

# Lecture 27: Parallel 3

Wednesday, March 13, 2019 9:41 AM

## Outline

- Warehouse scale computing and the cloud
- Accelerators

Course evals!

## Warehouse scale computing

NOTE: My pen died and I used the chalk board.

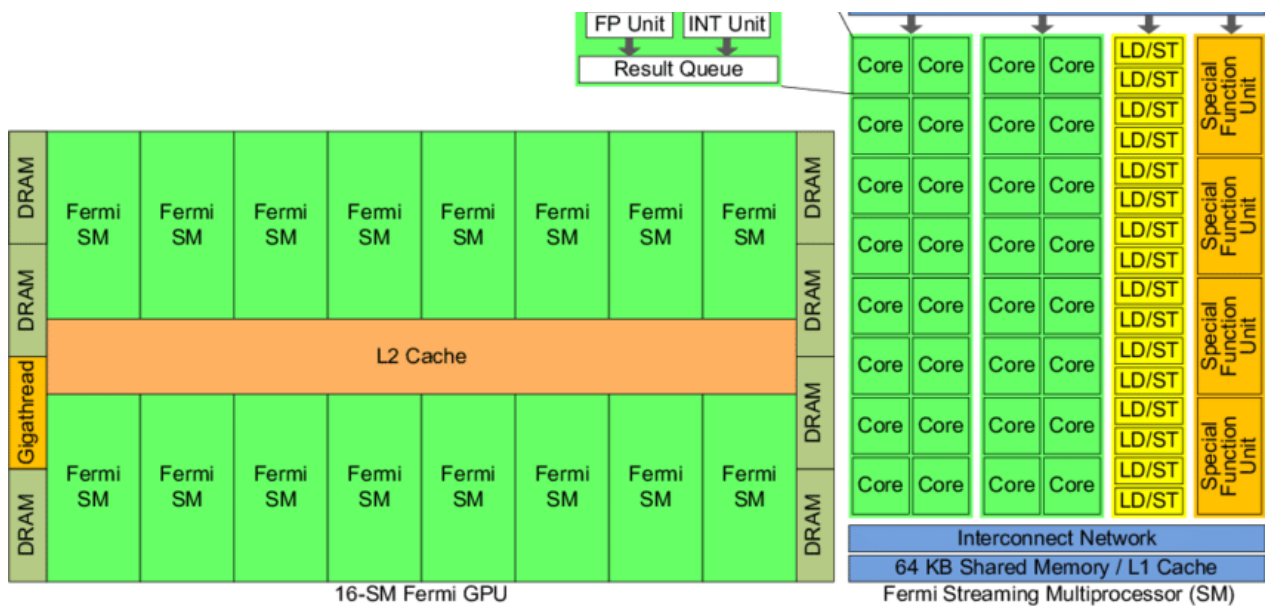
Here are my notes to myself 😊

- Warehouse scale computing
  - So far, we have looked at shared memory multiprocessing.
    - Communication happens on the CPU-side of the memory bus
    - Draw this.
  - Warehouse scale computing is different
    - Communication happens on the other side of the bus
    - Must send explicit messages
  - Ways to do it
    - send/rcv data
      - Messages, RDMA
      - Traditional HPC
    - RPC (remote procedure calls)
      - What Google/facebook/etc do
      - Micro services
      - "serverless" computing
  - The cloud
    - Create instances of VMs away from where you are
    - Could be just "containers"
    - Could be just "functions"
  - Often have things like memcached on the front end
  - Need load balancers to choose which machine to use
  - Map reduce
    - Example: word count
- ```
map(String key, String value) (key is document name, value is contents)
  for each word w in value: EmitIntermediate(w, "1")
reduce(String key, iterator values) (key is word, value is list of counts)
  result = 0
  for each v in values: results += ParseInt(v)
  Emit(result)
```

## Accelerators

### GPUs

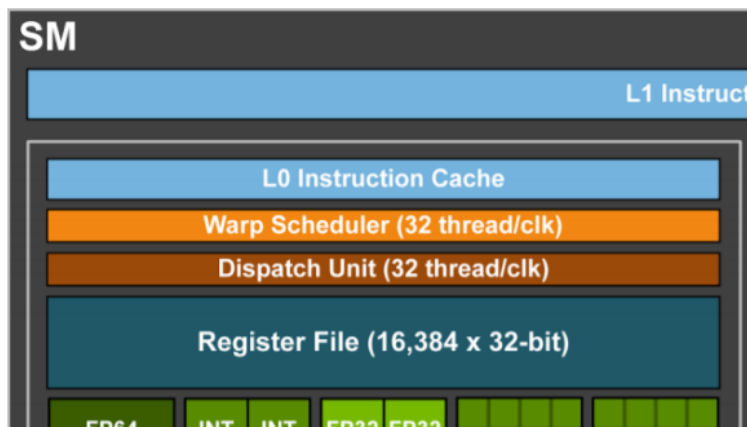


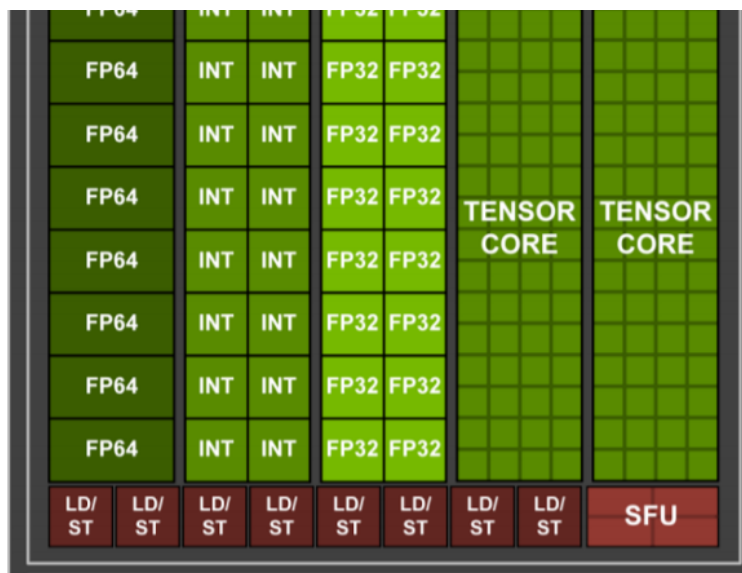


Current Volta/Turing

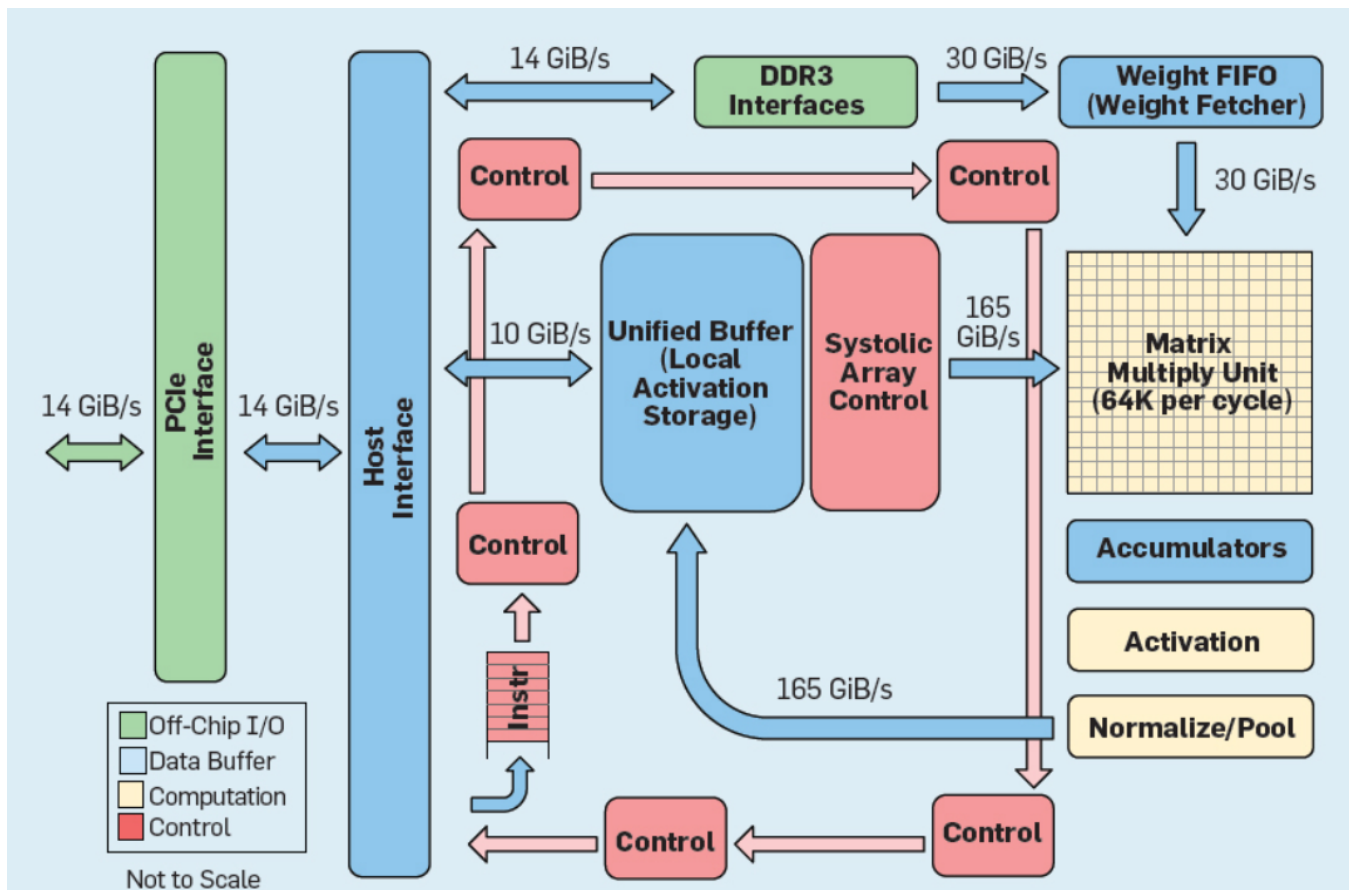


Figure 4. Volta GV100 Full GPU with 84 SM Units

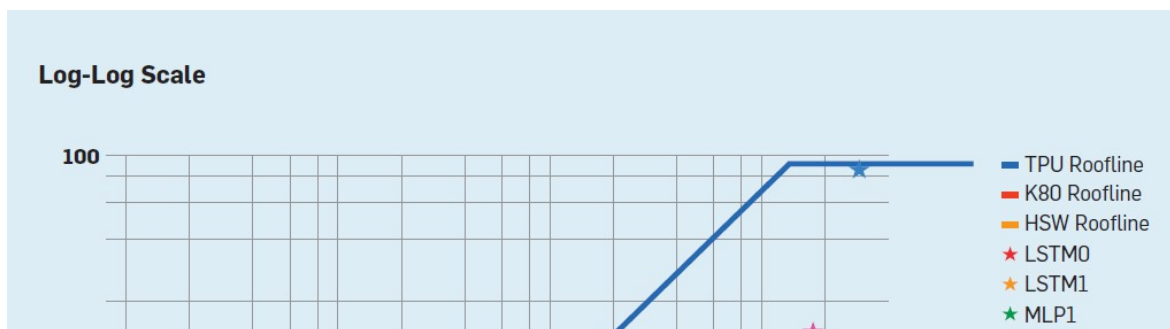


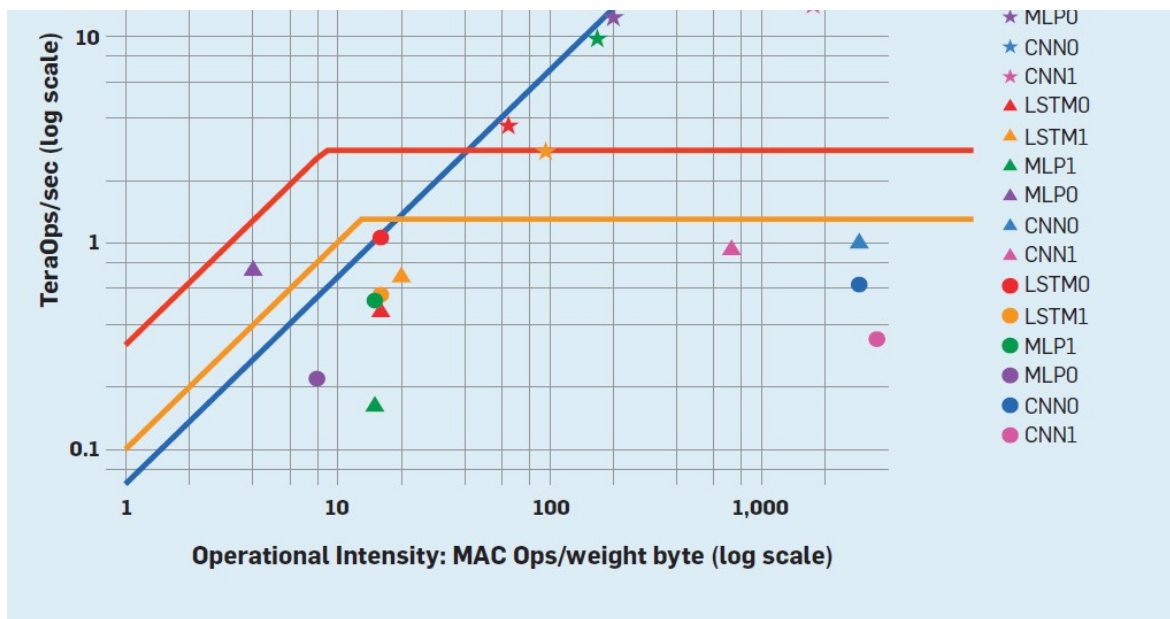


Tensor processing unit from Google



Performance





## For tomorrow

Read: <https://www.sigarch.org/lets-keep-it-to-ourselves-dont-disclose-vulnerabilities/>

Read: <https://www.sigarch.org/please-disclose-security-vulnerabilities/>

Pick a side and write 2-3 sentences as to why you agree with your side and disagree with the other.

I'll take this up at the end of class.

We'll have a discussion about this.