

СУПЕРКОМПЬЮТЕРНОЕ ОБРАЗОВАНИЕ: ПОНИМАНИЕ ПРЕДМЕТНОЙ ОБЛАСТИ

В.П. Гергель

Решительный переход компьютерной индустрии на многоядерные «рельсы», когда в рамках единой суперкомпьютерной системы могут использоваться сотни тысяч (а случае эксафлопности и миллиарды) вычислительных ядер, требует столь же радикальных перемен и в области программного обеспечения. Задействовать колоссальный вычислительный потенциал многопроцессорных многоядерных систем можно только при кардинальном "параллельном" обобщении традиционной - последовательной - технологии решения задач на ЭВМ. Так, численные методы в случае многопроцессорности и многоядерности должны проектироваться как системы параллельных и взаимодействующих между собой процессов, допускающих исполнение на независимых вычислительных процессорах (ядрах). Применяемые алгоритмические языки и системное программное обеспечение должны обеспечивать создание параллельных программ, организовывать синхронизацию и взаимоисключение асинхронных процессов и т.п.

Преодоление проблемы эффективного использования суперкомпьютерных систем и организации высокопроизводительных вычислений требует наличия новых знаний, владения новыми умениями и навыками. Причем этими новыми знаниями и умениями должны обладать не только «узкие» специалисты-разработчики параллельных программ – знания и умения в области суперкомпьютерных технологий (СКТ)(Будем использовать далее для краткости данную аббревиатуру СКТ как наиболее общую для обозначения предметной области, охватывающей суперкомпьютерные технологии, высокопроизводительные вычисления и параллельное программирование) и высокопроизводительных вычислений становятся важной квалификационной характеристикой любого современного специалиста. И, тем самым, чрезвычайно важной задачей на пути подготовки специалистов, успешно действующих в рамках многоядерного компьютерного мира, становится определение совокупного набора (Свода) знаний и умений, которые должны быть отнесены к данной новой предметной области.

Разработка Свода является сложной научно-образовательной проблемой. Свод должен быть полным, т.е. включать в себя все основные понятия предметной области. Вместе с этим, Свод должен быть достаточно компактным и не содержать излишние технические детали. И, конечно же, Свод должен признаваться научно-техническим сообществом определяемой предметной области.

Тем самым, Свод не может быть разработан отдельным исследователем. Свод не может быть также и разработкой небольшого творческого коллектива. Свод, как правило, должен являться результатом активной деятельности широкого круга профессиональных экспертов, и разработка Свода может занимать достаточно большие периоды времени.

Еще раз отметим важность понятие Свода. Именно Свод должен являться основой для определения профессиональных квалификационных требований. Только на основе Свода могут быть сформированы адекватные образовательные программы подготовки, переподготовки и повышения квалификации кадров в соответствующей предметной области.

В настоящее время в стране и за рубежом могут быть выделены следующие основные активности по определению предметной области суперкомпьютерных технологий и высокопроизводительных вычислений:

- Свод знаний и умений (далее Свод 1), разработанный в рамках Национальной программы суперкомпьютерного образования, инициированной Суперкомпьютерным консорциумом университетов России и поддержанной президентом и правительством страны [1].
- Рекомендации по составу предметной области параллельных и распределенных вычислений (далее Свод 2), подготовленные в рамках работ международных сообществ ACM и IEEE-CS по разработке учебных планов в области Компьютинга (Computing) [2-4].
- Рекомендации по составу предметной области параллельных и распределенных вычислений (далее Свод 3), разработанные в рамках проекта, выполняемого при поддержке научного фонда NSF (США) и комитета IEEE-TCPP (Technical Committee on Parallel Programming) [5].

Во всех перечисленных работах используется один и тот же - конструктивный - подход определения предметной области. Суть данного подхода состоит в следующем(В работах используется несколько различающаяся терминология. Для однородности изложения материала в рамках статьи далее будут использоваться приводимые ниже термины.):

1. Свод знаний и умений, который должен быть освоен специалистом для успешной профессиональной деятельности, определяется набором отдельных областей знаний, представляющих собой отдельные части определяемой сферы деятельности.
2. Далее, области знаний делятся на меньшие структуры, называемые разделами, которые представляют собой отдельные тематические модули внутри области.

3. Каждый раздел, в свою очередь, состоит из набора тем, представляющих собой нижний уровень этой иерархии в определяемой сфере деятельности.

Каждый раздел сопровождается указанием, является ли она обязательной или факультативной, а также рекомендуемым объемом учебного времени, необходимым для ее изучения.

Важно подчеркнуть, что подобная структура областей, разделов и тем определяет именно свод знаний, необходимый для освоения соответствующей сферы деятельности, а не представляет собой перечень учебных курсов. Учебные планы и необходимые учебные курсы разрабатываются далее уже на основе Сводов.

В данной статье осуществляется сопоставительная характеристика всех перечисленных выше работ по определению предметной области СКТ.

1. Свод знаний и умений в области СКТ

Разработка данного Свода проводилась в 2010-2012 гг. в рамках проекта «Создание системы подготовки высококвалифицированных кадров в области суперкомпьютерных технологий и специализированного программного обеспечения» Комиссии при Президенте РФ по модернизации и технологическому развитию экономики России (<http://hpc-education.ru>) [1]. Выполнение данного проекта осуществлялось Московским госуниверситетом при привлечении ведущих вузов страны (ННГУ, ТГУ, ЮУрГУ, СПбГУ ИТМО, МФТУ, ДФУ, ЮФУ) и участия многих других членов Суперкомпьютерного консорциума университетов России.

Разработанный Свод явился основой для выполненных в рамках проекта работ по анализу и расширению государственных образовательных стандартов, по разработке учебных планов и учебных курсов и по реализации программ подготовки, переподготовке и повышения квалификации в области суперкомпьютерных технологий и высокопроизводительных вычислений.

При разработке Свода широко использовался опыт исполнителей проекта по подготовке кадров в области СКТ [6-11]. В значительной степени учитывался также опыт разработки рекомендаций по составлению учебных планов в области информатики Computing Curricula международных сообществ ACM и IEEE [2-3].

На самом верхнем уровне рассмотрения в составе Свода были выделены 5 основных областей знаний, представляющих весь спектр проблематики компьютерных вычислений:

1. Математические основы параллельных вычислений.
2. Параллельные вычислительные системы (компьютерные основы).
3. Технологии параллельного программирования (основы программной инженерии).
4. Параллельные алгоритмы решения задач.
5. Параллельные вычисления, большие задачи и конкретные предметные области.

В полном виде разработанный Свод не может быть представлен в рамках статьи (39 разделов, более 300 тем) – полное содержание Свода приведено на сайте проекта <http://hpc-education.ru>.

Приведем состав сформированного Свода знаний и умений до уровня тематических разделов включительно.

№	Область знаний, раздел	Свод2(*)	Свод3
1	Математические основы параллельных вычислений		
1.1	Графовые модели программ	-	-
1.2	Концепция неограниченного параллелизма	-	-
1.3	Тонкая информационная структура программ	-	-
1.4	Эквивалентные преобразования программ	-	-
1.5	Модели вычислений для компьютерных систем	+	+
1.6	Математические модели параллельных вычислений	-/+	-/+
2	Параллельные вычислительные системы (компьютерные основы)		
2.1	Основа машинных вычислений	-	+
2.2	Основа построения компьютерных систем	-/+	+/-
2.3	Параллельные вычислительные системы	-/+	+/-
2.4	Многопроцессорные вычислительные системы	+/-	+/-
2.5	Многопроцессорные вычислительные системы с общей памятью	-/+	+
2.6	Многопроцессорные вычислительные системы с распределенной памятью	+	+
2.7	Графические процессоры	-/+	-
2.8	Вычислительные системы транспетафлопной и эксафлопной производительности	-	-
2.9	Распределенные вычислительные системы	+	-/+
2.10	Проблемы функционирования суперкомпьютерных центров и центров обработки данных	-	-
3.	Технологии параллельного программирования (основы программной инженерии)		
3.1	Общие принципы разработки параллельных программ	-/+	-/+
3.2	Основа параллельного программирования	-/+	+/-

3.3	Методы и технологии разработки параллельных программ	-	+/-
3.4	Параллельные проблемно-ориентированные библиотеки и комплексы программ	-	-
3.5	Инструментальные среды для разработки параллельных программ	-	-
3.6	Методы повышения эффективности параллельных программ	-	-
4.	Параллельные алгоритмы решения задач		
4.1	Общие принципы разработки параллельных алгоритмов	-/+	+/-
4.2	Учебные алгоритмы параллельного программирования	-	+
4.3	Параллельные алгоритмы матричных вычислений	-	+/-
4.4	Параллельные алгоритмы сортировки и поиска данных	-/+	+
4.5	Параллельные алгоритмы обработки графов	-/+	+
4.6	Параллельные алгоритмы решения дифференциальных уравнений в частных производных	-	-
4.7	Параллельные алгоритмы решения оптимизационных задач	-	-
4.8	Параллельные алгоритмы Монте-Карло	-	-
4.9	Параллельные алгоритмы для других классов вычислительно-трудоемких задач	-	-
5.	Параллельные вычисления, большие задачи и конкретные предметные области		
5.1	Параллельные методы решения вычислительно сложных задач наук о Земле.	-	-
5.2	Параллельные методы решения вычислительно сложных задач наук о жизни	-	-
5.3	Параллельные методы решения вычислительно сложных задач инженерных расчетов	-	-
5.4	Параллельные методы решения вычислительно сложных задач квантовой химии	-	-
5.5	Параллельные методы решения задач атомистического моделирования	-	-
5.6	Параллельные методы решения вычислительно сложных задач оборонной тематики	-	-
5.7	Прямые и обратные задачи механики реагирующих сред	-	-

* (В правых столбцах таблицы указывается наличие (+) или отсутствие (-) соответствующих тем Свода в других альтернативных разработках – т.е., в Своде 2 (Рекомендации Computer Science 2013) и в Своде 3 (Рекомендации проекта NSF/IEEE-TCPP) соответственно. Отметка «+/-» означает, что тема представлена в значительной степени, отметка «-/+» говорит о том, что тема раскрывается частично.)

После подготовки начального варианта Свод прошел тщательную экспертизу в Научно-образовательных центрах суперкомпьютерного образования, созданных в рамках выполнения проекта. Разработанный Свод был представлен и активно обсуждался на основных российских конференциях суперкомпьютерной тематики.

Как показывает сопоставительная характеристика, сформированный Свод знаний умений в значительной степени перекрывает предложения двух других разработок (Свода 1 и Свода 3 соответственно). Значительное количество разделов Свода представлено более полно (это разделы, например, Основы компьютерных систем, Общие принципы разработки параллельных программ и др.). Свод содержит также разделы, которые полностью отсутствуют в Сводах 2 и 3 (такие, например, как Вычислительные системы транспетафлопсной и экзафлопсной производительности, Проблемы функционирования суперкомпьютерных центров и центров обработки данных и др.). Кроме того, в сравниваемых Сводах 2 и 3 нет важной области знаний как Параллельные вычисления, большие задачи и конкретные предметные области.

Для конкретизации приведем пример структуру одного из разделов рассматриваемого Свода.

Методы и технологии разработки параллельных программ

1. Традиционные языки программирования и распараллеливающие компиляторы. Векторизация программ.
2. Программные библиотеки для разработки параллельных программ: Shmem, Linda, MPI, PVM, Intel TBB (Thread Building Blocks), Microsoft TPL (Task Parallel Library).
3. Надязыковые средства для организации параллелизма: Cray Fortran, DVM, OpenMP, HPF, Cilk.
4. Параллельные расширения традиционных языков программирования: CAF, UPC.
5. Параллельные языки программирования: Occam, SISAL, HOPMA.
6. Параллельные языки программирования для систем с распределенной общей памятью на основе модели разделённого глобального адресного пространства (partitioned global address space, PGAS): Chapel, X10.
7. Параллельные языки программирования для графических процессоров: CUDA, OpenCL.

8. Функциональные языки параллельного программирования: Parallel Haskell, Parlog, Erlang, T-Система.
9. Средства и технологии для поддержки метакомпьютинга и распределенных вычислений: Globus, gLite, UNICORE, BOINC, X-Com, Map/Reduce, ...
10. Технологии программирования FPGA-компьютеров.
11. Автоматизация распараллеливания и оптимизации программ.
12. Элементы схемотехники, языки описания электронных схем, VHDL.

2. Рекомендации учебных планов Computing Curricula

Деятельность международных сообществ ACM (Association for Computing Machinery) и IEEE Computer Society по подготовке рекомендаций по формированию учебных планов в области Компьютинга (Computing) (К сожалению, для Компьютинга (Computing) нет устоявшегося русскоязычного наименования. Так, при переводе международных рекомендаций Computing Curricula на русский язык использовалось понятие Информатика, а при разработке российских образовательных стандартов на основе Computing Curricula было применено наименование Информационные технологии. В данной работе для устранения разных толкований используется транслитерация **Компьютинг**.) является хорошим примером выполнения работ по определению и постоянному отслеживанию актуальности содержания предметной области. Первый разработанный Свод рекомендаций в области Компьютинга был подготовлен уже в 1968 г. Далее были подготовлены обновленные редакции в 1978, в 1991 и в 2001 гг. Начиная с 2001 г. новые редакции отдельных разделов Компьютинга стали появляться почти ежегодно.

Проблематика параллельных вычислений имеет отношение, прежде всего, к одной из составляющих частей Компьютинга, именуемой как Компьютерная наука (Computer Science) [2-4]. Впервые аспекты параллелизма стали упоминаться в редакции Computing Curricula от 2001 г., но и в 2001 г. [2] и в 2008 г. [3] параллелизм не являлся отдельной областью знаний – проблематика параллельных вычислений включалась как отдельные разделы в нескольких разных частях рекомендаций по компьютерных наукам.

И только в 2013 г. в разрабатываемой в данный момент времени редакции рекомендаций по компьютерных наукам организуется отдельная область знаний Параллельные и Распределенные вычисления (Parallel and Distributed Computing) [4]. В состав этой новой области знаний входят следующие разделы (Как и ранее, столбцы Свод1 (Рекомендации российского национального проекта суперкомпьютерного образования) и Свод3 (Рекомендации проекта NSF/IEEE-TCPP) используются для сопоставления различных вариантов разработанных определения предметной области суперкомпьютерных технологий, высокопроизводительных вычислений и параллельного программирования):

№	Раздел	Свод1	Свод3
1	Основы параллелизма (Parallelism Fundamentals).	+	+
2	Декомпозиция параллельности (Parallel Decomposition).	+	+
3	Коммуникация и координация (Communication and Coordination).	+	+
4	Параллельные алгоритмы, анализ и программирование (Parallel Algorithms, Analysis, and Programming).	+	+
5	Параллельная архитектура (Parallel Architecture).	+	+
6	Параллельная производительность (Parallel Performance).	+	+
7	Распределенные системы (Distributed Systems).	-/+	-/+
8	Облачные вычисления (Cloud Computing).	-/+	-/+
9	Формальные модели и Семантики (Formal Models and Semantics).	+/-	-/+

Выполненное сравнение показывает, что рекомендации Computer Science 2013 практически полностью перекрываются вариантами Сводов двух других рассматриваемых подходов. Можно отметить, что в рассматриваемых рекомендациях более широко рассматриваются разделы, связанные с Распределенными системами, Облачными вычислениями, Формальными моделями и семантикой.

В целом, как вывод, можно заключить, что данный вариант Свода является компактным, но не полным (по сравнению, например, со Сводом российского национального проекта суперкомпьютерного образования).

3. Рекомендации учебных планов NSF/IEEE-TCPP

Еще одна активность по определению предметной области параллельных и распределенных вычислений была предпринята в рамках проекта, поддержанного научным фондом NSF (США) и комитетом IEEE-TCPP (Technical Committee on Parallel Programming) [5]. Предварительный вариант рекомендаций был подготовлен в 2010 г. Рабочая редакция рекомендаций была опубликована в декабре 2012 г. Ожидается, что после активного обсуждения окончательная версия рекомендаций будет подготовлена в 2014 г.

В соответствии с данными рекомендациями в проблематике параллельных и распределенных вычислений выделены четыре области знаний:

1. Архитектура (Architecture).
2. Программирование (Programming).
3. Алгоритмы (Algorithms).
4. Дополнительные разделы (Cross Cutting and Advanced Topics).

Как можно заметить, выделенные области знаний во многом повторяют структуру Свода российского национального проекта суперкомпьютерного образования (Свод 1). Вместе с этим, две области знаний

Математические основы и Параллельные вычисления, большие задачи и конкретные предметные области, имеющиеся в Своде 1, отсутствуют в рекомендациях проекта NSF/IEEE-TCPP. С другой стороны, область Дополнительные разделы рассматриваемых рекомендаций как таковая отсутствует и распределена в других имеющихся областях знаний Свода 1.

К сожалению, в рекомендациях проекта NSF/IEEE-TCPP структурные элементы второго уровня – разделы – в явном виде не выделены. Фактически состав областей знаний определяется сразу в более мелких понятиях – темах, что, безусловно, усложняет использования данного определения предметной области. В ряде случаев предлагаемые темы объединены в группы – с использованием этих выделенных групп структура предметной области в рамках рассматриваемого подхода имеет следующий вид.

№	Область знаний, раздел	Свод1	Свод2
1.	Архитектура		
1.1	Параллелизм данных и управления	+	-/+
1.2	Общая и распределенная память	+	-/+
1.3	Иерархия памяти	+	-/+
1.4	Показатели производительности	+	-/+
1.5	Представление чисел с плавающей запятой	+	-
2.	Программирование		
2.1	Парадигмы параллельного программирования	+	-/+
2.2	Проблемы семантики и корректности	+	+
2.3	Проблемы производительности	+/-	-/+
3.	Алгоритмы		
3.1	Модели и оценки сложности	+	+/-
3.2	Алгоритмические парадигмы	+	-/+
3.3	Алгоритмические проблемы	+	-/+
4.	Дополнительные разделы (тематические группы отсутствуют)	-/+	-/+

Как можно заметить, сопоставление подходов показывает, что рекомендации рассматриваемого варианта (проекта NSF/IEEE-TCPP) практически полностью перекрываются вариантом Свода российского национального проекта суперкомпьютерного образования. С другой стороны, рекомендации проекта NSF/IEEE-TCPP шире и полнее рекомендаций Computer Science 2013. Лишь сводная область Дополнительные разделы рассматриваемых рекомендаций содержит темы, которые учтены не в полной мере в рекомендациях других сравниваемых вариантов (Сводов 1 и 2).

В целом можно заключить, что вариант Свода NSF/IEEE-TCPP является сравнительно полным, но не достаточно структурированным.

Кроме того следует особо отметить, что в рамках рассматриваемого подхода полагается, что большинство понятий, знаний и умений, составляющих предметную область параллельных и распределенных вычислений, должны быть освоены на уровне базовой общей подготовки (бакалавриата). В рекомендациях приводится перечень учебных курсов, которые могут быть использованы для обучения с указанием распределения тем предметной области по этим курсам.

Состав рекомендуемых курсов имеет вид:

- Основные курсы:
 - Введение в программирование,
 - Методы программирования,
 - Архитектура компьютерных систем,
 - Структуры данных и алгоритмы.
- Углубленные курсы (курсы по выбору):
 - Архитектура компьютерных систем (углубленный),
 - Анализ и разработка алгоритмов,
 - Языки программирования,
 - Программная инженерия,
 - Параллельные алгоритмы,
 - Параллельное программирование,
 - Компиляторы,
 - Компьютерные сети,
 - Распределенные сети.

Заключение

Полагаясь на проведенное рассмотрение можно заключить, что в настоящее время в стране и за рубежом активно выполняются работы по определению содержания предметной области суперкомпьютерных технологий, высокопроизводительных вычислений и параллельного программирования. Основными разработками в данном направлении являются:

- Свод знаний и умений, разработанный в рамках российского национального проекта суперкомпьютерного образования [1].
- Рекомендации по формированию учебных планов для компьютерных наук (Computer Science 2013) международных сообществ ACM и IEEE-CS [2-4].
- Рекомендации проекта NSF/IEEE-TCPP [5].

В целом проведенный анализ показывает, что разработанный Свод проекта суперкомпьютерного образования является более полным и во многом охватывает предложения других рассмотренных подходов. Тем не менее, международный опыт разработки рекомендаций является чрезвычайно полезным и обязательно должен учитываться при подготовке последующих редакций Свода. Кроме того, безусловно необходимым является организация взаимовыгодного взаимодействия всех профессиональных коллективов, занимающихся разработкой предложений состава предметной области СКТ.

ЛИТЕРАТУРА:

1. Воеводин В.В., Гергель В.П., Соколинский Л.Б., Демкин В.П., Попова Н.Н., Бухановский А.В. Развитие системы суперкомпьютерного образования в России: текущие результаты и перспективы // Вестник Нижегородского университета им. Н.И. Лобачевского. 2012. № 4. С. 268-274.
2. Рекомендации по преподаванию программной инженерии и информатики в университетах. - М.: ИНТУИТ.РУ «Интернет-Университет Информационных технологий», 2007 (электронный вариант издания содержится на <http://se.math.spbu.ru/SE>)
3. Computing Curricula Computer Science 2008 – см. http://www.computer.org/portal/c/document_library/get_file?p_l_id=2814020&folderId=3111026&name=DLFE-57604.pdf
4. Computing Curricula Computer Science 2013 – см. <http://ai.stanford.edu/users/sahami/CS2013/>
5. NSF/IEEE-TCPP Curriculum Initiative on Parallel and Distributed Computing – см. <http://www.cs.gsu.edu/~tcpp/curriculum/>
6. Воеводин В.В., Воеводин Вл.В. Параллельные вычисления. – СПб.: БХВ-Петербург, 2002.
7. Гергель В.П. Теория и практика параллельных вычислений. – М.: Интуит Бином. Лаборатория знаний, 2007.
8. Электронная энциклопедия по параллельным вычислениям Параллель. // Электронный ресурс <http://parallel.guru.ru>
9. Антонов А.С., Воеводин Вл.В., Гергель В.П., Соколинский Л.Б. Системный подход к суперкомпьютерному образованию // Вестник ЮУрГУ. Серия "Вычислительная математика и информатика". 2013. Т. 2. N 2. С. 5-17.
10. Антонов А.С., Артемьева И.Л., Бухановский А.В., Воеводин В.В., Гергель В.П., Демкин В.П., Коньков К.А., Крукиер Л.А., Попова Н.Н., Соколинский Л.Б., Сухинов А.И. Проект "Суперкомпьютерное образование": 2012 год // Вестник Нижегородского университета им. Н.И. Лобачевского. 2013. N 1-1. С. 12-16.
11. Воеводин В.В., Гергель В.П., Соколинский Л.Б., Демкин В.П., Попова Н.Н., Бухановский А.В. Развитие системы суперкомпьютерного образования в России: текущие результаты и перспективы // Вестник Нижегородского университета им. Н.И. Лобачевского. 2012. N 4-1. С. 268-274.