,48 ГГц на операциях с числами двойной точности (при выполнении операций FMA) составляет 5,3 ТФлопс. Сопроцессоры подключаются к центральным процессорам вычислительной системы по шине NVLink версии 1.0 с пропускной способностью в 80 ГБ/с.

Пиковая производительность системы IBM POWER8 составляет 11,2 ТФлопс, большая часть которой приходится на графические процессоры.

Система IBM POWER9 построена на основе гибридной архитектуры с использованием нового поколения суперскалярных процессоров IBM. В её состав (Рис. 1, б) входят два процессора IBM POWER9 [6] с максимальной частотой 3,5 ГГц и пиковой производительностью 0,56 ТФлопс, четыре сопроцессора NVIDIA Tesla V100 GPU, 1024 ГБ DDR4 ОЗУ, два твердотельных накопителя Micron 5100 Pro объемом 960 ГБ, а также контроллер EDR InfiniBand. Каждый из двух центральных процессоров имеет 20 вычислительных ядер, поддерживающих технологию SMT для четырех потоков (80 аппаратных потоков на сокет). Вычислительные ядра содержат 32 КБ кэша L1 инструкций и 32 КБ кэша L1 данных. На каждую пару ядер IBM POWER9 приходится 512 КБ кэша L2 и 10 МБ eDRAM кэша L3. В вычислительной системе используется SO (Scale-out) версия процессора IBM POWER9, отличающаяся от SU (Scale-up) версии тем, что оперативная память подключается к двум четырехканальным контроллерам памяти напрямую, без применения промежуточной микросхемы Centaur. При этом пиковая пропускная способность доступа к памяти составляет 340 ГБ/с [7]. К каждому центральному процессору сервера посредством шины NVLink 2.0 с пропускной способностью 150 ГБ/с подключено по два графических сопроцессора NVIDIA Tesla V100 (Tesla V100-SXM2-16GB; далее V100-NVL2), построенных на архитектуре Volta [8]. По сравнению с GPU Tesla P100, в указанных сопроцессорах число SM было увеличено до 80. Также в каждый SM было добавлено по 8 тензорных ядер, предназначенных для выполнения операций над матрицами. При этом количество ядер для выполнения операций над числами с плавающей запятой, приходящихся на SM, осталось прежним. Каждому потоковому процессору доступно 128 КБ комбинированного кэша, представляющего собой объединенные кэш L1 и разделяемую память. Объем разделяемого между всеми SM кэша L2 составляет 6 МБ. Сопроцессор содержит 16 ГБ HBM2

памяти с пропускной способностью 900 ГБ/с. По сравнению с предыдущим поколением графических процессоров, использующихся в системе IBM POWER8, максимальная частота графического процессора NVIDIA Tesla V100 была увеличена до 1530 МГц. Указанные изменения позволили поднять пиковую производительность GPU до 7,8 ТФлопс. Таким образом, пиковая производительность всей системы IBM POWER9 составляет 32,3 ТФлопс, что практически в 3 раза больше производительности системы IBM POWER8.

Используемая для сравнения система Intel Xeon Platinum 8160 (Рис. 2) создана на базе двух суперскалярных процессоров Intel Xeon Platinum 8160 с максимальной частотой 2,1 ГГц, основанных на микроархитектуре Skylake [9], и восьми сопроцессоров NVIDIA Tesla V100 (Tesla V100-SXM2-32GB; далее – V100-PE3). Для хранения данных используются два жестких диска Seagate 10000RPM 1,8 ТБ. Центральные процессоры вычислительной системы поддерживают технологию SMT (Hyper-Threading), позволяющую выполнять на каждом ядре до 2 аппаратных вычислительных потоков. Таким образом, на одном процессоре с 24 ядрами можно выполнять 48 аппаратных потоков. Каждое процессорное ядро содержит 32 КБ кэша L1 инструкций, 32 КБ кэша L1 данных, а также 1024 КБ кэша L2. Дополнительно на процессор приходится 33 МБ разделяемого кэша третьего уровня. Пиковая пропускная способность доступа к памяти объемом 1,5 ТБ для системы Intel Xeon Platinum 8160 составляет 256 ГБ/с.

В системе установлены сопроцессоры, аналогичны тем, что используются в системе IBM POWER9 за исключением несколько меньшей

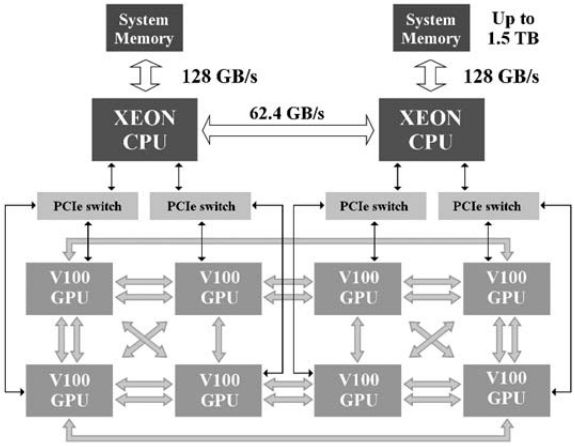



Рис. 2. Архитектура вычислительной системы Intel Xeon Platinum 8160

(каждая тонкая черная стрелка соответствует 16 линиям PCIe 3.0, каждая серая стрелка – одной линии NVLink 2.0; память сопроцессоров на схеме не отображена)

базовой частоты (1290 МГц) и удвоенного объема памяти HBM2. Еще одним отличием является то, что к центральным процессорам они подключаются с использованием шины PCIe 3.0 с пропускной способностью 32 ГБ/с.

Пиковая производительность системы Intel Xeon Platinum 8160 составляет 65,48 ТФлопс, из которых на каждый центральный процессор приходится 1,54 ТФлопс, а на каждый сопроцессор – 7,8 ТФлопс.

В качестве компиляторов и библиотек MPI, необходимых для проведения работ, использовались программные средства (Табл. 2), которые согласно ранее полученным оценкам [10, 11] обеспечивают оптимальный уровень производительности при расчетах на указанных типах вычислительных систем. Для сборки теста gearshifft использовался компилятор GCC C++ 8.3.1.

Табл. 2. Перечень программного обеспечения

| | IBM POWER8 | IBM POWER9 | Intel Xeon Platinum 8160 |
|---|---|---|---|
| Операционная система | CentOS 7.6 | RedHat 7.6 | CentOS 7.6 |
| Компилятор | IBM XL C/C++, Fortran 16.1.1 | IBM XL C/C++, Fortran 16.1.1 | Intel C/C++, Fortran 19.0.5 |
| Библиотека MPI | IBM Spectrum MPI 10.3 | IBM Spectrum MPI 10.3 | Intel MPI Library 2019 Update 5 |
| NVIDIA CUDA Toolkit | 10.1 | 10.1 | 10.1 |
| Драйвер GPU | 418.40.04 | 418.87.00 | 418.87.00 |

## 2. Методика оценки производительности вычислительных систем

В процессе исследований рассматривались вопросы производительности аппаратного обеспечения при выполнении параллельных приложений, а также различных реализаций математических библиотек. Оценивалась пропускная способность подсистем памяти, проводился анализ эффективности исследуемых вычислительных систем при проведении параллельных вычислений. В качестве тестовых задач также рассматривалось функционирование различных реализаций математических библиотек, предназначенных для выполнения быстрого преобразования Фурье (БПФ) и базовых операций линейной алгебры (BLAS), предусматривающих использование ресурсов центральных процессоров и ускорителей вычислений. В работе использовались общепризнанные тесты, перечисленные в Табл. 3.

Для оценки реальной пропускной способности подсистемы памяти использовался тест STREAM. Он позволяет изучить установившуюся пропускную способность при выполнении операций чтения и записи, выполняющихся совместно с арифметическими операциями. Тест содержит четыре вычислительных ядра Copy, Scale, Add и Triad, описание которых представлено в Табл. 4.

Вычислительные ядра работают с массивами чисел двойной точности (8 Б), размер которых устанавливается во время компиляции теста. Для корректной оценки пропускной способности памяти они должны быть как минимум в 4 раза большими объема всей доступной кэш памяти самого верхнего уровня. Поэтому при тестировании рассматриваемых вычислительных систем использовался массив размера 8,4E+7 элементов.

Результаты выполненных тестов сравнивались с теоретической пропускной способностью доступа к памяти для исследуемых систем.

Анализ эффективности исследуемых вычислительных систем при проведении параллельных вычислений с использованием технологий OpenMP и MPI выполнялся с использованием теста NPB. Он состоит из ряда простых задач: ядер и приложений. Ядра и приложения могут производить вычисления в классах сложности: S, W, A, B, C, D. С увеличением класса сложности возрастает и размерность основных массивов, данных и количество итераций в основных циклах программ. В работе использовались следующие ядра и приложения из состава NPB: EP, LU, MG, CG, FT и IS.

При выполнении указанных выше экспериментов на системе IBM POWER8 с числом потоков, не превышающим 20, применялся режим ST, а с числом потоков, равным 40, 80 и 160 – режимы SMT2, SMT4 и SMT8 соответственно. Процессоры IBM POWER9 и Intel Xeon Platinum 8160 имеют большее количество ядер, поэтому на использующих их системах при числе потоков, не превышающих 40, использовался режим ST. При числе потоков, равном 80 и 160, использовались режимы SMT2 и SMT4

Табл. 3. Перечень использованных тестов

| Исследуемая характеристика | Название теста | Версия |
|---|---|---|
| Пропускная способность подсистемы памяти | STREAM [12] | 5.10 |
| Производительность при выполнении параллельных приложений | NAS Parallel Benchmark (NPB) [13] | 3.4 |
| Производительности реализаций библиотек БПФ | gearshifft [14] | - |
| Производительности реализаций библиотеки BLAS | Crossroads/N9 DGEMM [15] | - |

Табл. 4. Вычислительные ядра теста STREAM

| Название | Операции | Байт/итерацию | Флопс/итерацию |
|---|---|---|---|
| Copy | a(i) = b(i) | 16 | 0 |
| Scale | a(i) = q*b(i) | 16 | 1 |
| Add | a(i) = b(i) + c(i) | 24 | 1 |
| Triad | a(i) = b(i) + q*c(i) | 24 | 2 |

(система IBM POWER9) соответственно. Вычислительные потоки равномерно распределялись по сокетам (на каждом запускалось по половине потоков).

Исследование производительности различных реализаций библиотеки БПФ выполнялось с использованием пакета gearshifft – расширяемой тестовой системы с открытым исходным кодом, написанной на языке C++. Запуск тестов, собранных с соответствующими версиями библиотеки БПФ (IBM ESSL 6.2, Intel MKL 2019 Update 5, а также cuFFT и cuFFTW, входящих в состав NVIDIA CUDA Toolkit), производился со следующими параметрами: основание – степень двойки, размерность задачи – 3D, тип входного сигнала – действительные двойной точности, тип преобразования – real-to-complex, режим использования памяти – in-place (структура входных данных используется для хранения выходных данных). В качестве входного сигнала использовался трехмерный массив вещественных чисел размерности от $2 \times 2 \times 2$ до $1024 \times 1024 \times 1024$, с равномерным увеличением размера задачи в 2 раза. Для устранения влияния на результаты случайных факторов тест запускался 20 раз, после чего вычислялось среднее время выполнения прямого БПФ, а также общее время работы теста, дополнительно включающее время, затрачиваемое на планирование преобразования, выделение памяти и т. д.

При выполнении экспериментов оценивалась производительность библиотек IBM ESSL и Intel MKL на центральных процессорах вычислительных систем в параллельном режиме. На графических сопроцессорах изучалась производительность библиотек cuFFT и cuFFTW.

Тестирование различных реализаций библиотеки BLAS (IBM ESSL 6.2, Intel MKL 2019 Update 5 и NVBLAS из состава NVIDIA CUDA Toolkit) проводилось с использованием пакета Crossroads/N9 DGEMM. Это простой многопоточный тест, выполняющий перемножение плотно заполненных матриц с использованием процедуры DGEMM. Указанная процедура является самой вычислительно затратной и, одновременно с этим, наиболее широко используемой подпрограммой BLAS. Этим объясняется возможность применения пакета Crossroads/N9 DGEMM для оценки производительности различных реализаций указанной библиотеки.

При выполнении экспериментов исследовалась производительность различных реализаций BLAS как на центральных процессорах в многопоточном режиме, так и на графических ускорителях. Для каждой комбинации исследуемых параметров производилось по 10 запусков теста, после чего полученные значения производительности усреднялись.

Во избежание получения ошибочных результатов, связанных с параллельным использованием вычислительного оборудования в иных целях, все численные расчеты производились в монопольном режиме. Динамическое управление частотой центральных процессоров отключалось. При этом устанавливалась максимальная частота процессоров, допустимая при одновременной максимальной загрузке всех вычислительных ядер.

## 3. Полученные результаты и их интерпретация

Проведенные исследования показали, что из трех рассматриваемых систем система IBM POWER9 имеет максимальную пропускную способность ОЗУ (Рис. 3). Это связано с бо́льшим числом каналов памяти, приходящимся на каждый процессор. При этом, несмотря на меньшую частоту микросхем памяти, система IBM POWER8 показывает более высокую измеренную пропускную способность ОЗУ при числе потоков, меньшем 10. Подобный результат можно объяснить используемыми в ней чипами Centaur, повышающими эффективность организации работы с памятью. Таким образом, вычислительные системы с архитектурой, схожей с архитектурой системы IBM POWER8, позволяют повысить производительность приложений с низкой степенью параллелизма и высокими требованиями к пропускной способности памяти. Вычислительная система Intel Xeon Platinum 8160 показала наихудшие результаты практически во всех ядрах STREAM. Лишь при использовании 40 потоков в ядрах Copy и Scale, наименее требовательных к пропускной способности ОЗУ, она оказалась быстрее системы IBM POWER8.

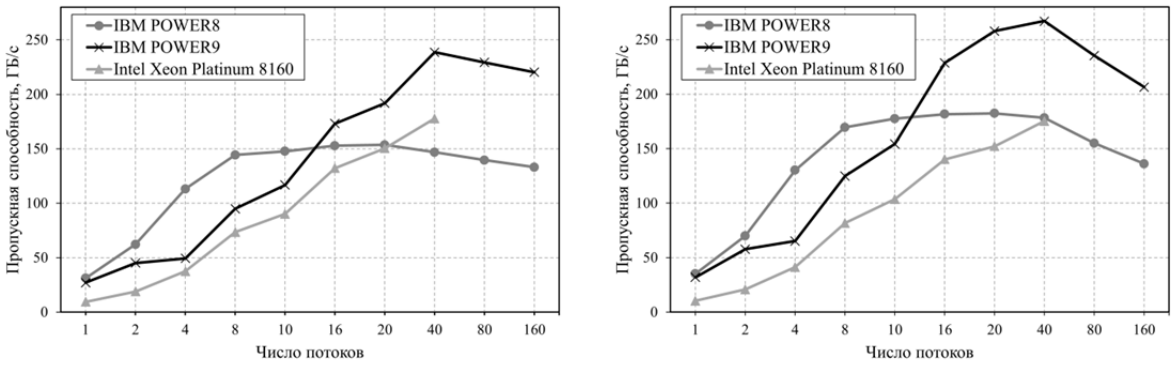



Рис. 3. Зависимость пропускной способности памяти от числа потоков в тесте STREAM Copy (а) и Triad (б)

Технология SMT повышает утилизацию процессорных ядер при выполнении не оптимизированных приложений, чья производительность ограничена скоростью выполнения вычислительных операций. Если же производительность приложения ограничена пропускной способностью памяти, то при использовании технологии одновременной многопоточности может наблюдаться снижение производительности из-за увеличения числа кэш-конфликтов, что приводит к снижению скорости доступа к ОЗУ (Рис. 3). Среди исследованных процессоров наибольшую производительность на ядро в тестах NPB продемонстрировал процессор системы IBM POWER8 благодаря большей частоте и возможности запуска до 8 аппаратных потоков на ядро. Однако, за счет большего числа вычислительных ядер, процессор IBM POWER9 оказался производительнее процессора IBM POWER8 в среднем в 1,3 раза при разнице в пиковой производительности между ними в 1,75 раза (Рис. 4). Такая разница в росте пиковой и реальной производительности может быть объяснена тем, что технология SMT на процессоре IBM POWER9 оказалась менее эффективной, чем на процессоре IBM POWER8.

Наихудшую эффективность при выполнении неоптимизированных приложений показала система Intel Xeon Platinum 8160. Несмотря на почти в три раза большую пиковую производительность установленных в ней процессоров, по сравнению с процессорами системы IBM POWER9, достигаемую за счет применения векторных инструкций длиной 512 бит, её реальная производительность в тесте NPB оказа-

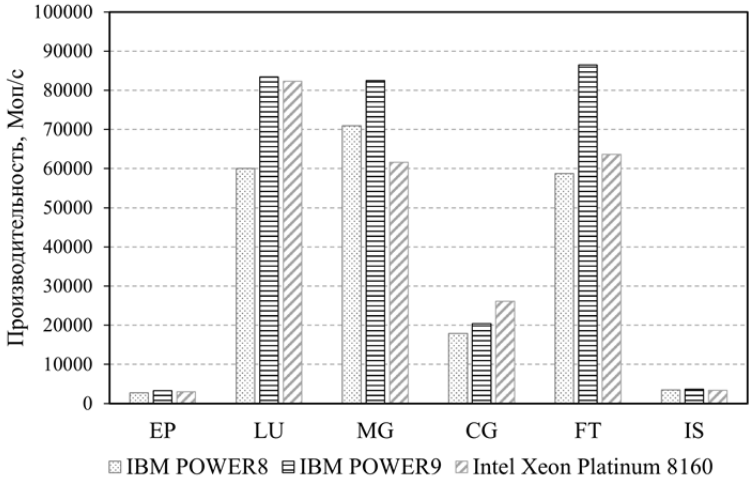



Рис. 4. Максимальная производительность вычислительных систем в тесте NPB (класс C)

результаты получены на системе IBM POWER8 с использованием 16 ядер, а на системах IBM POWER9 и Intel Xeon Platinum 8160 – с использованием 32 ядер, что связано с ограничениями теста

лась сопоставима (длина векторных инструкций процессоров IBM POWER8 и POWER9 составляет 128 бит).

На всех системах большинство исследованных приложений, разработанных с использованием технологии OpenMP, демонстрируют несколько более высокий уровень производительности нежели приложения, разработанные с использованием технологии MPI. При этом наибольший прирост производительности технология OpenMP обеспечивает в случаях наличия интенсивного взаимодействия между вычислительными потоками. С другой стороны, за счет того, что технология MPI не подразумевает использования общей памяти, а все взаимодействия между вычислительными процессами происходят за счет обмена сообщениями, приложения, разработанные с её использованием, демонстрируют более высокий уровень локальности данных, чем приложения, основанные на технологии OpenMP. Это позволяет повышать эффективность применения технологии SMT на системах с большими объемами кэш памяти за счет снижения числа кэш-конфликтов при использовании большого числа процессов.

Исследование производительности различных реализаций математических библиотек, результаты которого представлены на Рис. 5-10, показало, что на вычислительных системах IBM POWER при использовании центральных процессоров библиотеки БПФ (Рис. 5) демонстрируют б*о*льшую (при учете вспомогательных операций), а библиотеки BLAS (Рис. 8) – меньшую производительность по сравнению с производительностью библиотеки Intel MKL на вычислительной системе Intel Xeon Platinum 8160. При этом использование графических сопроцессоров позволяет значительно повысить скорость их работы (Рис. 6 и 9).

Применение шины NVLink в системах IBM существенно сокращает время на копирование данных между процессором и сопроцессором, что уменьшает общее время выполнения БПФ и подпрограммы DGEMM. Несмотря на это автоматическая выгрузка вычислений на сопроцессоры с использованием библиотеки cuFFTW не позволила получить преимущества по сравнению с выполнением БПФ на центральных процессорах (Рис. 7) из-за значительных временных затрат на копирование данных из оперативной памяти в память сопроцессора. Поэтому использование указанной библиотеки оправдано лишь при необходимости освобождения центрального процессора для других вычислений и невозможности или нецелесообразности модификации исходного кода используемого приложения.

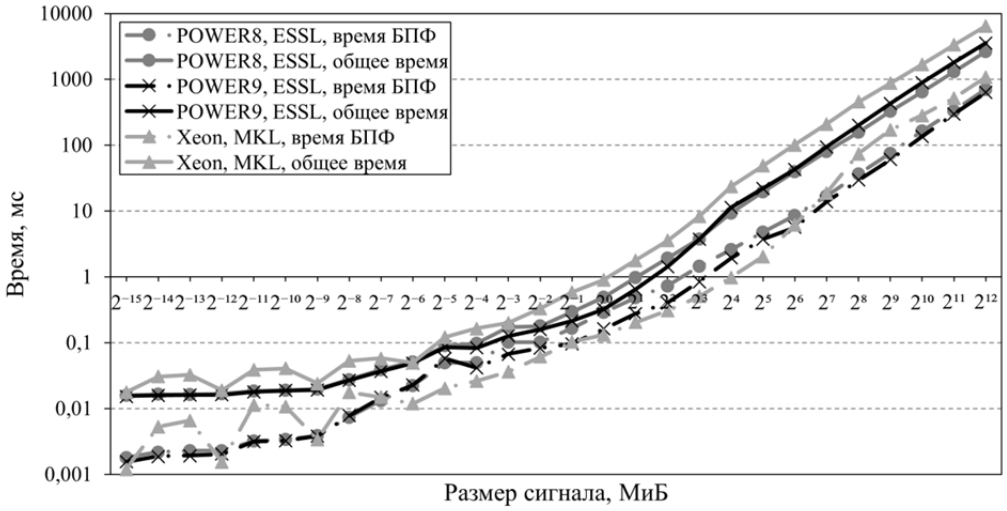



Рис. 5. Зависимость минимального времени выполнения прямого БПФ и общего минимального времени выполнения теста от размера сигнала на центральных процессорах вычислительных систем

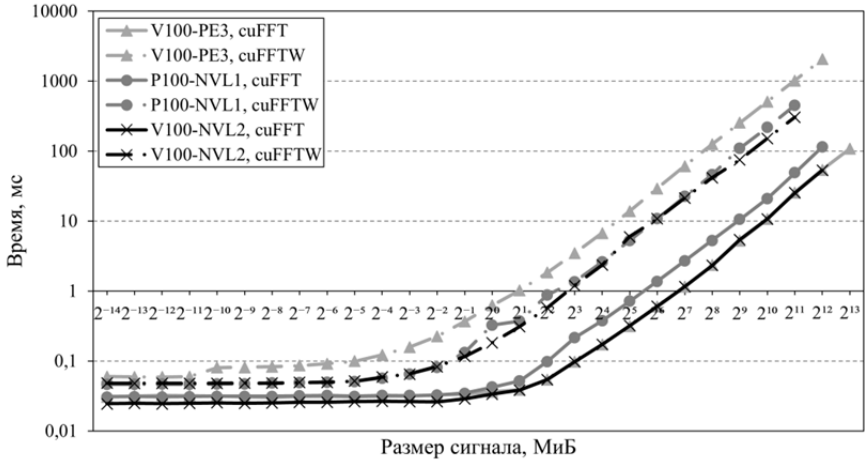



Рис. 6. Зависимость времени выполнения прямого БПФ от размера сигнала на сопроцессорах вычислительных систем

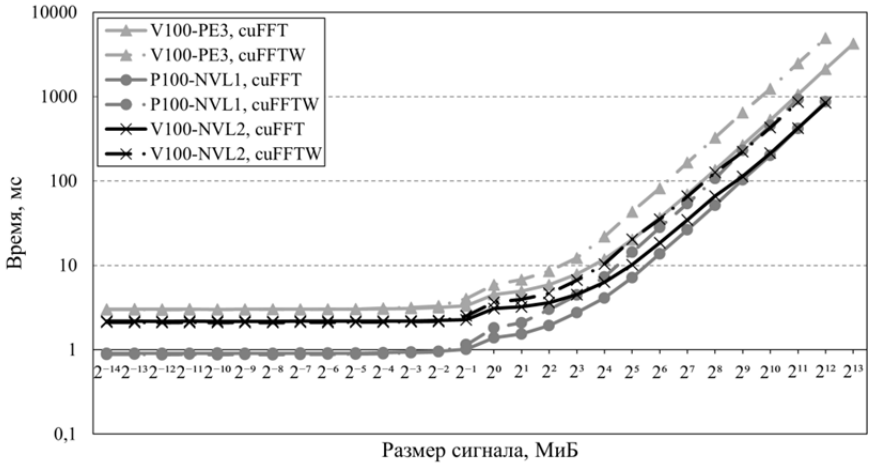



Рис. 7. Зависимость общего времени выполнения теста от размера сигнала на сопроцессорах вычислительных систем

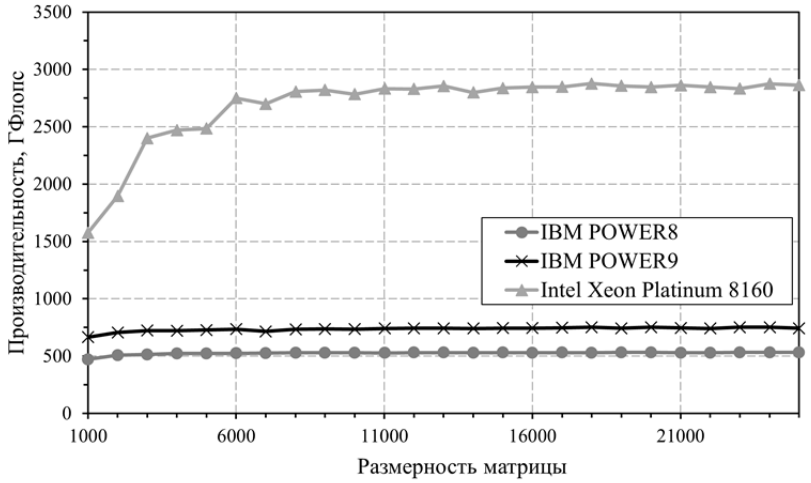



Рис. 8. Зависимость производительности, достигаемой центральными процессорами вычислительных систем, от размерности матриц в тесте DGEMM при использовании всех доступных ядер (режим ST); на системах IBM POWER8 и POWER9 использовалась библиотека IBM ESSL, на системе Intel Xeon Platinum 8160 – Intel MKL

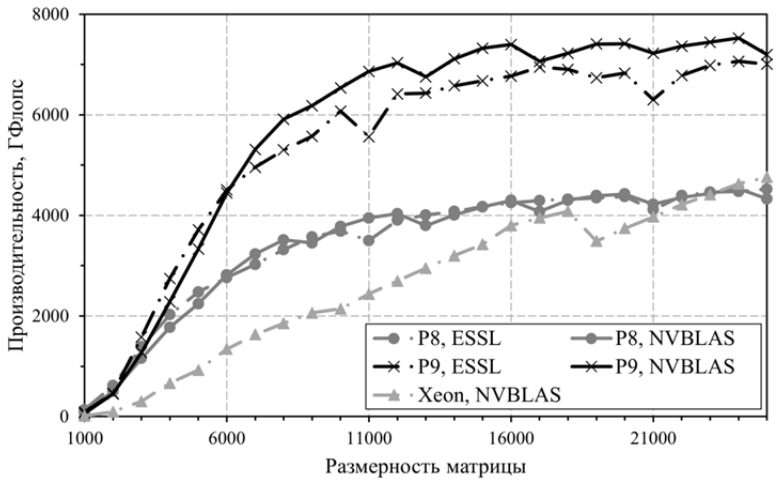



Рис. 9. Зависимость производительности, достигаемой при использовании одного сопроцессора, от размерности матриц в тесте DGEMM

В отличие от библиотеки cuFFTW, библиотеки BLAS (IBM ESSL и NVBLAS) допускают эффективную автоматическую выгрузку вычислений на сопроцессоры. Используя эту возможность, а также высокую пропускную способность шины NVLink на системах IBM POWER, можно построить высокопроизводительное комплексное решение для задач, использующих подпрограммы BLAS третьего уровня, без необходимости значительной переработки существующего программного обеспечения. Указанная шина также позволяет повысить эффективность распараллеливания вычислений на несколько GPU и увеличить утилизацию сопроцессоров при их совместном использовании несколькими процессами. В качестве примера на Рис. 10 приведены результаты теста DGEMM в режиме автоматической выгрузки вычислений на два сопроцессора. Из них видно, что обе системы IBM демонстрируют значительный прирост производительности по сравнению с выполнением вычислений на одном сопроцессоре (на 59% и 92% на вычислительных системах IBM POWER8 и IBM POWER9 соответственно). При этом использование шины PCIe 3.0 для связи центрального процессора и сопроцессора на системе Intel Xeon Platinum 8160 позволяет поднять произ-

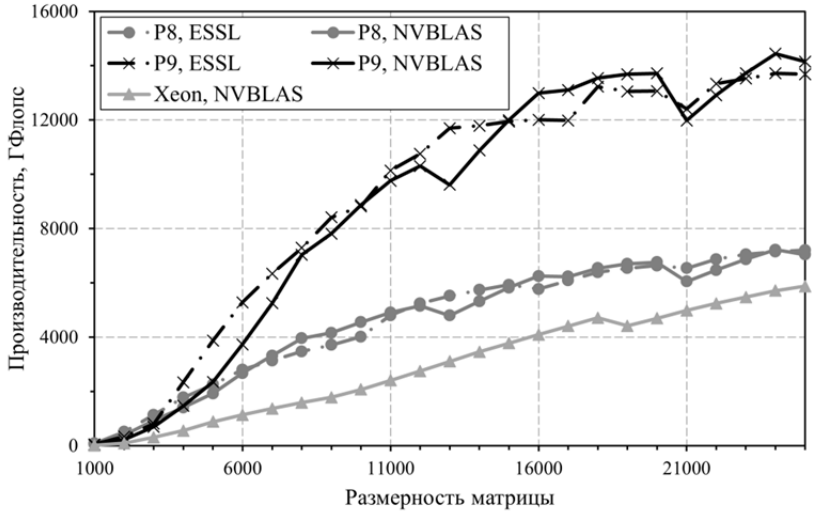



Рис. 10. Зависимость производительности, достигаемой при использовании двух сопроцессоров, от размерности матриц в тесте DGEMM

водительность при добавлении второго GPU лишь на 22%, а дальнейшее увеличение их числа приводит к её снижению.

## Заключение

В настоящее время высокопроизводительные вычислительные системы являются одним из наиболее востребованных элементов научной инфраструктуры, обеспечивающим проведение современных исследований в различных областях знаний. При этом наблюдавшееся в последние годы активное развитие систем с гибридной архитектурой, основанных на совместном использовании центральных процессоров и различных сопроцессоров (в первую очередь графических), а также создание для них эффективных алгоритмов параллельной обработки данных и средств разработки приложений, открыли дополнительные возможности по применению в науке современных вычислительных технологий.

Однако с точки зрения центров коллективного пользования (ЦКП), при внедрении новых суперкомпьютеров возникают вопросы, связанные с выбором подходов к организации эффективной вычислительной среды, удовлетворяющей требованиям отдельных научных задач и использующихся при их решении компьютерных алгоритмов. Для получения ответа на них и было проведено исследование вычислительных систем на базе современных процессоров IBM POWER и сопроцессоров NVIDIA Tesla. Оно показало их высокую эффективность в решении как задач, связанных с использованием методов машинного обучения и глубокого обучения [16], так и классических вычислительных задач (первопринципные расчеты свойств новых материалов [17], численное моделирование сложных физических процессов [18] и т. д.) в ЦКП. Большой вклад в достижении таких результатов вносят доступность для указанных систем стандартного системного программного обеспечения, широко применяющегося на суперкомпьютерах (операционная система GNU/Linux, планировщики заданий, системы мониторинга ресурсов, библиотеки MPI), использование высокоскоростной шины NVLink для связи процессоров и сопроцессоров, а также реализованные возможности математических библиотек и компиляторов, обеспечивающих выгрузку вычислений на ускорители даже для немодифицированного программного кода.

Перечисленные особенности систем IBM POWER позволяют рассматривать их как основу для организации универсальной вычислительной платформы на базе гибридных архитектур. Такой вывод подтверждают полученные результаты экспериментальных исследований, которые показывают, что при выполнении неоптимизированного программного кода более эффективными по сравнению с системой Intel являются системы IBM. Это связано как с более высокой тактовой частотой использующихся в них процессоров, так и с возможностями технологии SMT, повышающей утилизацию исполнительных устройств CPU. При этом б*о*льшая производительность на ядро, продемонстрированная системами IBM в большинстве тестов, позволяет эффективно использовать ресурсы центральных процессоров во время выполнения гибридных приложений, для работы которых достаточно запуска одного вычислительного процесса на сопроцессор. Стоит также отметить, что используемая в этих системах шина NVLink и более высокая по сравнению с системой Intel пропускная способность оперативной памяти решает проблему низкой скорости загрузки данных в память ускорителя. Указанные особенности позволяют эффективно выгружать массивно-параллельные вычисления на один или несколько сопроцессоров, что компенсирует отсутствие поддержки векторных инструкций большой длины процессорами IBM. Это подтверждается результатами, продемонстрированными системами IBM в тесте DGEMM.

Важно отметить, что при выполнении параллельных вычислений исключительно на центральном процессоре с использованием уже оптимизированного прикладного программного обеспечения (векторизация вычислений, применение оптимизированных библиотек, например, Intel MKL), б*о*льшую производительность будут иметь системы на базе процессоров Intel. Применение в указанных системах сопроцессоров ограничено использованием низкоскоростной шины PCIe, которая затрудняет оперативную загрузку больших объемов данных в память

GPU. Это увеличивает трудозатраты на разработку приложений, работающих в режиме выгрузки вычислений, из-за необходимости оптимизации работы с памятью и, в общем случае, снижает эффективность работы сопроцессоров.

При проведении численных расчетов использовано оборудование ЦКП "Центр данных ДВО РАН" (ВЦ ДВО РАН, г. Хабаровск) [19] и ЦКП "Информатика" Федерального исследовательского центра "Информатика и управление РАН" (г. Москва) [20].

## Литература


1. Brodtkorb A.R., Dyken C., Hagen T.R., Hjelmervik J.M., Storaasli, O.O. State-of-the-art in Heterogeneous Computing // Scientific Programming. 2010. V. 18. No. 1. P. 1–33. DOI: 10.1155/2010/540159.
2. Sinharoy B., Van Norstrand J.A., Eickemeyer R.J., Le H.Q., Leenstra J., Nguyen D.Q., Konigsburg B., Ward K., Brown M.D., Moreira J.E., Levitan D., Tung S., Hrusecky D., Bishop J.W., Gschwind M., Boersma M., Kroener M., Kaltenbach M., Karkhanis T., Fernsler K.M. IBM POWER8 processor core microarchitecture // IBM Journal of Research and Development. 2015. Vol. 59, No. 1. P. 2:1–2:21. DOI: 10.1147/JRD.2014.2376112.
3. Eggers S.J., Emer J.S., Levy H.M., Lo J.L., Stamm R.L., Tullsen D.M. Simultaneous multithreading: a platform for next-generation processors // IEEE Micro. 1997. Vol. 17, No. 5. P. 12–19. DOI: 10.1109/40.621209.
4. Starke W.J., Stuecheli J., Daly D.M., Dodson J.S., Auernhammer F., Sagmeister P.M., Guthrie G.L., Marino C.F., Siegel M., Blaner B. The cache and memory subsystems of the IBM POWER8 processor // IBM Journal of Research and Development. 2015. Vol. 59, No. 1. P. 3:1-3:13. DOI: 10.1147/JRD.2014.2376131.
5. Foley D., Danskin J. Ultra-Performance Pascal GPU and NVLink Interconnect // IEEE Micro. 2017. Vol. 37. No. 2. P. 7–17. DOI: 10.1109/MM.2017.37.
6. Sadasivam S.K., Thompto B.W., Kalla R., Starke W.J. IBM Power9 Processor Architecture // IEEE Micro. 2017. Vol. 37. No. 2. P. 40–51. DOI: 10.1109/MM.2017.40.
7. Starke W.J., Dodson J.S., Stuecheli J., Retter E., Michael B.W., Powell S.J., Marcella J.A. IBM POWER9 memory architectures for optimized systems // IBM Journal of Research and Development. 2018. Vol. 62. No. 4/5. P 3:1–3:13. DOI: 10.1147/JRD.2018.2846159.
8. Choquette J., Giroux O., Foley D. Volta: Performance and Programmability // IEEE Micro. 2018. Vol. 38. No. 2. P. 42–52. DOI: 10.1109/MM.2018.022071134.
9. Mulnix D. Intel Xeon Processor Scalable Family Technical Overview // [Электронный ресурс] – Режим доступа https://software.intel.com/ru-ru/articles/intel-xeon-processor-scalable-family-technical-overview (дата обращения 08.04.2020).
10. Mal'kovskii S. I., Sorokin A. A., Korolev S. P., Zatsarinnyi A. A., Tsoi G. I. Performance Evaluation of a Hybrid Computer Cluster Built on IBM POWER8 Microprocessors // Programming and Computer Software. 2019. Vol. 45. No. 6, P. 324-332. DOI: 10.1134/S0361768819060057.
11. Мальковский С.И., Пересветов В.В. Оценка производительности вычислительного кластера на четырехъядерных процессорах // Материалы межрегиональной научно-практической конференции «Информационные и коммуникационные технологии в образовании и научной деятельности» 21-23 сентября 2009 года, г. Хабаровск. 2009. С. 261–268.
12. McCalpin J.D. Memory Bandwidth and Machine Balance in Current High Performance Computers // IEEE Technical Committee on Computer Architecture Newsletter. 1995. P. 19-25.
13. Bailey, D.; Barszcz, E.; Barton, J.; Browning, D.; Carter, R.; Dagum, L.; Fatoohi, R.; Fineberg, S.; Frederickson, P.; Lasinski, T.; Schreiber, R.; Simon, H.; Venkatakrishnan, V.; Weeratunga, S. The NAS Parallel Benchmarks. RNR Technical Report RNR 94-007 // [Электронный ресурс] – Режим доступа https://www.davidhbailey.com/dhbpapers/npb.pdf (дата обращения 07.05.2020).
14. Steinbach P., Werner M. gearshifft – The FFT Benchmark Suite for Heterogeneous Platforms. In: Kunkel J., Yokota R., Balaji P., Keyes D. (eds) High Performance Computing. ISC 2017. Lecture Notes in Computer Science. 2017. Vol 10266. Springer, Cham. P. 199–216. DOI: 10.1007/978-3-319-58667-0_11.
15. DGEMM // Электронный ресурс – Режим доступа https://web.archive.org/web/20180408033423/http://www.nersc.gov/research-and-development/apex/apex-benchmarks/dgemm/ (Дата обращения 08.04.2018).
16. Никитин О.Ю., Лукьянова О.А. Анализ ускорения глубокого обучения на основе вычислительной системы IBM POWER8 // Материалы V международной научно-практической конференции «Информационные технологии и высокопроизводительные вычисления» 16–19 сентября 2019 года, г. Хабаровск. 2019. С. 199–203.
17. Карцев А.И., Мальковский С.И., Волович К.И., Сорокин А.А. Исследование производительности и масштабируемости пакета Quantum ESPRESSO при изучении низкоразмерных систем на гибридных вычислительных системах // Материалы I международной конференции «Математическое моделирование в материаловедении электронных компонентов» 21–23 октября 2019 года, г. Москва. 2019. С. 18–20.
18. Волков К.Н., Добров Ю.В., Карпенко А.Г., Мальковский С.И., Сорокин А.А. Моделирование газовой динамики гиперзвуковых летательных аппаратов с использованием модели высокотемпературного воздуха и графических процессоров // Вычислительные методы и программирование. 2021. Т. 22. С. 29–46. DOI: 10.26089/NumMet.v22r103.
19. Sorokin A.A., Makogonov S.V., Korolev S.P. The Information Infrastructure for Collective Scientific Work in the Far East of Russia // Scientific and Technical Information Processing. 2017. Vol. 4. P. 302–304. DOI: 10.3103/S0147688217040153.
20. Положение о ЦКП «Информатика» // [Электронный ресурс] – Режим доступа http://www.frccsc.ru/ckp (дата обращения 22.01.2020).

**Сорокин Алексей Анатольевич.** Вычислительный центр Дальневосточного отделения Российской академии наук, г. Хабаровск, Россия. Главный научный сотрудник, кандидат технических наук, автор более 100 печатных работ. Область научных интересов: цифровые платформы, высокопроизводительные вычисления, проблемно-ориентированные базы данных, системы компьютерной поддержки научных исследований, информационно-телекоммуникационные системы. E-mail: alsor@febras.net

**Мальковский Сергей Иванович.** Вычислительный центр Дальневосточного отделения Российской академии наук, г. Хабаровск, Россия. Научный сотрудник, кандидат технических наук. Количество печатных работ: 23 (в т. ч. одна монография). Область научных интересов: высокопроизводительные вычисления, гибридные вычислительные системы, GRID, параллельные алгоритмы, нейронные сети, искусственный интеллект, машинное обучение, глубокое обучение, математическое моделирование. E-mail: sergey.malkovsky@ccfebras.ru

# Performance Evaluation of Heterogeneous Computing Systems Based on Modern IBM POWER Processors

A. A. Sorokin, S. I. Malkovsky

Computing Center of the Far Eastern Branch of the Russian Academy of Sciences, Khabarovsk, Russia

**Abstract.** The article is devoted to the complex study of hardware and software of heterogeneous computing systems based on modern IBM POWER processors and NVIDIA Tesla graphics coprocessors. Using the various parallel programming technologies, the performance of the memory subsystem and central processors is investigated in parallel mode. The effectiveness of the functioning of math libraries has been studied, including those providing offloading calculations to the coprocessor. Basic recommendations on the use of this class equipment for solving various scientific problems are given, based on the results of the work carried out.

**Keywords:** heterogeneous computing system, computer architecture, IBM POWER8, IBM POWER9, Intel Xeon Platinum 8160, GPU, math library, simultaneous multithreading, performance, benchmark.

**DOI** 10.14357/20718632210303

## References

1. Brodtkorb A.R., Dyken C., Hagen T.R., Hjelmervik J.M., Storaasli, O.O. 2010. State-of-the-art in Heterogeneous Computing. Scientific Programming. 18(1):1–33. DOI: 10.1155/2010/540159.
2. Sinharoy B., Van Norstrand J.A., Eickemeyer R.J., Le H.Q., Leenstra J., Nguyen D.Q., Konigsburg B., Ward K., Brown M.D., Moreira J.E., Levitan D., Tung S., Hrusecky D., Bishop J.W., Gschwind M., Boersma M., Kroener M., Kaltenbach M., Karkhanis T., Fernsler K.M. 2015. IBM POWER8 processor core microarchitecture. IBM Journal of Research and Development. 59(1):2:1–2:21. DOI: 10.1147/JRD.2014.2376112.
3. Eggers S.J., Emer J.S., Levy H.M., Lo J.L., Stamm R.L., Tullsen D.M. 1997. Simultaneous multithreading: a platform for next-generation processors. IEEE Micro. 17(5):12–19. DOI: 10.1109/40.621209.
4. Starke W.J., Stuecheli J., Daly D.M., Dodson J.S., Auernhammer F., Sagmeister P.M., Guthrie G.L., Marino C.F., Siegel M., Blaner B. 2015. The cache and memory subsystems of the IBM POWER8 processor. IBM Journal of Research and Development. 59(1):3:1-3:13. DOI: 10.1147/JRD.2014.2376131.
5. Foley D., Danskin J. Ultra-Performance Pascal GPU and NVLink Interconnect 2017. IEEE Micro. 37(2):7–17. DOI: 10.1109/MM.2017.37.
6. Sadasivam S.K., Thompto B.W., Kalla R., Starke W.J. 2017. IBM Power9 Processor Architecture. IEEE Micro. 37(2):40–51. DOI: 10.1109/MM.2017.40.
7. Starke W.J., Dodson J.S., Stuecheli J., Retter E., Michael B.W., Powell S.J., Marcella J.A. 2018. IBM POWER9 memory architectures for optimized systems. IBM Journal of Research and Development. 62(4/5):3:1–3:13. DOI: 10.1147/JRD.2018.2846159.
8. Choquette J., Giroux O., Foley D. 2018. Volta: Performance and Programmability. IEEE Micro. 38(2):42–52. DOI: 10.1109/MM.2018.022071134.

9. Mulnix D. Intel Xeon Processor Scalable Family Technical Overview. 2017. Available at: https://software.intel.com/ru-ru/articles/intel-xeon-processor-scalable-family-technical-overview (accessed April 8, 2020).
10. Mal'kovskii S. I., Sorokin A. A., Korolev S. P., Zatsarinnyi A. A., Tsoi G. I. 2019. Performance Evaluation of a Hybrid Computer Cluster Built on IBM POWER8 Microprocessors. Programming and Computer Software. 45(6):324-332. DOI: 10.1134/S0361768819060057.
11. Malkovsky S.I., Peresvetov V.V. 2009. Ocenka proizvoditel'nosti vichislitel'nogo klastera na chetyrehyadenyh processorah [Evaluating the performance of a computing cluster on quad-core processors]. Materialy mezhregional'noi nauchno-prakticheskoy konferencii "Informacionnye I kommunikacionnye tehnologii v obrazovanii i nauchnoy deyatel'nosti" [Scientific and Practical Conference (Interregional) "Information and Communication Technologies in Education and Scientific Activity" Proceedings]. Khabarovsk. 261–268.
12. McCalpin J.D. 1995. Memory Bandwidth and Machine Balance in Current High Performance Computers. IEEE Technical Committee on Computer Architecture Newsletter. 19-25.
13. Bailey, D.; Barszcz, E.; Barton, J.; Browning, D.; Carter, R.; Dagum, L.; Fatoohi, R.; Fineberg, S.; Frederickson, P.; Lasinski, T.; Schreiber, R.; Simon, H.; Venkatakrishnan, V.; Weeratunga, S. The NAS Parallel Benchmarks. RNR Technical Report RNR 94-007. Available at: https://www.davidhbailey.com/dhbpapers/npb.pdf (accessed May 7, 2020).
14. Steinbach P., Werner M. 2017. gearshifft – The FFT Benchmark Suite for Heterogeneous Platforms. In: Kunkel J., Yokota R., Balaji P., Keyes D. (eds) High Performance Computing. ISC 2017. Lecture Notes in Computer Science. Vol 10266. Springer, Cham. 199–216. DOI: 10.1007/978-3-319-58667-0_11.
15. DGEMM. 2015. Available at: https://web.archive.org/web/20180408033423/http://www.nersc.gov/research-and-development/apex/apex-benchmarks/dgemm/ (accessed April 8, 2018).
16. Nikitin O.U., Lukyanova O.A. 2019. Analiz uskoreniya glubokogo obucheniya na osnove vichislitelnoy sistemy IBM POWER8 [Analysis of Deep Learning Acceleration with IBM POWER8 Computing System]. Materyaly V mezdunarodnoy nauchno-prakticheskoy kovferencii "Informacionnye tehnologii i visokoproizvoditel'nye vichislenia" [5th Scientific and Practical Conference (International) "Information Technologies and High Performance Computing" Proceedings]. Khabarovsk. 199–203.
17. Kartsev A.I., Malkovsky S.I., Sorokin A.A., Volovich K.I. 2019. Issledovanie proizvoditel'nosti i masshtabiruemosti paketa Quantum ESPRESSO pri izuchenii nizkorazmernyh system na gibridnyh vichislitelnyh sistemah [Scaling and Productivity of Quantum ESPRESSO Package Based on the GPU-enabled Systems: the Case Study of Low-dimensional Systems Design]. Materialy I mezdunarodnoi konferencii "Matematicheskoe modelirovanie v materialovedenii elektronnyh komponentov" [1th Conference (International) "Mathematical Modeling in Materials Science of Electronic Components" Proceedings]. Moscow. 18–20.
18. Volkov K.N., Dobrov Yu.V., Karpenko A.G., Malkovsky S.I., Sorokin A.A. Simulation of Gas Dynamics of Hypersonic Aircrafts with the Use of Model of High-Temperature Air and Graphics Processor Units. Numerical Methods and Programming (Vychislitel'nye Metody i Programmirovanie). 22:29–46. DOI: 10.26089/NumMet.v22r103.
19. Sorokin A.A., Makogonov S.V., Korolev S.P. 2017. The Information Infrastructure for Collective Scientific Work in the Far East of Russia. Scientific and Technical Information Processing. 4:302–304. DOI: 10.3103/S0147688217040153.
20. Polozenie o CKP "Informatika" [Regulations on the Center for Collective Use "Informatics"]. Available at: http://www.frccsc.ru/ckp (accessed January 22, 2020).

**Sorokin A. A.** PhD, Computing Center of the Far Eastern Branch of the Russian Academy of Sciences, 65 Kim U Chen Str., Khabarovsk, 680000, Russian Federation, e-mail: alsor@febras.net

**Malkovsky S. I.** PhD, Computing Center of the Far Eastern Branch of the Russian Academy of Sciences, 65 Kim U Chen Str., Khabarovsk, 680000, Russian Federation, e-mail: sergey.malkovsky@ccfebras.ru