I watched the recommended presentation on “Enabling In-Memory Computation” in the slides and noted what I learned.

**Main memory is a critical component of all computing systems**: server, mobile, embedded, desktop, sensor, etc.

Main memory must scale in size, technology, efficiency, cost, and management algorithm to maintain **performance growth and technology scaling benefits**

DRAM and memory controllers, as we know them today, are (will be) unlikely to satisfy all requirements

Some emerging non-volatile memory technologies like PCM enable new opportunities: memory + storage merging 🡪 we need to rethink the main memory system to fix DRAM issues and enable emerging technologies

**Major trends affecting main memory**

1. Need for main memory capacity, bandwidth, QoS increasing
   * **Multi-core**: the increasing number of cores/agents
   * **Data-intensive applications**: increasing demand/hunger for data
   * **Consolidation**: cloud computing, GPUs, mobile, heterogeneity
2. Main memory energy/power is a key system design concern
   * ~40-50% energy spent on off-chip memory hierarchy [Lefurgy, IEEE computer’03]
   * >40% on DRAM [Ware, HPCA’10][Paul, ISCA’15]
   * DRAM consumes power even when not used because of periodic refreshing
3. DRAM technology scaling is ending
   * ITRS projects that DRAM will not scale easily below X nm
   * Scaling has provided many benefits
     1. The higher capacity (density)
     2. Lower cost
     3. Lower energy

**2000 ISCA work by Lim et al.**: Core count doubles every 2 years, although DRAM DIMM capacity doubles every 3 years. **In other words, memory capacity per core is expected to drop by 30% every two years**.

A work by the prof’s group in collaboration with Google in “Google workloads for consumer devices: mitigating data movement bottlenecks” title showed **that 62.7% of the total system energy is spent on data movement**.

Major trend: hybrid main memory (traditional DRAM + PCM for example)

Main memory needs **intelligent controllers**. One can predictably induce errors in most DRAM memory chips (RowHammer). If we don’t have an intelligent memory controller, increasing memory refresh rate can be a solution to RowHammer problem. However, actually we want to decrease this rate. Safari group’s solution for this problem was “**PARA: Probabilistic Adjacent Row Activation**”. Key idea is that after closing a row, they activate one of its neighbors with a low probability (p = 0.005).

What I understood generally in this presentation is that we have to try to substitute DRAM with other technology instead of complexing our architecture to adapt to using this subsystem in our computing systems.

I have to re-watch it later.

Then, I watched the recommended presentation on “**Accelerating Genome Analysis: A Primer on an Ongoing Journey”** in the slides and noted what I learned.

System design for bioinformatics is a critical problem because it has large scientific, medical, societal, personal implications. In genomic processing, DNA read mapping is the bottleneck. For sequencing we can sequence fast. Many bottlenecks exist in accessing and manipulating huge amounts of genomic data during analysis.

Bioinformatics dream in 2007 was to possess an embedded device, which can perform genomic analysis in real time within a minute.

* Which of DNAs does this DNA segment match with?
* What is the likely genetic disposition (behavior) of this patient to this drug?

**Algorithmic acceleration**

* Exploiting structure of the Genome
* Exploiting SIMD structure

**Hardware Acceleration**

* Specialized Architectures
* Processing in memory

**Future opportunities**: New sequencing technologies

**DNA Sequencing**: our goal in DNA sequencing is to find the complete sequence of A, C, G, and T’s in DNA.

**Challenge of DNA sequencing**: There is not machine that takes long DNA as an input, and gives the complete sequence as output. All sequencing machines chop DNA into pieces and identify relatively small pieces (but now how they fit together)

Development of high-throughput sequencing (HTS) technologies decreased the cost of sequencing so much. Today, with the help of this technology we can sequence more genomes than we can process. Some of high-throughput sequencing technologies: Illumina, Roche 54, Ion Torrent, SOLID …

In fact what goes on in this technology is that small DNA fragments are first amplified and then sequenced in parallel which results in:

* High Throughput
* High Speed
* Low Cost
* Short reads

Anyway, today we have tradeoffs about developing devices for sequencing genome and we can do it fast. However, our today bottleneck is DNA read mapping.

**Example question**: If I give you a bunch of sequences, tell me where they are the same and where they are different.

**Example question 2**: Given a bunch of short sequences, can you identify the approximate species cluster for gnomically unknown organisms (bacteria)?

**Answer**: If we want to answer these questions we need to construct the entire genome from many reads. However, this is where our bottleneck emerges. DNA read mapping is 150X slower than read sequencing. Mapping short reads to reference genome is challenging (billions of 50-300 base pair reads).