

Цели и задачи курса

В учебном курсе изучаются технологии параллельного программирования, предназначенные для разработки высокопроизводительных реализаций вычислительно трудоемких алгоритмов для выполнения на параллельных вычислительных системах различной архитектуры. Технологии ориентированы на кластеры/суперкомпьютеры (технология MPI), состоящие из вычислительных узлов на базе традиционных многоядерных центральных процессоров (технологии OpenMP, Intel Cilk Plus, Intel TBB, Intel ArBB, OpenCL), а также гетерогенных узлов с использованием графических процессоров (технологии NVIDIA CUDA, OpenCL).

Лекционная часть предполагает изучение областей применения и основных возможностей разных технологий параллельного программирования. В рамках лабораторного практикума рассматривается цикл работ, включая создание последовательной реализации в качестве базы для сравнения, подготовку параллельной версии, ее анализ и, при необходимости, оптимизацию. Изучение проводится на модельных задачах, не требующих наличия специфических знаний из конкретных предметных областей, за исключением информации, присутствующей в работе.

Будучи изначально ориентированным на студентов физико-математического направления (бакалавриат, специалитет, магистратура), курс рассчитан также на преподавателей и научных сотрудников, а также аспирантов высших учебных заведений, изучающих параллельное программирование.

Требования к слушателям

Знание учебного материала курсов K101 «Введение в методы программирования-1», K102 «Введение в методы программирования-2», K206 «Архитектура вычислительных систем», K202 «Операционные системы», П101 «Введение в параллельное программирование».

Курс предполагает наличие у слушателей базовых знаний и навыков структурного, модульного и (в некоторых разделах) объектно-ориентированного программирования. В качестве базового языка рассматривается C/C++. Многие идеи могут быть с успехом перенесены на другой язык, поддерживающий параллельное программирование, в частности Fortran.

Описание курса

Основными учебными разделами курса являются:

- **Введение в предмет** – 2 часа

Понятия последовательного и параллельного алгоритма, последовательной и параллельной программы. Закон Амдаля. Цели создания параллельных программ. Способы и инструментальные средства разработки параллельных программ, общая характеристика.

- **Параллельное программирование для кластерных систем на основе технологии MPI** – 12 часов

Структура программы на MPI, принципы организации параллелизма. Настройка средств разработки, способы запуска. Состав MPI. Режимы передачи данных. Группы процессов и коммутаторы. Парные и коллективные операции. Создание и использование собственных типов данных в MPI для передачи данных. Работа с виртуальными топологиями.

- **Параллельное программирование для систем с общей памятью на основе технологии OpenMP** – 12 часов

Структура OpenMP-программы, принципы организации параллелизма. Составные части OpenMP: директивы компилятора, функции run-time библиотеки, переменные окружения. Основные директивы OpenMP, формат записи. Типы директив. Распределение вычислений между потоками.

Управление областью видимости данных. Синхронизация как задача параллельного программирования. Средства синхронизации в OpenMP. Библиотека функций OpenMP.

- **Параллельное программирование для систем с общей памятью на основе технологии Intel Cilk Plus** – 7 часов

Intel Cilk Plus – расширение C/C++. Основные возможности Cilk Plus. Подключение Cilk Plus. Конструкции `cilk_spawn` и `cilk_sync`, синтаксис и механизм исполнения. Конструкция `cilk_for`, синтаксис и механизм исполнения.

- **Параллельное программирование для систем с общей памятью на основе технологии Intel Array Building Blocks (ArBB)** – 4 часа

Intel ArBB – библиотека векторного параллелизма. Основные возможности ArBB. Подключение библиотеки ArBB. Некоторые синтаксические конструкции ArBB. Основные типы и структуры данных, операции над ними. Некоторые функции работы с контейнерами. Управляющие конструкции библиотеки ArBB. Объявление и вызов функций при использовании библиотеки ArBB.

- **Параллельное программирование для систем с общей памятью на основе технологии Intel Threading Building Blocks (TBB)** – 14 часов

Инициализация и завершение библиотеки. Распараллеливание простых циклов. Распараллеливание сложных конструкций: сортировка, циклы с условием. Механизм логических задач.

- **Параллельное программирование для графических процессоров на основе технологии NVIDIA CUDA** – 12 часов

Введение в GPGPU и гетерогенное программирование. Обзор архитектуры графических процессоров. Понятие ядра. Поток и блоки потоков. CUDA C – язык параллельного программирования для графических процессоров. Исполнение блоков и потоков, связь с архитектурой. Иерархия памяти CUDA. Синхронизация и организация взаимодействия между потоками. Общие принципы создания параллельных программ для графических процессоров. Основные приемы оптимизации приложений на CUDA.

- **Параллельное программирование для гетерогенных систем на основе технологии OpenCL** – 8 часов

OpenCL как стандарт гетерогенных вычислений. Модели платформы, исполнения, памяти, программирования. Язык OpenCL C.

- **Параллельное программирование для гетерогенных систем с одновременным использованием центральных и графических процессоров** – 1 час

Одновременное использование всех доступных вычислительных ресурсов в гетерогенных системах: возможности, проблемы, перспективы. Примеры гетерогенных приложений с одновременным использованием центральных и графических процессоров.

Учебный курс включает расширенный лабораторный практикум по параллельному программированию. Данный практикум содержит как технологические лабораторные работы по освоению технологий MPI и OpenMP, так и задания по разработке параллельных программ для решения сложных вычислительно-трудоемких задач.

Учебно-методическое обеспечение курса содержит:

- комплект электронных учебных материалов;
- презентации для проведения лекционных занятий;
- комплект лабораторных работ.

Ожидаемые результаты обучения

Обучаемые, успешно освоившие учебный курс, будут:

- применять технологии разработки параллельных программ для решения типовых задач вычислительной математики (матричные вычисления, сортировка, обработка графов, оптимизация);
- разрабатывать, отлаживать и оптимизировать параллельные программы для традиционных и гетерогенных кластерных/супер-компьютерных систем;
- использовать при разработке параллельных программ технологии MPI, OpenMP, Intel Cilk Plus, Intel TBB, Intel ArBB, OpenCL, NVIDIA CUDA;
- выполнять вычислительные эксперименты на высокопроизводительных вычислительных системах;
- проводить оценку эффективности выполняемых параллельных вычислений.

Учебная литература

1. Гергель В.П. Теория и практика параллельных вычислений. М.: Интернет-Университет Информационных технологий; Бином, 2007. 424 с.
2. Корняков К. В., Кустикова В. Д., Мееров И. Б., Сиднев А. А., Сысоев А. В., Шишков А. В. Инструменты параллельного программирования в системах с общей памятью: учебник / Под ред. проф. В.П. Гергеля. М.: Изд-во Московского университета, 2010. 272 с.
3. Линева А.В., Боголепов Д.К., Бахраков С.И. Технологии параллельного программирования для процессоров новых архитектур: учебник / Под ред. В.П. Гергеля. М.: Изд-во Московского университета, 2010. 160 с.
4. Сандерс Д., Кэндрот Э. Технология CUDA в примерах: введение в программирование графических процессоров. М.: ДМК, 2011.

Дополнительные сведения

Курс прошел успешную апробацию в рамках научных школ, конференций, программ повышения квалификации ППС, проводимых в Москве, Санкт-Петербурге, Нижнем Новгороде, Перми, Белгороде и др., и учебного процесса на факультете ВМК ННГУ.

Разработка учебно-методического и программного обеспечения данного курса была поддержана компанией Интел (2009-2011).

Контактная информация:

ст. преп. каф. МО ЭВМ, к.т.н. А.В. Сысоев, sysoyev@vmk.unn.ru