EC2.204
Introduction to Processor Architecture (IPA)

Spring 2026

Lecture – 01

- ✓ **Coursework logistics**
- ✓ **Introduction to IPA**

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY

HYDERABAD

05th Jan 2026

About Instructor

Priyesh Shukla

Assistant Professor, IIIT Hyderabad (April 2025 – current)
Head of Sustainable, Advanced and Robust Computing Systems (SARCS) Research Group
Center for VLSI and Embedded Systems Technology (CVEST), and Computer Systems Group (CSG)
priyesh.shukla@iiit.ac.in

Past experience

Chief Engineer (Advanced Research Group), Samsung Research – Bangalore (2022 – 2025)
Research Co-op at Bose Corporation - Massachusetts (2022) & Qualcomm - San Diego (2020)
Engineer, Qualcomm – Bangalore (2017-18)

Education

Doctor of Philosophy in Electrical and Computer Engineering from **University of Illinois, Chicago, USA**

Master of Engineering in Microelectronics from **BITS-Pilani**

Bachelor of Engineering (Hons.) in Electrical and Electronics from **BITS-Pilani**

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
HYDERABAD

Coursework Logistics

Course Pre-requisites

Digital Design (Digital Systems and Microcontrollers)

Hardware Description Language (HDL) – Verilog/VHDL

C/Cpp programming – preferred

Lectures and Tutorials

- **Lectures: 2pm to 3:25pm (Every Monday and Thursday)**

- **Tutorials: Weekly (TBA)**
  - ✓ **For project discussions**
  - ✓ **Clarification of doubts from lectures (if any)**

- **All communications will be through Moodle**

Course material

Reference book

Computer Architecture: A Quantitative Approach by *Patterson, Hennessy and Kozyrakis* 7th edition (6th ed. works too)

Links

https://riscv.org/

https://github.com/riscv/learn

https://riscv.org/community/training/

Grading plan

Type of Evaluation	Weightage (in %)
Quiz-1	10
End Sem Exam	30
Project	60

A few Assignments may be floated - adjusted within project component

Project

- **RISC-V Processor Design**
- **3 students per group**
- More details to be shared soon…

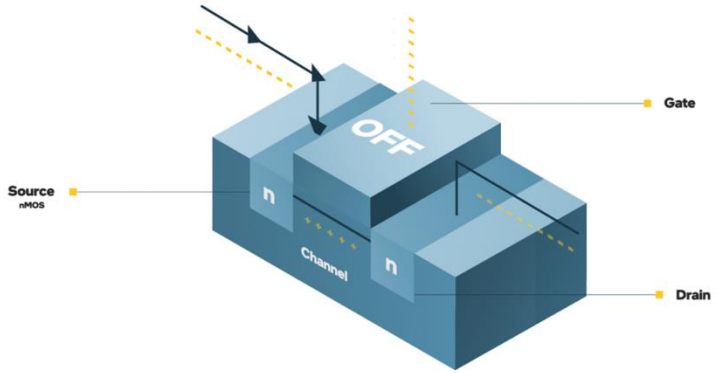
Course introduction and overview

Fundamental building block of modern computers

https://www.youtube.com/watch?v=Q5paWn7bFg4

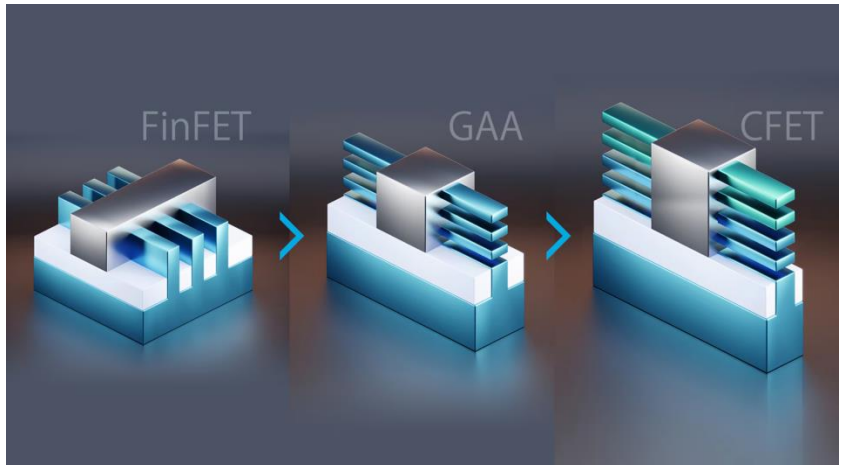
Julius Edgar Lilienfeld proposed the concept of a field-effect transistor in 1925.

John Bardeen, William Shockley, and Walter Brattain at Bell Labs in 1948; Bardeen and Brattain invented the point-contact transistor in 1947 and Shockley invented the bipolar junction transistor in 1948.



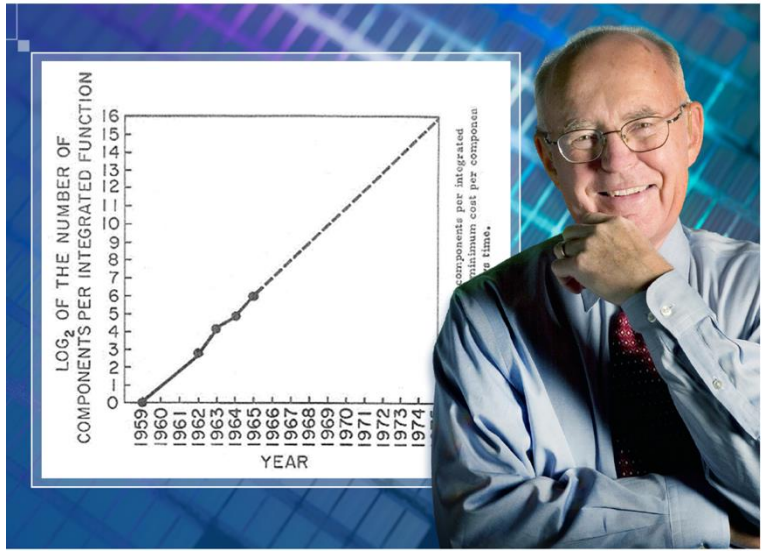
MOSFET

*(introduced in **1959**)*



Source: Wiki, Intel & Telescope Magazine

Moore’s law



Gordon Moore (Intel)

Moore’s Law: The number of transistors on microchips has doubled 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

Transistor count

50,000,000,000
10,000,000,000
5,000,000,000
1,000,000,000
500,000,000
100,000,000
50,000,000
10,000,000
5,000,000
1,000,000
500,000
100,000
50,000
10,000
5,000
1,000

1970 1975 1980 1985 1990 1995 2000 2005 2010 2015 2020

Year in which the microchip was first introduced

Intel 4004
Intel 8008
Intel 8080
Intel 8085
Intel 8086
Intel 8088
Intel 80186
Intel 80286
Intel 80386
Intel 80486
Intel i960
ARM 1
ARM 2
ARM 3
ARM 6
ARM 7TDMI
ARM700
ARM Cortex-A9
R4000
SA-110
AMD K5
AMD K6
AMD K6-III
AMD K7
AMD K8
Pentium
Pentium Pro
Pentium II
Klamath
Pentium II Deschutes
Pentium II Mobile Dixon
Pentium III Coppermine
Pentium III Katmai
Pentium III Tualatin
Pentium 4 Willamette
Pentium 4 Northwood
Barton
Atom
Pentium 4 Prescott
Pentium 4 Cedar Mill
Pentium 4 Prescott-2M
Core 2 Duo Allendale
Pentium D Smithfield
Core 2 Duo Conroe
Cell
Core 2 Duo Wolfdale 3M
Itanium 2 Madison 6M
Core 2 Duo Wolfdale
Itanium 2 with 9 MB cache
AMD K10 quad-core 2M L3
Core i7 (Quad)
Pentium D Presler
POWER6
Itanium 2 McKinley
Dual-core Itanium 2
8-core Xeon Nehalem-EX
Six-core Xeon 7400
12-core POWER8
61-core Xeon Phi
Xbox One main SoC
18-core Xeon Haswell-E5
IBM z13 Storage Controller
72-core Xeon Phi
SPARC M7
GC2 IPU
Centriq 2400
AMD Epyc Rome
AWS Graviton2
32-core AMD Epyc
Apple A12X Bionic
HiSilicon Kirin 990 5G
Apple A13 (iPhone 11 Pro)
AMD Ryzen 7 3700X
HiSilicon Kirin 710
10-core Core i7 Broadwell-E
Qualcomm Snapdragon 835
Dual-core + GPU Iris Core i7 Broadwell-U
Quad-core + GPU GT2 Core i7 Skylake K
Quad-core + GPU Core i7 Haswell
Apple A7 (dual-core ARM64 "mobile SoC")
TMS 1000
Zilog Z80
RCA 1802
Motorola 6809
Motorola 6800
MOS Technology 6502
WDC 65C02
WDC 65C816
Novix NC4016
Motorola 68000
Motorola 68020
TI Explorer's 32-bit Lisp machine chip
DEC WRL MultiTitan

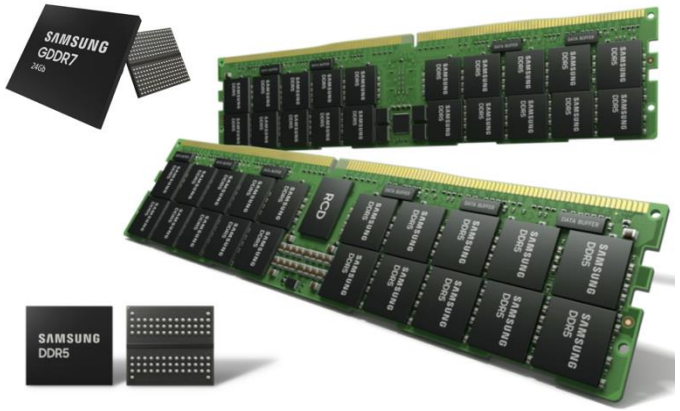
Data source: Wikipedia (wikipedia.org/wiki/Transistor_count)

OurWorldinData.org – Research and data to make progress against the world’s largest problems.

Microprocessor revolution

Era	Key Milestones	Architecture/Tech	Clock Speed	Transistors	Trends
1970s	**Intel 4004 (1971) Intel 8086 (1978)**	4-bit → 16-bit	~0.7–10 MHz	2,300–29,000	Birth of microprocessors
1980s	**Intel 80286, 80386 Motorola 68000**	16-bit → 32-bit	~10–40 MHz	~134K–275K	Desktop revolution
1990s	**Pentium, PowerPC**	Superscalar, MMX	60 MHz → 1 GHz	~1M–42M	Increased parallelism
2000s	**Intel Core, AMD Athlon Multicore chips**	64-bit, Dual-Core	1–3 GHz	~42M–820M	Multicore shift
2010s	**Intel i-series, AMD Ryzen, ARM in mobile**	Up to 64-core (server)	1.5–5 GHz	~1B–10B+	Efficiency, mobile power
2020s	**Apple M1/M2, AMD Ryzen 7000, Intel Core Ultra**	Heterogeneous (big.LITTLE), AI accelerators	2–5+ GHz	10B–100B+	AI, power efficiency, chiplets

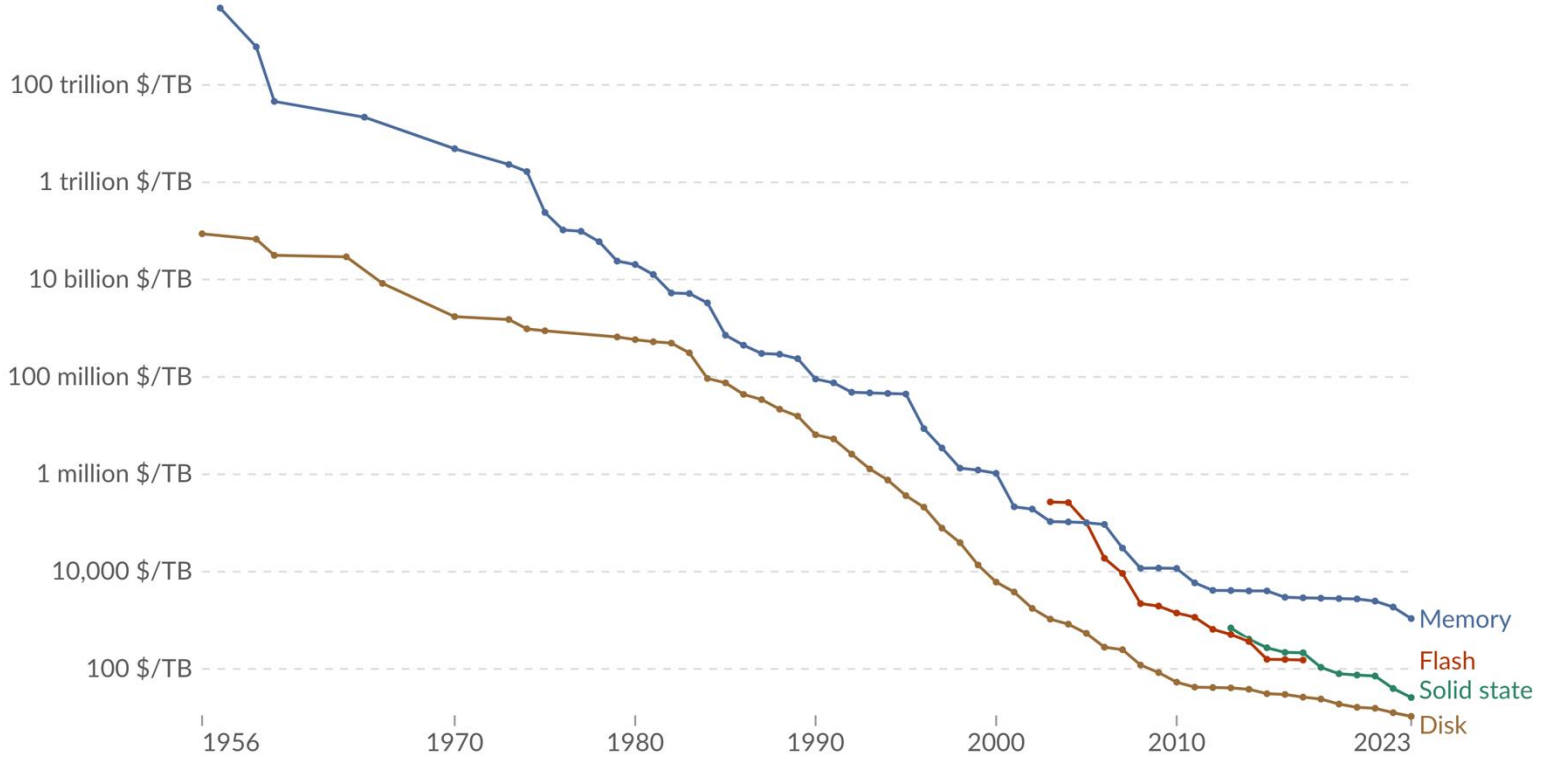
Memory and cost



Historical price of computer memory and storage

Our World in Data

This data is expressed in US dollars per terabyte (TB), adjusted for inflation. "Memory" refers to random access memory (RAM), "disk" to magnetic storage, "flash" to special memory used for rapid data access and rewriting, and "solid state" to solid-state drives (SSDs).

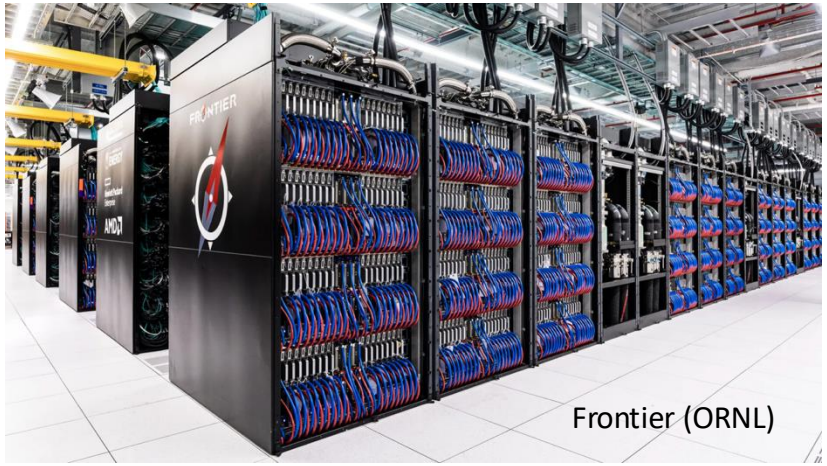


Data source: John C. McCallum (2023); U.S. Bureau of Labor Statistics (2024)

OurWorldinData.org/technological-change | CC BY

Note: For each year, the time series shows the cheapest historical price recorded until that year. This data is expressed in constant 2020 US$.

Computational capacity



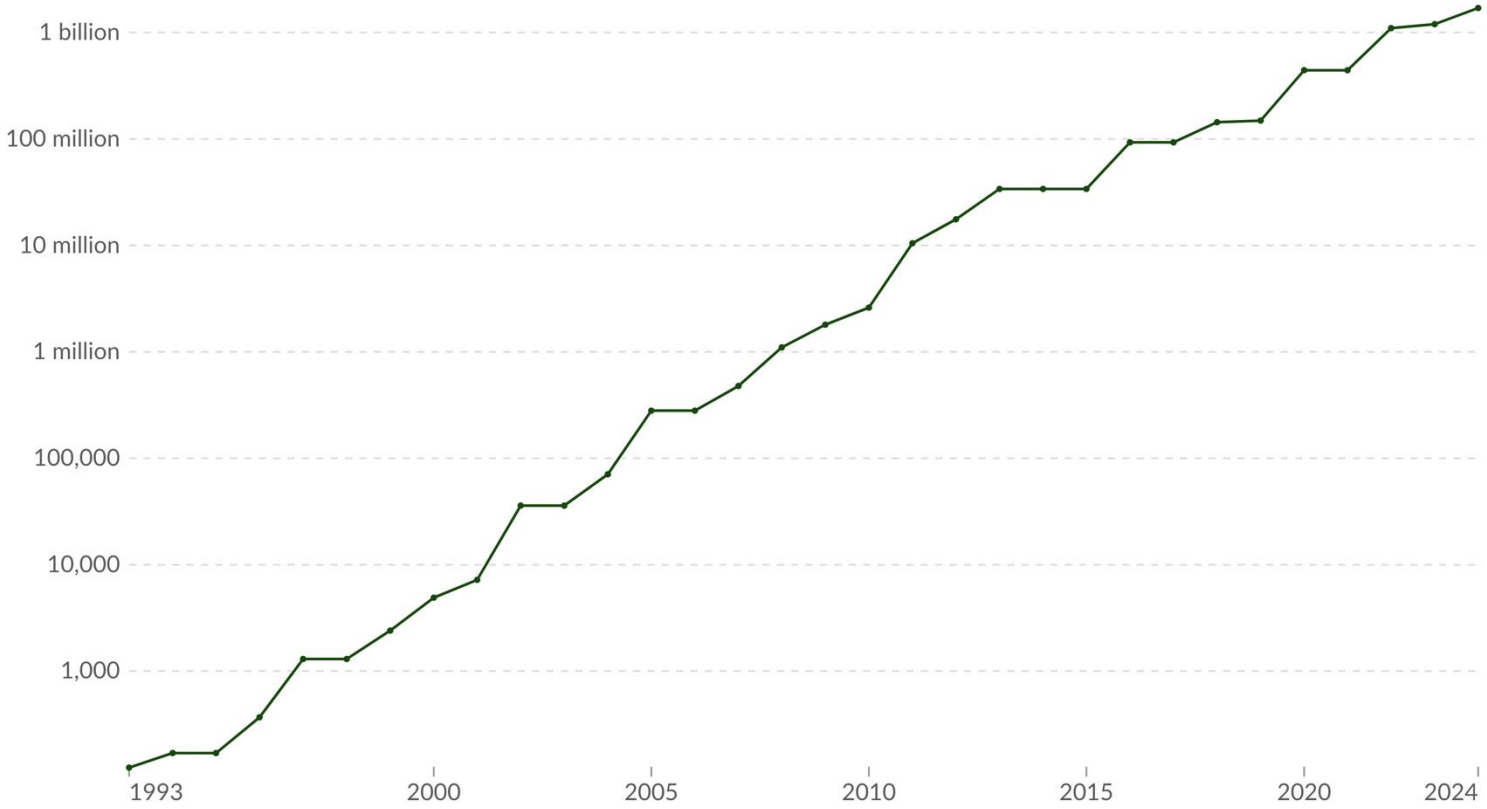
Frontier (ORNL)

El Capitan (LLNL)

Computational capacity of the fastest supercomputers

Our World in Data

The number of floating-point operations[1] carried out per second by the fastest supercomputer in any given year. This is expressed in gigaFLOPS, equivalent to 10^9 floating-point operations per second.

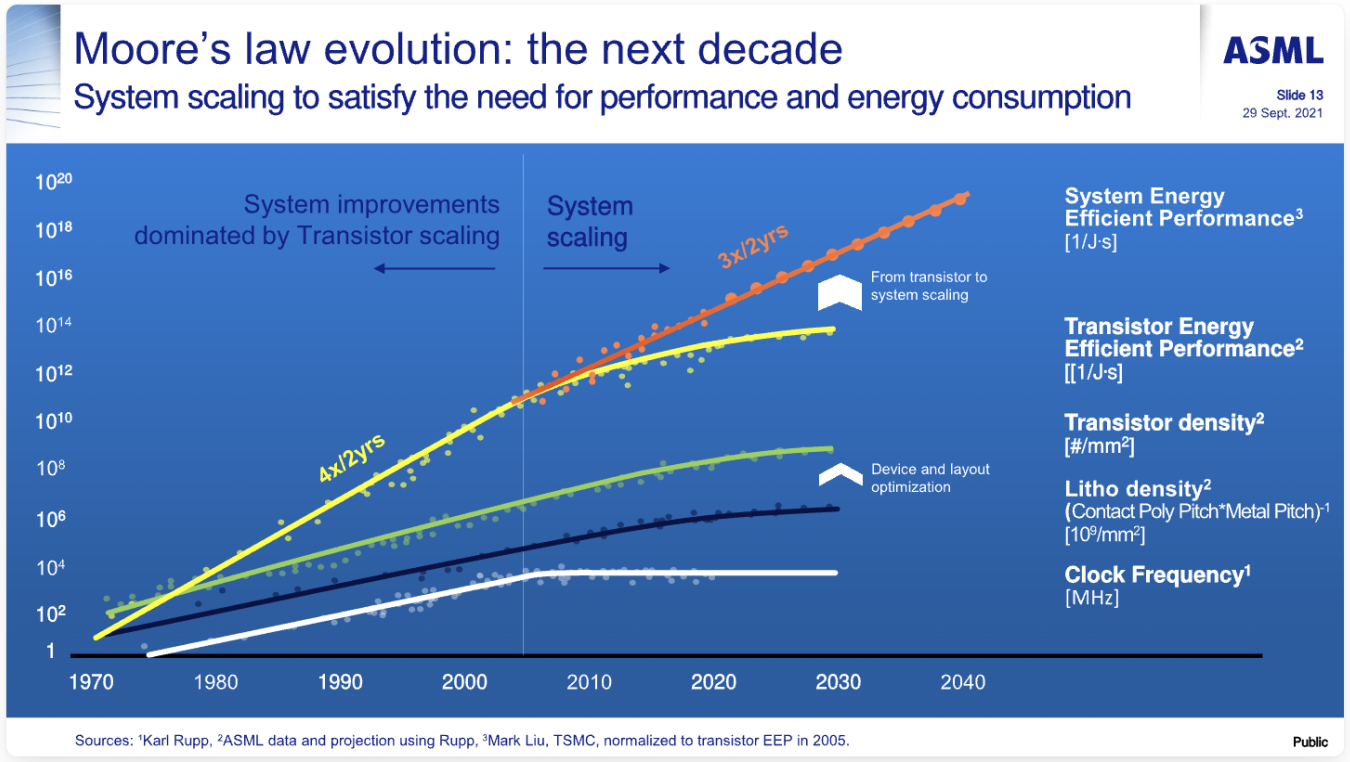


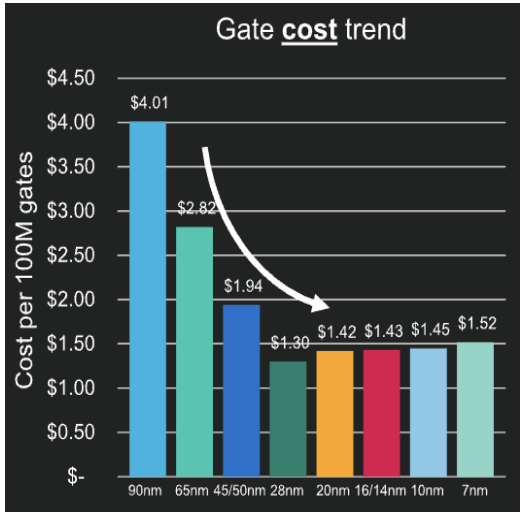
Data source: Dongarra et al. (2024)

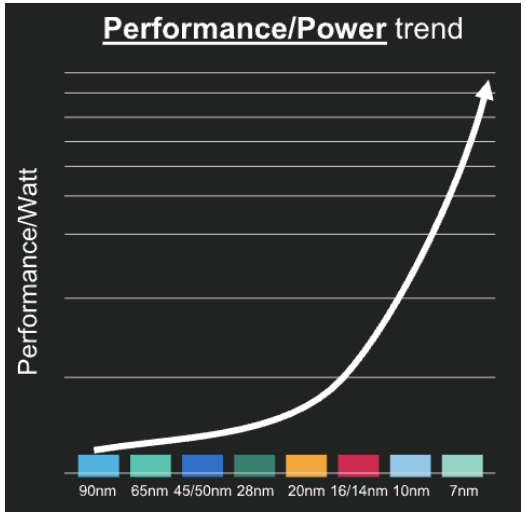
OurWorldinData.org/technological-change | CC BY

1. Floating-point operation A floating-point operation (FLOP) is a type of computer operation. One FLOP represents a single arithmetic operation involving floating-point numbers, such as addition, subtraction, multiplication, or division.

Slowdown of Moore’s law and Price Rise







Moore’s law is slowing down but computational demand continues…

Even if performance is achieved by system scaling, cost increases…

No longer meaningful and attractive for ordinary consumers!

THE FUTURE OF COMPUTING

Computer Architecture in the Post-Moore Era

Evolution, Challenges, and Opportunities in Modern Processor Design

The Foundation: Moore's Law & Dennard Scaling

MOORE'S LAW

Transistor count in integrated circuits doubles approximately every **two years**

DENNARD SCALING

As transistors shrink, voltage scales down, keeping **power density constant**

IMPACT

Together, these principles drove **40 years** of exponential performance growth

Source: Historical semiconductor industry observations

The End of an Era

~2006

> DECADE AGO

Dennard Scaling Ends

Power density no longer stays constant as transistors shrink. Heat dissipation becomes the primary constraint.

Recent

SLOWDOWN

Moore's Law Slowing

Physical limitations and economic factors now constrain further transistor scaling.

Physical Limits

Quantum effects at atomic scales

Economic Factors

Exponentially rising fab costs

Thermal Wall

Heat management challenges

The Solution: Domain-Specific Architectures

PERFORMANCE BENEFIT

3+ Generations

Equivalent to three or more generations of Moore's Law

WHY NOW?

- General-purpose performance gains are slowing
- Custom architectures won't be obsoleted quickly
- Diverse applications need specialized solutions

KEY APPLICATION DOMAINS

Deep Neural Networks and GenAI

High computation, lower precision. Separate architectures for inference vs. training.

Image Processing

High computation, lower precision. Critical for mobile devices where power is limited.

FPGA Accelerators

Reprogrammable for various domains. Ideal for irregular, frequently updated apps.

Beyond Moore's Law: New Technologies

Heterogeneous scaling and emerging technologies provide capabilities beyond traditional transistor scaling

2.5D Stacking

Vertically integrated chip packages with high-bandwidth interconnects between layers

New NV Memories

Emerging nonvolatile memory technologies bridging the gap between DRAM and storage

Optical Interconnects

Light-based communication for high-bandwidth, low-latency data transfer

Key Insight: Fundamental design decisions must be reexamined from first principles to effectively utilize these technologies

Students and practitioners need expertise in both traditional and emerging techniques

Open Source Architecture: RISC-V

OPEN STANDARD ISA

The sixth edition updates all examples to use the RISC-V instruction set architecture

WHY OPEN SOURCE NOW?

With Moore's Law slowing, implementations have **longer lifetimes**, giving open-source architectures more time for optimization.

ADVANTAGES OF OPEN-SOURCE ARCHITECTURES

Extended Lifetime

Slower scaling means implementations remain competitive longer

Continued Optimization

Community-driven refinement improves quality implementations

Educational Value

Hands-on experience with real architecture, no licensing barriers

Industry Adoption

Growing ecosystem of tools, compilers, and implementations

An Exciting Era for Computer Architecture

> *"This is the most exciting time in computer architecture since the industrial exploitation of instruction-level parallelism in microprocessors 25 years ago."*

KEY TAKEAWAYS

Domain-Specific is Key

Custom architectures now outperform general-purpose for specific workloads

Open Source Rising

RISC-V and open architectures have longer viable lifetimes

New Technologies

2.5D stacking, NV memories, and optical interconnects expand possibilities

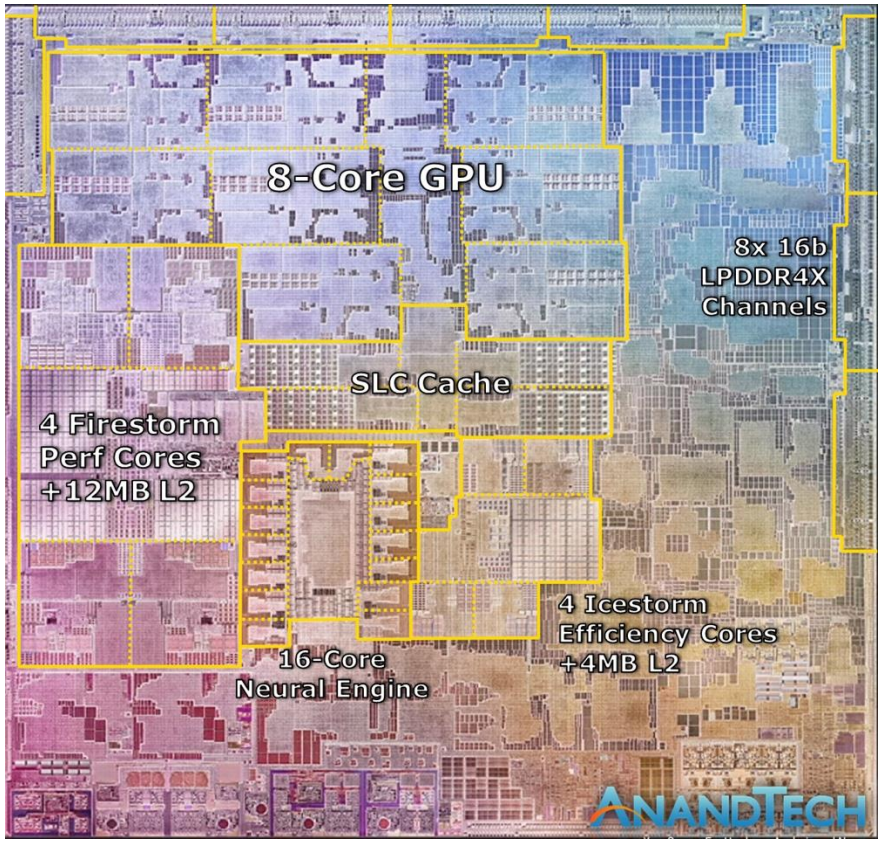
First Principles

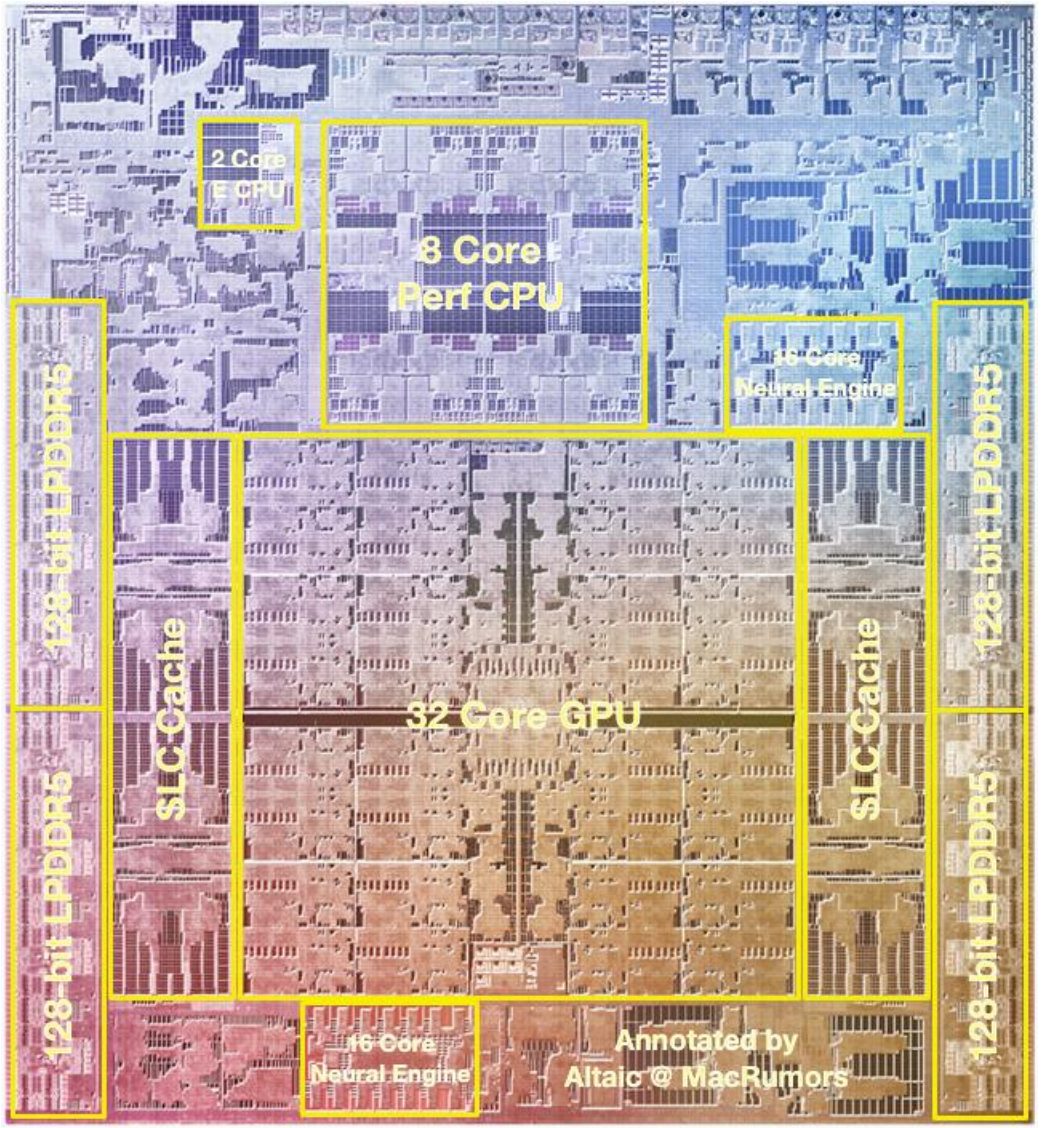
Fundamental design decisions must be reexamined in this new era

Textbook: The Sixth Edition and beyond

Updated for the post-Moore era with new chapters on domain-specific architectures

Processing Chip – Die Shots





Thank you!

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY

HYDERABAD