18-447 Lecture 1: Intro to Computer Architecture

James C. Hoe
Department of ECE
Carnegie Mellon University

Housekeeping

- Your goal today
  - know what you are getting into
  - decide if you are staying . . .
- Notices
  - Student survey on Canvas, past due
  - Handout #4: HW1, due noon 2/5
  - Lab 1, Part A, due Week 3
  - Lab 1, Part B, due Week 4
- Readings (for Today)
  - P&H Ch1
  - Architecture* of the IBM System/360, Amdahl, et al

What is 18-447?

- 18-213: Introduction to Computer Systems
  - “C” as the model of computation
  - interact with the computer hardware through OS
  - what about the details below the abstraction?

Somehow a program ends
up executing as digital logic

- 18-240: Fundamentals of Computer Engineering
  - digital logic as the model of computation
  - gates and wires as building blocks
  - what about the details below this abstraction?

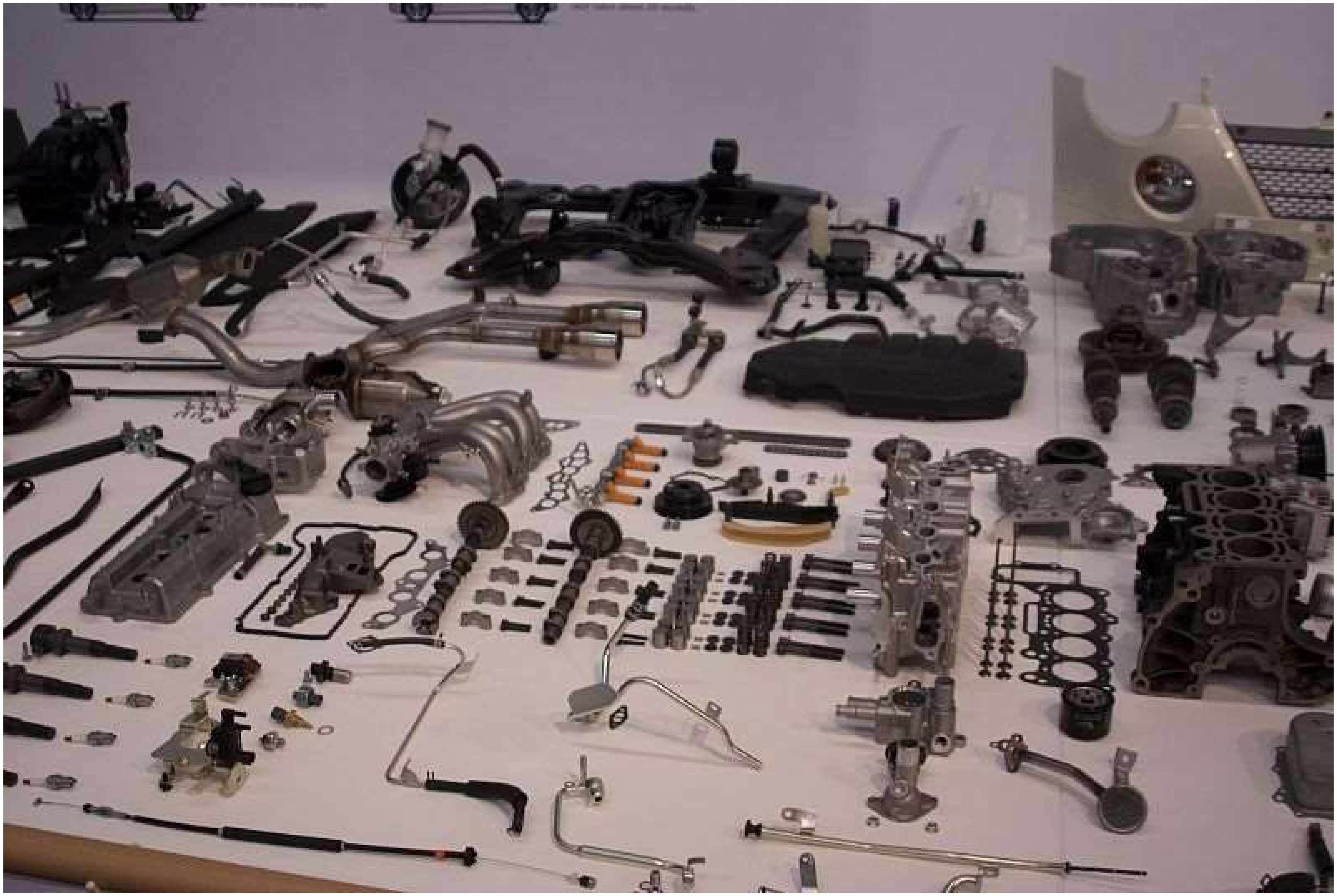
18-447: Fuzzy to Concrete

18-213

- "Computer Architecture"
  - functional spec for software and programmers
  - design spec for the hardware people
- Computer Organization
  - take architecture to "micro"architecture
  - how to assemble/evaluate/tune
- Computation Structures
  - digital representations
  - processing, storage and I/O elements

18-240

This is not 18-447

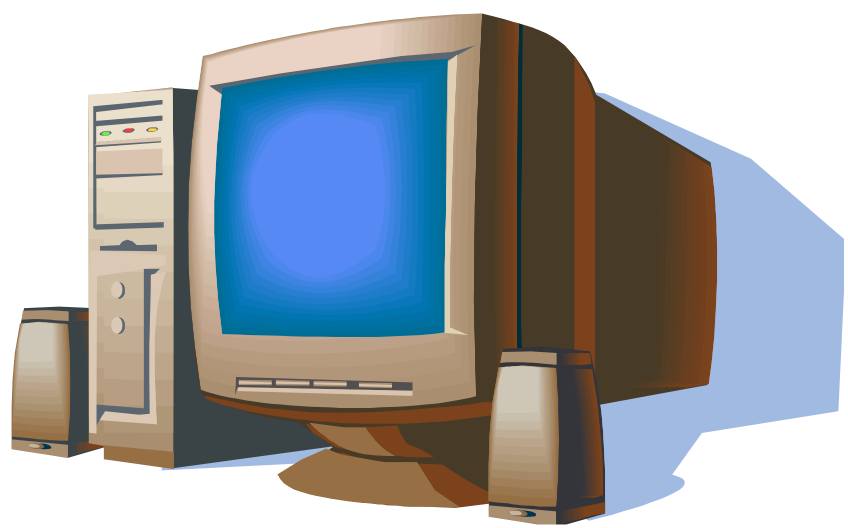


[Wikimedia Creative Commons]

What is a Computer?

- **Computer, 2. a.** A calculating-machine; esp. an automatic electronic device for performing mathematical or logical operations; freq. with defining word prefixed, as *analogue*, *digital*, *electronic computer*.

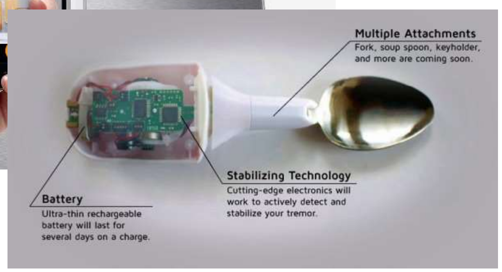
--- Oxford English Dictionary, circa 2000

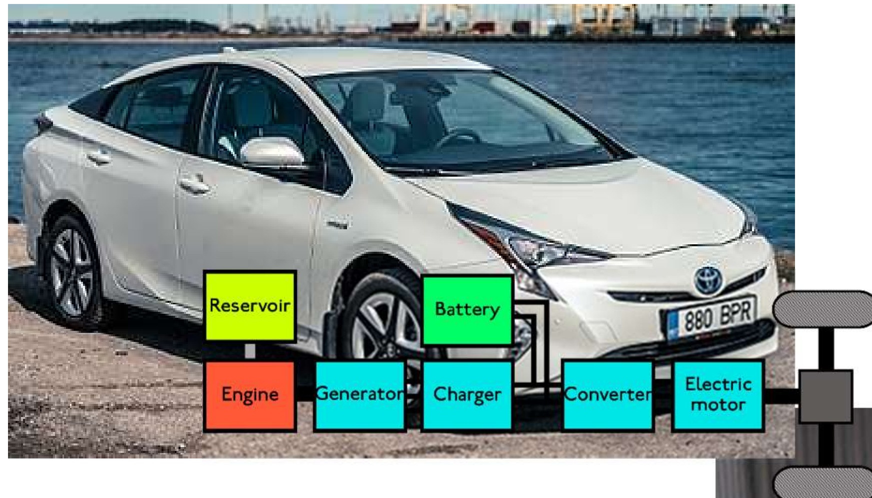


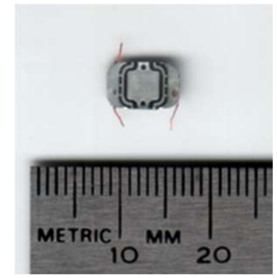
Some Familiar Computers

[images from Wikipedia]

Where is the computer?







[images from Wikipedia]

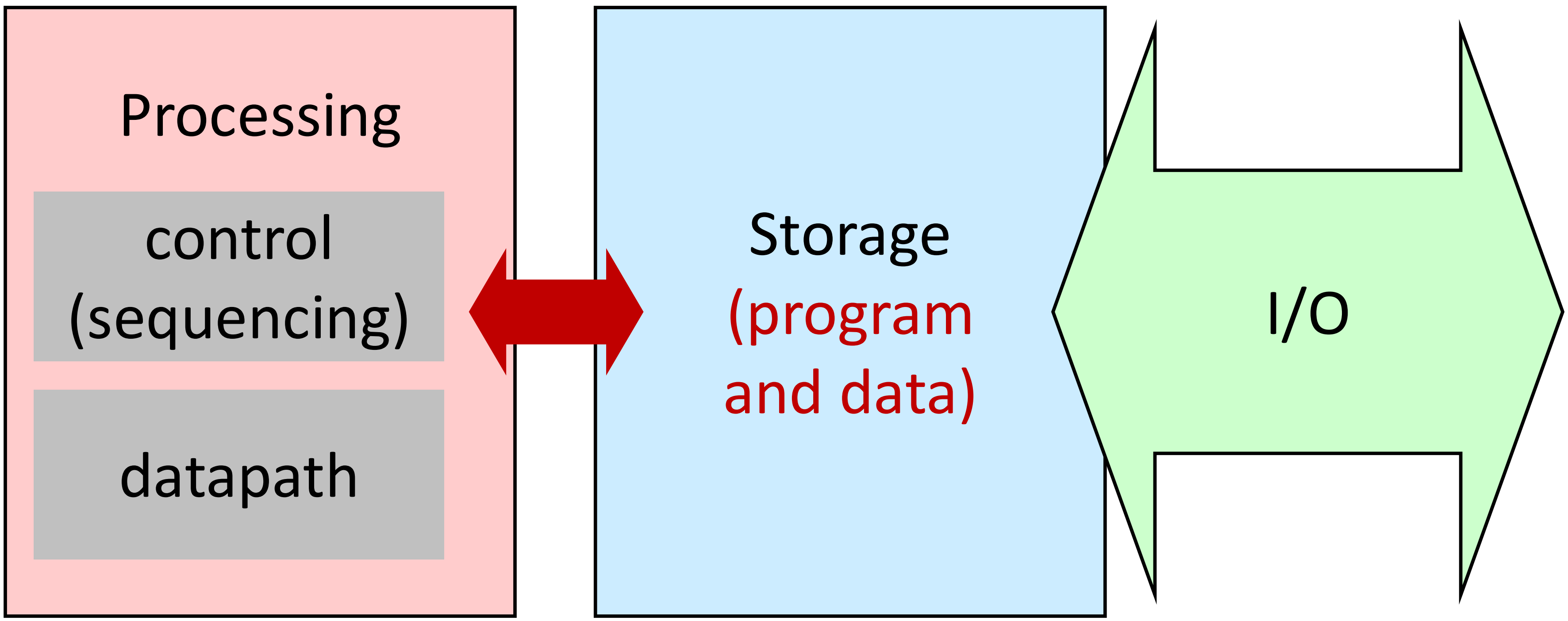
Modern computing is as much about enhancing capabilities as data processing!!

Keeping up with the times

- Computer, 3. An electronic device (or system of devices) which is used to store, manipulate, and communicate information, perform complex calculations, or control or regulate other devices or machines, and is capable of receiving information . . . and of processing it in accordance with variable procedural instructions . . . used esp. for handling text, images, music, and video, accessing and **using the Internet, communicating with other people (e.g. by means of email), and playing games**.

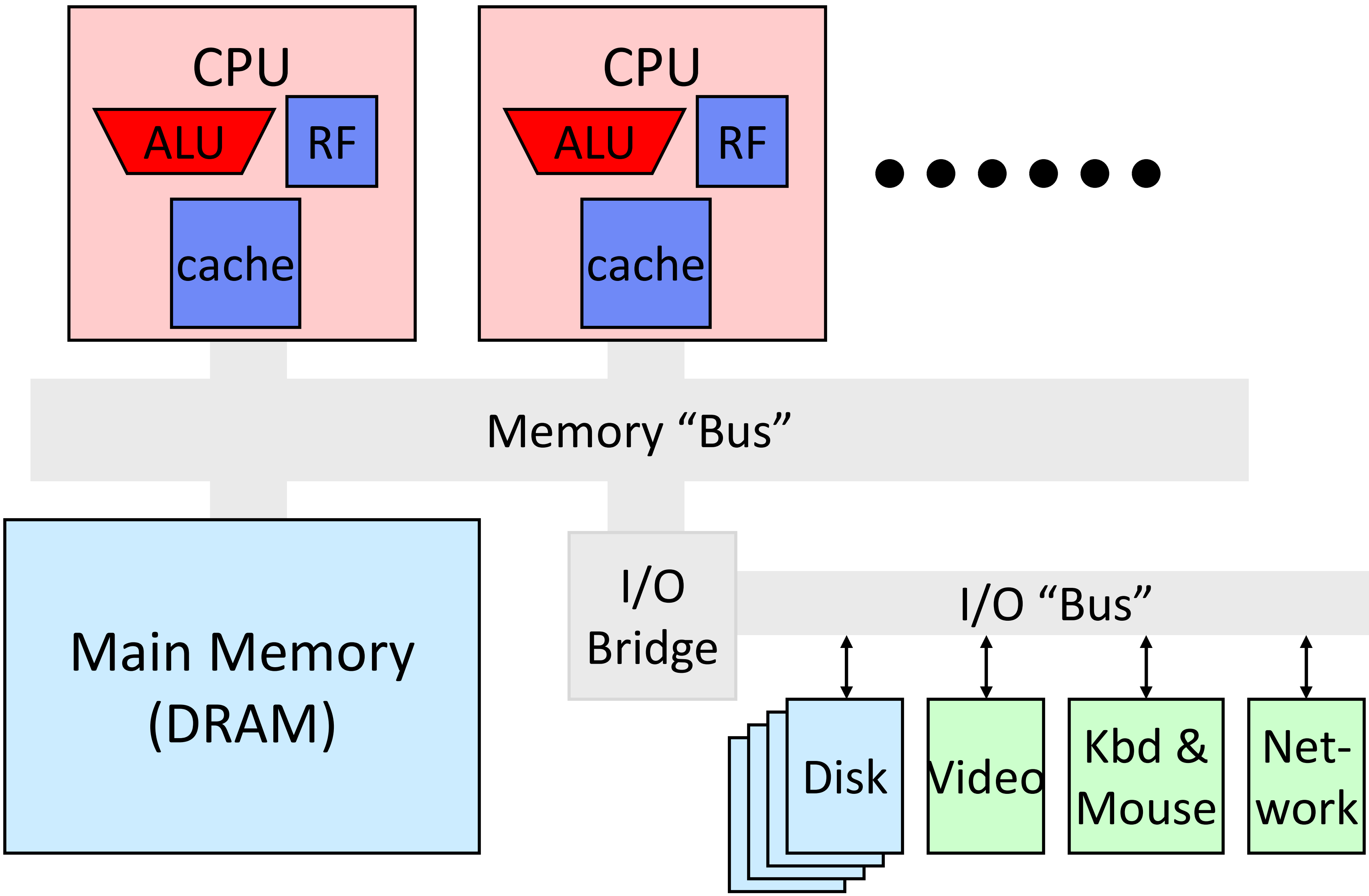
--- Oxford English Dictionary, circa 2018

So what makes a computer a computer?

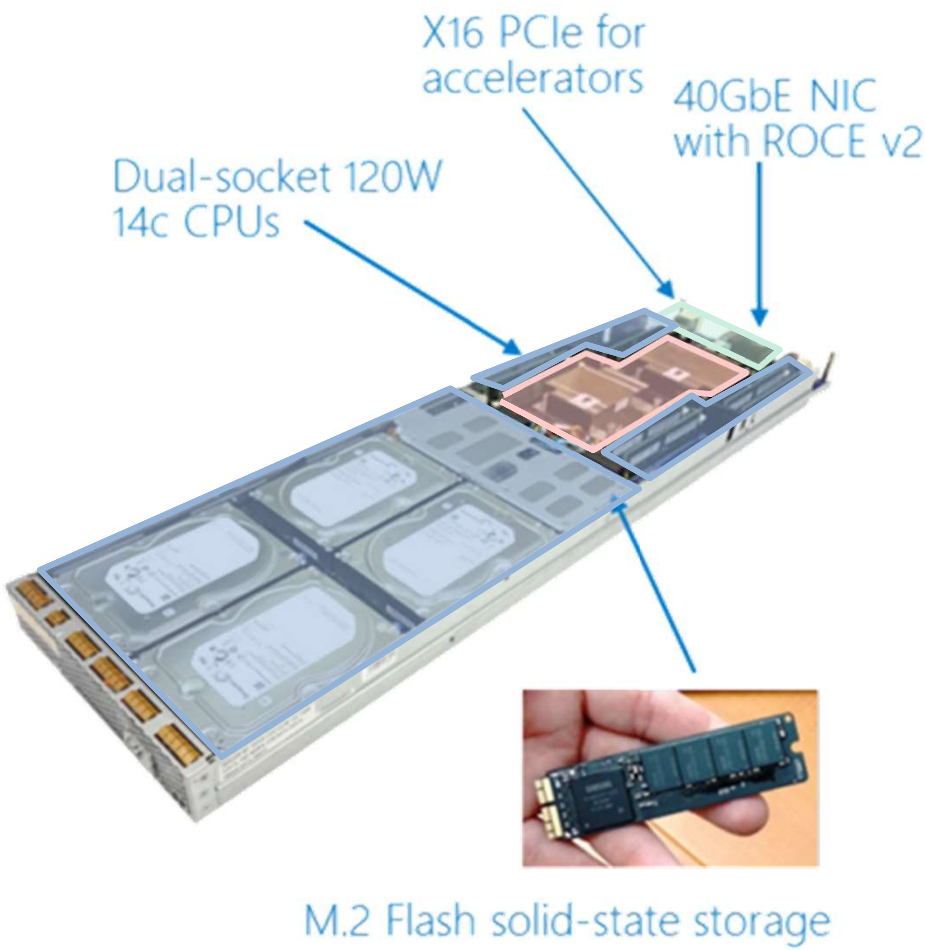


Having program stored as data is an extremely important step in the evolution of computer architectures

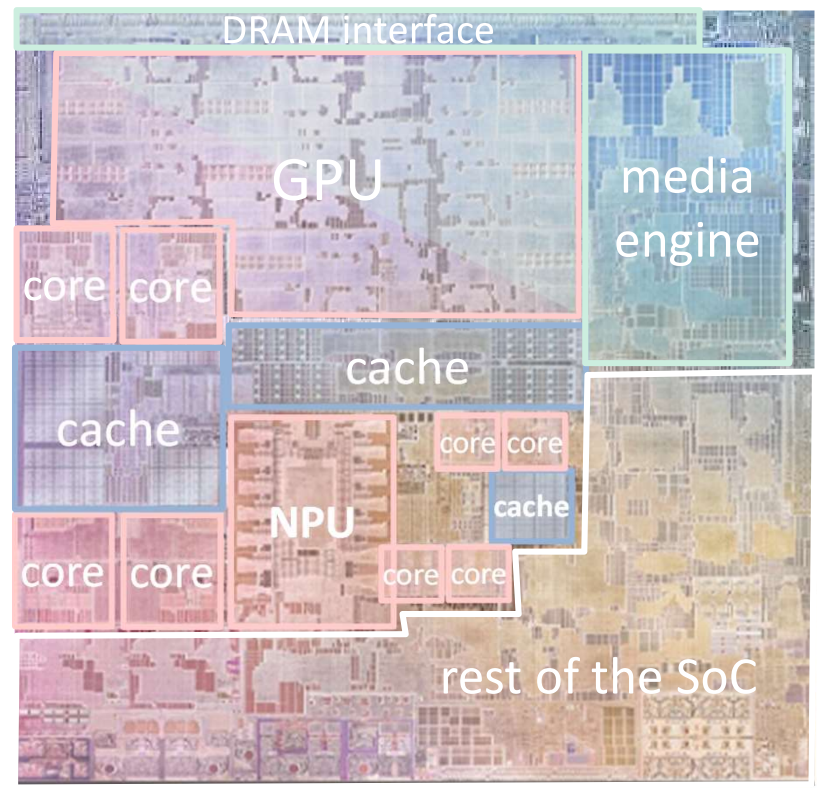
“Classic” Computer Organization



Greatly Varied Manifestations

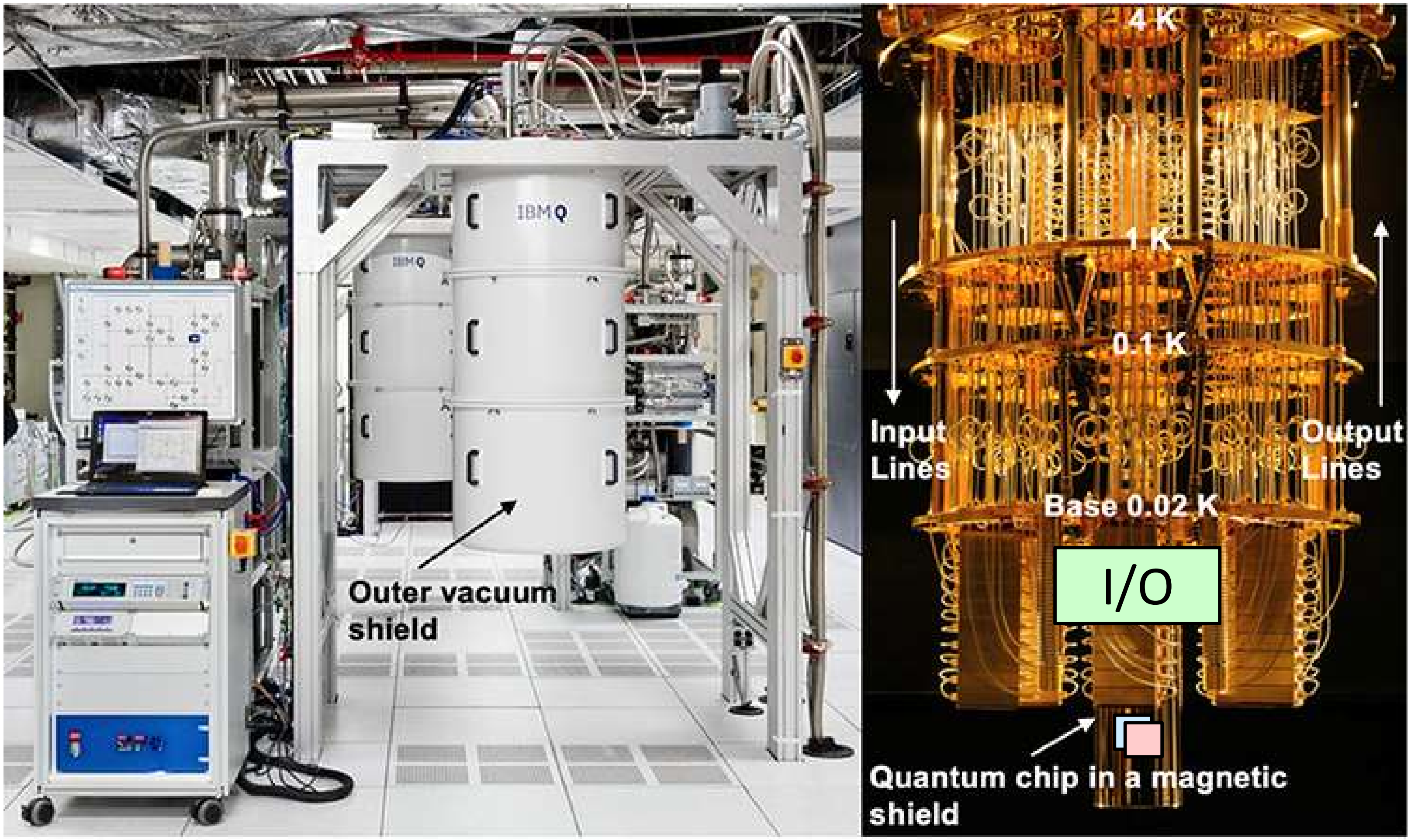


[Microsoft's Open CloudServer]



[raw M1 die photo from apple.com]

IBM Q

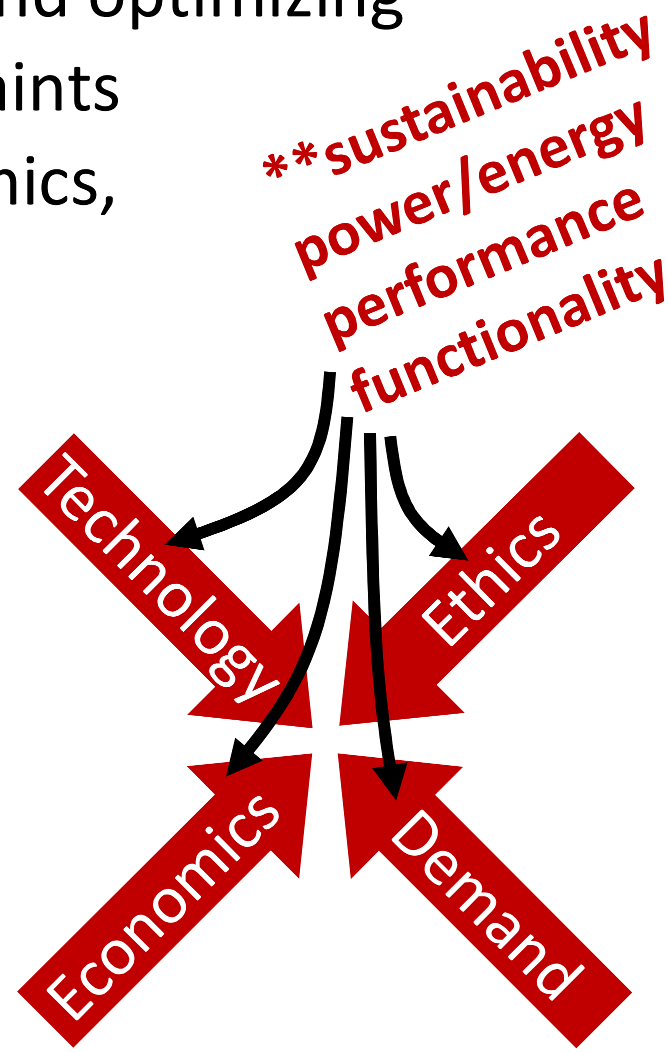


Dilution fridge setup: outside view **Dilution fridge setup:** inside view

[image from IBM.com]

Computer Architecture is Engineering

- An applied discipline of finding and optimizing solutions under the joint constraints of demand, technology, economics, and ethics
- Instances of what we practice evolve continuously
- Need to learn the principles that govern how to develop solutions to meet constraints
- **Don't memorize instances; understand why it is that way**

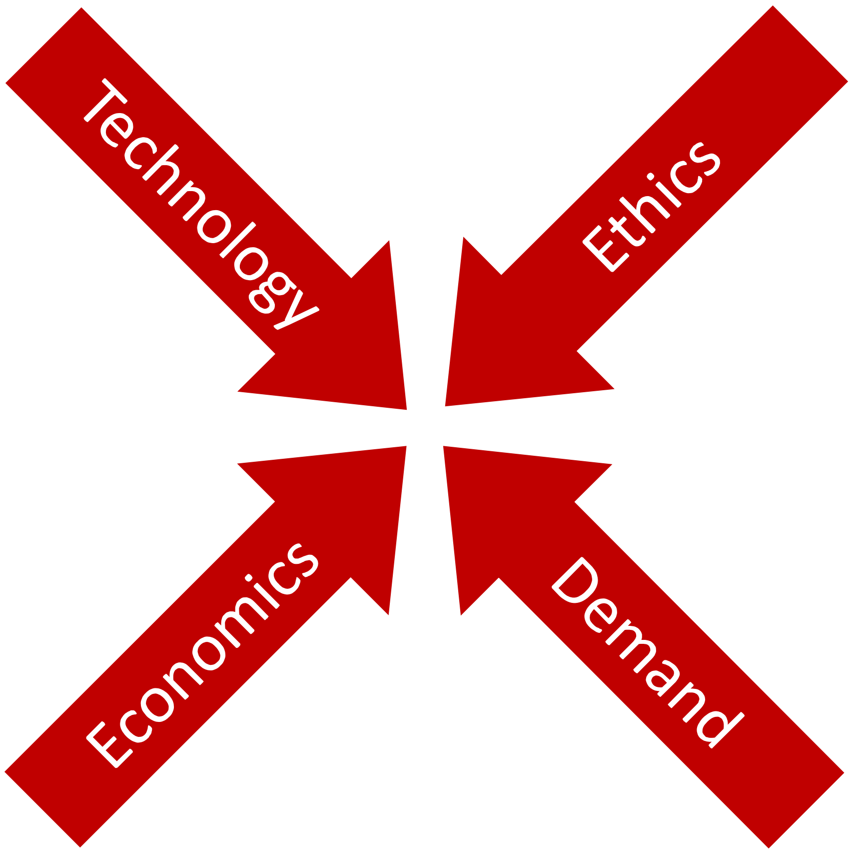


Historical Perspectives:
prelude to modern computer architecture

You should read “Historical Perspectives” at end of P&H chapters.
For more, read “A History of Modern Computing” by Ceruzzi.

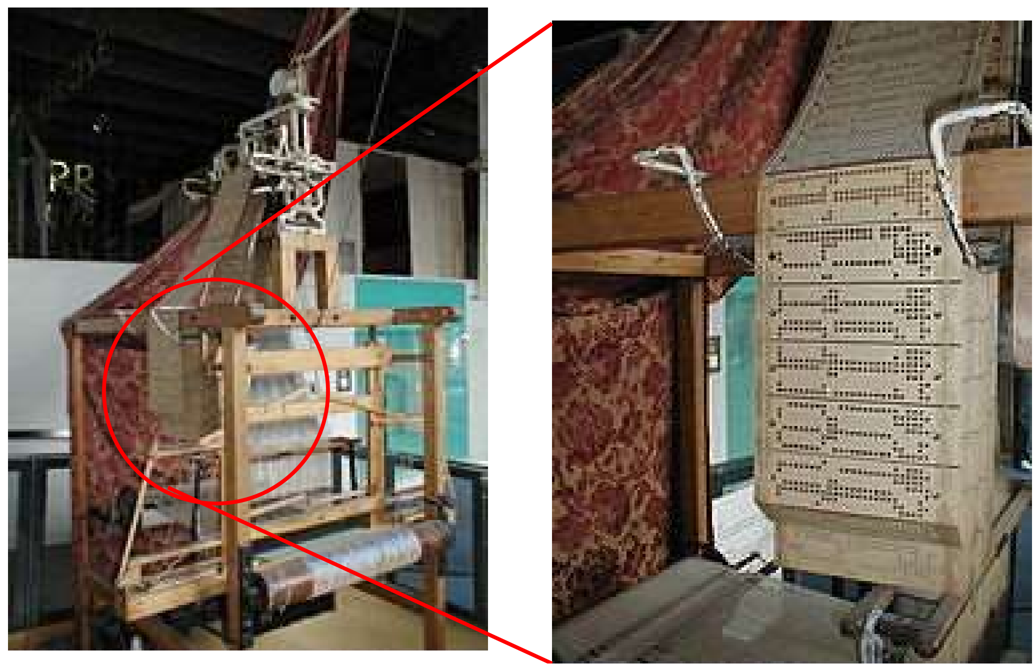
Forces on Innovation

- Timely innovations are rarely unique or original
- Similar constraints lead to similar engineering solutions



Beginnings of Digital Computing

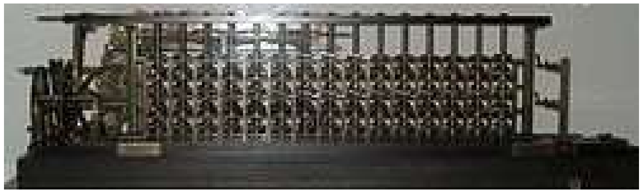
- Industrial Revolution era's "hi-tech" in mechanization
  - steam engines
  - mechanical calculators,
  - Jacquard's loom:
    gears, pulleys,
    chains and
    punch cards



[images from Wikipedia]

Charles Babbage (1791-1871)

- Difference Engine, 1823: a special-purpose computer
  - evaluated polynomial functions by method of successive differences (requiring only additions)
  - eventually built by Georg and Edvard Schuetz in 1855



- Analytical Engine, 1833: a general-purpose computer
  - programmed by punch cards, "assembly language" included loops and branches
  - 1000 word data store, punch card I/O
  - unfortunately never completed (would have been 10x30 meters, steam-engine powered)

[images from Wikipedia]

100 Years of Technology Advances

- Mechanical, 1800s
  - gears, chains, pulleys, and steam power
  - punch cards!!
- Electromechanical, early 1900s
  - switches, relays, acoustic delay line memory
  - e.g. Harvard/IBM Mark 1, Aiken 1939~1944, 50ft long, 5ton, 750K parts, 3~6 sec per addition

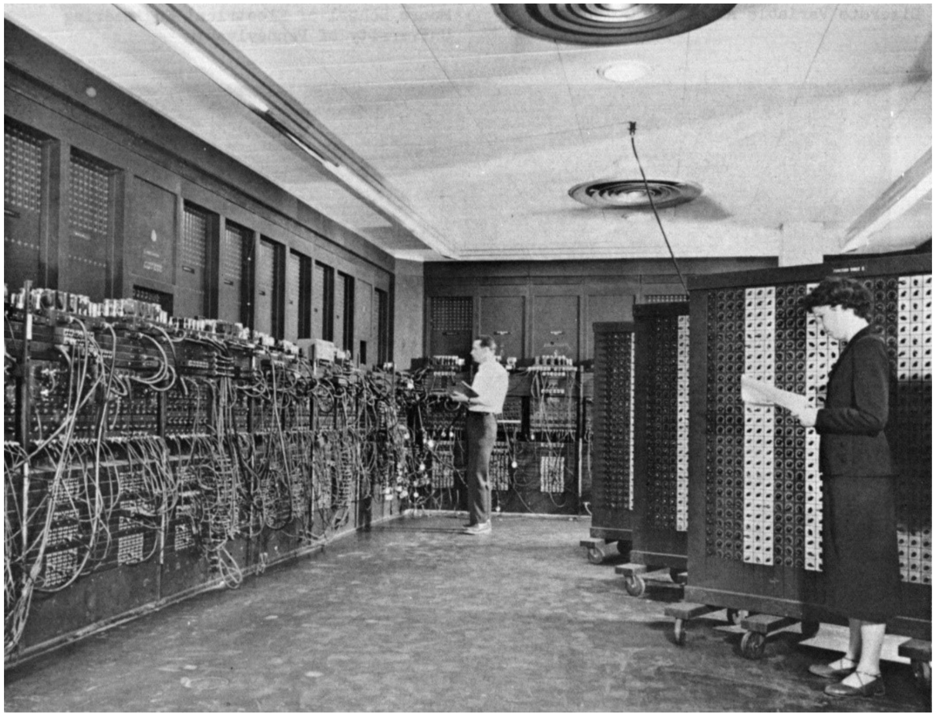
Used ideas from Analytical Engine

- Electronic, mid 1900s and on
  - plugboards, vacuum tubes, cathode ray tube memory
  - and later DRUM, core, transistors and so on . . . . .

Changing demands and economics?

ENIAC, 1946

Eckert and Mauchly, U of Penn



[images from Wikipedia]

- the first programmable electronic digital computer
- 18,000 vacuum tubes
- 30 ton, 80 by 8.5 feet
- 1900 additions per second
- 20 10-decimal-digit words (100-word core by 1952)
- Programmed by 3000 switches in the function table and plug-cables (later became stored program for faster program loading)

Proliferation in 40s and 50s

- From "Moore School Lectures"
  - ENIAC, Eckert & Mauchly, 1946 (revealed)
  - EDVAC, von Neumann, 1944~1952
  - EDSAC, Wilkes, 1949 (first stored-program built)
  - IAS, Bigelow, 1952
  - ORDVAC, SEAC, MANIAC, JOHNIAC, ILLIAC ...
- They were not alone:
  - ABC, Atanasoff and Berry, 39~42
  - Z3, Z4, Konrad Zuse late 30's early 40's
  - Colossus, Alan Turing, 1943
- Don't forget software advances---Fortran was first done in 1954

Commercialization in the 50s

- UNIVAC (1951) the first commercial computer
  contract price $400K, actual cost ~$1M, sold 48 copies
- IBM 701 (1952) “leased” 19 units, $12K per month
  (www-1.ibm.com/ibm/history/exhibits/701/701_customers.html)
- IBM 650 (1953) sold ~2000 units at $200K ~ 400K
- IBM System/360, 1964 Redefined Industry!!
  – a family of **binary compatible** computers (previously, IBM had 4 incompatible lines)
  – 19 combinations of varying speed and memory capacity from $200K ~ $2M
  – ISA still alive today in *z/Architecture* mainframes

Cheaper or Faster in 60s and 70s

- Minicomputers
  - DEC PDP-8, 1965, $20K, size of large refrigerators
  - less powerful than “mainframes”, 10x cheaper
  - departmental computers, timesharing---PDP-11 and VAXs enjoyed extreme popularity in the 70s and 80s
- Supercomputers
  - performance at all cost!! (ECL, liquid-cool, hand-built)
  - biggest customers: national security, nuclear weapons, cryptography, (also aerospace, petroleum, automotive, pharmaceutical, sciences)
  - see Seymour Cray (1925~1996) on Wikipedia

What happened to these computer lines?

Early Examples

DEC PDP 8, 1963
an early mini

Xerox Alto, 1973
an early "PC" with
mouse and GUI

[images from Wikipedia]

Cray 3, 1993

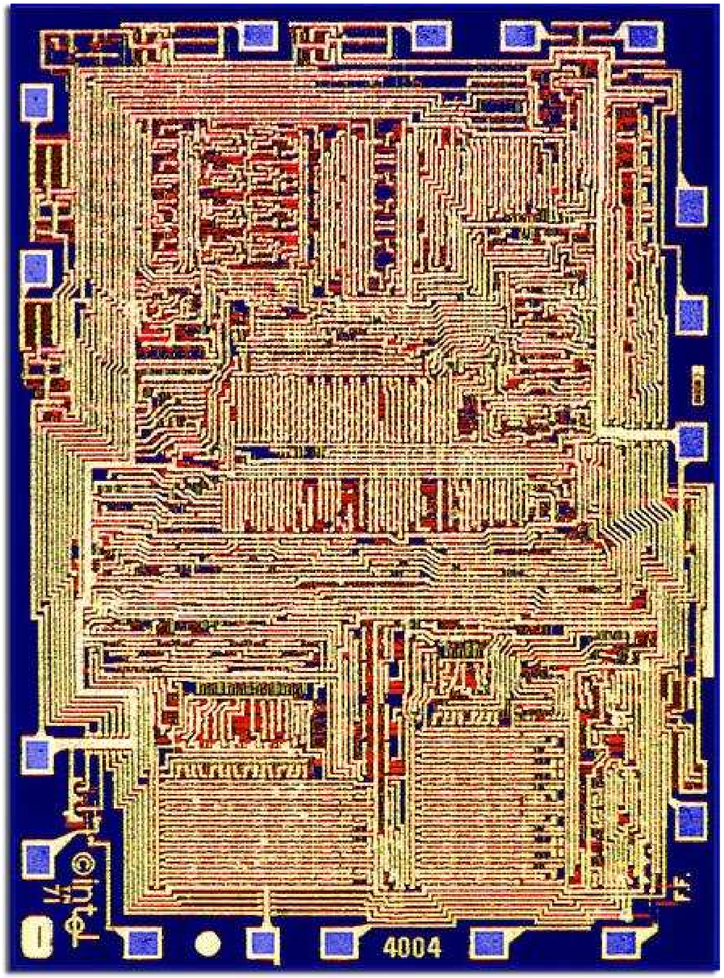
90KW: liquid cooled by “Fluorinert”

$30,000,000

15 GFLOPS (1 sec on Cray3 ≈ 67 years ENIAC)

[images from Wikipedia]

The “Killer Micros” from 70s and on

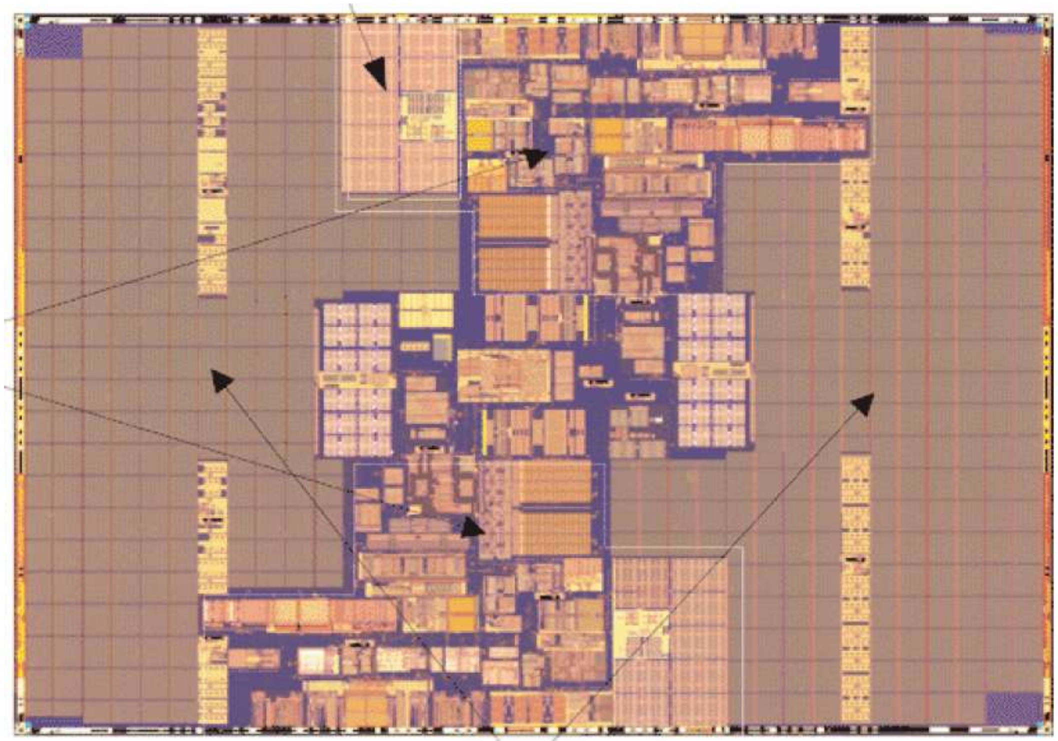


[from Molecular Expressions]

- **Intel 4004, first single chip CPU**
  - 4-bit processor for calculator
  - 2,300 transistors
  - 16-pin DIP package
  - 740kHz (eight clock cycles per CPU cycle of 10.8 μsec)
  - ~100K OPs per second

download the actual schematic from www.4004.com

Intel Itanium (Montecito) 2004

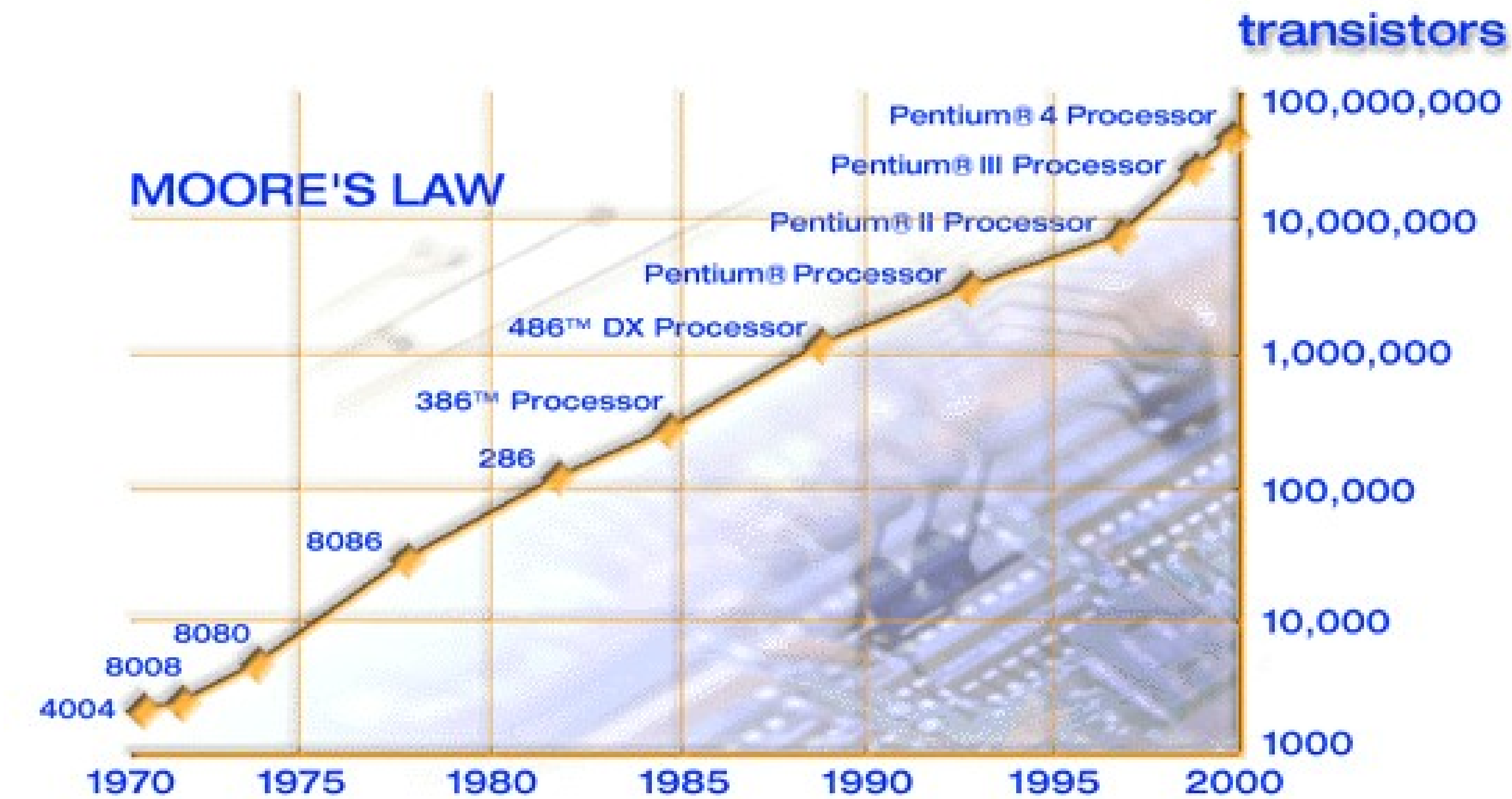


[from Best Servers of 2004, Microprocessor Report, January 2005.]

- 64-bit processor
- 1.7 billion transistors
- 1.7 GHz, issue up to 8 instructions per cycle
- 26 MByte of cache!!

In ~30 years, about 100,000 fold growth in transistor count and performance!

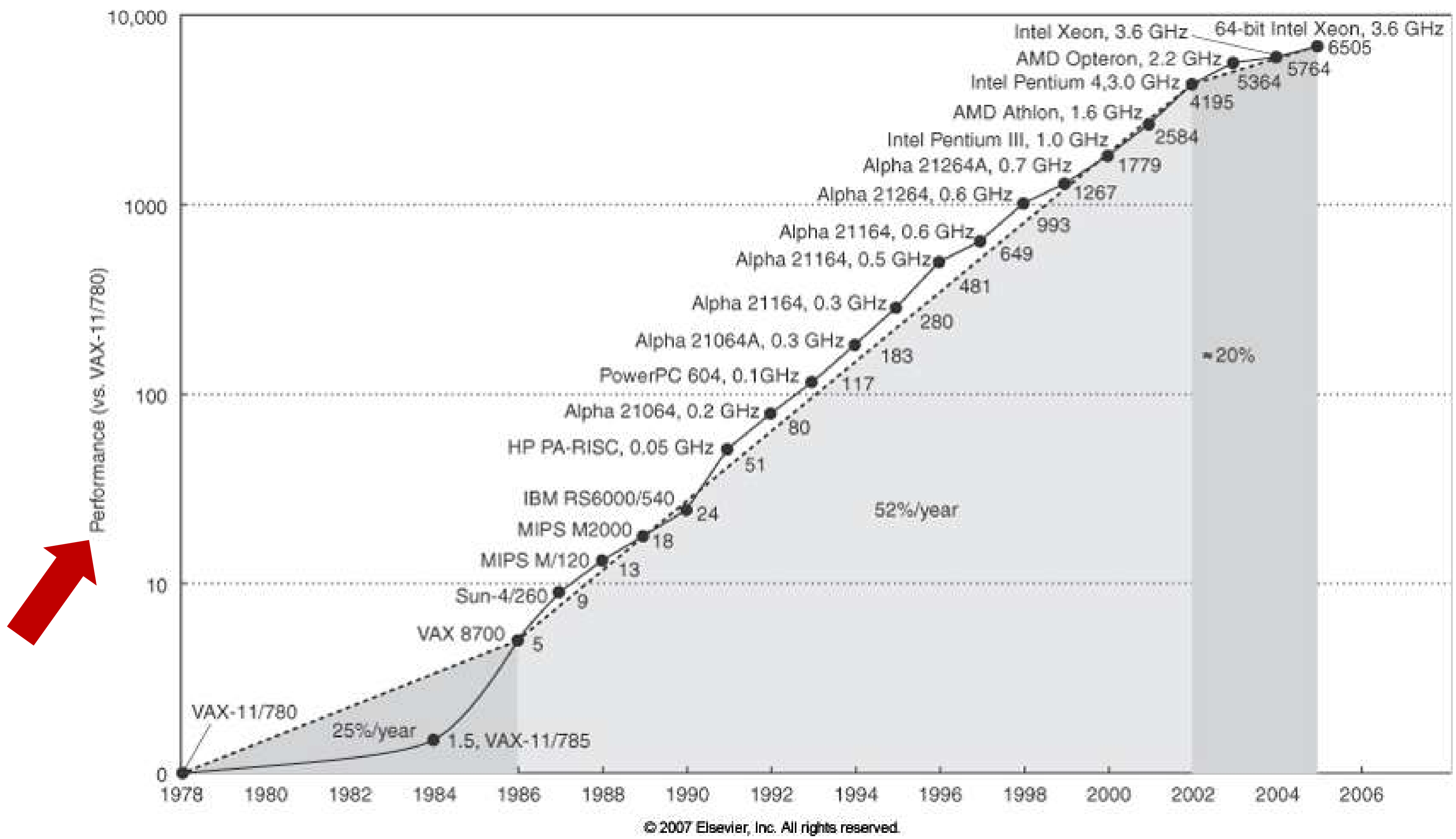
The Era of Moore's Law



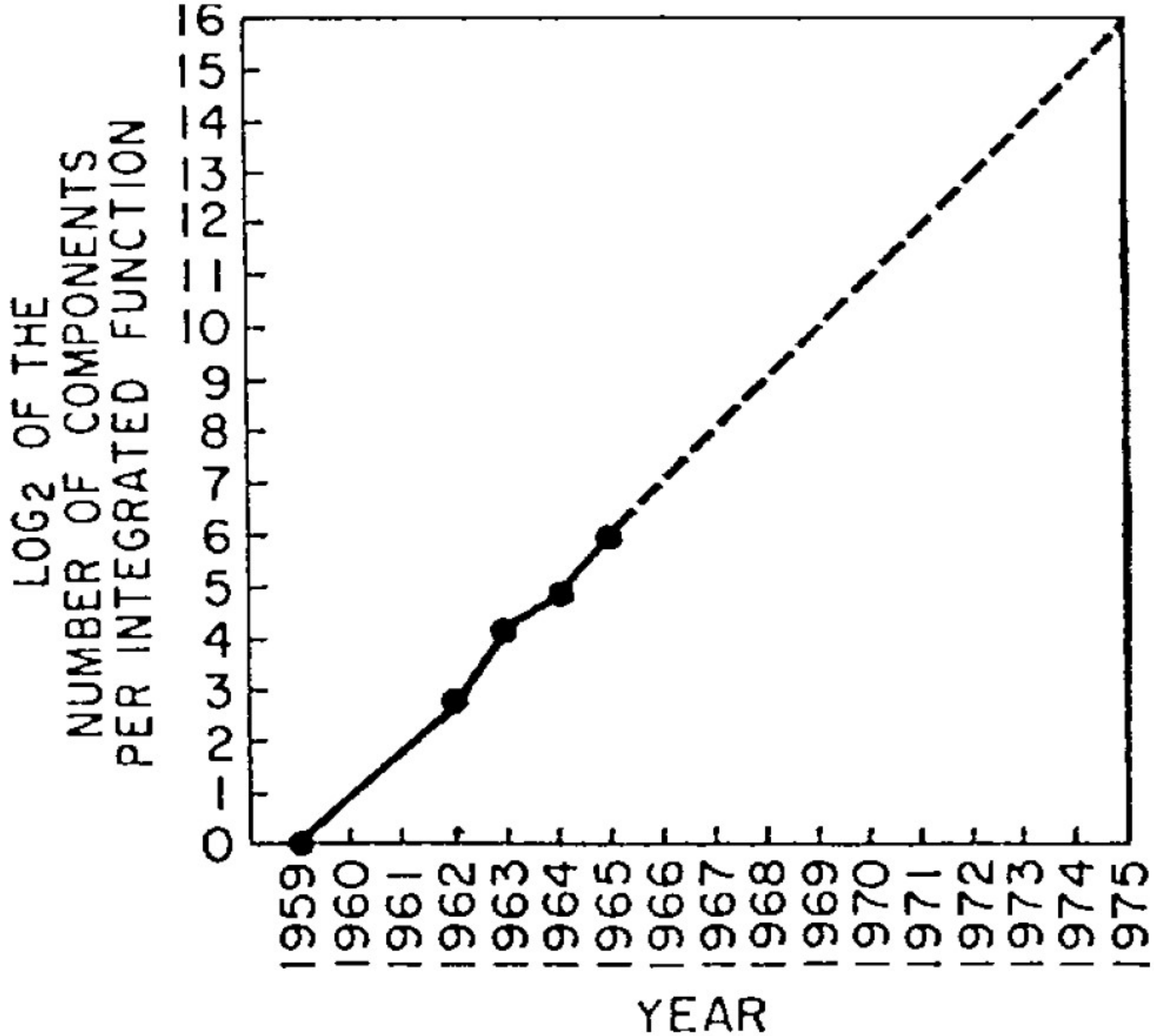
[http://www.intel.com/research/silicon/mooreslaw.htm]

Original article at http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=658762

The "Other" Moore's Law

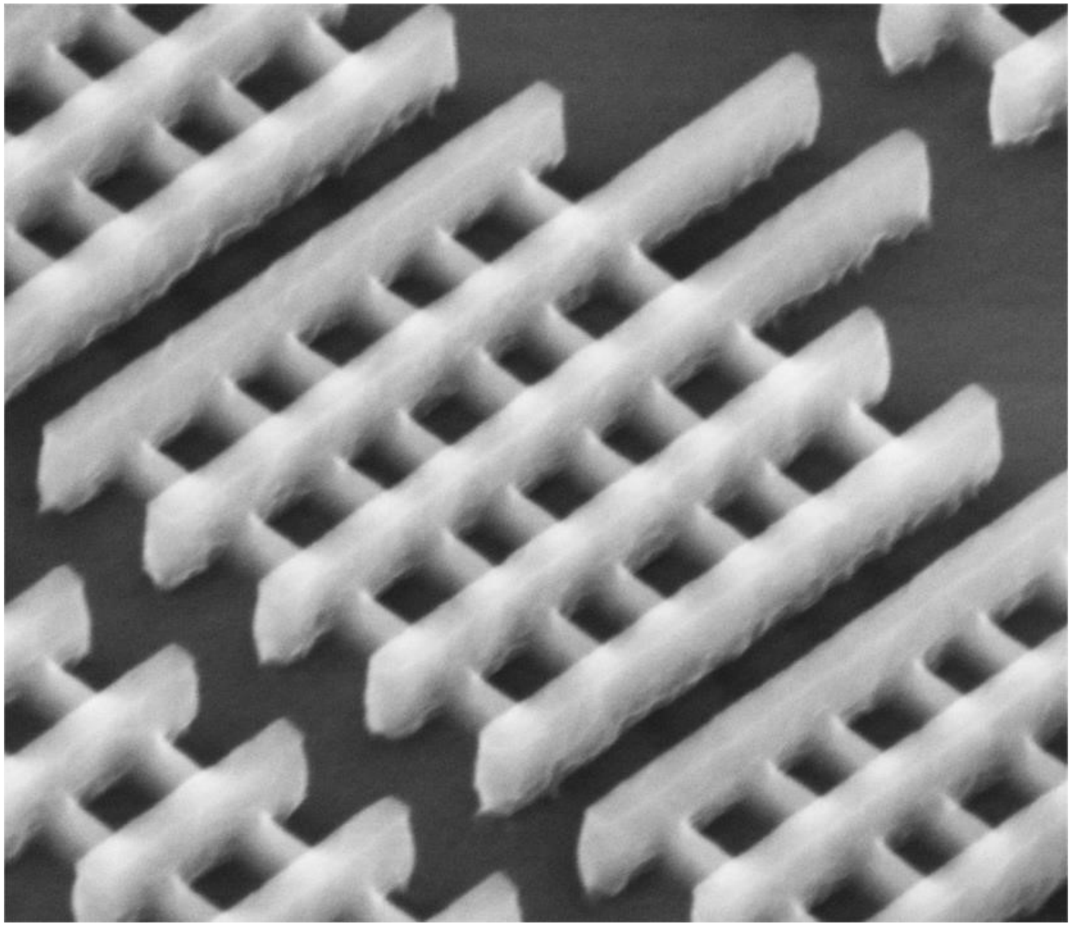


The Actual Moore's Law



[Cramming More Components Onto Integrated Circuits, G. E. Moore, 1965]

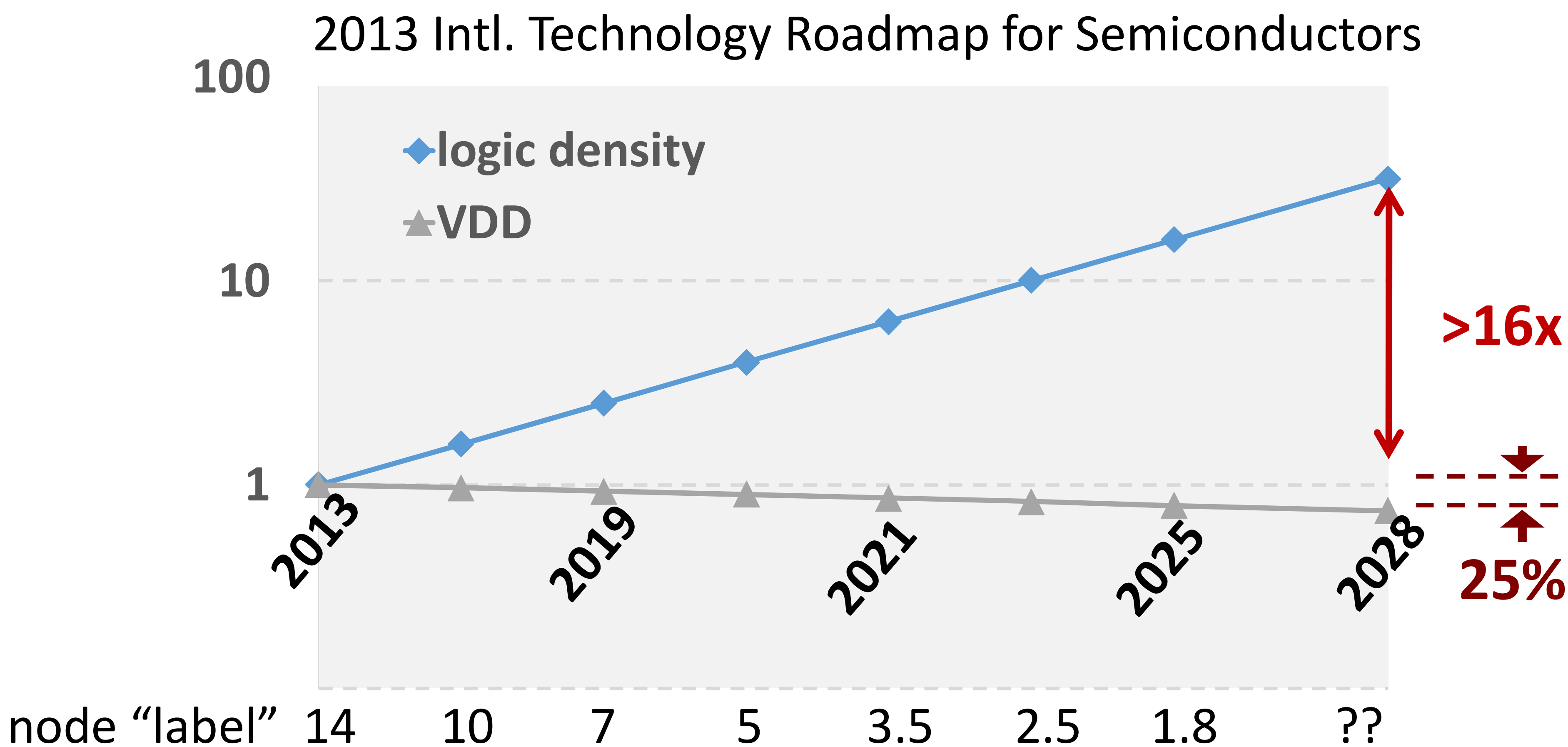
The End of Moore's Law?



[Tri-gate FinFET, Intel Newsroom, 2015]

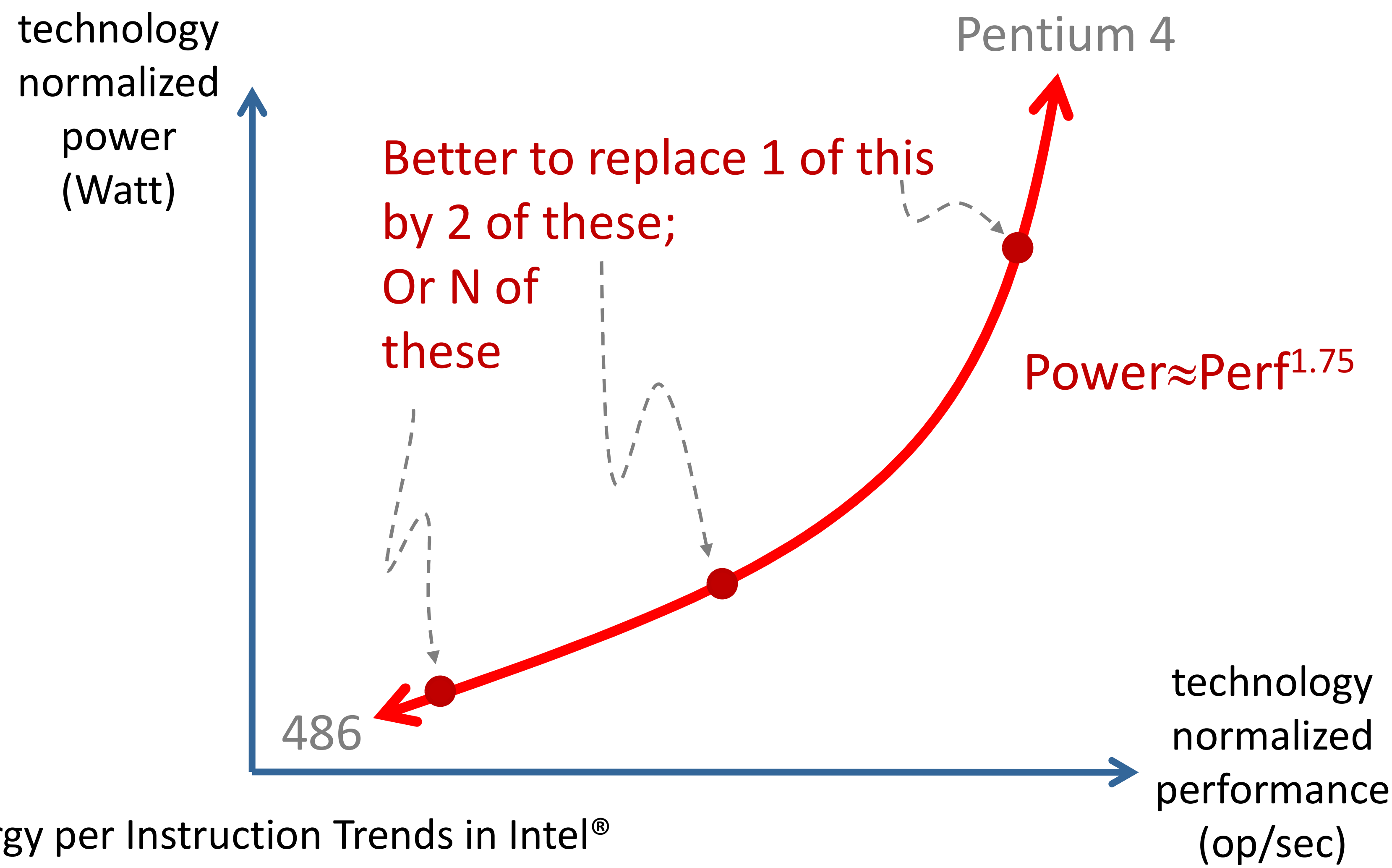
Distance between silicon atoms ~ 250 pm

Moore’s Law without Dennard Scaling



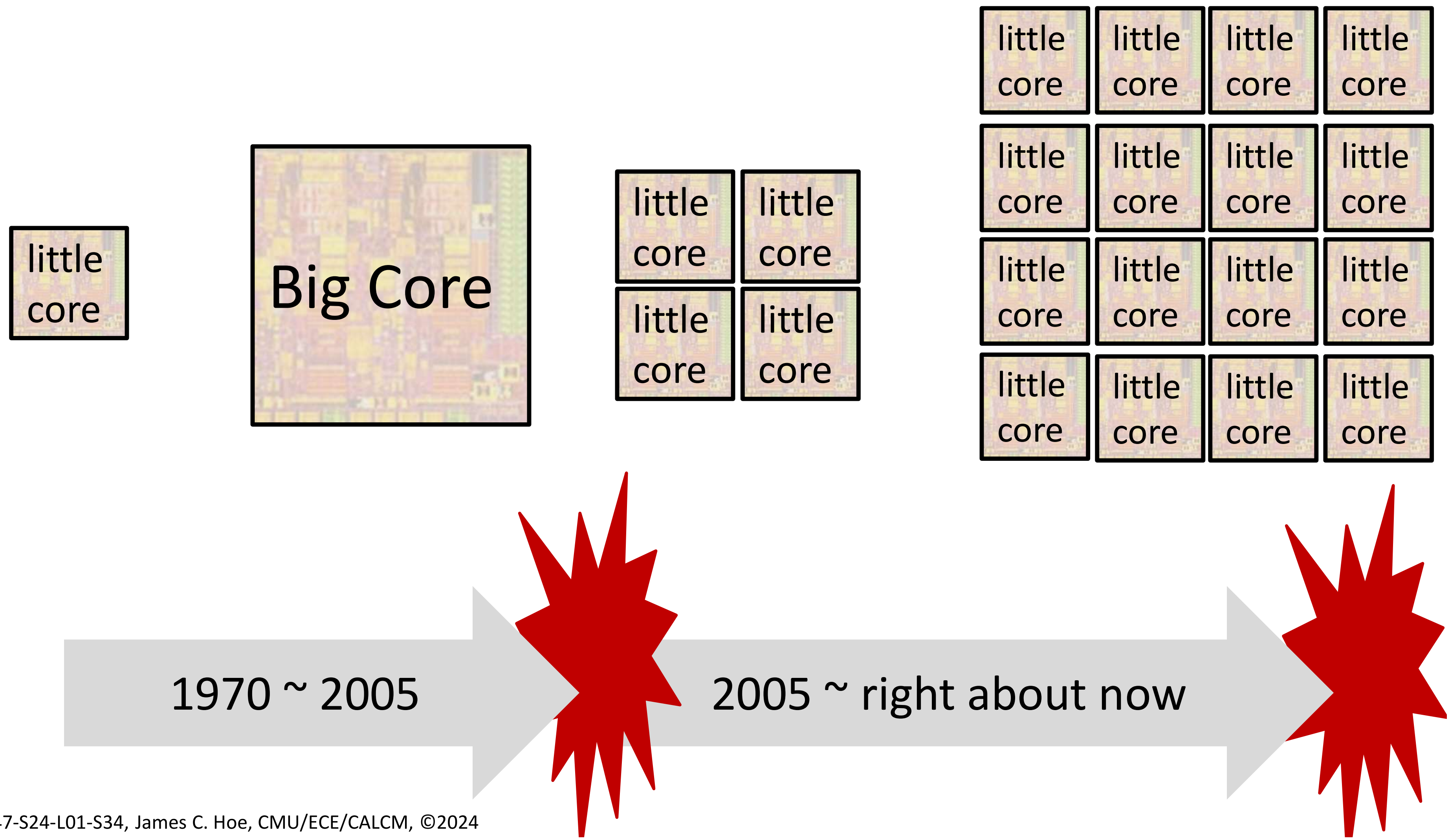
Under fixed power ceiling, more ops/second only achievable if less Joules/op?

Why multicores everywhere?

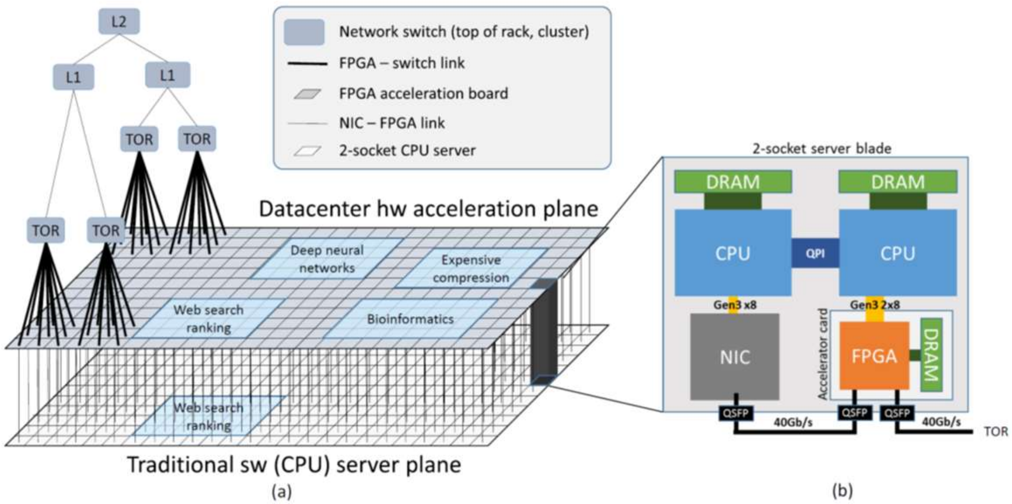


Energy per Instruction Trends in Intel®
Microprocessors, Grochowski et al., 2006

Moore’s Law Scaling with Cores

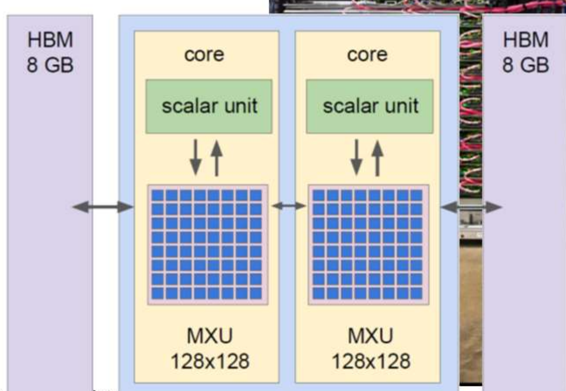


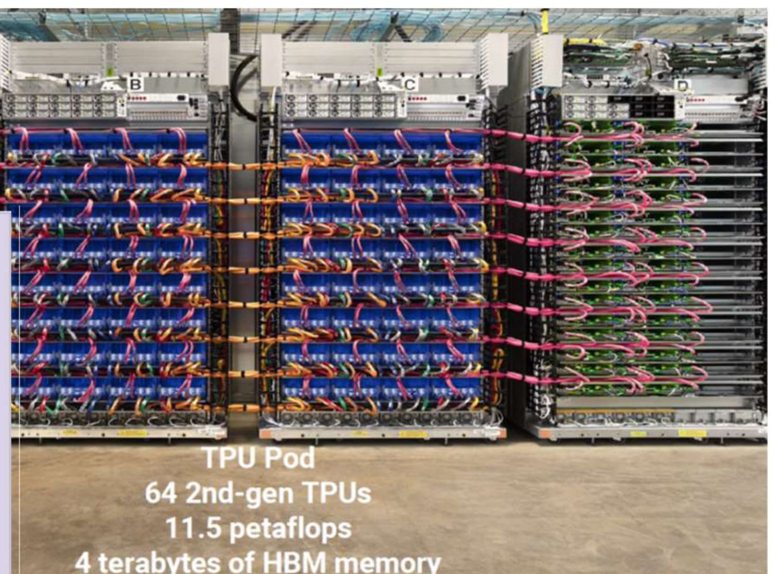
Computing's Brave New World



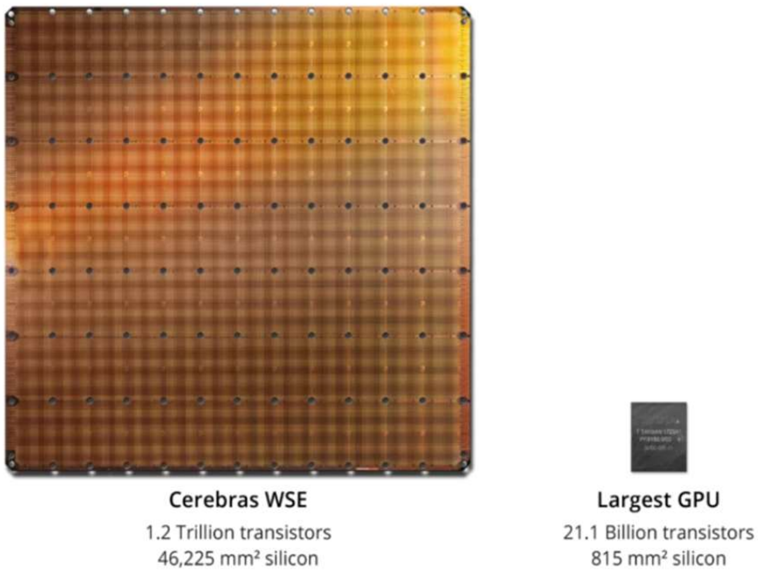
Microsoft Catapult [MICRO 2016, Caulfield, et al.]

Google TPU [Hotchips, 2017, Jeff Dean]





Computing's Brave New World

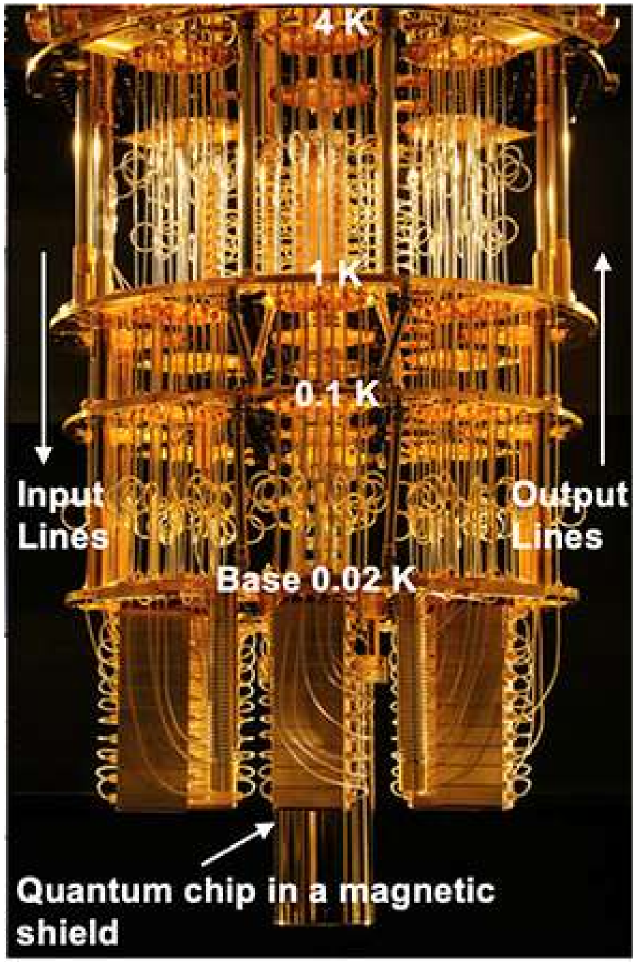


[Cerebras Wafer-Scale AI Processor]

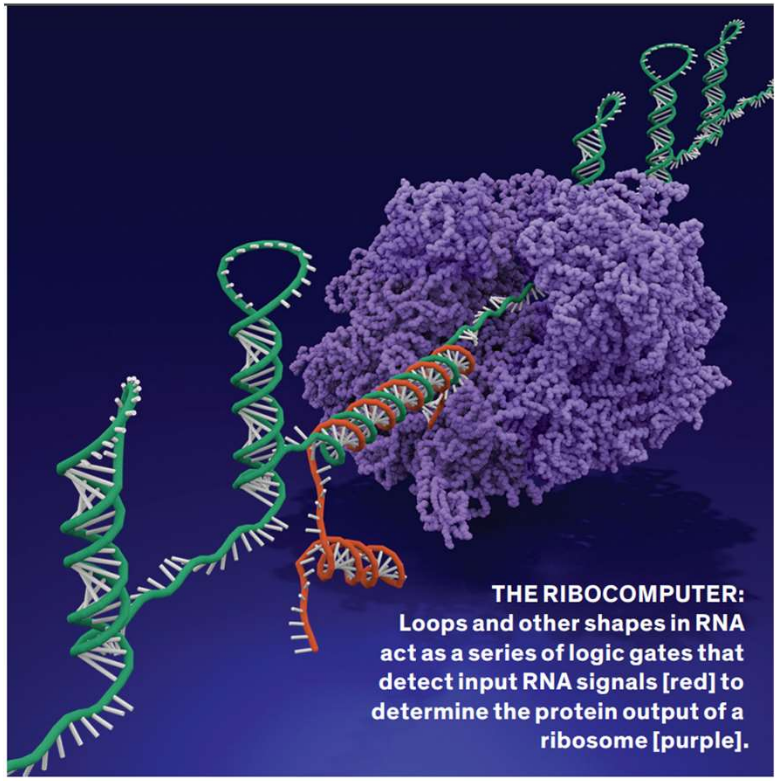
Rank	System	Cores	Rmax (PFlop/s)	Rpeak (PFlop/s)	Power (kW)
9	**Selene** - NVIDIA DGX A100, AMD EPYC 7742 64C 2.25GHz, NVIDIA A100, Mellanox HDR Infiniband, Nvidia NVIDIA Corporation United States	555,520	63.46	79.22	2,646

[Selene AI Supercomputer]

New Frontiers

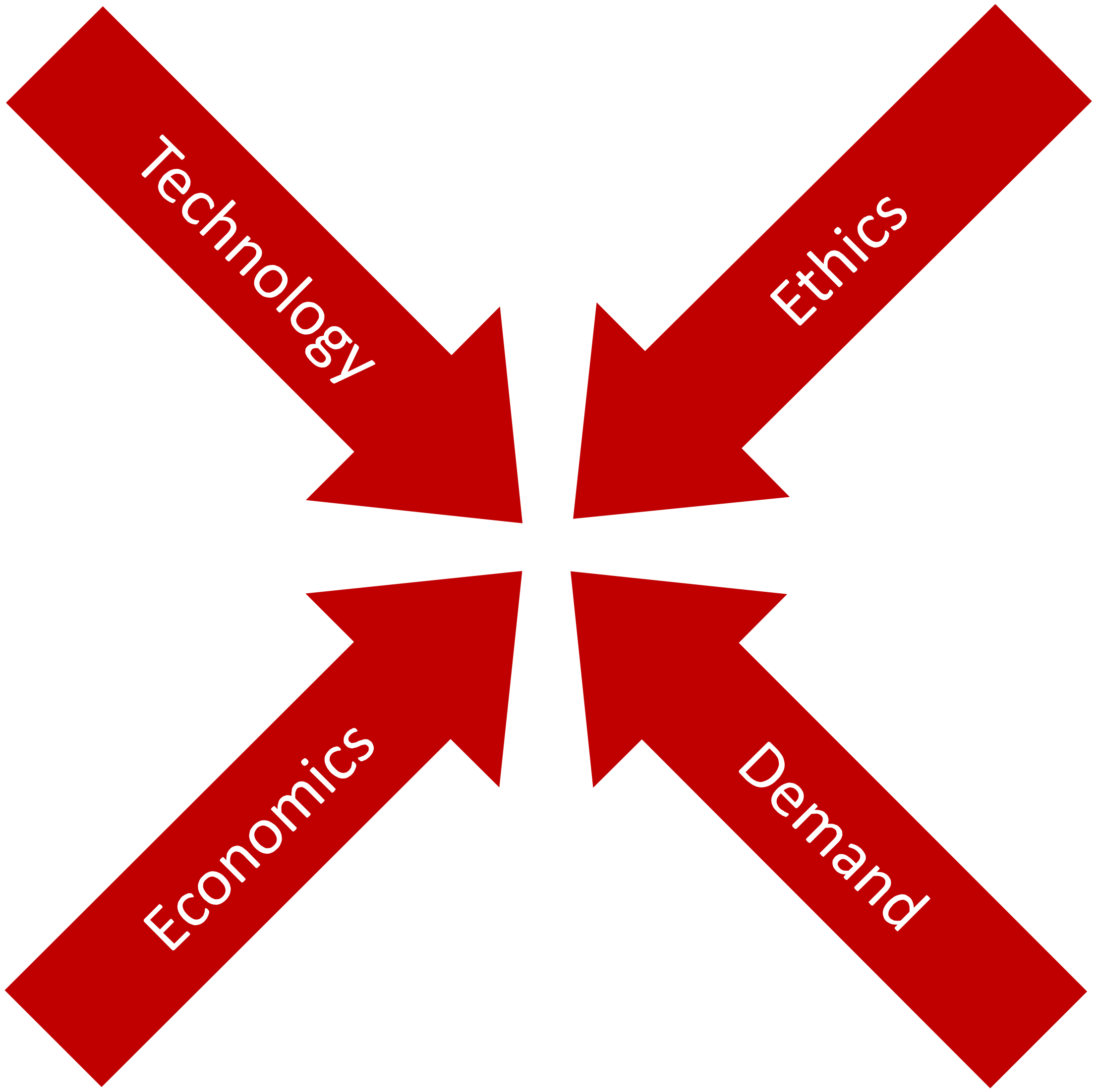


[IBM Q]



[IEEE Spectrum, Biocomputer and Memory Built Inside Living Bacteria]

Forces on Innovation



Where do we go from here?

(http://www.ece.cmu.edu/~ece447/schedule.html)

Course Logistics

- Please pay attention to Canvas and Piazza for updates and announcements
- H01: Syllabus
  - this is our contract for the term
  - please read it
- Lecture schedule online
  - http://www.ece.cmu.edu/~ece447/schedule.html
  - reading assignments are to be completed before lecture
  - pay attention to midterm date; the time to resolve conflicts is right now