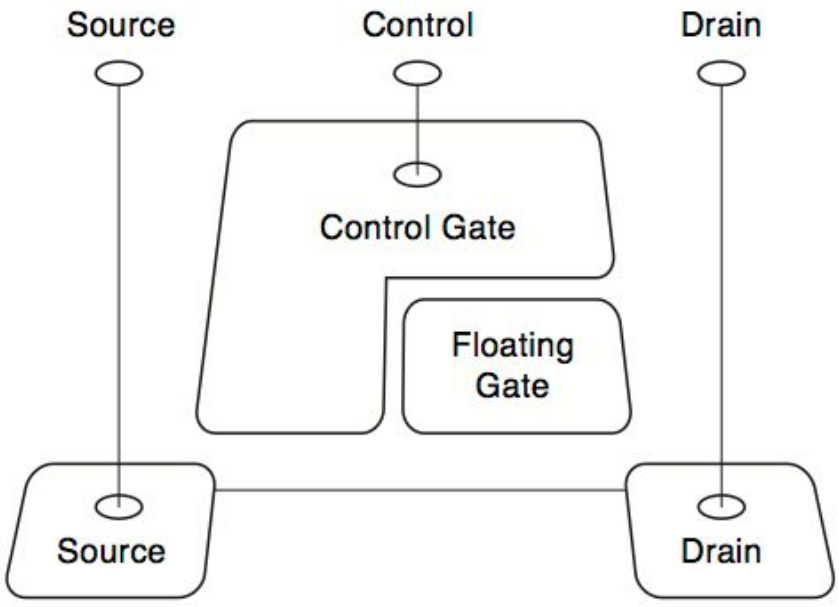
Topic 18: Flash Storage

Reading: 12.2-12.3

- Hard drives are mechanical and slow.
- Over the years there have been many attempts to replace it with something that is electronic and therefore faster.
- Flash storage has been the most successful of these.
    - It is *solid state*, i.e. electronic, so it has no moving parts.
    - It has a floating gate that can hold its charge for a very long time.



- Floating gate is not connected to anything and is separated from the control gate by insulator material. However, a large enough voltage to the control gate can charge or discharge the floating gate.
- Floating gate charge determines whether or not the transistor turns on when normal voltage is applied to the control gate.
- We are not electrical engineers. What this means to us is:
  - Flash must be read and written in whole 2KB - 16KB *pages*.
  - Flash must be erased before it can be written. It must be erased in *erasure block* units of 128KB to 512KB depending on the device.
    - Cannot overwrite individual pages.
    - An erasure block can only be erased a finite number of times. It can *wear out*.

- Flash is usually packaged as a disk (SSD).
  - It provides pretty much the same interface to the OS as a hard drive, e.g. read and write sectors.
  - Much higher performance vs magnetic disk (HDD).
    - Random reads: 10K/sec vs. 100

- Random writes: 2K/sec vs 100
- Sequential: 250MB/sec vs. 100MB/sec.

- Much less power.
- Much more durable.
- Much more expensive ($0.10/GB vs $0.025/GB)

- *Flash Translation Layer (FTL)*
  - Maps logical sectors to physical flash pages.
    - Giant indirection table
  - When a sector is written by the OS, the SSD chooses a "good" page in a "good" erasure block in which to write it.
    - Must erase entire erasure block.
      - Must move any live sectors out of it first.
    - Ideally SSD can cache entire pages of sectors, or even entire erasure blocks of sectors, before actually writing them.
      - Volatile memory backed by capacitors.
    - *Wear leveling*: choose erasure block based on how worn out it is. A "good" erasure block is one that has a low wear level.
      - Ensure that erasure blocks wear out evenly.

- Even if filesystem writes the same sector(s) repeatedly, e.g. the sectors containing the FAT. Filesystems were not designed with wear limits in mind.

■ To move all live sectors out of an erasure block must first know which sectors are live (i.e. not considered free by the filesystem).

- How does the SSD know this? Traditionally file systems did not tell the disk when a sector is no longer in use.

  - SSDs added a “trim” operation so the SSD can know which sectors are live and which are dead.

    - File system uses trim when a sector is added to the free list.