

FUNDAMENTALS OF QUANTITATIVE DESIGN

Contents and objective

- Getting familiar with terms & concepts
- Overall trends
- Figure of merit (metrics)
- Computer design principles

Performance (vs. VAX-11/780)

100000

10000

1000

100

10

1

1978 1980 1982 1984 1986 1988 1990 1992 1994 1996 1998 2000 2002 2004 2006 2008 2010 2012

AX-11/780, 5 MHz

25%/year

1.5, VAX-11/785

52%/year

22%/year

Intel D850EMVR motherboard (3.06 GHz, Pentium 4 processor with Hyper-Threading Technology)

Intel VC820 motherboard, 1.0 GHz Pentium III processor

Professional Workstation XP1000, 667 MHz 21264A

Digital AlphaServer 8400 6/575, 575 MHz 21264

AlphaServer 4000 5/600, 600 MHz 21164

Digital Alphastation 5/500, 500 MHz

Digital Alphastation 5/300, 300 MHz

Digital Alphastation 4/266, 266 MHz

IBM POWERstation 100, 150 MHz

Digital 3000 AXP/500, 150 MHz

HP 9000/750, 66 MHz

IBM RS6000/540, 30 MHz

MIPS M2000, 25 MHz

MIPS M/120, 16.7 MHz

Sun-4/260, 16.7 MHz

VAX 8700, 22 MHz

Intel Xeon 6 cores, 3.3 GHz (boost to 3.6 GHz)

Intel Xeon 4 cores, 3.3 GHz (boost to 3.6 GHz)

Intel Core i7 Extreme 4 cores 3.2 GHz (boost to 3.5 GHz)

Intel Core Duo Extreme 2 cores, 3.0 GHz

Intel Core 2 Extreme 2 cores, 2.9 GHz

AMD Athlon 64, 2.8 GHz

AMD Athlon, 2.6 GHz

Intel Xeon EE 3.2 GHz

IBM Power4, 1.3 GHz

24,129

21,871

19,484

14,387

11,865

7,108

6,681

6,043

4,195

3,016

1,779

1,267

993

649

481

280

183

117

80

51

24

18

13

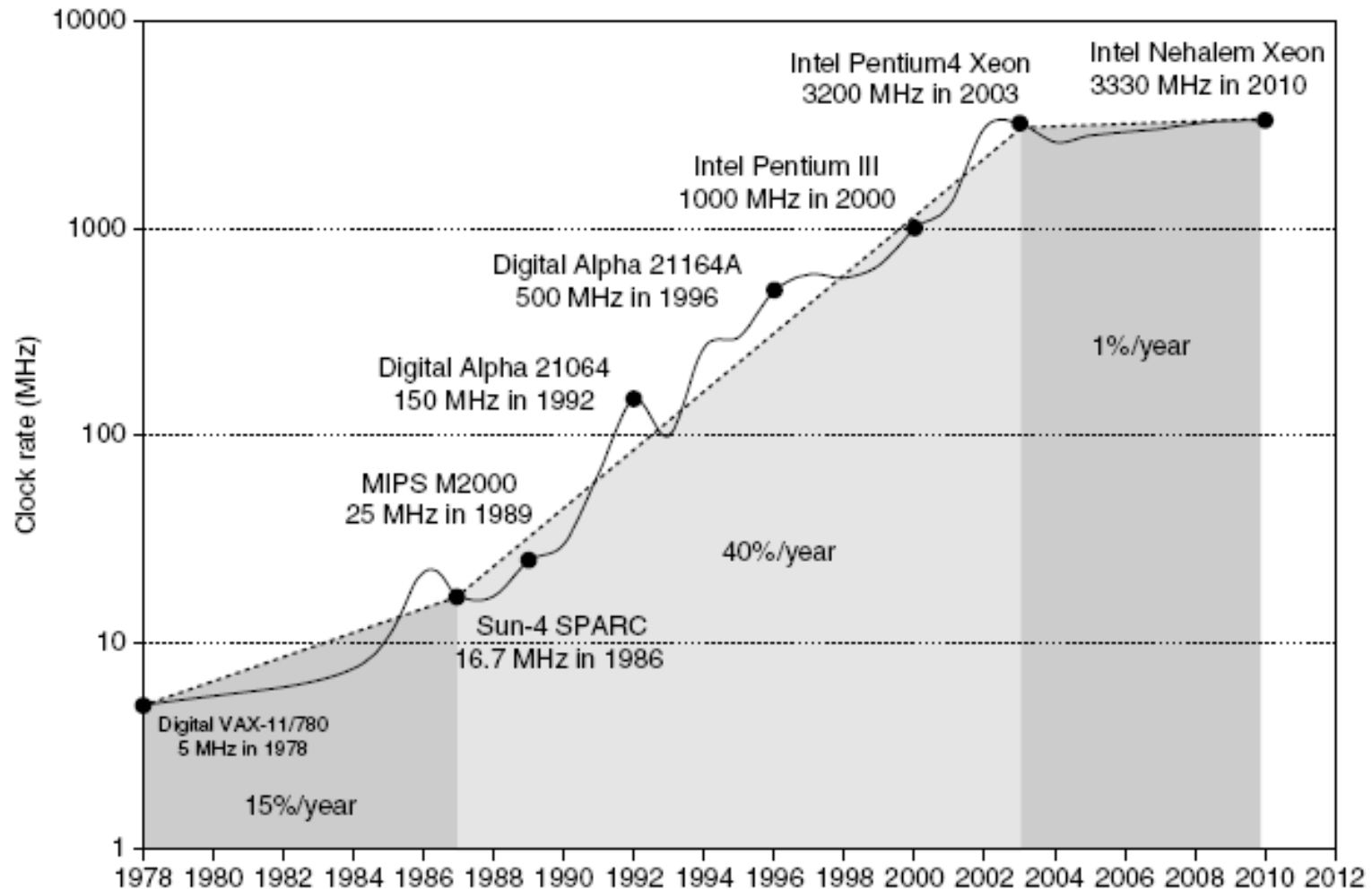
9

5

Points to Note

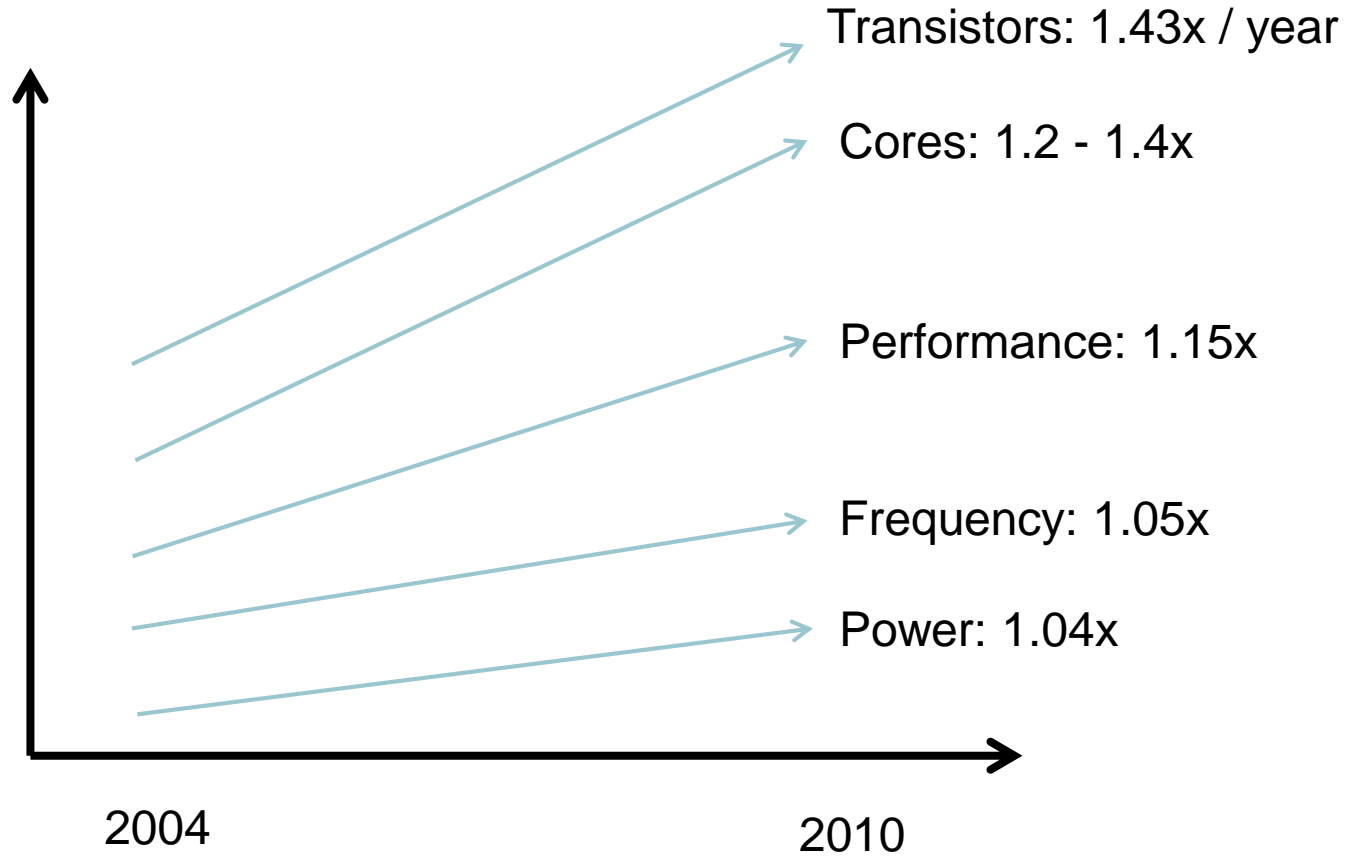
- ⑩ The 52% growth per year is because of faster clock speeds and architectural innovations (led to 25x higher speed)
- ⑩ Clock speed increases have dropped to 1% per year in recent years
- ⑩ The 22% growth includes the parallelization from multiple cores
- ⑩ Moore's Law: transistors on a chip double every 18-24 months

Clock Speed Increases



Source: H&P textbook

Recent Microprocessor Trends



Source: Micron University Symp.

Classes of Computers

- Personal Mobile Device (PMD)
 - e.g. smart phones, tablet computers
 - Emphasis on energy efficiency and real-time
- Desktop Computing
 - Emphasis on price-performance
- Servers
 - Emphasis on availability, scalability, throughput
- Clusters / Warehouse Scale Computers
 - Emphasis on availability and price-performance
 - Sub-class: Supercomputers, emphasis: floating-point performance and fast internal networks
- Embedded Computers
 - Emphasis: price

Choosing Programs to Evaluate Perf.

- Toy benchmarks
 - e.g., quicksort, puzzle
 - No one really runs. Scary fact: used to prove the value of RISC in early 80's
- Synthetic benchmarks
 - Attempt to match average frequencies of operations and operands in real workloads.
 - e.g., Whetstone, Dhrystone
 - Often slightly more complex than kernels; But do not represent real programs
- Kernels
 - Most frequently executed pieces of real programs
 - Good for focusing on individual features not big picture
 - Tend to over-emphasize target feature
- Real programs
 - e.g., gcc, spice, SPEC2006 (standard performance evaluation corporation), TPCC, TPCD, PARSEC, SPLASH

METRICS

Transistor dimension (Area)

- Feature size
 - Also called geometry, process node
 - Minimum size of transistor or wire in x or y dimension
- 10 microns in 1971 to .032 microns in 2011, .016 microns in 2016
- Leads to chip-miniaturization
- Allows fitting more transistors on a chip

Throughput and Latency

- Bandwidth or throughput
 - Total work done in a given time
 - 10,000-25,000X improvement for processors
 - 300-1200X improvement for memory and disks
- Latency or response time
 - Time between start and completion of an event
 - 30-80X improvement for processors
 - 6-8X improvement for memory and disks

Measuring Performance

- Typical performance metrics:
 - Response time
 - Throughput
- Speedup of X relative to Y
 - $\text{Execution time}_Y / \text{Execution time}_X$
- Execution time
 - Wall clock time: includes all system overheads
 - CPU time: only computation time