

Министерство образования и науки Российской Федерации
Уральский федеральный университет
имени первого Президента России Б. Н. Ельцина

К. А. Некрасов, С. И. Поташников,
А. С. Боярченков, А. Я. Купряжкин

Параллельные вычисления общего назначения на графических процессорах

Учебное пособие

Рекомендовано методическим советом УрФУ
для студентов, обучающихся по направлениям подготовки
14.04.02 — Ядерная физика и технологии;
09.04.02 — Информационные системы и технологии;
14.04.01 — Ядерные реакторы и материалы

2-е издание, стереотипное

Москва
Издательство «ФЛИНТА»
Издательство Уральского университета
2017

УДК 004.032.24:004.383.5(075.8)

ББК 32.97я73

П18

Авторы:

Некрасов К. А., Поташников С. И., Боярченков А. С., Купряжкин А. Я.

Рецензенты:

Институт теплофизики УрО РАН (д-р физ.-мат. наук, проф. В. Г. Байдаков); гл. науч. сотр. лаборатории математического моделирования Института промышленной экологии УрО РАН д-р физ.-мат. наук, проф. А. Н. Вадаксин

**Параллельные вычисления общего назначения на графических процессо-
П18рах** [Электронный ресурс]: учебное пособие / К. А. Некрасов, С. И. Поташников, А. С. Боярченков, А. Я. Купряжкин. — 2-е изд., стер. — М. : ФЛИНТА : Изд-во Урал. ун-та, 2017. — 102 с.

ISBN 978-5-9765-3182-6 (ФЛИНТА)

ISBN 978-5-7996-1722-6 (Изд-во Урал. ун-та)

В учебном пособии изложены основные принципы организации высокоскоростных параллельных вычислений на графических процессорах. Рассмотрены подходы к программированию графических процессоров с использованием шейдерной модели и NVIDIA CUDA. Проанализированы примеры.

Пособие предназначено для проведения практических занятий по программированию графических процессоров для магистрантов.

Библиогр.: 22 назв. Рис. 24. Табл. 3. Прил. 1.

УДК 004.032.24:004.383.5(075.8)

ББК 32.97я73

ISBN 978-5-9765-3182-6 (ФЛИНТА)

ISBN 978-5-7996-1722-6 (Изд-во Урал. ун-та)

© Уральский федеральный университет, 2016

Оглавление

Введение	3
1. Структура и возможности вычислительной системы с графическим процессором	4
1.1. Задача компьютерной визуализации трехмерных сцен.....	4
1.2. Архитектура графического процессора (GPU)	5
1.2.1. Распараллеливание вычислений по данным	5
1.2.2. Взаимодействие графического и центрального процессоров ...	8
1.2.3. Иерархия памяти, доступной центральному и графическому процессорам	9
1.2.4. Конвейерная обработка данных	12
1.3. Уровни управления графическим процессором и основные системы программирования GPU.....	14
1.3.1. Уровни управления графическим процессором	14
1.3.2. Драйвер графического процессора.....	14
1.3.3. Интерфейсы программирования приложений	15
1.3.4. Пользовательское приложение.....	17
1.3.5. Программно-аппаратная платформа NVIDIA CUDA.....	19
1.3.6. Выбор платформы программирования GPU	20
1.4. Области применения графических процессоров	21
1.5. Необходимое аппаратное и программное обеспечение	21
Средства программирования.....	23
2. Поточно-параллельное программирование GPU.....	25
2.1. Распараллеливание расчетов	25
2.1.1. Распараллеливание по задачам.....	25
2.1.2. Распараллеливание по инструкциям.....	26
2.1.3. Распараллеливание по данным.....	27
2.2. Преимущества графических процессоров при параллельных расчетах	28
2.3. Принцип программирования SIMD на примере пиксельного шейдера	30
2.4. Пример сложения матриц.....	32
2.4.1. Распараллеливание независимых вычислений.....	32
2.4.2. Сложение матриц в рамках шейдерной модели 3.0	34
2.4.3. Структура программы для центрального процессора.....	36
2.4.4. Реализация программы для центрального процессора на C# ...	37

2.4.5. Программа для графического процессора	43
2.4.6. Вычислительные шейдеры модели 5.0	49
3. Программирование графических процессоров на CUDA.....	50
3.1. Модель программирования графических процессоров как универсальных вычислительных систем	50
3.1.1. Взаимодействие параллельных вычислительных процессов ..	50
3.1.2. Концепция универсального вычислительного устройства CUDA.....	51
3.1.3. Иерархия вычислительных процессов и памяти CUDA.....	53
3.1.4. Возможности и ограничения процессоров архитектуры CUDA	55
3.1.5. Конвейерная обработка данных в архитектуре CUDA.....	56
3.2. Особенности программирования на CUDA	56
3.2.1. Идентификация вычислительного потока.....	56
3.2.2. Совместимость с шейдерными моделями.....	58
3.2.3. Язык программирования CUDA	58
3.2.4. Структура программы на CUDA.....	59
3.3. Анализ алгоритма параллельного перемножения матриц.....	62
3.3.1. Алгоритм перемножения матриц	62
3.3.2. Процедура перемножения матриц на CUDA.....	64
3.3.3. Оптимизация доступа к памяти при умножении матриц.....	69
3.4. Динамика N тел на CUDA. Пример ускорения программы за счет скорости GPU	73
3.5. Распараллеливание алгоритмов сортировки. Пример ускорения программы за счет скорости GPU.....	76
Заключение	79
Приложение	80
Перемножение матриц на CUDA. Программа, исполняемая центральный процессором.....	80
Вычисление скалярного произведения векторов на CUDA	84
Компиляция программ на CUDA.....	88
Профайлер	94
Библиографический список.....	98