

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЗАХІДНОУКРАЇНСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ЄРЕВАНСЬКИЙ НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ**

ЗАТВЕРДЖУЮ

Директор Єреванського навчально-
наукового інституту ЗУНУ

Саак ГУДРАТЯН
_____ 2024 р.



ЗАТВЕРДЖУЮ

Проректор з науково-педагогічної роботи

Віктор ОСТРОВЕРХОВ
_____ 2024 р.



РОБОЧА ПРОГРАМА

з дисципліни

«Проектування комп'ютерних систем на основі графічних процесорів»

ступінь вищої освіти – бакалавр

галузь знань - 12 - Інформаційні технології

спеціальність - 123 - Комп'ютерна інженерія

освітньо-професійна програма - «Комп'ютерна інженерія»

Кафедра фундаментальних дисциплін

Форма навчання	Курс	Семестр	Лекції (год.)	Лаб. (год.)	ІРС (год.)	Тренінг (год.)	Самост. робота студ. (год.)	Разом (год.)	Залік (сем.)	Екз. (сем.)
Заочна	4	7	8	4	-	-	138	150	8	-

Робочу програму розробив к.тех.н., доцент кафедри фундаментальних дисциплін
ЄННІ ЗУНУ Армен МУРАДЯН.



Робоча програма затверджена на засіданні кафедри фундаментальних дисциплін,
протокол № 1 від 27.08.2024 року.

Завідувач кафедри
фундаментальних дисциплін,
доктор технічних наук, професор



Гамлет АРУТЮНЯН

Розглянуто та схвалено групою забезпечення спеціальності 123 «Комп'ютерна
інженерія», протокол № 1 від 30.08.2024 р.

Голова ГЗС,
д.т.н., професор



Гамлет АРУТЮНЯН

Гарант
д.т.н., професор



Гамлет АРУТЮНЯН

СТРУКТУРА РОБОЧОЇ ПРОГРАМИ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ "ПРОЕКТУВАННЯ КОМП'ЮТЕРНИХ СИСТЕМ НА ОСНОВІ ГРАФІЧНИХ ПРОЦЕСОРІВ"

1. Опис дисципліни «Проектування комп'ютерних систем на основі графічних процесорів»

Дисципліна «Проектування комп'ютерних систем на основі графічних процесорів»	Галузь знань, спеціальність, ступінь вищої освіти	Характеристика навчальної дисципліни
Кількість кредитів – 5	Галузь знань – 12 „Інформаційні технології”	Статус дисципліни – вибіркова Мова навчання – українська
	Спеціальність – 123 „Комп'ютерна інженерія”	Рік підготовки: 4 Семестр: 7
Кількість змістових модулів – 4	Ступінь вищої освіти - бакалавр	Лекції: 8 год. Практичні заняття: 4 год.
Загальна кількість годин – 150 год.		Самостійна робота: 138 год.
		Вид підсумкового контролю: залік

2. Мета й завдання дисципліни

2.1. Мета вивчення дисципліни

Метою курсу «Проектування комп'ютерних систем на основі графічних процесорів» є отримання знань та навиків розробки програмних систем, які призначені для обробки з допомогою графічних процесорів.

В процесі роботи студенти отримають знання щодо принципів обробки даних з допомогою центрального та графічного процесорів для розуміння принципу роботи сучасних систем. Додатково студенти отримують знання щодо підходів до розробки розпаралелених алгоритмів та отримають практичні навички для програмної реалізації.

2.2 Завдання вивчення дисципліни

Завданням дисципліни визначається тим, щоб дати студентам теоретичну та практичну підготовку із проектування, розробки та програмної реалізації програмних систем з допомогою графічних процесорів.

2.3. Знання та вміння, формування котрих забезпечує вивчення дисципліни:

В процесі роботи студенти отримають знання щодо принципів обробки даних з допомогою центрального та графічного процесорів для розуміння принципу роботи сучасних систем. Додатково студенти отримують знання щодо підходів до розробки розпаралелених алгоритмів та отримають практичні навички для програмної реалізації.

3. Програма навчальної дисципліни «Проектування комп'ютерних систем на основі графічних процесорів»

Змістовий модуль 1. Вступ до складності обчислювальних задач і технології CUDA

Тема 1. Вступ. Поняття складності обчислювальних задач

Класифікація обчислювальних задач. Приклади складних обчислювальних задач. Визначення обчислювальної складності задач. Часові та просторові обмеження.

Тема 2. Підходи до декомпозиції складних задач

Метод декомпозиції задач. Розбиття задач на підзадачі. Принципи модульного підходу. Визначення граничних умов.

Тема 3. Підходи до виконання складних програм

Паралельні та послідовні програми. Стратегії ефективного виконання програм. Використання багатоядерних систем. Особливості багатопотокових систем.

Тема 4. Технологія CUDA: Вступ і архітектура

Вступ до CUDA. Архітектура CUDA. Роль графічних процесорів у паралельних обчисленнях. Переваги використання GPU.

Тема 5. Вступ до моделей паралельних обчислень

Принципи паралельних обчислень. Загальна архітектура паралельних систем. Основні поняття багатопроесорних систем. Переваги та обмеження паралелізму.

Змістовий модуль 2. Паралельні обчислення та багатопоточність

Тема 6. Види розпаралелення обчислювальних задач

Поняття паралелізму. Рівні розпаралелення: блоки, потоки. Типи паралельних задач. Масштабування обчислювальних процесів.

Тема 7. Синхронізація, потоки та процеси

Визначення потоків та процесів. Основні підходи до синхронізації. Засоби для синхронізації процесів. Приклади використання потоків.

Тема 8. Складність обчислень паралельних задач

Оцінка складності паралельних алгоритмів. Часові та ресурсні обмеження. Співвідношення між кількістю процесів і швидкістю виконання. Блокування та конкуренція потоків.

Тема 9. Багатопоточність: основи та бібліотеки

Моделі багатопоточності. Огляд бібліотек для багатопоточних систем. Взаємодія між потоками в багатопотокових програмах. Приклади

багатопоточних програм.

Тема 10. Паралелізм на рівні даних

Основи паралельних обчислень з використанням даних. Паралельні структури даних. Приклади паралелізму на рівні даних. Оптимізація розподілу даних.

Змістовий модуль 3. Основи технології CUDA

Тема 11. CUDA: Ядро та функції

Ядро CUDA. Оголошення функцій у CUDA. Взаємодія хоста і девайса. Стратегії для ефективної роботи з функціями.

Тема 12. CUDA: Потoki та типи пам'яті

Організація потоків у CUDA. Різні типи пам'яті в CUDA. Приклади роботи з потоками. Оптимізація використання пам'яті в GPU.

Тема 13. CUDA: Блоки та потоки

Структура блоків і потоків у CUDA. Програмування ідентифікаторів блоків і потоків. Взаємодія між блоками. Паралелізм на рівні блоків.

Тема 14. CUDA: Атомарні операції та синхронізація

Основи атомарних операцій у CUDA. Важливість синхронізації потоків. Підходи до синхронізації на рівні GPU. Приклади застосування атомарних операцій.

Тема 15. CUDA: Оптимізація обчислень

Методи оптимізації програм на базі CUDA. Визначення вузьких місць в обчислювальних процесах. Стратегії оптимізації пам'яті. Приклади практичної оптимізації.

Змістовий модуль 4. Практичне використання технології CUDA

Тема 16. CUDA: Обробка зображень

Методи обробки зображень на GPU. Впровадження фільтрів і алгоритмів обробки. Прискорення обробки зображень за допомогою CUDA. Приклади з реальних задач.

Тема 17. Алгоритми сортування даних з допомогою CUDA

Типи алгоритмів сортування. Використання CUDA для швидкого сортування. Оптимізація процесу сортування на GPU. Приклади практичних задач.

Тема 18. CUDA: Робота з нейронними мережами

Нейронні мережі та GPU. Використання CUDA для навчання нейронних мереж. Прискорення обчислень у нейронних мережах за допомогою GPU. Огляд бібліотек для нейронних мереж.

Тема 19. CUDA: Використання для обробки великих даних

Особливості паралельної обробки великих обсягів даних на GPU. Оптимізація

завдань великої вибірки з використанням CUDA. Методи зберігання та доступу до великих даних у CUDA. Приклади обробки великих даних у реальних проектах.

Тема 20. Практичне застосування CUDA у задачах штучного інтелекту

Впровадження обчислень для AI задач на базі CUDA. Оптимізація нейронних мереж для GPU. Розробка програмного забезпечення для AI з використанням GPU. Використання CUDA у реальних AI задачах.

4. Структура залікового кредиту дисципліни «Проектування комп'ютерних систем на основі графічних процесорів»

(заочна форма навчання)

	Кількість годин		
	Лекції	Лабораторні заняття	Самостійна робота
Змістовий модуль 1			
Тема 1. Вступ. Поняття складності обчислювальних задач	1		10
Тема 2. Підходи до декомпозиції складних задач			6
Тема 3. Підходи до виконання складних програм			6
Тема 4. Технологія CUDA: Вступ і архітектура	1	1	10
Тема 5. Вступ до моделей паралельних обчислень			6
Змістовий модуль 2			
Тема 6. Види розпаралелення обчислювальних задач	1		10
Тема 7. Синхронізація, потоки та процеси	1		10
Тема 8. Складність обчислень паралельних задач			6
Тема 9. Багатопоточність: основи та бібліотеки			8
Тема 10. Паралелізм на рівні даних			6
Змістовий модуль 3			
Тема 11. CUDA: Ядро та функції	1	1	6
Тема 12. CUDA: Потоки та типи пам'яті			6
Тема 13. CUDA: Блоки та потоки			6
Тема 14. CUDA: Атомарні операції та синхронізація	1	1	6
Тема 15. CUDA: Оптимізація обчислень			6
Змістовий модуль 4			
Тема 16. CUDA: Обробка зображень			6
Тема 17. Алгоритми сортування даних з допомогою CUDA	1		6
Тема 18. CUDA: Робота з нейронними мережами			6
Тема 19. CUDA: Використання для обробки великих даних	1	1	6

Тема 20. Практичне застосування CUDA у задачах штучного інтелекту			6
Разом	8	4	138

Тематика практичних занять

Практична робота №1.

Тема: CUDA: Інсталювання та основи технології.

Мета: Ознайомитись інструментами для запуску задач під CUDA.

Питання для обговорення:

1. Синтаксис Python для запуску CUDA задач
2. запуск задач у хмарному середовищі
3. запуск задач в локальному середовищі

Практична робота №2.

Тема: CUDA: Ядра та компілювання програм.

Мета: Навчитись писати та запусками власні програми з допомогою CUDA

Питання для обговорення:

1. Функції для роботи зі стрічками
2. Функції для роботи з масивами
3. Функціональний підхід в програмуванні

Практична робота №3.

Тема: CUDA: Потоки та типи пам'яті.

Мета: Ознайомитись із роботою з потоками.

Питання для обговорення:

1. Поняття потоків
2. Типи пам'яті
3. Патери проектування

Практична робота №4.

Тема: CUDA: Атомарні операції та синхронізація потоків.

Мета: Набуття практичних навичок забезпечення синхронізації потоків в CUDA.

Питання для обговорення:

1. синхронізація потоків
2. види потоків
3. основні задачі синхронізації.

5. Самостійна робота студентів (заочна форма навчання)

№ п/п	Тематика
1	Класифікація обчислювальних кластерів
2	Грид та хмарні обчислення
3	CUDA: Атомарні операції та синхронізація потоків
4	Робота з нейронними мережами на основі технології CUDA
5	CUDA: Потоки та типи пам'яті
6	CUDA: Робота з блоками та потоками
7	Архітектури високопродуктивних обчислювальних систем
8	Класифікація обчислювальних кластерів
9	CUDA: Атомарні операції та синхронізація потоків
10	CUDA: Ядра та компілювання програм
11	CUDA: Обробка зображень
12	Порівняння CPU та GPU для обчислювальних задач
13	Методи оптимізації використання пам'яті в GPU
14	Моделі багатопоточності в графічних процесорах
15	Оптимізація алгоритмів сортування на GPU
16	Обробка великих даних за допомогою графічних процесорів

6. Засоби оцінювання та методи демонстрування результатів навчання

У навчальному процесі застосовуються: лекції, в тому числі з використанням мультимедіа проектора та інших ТЗН; практичні заняття; індивідуальні заняття, самостійна робота студента, робота в мережі Інтернет.

У процесі вивчення дисципліни "Проектування Інтернет додатків для комп'ютерної інженерії" використовуються наступні засоби оцінювання та методи демонстрування результатів навчання:

- поточні опитування;
- презентації результатів виконання завдань та досліджень;
- оцінювання результатів самостійної роботи;
- залік.

7. Критерії, форми поточного та підсумкового контролю

Підсумковий бал (за 100-бальною шкалою) з дисципліни "Проектування комп'ютерних систем на основі графічних процесорів" визначається як середньозважена величина, в залежності від питомої ваги кожної складової залікового кредиту:

В процесі вивчення дисципліни "Проектування комп'ютерних систем на основі графічних процесорів" використовуються наступні методи оцінювання навчальної роботи студента:

- поточне тестування та опитування;
- оцінювання наскрізного проекту у результаті самостійної роботи;
- залік.

Шкала оцінювання:

За шкалою університету	За національною шкалою	За шкалою ECTS
90-100	Відмінно	A (відмінно)
85-89	Добре	B (дуже добре)
75-84		C (добре)
65-74	Задовільно	D (задовільно)
60-64		E (достатньо)
35-59	Незадовільно	FX (незадовільно, з можливістю повторного складання)
1-34		F (незадовільно, з обов'язковим повторним курсом)

8. Інструменти, обладнання та програмне забезпечення, використання яких передбачає навчальна дисципліна

№	Найменування	Номер теми
1.	Google Colaboratory, Python	11-20

Рекомендовані джерела інформації

1. Березький О.М. Методи, алгоритми та програмні засоби опрацювання біомедичних зображень: монографія / [О.М. Березький, Ю.М. Батько, К.М. Березька, С.О. Вербовий, Т.В. Дацко, Л.О. Дубчак, І.В. Ігнатєв, Г.М. Мельник, В.Д. Ніколюк, О.Й. Піцун]; під наук. ред. Березький О.М., Тернопіль . ТНЕУ «Економічна думка», 2017. - 330 с
2. Berezsky O. GPU – based biomedical image processing / O. Berezsky, O. Pitsun, L. Dubchak, P. Lyaschynsky, P. Lyaschynsky // Proceedings of XIV International Conference Perspective Technologies and methods in mems design (MEMSTECH 2018) 18-22 April, 2018, Lviv-Polyana, Ukraine, pp. 96- 99
3. Ryan N Gutenkunst, dadi.CUDA: Accelerating Population Genetics Inference with Graphics Processing Units, Molecular Biology and Evolution, Volume 38, Issue 5, May 2021, Pages 2177–2178.
4. Duane Storti. CUDA for Engineers: An Introduction to High-Performance Parallel Computing 1st Edition/ Addison-Wesley Professional; 1st edition (November 2, 2015) – 452 p.
5. Xu, Derong, Y. Luo, C. Montag, Y. Hao, and J. Qiang. Full Range Tune Scan Studies Using Graphics Processing Units With CUDA in EIC beam-beam simulations. No. BNL-222292-2021-COPA. Brookhaven National Lab.(BNL), Upton, NY (United States), 2021..
6. Brian Tuomanen Hands-On GPU Programming with Python and CUDA: Explore high- performance parallel computing with CUDA 1st Edition, Kindle Edition – 2018
7. Jaegeun Han, Bharatkumar Sharma, Learn CUDA Programming - Packt Publishing (September 27, 2019) – 508p.
8. Bhaumik Vaidya. Hands-On GPU-Accelerated Computer Vision with OpenCV and CUDA: Effective techniques for processing complex image data in real time using

GPUs Kindle Edition - Packt Publishing; 1st edition (September 26, 2018) – 382 p.

9. Pitsun O. Multi-threaded Parallelization of Automatic Immunohistochemical Image Segmentation. In: Hu, Z., Wang, Y., He, M. (eds) Advances in Intelligent Systems, Computer Science and Digital Economics IV. CSDEIS 2022. Lecture Notes on Data Engineering and Communications Technologies, vol 158. Springer, Cham

10. Aslam, Maria, Omer Riaz, Shahzad Mumtaz, and Ali Daniyal Asif. "Performance comparison of GPU-based jacobi solvers using CUDA provided synchronization methods." IEEE Access 8 (2020): 31792-31812.