

# Homework : High Performance Computing in Earth Observation and Remote Sensing, HPC

111C51502, CY Chingyao Fu, NTUT

Oct 2023

## Introduction

這是圖像識別Pattern Recognition第三節課的課後作業, 主要是關於高效能運算HPC領域的一些知識和常見的字匯解釋, 包含CPU/GPU, CUDA, 記憶體架構, 高光譜影像Hyperspectral Image 和多孔徑雷達SAR等。

## Reference

- Programming Massively Parallel Processors, A Hands-on Approach  
David B. Kirk, NVIDIA Corporation  
Wen-mei W. Hwu  
Third Edition, Year 2017
- Computer Organization Design, THE HARDWARE/SOFTWARE INTERFACE  
David A. Patterson, University of California, Berkeley  
John L. Hennessy, Stanford University  
Year 2005
- Hyperspectral Data Processing: Algorithm Design and Analysis  
Chein-I Chang, University of Maryland, Baltimore County(UMBC)  
Year 2013
- Digital Processing of Synthetic Aperture Radar Data  
Ian G. Cumming, University of British Columbia  
Frank H. Wong, University of British Columbia  
Year 2005
- Practice of Analyzing EO-1 Hyperion Multi-spectral Satellite Images Using PCA  
CY Chingyao Fu, AI EMBA Program, NTUT  
Year 2023

## 1 CPU and GPU

- CPU(Central Processing Unit, 中央處理單元)
  - The design of a CPU is optimized for sequential code performance. It makes use of sophisticated control logic to allow instructions from a single thread to execute in parallel or even out of their sequential order while maintaining the appearance of sequential execution.

- The CPUs, are designed to minimize the execution latency of a single thread. Large last-level on-chip caches are designed to capture frequently accessed data and convert some of the long-latency memory accesses into short-latency cache accesses. The arithmetic units and operand data delivery logic are also designed to minimize the effective latency of operation at the cost of increased use of chip area and power.
- Memory bandwidth is another important issue. The speed of many applications is limited by the rate at which data can be delivered from the memory system into the processors.
- GPU(Graphics Processing Unit, 圖像處理單元)
  - The design philosophy of the GPUs has been shaped by the fast growing video game industry that exerts tremendous economic pressure for the ability to perform a massive number of floating-point calculations per video frame in advanced games. This demand motivates GPU vendors to look for ways to maximize the chip area and power budget dedicated to floating-point calculations.
  - It should be clear now that GPUs are designed as parallel, throughput-oriented computing engines and they will not perform well on some tasks on which CPUs are designed to perform well.
  - For programs that have one or very few threads, CPUs with lower operation latencies can achieve much higher performance than GPUs. When a program has a large number of threads, GPUs with higher execution throughput can achieve much higher performance than CPUs.

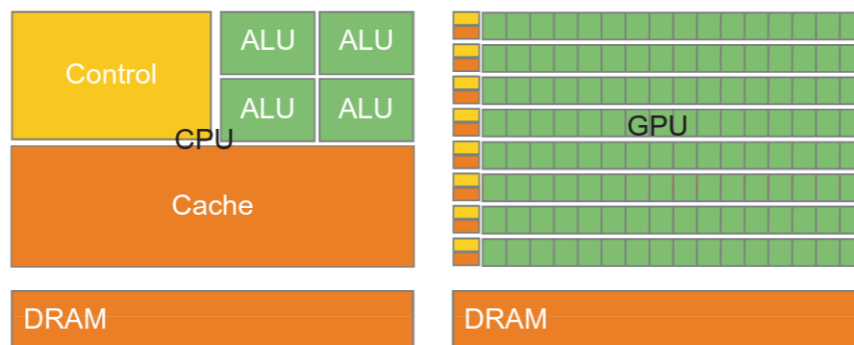


Figure 1: CPUs and GPUs have fundamentally different design philosophies

## 2 CUDA

- CUDA(Compute Unified Device Architecture)
  - CUDA is currently the most mature framework for massively parallel computing. It is broadly used in the high-performance computing industry, with sophisticated tools such as compilers, debuggers, and profilers available on the most common operating systems.
  - An important point: while our examples will mostly use CUDA C for its simplicity and ubiquity, the CUDA platform supports many languages and application programming interfaces (APIs) including C++, Python, Fortran, OpenCL, OpenACC, OpenMP, and more.
  - CUDA is really an architecture that supports a set of concepts for organizing and expressing massively parallel computation.

### 3 Memory Hierarchy

- Structure
  - The memory system is organized as a hierarchy: a level closer to the processor is generally a subset of any level further away, and all the data is stored at the lowest level.
  - Accesses that hit in the highest level of the hierarchy can be processed quickly. Accesses that miss go to lower levels of the hierarchy, which are larger but slower.

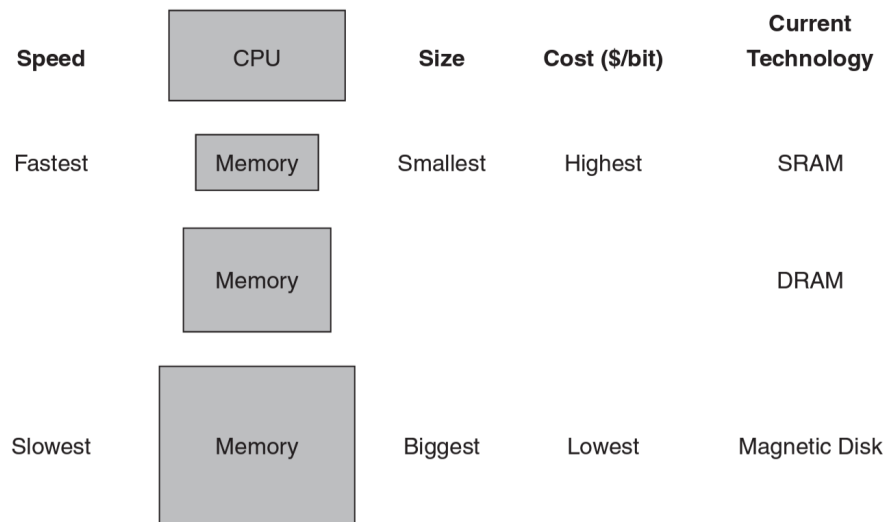


Figure 2: The basic structure of a memory hierarchy

- Performance
  - Memory hierarchies take advantage of temporal locality by keeping more recently accessed data items closer to the processor. Memory hierarchies take advantage of spatial locality by moving blocks consisting of multiple contiguous words in memory to upper levels of the hierarchy.
  - If the hit rate is high enough, the memory hierarchy has an effective access time close to that of the highest (and fastest) level and a size equal to that of the lowest (and largest) level.
- Implement and Challenge
  - Main memory is implemented from DRAM (dynamic random access memory), while levels closer to the processor (caches) use SRAM (static random access memory).
  - The challenge in designing memory hierarchies to close this growing gap. All the hardware design choices for memory hierarchies have both a positive and negative effect on performance. This means that for each level of the hierarchy there is an optimal performance point per program, which must include some misses.
  - How can we overcome the growing gap between processor speeds and lower levels of the hierarchy? This question is currently the topic of much research.

### 4 Hyperspectral Image

- Hyperspectral Image(高光譜影像)

- 高光譜影像是由數百個連續的光譜通道採集的，具有非常高的光譜解析度。這與多光譜影像(Multispectral Image)不同，後者通常只由數十個離散的光譜通道/帶採集，並具有較低的光譜解析度。
  - 高光譜影像是由數百個連續的光譜通道提供的光譜特性。許多感興趣的物質只能通過它們的光譜特性來探索，而不是3D空間特性。
  - 由於高光譜傳感器提供的非常高的光譜解析度，許多通常未知或無法通過視覺評估識別的微妙物質現在可以用於數據分析。這些應用可能包括農業、生態學、地質學、環境監測、執法、軍事和醫學診斷等。
- 在PR\_HW02\_C已練習過高光譜影像主成分分析PCA實作。其中取用2008年6月拍攝到台灣雪山國家公園附近的EO-1衛星242頻段的高光譜影像。參考: Practice of Analyzing EO-1 Hyperion Multi-spectral Satellite Images Using PCA

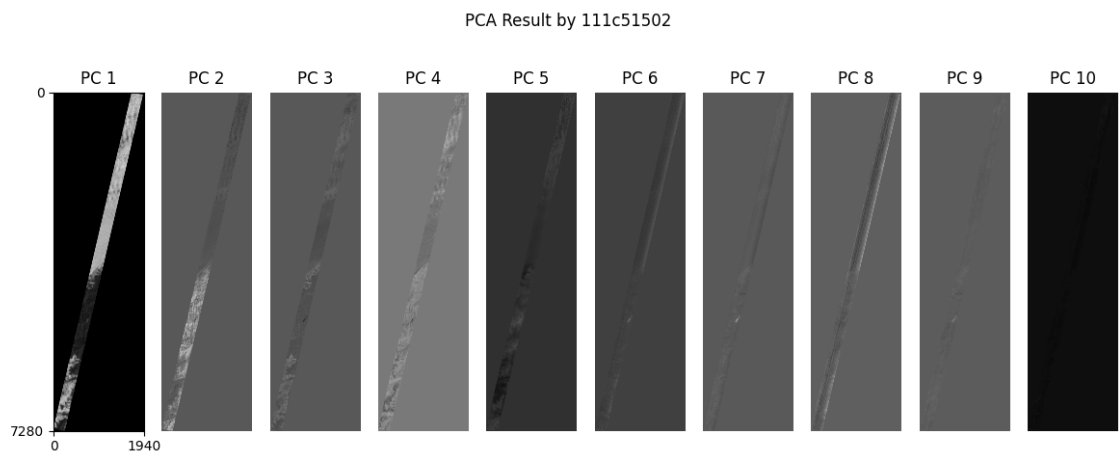


Figure 3: PCA Result

## 5 SAR

- Synthetic Aperture Radar(合成孔徑雷達, SAR)
  - Synthetic Aperture Radar (SAR) was originally developed for military purposes during World War II. Its initial purpose was to track aircraft and ships through severe weather conditions and darkness. With advancements in radio frequency (RF) technology, antennas, and more recently, digital technology, SAR has seen steady growth.
- Fundamental Principles
  - SAR measures the range to a target (known as a radar scatterer) via time delay and the direction of a target via antenna directivity.
  - Doppler shifts are used to measure the speed of a target.
  - By processing these Doppler shifts, fine resolution can be obtained in a direction perpendicular to the range or beam direction.
  - This latter concept, often credited to Carl Wiley of Goodyear Aerospace in 1951, led to the discovery that two-dimensional images could be made of the targets and of the Earth's surface.
- Applications and Development

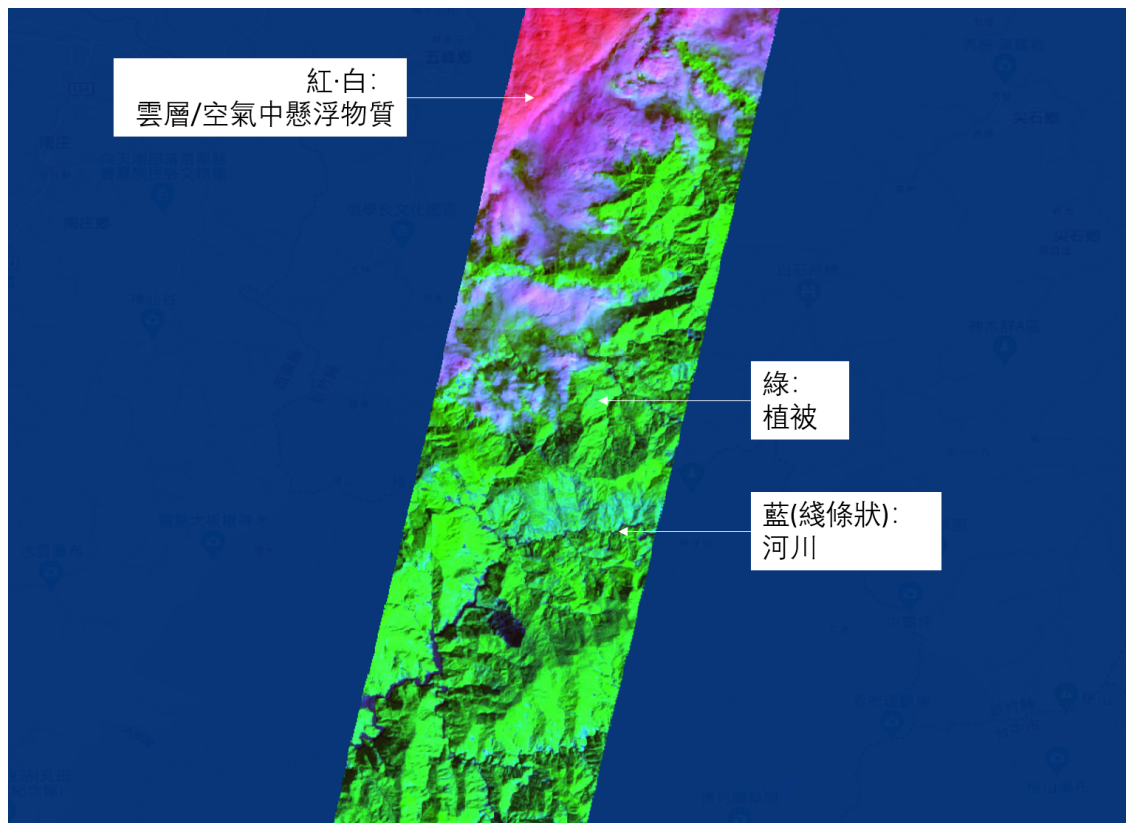


Figure 4: PCA Stacked Image

- SAR has been prominent in the remote sensing community, and there are various main SAR processing algorithms used in satellite remote sensing.
- In addition to basic SAR processing algorithms, there are algorithms for determining the main Doppler parameters, the Doppler centroid, and the azimuth FM rate.

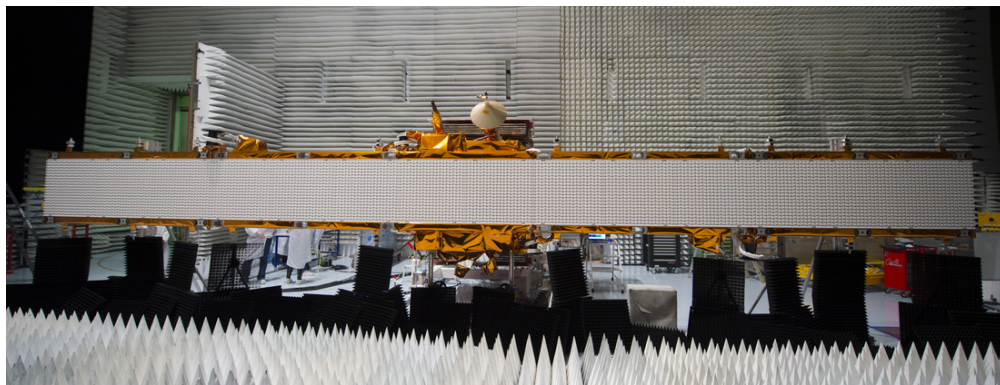


Figure 5: C-Band Synthetic Aperture Radar, or C-SAR, installed on Sentinel-1