

ФОРМИРОВАНИЕ ИНФОРМАЦИОННЫХ РЕСУРСОВ НА ОСНОВЕ ПАРАЛЛЕЛЬНЫХ ВЫЧИСЛЕНИЙ

Раскрываются формирования информационных ресурсов на основе параллельных вычислений. Раскрыта проблема семантического разрыва. Отмечен антропологический подход оценки производительности вычислительных систем. Показана целесообразность применения тензорных методов и сетей Петри для формирования информационных ресурсов.

Ключевые слова: информационные ресурсы, решение сложных задач, большие объемы вычислений, параллельная обработка информации, сети Петри.

V. P. Kulagin

THE CREATION OF INFORMATION RESOURCES ON THE BASIS OF PARALLEL COMPUTATIONS

The article describes the methods of formation of information resources on the basis of parallel computing. This article describes the semantic gap. The article describes the anthropological approach evaluate the performance of computing systems. The article shows the feasibility of tensor methods and Petri nets for the formation of information resources.

Key words: information resources, solve complex problems, large amounts of computation, parallel processing of information, Petri nets

В практике ряда исследований, например, при обработке фотограмметрической космической информации или при сложных преобразованиях [1], возникает необходимость обработки больших объемов информации. Например, решение задач, возникающих в таких областях, как обработка космических изображений [2], машинное зрение, ядерная физика, структурный анализ, обработка речевых, радиолокационных, сейсмических, метеорологических, медицинских и других данных, требует от современных вычислительных систем производительности от 25 млрд. до 1000 трлн. операций/секунду [3]. Другими словами, возникает ситуация информационного барьера [4]. Для преодоления этого информационного барьера применяют параллельные вычисления. Особенность параллельных вычислений в том, что они требуют разработки специальных алгоритмов. Обычные алгоритмы ориентированы на последовательную обработку. При этом возникает необходимость максимально эффективно использовать аппаратные ресурсы

современной вычислительной техники. Задача разработки параллельных алгоритмов для повышения эффективности вычислений может быть сравнима по сложности с решением самой прикладной задачи.

Особенности параллельных вычислительных систем. Информационные ресурсы включают не только данные и программы, но и результаты компьютерной обработки данных, а также результаты решения задач, которые впоследствии используются для анализа и получения новых знаний. Решения ряда современных сложных и объемных задач может быть получены только с помощью параллельных вычислительных систем.

Параллельной вычислительной системой (ПВС) [2, 3] называют вычислительную систему, закон функционирования которой допускает ее декомпозицию на отдельные составляющие, при сохранении свойства эмерджентности. Эти составляющие осуществляют независимую (параллельную) обработку информации, но синхронно в рамках единой функциональной системы.

Отличие структуры ПВС от структуры обычной вычислительной системы заключается в дополнительном описании типов связей и распределении функций обработки.

Главный вопрос при параллельных вычислениях состоит в распараллеливании последовательных алгоритмов таким образом, чтобы они имели возможность эффективной реализации на ПВС. Проблемы анализа и синтеза структур ПВС взаимосвязаны и образуют в совокупности проблему, которая в полном объеме не решена и в настоящее время.

Существует ряд недостатков последовательных вычислительных систем, которые определяют как семантический разрыв [5, 6]. Семантический разрыв обуславливает возникновение ряда проблем, к которым относятся высокая стоимость разработки программного обеспечения, его ненадежность и эксплуатационная неэффективность, чрезмерный объем программ, сложность компиляторов и операционных систем, и др. [5].

Семантический разрыв оказывает косвенное влияние и на понятие производительность ЭВМ, которое разные эксперты трактуют по-разному. Причина разной интерпретации производительности в том, что ее можно определять как на макро-, так и на микроуровне. Другими словами, при решении задач существует цепочка действий, которая включает человеческий фактор, как компонент общего решения.

«постановка задачи → задача → вычисление → решение»

Общая цепочка представляет собой макроуровень, компоненты цепочки «вычисление → решение» определяют микроуровень.

Многие разработчики, используя традиционный подход к проектированию "снизу-вверх" [7] и стремясь минимизировать стоимость аппаратных средств, концентрируют свое внимание только на микроуровне (вычисление) и возлагают на плечи аналитиков и программистов решение всех остальных трудных проблем.

В качестве примера подобного несоответствия возьмем данные из работы [8]. Какая ВС имеет более высокую производительность: система, выполняющая программу X за 1 минуту, но требующая 10 человеко-недель для программирования, или система Y, выполняющая ту же программу X за 2 минуты при затратах на программирование 2 человеко-недели?

С позиций микроуровня быстродействие первой системы по сравнению со второй в два раза выше, с позиций макроуровня - в пять раз ниже. Если же исходить из известных статистических данных о том, что 85%

всех программ выполняется только один раз, то такая характеристика архитектуры ЭВМ, как простота программирования, является не менее важной, чем скорость выполнения команд.

Изложенное выше не означает, что скорость выполнения программы является несущественным фактором. Важно понимать, что быстродействие ВС зависит как от структурных особенностей, так и от вида представления решаемой задачи.

В работах [7-10] и многих других показано, что традиционная технология проектирования универсальных ЭВМ снизу-вверх, т.е. от архитектуры к численным методам и программам решения задач, неэффективна при разработке параллельных ВС. В работах отмечено, что для эффективного применения параллельных ВС необходима единая форма описания алгоритмов и структур ВС. Как решение подобной проблемы в настоящее время практикуется метод «встречных потоков» [11], который синтезирует методы «сверху вниз» и «снизу-вверх». В настоящее время применяют специальные современные подходы к построению параллельных ВС.

Локальность связей. Традиционный способ построения параллельных ВС и языков программирования не подходит для системы обработки данных в реальном масштабе времени из-за большого числа вспомогательных операций, связанных с распределением вычислительных ресурсов и препятствующих повышению производительности [3]. Поэтому в многопроцессорных системах стало целесообразно использовать локальные связи, что позволяет уменьшить взаимную зависимость процессоров и задержки, вносимые длинными соединениями. Такой подход привел к разработке нескольких специализированных устройств обработки данных с локальными связями и с архитектурой, ориентированной на исполнение в виде СБИС: трансверсальная фильтрация, систолические матрицы, потоковые процессорные матрицы [7, 8, 10].

Программируемая архитектура и динамическая реконфигурация вычислительных систем. Из многочисленных архитектур построения суперЭВМ необходимо отметить как одно из направлений – создание многопроцессорных систем (МПС). Начало работам в данном направлении положено отечественными разработками ВС с перестраиваемой архитектурой [3]. Поиск путей совершенствования процесса проектирования ЭВМ привели к созданию фирмой MEIKO Ltd. (Великобритания) новой технологии разработки суперЭВМ, опирающейся

на "бесшовную" архитектуру, позволяющую плавно наращивать вычислительную мощность системы для точного соответствия требованиям заказчика ("обычные" суперЭВМ допускают расширение только модулями, причем довольно большими и дорогими).

МПС с перестраиваемой архитектурой можно разделить на два основных класса: МПС с программируемой архитектурой и МПС с динамической реконфигурацией. В МПС с программируемой архитектурой [12] нужная конфигурация определяется программным путем, когда пользователь, например, может создавать необходимую группу процессоров с требуемой длиной слова. Передача и распределение информации между множествами процессоров и комплексов памяти происходит через программируемый параллельный коммутатор, называемый иногда коммутационной структурой (КС) [13].

Реализация программируемых коммутаторов сопряжена с определенными трудностями, основная из которых заключается в том, что время передачи данных от одного микропроцессора (МП) к другому через КС существенно больше времени обработки информации в самом МП [13]. Предлагаемые приемы для уменьшения времени передачи данных связаны либо с изменением, либо с поиском более эффективного алгоритма, что приводит к достижению поставленной цели, но либо за счет повышения трудоемкости программирования, либо за счет увеличения времени построения алгоритма, что не является удовлетворительным.

Пространственно-временное распределение данных. В каждом алгоритме, реализованном на параллельной ВС, важно точно определить момент времени, когда данные достигают процессорного элемента (ПЭ). В программе, написанной на обычном языке высокого уровня, нет взаимодействия пространства и времени, так как в ней рассматривается только одно действие в единицу времени и отсутствует информация, где физически расположены данные. Иначе обстоит дело при разработке систем на процессорных и волновых матрицах. Данные, которые расположены далеко от ПЭ, требуют больших энергетических затрат и имеют большее время доступа. Поэтому алгоритм, считающийся хорошим для традиционного языка программирования, не обязательно хорош, когда в расчет принимаются явления, происходящие на физическом уровне. Таким образом, при проектировании параллельных систем на матричных структурах необходимо учитывать пространственное и временное распределение данных [3, 7, 14].

Декомпозиция вычислительных структур и данных. При обработке данных на параллельных ВС часто возникают вопросы отображения вычислительных задач большой размерности на ограниченные структуры ПЭ. Поэтому необходимо проектировать только такие специализированные устройства, которые могут быть использованы для решения задачи произвольного размера и делать это не хуже, чем программные средства универсальных компьютеров. Для решения данных вопросов часто используются аддитивное разбиение и тензорное исчисление [3, 15].

Параллельное представление алгоритмов. Многие важные алгоритмы обработки данных имеют очень специализированные структуры или совсем не структурированы. Вопросы реализации таких алгоритмов на параллельных и конвейерных устройствах достаточно сложны. Поэтому для достижения желаемой скорости обработки необходимы средства, позволяющие максимально использовать возможности параллелизма и конвейерной организации. Иначе говоря, задача нахождения отношения параллельности реализуемых алгоритмов на множестве ПЭ весьма актуальна.

Аппаратные средства параллельной обработки. Из числа новых разработок элементной базы для высокопроизводительных ЭВМ следует отметить создание СБИС центральных процессоров с высоким быстродействием, создание высокопроизводительных МПС с ограниченным набором команд (RISC-процессоров), реализация параллельных процессоров в виде одной СБИС и появление транспьютеров – нового мощного направления развития элементной базы.

Термин "транспьютер" в настоящее время используется в двух значениях: 1) как наименование программируемых СБИС, разработанных фирмой INMOS и предназначенных в качестве вычислительных модулей для создания сверхпроизводительных параллельных систем; 2) как наименование любых подобных транспьютерам вычислительных модулей. Основное назначение транспьютеров заключается в построении параллельных вычислительных систем высокой производительности. Транспьютерную систему можно использовать для организации конвейерных вычислений, систолической обработки, векторной и матричной обработки, обработки методом потока данных.

Обобщая результаты анализа основных тенденций в развитии современных ВС, следует отметить, что наиболее перспективным способом достижения высокой производи-

тельности является построение ВС на основе большого числа однородных ЭВМ. Данные системы привлекательны из-за локальности линий связи, что обеспечивает большую пропускную способность. Стремление эффективно решать задачи на таких системах должно обязательно сопровождаться согласованием структуры численных методов и архитектуры ВС.

Формализованное описание структуры. Целью формализованного описания структур ВС является представление имеющихся данных и параллельных процессов в виде специальных формальных объектов, удобных для проведения над ними вычислительных и имитационных экспериментов на ЭВМ. Поэтому выбор формализованного языка, в наибольшей степени учитывающего особенности параллельных ВС, является основной задачей начального этапа проектирования.

Решение таких задач связано с применением специальных методов построения синхронных и асинхронных моделей дискретных систем. Среди этих методов наибольшую известность получили методы алгоритмизации систем массового обслуживания, автоматного и агрегативного моделирования, расширения известных языков программирования [16], структурные нотации [17], сетевой и алгебраический подходы [18, 19], графовые модели [19, 20].

Подход, предложенный в работе [9] для формализации анализа сложной системы и состоящий из трех шагов: структуризации объекта, формализации элементов сложной системы и взаимодействия между этими объектами, обладает существенным недостатком. Данный недостаток заключается в том, что такая формализация элементов системы и взаимодействия между ними обладает "неформульным" заданием схемы сопряжения (в виде рисунков) и операторов сопряжения (в виде таблиц).

Подобная формализация не представляет возможность описать область эквивалентных структурных преобразований схемы сопряжения. Этих недостатков лишена формализация элементов сложной системы и взаимодействия между ними с помощью так называемых R-модулей [3], которые используются не только при описании детерминированных динамических систем, функционирующих в дискретном времени, но и при описании стохастических систем, представляемых вероятностными автоматами. Однако, подобные автоматные модели не перекрывают все возникающие задачи.

Использование параллельных языков программирования и структурных нотаций

для формализованного описания структур ВС эффективно лишь для анализа одной структуры ВС, так как данные описания громоздки и не приспособлены для поддержки процедур синтеза новых структур ВС. Если ВС рассматривать только как множество взаимодействующих функциональных блоков (объектов), а не как вычислительную сеть или многопроцессорную систему, то для исследования процессов, протекающих в ВС, может быть использован сетевой подход [18, 19]. В работе [10] для описания параллельных процессов с синхронизацией описаны предложены OS-сети. Потребляемые ресурсы в этой модели явно не задаются, повторно используемые ресурсы используются самые простые. Предложенные OS-сети используются, в основном, для анализа тупиковых ситуаций, которые могут возникнуть в параллельных ВС.

Эффективным средством анализа и синтеза параллельных ВС и процессов является алгебраический подход [18], который основан на формульном выражении сетевых моделей. Использование алгебраического подхода позволяет аналитическими методами путем проведения эквивалентных преобразований формул получать оптимальные структуры ВС [9]. Недостатками данного подхода являются: во-первых, ограниченность, т.к. не все сетевые структуры могут быть описаны алгебраически; во-вторых, сложность, т.к. процесс построения алгебраического описания сложной сетевой структуры требует определенных навыков и умения.

В последнее время во многих работах отмечается тот факт, что графовые модели являются наиболее удобными и эффективными средствами описания и исследования параллельных структур и процессов. К настоящему времени существует несколько формализмов, основанных на графовых моделях и служащих для описания параллельных процессов.

К концу 70-х годов указанные модели были практически вытеснены сетями Петри (СП) [19, 20] – формализмом, описывающим структуру и взаимодействие параллельных процессов.

Выводы.

1. Применение параллельных вычислительных систем является обязательным условием решения сложных задач и формирования информационных ресурсов на этой основе.

2. Эффективное решение задач на ПВС должно сопровождаться согласованием структуры численных методов и архитектуры ВС.

3. При решении проблемы синтеза эффективных вычислительных структур в качестве математического фундамента целесообразно использовать тензорное исчисление.

4. Анализ методов формализованного описания современных структур ПВС и их особенностей, параллельных процессов

и алгоритмов, дает основание рекомендовать аппарат сетей Петри, который дает возможность описывать разнообразные объекты и предоставляет развитые методы анализа параллельных процессов и, обладая однородностью и аналитическими зависимостями, удовлетворяет условиям для использования в тензорной методологии.

ЛИТЕРАТУРА

1. Vladimir P. Kulagin, Victor Y. Tsvetkov, Maiya V. Maksimova The Convergence in Spatial Tasks // European Researcher, 2013, Vol.(58), no.9-1, p.2179-2184.
2. Цветков В.Я. Перспективы применения методов параллельного программирования в топографо-геодезическом производстве. /Обзорная информация. - М.: ЦНИИГАиК, 1986. – 49 с.
3. Суперкомпьютерные технологии в науке, образовании и промышленности / Под редакцией: академика В. А. Садовниченко, академика Г. И. Савина, чл.-корр. РАН Вл. В. Воеводина.-М.: Издательство Московского университета, 2009.-232 с.
4. Цветков В.Я Маркелов В.М., Романов И.А. Преодоление информационных барьеров // Дистанционное и виртуальное обучение. 2012. № 11. С. 4-7.
5. Dorai C, Venkatesh S. Bridging the Semantic Gap with Computational Media Aesthetics.
6. V. Y. Tsvetkov. Information Interaction as a Mechanism of Semantic Gap Elimination // European Researcher, 2013, Vol.(45), № 4-1, p.782-786.
7. Цапко И. В., Цапко С. Г. Архитектура вычислительных систем. – Томск, Из-во Томского политехнического университета, 2011. –168 с.
8. Майерс Г. Архитектура современных ЭВМ: В 2-кн. Кн.1. - М.: Мир, 1985. - 364 с.
9. Букалов А. А., Дацюк В. Н., Жегуло А. И. Программирование многопроцессорных вычислительных систем. Ростов-на-Дону. Издательство ООО «ЦВВР», 2003, 208 с.
10. Воеводин В.В. Параллельные структуры алгоритмов и программ. – М.: ОВМ АН СССР, 1987. – 148 с.
11. Цветков В.Я., Вознесенская М.Е. Метод встречных потоков при проектировании программных продуктов // Успехи современного естествознания. – №3. – 2010. – с.138-139.
12. Митропольский Ю.И. Концепции иерархического построения высокопроизводительных вычислительных систем. // Кибернетика и вычислительная техника (Москва). – 1991. – №5. – С.7-14.
13. Мальцева Н.С., Коммутационная структура с параллельной идентификацией для многопроцессорных вычислительных систем / дис. К.т.н., специальность 05.13.05- Астрахань, 2008, 140 с.
14. Кулагин В.П. Исследование моделей вычислительных систем методом преобразования координат. // Вычислительная техника в автоматизированных системах контроля и управления: Межвуз. сб. научн. тр. - Вып.20. / Пенза: Пенз. политехн. ин-т. - 1990. – С.4-7.
15. Кулагин В.П. Тензорные методы проектирования структур вычислительных систем.//АВТ. – 1989. – №.2. – С.64-71.
16. Армстронг Дж.Р. Моделирование цифровых схем на языке VHDL: концепция моделирования на уровне интегральных схем. – М.: Мир, 1992. – 174 с.
17. Гергель В.П., Стронгин Р.Г. Основы параллельных вычислений для многопроцессорных вычислительных систем. – Н.Новгород, ННГУ, 2001.
18. Кулагин В.П. Алгебра сетевых моделей для описания параллельных вычислительных систем.// Автоматизация и современные технологии. – 1993. – №.2. – С.25-30.
19. Кулагин В.П. Моделирование структур параллельных ВС на основе сетевых моделей. – М.: МИЭМ, 1998 – 102 с.
20. Питерсон Дж. Теория сетей Петри и моделирование систем. – М.: Мир, 1984. – 264.

REFERENCES

1. Vladimir P. Kulagin, Victor Y. Tsvetkov, Maiya V. Maksimova The Convergence in Spatial Tasks. European Researcher, 2013, Vol.(58), no.9-1, p.2179-2184.
2. Tsvetkov V.Ia. *Perspektivy primeneniia metodov parallel'nogo programmirovaniia v topografo-geodezicheskom proizvodstve / Obzornaia informatsiia* [Prospects for the use of parallel programming techniques in the production of topographic and geodetic / Review]. Moscow, TsNIIGAiK, 1986. 49 p.
3. *Superkomp'uternye tekhnologii v nauke, obrazovanii i promyshlennosti / Pod redaktsiei: akademika V. A. Sadovnichego, akademika G.I. Savina, chl.-korr. RAN V.I. Voevodina* [Supercomputer technologies in science, education and industry / Edited by Academician V.A.Sadovnichii, academician G.I.Savin, Corr. RAS V.B.Voevodin]. Moscow, Izdatel'stvo Moskovskogo universiteta, 2009.-232 s.
4. Tsvetkov V.Ia Markelov V.M., Romanov I.A. Overcoming information barriers. *Distantionnoe i virtual'noe obuchenie - Distance and virtual learning*, 2012, no.11. pp. 4-7 (in Russian).
5. Dorai C, Venkatesh S. Bridging the Semantic Gap with Computational Media Aesthetics.

6. V. Y. Tsvetkov. Information Interaction as a Mechanism of Semantic Gap Elimination. European Researcher, 2013, Vol.(45), no,4-1, p.782-786.
7. Tsapko I. V., Tsapko S. G. *Arkhitektura vychislitel'nykh sistem* [Architecture of computing systems]. Tomsk, Iz-vo TPU, 2011. 168 p.
8. Maier G. *Arkhitektura sovremennykh EVM* [Architecture of modern computers]. Moscow, Mir, 1985. 364 p.
9. Bukatov A.A., Datsiuk V.N., Zhegulo A.I. *Programmirovaniye mnogoprotsessornykh vychislitel'nykh sistem* [Programming multiprocessor systems]. Rostov-na-Donu. TsVVR, 2003, 208 p.
10. Voevodin V.V. *Parallelnyye struktury algoritmov i programm* [Parallel structures algorithms and programs]. Moscow, OVM AN SSSR, 1987. 148 p.
11. Tsvetkov V.Ia., Voznesenskaia M.E. Method of counter streams when designing software products. *Uspekhi sovremennogo estestvoznaniia - Successes of modern science*, 2010, no.3. pp.138-139 (in Russian).
12. Mitropol'skii Iu.I. The concept of hierarchical construction of high-performance computing systems. *Kibernetika i vychislitel'naya tekhnika - Cybernetics and Computer Science*, 1991, no.5. pp.7-14 (in Russian).
13. Mal'tseva N.S. *Kommutatsionnaya struktura s parallel'noi identifikatsiei dlia mnogoprotsessornykh vychislitel'nykh sistem. diss. K.t.n., 05.13.05* [Switching structure with parallel identification for multiprocessor systems. Diss. Ph.D., 05.13.05]. Astrakhan', 2008, 140 p.
14. Kulagin V.P. *Issledovanie modelei vychislitel'nykh sistem metodom preobrazovaniia koordinat* [The study models of computing systems by the coordinate transformation]. *Vychislitel'naya tekhnika v avtomatizirovannykh sistemakh kontrolya i upravleniia: Mezhd. sb. nauchn. tr. - Vyp.20.* [Computers in automation and control systems: Intercollege. Vyp.20]. Penza, PPI, 1990. pp.4-7.
15. Kulagin V.P. *Tenzornyye metody proektirovaniia struktur vychislitel'nykh sistem* [Tensor methods of designing structures computing]. 1989. no.2. pp.64-71.
16. Armstrong Dzh.R. *Modelirovaniye tsifrovyykh skhem na iazyke VHDL: kontseptsii modelirovaniia na urovne integral'nykh skhem* [Simulation of digital circuits in the language VHDL: the concept of modeling on the integrated circuit]. Moscow, Mir, 1992. 174 p.
17. Gergel' V.P., Strongin R.G. *Osnovy parallel'nykh vychislenii dlia mnogoprotsessornykh vychislitel'nykh sistem* [Introduction to Parallel Programming for multiprocessor systems]. N.Novgorod, NNGU, 2001.
18. Kulagin V.P. Algebra network models for parallel computing. *Avtomatizatsiia i sovremennyye tekhnologii - Automation and modern technology*, 1993, no.2. pp.25-30 (in Russian).
19. Kulagin V.P. *Modelirovaniye struktur parallel'nykh VS na osnove setevykh modelei* [Modeling of structures parallel to the aircraft based on network models]. Moscow, MIEM, 1998. 102 p.
20. Piterson Dzh. *Teoriia setei Petri i modelirovaniye sistem* [The Theory of Petri nets and simulation systems]. Moscow, Mir, 1984. 264 p.

Информация об авторе

Кулагин Владимир Петрович (Россия, Москва) – Профессор, доктор технических наук. Заместитель директора. Московский институт электроники и математики НИУ ВШЭ. E-mail: kvp@miem.ru

Information about the author

Kulagin Vladimir Petrovich (Russia, Moscow) – Professor, doctor of technical sciences. Deputy Director. Moscow Institute of Electronics and Mathematics HSE NRU (Higher School of Economics - National Research University). E-mail: kvp@miem.ru