

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«Тольяттинский государственный университет»

Б1.О.05

(индекс дисциплины)

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Архитектура и программное обеспечение высокопроизводительных вычислительных систем

(наименование дисциплины)

по направлению подготовки
01.04.02 Прикладная математика и информатика

направленность (профиль)/специализация
Математическое моделирование

Форма обучения: очная

Год набора: 2019

Общая трудоемкость: 4Е

Распределение часов дисциплины по семестрам

Семестр	3	Итого
Форма контроля	зачет	
Вид занятий		
Лекции	18	18
Лабораторные		
Практические	18	18
Руководство: курсовые работы (проекты) / РГР		
Промежуточная аттестация	0,25	0,25
Контактная работа	36,25	36,25
Самостоятельная работа	107,75	107,75
Контроль		
Итого	144	144

Рабочую программу составил(и):

доцент кафедры ПМИИ, к.т.н., Очеповский А.В.

(должность, ученое звание, степень, Фамилия И.О.)

Рецензирование рабочей программы дисциплины:



Отсутствует



Рецензент

(должность, ученое звание, степень, Фамилия И.О.)

Рабочая программа дисциплины составлена на основании ФГОС ВО и учебного плана направления подготовки (специальности)

01.04.02 Прикладная математика и информатика

(код и наименование направления подготовки, специальности в соответствии с ФГОС ВПО)

Срок действия программы дисциплины до «31» августа 2021 г.

УТВЕРЖДЕНО

На заседании кафедры «Прикладная математика и информатика»

(протокол заседания № 1 от «30» августа 2018г.).

1. Цель освоения дисциплины

Цель – формирование у студентов научных представлений о принципах разработки и реализации высокопроизводительных многопоточных алгоритмов для платформы CUDA Nvidia и практических навыков применения полученных знаний для решения задач в области прикладной математики и информатики.

2. Место дисциплины в структуре ОПОП ВО

Дисциплины и практики, на освоении которых базируется данная дисциплина:

- Избранные вопросы теоретической информатики;
- Компьютерное моделирование.

Дисциплины и практики, для которых освоение данной дисциплины необходимо как предшествующее:

- Производственная практика (научно-исследовательская работа) 4;
- Производственная практика (преддипломная практика);
- государственная итоговая аттестация.

3. Планируемые результаты обучения

Формируемые и контролируемые компетенции (код и наименование)	Индикаторы достижения компетенций (код и наименование)	Планируемые результаты обучения
(ОПК-1) Способен решать актуальные задачи фундаментальной и прикладной математики	(ИОПК-1.1) Демонстрирует знания в области фундаментальной и прикладной математики	Знать: методы прикладной математики и информатики
	(ИОПК-1.2) Демонстрирует знания актуальных нерешенных проблем организационно-технических и экономических процессов	Уметь: использовать новые знания и применять их в профессиональной деятельности
	(ИОПК-1.3) Демонстрирует умения применять математические методы при решении задач фундаментальной и прикладной математики	Владеть: методологией и навыками решения научных и практических задач
(ОПК-2) Способен совершенствовать и реализовывать новые математические методы решения прикладных задач	(ИОПК-2.1) Демонстрирует понимание теории систем и системного анализа, дискретной математики, теории вероятностей и математической статистики, методов оптимизации и исследования операций, нечетких вычислений,	Знать: современное состояние и принципиальные возможности языков и систем программирования на GPU;

Формируемые и контролируемые компетенции (код и наименование)	Индикаторы достижения компетенций (код и наименование)	Планируемые результаты обучения
	математического и имитационного моделирования	
	(ИОПК-2.2) Анализирует существующие методы решения прикладных задач для выбора рационального решения.	Уметь: использовать современные теории, методы, системы и средства прикладной математики и информационных технологий для решения научно-исследовательских и прикладных задач
	(ИОПК-2.3.) Демонстрирует способности совершенствовать существующие методы прикладной математики, а также реализовывать новые математические методы решения прикладных задач.	Владеть: способностью интерпретировать полученные данные по соответствующим научным исследованиям
(ОПК-4) Способен комбинировать и адаптировать существующие информационно-коммуникационные технологии для решения задач в области профессиональной деятельности с учетом требований информационной безопасности	(ИОПК-4.1) Анализирует методы и средства решения актуальных задач профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности	Знать: архитектуры гетерогенных параллельных вычислительных систем; основные методы обработки данных.
	(ИОПК-4.2.) Демонстрирует возможности комбинировать и адаптировать существующие информационно-коммуникационные технологии для решения задач в области профессиональной деятельности с учетом требований информационной безопасности	Уметь: использовать найденные и полученные данные для исследования и решения научных и практических задач
	(ИОПК-4.3) Оценивает	Владеть: навыками работы с

Формируемые и контролируемые компетенции (код и наименование)	Индикаторы достижения компетенций (код и наименование)	Планируемые результаты обучения
	стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности	современными программными и аппаратными средствами информационных технологий для выполнения научных исследований

4. Структура и содержание дисциплины

Модуль (раздел)	Вид учебной работы	Наименование тем занятий (учебной работы)	Семестр	Объем, ч.	Баллы	Интерактив, ч.	Формы текущего контроля (наименование оценочного)
Модуль 1. История и архитектура параллельных вычислений на CUDA	Лек.	Тема 1. Гетерогенные параллельные архитектуры	3	2	-	-	
	Пр.	Практическое занятие №1. Технология программирования на CUDA	3	2	-	-	Отчет по практической работе 1
Модуль 2. Параллельные вычисления CUDA	Лек.	Тема 2. Модель вычислений и памяти CUDA	3	4	-	-	
	Лек.	Тема 3. Потоки и параллелизм CUDA	3	4	-	-	
	Лек.	Тема 4. Концепции реализации базовых параллельных алгоритмов на Cuda	3	4	-	-	
	Пр.	Практическое занятие №2. Вычислительный эксперимент для оценки эффективности параллельных вычислений по технологии CUDA	3	4	-	-	Отчет по практической работе 2
	Пр.	Практическое занятие №3. Визуализация анимированных графиков функций при помощи CUDA	3	4	-	-	Отчет по практической работе 3
	Лек.	Тема 5. Оптимизация алгоритмов: CUDA libraries	3	4	-	-	
	Пр.	Практическое занятие №4. Вычислительный эксперимент для оценки эффективности обработки изображений с распараллеливанием вычислений с помощью технологии CUDA	3	4	-	-	Отчет по практической работе 4

Модуль (раздел)	Вид учебной работы	Наименование тем занятий (учебной работы)	Семестр	Объем, ч.	Баллы	Интерактив, ч.	Формы текущего контроля (наименование оценочного)
	Пр.	Практическое занятие №5. Вычислительный эксперимент для оценки эффективности использования высокопроизводительных библиотек для распараллеливания вычислений с помощью технологии CUDA	3	4	-	-	Отчет по практической работе 5
	ПА		3	0,25	-	-	
	СР		3	107,75	-	-	
		Контроль			-	-	
Итого:				144			

5. Образовательные технологии

В рамках изучения дисциплины предусмотрено использование следующих образовательных технологий:

- технология традиционного обучения: лекции и практические работы, самостоятельная работа;
- технология проектного обучения: реализация и защита отчетов по практическим работам.

6. Методические указания по освоению дисциплины

Изучение дисциплины требует систематического и последовательного накопления знаний, следовательно, пропуски отдельных тем не позволяют глубоко освоить предмет.

В ходе лекционных занятий рекомендуется задавать преподавателю уточняющие вопросы с целью уяснения теоретических положений, разрешения спорных ситуаций.

Студент может дополнить список использованной литературы современными источниками, не представленными в списке рекомендованной литературы, и в дальнейшем использовать собственные подготовленные учебные материалы при написании курсовых и выпускных квалификационных работ.

6.1. Рекомендации по подготовке к практическим занятиям

Студентам следует:

- при подготовке к практическим занятиям следует обязательно использовать не только лекции, учебную литературу, но и другие источники;
- в начале занятий задать преподавателю вопросы по материалу, вызвавшему затруднения в его понимании и освоении при решении задач, заданных для самостоятельного решения;
- на занятии доводить каждую задачу до окончательного решения, демонстрировать понимание проведенных расчетов (анализов, ситуаций), в случае затруднений обращаться к преподавателю.

Для того чтобы практические занятия приносили максимальную пользу, необходимо помнить, что упражнение и решение задач проводятся по рассмотренному на лекциях материалу и связаны, как правило, с детальным разбором отдельных вопросов лекционного курса. Следует подчеркнуть, что только после усвоения лекционного материала с определенной точки зрения (а именно с той, с которой он излагается на лекциях) он будет закрепляться студентом на практических занятиях как в результате обсуждения и анализа лекционного материала, так и с помощью решения проблемных ситуаций, задач. При этих условиях студент не только хорошо усвоит материал, но и научится применять его на практике, а также получит дополнительный стимул (и это очень важно) для активной работы на лекции.

При самостоятельном решении задач нужно обосновывать каждый этап решения, исходя из теоретических положений курса. Если студент видит несколько путей решения проблемы (задачи), то нужно сравнить их и выбрать самый рациональный. Полезно до начала вычислений составить краткий план решения проблемы (задачи). Решение проблемных задач или примеров следует излагать подробно, вычисления располагать в строгом порядке, отделяя вспомогательные вычисления от основных. Решения при необходимости нужно сопровождать комментариями, схемами, чертежами и рисунками.

Следует помнить, что решение каждой учебной задачи должно доводиться до

окончательного логического ответа, которого требует условие, и по возможности с выводом. Полученный ответ следует проверить способами, вытекающими из существа данной задачи. Полезно также (если возможно) решать несколькими способами и сравнить полученные результаты. Решение задач данного типа нужно продолжать до приобретения твердых навыков в их решении.

7. Оценочные средства

7.1. Паспорт оценочных средств

Семестр	Код контролируемой компетенции (или ее части)	Наименование оценочного средства
3	ОПК-1, ОПК-2, ОПК-4	Собеседование по модулю 1 Собеседование по модулю 2 Отчеты по практическим работам 1-5

7.2. Типовые задания или иные материалы, необходимые для текущего контроля

7.2.1. Вопросы для собеседования по модулю (примеры вопросов) (наименование оценочного средства)

Модуль 1. История и архитектура параллельных вычислений на CUDA

1. Предпосылки вычислений на GPU
2. Особенности применения видеокарт в 2001-2006 гг для вычислений
3. «Многообещающие» результаты применения видеокарт в 2001-2006 гг. для вычислений
4. Проблемы начального этапа применения видеокарт для вычислений
5. 2006г: появление CUDA и G80
6. 2006г: другие производители видеокарт в борьбе за новую нишу гетерогенных вычислений
7. 2007-2008гг: быстрый рост производительности видеокарт и их применение для вычислений
8. Современный этап: оборудование, программы, производительность

Критерии оценки:

- оценка «зачтено» выставляется студенту, если продемонстрированы всесторонние, систематизированные, глубокие знания по истории вычислений на GPU;
- оценка «не зачтено» выставляется студенту, если продемонстрированы фрагментарные, несистематизированные знания по истории вычислений на GPU, допускаются неточности и ошибки причинно-следственных связей.

Модуль 2. Параллельные вычисления CUDA

1. Преимущества CUDA, место CUDA в современном HPC
2. Архитектура GPU в сравнении с архитектурой CPU
3. Понятие ComputeCapability, Fermi и Kepler
4. Гибридное программирование CPU+GPU: сетки, ядра, запуск ядер
5. Гибридное программирование CPU+GPU: выбор устройства, обработка ошибок, асинхронность в CUDA

Критерии оценки:

- оценка «зачтено» выставляется студенту, если продемонстрированы всесторонние, систематизированные, глубокие знания по гетерогенным параллельным архитектурам;
- оценка «не зачтено» выставляется студенту, если продемонстрированы фрагментарные, несистематизированные знания Гетерогенные параллельные архитектуры, допускаются неточности и ошибки.

7.2.2. Комплект отчетов по практическим работам (примеры). *(наименование оценочного средства)*

Практическая работа №1 «Технология программирования на CUDA»

Форма отчета по практической работе

- титульный лист;
- задание;
- основные концепции решения;
- код программы;
- экранная форма разработанного приложения (с результатами консольного вывода);
- выводы.

Практическая работа №2 «Вычислительный эксперимент для оценки эффективности параллельных вычислений по технологии CUDA»

Форма отчета по практической работе

- титульный лист;
- задание;
- код программы;
- экранная форма разработанного приложения;
- результат эксперимента (таблицы и графики);
- выводы

Практическая работа №3 «Визуализация анимированных графиков функций при помощи CUDA»

Форма отчета по практической работе

- титульный лист;
- задание;
- код программы;
- экранная форма разработанного приложения;
- выводы.

Практическая работа №4 «Вычислительный эксперимент для оценки эффективности обработки изображений с распараллеливанием вычислений с помощью технологии CUDA»

Форма отчета по практической работе

- титульный лист;
- задание;
- код программы;
- экранная форма разработанного приложения;
- результат эксперимента (таблицы и графики);
- выводы.

Практическая работа №5 «Вычислительный эксперимент для оценки эффективности использования высокопроизводительных библиотек для распараллеливания вычислений с помощью технологии CUDA»

Форма отчета по практической работе

- титульный лист;
- задание;
- код программы;
- экранная форма разработанного приложения;
- результат эксперимента (таблицы и графики);
- выводы.

Требования к оформлению

Отчёт по практической работе выполняется на страницах формата А4 в электронном виде.

При оформлении отчёта используется сквозная нумерация страниц, считая титульный лист первой страницей. Номер страницы на титульном листе не ставится. Номера страницы ставятся по центру сверху.

При оформлении отчёта соблюдать следующие требования:

- Для заголовков: полужирный шрифт, 14 пт, центрированный.
- Для основного текста: нежирный шрифт, 14 пт, выравнивание по ширине.
- Во всех случаях тип шрифта – Times New Roman, отступ абзаца 1.25 см, полуторный междустрочный интервал.
- Поля: левое – 2 см, правое, верхнее и нижнее – 1 см.

Процедура оценивания

Оценка выполненной работы проводится по следующим критериям:

1. Наличие всей существенной информации по работе
2. Точность и полнота предоставляемых сведений
3. Непротиворечивость приводимой информации
4. Правильность интерпретаций и выводов, которые сделаны по результатам работы
5. Степень достижения студентом поставленной цели
6. Обоснованность применяемого решения
7. Грамотность (содержательная) используемых формулировок

Критерии оценки за отчеты по практическим работам:

– оценка «зачтено» ставится студенту, который продемонстрировал результаты выполнения практической работы, соответствующие поставленным задачам, и представил отчет, оформленный должным образом и содержащий краткое описание полученных результатов;

– оценка «не зачтено» ставится студенту, который не продемонстрировал результаты выполнения практической работы или не представил по ней отчет или представленный отчет не соответствует требованиям по оформлению.

7.3. Оценочные средства для промежуточной аттестации по итогам освоения дисциплины

7.3.1. Вопросы к промежуточной аттестации

Семестр 3

№ п/п	Вопросы
1	Концептуальное понятие о вычислительных системах.
2	Понятие о вычислительных системах.
3	Типы архитектур: MISD, SIMD, MIMD. Классификация ВС
4	Принципы построения вычислительных систем.
5	Структура ВС: типовые структуры сетей межвычислительных связей
6	История вычислений на GPU.
7	Отличия архитектуры GPU от архитектуры CPU.
8	Гетерогенные параллельные архитектуры.
9	Преимущества CUDA, место CUDA в современном HPC.
10	Утилизация латентности памяти в CUDA.
11	Составные части программы с использованием CUDA.
12	Nvidia SIMT. Грид, блок, варп, дивергентное выполнение.
13	Аппаратная и виртуальная масштабируемость CUDA.
14	Кеширование глобальной памяти.
15	Определение транзакции. Шаблоны доступа.
16	Набивка (padding) матриц и раздельное хранение полей структур
17	Концепция общей памяти в CUDA.
18	Синхронизации, стратегия использования.
19	Банки общей памяти.
20	Различия между Fermi и Kepler. Примеры банк-конфликтов
21	Константная память и однородные обращения.
22	Примеры параллельных алгоритмов, допускающих использование константной памяти.
23	Регистры и локальная память.
24	Назначение локальной памяти и случаи её использования.
25	Расчет и оптимизация Occupancy.
26	Средства профилирования в CUDA.
27	События и метрики. Основные метрики.
28	Удаленное профилирование.
29	Средства отладки в CUDA.
30	Быстрый поиск ошибок использования API и некорректных обращений в память.
31	Pinned-память.
32	Неблокирующие копирования, обращения в память CPU напрямую из ядер
33	UVA. Примеры упрощения кода при использовании UVA.
34	Концепция потоков.
35	Примеры выгодного и невыгодного использования потоков.

36	CUDA-контекст. Схемы использования нескольких GPU: с одной CPU-нитью, с OpenMP, с MPI.
37	Peer-to-peer копирования между GPU.
38	Прямые обращения в память другого устройства.
39	Схема работы NVCC.
40	Совместимость, виртуальные и физические архитектуры.
41	Двухэтапная компиляция с промежуточным представлением
42	Неявная JIT-компиляция.
43	Линковка кода для графических ускорителей NVIDIA.
44	Низкоуровневый анализ программ.
45	CUDA driver API, явная работа с объектными модулями
46	Явная JIT-компиляция.
47	Высокопроизводительные библиотеки CUDA.
48	Основы использования Thrust.
49	Параллельные алгоритмы.
50	Элементарные понятия параллельного программирования.
51	Параллельный алгоритм умножения матриц.
52	Показатели эффективности параллельных алгоритмов: коэффициенты накладных расходов, ускорения и эффективности.
53	Парадокс параллелизма.
54	Понятие о сложных задачах.
55	Схемы обмена информацией между ветвями параллельных алгоритмов.
56	Методика крупноблочного распараллеливания сложных задач.

7.3.2. Критерии и нормы оценки

Семестр	Форма проведения промежуточной аттестации	Критерии и нормы оценки	
3	Зачет (устно)	«зачтено»	Студент понимает основные положения технологии разработки алгоритмов на платформе CUDA
		«не зачтено»	Студент не знает <i>наиболее важные</i> определения, не может ориентироваться в материале
3	Пересдача	«зачтено»	Студент понимает основные положения технологии разработки алгоритмов на платформе CUDA
		«не зачтено»	Студент не знает <i>наиболее важные</i> определения, не может ориентироваться в материале

8. Учебно-методическое и информационное обеспечение дисциплины

8.1. Обязательная литература

№ п/п	Авторы, составители	Заглавие (заголовок)	Тип (учебник, учебное пособие, учебно-методическое пособие, практикум, др.)	Год издания	Количество в научной библиотеке / Наименование ЭБС
1	Гергель В. П.	Теория и практика параллельных вычислений [Электронный ресурс] : [учебное пособие] / В. П. Гергель. - 2-е изд., испр. - Москва : ИНТУИТ, 2016. - 480 с. : ил. - (Основы информационных технологий). - ISBN 978-5-94774-645-7.	учебное пособие	2016	ЭБС «IPRbooks»

8.2. Дополнительная литература

№ п/п	Авторы, составители	Заглавие (заголовок)	Тип (учебник, учебное пособие, учебно-методическое пособие, практикум, др.)	Год издания	Количество в научной библиотеке / Наименование ЭБС
1	Федотов И. Е.	Приемы параллельного программирования [Электронный ресурс] : учеб. пособие по спец. 010503 - "Математическое обеспечение и администрирование информационных систем" / И. Е. Федотов. - Москва : РосНОУ, 2009. - 184 с. : ил. - ISBN 978-5-89789-048-4.	учебное пособие	2009	ЭБС «IPRbooks»

8.3. Перечень профессиональных баз данных и информационных справочных систем

- Springer International Publishing, Part of Springer Science+Business Media
- NVIDIA GPU Educators Program. [Electronic resource] : [Образовательные программы NVIDIA GPU] . – Electronic data. [2018]. – Mode of access : <https://developer.nvidia.com/educators>
- The NVIDIA® CUDA® Toolkit. [Electronic resource] : [Инструментарий NVIDIA® CUDA®] . – Electronic data. [2018]. – Mode of access : <https://developer.nvidia.com/cuda-toolkit>
- Курс лекций по CUDA. [Электронный ресурс] . – Электронные данные. [2018]. – Режим доступа : http://www.nvidia.ru/object/cuda_state_university_courses_new_ru.html

8.4. Перечень программного обеспечения

№ п/п	Наименование ПО	Количество лицензий	Реквизиты договора (дата, номер, срок действия)
1	CUDA Toolkit	Не ограничено	Свободно ПО

8.5. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

№ п/п	Наименование оборудованных учебных кабинетов, лабораторий, мастерских и др. объектов для проведения практических и лабораторных занятий	Перечень основного оборудования
1.	Компьютерный класс. Учебная аудитория для проведения занятий лекционного типа. Учебная аудитория для проведения занятий семинарского типа. Учебная аудитория для проведения лабораторных работ. Учебная аудитория для курсового проектирования (выполнения курсовых работ). Учебная аудитория для проведения групповых и индивидуальных консультаций Учебная аудитория для проведения занятий текущего контроля и промежуточной аттестации.(УЛК-312)	Столы ученические, переносной проектор, экран, стол преподавательский, стулья, доска аудиторная (маркерная), ПК с выходом в сеть Интернет
2.	Учебная аудитория для проведения занятий лекционного типа. Учебная аудитория для проведения занятий семинарского типа. Учебная аудитория для курсового проектирования (выполнения курсовых работ). Учебная аудитория для проведения групповых и индивидуальных консультаций. Учебная аудитория для проведения занятий текущего контроля и промежуточной аттестации. (УЛК-418)	Столы ученические двухместные (моноблок), доска аудиторная 3-х секционная (меловая), стол преподавательский , стулья, проектор Acer

3.	Компьютерный класс. Помещение для самостоятельной работы. Учебная аудитория для проведения занятий семинарского типа. Учебная аудитория для курсового проектирования (выполнения курсовых работ). Учебная аудитория для проведения групповых и индивидуальных консультаций. Учебная аудитория для проведения занятий текущего контроля и промежуточной аттестации. (Г-401)	Столы ученические, стулья ученические, ПК с выходом в сеть Интернет
----	--	---