ECE 571 – Advanced Microprocessor-Based Design Lecture 27

Vince Weaver

https://web.eece.maine.edu/~vweaver

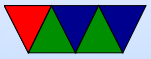
vincent.weaver@maine.edu

13 November 2024

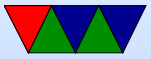
Announcements

- Homework 9 assigned
- Don't forget the project ideas
- Further readings:
  - https://semiengineering.com/dram-thermal-issues-reach-crisis-point/
  - https://www.semianalysis.com/p/the-memory-wall
  - *An Experimental Setup to Evaluate RAPL Energy Counters for Heterogeneous Memory* paper

    https://dl.acm.org/doi/10.1145/3629526.3645052

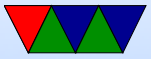


Future Memory

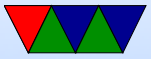


CXL Interconnect

- Compute Express Link
- By Intel and others, initially to get cache coherent memory on PCIe cards (like GPUs)
- RAM attached via PCIe
- Cache coherent
- 200ns latency
- `https://www.theregister.com/2022/08/04/cxl_time_now/`
- Intel Sapphire Rapids and AMD Zen4 have support
- CXL.io, CXL.cache, CXL.mem

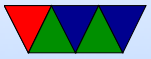


- 1.0 based on PCIe 5.0, 3.0 PCIe 6.0

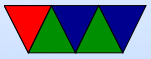


Non-volatile RAM (NVRAM)

- Core Memory
  - Old days, tiny ferrite cores on wire
  - Low density
- MASK ROM/EPROM/EEPROM
- Battery backed (CMOS) RAM
- FeRAM/Magnetoram – store in magnetic field
- MRAM (magnetic RAM), Spin-transfer-torque (STT-MRAM)
- Flash NAND/NOR (will talk about more with disks)

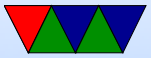


  - Only so many write cycles (thousands) as opposed to billions+ for DRAM
  - High power to erase
  - Often have to erase in large blocks, not bit by bit
  - Wear leveling
- Millipede memory, tiny bumps, MEMS devices to read
- Phase change RAM (see below)
- Memristors (see below)
- Intel/Micron Optane/3D-Xpoint (see below)



Phase Change RAM

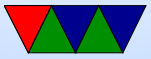
- Material
  - bit of material can be crystalline or amorphous
  - resistance is different based on which
  - need a heater to change shape
  - needs a lot of current to change phase
  - chalcogenide glass – used in CD-Rs
  - heating element change from amorphous (high resistance, 0) to crystalline (low resistance, 1)
  - Amorphous if you heat and quench, crystal if cook a



while

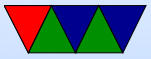
- temp sensitive, values lost when soldering to board (unlike flash)
- Newer methods might involve lasers and no phase change?

- Features
  - Faster write performance than flash (slower than DRAM)
  - Can change individual bits (flash need to erase in blocks)
  - can potentially store more than one bit per cell



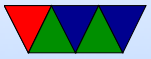
- better than flash (takes .1ms to write, write whole blocks at once
- 100ns (compared to 2ns of DRAM) latency

- Longevity
  - Flash wears out after 5000 writes, PCM millions
  - Flash fades over time. Phase change lasts longer as long as it doesn't get too hot.
  - But also, unlike DRAM, a limit on how many times can be written.
- Can you buy phase change ram?
  - Micron sold from 2012-2014? Not much demand



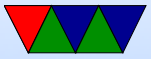
Memristors

- resistors, relationship between voltage and current
- capacitors, relationship between voltage and charge
- inductors, relationship between current and magnetic flux
- memristor, relationship between charge and magnetic flux; "remembers" the current that last flowed through it
- Lot of debate about whether possible. HP working on memristor based NVRAM



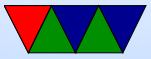
Intel/Micron Optane/3D-Xpoint/QuantX

- `https://specbranch.com/posts/rip-optane/`
- Note: Intel finally cancelled Optane in 2022
- 3D-Crosspoint (intel) QuantX (Micron)
- Faster than flash, more dense than DRAM
- Can get it in an SSD (so no special hardware needed)
- Also as special slot on motherboard (or even DIMM)
- 3D grid, not every bit needs a transistor so can be 4x denser than DRAM. Bit addressable.
- Intel very mysterious about exactly how it works



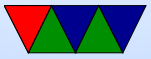
ReRAM (store in changed resistance) but is it phase-change?

- Intel denies everything
- ReRAM works by having a dielectric layer and blasting channels through it.
- Can you buy Optane?
  April 24th 2019? Special M.2 slot on Gen7 (Kaby lake? motherboards)
  For now, 16GB and 32GB modules, using like a cache of your hard disk.

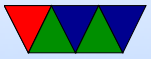


NVRAM Operating System Challenges

- How do you treat it? Like disk? Like RAM?
- Do you still need RAM? What happens when OS crashes?
- Problem with treating like disk is the OS by default caches/ copies disk pages to RAM which is not necessary if the data is already mapped into address space
- Challenges: Mapping into memory? No need to copy from disk?

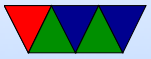


- Problems with NVRAM: caches.
- Memory is there when reboot like it was, but things in caches lost.
- So like with disks, if the cache and memory don't match you're going to have problems trying to pick up the pieces.

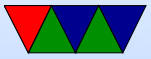


Saving Power/Energy with RAM

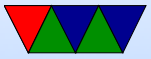
- AVATAR: A Variable retention time aware refresh for DRAM systems by Qureshi et al.
  - JEDEC standard: cell must have 64ms retention time
  - Why refresh bad? Block memory, preventing read/write requests
  - Consume energy (6,28,35)
  - The bigger DRAM gets, more refresh needed
  - predict that in 64Gb chips 50% of Energy will be in refresh



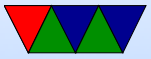
- Multi-rate refresh possible – detect which cells need more and refresh them more often (can be a 4-8x difference)
- VRT (variable retention rate) a problem. Some cells switch back and forth between. So when you probe it might check fine, but then fail later.
- They find that addition of cells stabilized to one new cell/15 mins over time
- Use ECC to catch these errors, though relying on ECC in this case can lead to uncorrectable error every 6 months



- They propose using ECC to adjust the VRT at runtime based on errors that are found
- They find on a 64Gb chip improves perf by 35% and **Energy-Delay** by 55%
- "Refresh-wall"
- Memory controller keeps track of this info
- VRT first reported in 1987. Fluctuations in GIDL (gate-induced drain leakage) presence of "trap" near the gate region
- Intel and Samsung say VRT one of biggest challenge in scaling DRAM

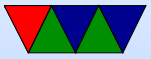


- VRT not necessarily bad – can cause retention to get better!
- Test – use FPGA to talk to 24 different DRAM chips, at controlled temperatures.
  Why do they use an FPGA?
- Actually it's just 3 chips from different vendors, each with 8 chips (for 24)
- Look into ECC. Soft-error rate is 200-5000 FIT/Mbit. Every 3-75 hours for 8GB DIMM. Soft errors happen 54x-2700x lower rate than VRT
- Downside of ECC ... have to scrub memory to check

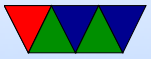


for errors. Also has energy/perf overhead. Energy to refresh DIMM 1.1mJ, energy to scrub 161mJ (150x) but if you scrub every 15 minutes it's a win.

– Use memory system simulator USIMM

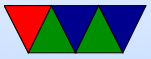


DRAM – Power Saving



DRAM – Mobile DRAM

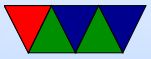
- From Micron: "TN-46-12: Mobile DRAM Power-Saving Features", 2009
- Temperature-Compensated Self Refresh (TCSR) – Auto adjust refresh timings based on temperature
- Partial Array Self Refresh (PASR) – only refresh parts of RAM that have data in them
- Deep Power Down (DPD) – enable turning off the voltage generators when maintaining DRAM not needed
- Has equations for estimating power usage



DRAM – Elsewhere

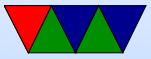
- Tom's Hardware. 2010. "How Much Power Does Low-Voltage DDR3 Memory Really Save?" Using low-voltage (1.25 or 1.35 rather than 1.5) DDR3 DRAM can reduce power by 0.5-1W. Slower performance settings, but not really noticeable.

- Linus Torvalds Rant from 2007: DRAM Energy not a prime concern. Just don't use FBDIMMs if you want low-power.



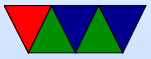
DRAM – Recent Academic

- "Rethinking DRAM Power Modes for Energy Proportionality", Malladi et al, Micro 2012.
  - DRAM spends lots of time idle, but latency is so high for wakeup it cannot utilize powerdown modes
  - Reference 25% of data-center energy usage is DRAM?
  - Use LPDDR2 trades bandwidth for efficiency
  - Current modes involve turning off DLLs (Delay-locked loops?) which are slow to turn on again, 700ns+
  - some background on DRAM operation

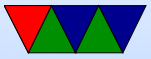


- Low-power mode sounds good, but then it takes 512 memory cycles of power to re-start (a lot of energy)
- Propose MemBLAZE. Moves clock generation out of DIMM and into memory controller, allowing fast wakeup

- "Towards Energy-Proportional Datacenter Memory with Mobile DRAM", Malladi et al, ISCA 2012.
  - Look at using LPDDR2 in servers rather than DDR3.
  - DDR3 often in "Active-idle" as many workloads do not allow sleep.



- "A Predictor-based Power-Saving Policy for DRAM Memories", Thomas et al, EuroMicro 2012.
  - Use a history based predictor to pick when to powerdown.
  - Say up to 20% of mobile devices and 25% of data center is DRAM
- "Rethinking DRAM Design and Organization for Energy-Constrained Multi-Cores", Udipi et al., ISCA 2010
  - DRAMs "overfetch" which hurts energy
- "A Comprehensive Approach to DRAM Power



Management", Hur and Lin, HPCA2008.

- Throttling and Power Shifting – slowing down to fit in power budget
- Put DRAMs in low power mode – available commercially but no one seems to use this yet
- Simulate for Power5 and DDR2-533
- Modify the memory controller