

# GPGPU Programming

Fall-2018

## Instructor Information

### Instructor

Dr. Ayaz ul Hassan Khan

### Email

[ayazhk@gmail.com](mailto:ayazhk@gmail.com)

### Office Location & Hours

TBA

## General Information

### Description

Learn how to program heterogeneous parallel computing systems and achieve: high performance and energy-efficiency, functionality and maintainability, scalability across future generations, parallel programming API, tools and techniques, principles and patterns of parallel algorithms, processor architecture features and constraints

### Expectations and Goals

- List the major difference between latency devices (CPU cores) and throughput devices (GPU cores)
- State the importance and nature of scalability and portability in parallel programming
- Recognize the main venues and developer resources for GPU computing, write CUDA programs
- Estimate memory bandwidth requirements, analyze the problem and exploit data parallelism
- Evaluate and analyze the implementation of parallel numerical algorithms
- Evaluate some valuable tools and resources from the CUDA toolkit
- Recognize the effectiveness of related parallel programming model: OpenACC

## Course Materials

### Required Text

- Programming Massively Parallel Processors: A Hands-on Approach, 2nd Edition by David Kirk and Wen-Mei Hwu
- CUDA by Example: An Introduction to General - Purpose GPU Programming, Jason Sanders and Edward Kandrot

### Reference Text

- CUDA C Programming Guide, NVIDIA
- <https://nvidia.gwlab.com/>
- [www.openacc.org](http://www.openacc.org), [http://www.nvidia.com/object/cuda\\_home\\_new.html](http://www.nvidia.com/object/cuda_home_new.html)
- Lecture Notes, Handouts, and Selected Research Papers

## Pre-Requisites: Computer Programming, Computer Architecture

### Brief List of Topics to be covered:

#### Topic

---

*Introduction to Heterogeneous Parallel Computing*

---

*Introduction to CUDA C*

---

*CUDA Parallelism Model*

---

*Memory Model and Locality*

---

*Kernel - Based Parallel Programming*

---

*Performance Considerations: Memory*

---

*Atomic Operations*

---

*Parallel Computation Patterns: Stencil, Reduction, Scan*

---

*Related Programming Models: MPI, OpenACC*

---

### About Course Instructor:

Dr. Ayaz ul Hassan Khan received his BS degree from NED-Pakistan, MS degree in Computer Science from LUMS-Pakistan and PhD degree in Computer Science and Engineering with the specialization in Parallel Computing from King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia. He has seven years of experience in both industry and academics. He has hands-on experience in Database programming, Network Programming, Embedded Systems Programming, and Parallel Programming. His current areas of interest include Parallel and Distributed Computing, High Performance Computing, Computer Architecture, Operating Systems, Deep Learning and Big Data Analytics. He has published 6 journals and 11 conference papers/posters in the field of his research areas in recent years. Check out the following for details:

<https://sites.google.com/site/ayazresearch/>

### Related Links:

- <https://www.nvidia.com/en-us/gtc/>
- <https://sc-asia.org/>
- <http://www.supercomp.org/>
- <https://link.springer.com/journal/10766>
- <https://www.journals.elsevier.com/parallel-computing/>
- <https://developer.nvidia.com/cuda-zone>
- <https://developer.nvidia.com/cuda-toolkit>
- <https://www.openacc.org/>
- <https://www.open-mpi.org/>