

УДК 004.3'144: 004.383.5CUDA
ББК 32.973.26-04
С18

- С18 Сандерс Дж., Кэндрот Э.
Технология CUDA в примерах: введение в программирование графических процессоров: Пер. с англ. Слинкина А. А., научный редактор Боресков А. В. – М.: ДМК Пресс, 2018. – 232 с.: ил.
ISBN 978-5-97060-581-3

CUDA – вычислительная архитектура, разработанная компанией NVIDIA и предназначенная для разработки параллельных программ. В сочетании с развитой программной платформой архитектура CUDA позволяет программисту задействовать невероятную мощь графических процессоров для создания высокопроизводительных приложений, включая научные, инженерные и финансовые приложения.

Книга написана двумя старшими членами команды по разработке программной платформы CUDA. Новая технология представлена в ней с точки зрения программиста. Авторы рассматривают все аспекты разработки на CUDA, иллюстрируя изложение работающими примерами. После краткого введения в саму платформу и архитектуру CUDA, а также беглого обзора языка CUDA C, начинается подробное обсуждение различных функциональных возможностей CUDA и связанных с ними компромиссов. Вы узнаете, когда следует использовать то или иное средство и как писать программы, демонстрирующие поистине выдающуюся производительность.

Издание предназначено для программистов, а также будет полезно инженерам, научным работникам и студентам вузов.

УДК 004.3'144: 004.383.5CUDA
ББК 32.973.26-04

Все права защищены. Любая часть этой книги не может быть воспроизведена в какой бы то ни было форме и какими бы то ни было средствами без письменного разрешения владельцев авторских прав.

Материал, изложенный в данной книге, многократно проверен. Но поскольку вероятность технических ошибок все равно существует, издательство не может гарантировать абсолютную точность и правильность приводимых сведений. В связи с этим издательство не несет ответственности за возможные ошибки, связанные с использованием книги.

ISBN 978-0-13-138768-3 (анг.)
ISBN 978-5-97060-581-3 (рус.)

© NVIDIA Corporation
© Оформление, ДМК Пресс



Содержание

Предисловие	10
Вступление	12
Благодарности	14
Об авторах	15
Глава 1. Почему CUDA? Почему именно теперь?	16
1.1. О чем эта глава	16
1.2. Век параллельной обработки	16
1.2.1. Центральные процессоры	17
1.3. Развитие GPU-вычислений	18
1.3.1. Краткая история GPU	18
1.3.2. Ранние этапы GPU-вычислений	19
1.4. Технология CUDA	20
1.4.1. Что такое архитектура CUDA?	21
1.4.2. Использование архитектуры CUDA	21
1.5. Применение CUDA	22
1.5.1. Обработка медицинских изображений	22
1.5.2. Вычислительная гидродинамика	23
1.5.3. Науки об окружающей среде	24
1.6. Резюме	24
Глава 2. Приступая к работе	26
2.1. О чем эта глава	26
2.2. Среда разработки	26
2.2.1. Графические процессоры, поддерживающие архитектуру CUDA	26
2.2.2. Драйвер устройства NVIDIA	28
2.2.3. Комплект средств разработки CUDA Development Toolkit	28
2.2.4. Стандартный компилятор	30
Windows	30
Linux	30
Macintosh OS X	31
2.3. Резюме	31
Глава 3. Введение в CUDA C	32
3.1. О чем эта глава	32

Глава 4. Параллельное программирование на CUDA C..... 43

Глава 5. Взаимодействие нитей 59

Глава 6. Константная память и события 88

Глава 7. Текстурная память 104

7.1. О чем эта глава	104
7.2. Обзор текстурной памяти	104
7.3. Моделирование теплообмена.....	105

7.3.1. Простая модель теплообмена	105
7.3.2. Обновление температур	106
7.3.3. Анимация моделирования	108
7.3.4. Применение текстурной памяти.....	112
7.3.5. Использование двумерной текстурной памяти	116
7.4. Резюме	120

Глава 8. Интероперабельность с графикой..... 121

8.1. О чем эта глава	121
8.2. Взаимодействие с графикой.....	122
8.3. Анимация волн на GPU с применением интероперабельности с графикой	128
8.3.1. Структура GPUAnimBitmap	129
8.3.2. И снова об анимации волн на GPU	132
8.4. Моделирование теплообмена с использованием интероперабельности с графикой	134
8.5. Интероперабельность с DirectX.....	138
8.6. Резюме	138

Глава 9. Атомарные операции..... 140

9.1. О чем эта глава	140
9.2. Вычислительные возможности	140
9.2.1. Вычислительные возможности NVIDIA GPU	141
9.2.2. Компиляция программы для GPU с заданным минимальным уровнем вычислительных возможностей.....	142
9.3. Обзор атомарных операций.....	143
9.4. Вычисление гистограмм	145
9.4.1. Вычисление гистограммы на CPU	145
9.4.2. Вычисление гистограммы на GPU	147
Ядро вычисления гистограммы с применением атомарных операций с глобальной памятью	151
Ядро вычисления гистограммы с применением атомарных операций с глобальной и разделяемой памятью	153
9.5. Резюме	155

Глава 10. Поток..... 156

10.1. О чем эта глава	156
10.2. Блокированная память CPU	156
10.3. Поток CUDA.....	161
10.4. Использование одного потока CUDA.....	161
10.5. Использование нескольких потоков CUDA	166
10.6. Планирование задач на GPU	171
10.7. Эффективное использование потоков CUDA.....	173
10.8. Резюме	175

Глава 11. CUDA C на нескольких GPU..... 176

11.1. О чем эта глава	176
11.2. Нуль-копируемая память CPU	176

Глава 12. Последние штрихи	194
12.1. О чем эта глава	194
12.2. Инструментальные средства CUDA	195
12.2.1. CUDA Toolkit	195
12.2.2. Библиотека CUFFT	195
12.2.3. Библиотека CUBLAS	196
12.2.4. Комплект NVIDIA GPU Computing SDK	196
12.2.5. Библиотека NVIDIA Performance Primitives	197
12.2.6. Отладка программ на языке CUDA C	197
CUDA-GDB	197
NVIDIA Parallel Nsight	198
12.2.7. CUDA Visual Profiler	198
12.3. Текстовые ресурсы	200
12.3.1. Programming Massively Parallel Processors: A Hands-On Approach	200
12.3.2. CUDA U	200
Материалы университетских курсов	201
Журнал DR. DOBB'S	201
12.3.3. Форумы NVIDIA	201
12.4. Программные ресурсы	202
12.4.1. Библиотека CUDA Data Parallel Primitives Library	202
12.4.2. CULAtools	202
12.4.3. Интерфейсы к другим языкам	203
12.5. Резюме	203

Приложение А. Еще об атомарных операциях	204
А.1. И снова скалярное произведение	204
А.1.1. Атомарные блокировки	206
А.1.2. Возвращаясь к скалярному произведению: атомарная блокировка	208
А.2. Реализация хеш-таблицы	211
А.2.1. Обзор хеш-таблиц	212
А.2.2. Реализация хеш-таблицы на CPU	214
А.2.3. Многонитевая реализация хеш-таблицы	218
А.2.4. Реализация хеш-таблицы на GPU	219
А.2.5. Производительность хеш-таблицы	225
А.3. Резюме	226

Предметный указатель227