CS60003: High Performance Computer Architecture

Introduction and Current Trends in Computer Architecture

IIT KHARAGPUR

Instructor:
Prof. Rajat Subhra Chakraborty
Professor
Dept. of Computer Science and Engineering
Indian Institute of Technology Kharagpur
Kharagpur, West Bengal, India 721302
E-mail: rschakraborty@cse.iitkgp.ac.in

What is a Digital Computer?

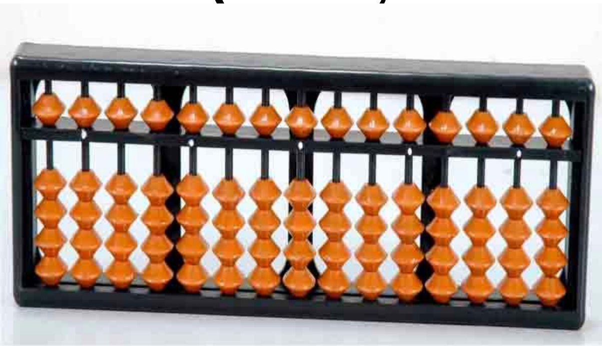
- **<u>(Modern) Definition:</u> a programmable electronic device capable of storing and processing (digitized) information**

https://unsplash.com/s/photos/desktop-computer

However, computers were not always "programmable", nor "electronic", nor "digital"!

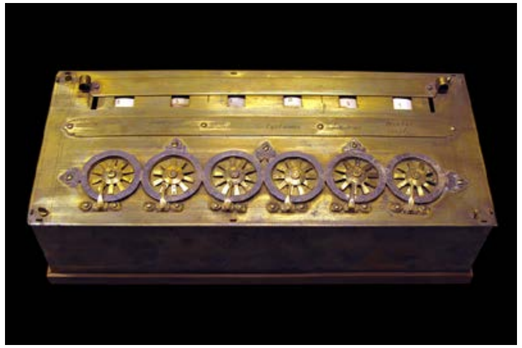
Computers Since Ancient Age

- ***Abacus*** **(China, ~3000 BC), still in use**



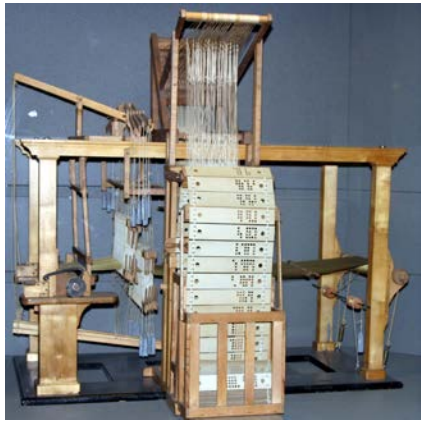
https://supermaths.co.uk/abacus-tool/

- ***Pascaline*** **(France, 1642), still usable**



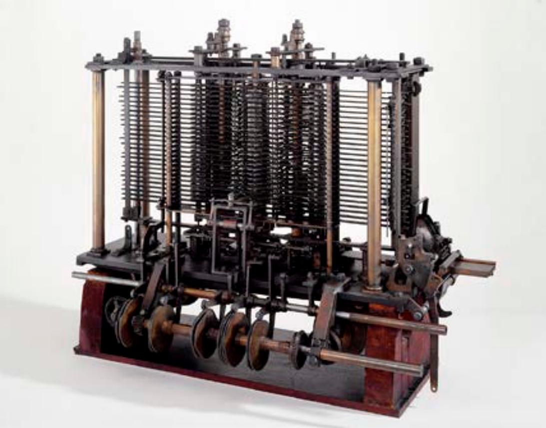
https://commons.wikimedia.org/wiki/File:Pascaline-CnAM_823-1-IMG_1506-black.jpg

- ***Jacquard's Loom*** **(*Jacquard,* France, 1801)**



https://addiator.blogspot.com/2011/10/jacquards-loom-and-stored-programme.html

- ***Analytical Engine*** **(Babbage, Britain, 1837)**



https://commons.wikimedia.org/wiki/File:Babbages_Analytical_Engine,_1834-1871._(9660574685).jpg

Computing in the 21st Century

- **Computers come in a bewildering variety….**

- **The traditional role of a computer as an accelerator for “mathematical computations” is limited nowadays**
- **A vast majority of computers are used nowadays to access information on the information, multimedia and entertainment**
- **The exponential growth on the Internet and fast pace of VLSI development has helped**

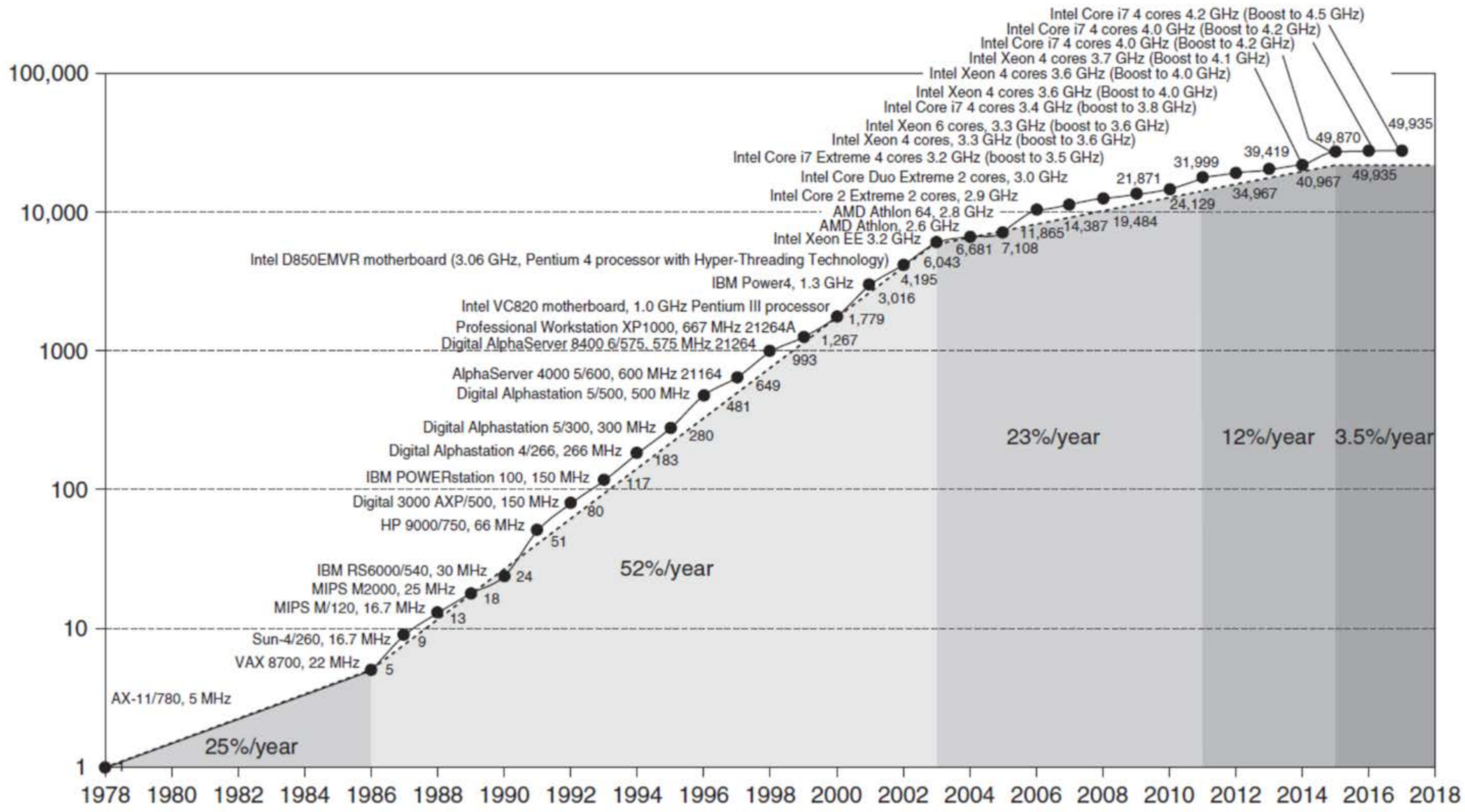
Computer Technology

- **Performance improvements caused by:**
  - **Improvements in semiconductor technology**
    - **Transistor size decrease, clock speed increase, …..**

- **Improvements in Computer Architecture**
  - **Enabled by platform-independent OSes (e.g. UNIX, Linux, ….)**
  - **Adoption of RISC architectures (explicit/implicit, e.g. x86)**

- **Together have enabled:**
  - **Lightweight computers**
  - **Productivity-based managed/interpreted programming languages**

Improvement to Single Processor Performance



H&P CA:A QA (6th. Ed.)

For SPEC integer benchmarks

Factor of 10,000 improvement over 40 years (1978-2018)!

However, rate of improvement is decreasing in recent past!

Clock speed has almost stagnated!

Parallelism

- **Parallelism in Computer Architecture**
  - **Instruction-Level Parallelism (ILP)**
    - **exploited by almost all modern processors**
  - **Thread-Level Parallelism**
    - **exploited by multi-core processors**
  - **Data-level Parallelism**
    - **exploited by Vector architectures, Graphic Processor Units (GPUs), multimedia extensions of instruction sets**
  - **Request-Level Parallelism**
    - **exploited by clusters and warehouse-scale computers**

- **We will cover the first two types of parallelism in class**
- **For data-level parallelism: CS60104 (High Perf. Para. Prog.)**

Flynn’s Taxonomy (1966)

- **Single instruction stream, single data stream (SISD)**
  - **Basic approach till ~2004**
  - **Exploited by almost all modern processors**
  - **Uses ILP techniques for high performance**
- **Single instruction stream, multiple data streams (SIMD)**
  - **Vector architectures**
  - **Multimedia extensions**
  - **Graphics processor units**
- **Multiple instruction streams, single data stream (MISD)**
  - **No commercial implementation**
- **Multiple instruction streams, multiple data streams (MIMD)**
  - **Tightly-coupled MIMD (e.g. multi-core processors)**
  - **Loosely-coupled MIMD**

RISC-V Instruction Set Architecture (ISA)

- **RISC-V introduced at UC Berkeley (Prof. Patterson and his team)**
  - **Classic RISC design**
  - **Load-Store design**
  - **Plenty of general-purpose registers**
  - **Novelty: extensible for advanced features**
- **RISC-V Registers (32 g.p., 32 f.p.)**

Register	Name	Use	Saver
x0	zero	constant 0	n/a
x1	ra	return addr	caller
x2	sp	stack ptr	callee
x3	gp	gbl ptr	
x4	tp	thread ptr	
x5-x7	t0-t2	temporaries	caller
x8	s0/fp	saved/ frame ptr	callee

Register	Name	Use	Saver
x9	s1	saved	callee
x10-x17	a0-a7	arguments	caller
x18-x27	s2-s11	saved	callee
x28-x31	t3-t6	temporaries	caller
f0-f7	ft0-ft7	FP temps	caller
f8-f9	fs0-fs1	FP saved	callee
f10-f17	fa0-fa7	FP arguments	callee
f18-f27	fs2-fs21	FP saved	callee
f28-f31	ft8-ft11	FP temps	caller

H&P CA:A QA (6^{th}. Ed.)

Trends in Technology

- Integrated circuit technology (Moore's Law)
  - Transistor density: 35%/year
  - Die size: 10-20%/year
  - Integration overall: 40-55%/year

- DRAM capacity: 25-40%/year (slowing)
  - 8 Gb (2014), 16 Gb (2019), possibly no 32 Gb

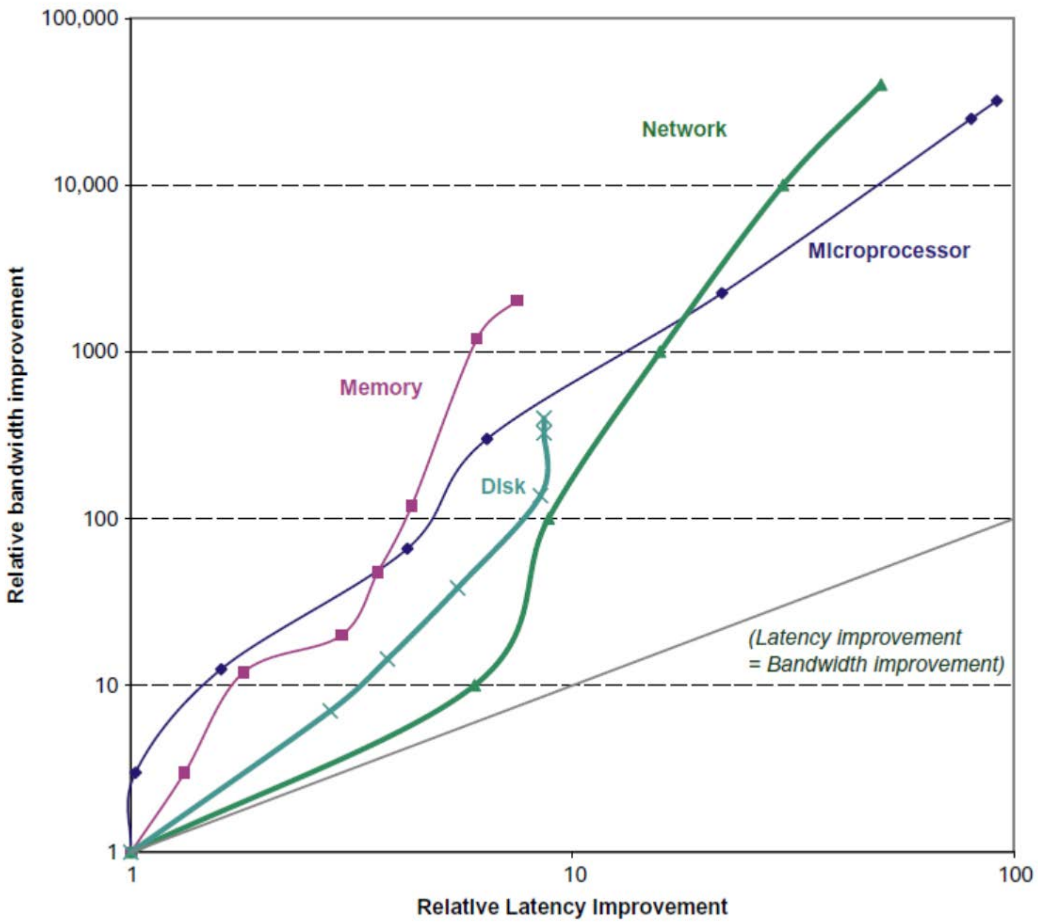
- Flash capacity: 50-60%/year (most promising!)
  - 8-10X cheaper/bit than DRAM

- Magnetic disk capacity: recently slowed to less than 5%/year
  - 8-10X cheaper/bit then Flash
  - 200-300X cheaper/bit than DRAM

Bandwidth and Latency

- **“Bandwidth”/”Throughput”**
  - **Total work done in a given time**
  - **32,000-40,000X improvement for processors**
  - **300-1200X improvement for memory and disks**

- **“Latency”/”Response Time”**
  - **Time between start and completion of an event**
  - **50-90X improvement for processors**
  - **6-8X improvement for memory and disks**

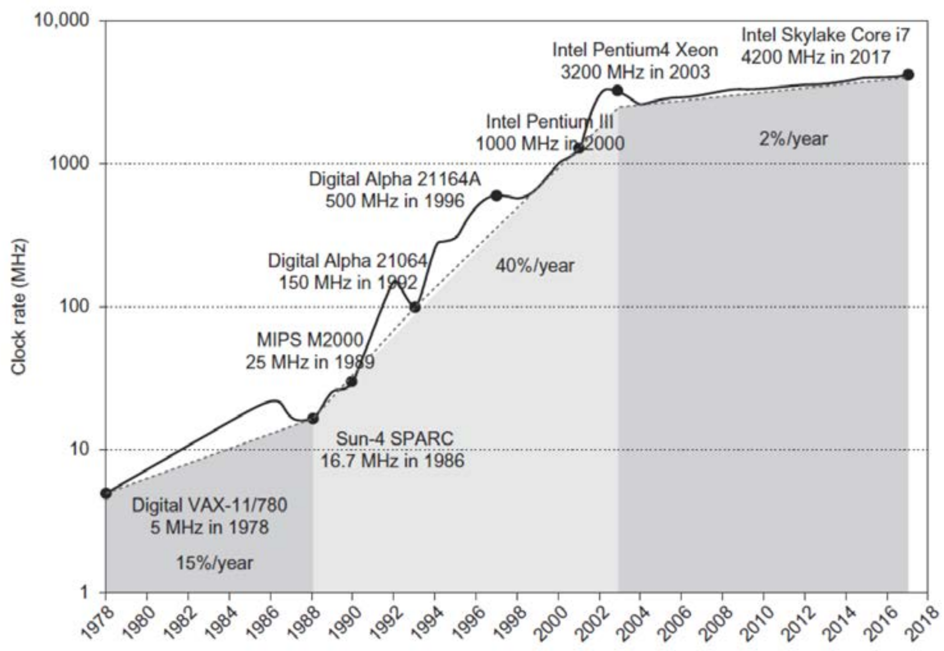


H&P CA:A QA (6th. Ed.)

Dynamic Energy and Power

- Dynamic energy
  - Transistor switch from 0 → 1 or 1 → 0
  - ½ x Capacitive load x Voltage^2
- Dynamic power
  - ½ x Capacitive load x Voltage^2 x Frequency switched
- Reducing clock rate reduces power, not energy

- Intel 80386 consumed ~2 W
- 3.3 GHz Intel Core i7 consumes 130 W
- Heat must be dissipated from 1.5 x 1.5 cm chip
- This is the limit of what can be cooled by air

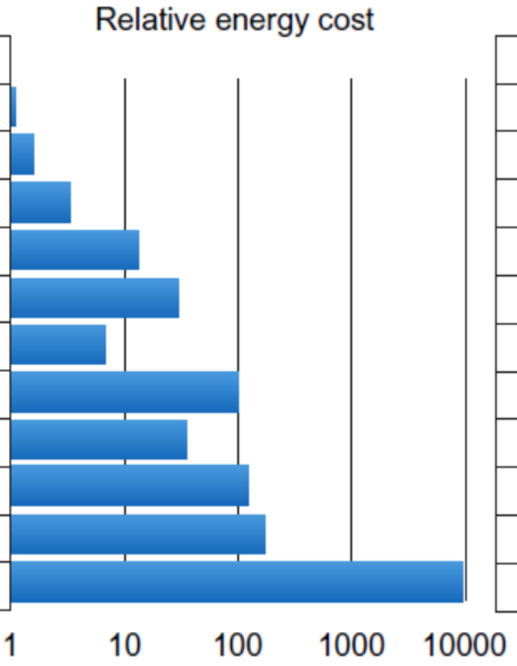


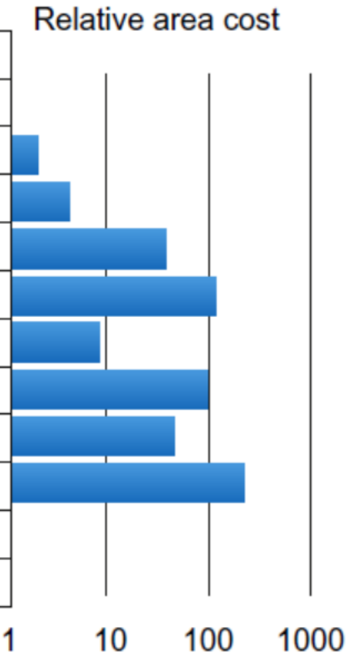
H&P CA:A QA(6th. Ed.)

Static Power

- **Static power consumption**
  - **25-50% of total power**
  - **Current$_{static}$ x Voltage**
  - **Scales with number of transistors**
  - **To reduce: power gating**

Operation:	Energy (pJ)	Area (μm^2)
8b Add	0.03	36
16b Add	0.05	67
32b Add	0.1	137
16b FB Add	0.4	1360
32b FB Add	0.9	4184
8b Mult	0.2	282
32b Mult	3.1	3495
16b FB Mult	1.1	1640
32b FB Mult	3.7	7700
32b SRAM Read (8KB)	5	N/A
32b DRAM Read	640	N/A

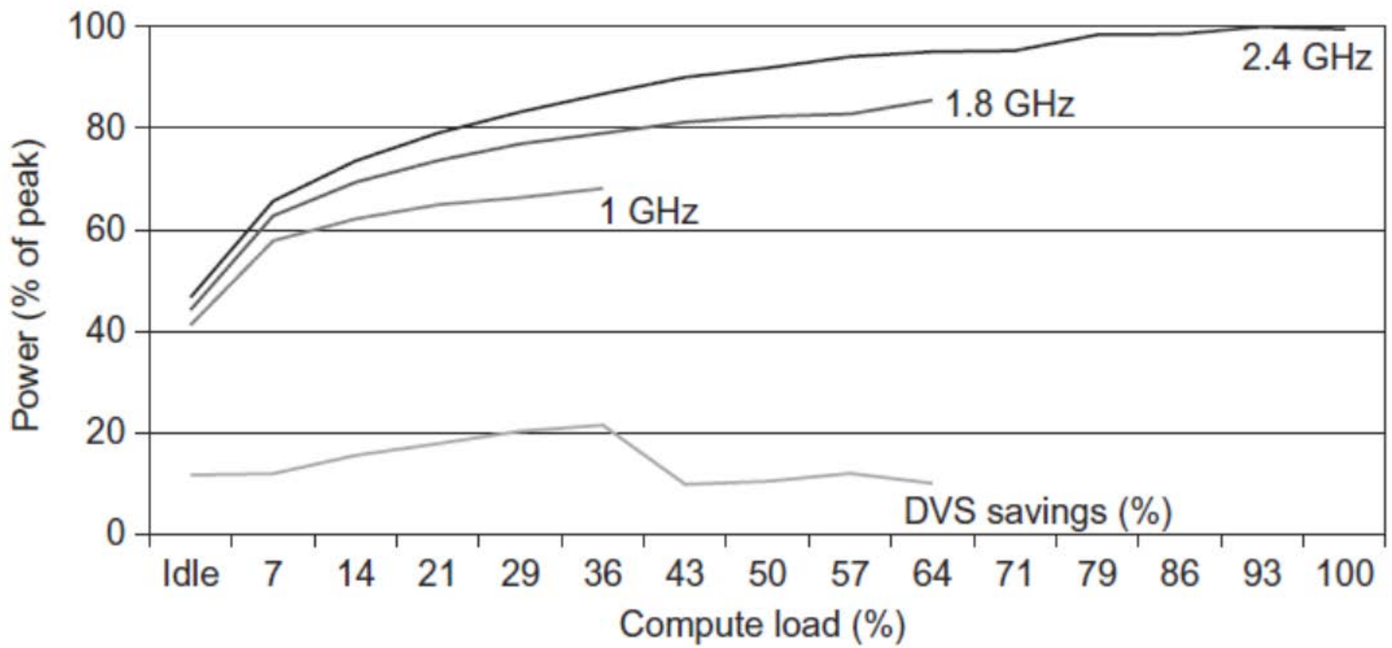




H&P CA:A QA (6th. Ed.)

Reducing Power

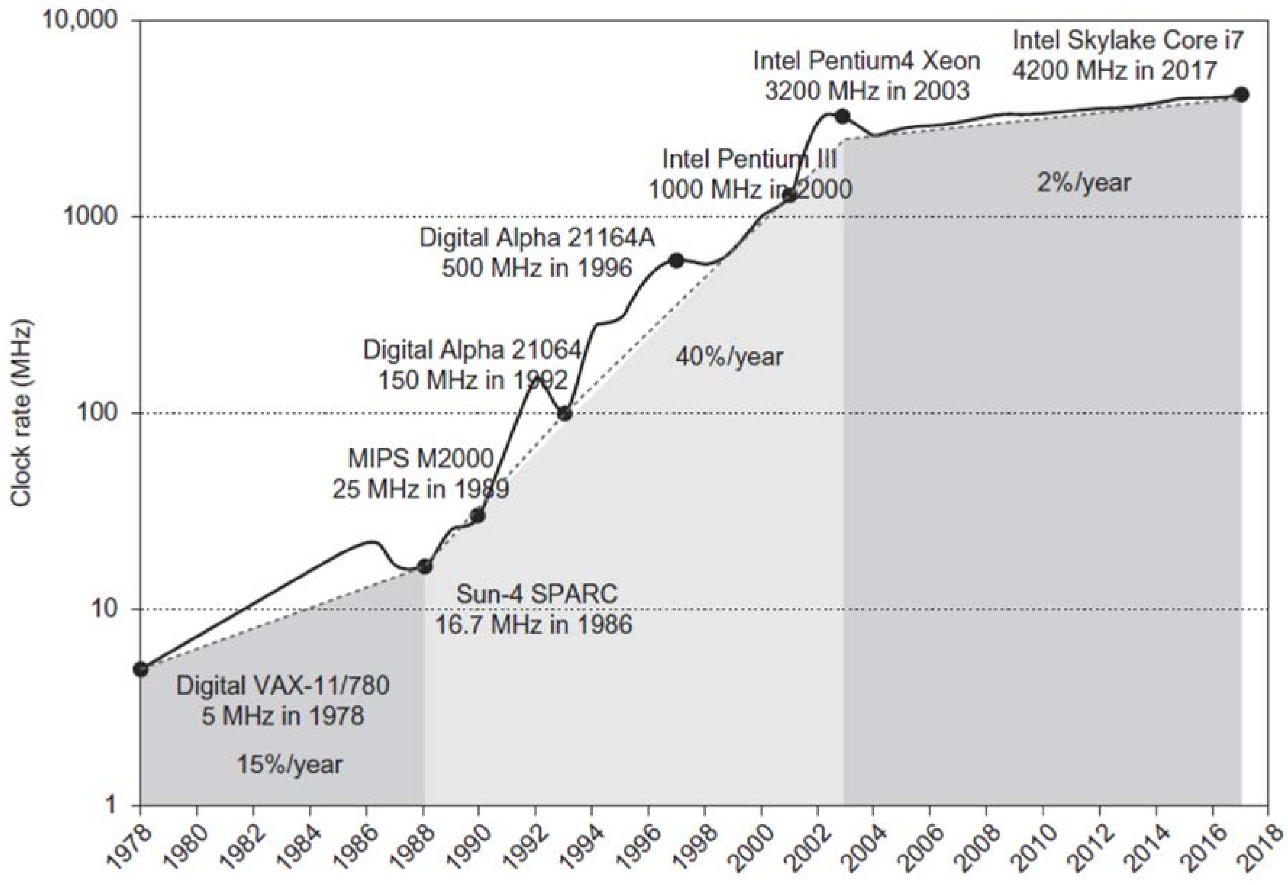
- **Techniques for reducing power:**
  - **Do nothing well**
  - **Dynamic Voltage-Frequency Scaling**
  - **Low power state for DRAM, disks**
  - **Overclocking, turning off cores**



H&P CA:A QA (6th. Ed.)

CPU Clock Frequency Trends (dictated by Power)

- **Intel 80386 consumed ~2 W**
- **3.3 GHz Intel Core i7 consumes ~130 W**
- **Heat must be dissipated from 1.5 x 1.5 cm chip**
- **This is the limit of what can be cooled by air**



H&P CA:A QA (6th. Ed.)

What will We Learn in This Course?

- Principles and techniques of high-level design ("architecture") and some low-level hardware details ("microarchitecture") of modern CPUs
- This is PG level course that assumes you have already been exposed to basics of digital logic, computer organization, Instruction Set, etc.
  - e.g. you understand what is a register, an ALU, assembly language programs, etc.
- Quantitative analysis would be emphasized throughout the course
  - Always keep a calculator handy ☺
- Analysis would be primarily based on experimental/simulation data
- Focus on understanding how processor design has evolved to improve performance

Course Logistics

- **4 hours of lecture every week**
- **Each major topic will be followed by a class-test during regular class hour**
  - **Class-test would be announced 7-10 days in advance**
- **Programming assignments: 1-2 weeks deadline**
  - **gem5 simulator (possibly)**
- **Teacher's assessment: 30 marks**
  - **Class-tests, programming assignments, class participation, etc.**
- **Mid-semester Exam: 30 marks (after scaling down)**
- **End-semester Exam: 40 marks (after scaling down)**
  - **End-semester exam will cover entire syllabus**

Textbook and Pre-requisite

- **<u>Required Textbook:</u>**

  Computer Architecture: A Quantitative Approach **(6th ed.) [Indian edition]**

  by John L. Hennessy and David A. Patterson

  Publisher: Morgan Kaufman
- **Example Architecture: RISC-V**

- **Other study materials:**
  - **Class notes**
  - **Handouts**
  - **Online resources for simulators, etc.**
- **Pre-requisite: Appendix-A to Appendix-C of the textbook**

- **Don't worry if you do not have all the pre-requisite knowledge, we will cover the background as needed!**


Thank
you