ECE-6913 Computer System Architecture
Homework Extra Credit

New York University, Fall 2025

Due on 12/19, 11:59 PM

Name: Raman Kumar Jha
NYU ID: N13866145

Extra Credit: Top 5 AI Hardware Corporations Ranking

Ranking Strategy

The following ranking is prioritized based on three key factors: **current market dominance** (software/hardware ecosystem), **architectural innovation** (solving bottlenecks like memory bandwidth or latency), and **scalability** (data center vs. edge deployment).

1. **Nvidia**

2. **AMD (Advanced Micro Devices)**

3. **Broadcom**

4. **Google**

5. **Qualcomm**

Justifications

1. Nvidia

Justification for Inclusion: Nvidia is currently the undisputed leader in AI hardware. Their Graphics Processing Units (GPUs), such as the H100 and the upcoming Blackwell architecture, are the industry standard for training massive Large Language Models (LLMs). Crucially, their "moat" is not just hardware but software; the CUDA ecosystem is deeply entrenched in AI research and development.
Ranking Justification (#1): They are ranked first because they effectively hold a monopoly on AI training hardware today. For the next 5 years, legacy code compatibility and their aggressive innovation cycle ensure they will remain the default choice for the majority of data centers.

2. AMD

Justification for Inclusion: AMD is the primary direct competitor to Nvidia in the high-performance GPU space. Their MI300 series accelerators offer competitive memory bandwidth and capacity, which are critical for running large models.
Ranking Justification (#2): They are ranked second because the tech industry is desperate for a "second source" to reduce reliance on Nvidia. As AMD improves its software stack (ROCm), it is best positioned to capture significant market share from hyperscalers looking for cost-effective alternatives to Nvidia.

3. Broadcom

Justification for Inclusion: While less visible than GPU makers, Broadcom is the backbone of AI infrastructure. They specialize in high-speed networking (Ethernet/PCIe switches) required to connect thousands of GPUs together. Furthermore, they are the design partner for "Custom Silicon" (ASICs), helping companies like Google create their TPUs and Meta create their MTIA chips.

Ranking Justification (#3): As AI models grow, the bottleneck shifts from the chip itself to the network connecting the chips. Broadcom's dominance in networking and custom ASIC design secures it a critical, high-volume role in the future of AI clusters.

4. Google

Justification for Inclusion: Google was an early pioneer with its Tensor Processing Unit (TPU). Unlike general-purpose GPUs, TPUs are Application-Specific Integrated Circuits (ASICs) designed specifically for tensor operations, offering high efficiency for Google's massive internal workloads (Search, Gemini, YouTube).

Ranking Justification (#4): They are ranked fourth because they represent the successful "vertical integration" model. While they don't sell chips openly like Nvidia, the sheer scale of their infrastructure and the maturity of their TPU architecture (now in its 5th/6th generation) make them a powerhouse in AI hardware prospects.

5. Qualcomm

Justification for Inclusion: While the companies above focus on Data Centers (Cloud AI), Qualcomm is the leader in "Edge AI." Their Snapdragon platforms with integrated NPUs (Neural Processing Units) are enabling AI models to run locally on smartphones, laptops, and cars without needing an internet connection.

Ranking Justification (#5): They round out the top 5 because the next wave of AI adoption will likely move from the cloud to the device (Edge). As privacy concerns and latency requirements grow, Qualcomm's dominance in low-power, high-efficiency mobile inference puts it in a unique position for growth.