What does the Future Hold?

CS 3410: Computer System Organization & Programming

Spring 2025

[K. Bala, A. Bracy, E. Sirer, Z. Susag, and H. Weatherspoon]

Logistics

Final Exam is on Sat. May 17 at 2pm in Statler Auditorium (STL185)

- Review Sessions …

A13: Raycasting due tonight (5/6) at 11:59pm!

Course evaluations are out now! Please complete them!!

- Due Friday, May 9 at 11:59pm
- Part of the *Survey* portion of your final grade

Big Pic

TECH

Satya Nadella says as much as 30% of Microsoft code is written by AI

PUBLISHED TUE, APR 29 2025·9:33 PM EDT | UPDATED TUE, APR 29 2025·9:58 PM EDT

Jordan Novet @JORDANNOVET | Jonathan Vanian @IN/JONATHAN-VANIAN-B704432/

KEY POINTS

- Microsoft CEO Satya Nadella on Tuesday said that as much as 30% of the company's code is now written by artificial intelligence.
- Nadella made the comments during a conversation before a live audience with Meta CEO Mark Zuckerberg at the social media company's LlamaCon AI developer event.
- Zuckerberg said Meta is focused on developing an AI model that can in turn build as much as half of other AI models within the next year.

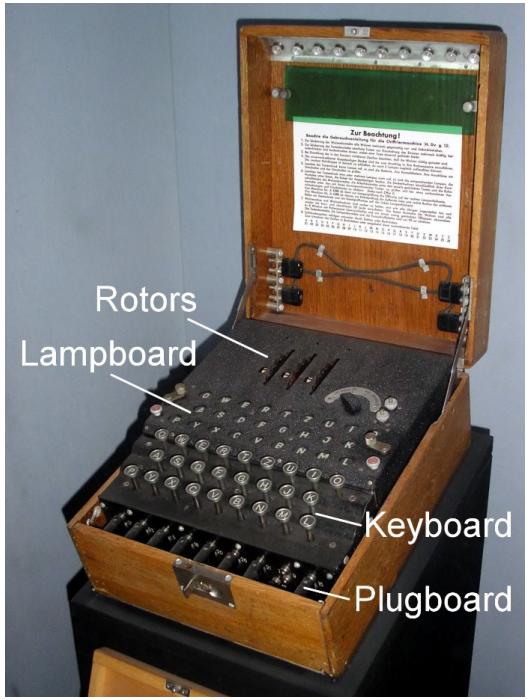
Facebook's CEO Mark Zuckerberg (L) speaks with Microsoft's CEO Satya Nadella after posing for a family picture with guests who attend the "Tech for Good" Summit at the Elysee Palace in Paris, on May 23, 2018.
Charles Platiau | AFP | Getty Images

“Sometimes it is the people that no one imagines anything of who do the things that no one can imagine”

-- The Imitation Game

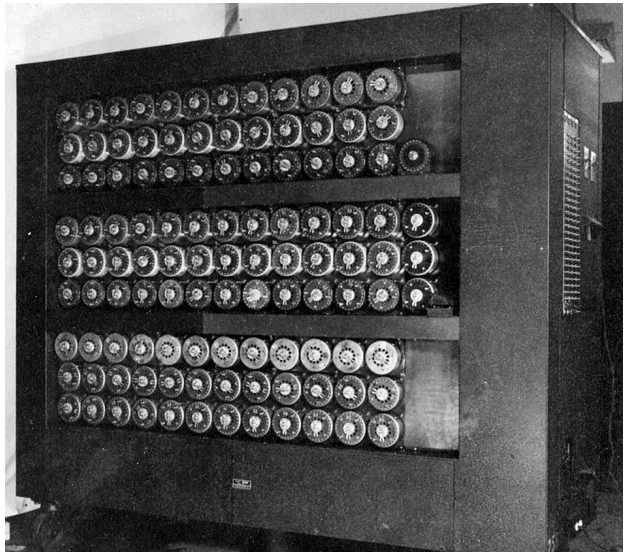
The Enigma Machine & The Bombe

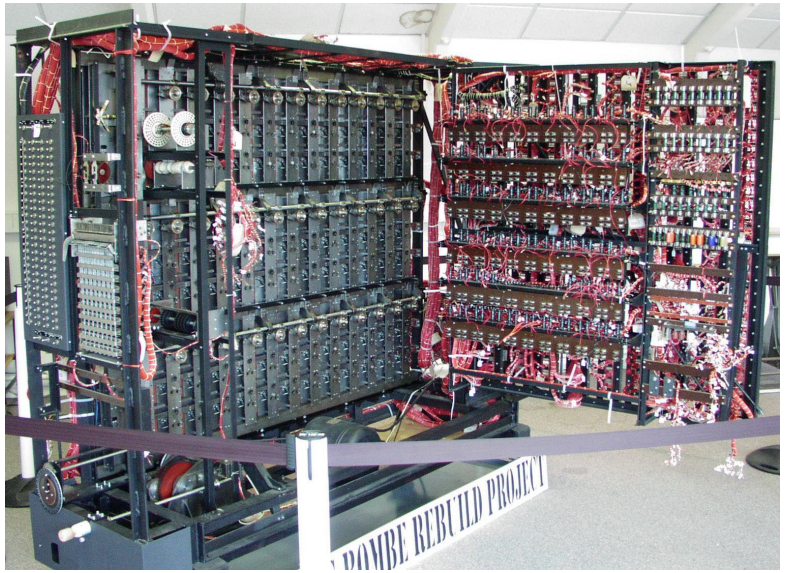
Enigma Machine



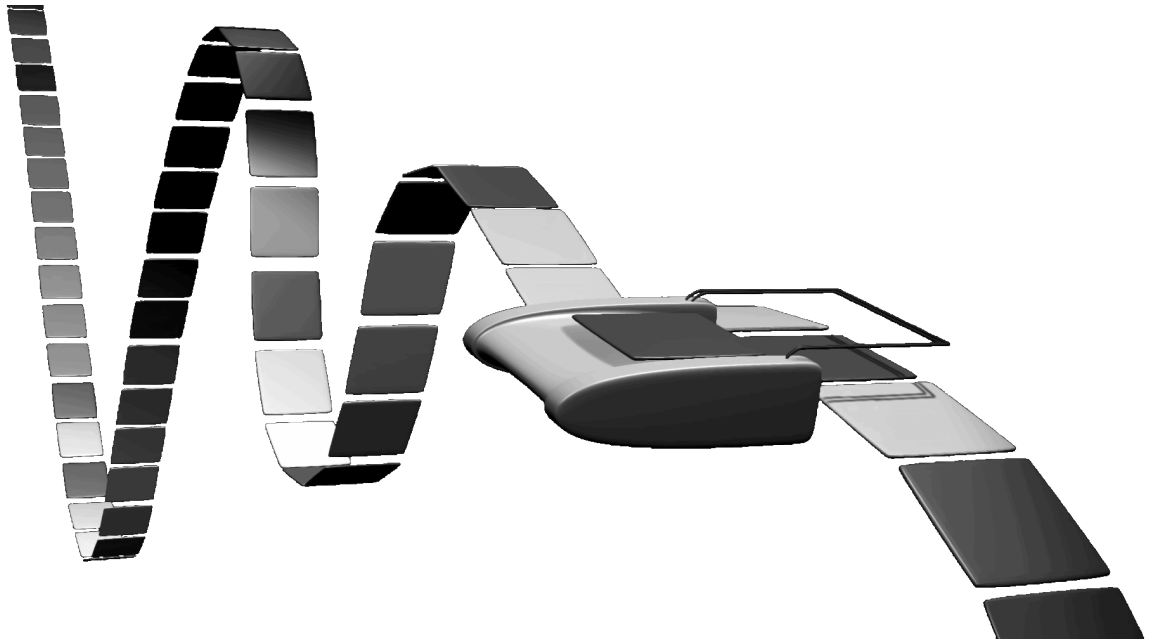
Used by the Germans during World War II to encrypt and exchange secret messages

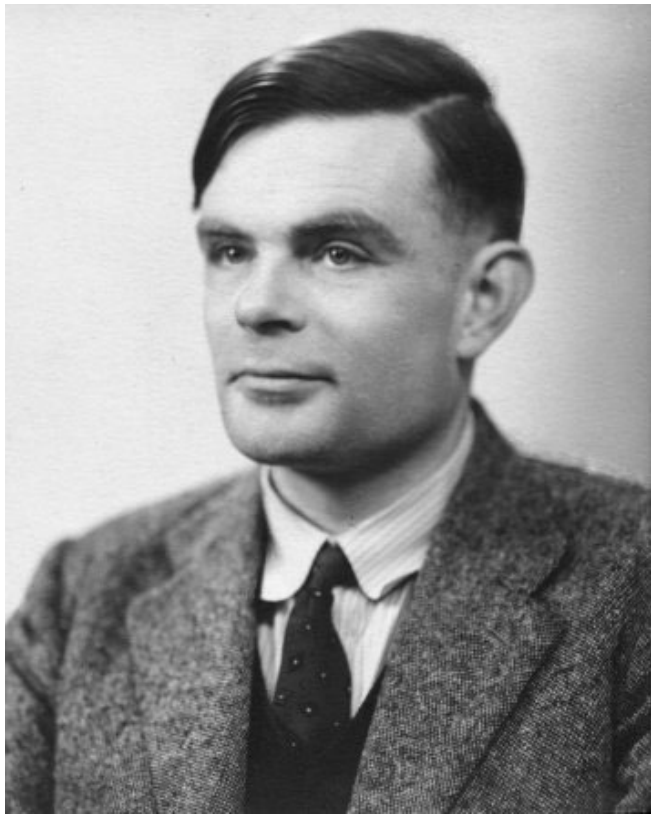
The Bombe





Used by the Allies to break the German Enigma machine during World War II





Turing Machine
1936

= abstract model for CPU that can simulate any algorithm

Alan Turing

The Big Picture

What is this?

```
#include <stdio.h>

int main() {
   printf("Hello world!\n");
   return 0;
}
```

How does it work?

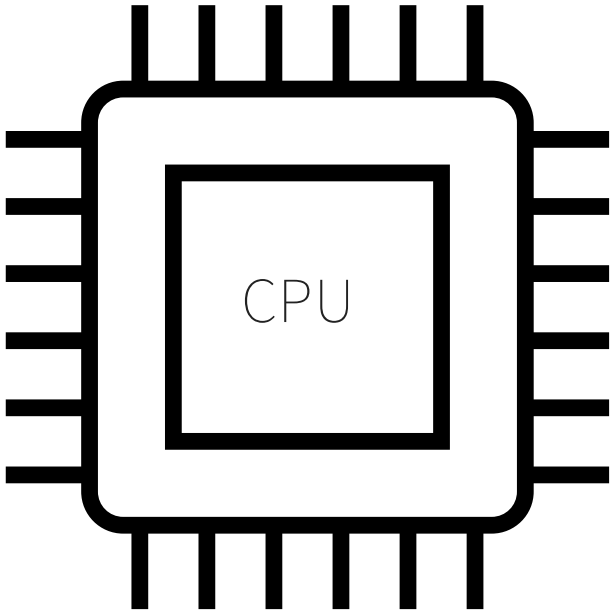
I'm glad you asked…

15 weeks later and you'll know!

"I know Kung Fu."

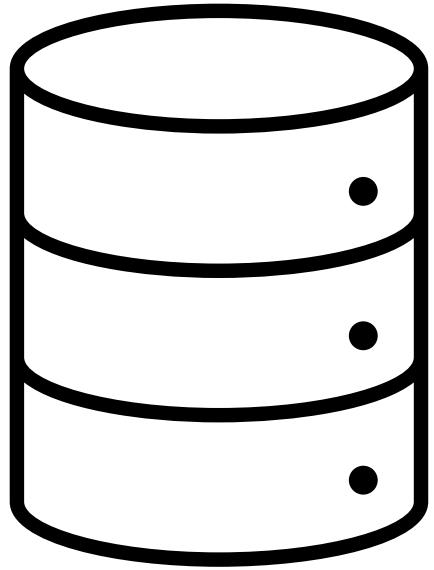
Big Picture

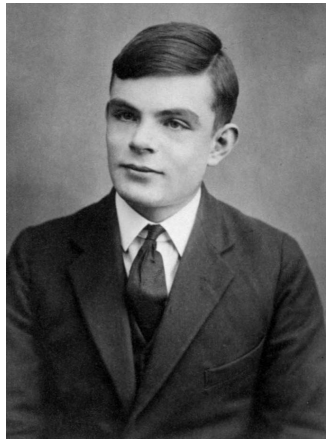
Processor



- ✓ Runs code; does computations
- ✗ Doesn't remember anything

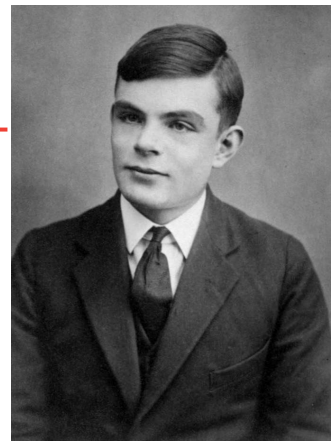
Memory



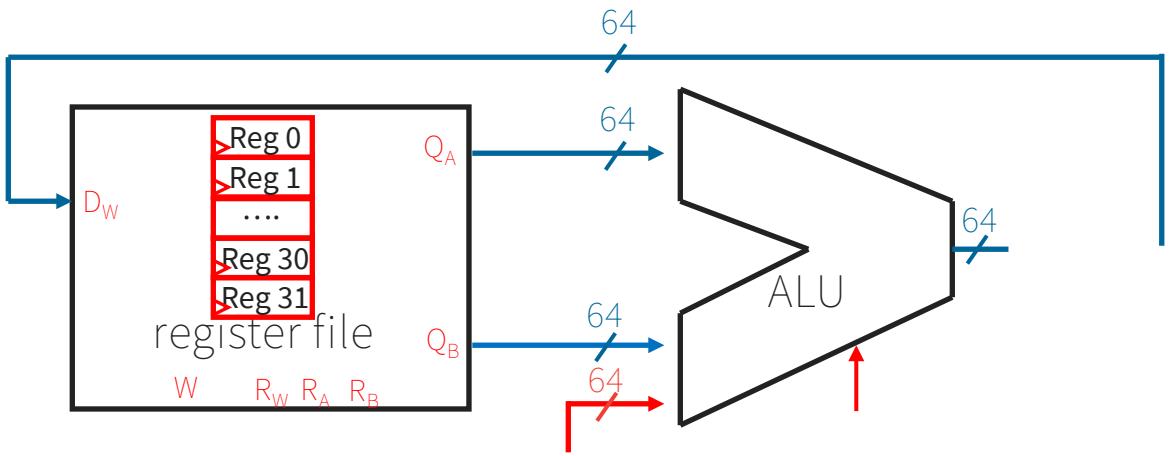


- ✗ Can't compute anything
- ✓ Stores data

Big Picture

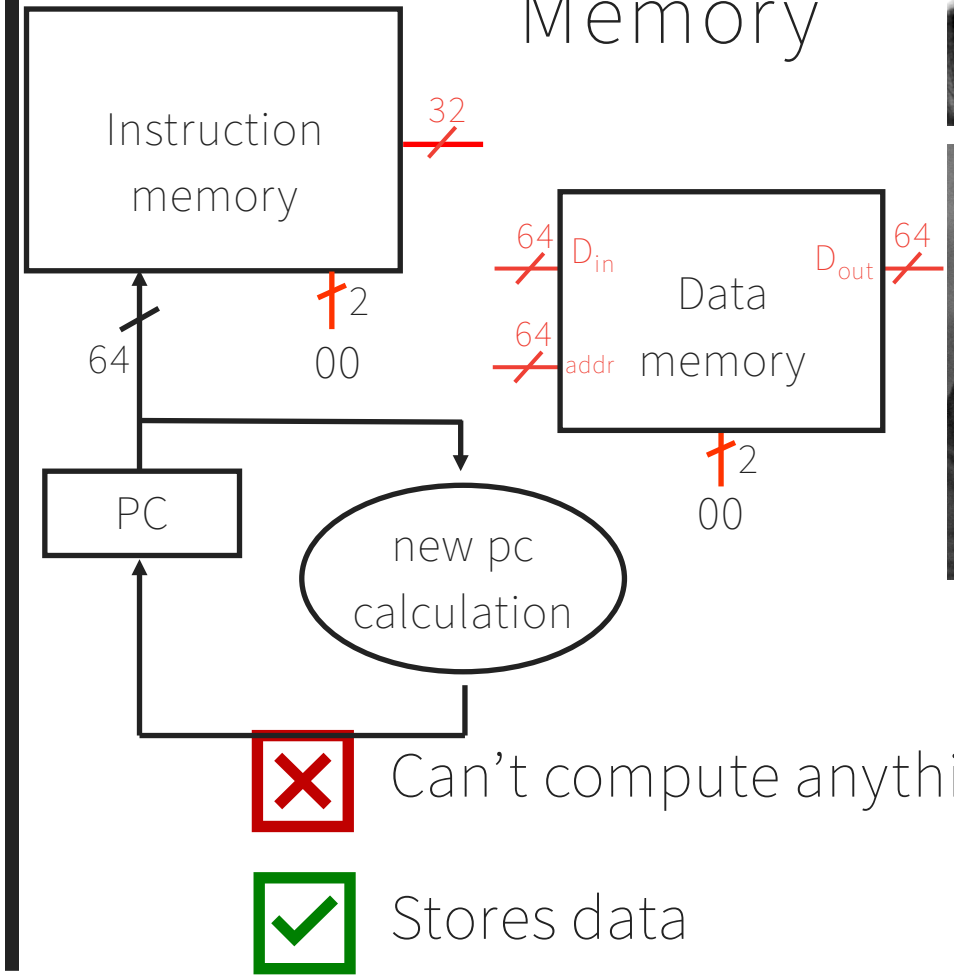


Processor



- ✓ Runs code; does computations
- ✗ Doesn't remember anything

Memory



- ✗ Can't compute anything
- ✓ Stores data

C CS 3410 in 1 slide

compiler

RISC-V assembly

assembler

machine code

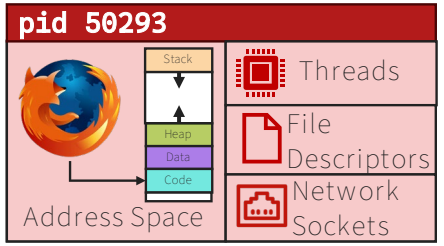
loader

```
int x = 10;
x = 2* x + 15;
```

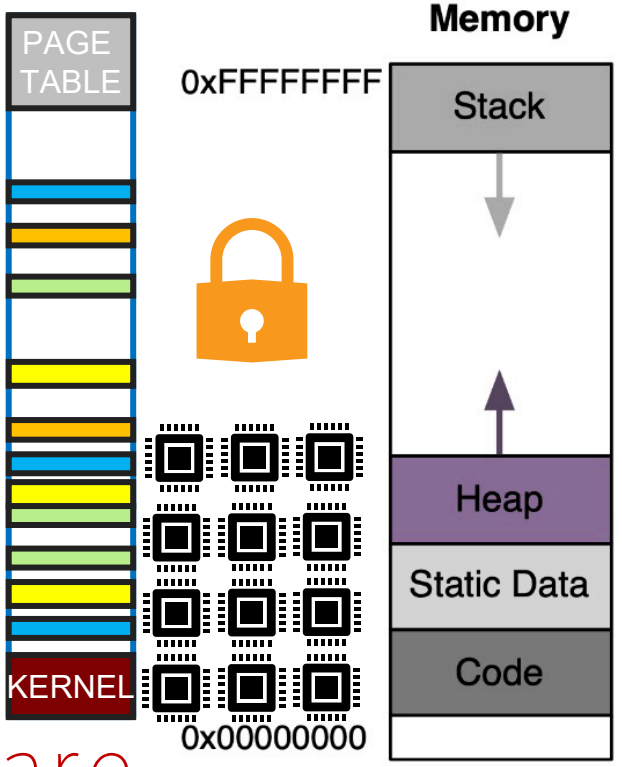
```
addi x5, x0, 10
muli x5, x5, 2
addi x5, x5, 15
```

```
00000000101000000000001010010011
00000000001000101000001010000000
00000000111100101000001010010011
```

Calling Conventions



System calls
Exceptions
Interrupts
I/O



Software

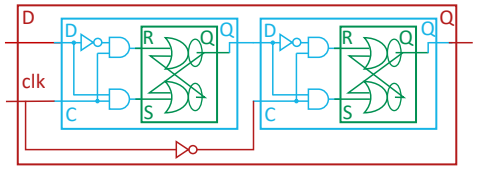
CPU

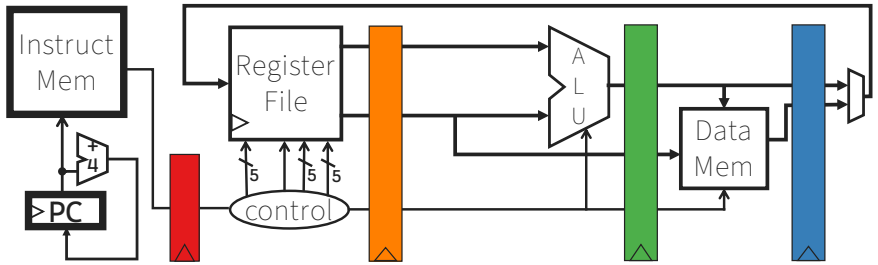
Circuits

Gates

Transistors

Silicon



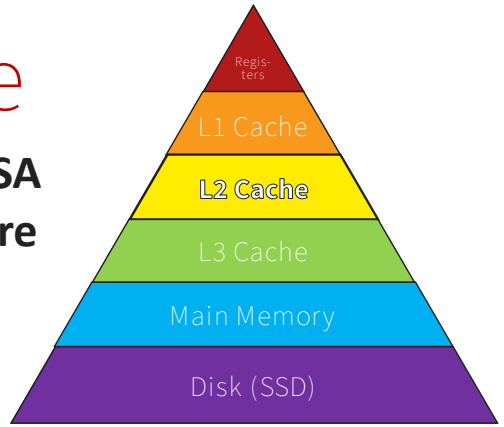


Hardware

Architecture -- ISA

Microarchitecture

- **Pipelining**
- **Caches**



More Than Moore

More computation, more storage, more data, more electricity, more parallelism, more concurrency, more heat, more networks, more domains, more architectures, more attack surfaces, more code, more bugs, more integration, more bits, more of more…

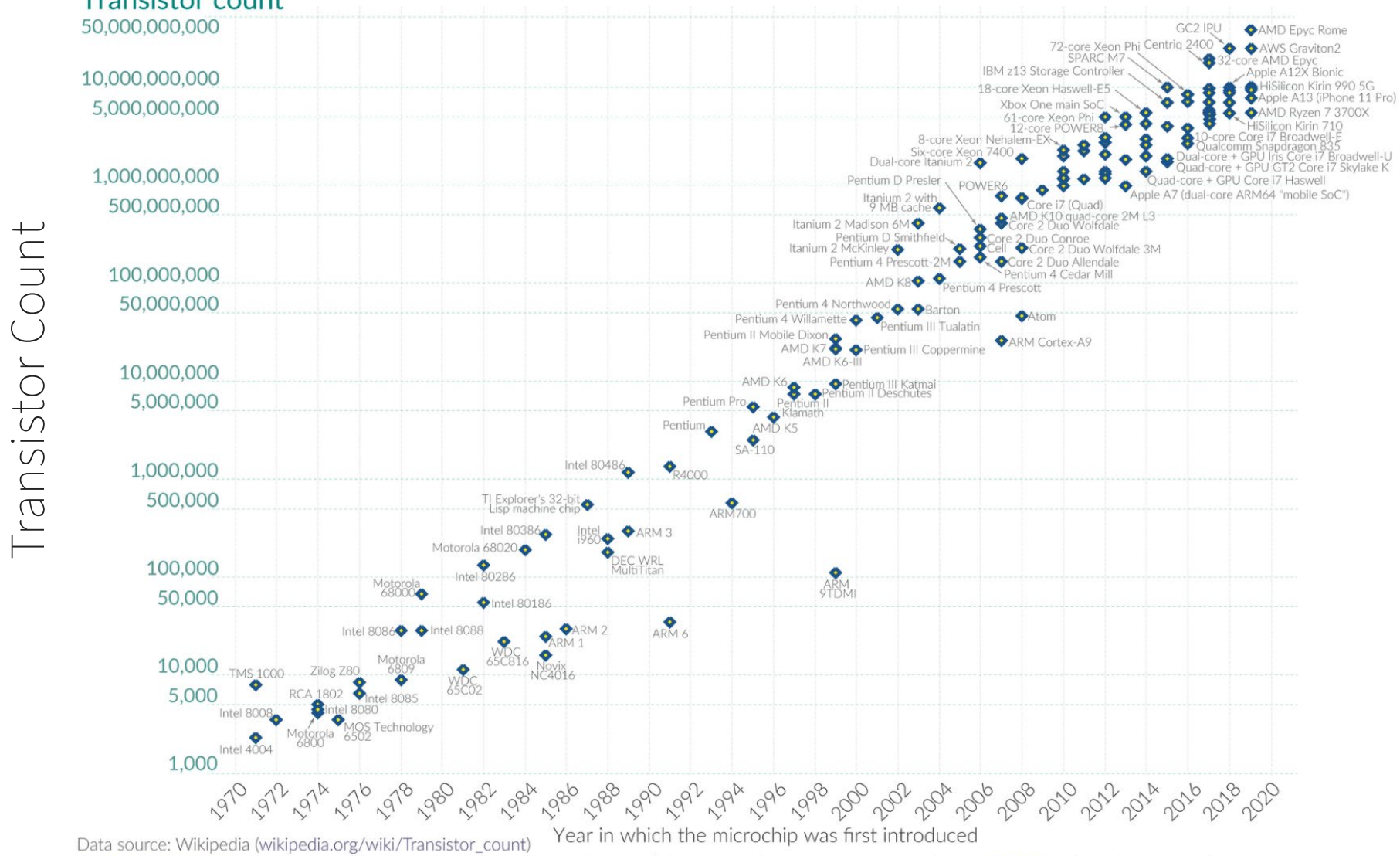
Moore's Law

- Moore's Law introduced in 1965
  - Number of transistors that can be integrated on a single die would double every 18 to 24 months (i.e., grow exponentially with time).
- Amazingly visionary
  - 2300 transistors, 1 MHz clock (Intel 4004) - 1971
  - 16 Million transistors (Ultra Sparc III)
  - 42 Million transistors, 2 GHz clock (Intel Xeon) – 2001
  - 55 Million transistors, 3 GHz, 130nm technology, 250mm2 die (Intel Pentium 4) – 2004
  - 290+ Million transistors, 3 GHz (Intel Core 2 Duo) – 2007
  - 731 Million transistors, 2-3Ghz (Intel Nehalem) – 2009
  - 1.4 Billion transistors, 2-3Ghz (Intel Ivy Bridge) – 2012
  - 7.2 Billion transistors, 3-3.9 GHz Intel Broadwell (22-core) – 2016

Moore's Law: The number of transistors on microchips doubles every two years

Moore's law describes the empirical regularity that the number of transistors on integrated circuits doubles approximately every two years.
This advancement is important for other aspects of technological progress in computing – such as processing speed or the price of computers.

Our World in Data



Data source: Wikipedia (wikipedia.org/wiki/Transistor_count)
OurWorldinData.org – Research and data to make progress against the world's largest problems.

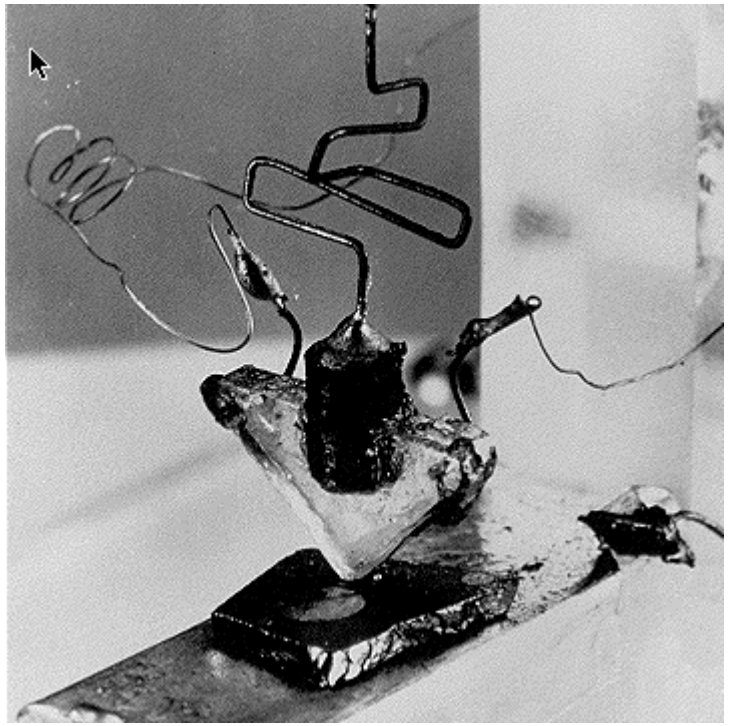
https://en.wikipedia.org/wiki/Moore%27s_law

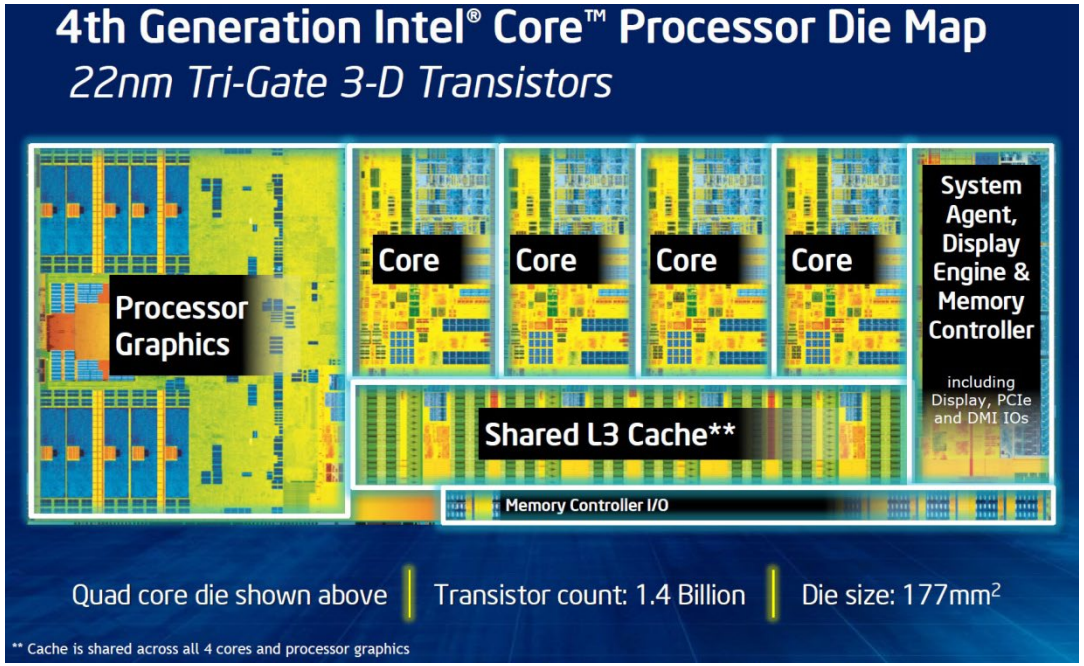
Why Multicore?

- Moore's law
  - A law about transistors
  - Smaller means more transistors per die
  - And smaller means faster too

- But: Power consumption growing too…

Multi-core



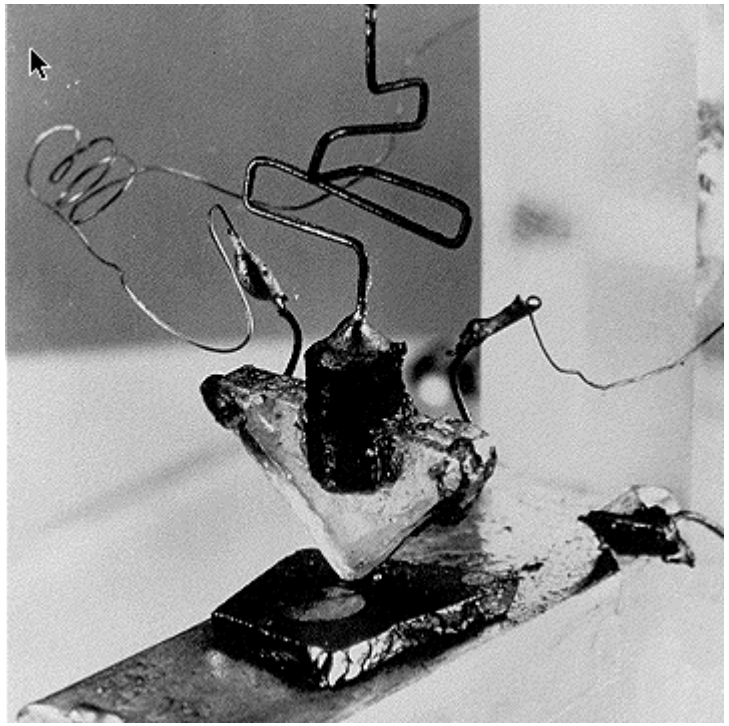


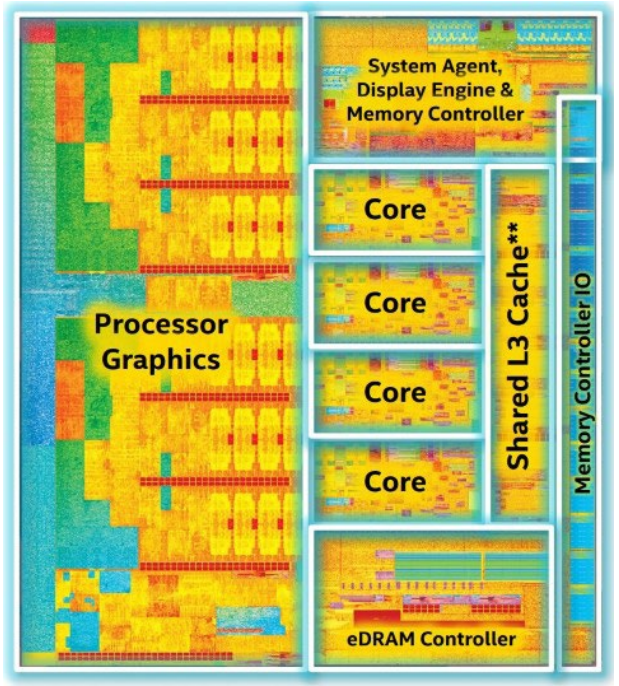
http://techguru3d.com/4th-gen-intel-haswell-processors-architecture-and-lineup/

- The first transistor
  - One workbench at AT&T Bell Labs
  - 1947
  - Bardeen, Brattain, and Shockley
- Intel Haswell
  - 1.4 billion transistors, 22nm
  - 177 square millimeters
  - Four processing cores

https://en.wikipedia.org/wiki/Transistor_count

Multi-core



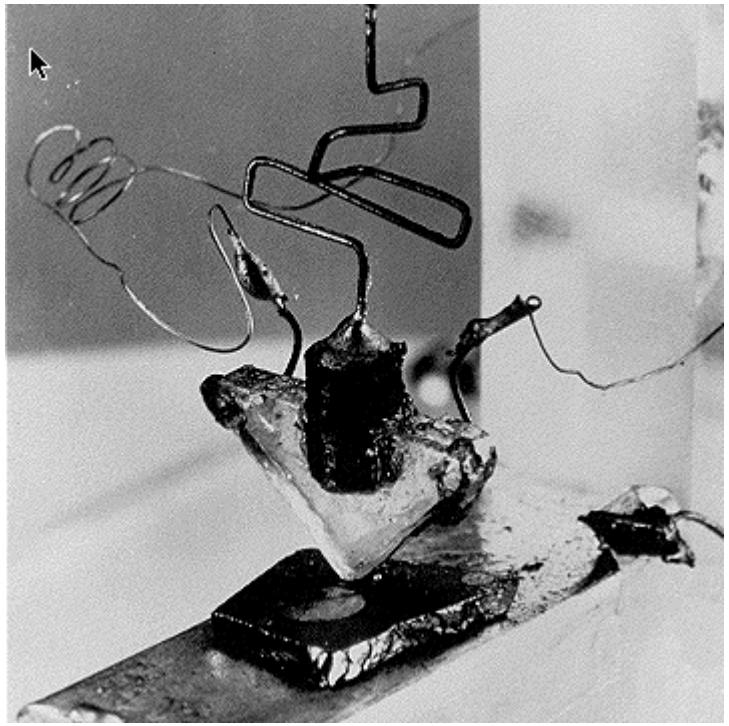


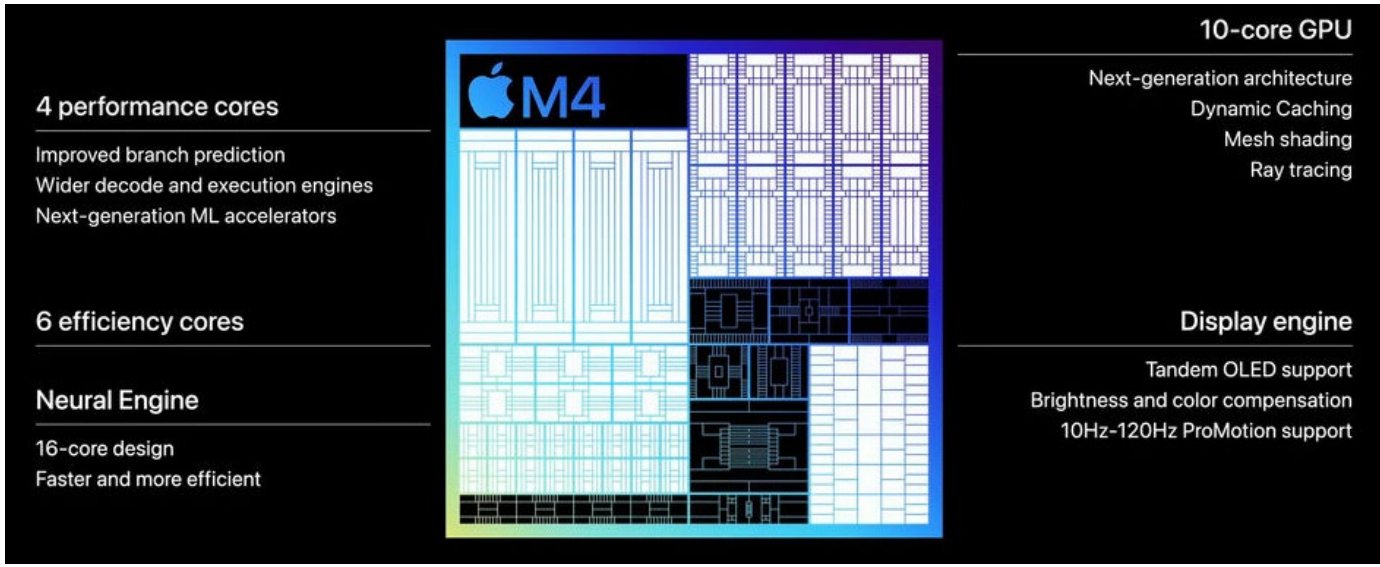
https://www.computershopper.com/computex-2015/performance-preview-desktop-broadwell-at-computex-2015

- The first transistor
  - One workbench at AT&T Bell Labs
  - 1947
  - Bardeen, Brattain, and Shockley
- Intel Broadwell
  - 7.2 billion transistors, 14nm
  - 456 square millimeters
  - Up to 22 processing cores

https://en.wikipedia.org/wiki/Transistor_count

Multi-core





- The first transistor
  - One workbench at AT&T Bell Labs
  - 1947
  - Bardeen, Brattain, and Shockley
- Apple M4
  - 28 billion transistors, 3nm
  - 177 square millimeters
  - 4x-10x performance, 4x-6x efficiency, 8x-40x GPU, 16x Neural processing cores

https://en.wikipedia.org/wiki/Transistor_count

https://en.wikipedia.org/wiki/Transistor_count
https://en.wikipedia.org/wiki/Apple_M4

Cloud Computing

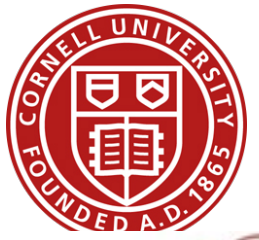
Cloud Computing

- The Cloud
  - A computer utility, warehouse computers; a commodity
  - Catalyst for technology economy
  - Revolutionizing artificial intelligence, machine learning, health care, financial systems, scientific research, and society

Cloud Computing

- The Cloud
  - *ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.*

NIST Cloud Definition

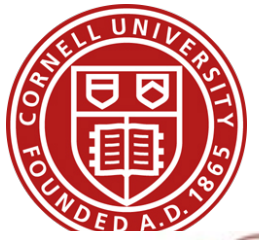


Windows Azure

Cloud Computing

- The Cloud
  - *ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.*

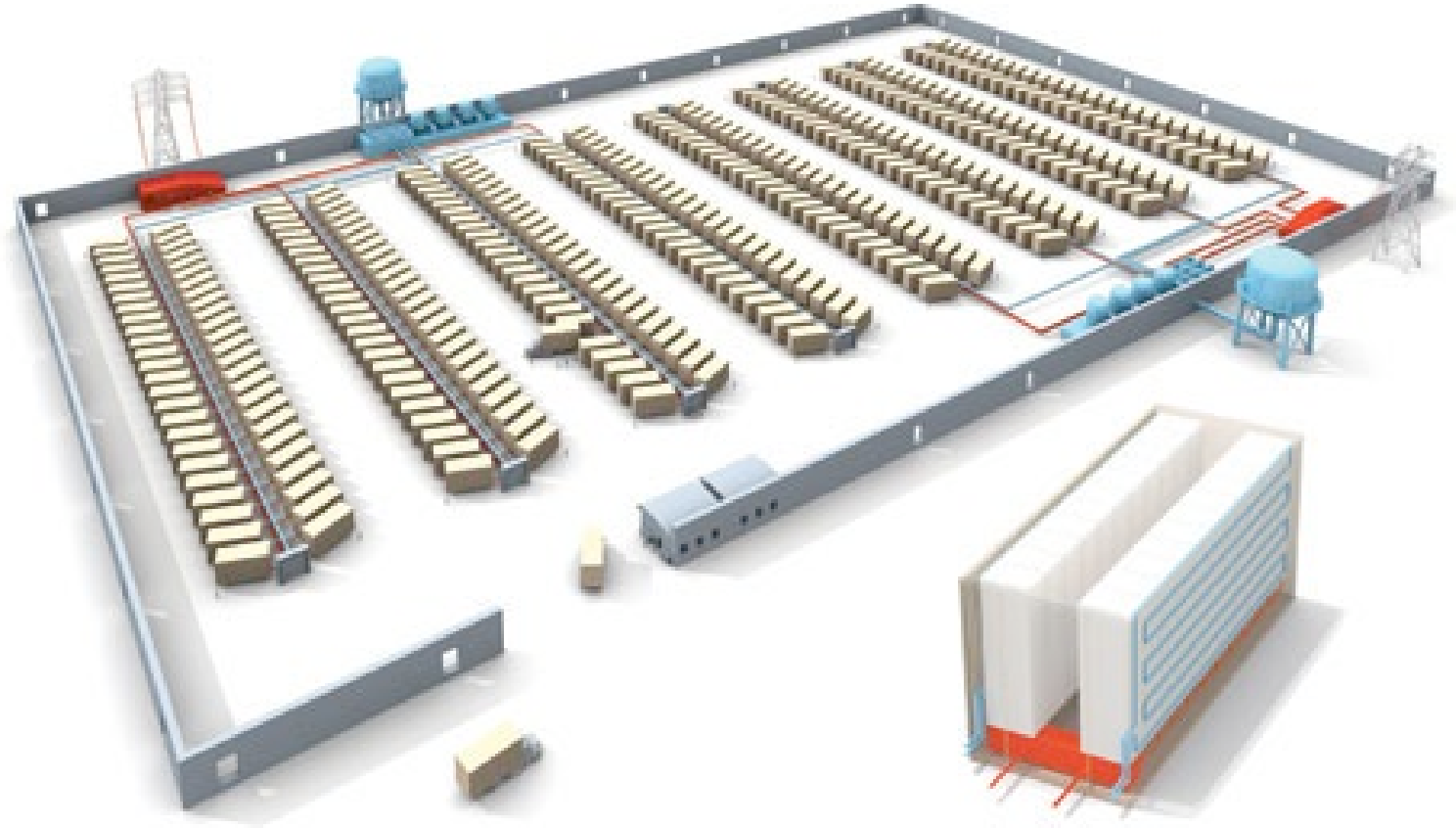
NIST Cloud Definition



Windows Azure

Cloud Computing

- How big is Big Data in the Cloud?
  - Exabytes: Delivery of petabytes of storage daily



Titan tech boom, randy katz, 2008

Cloud Computing

- How big is Big Data in the Cloud?
  - Most of the worlds data (and computation) hosted by few companies

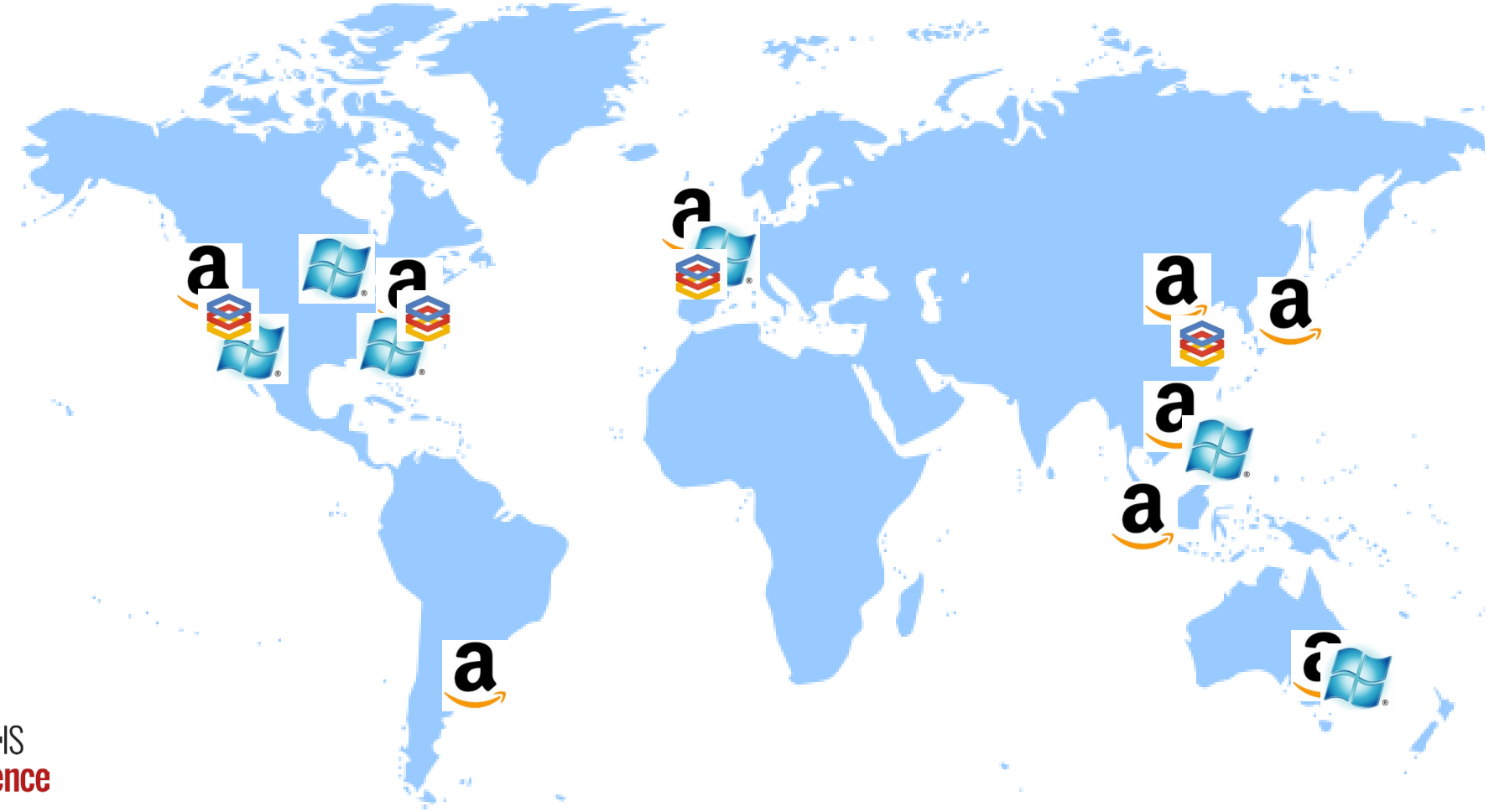
Cloud Computing

- How big is Big Data in the Cloud?
  - Most of the worlds data (and computation) hosted by few companies

- Currently 4.72 billion internet users
  - 900,000 new users each day [Hootsuite]
- Growing to 175 zettabytes in 2025
- 65% of this data will stored and processed in datacenters [IDC]

Cloud Computing

- How big is Big Data in the Cloud?
  - Most of the worlds data (and computation) hosted by few companies



Power Demands Rapidly Increasing

npr

NEWSLETTERS SIGN IN NPR SHOP DONATE

NEWS CULTURE MUSIC PODCASTS & SHOWS SEARCH

NATIONAL

Three Mile Island nuclear plant will reopen to power Microsoft data centers

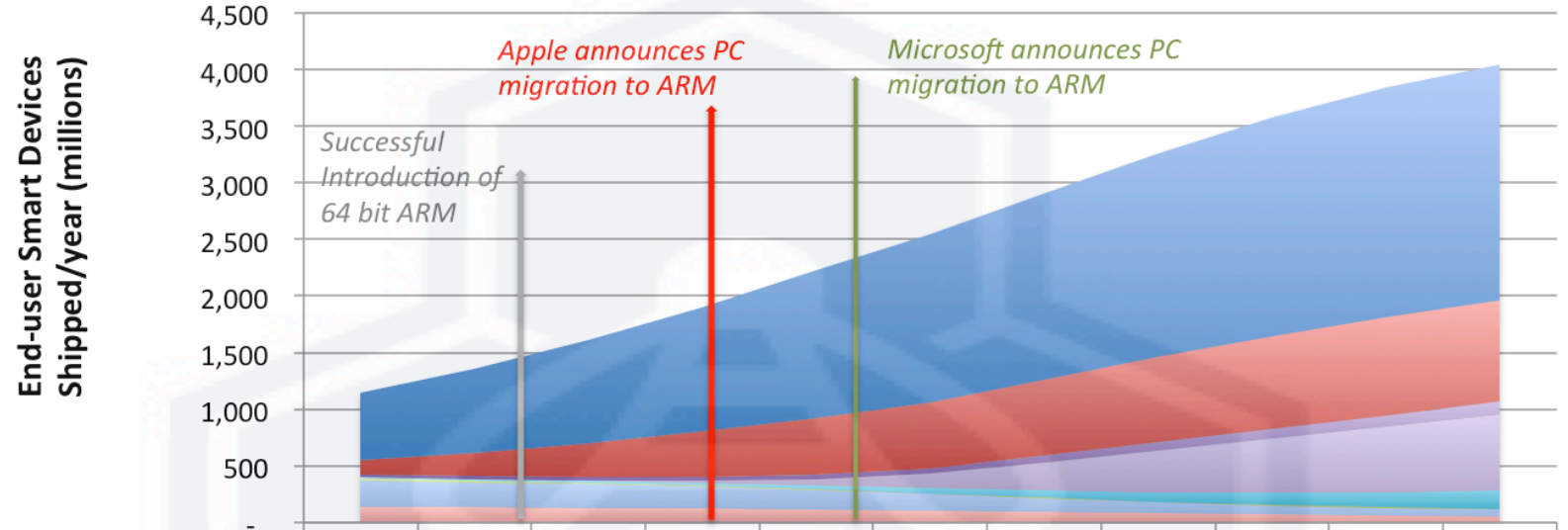
SEPTEMBER 20, 2024 · 1:40 PM ET

By C Mandler

Embedded Processors

Where is the Market?

End-user device shipments per year



	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
ARM Smartphones	604	743	906	1,097	1,305	1,488	1,651	1,800	1,926	2,022	2,083
ARM Tablets	121	200	300	405	502	583	670	750	818	867	893
ARM PCs, Chromebooks, etc.	22	25	29	34	39	46	55	70	85	100	115
ARM Smart Wearable Devices	-	-	5	17	50	124	248	371	483	579	666
Other ARM Devices	15	18	22	29	38	51	67	86	109	136	168
eReaders	20	15	12	9	7	6	5	5	4	4	3
x86 Laptop PCs	229	217	204	188	167	144	119	96	79	68	61
x86 Desktop PCs, Workstations	137	132	127	120	111	100	88	75	65	57	51
x86 Tablets	2	3	4	6	7	9	9	8	6	4	3
Ratio ARM/x86	2.1	2.8	3.8	5.1	6.8	9.1	12.5	17.3	22.8	28.6	34.3

Year

Source: © Wikibon 2013, IDC & Gartner 2013 shipments & forecast, & Wikibon 2014-2022 projections. Assumption: Apple & Microsoft migrate to 64-bit ARM

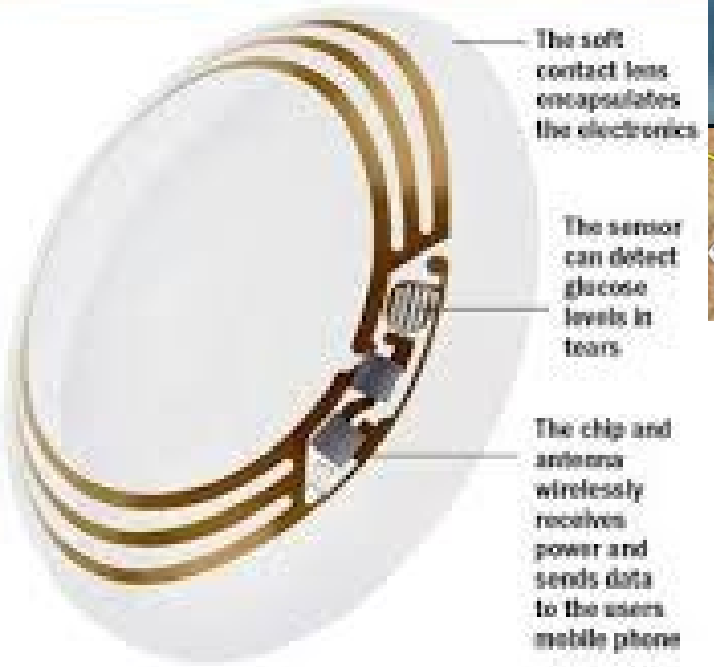
http://wikibon.org/wiki/v/64-bit_ARM_Ushers_in_the_Mobile-Cloud_Application_Model

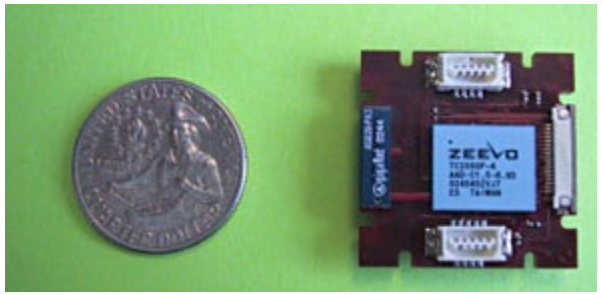
Internet of Things (IoT)

- E.g. Cornell Institute for Digital Agriculture (CIDA)

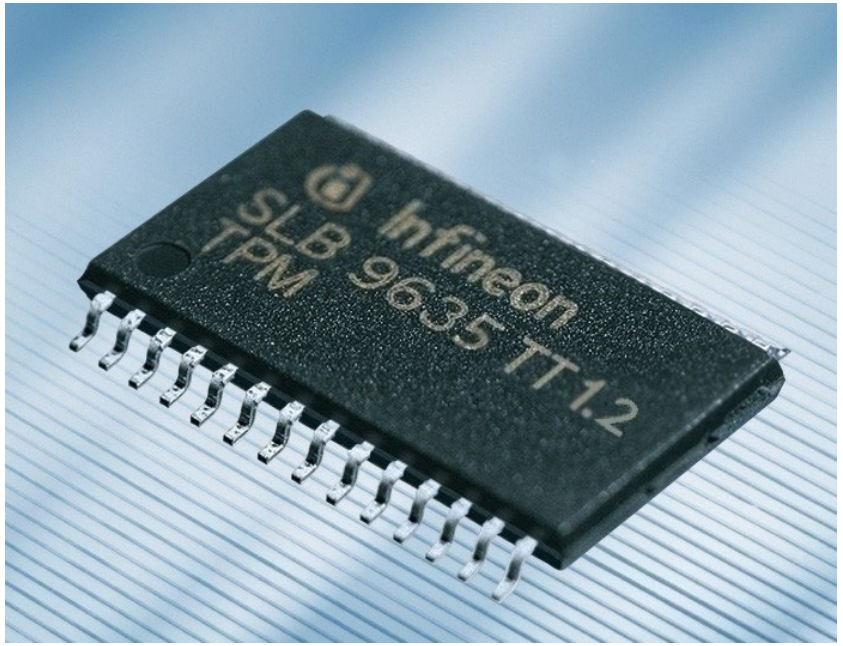
Where to?

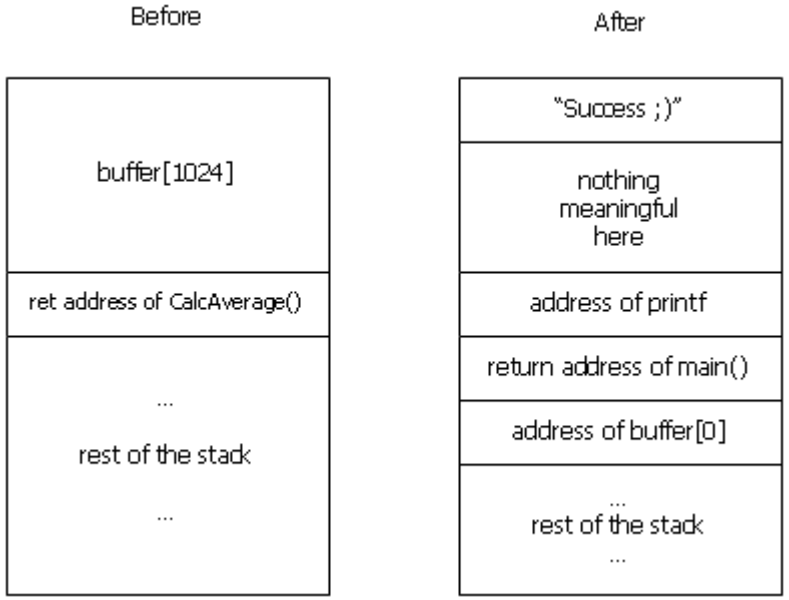
- Smart Du

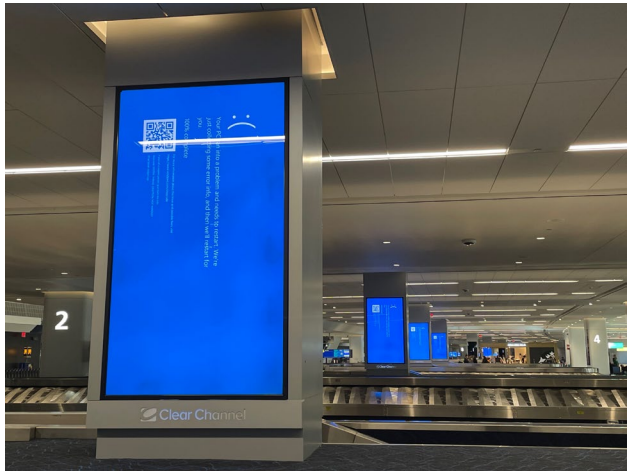


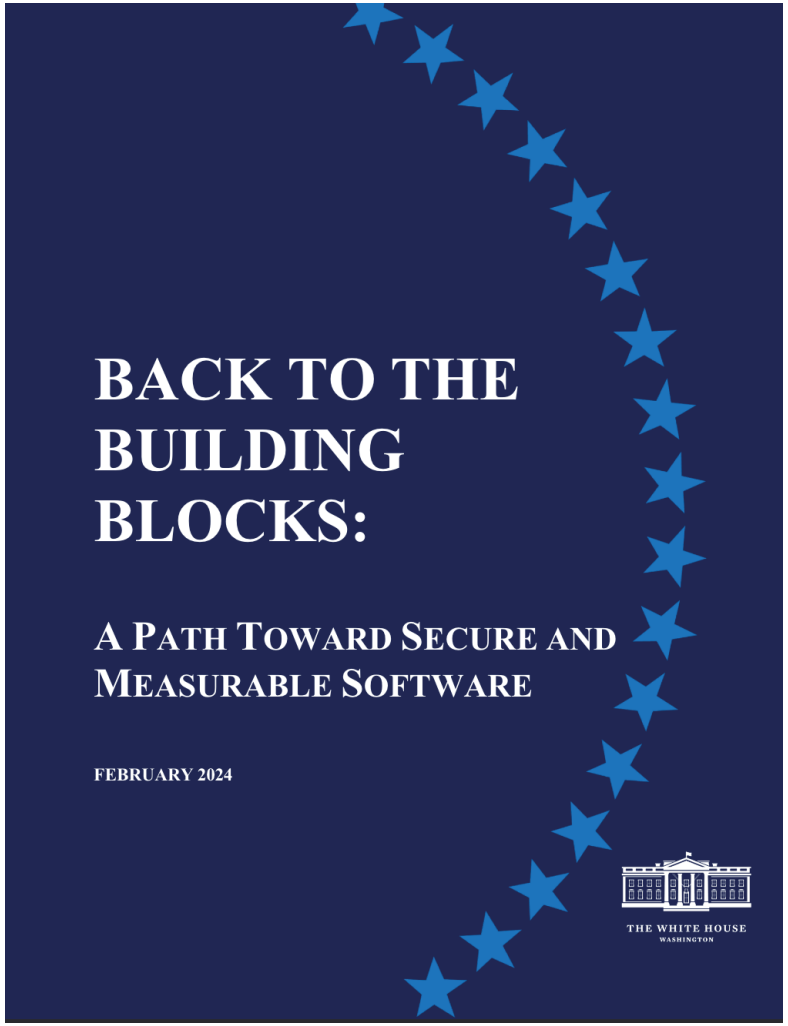


Security


Infineon
SLB 9635 TT1.2
TPM


Before
buffer[1024]
ret address of CalcAverage()
...
rest of the stack
...
After
"Success ;)"
nothing meaningful here
address of printf
return address of main()
address of buffer[0]
...
rest of the stack
...


2
Clear Channel

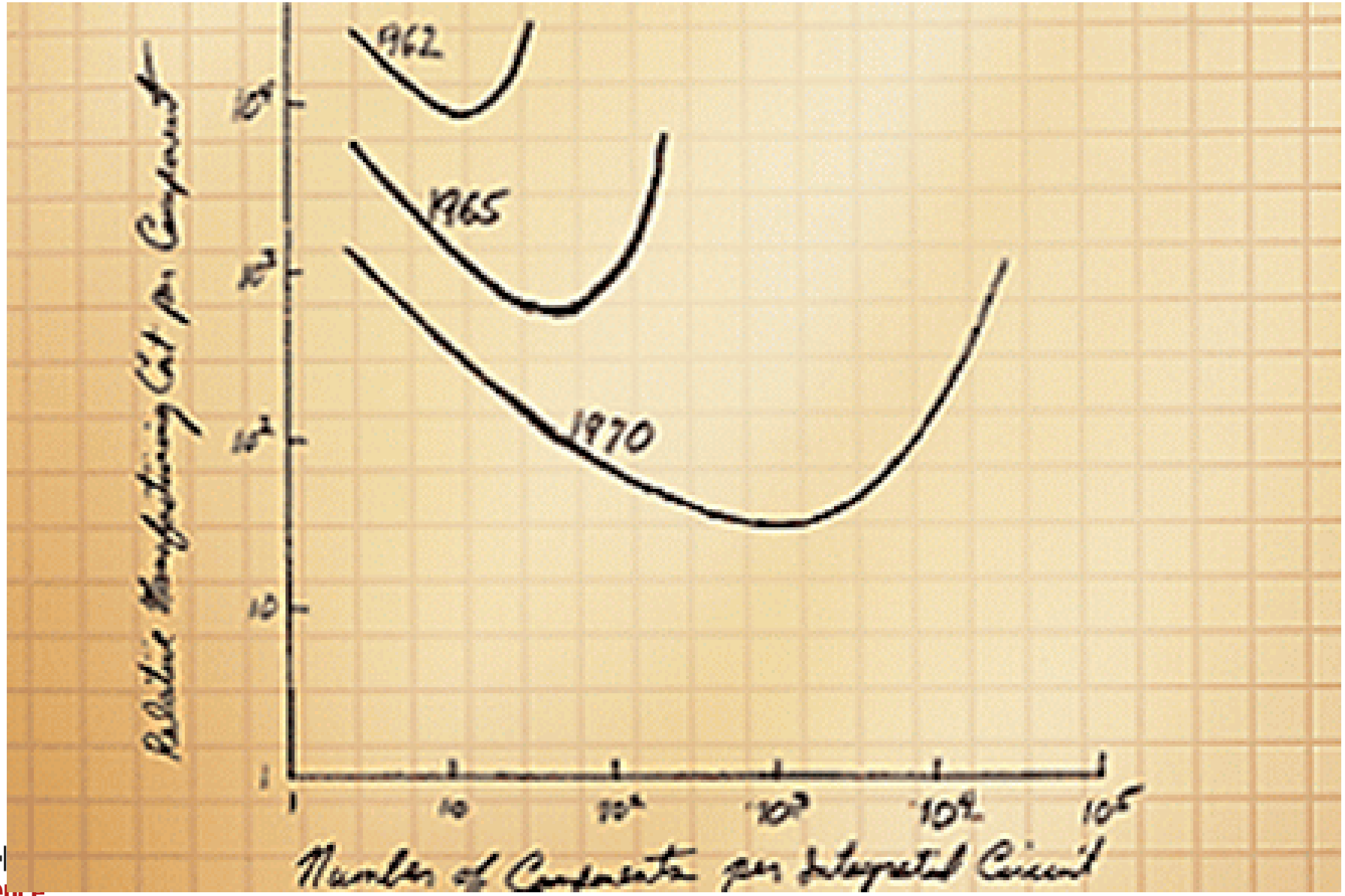

BACK TO THE BUILDING BLOCKS:
A PATH TOWARD SECURE AND MEASURABLE SOFTWARE
FEBRUARY 2024
THE WHITE HOUSE
WASHINGTON

Moore's Law

- Moore's Law introduced in 1965
  - Number of transistors that can be integrated on a single die would double every 18 to 24 months (i.e., grow exponentially with time)

- Amazingly visionary
  - 2300 transistors, 1 MHz clock (Intel 4004) - 1971
  - 16 Million transistors (Ultra Sparc III)
  - 42 Million transistors, 2 GHz clock (Intel Xeon) – 2001
  - 55 Million transistors, 3 GHz, 130nm technology, 250mm2 die (Intel Pentium 4) – 2004
  - 290+ Million transistors, 3 GHz (Intel Core 2 Duo) – 2007
  - 731 Million transistors, 2-3Ghz (Intel Nehalem) – 2009
  - 1.4 Billion transistors, 2-3Ghz (Intel Ivy Bridge) – 2012
  - 7.2 Billion transistors, 3-3.9 GHz Intel Broadwell (22-core) – 2016

Moore's Law



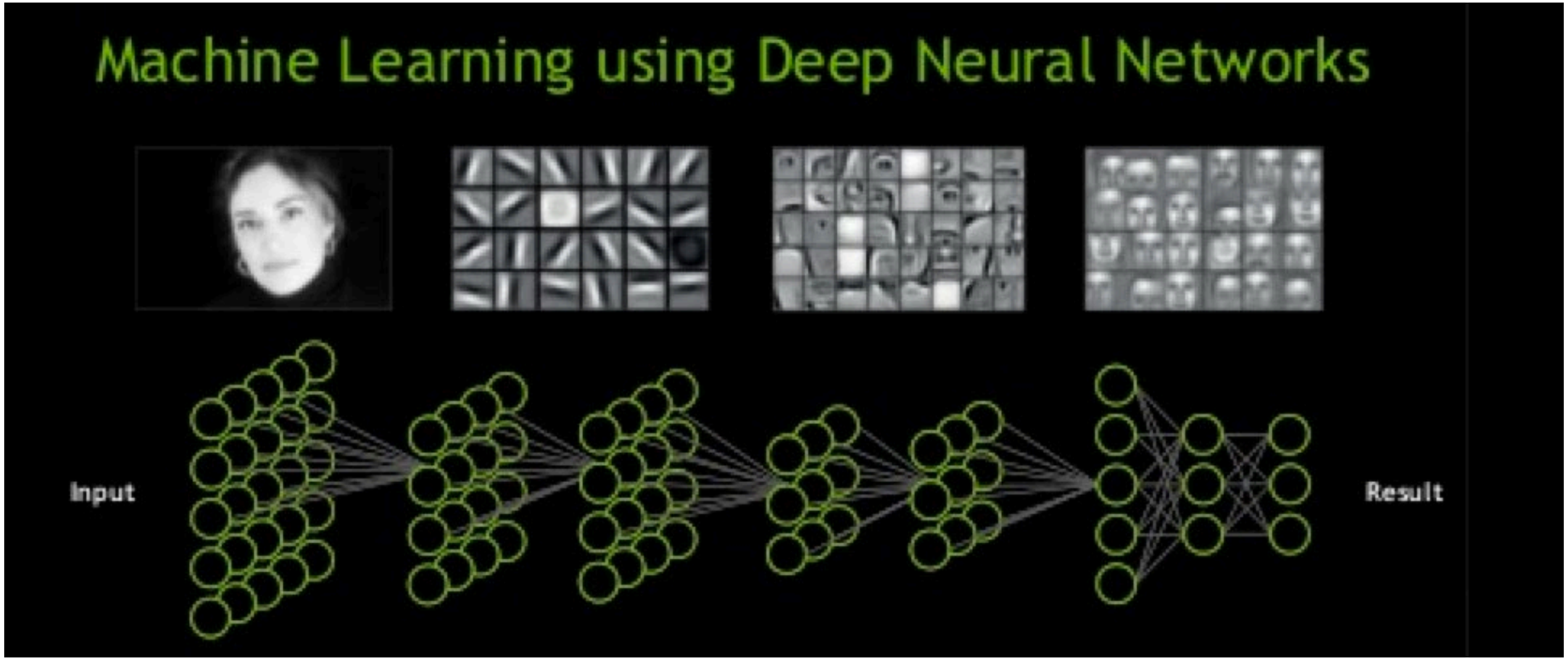
Parallelism

- Dennard scaling: power

- Must exploit parallelism for performance

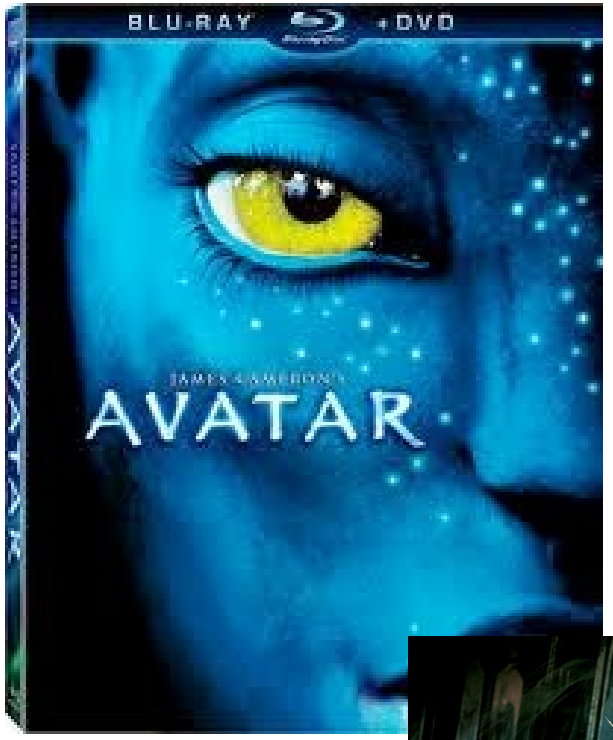
- AI
- AI
- AI!

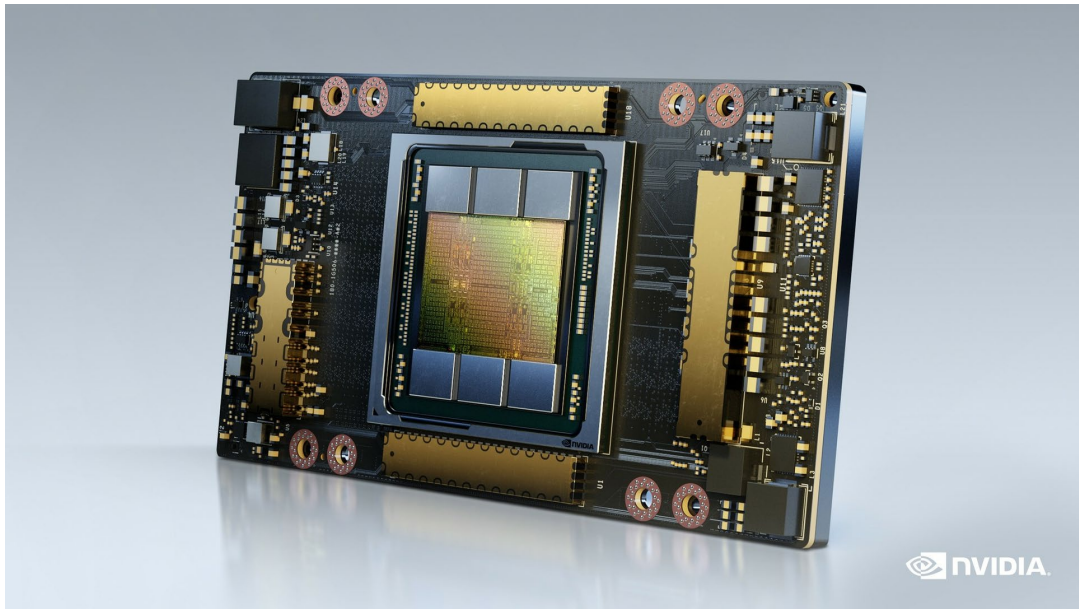
Systems for AI and Deep Learning




Always has been
Wait, AI is just Systems?
imgflip.com

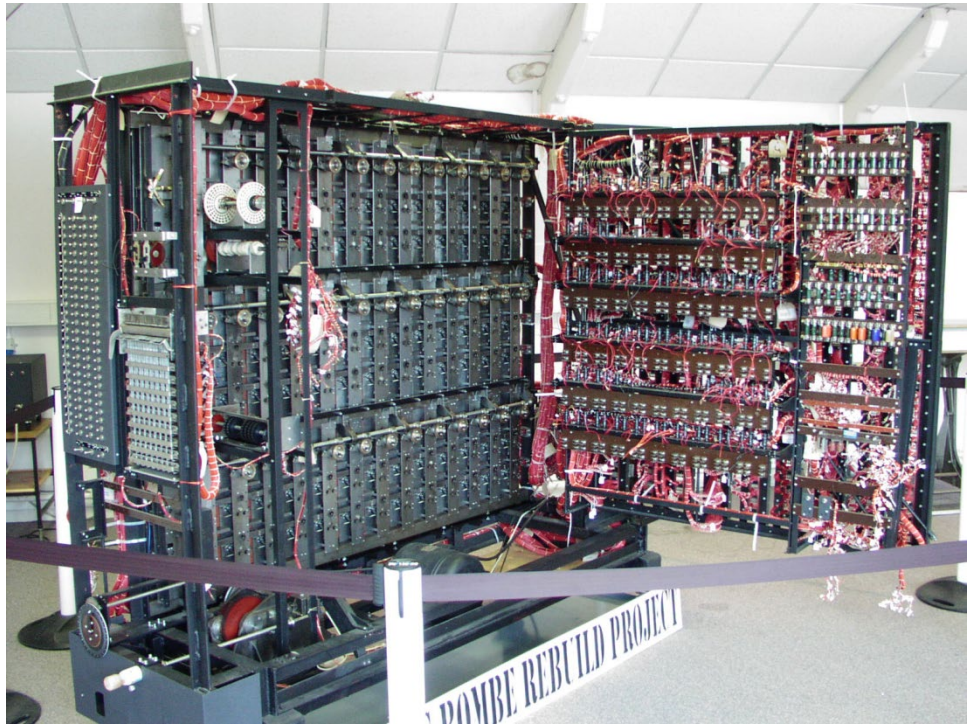
GPUs for Graphics, of course





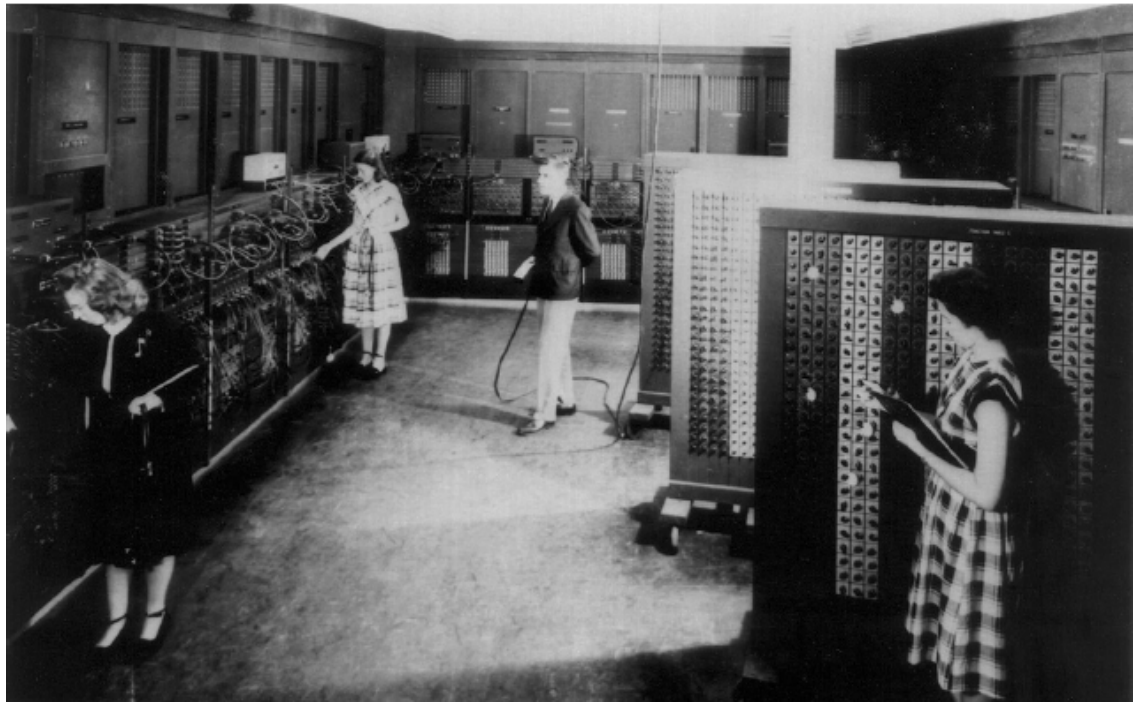
NVIDIA A100

You Could Save The World One Day!



ENIAC - 1946

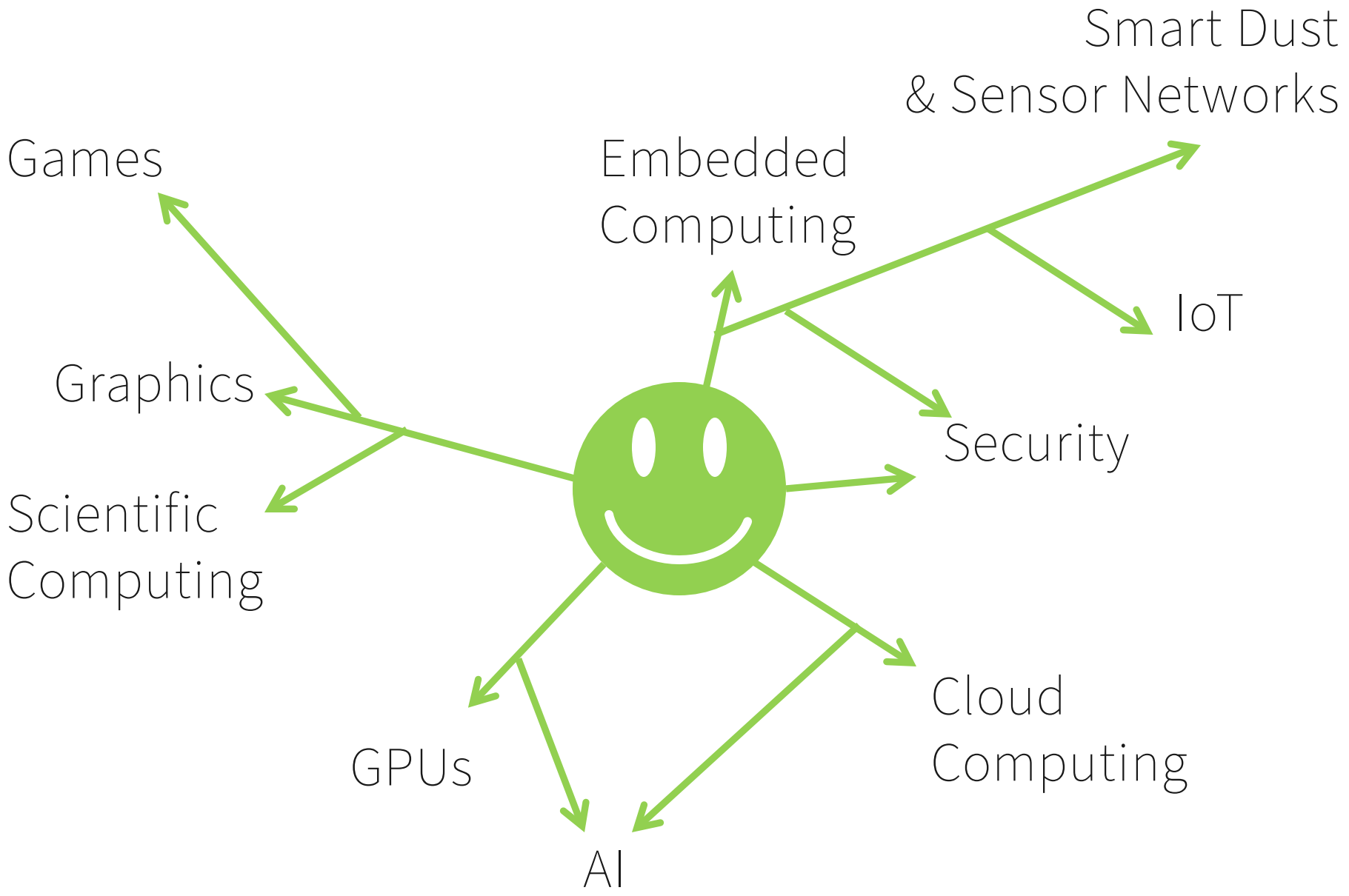
First general purpose electronic computer. Designed to calculate ballistic trajectories



Human Computers programming the IBM 7090

Mary Jackson

Hidden Figures

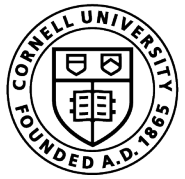

Smart Dust
& Sensor Networks
Games
Embedded
Computing
IoT
Graphics
Security
Scientific
Computing
Cloud
Computing
GPUs
AI

Courses Offered in Fall 2025

- CS 4410: Operating Systems (not offered in the Fall 2025, but offered Spring 2026)
- CS 4411: Practicum in Operating Systems
- CS 4414: Systems Programming
- **CS 4420 (ECE 4750): Computer Architecture
- CS 5414 - Distributed Computing Principles
- CS 5416 - Cloud Computing and ML Hosting

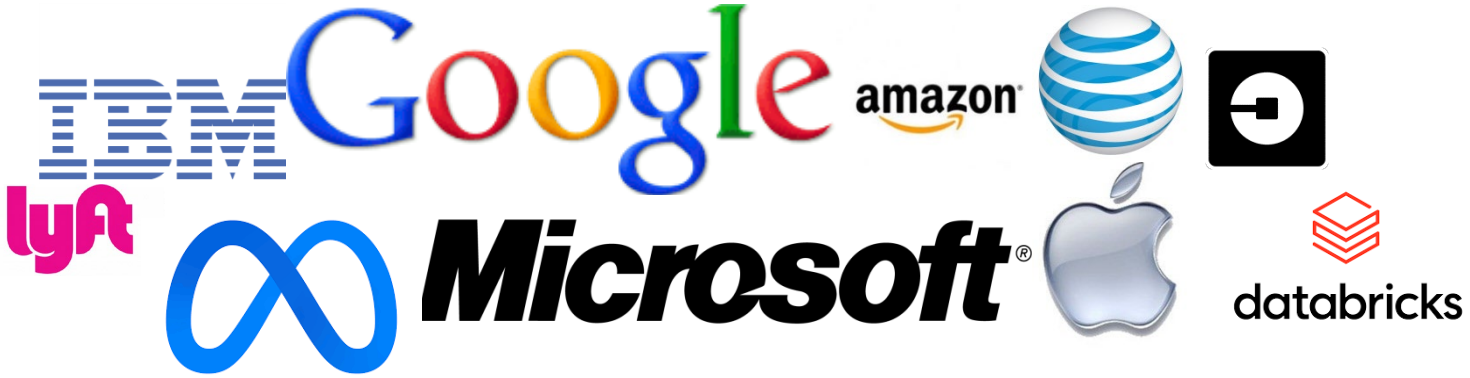
Where to?

- Your job as a computer scientist will require knowledge the computer
- Research/University



Cornell Bowers C·IS
Computer Science

- Industry



- Government

"If you want to make an apple pie from scratch, you must first create the universe."

– Carl Sagan