

**ПОДХОД К ОРГАНИЗАЦИИ ВЫЧИСЛИТЕЛЬНОГО
ПРОЦЕССА ГИБРИДНОЙ ВЫЧИСЛИТЕЛЬНОЙ СИСТЕМЫ
С ИСПОЛЬЗОВАНИЕМ INTEL XEON PHI**

Л. А. Васин

**APPROACH TO THE ORGANIZATION OF A COMPUTATIONAL
PROCESS OF A HYBRID COMPUTING SYSTEM USING INTEL
XEON PHI**

L. A. Vasin

Аннотация. *Предмет.* Обработка больших массивов информации, выполнение различных видов прогнозирования, проведение машинного обучения – все эти задачи требуют организации эффективного параллельного вычислительного процесса, ориентированного на высокопроизводительные вычисления. Для повышения производительности вычислительной системы может использоваться принцип гибридности, сочетая возможности центральных, графических и других сопроцессорных устройств. Для увеличения эффективности и точности вычислительного процесса могут использоваться нетрадиционные арифметические системы, например система остаточных классов. В работе показан вариант построения гибридной параллельной вычислительной структуры с дополнительными сопроцессорами и графическими процессорами. Целью работы являются разработка и исследование способов организации параллельных вычислительных структур и вычислительного процесса. *Методы.* Проводимое исследование основано на применении теории построения параллельных вычислительных систем, а также на принципах функционирования системы остаточных классов. *Результаты.* Включение в состав вычислительной системы потокового планировщика данных между графическим и сопроцессорным блоком позволяет осуществлять необходимую коммутацию данных в зависимости от выполняемой арифметической операции. Это позволяет использовать графические процессоры только для выполнения операций в базисе системы остаточных классов, что обусловлено наличием большого количества вычислительных ядер. Преобразование данных из позиционной системы счисления в непозиционную выполняется средствами математического сопроцессора, в роли которого используется Intel Xeon Phi. Использование гибридной схемы организации вычислительной системы позволяет более эффективно выполнять вычислительный процесс с использованием системы остаточных классов. *Выводы.* Предложенный подход заключается в комплексном использовании графических и сопроцессорных устройств, что позволяет эффективно организовать вычислительный процесс на основе модулярной арифметики за счет ее естественного параллелизма, а также за счет выполнения процедуры перевода чисел в СОК и обратно на сопроцессоре Intel Xeon Phi.

Ключевые слова: архитектура вычислительной системы, организация вычислительного процесса, графический процессор, система остаточных классов, модулярные операции, арифметический базис.

Abstract. *Background.* Processing large amounts of information, performing various types of forecasting, carrying out machine learning, all these tasks require the organization of an effective parallel computing process focused on high-performance computing. To improve the performance of a computing system, the principle of hybridity can be used, combining the capabilities of central, graphic, and other co-processor devices. To increase the

efficiency and accuracy of the computational process, unconventional arithmetic systems can be used, for example, a system of residue number system. The paper shows a variant of constructing a hybrid parallel computing structure with additional coprocessors and graphics processors. The aim of the work is to develop and study ways of organizing parallel computing structures and the computing process. *Methods.* The conducted research is based on the application of the theory of construction of parallel computing systems, as well as on the principles of functioning of the residue number system. *Results.* The inclusion of a streaming data scheduler between the graphic and coprocessor unit in the computer system allows the necessary data switching depending on the arithmetic operation being performed. This allows you to use graphics processors only to perform operations in the basis of the system of residual classes, due to the presence of a larger number of processing cores. Transformation of data from positional number system to non-positional one is performed by means of a mathematical coprocessor, in the role of which Intel Xeon Phi is used. The use of a hybrid scheme for the organization of a computing system allows one to more efficiently perform a computational process using a system of residual classes. *Conclusions.* The proposed approach consists in the integrated use of graphics and coprocessor devices, which makes it possible to efficiently organize a computational process based on modular arithmetic due to its natural parallelism, as well as by performing the procedure of converting numbers to residue number system and back on an Intel Xeon Phi coprocessor.

Key words: computing system architecture, computing process organization, graphics processor, residual class system, modular operations, arithmetic basis.

Введение

Для выполнения научных расчетов, проведения моделирования в различных областях науки, обработки медицинских данных необходима организация вычислительного процесса над данными большой размерности, что приводит к использованию значительных вычислительных ресурсов. Для повышения эффективности могут использоваться параллельные вычислительные структуры с различной параллельной архитектурой (многопроцессорные, кластерные). Создание технологий CUDA и OpenCL позволило использовать для организации вычислительного процесса графические процессоры, которые имеют большое количество вычислительных ядер. Стало возможным создание гибридных вычислительных систем, в которых, помимо центрального процесса, присутствуют как графические (NVIDIA, AMD), так и дополнительные сопроцессорные устройства (Intel Xeon Phi, NVIDIA Tesla). Это позволило значительно увеличить производительность массивно-параллельных вычислений. Увеличение количества вычислительных ядер, а также количества вычислительных узлов в кластерных системах не всегда позволяет пропорционально увеличить общую производительность [1].

Повышение производительности вычислительной системы можно достичь путем использования нетрадиционной вычислительной арифметики, например, на непозиционную, представленную системой остаточных классов (СОК), в которой используется модулярный формат данных [2]. Использование СОК в вычислительной системе позволяет достичь более высокой производительности за счет естественного параллелизма, отсутствия переноса, уменьшения количества используемых данных и более простого представления чисел. Особенностью организации вычислительного процесса в модулярной арифметике является необходимость перевода чисел из позиционной системы счисления в непозиционную [2].

Предлагается в гибридных вычислительных структурах, состоящих из CPU, GPU, разделить вычислительный процесс по всем процессорным элементам с выполнением расчетов в модульном формате на GPU, а перевод из позиционной системы и обратно выполнять на дополнительном сопроцессоре Intel Xeon Phi. В роли потокового планировщика используется центральный процессор вычислительной системы. Такая организация вычислительного процесса позволит оптимально использовать все имеющиеся вычислительные устройства.

Постановка задачи

Данная статья носит исследовательский характер. Проанализированы публикации, посвященные организации параллельных вычислительных систем, а также применение в них СОК [2–4]. Рассмотрены вопросы, касающиеся перевода чисел в СОК, а также ее использование на графических процессорных структурах и сопроцессоре Intel Xeon Phi, технология и особенности его программирования [3].

При организации массивно-параллельного вычислительного процесса в вычислительной системе, где присутствуют GPU, основная часть вычислений производится именно на них, а CPU вычислительной системы выполняет роль обработчика и арбитра входных данных для осуществления взаимодействия между входными и выходными данными с GPU. При использовании СОК необходимо осуществить перевод входных данных в модульный формат. Этот процесс можно выполнить на CPU, при этом оставляя ресурсы GPU свободными, для основной части вычислений в модульном формате. Современные процессоры имеют ограниченный набор вычислительных ядер, например Intel Xeon последнего поколения имеет их до двух десятков. Для повышения производительности традиционно увеличивают количество вычислительных ядер путем различных комбинаций CPU и GPU. Такая система с гетерогенной вычислительной структурой является гибридной. Существуют варианты с дополнительными сопроцессорными устройствами, например, Intel Xeon Phi 7120p, в зависимости от модели имеется до 61 вычислительного ядра, общую высокоскоростную кольцевую шину передачи, встроенную память объемом 16 Gb. Он реализован на основе архитектуры Intel MIC (Many Integrated Core), отличительная особенность которой является техническая реализация множества ядер, выполненных на одном кристалле. Преимущество его использования перед CPU заключается в наличии большого количества ядер, на которых можно реализовать выполнение хорошо распараллеливаемых программ и векторных вычислений, используемых преимущество высокоскоростной кольцевой шины.

Решение проблемы более эффективной организации вычислительного процесса с гибридной ВС с использованием модульной арифметики представленной СОК заключается в использовании сопроцессорного устройства в качестве потокового преобразователя чисел в модульный формат, при этом основной набор арифметических операций будет выполняться на GPU. Выполнение немодулярных операций по переводу из позиционной системы счисления в непозиционную производится на дополнительном сопроцессоре, с полным использованием вычислительных ядер, чем достигается уменьшение времени выполнения таких операций [5].

Вычислительные ядра GPU используются для выполнения модулярных операций, которые хорошо выполняются параллельно, например арифметическое сложение, вычитание без знака и др. Выполнение немодулярных операций таких, как сравнение, вычитание с получением отрицательного результата, контроль переполнения модулярного числа и т.д., основано на вычислении точных позиционных характеристик модулярных чисел и производится также на ядрах GPU, но с меньшей производительностью [6].

Особенности модулярной арифметики

Из существующих непозиционных систем счисления применение в вычислительных системах нашла система остаточных классов. СОК основывается на основе понятий вычета китайской теоремы об остатках. Она позволяет представить любое число:

$$x = x_1 * e_1 + \dots + x_n * e_n \pmod{M}, \quad (1)$$

где

$$e_i = M / P_i * ((M / P_i) - 1 \pmod{P_i}), (1 \leq i \leq n). \quad (2)$$

Исходное число может быть представлено в простой малоразрядной форме, получаемой путем деления с остатком (по модулю) на заранее выбранные модулярные основания. В итоге получатся остатки (вычеты), которые и являются простыми малоразрядными числами. Такие числа являются основанием данной арифметики. Основное преимущество СОК применительно к массивно-параллельным вычислениям заключается в параллельном вычислении остатков, и их сумма дает итоговое значение числа. Это позволяет оперировать длинными числами и использовать СОК в высокоточной арифметике.

Перевод чисел в модулярную структуру на сопроцессоре производится способом прямого преобразования, при котором двоичные числа в модулярный формат переводятся с помощью модульного суммирования остатков по модулю p_i ($i = 1, 2, \dots, n$) для B разрядов и n / B форматов с учетом их весов [7].

Особенности организации вычислительного процесса

Вычислительный процесс в гибридной ВС, в состав которой входят NVIDIA Kepler GPU и сопроцессор Intel Xeon Phi с использованием СОК, выполняется следующим способом:

1. Ввод чисел в позиционной системе счисления.
2. Преобразование чисел в модулярную форму на вычислительных ядрах Intel Xeon Phi табличным способом.
3. Выполнение арифметических операций с основаниями модулярных чисел.
4. Преобразование получившегося числа из модулярного формата в позиционную систему счисления, используя сопроцессор Intel Xeon Phi.
5. Вывод результата.

При такой организации вычислительные ядра графических процессоров полностью освобождаются от процесса перевода в разные системы счисления, при этом CPU выступает в роли потокового планировщика и организует

разделение этапов вычислительных процессов на модулярную составляющую и немодулярную и диспетчеризацию вычислительными ядрами GPU и сопроцессора. Такое использование гибридной вычислительной системы является перспективным направлением в планировании вычислительных операций в СОК.

Предлагаемый подход к организации вычислительного процесса в гибридной вычислительной структуре показан на рис. 1.

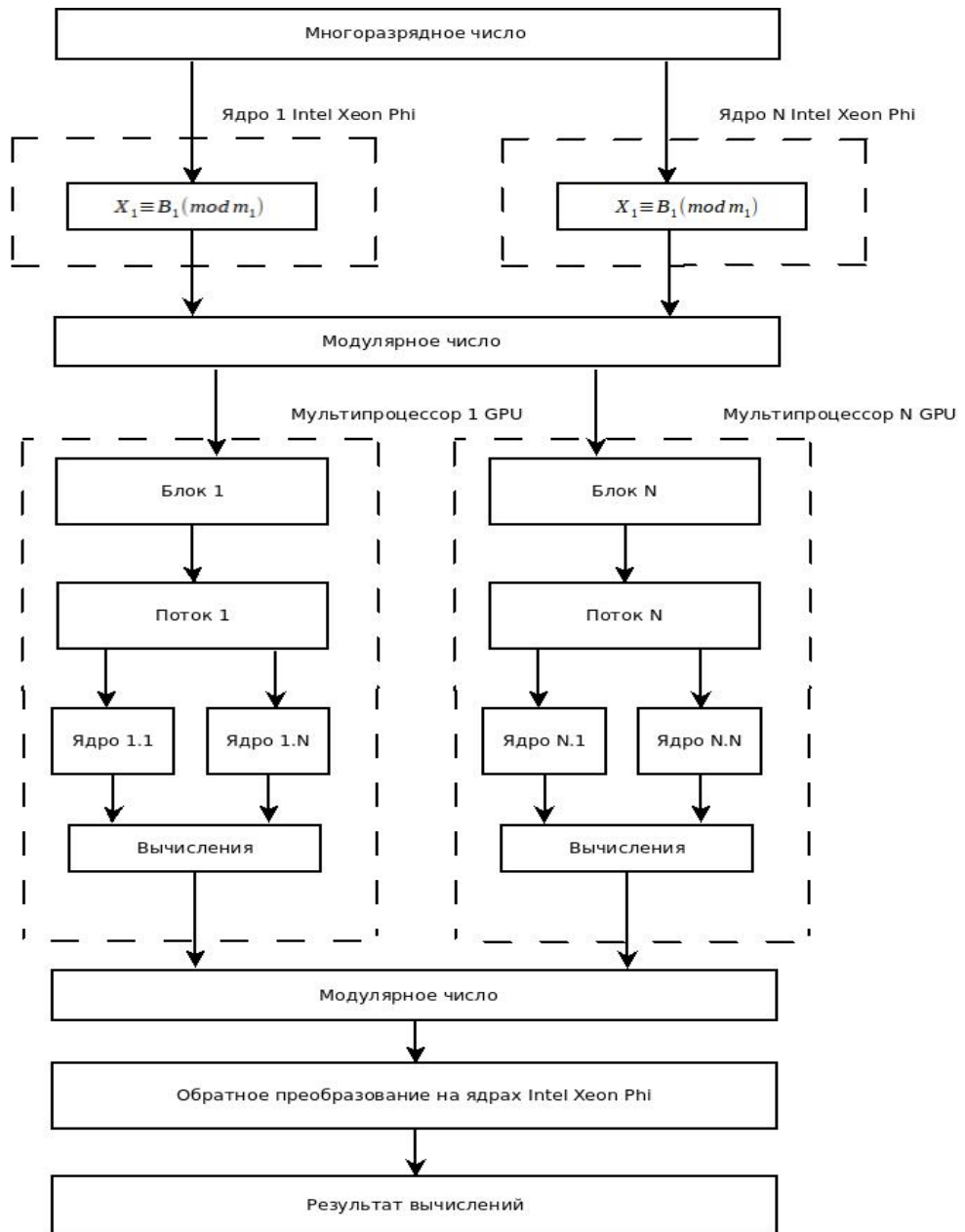


Рис. 1. Схема организации вычислительного процесса в гибридной вычислительной системе

Выполнение вычислительного процесса происходит следующим образом. Данные по кольцевой шине поступают на CPU сопроцессорного устройства Intel Xeon Phi, на ядрах которого происходит операция преобразования в модулярный формат на основе заранее выбранных модулей. При этом можно использовать длинные числа, обеспечивая загрузкой все вычислительные ядра. Общее количество потоков можно увеличивать путем добавления количества сопроцессорных ядер.

Модулярные данные, представленные в виде массива блоков, передаются на вычислительные ядра GPU. Управление данными организует потоковый планировщик CPU. При наличии нескольких GPU планировщик диспетчеризирует поток между ними. Особенностью архитектуры GPU является наличие у каждого вычислительного ядра небольшого количества оперативной памяти, что более подходит при работе с вычислениями, так как позволяет уменьшить временную сложность вычислений за счет работы с более простыми и малоразрядными числами. Достигается более высокое быстродействие вычислительного процесса, организованного в СОК при использовании GPU. При этом этапы преобразования форматов данных осуществляются на вычислительных ядрах традиционной архитектуры.

Заключение

В работе показан теоретический подход к организации вычислительного процесса, организованного по гибридной схеме с использованием графических и сопроцессорных узлов. В качестве используемой машинной арифметики используется непозиционная система счисления СОК. Показано, что:

1. Использование CPU в качестве потокового планировщика вычислительного процесса позволяет осуществлять диспетчеризацию между его частями, модулярной и немодулярной составляющими.
2. Использование модулярного формата данных позволяет эффективно организовать вычисление на основе GPU с использованием малоразрядных чисел.
3. Преобразование данных в модулярный формат производится на сопроцессорном устройстве Intel Xeon Phi.

Библиографический список

1. Баденко, В. Л. Высокопроизводительные вычисления : учеб. пособие / В. Л. Баденко. – СПб. : Изд-во Политехн. ун-та, 2010. – 180 с.
2. Акушский, И. Д. Машинная арифметика в остаточных классах / И. Д. Акушский, Д. И. Юдицкий. – М. : Сов. Радио, 1968.
3. Программирование на современных мультиядерных архитектурах (на примере Intel Xeon Phi). – URL: <http://www.intuit.ru/studies/courses/10612/1096/info> (дата обращения: 20.10.2018).
4. Червяков, Н. И. Модулярные параллельные вычислительные структуры нейропроцессорных систем / Н. И. Червяков, С. А. Ряднов, П. А. Сахнюк, А. В. Шапошников. – М. : Физматлит, 2003. – 288 с.
5. Модулярная арифметика и ее приложения в инфокоммуникационных технологиях / Н. И. Червяков, И. Н. Лавриненко, А. В. Лавриненко, А. А. Коляда, П. А. Ляхов, М. Г. Бабенко. – М. : Физматлит, 2017. – 402 с.
6. Исупов, К. С. Методика выполнения базовых немодульных операций в модулярной арифметике с применением интервальных позиционных характеристик /

К. С. Исупов // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2013. – № 3 (27). – С. 26–39.

7. Червяков, Н. И. Методы, алгоритмы и техническая реализация основных проблемных операций, выполняемых в системе остаточных классов / Н. И. Червяков // Инфокоммуникационные технологии. – 2011. – № 4. – С. 4–12.

References

1. Badenko V. L. *Vysokoproizvoditel'nye vychisleniya: ucheb. posobie* [High performance computing: textbook]. Saint-Petersburg: Izd-vo Politekhn. un-ta, 2010, 180 p.
2. Akushskiy I. D., Yuditskiy D. I. *Mashinnaya arifmetika v ostatochnykh klassakh* [Machine arithmetic in residual classes]. Moscow: Sov. Radio, 1968.
3. *Programmirovaniye na sovremennykh mul'tiyadernnykh arkhitekturakh (na primere Intel Xeon Phi)* [Programming on modern multi-core architectures (by the example of Intel Xeon Phi)]. Available at: <http://www.intuit.ru/studies/courses/10612/1096/info> (accessed Oct. 20, 2018).
4. Chervyakov N. I., Ryadnov S. A., Sakhnyuk P. A., Shaposhnikov A. V. *Modulyarnye parallel'nye vychislitel'nye struktury neyro-protsessornykh sistem* [Modular parallel computing structures of connected systems]. Moscow: Fizmatlit, 2003, 288 p.
5. Chervyakov N. I., Lavrinenko I. N., Lavrinenko A. V., Kolyada A. A., Lyakhov P. A., Babenko M. G. *Modulyarnaya arifmetika i ee prilozheniya v infokommunikatsionnykh tekhnologiyakh* [Modular arithmetic and its applications in information and communication technologies]. Moscow: Fizmatlit, 2017, 402 p.
6. Isupov K. S. *Izvestiya vysshikh uchebnykh zavedeniy. Povolzhskiy region. Tekhnicheskie nauki* [News of higher educational institutions. Volga region. Technical science]. 2013, no. 3 (27), pp. 26–39.
7. Chervyakov N. I. *Infokommunikatsionnye tekhnologii* [Information and communication technologies]. 2011, no. 4, pp. 4–12.

Васин Леонид Анатольевич

кандидат технических наук,
заведующий кафедрой информационно-
вычислительных систем,
Пензенский государственный
университет архитектуры
и строительства
(Россия, г. Пенза, ул. Германа Титова, 28)
E-mail: leo.vasin@gmail.com

Vasin Leonid Anatolievich

candidate of technical sciences,
head of sub-department
of information-computing systems,
Penza State University
of Architecture and Construction
(28 Herman Titov street,
Penza, Russia)

УДК 004.4

Васин, Л. А.

Подход к организации вычислительного процесса гибридной вычислительной системы с использованием Intel Xeon Phi / Л. А. Васин // Модели, системы, сети в экономике, технике, природе и обществе. – 2018. – № 3 (27). – С. 120–126.