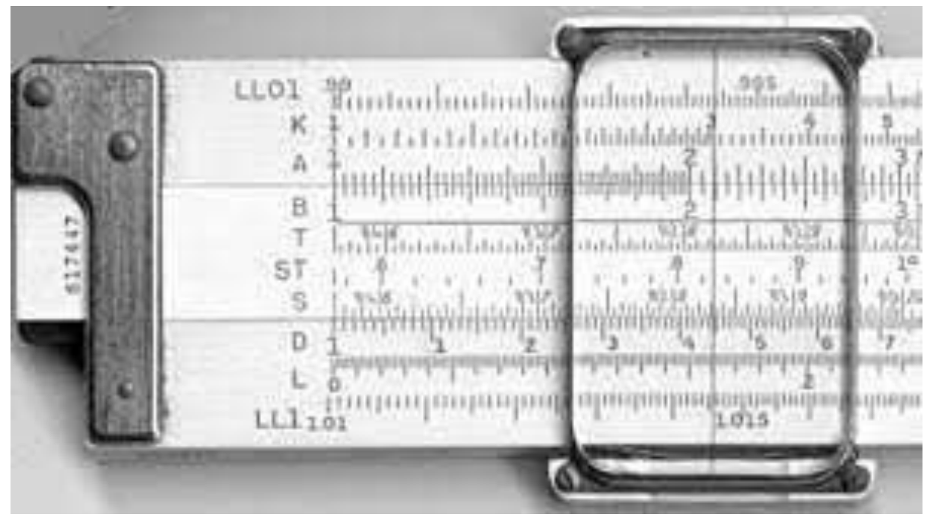
CSCI 381/780: Parallel and Distributed Computing
Why Parallel (& Distributed) Computing?

Jun Li
Queens College & Graduate Center
jun.li@qc.cuny.edu

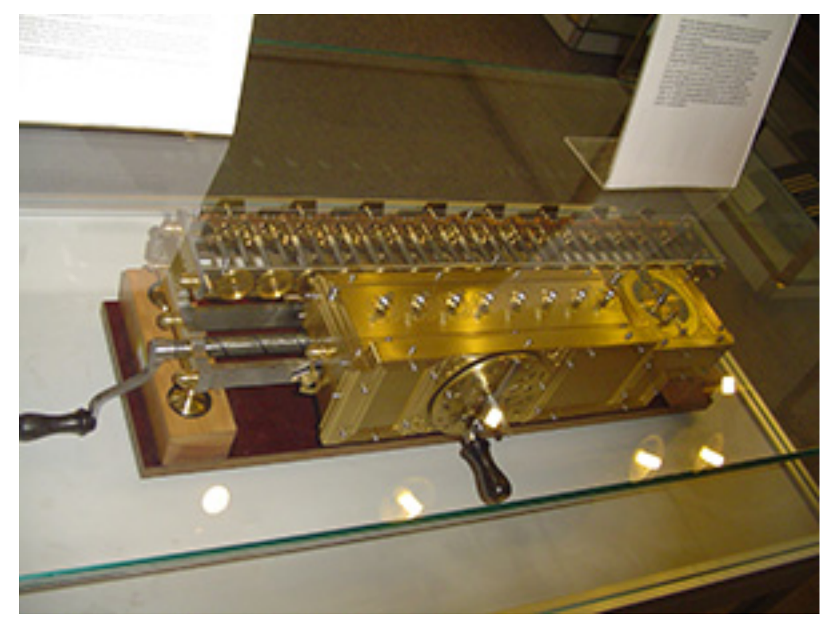
Mechanical "Computer"

abacus



slide rule

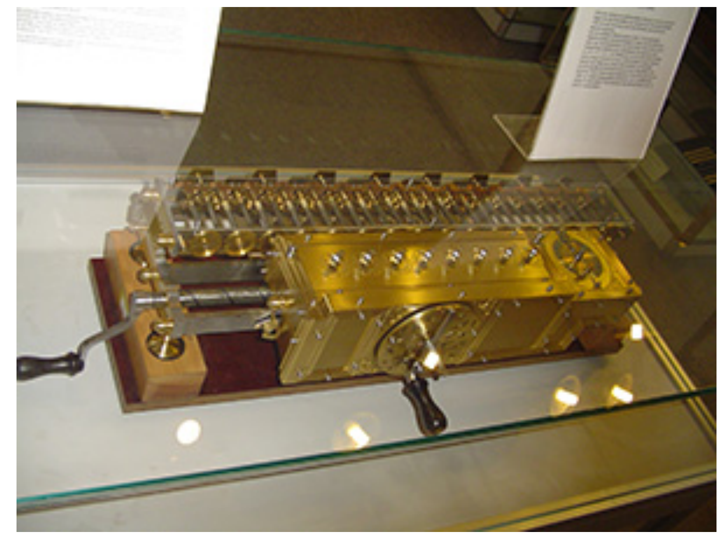
Pascal calculator



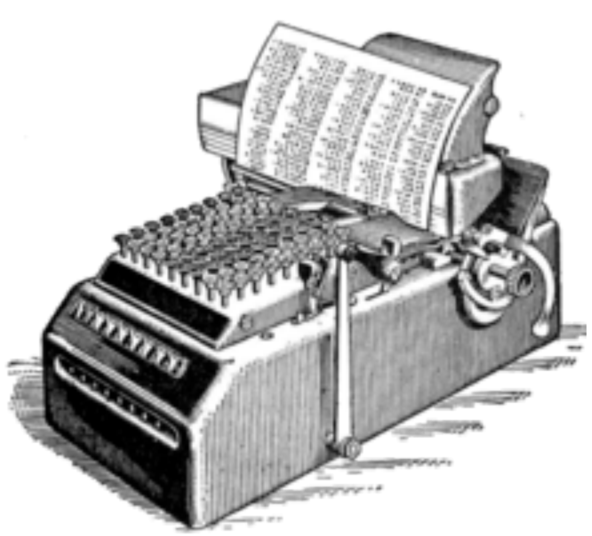
Leibniz's Calculating Machine

Wheel/Key-based Calculator

Pascal calculator (1652)



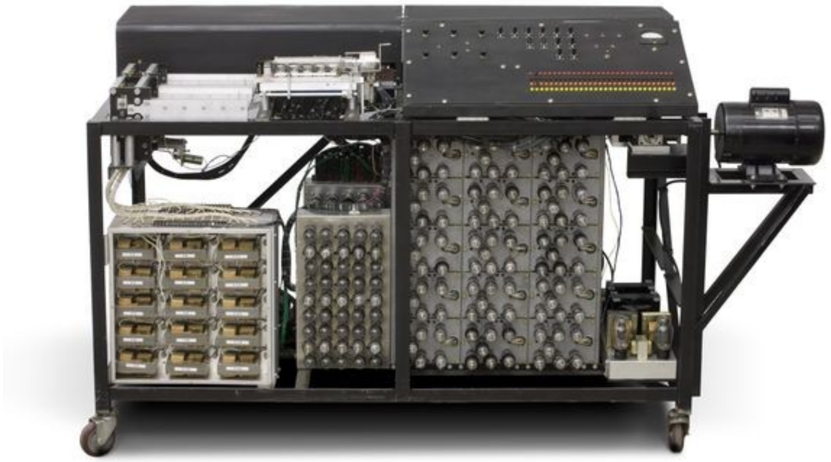
Leibniz's Calculating Machine (1694)



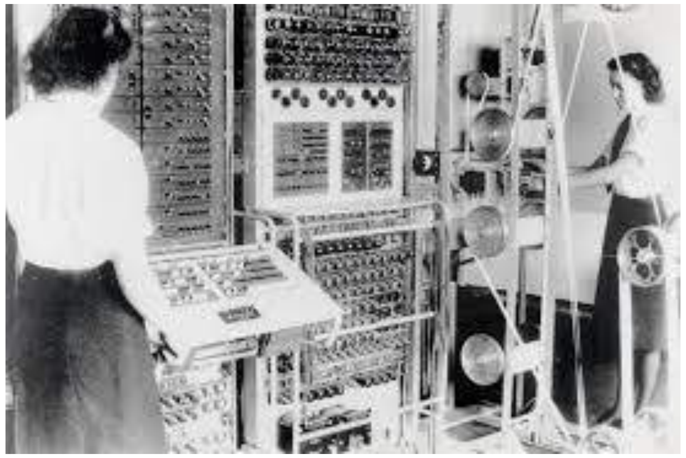
Comptormeter (1887)

Dalton’s Adding Machine (1903)

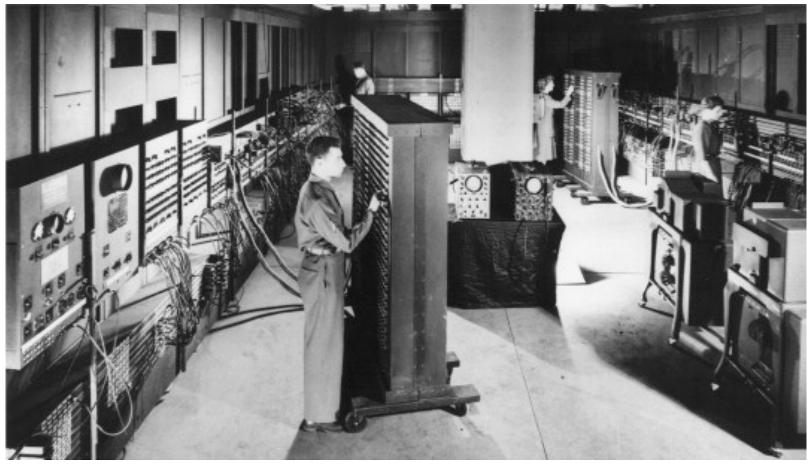
Electronic Computer



Atanasoff-Berry Computer (1942)



Colossus (1943)



ENIAC (1946)

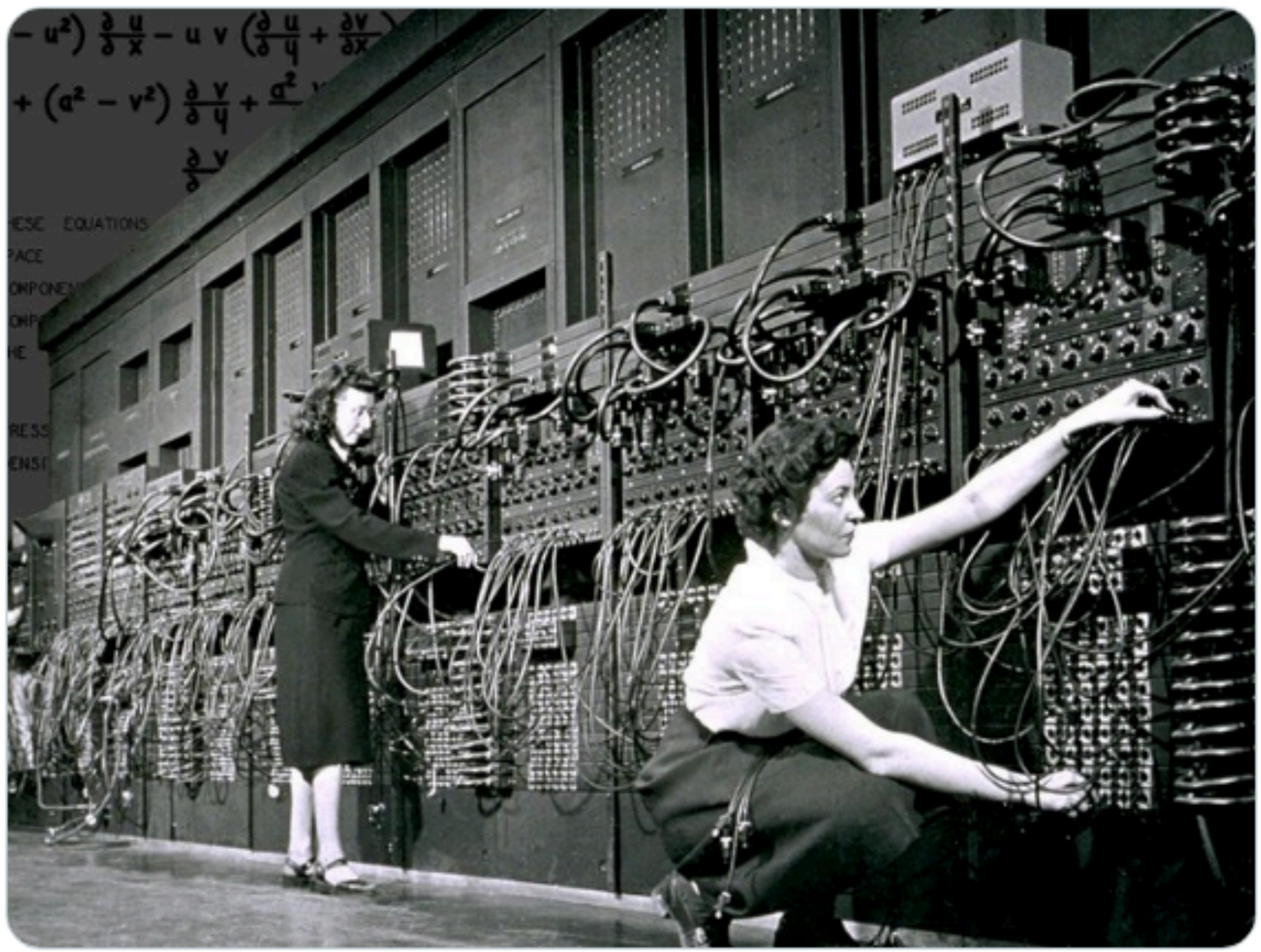
ENIAC can perform 1900 additions per second, but ENIAC calculated a trajectory in 30 seconds that took a human 20 hours (allowing one ENIAC hour to displace 2,400 human hours)

MIT CSAIL
@MIT_CSAIL

Today's the 75th anniversary of ENIAC, widely considered to be the first digital computer, coded by six female programmers. bit.ly/1i95x4o
@DigitalTrends

#EniacDay

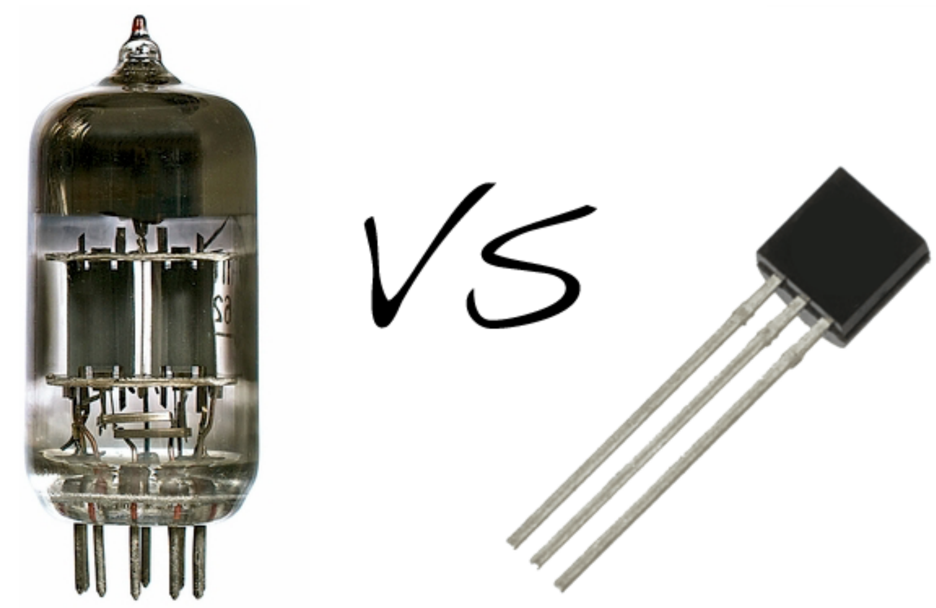
(photo: Corbis/Getty Images)



11:53 AM · Feb 15, 2021 · TweetDeck

From Cubes to Transistors

- **The first-generation computers are made with vacuum tubes.**



vacuum computer

Property of Museum of History & Industry, Seattle

transistor computer

Commercialization

IBM System/360 (1964)

Mainframe

DEC PDP-8 (1965)

Mini-computer

Cray-1 (1976)

Supercomputer

Supercomputers

Rank	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
1	**Supercomputer Fugaku** - Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D, Fujitsu RIKEN Center for Computational Science Japan	7,630,848	442,010.0	537,212.0	29,899
2	**Summit** - IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband, IBM DOE/SC/Oak Ridge National Laboratory United States	2,414,592	148,600.0	200,794.9	10,096
3	**Sierra** - IBM Power System AC922, IBM POWER9 22C 3.1GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband, IBM / NVIDIA / Mellanox DOE/NNSA/LLNL United States	1,572,480	94,640.0	125,712.0	7,438

LINPACK benchmark :

Rmax = maximal performance obtained

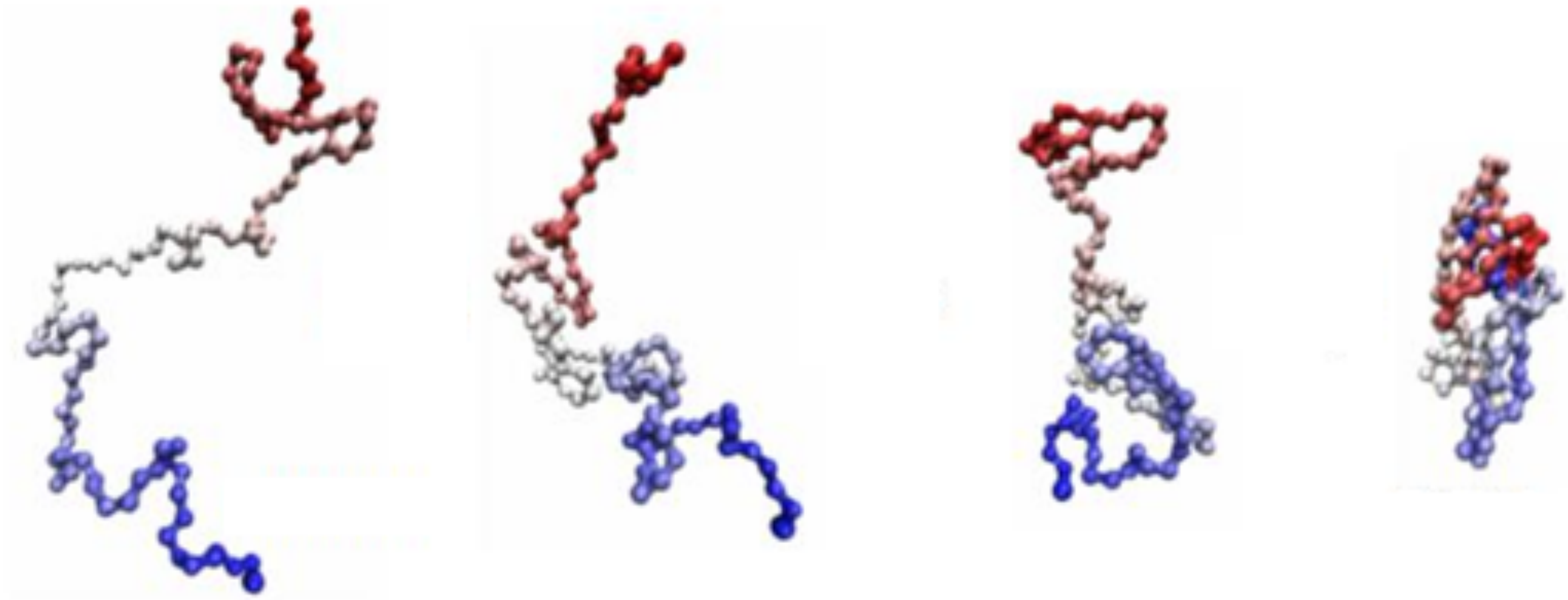
Rpeak = theoretical maximal performance

A 1 teraFLOPS (TFLOPS) computer system is capable of performing one trillion (10^{12}) floating-point operations per second

Why we need ever-increasing performance

- Computational power is increasing, but so are our computation problems and needs.
- Problems we never dreamed of have been solved because of past increases, such as decoding the human genome.
- More complex problems are still waiting to be solved.

Protein folding



Highly accurate protein structure prediction with AlphaFold

John Jumper, Richard Evans, ... Demis Hassabis + Show authors

Nature 596, 583–589 (2021) | Cite this article

Abstract

Proteins are essential to life, and understanding their structure can facilitate a mechanistic understanding of their function. Through an enormous experimental effort[1,2,3,4], the structures of around 100,000 unique proteins have been determined[5], but this represents a small fraction of the billions of known protein sequences[6,7]. Structural coverage is bottlenecked by the months to years of painstaking effort required to determine a single protein structure. Accurate computational approaches are needed to address this gap and to enable large-scale structural bioinformatics. Predicting the three-dimensional structure that a protein will adopt based solely on its amino acid sequence—the structure prediction component of the 'protein folding problem'[8]—has been an important open research problem for more than 50 years[9]. Despite recent progress[10,11,12,13,14], existing methods fall far short of atomic accuracy, especially when no homologous structure is available. Here we provide the first computational method that can regularly predict protein structures with atomic accuracy even in cases in which no similar structure is known. We validated an entirely redesigned version of our neural network-based model, AlphaFold, in the challenging 14th Critical Assessment of protein Structure Prediction (CASP14)[15], demonstrating accuracy competitive with experimental structures in a majority of cases and greatly outperforming other methods. Underpinning the latest version of AlphaFold is a novel machine learning approach that incorporates physical and biological knowledge about protein structure, leveraging multi-sequence alignments, into the design of the deep learning algorithm.

Drug discovery

Russell Kaplan
@russelljkaplan

1/ A friend runs a biotech startup designing drugs to
fight cancer.
make a protei
Binding to jus
growth, but b
faster.

5:02 AM · Jan 23, 20

Russell Kaplan
@russelljkaplan

2/ If they
one recep
might be a

5:02 AM · Jan 2

Russell Kaplan
@russelljkaplan

3/ Before AlphaFold, finding such a protein would take
~1 month:
them into
receptor b
DNA.

5:02 AM · Jan 2

Russell Kaplan
@russelljkaplan

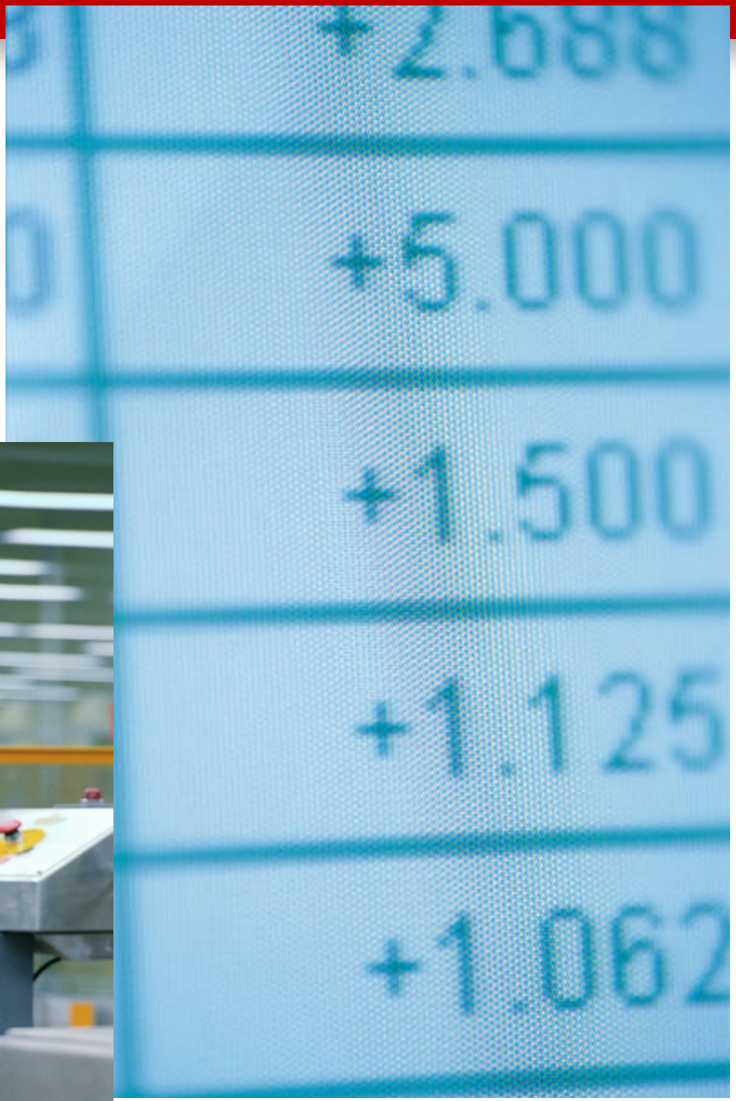
4/ AlphaFold unlocked a different approach: it found the 3D structure of the existing protein & receptors, which was unknown. With the structure + another ML model, they saw how binding affinities would change with different mutations. This led to an ideal candidate in 8 hours.

5:02 AM · Jan 23, 2022 · Twitter Web App

Energy research

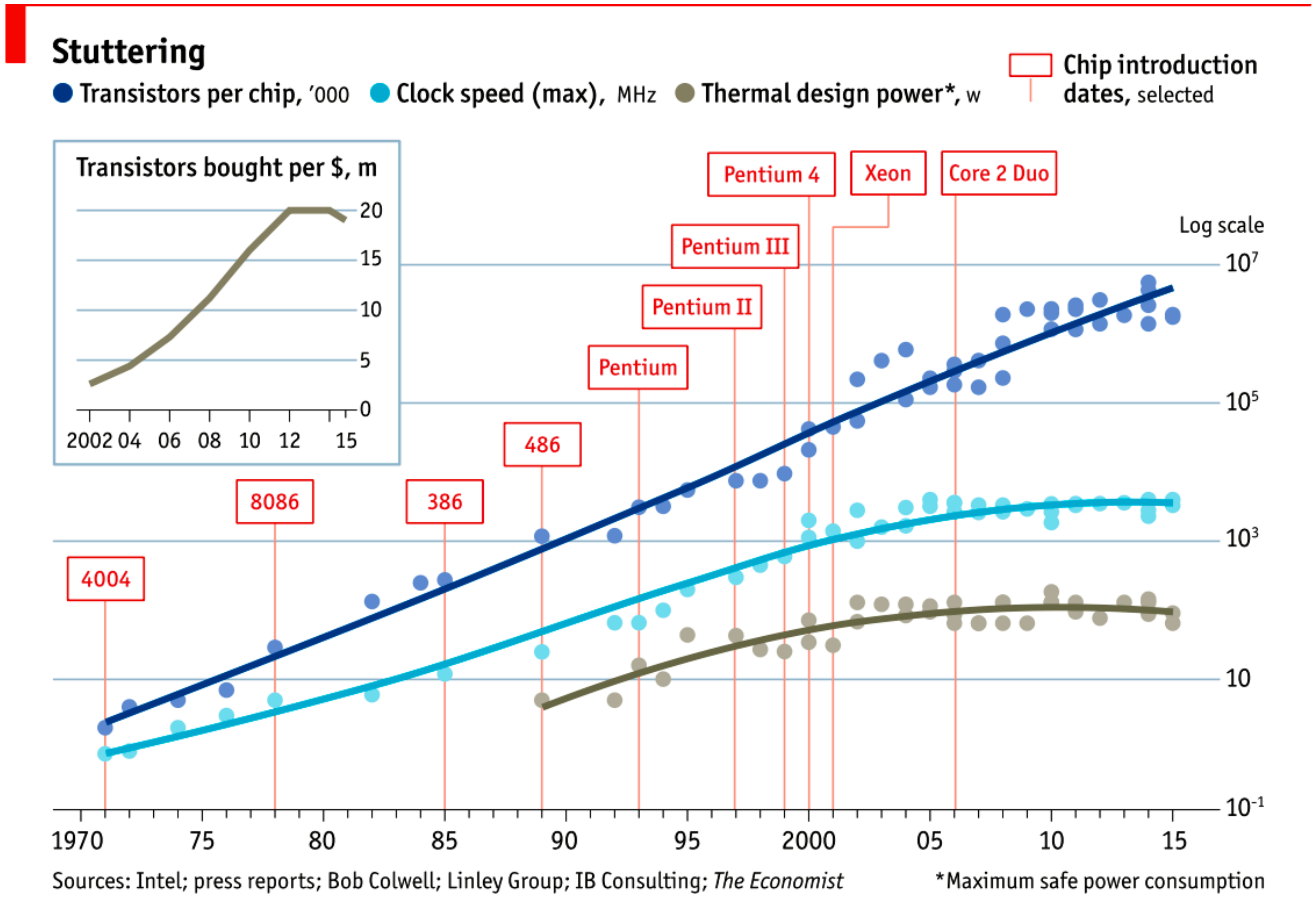
Climate modeling

Data analysis



Moore's Law

- From 1986 to 2002, microprocessors were speeding like a rocket, increasing in performance an average of 50% per year.

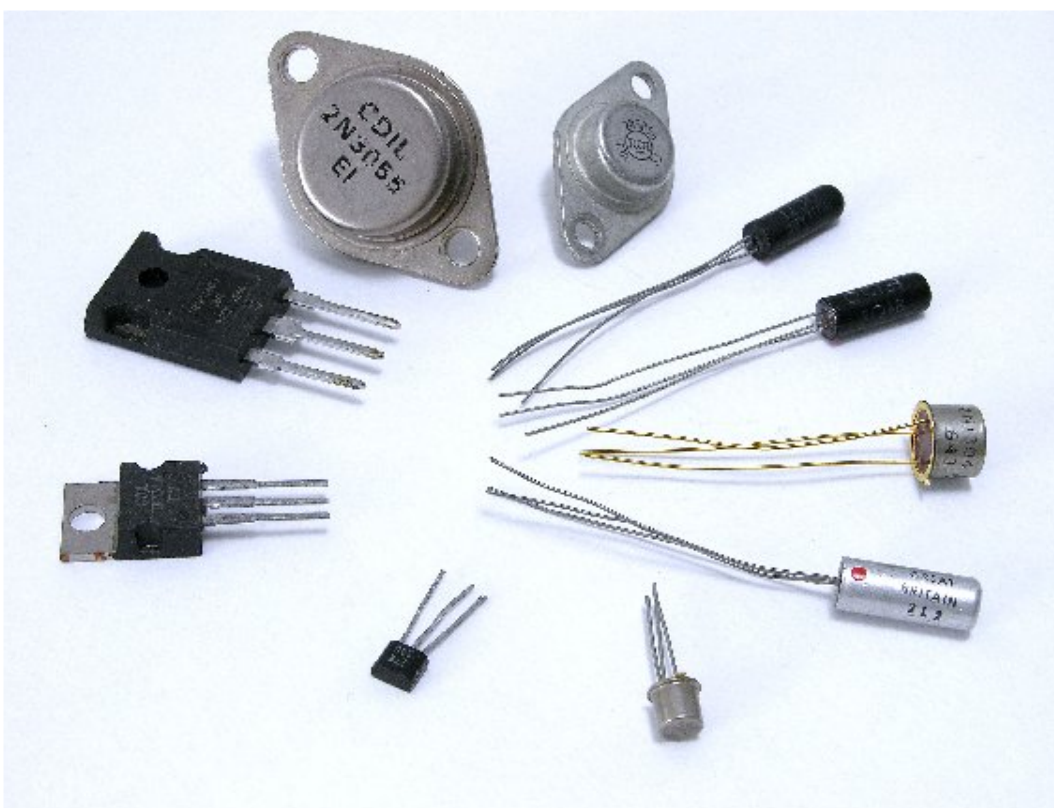


- Since then, it's dropped to about a 20% increase per year.

Why we're building parallel systems

- **Up to now, performance increases have been attributable to the increasing density of transistors.**

- **But there are inherent problems.**



discrete transistors

integrated circuits

A little physics lesson

- Smaller transistors = faster processors.
- Faster processors = increased power consumption.
- Increased power consumption = increased heat.
- Increased heat = unreliable processors.

Power consumption

- **The dominant technology for integrated circuits is called CMOS (complementary metal oxide semiconductor). For CMOS, the primary source of energy consumption is so-called dynamic energy:**
  - Energy is consumed when transistors switch states from 0 to 1 and vice versa.

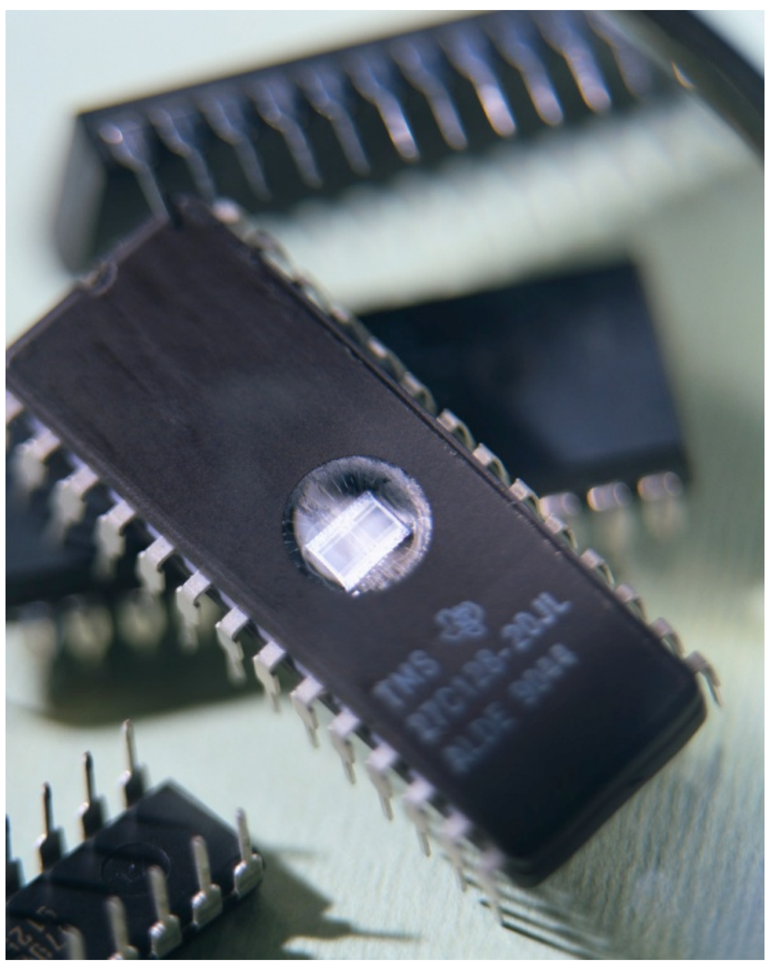
$$Energy \propto 1/2 \times Capacitive\ load \times Voltage^2$$

$$Power \propto 1/2 \times Capacitive\ load \times Voltage^2 \times Frequency\ switched$$

- **Further lowering of the voltage appears to make the transistors too leaky.**

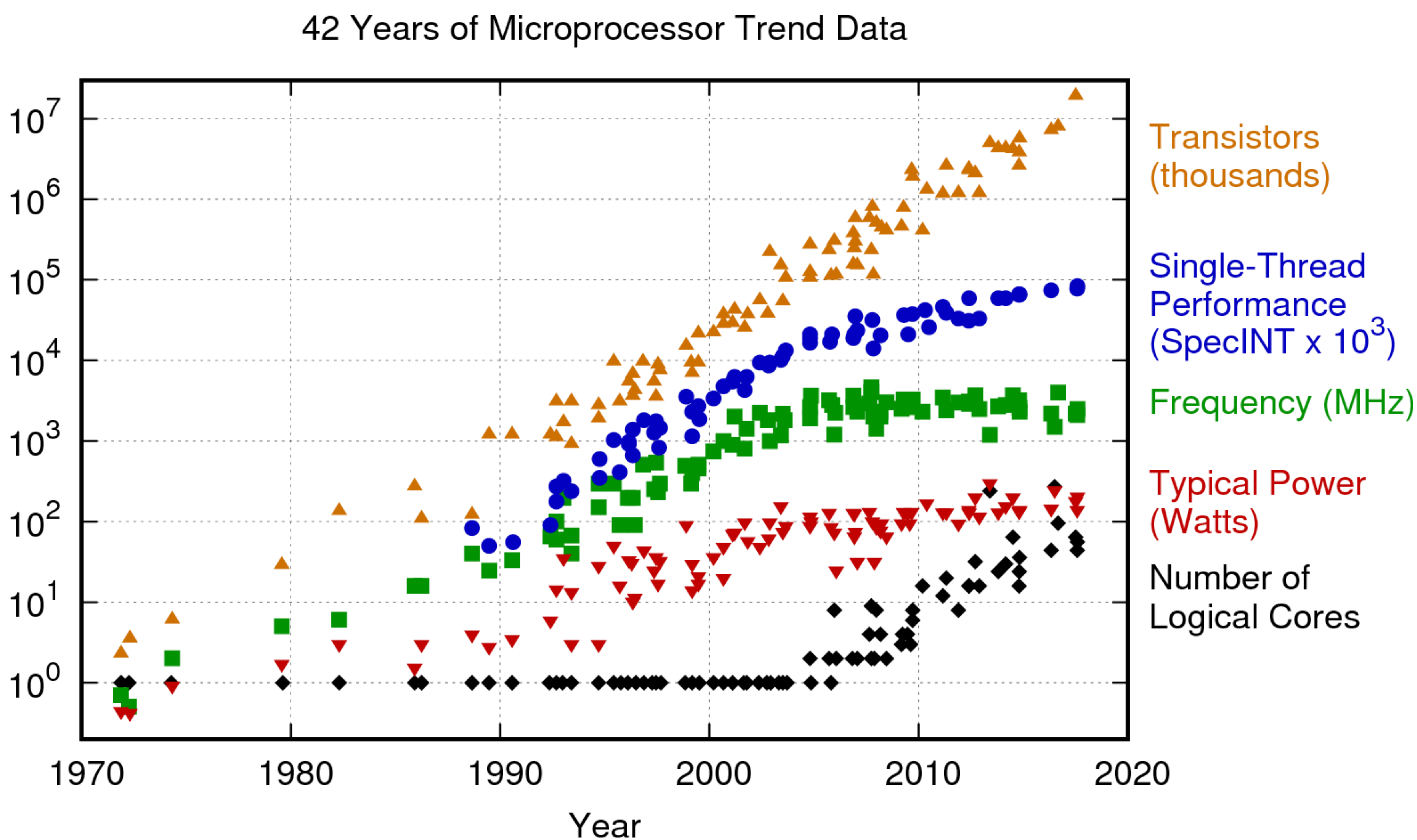
Solution

- Move away from single-core systems to multicore processors.
- “core” = central processing unit (CPU)



Introducing parallelism!!!

Multiprocessor



Original data up to the year 2010 collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, and C. Batten
New plot and data collected for 2010-2017 by K. Rupp

Multiprocessor

- **Rather than continuing to decrease the response time of a single program running on the single processor, as of 2006 all desktop and server companies are shipping microprocessors with multiple processors ("cores") per chip, where the benefit is often more on throughput than on response time.**
  - For programmers to get significant improvement in response time, they need to rewrite their programs to take advantage of multiple processors.