Computer Architecture

Lecture 1: Introduction

Iakovos Mavroidis

Computer Science Department

University of Crete

Outline

- Logistics
- CPU Evolution
- Course goal (what is Computer Architecture?)

Course Administration

- Instructors
  - Iakovos Mavroidis (jacob@ics.forth.gr)
  - Christos Sotiriou(sotiriou@csd.uoc.gr,)
- Teaching Assistant
  - Βασιλάκης Ευάγγελος
- Lectures
  - Monday-Wednesday, 13:00-15:00
  - Friday, 13:00-15:00 on a need basis
    - Monday 8/10 → Friday 12/10
    - Monday 15/10 → Friday 19/10
    - Wednesday 17/10 → TBD

Course Administration

- Website
  - http://www.csd.uoc.gr/~hy425
- Project
  - TBA
- Sources
  - **Course textbook:** Hennessy and Patterson, Computer Architecture, A Quantitative Approach.

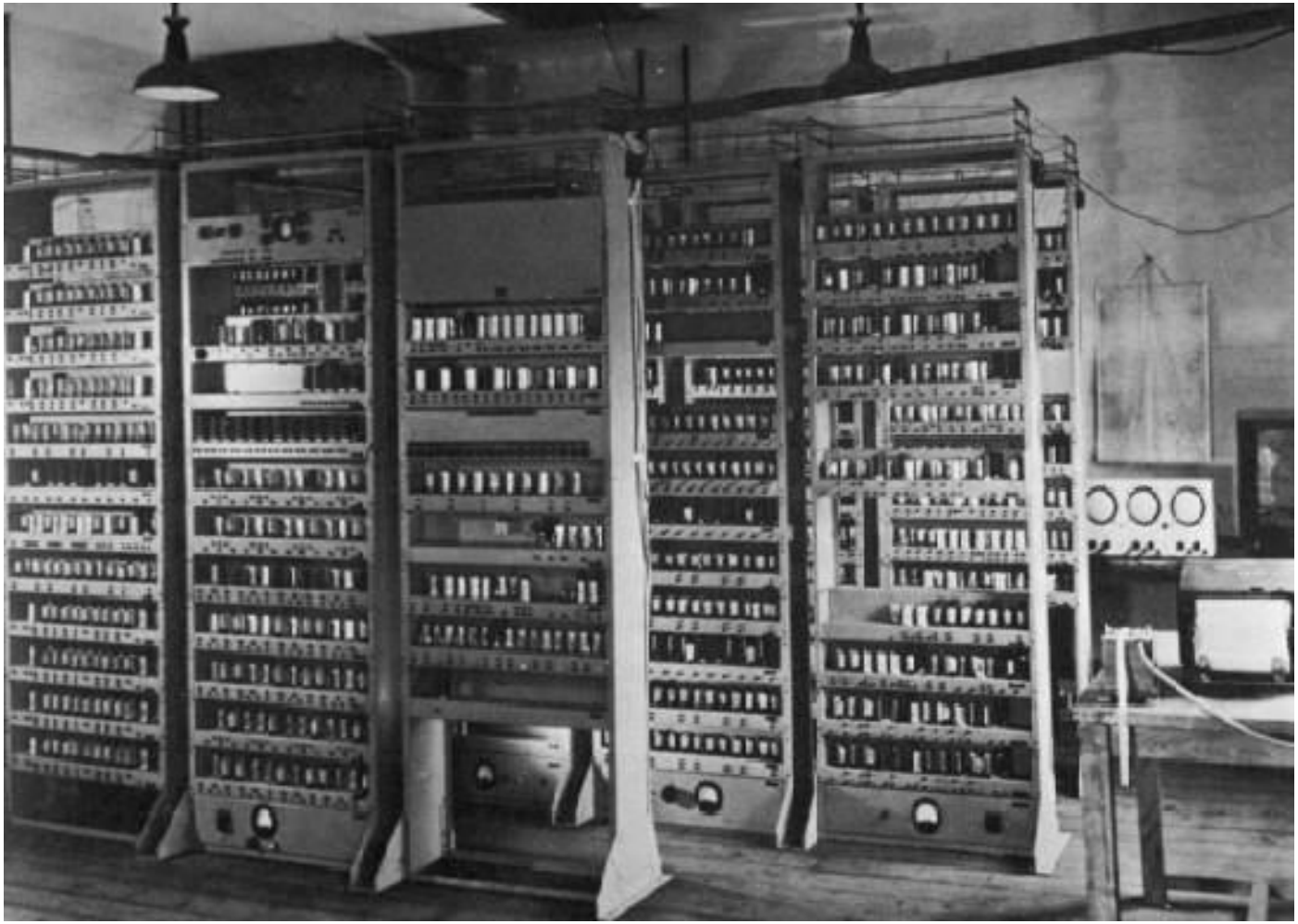
    3rd Edition Available in Greek (Tziolas publishers, translation by D. Pnevmatikatos, D. Serpanos and G. Stamoulis). ISBN 97896041807693

Tentative Topics

- 1.5 weeks: Fundamentals, ISA, Pipelining (*review*)
- 2 weeks: Instructional Level Parallelism
- 1 weeks: Branch Prediction
- 2 week: Multiple Issue, VLIW, Vector Processors
- 1.5 weeks: Multithreading, Latency Tolerance
- 1 week: Memory Hierarchy
- 1 week: Multi-processors

10-11 weeks, 20-22 lectures

History in Computer Devices



EDSAC, University of Cambridge, UK, 1949

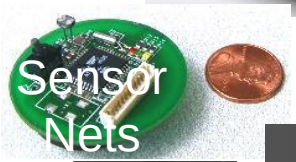
Computing Systems Today

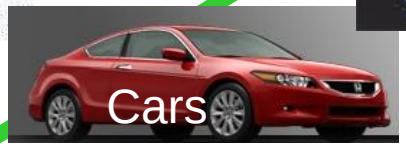
- The world is a large parallel system
  - Microprocessors in everything
  - Vast infrastructure behind them

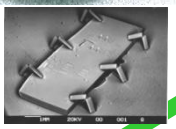
Internet Connectivity

Scalable, Reliable, Secure Services

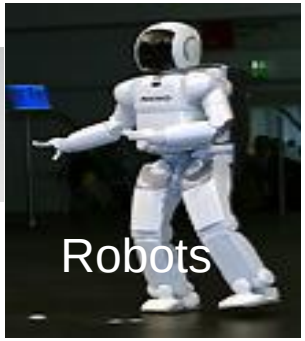
Databases
Information Collection
Remote Storage
Online Games
Commerce
…







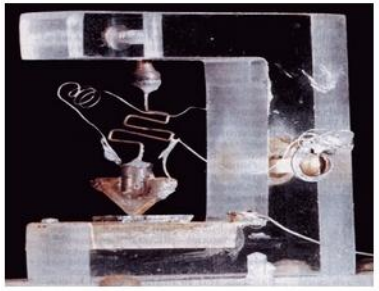
MEMS for Sensor Nets



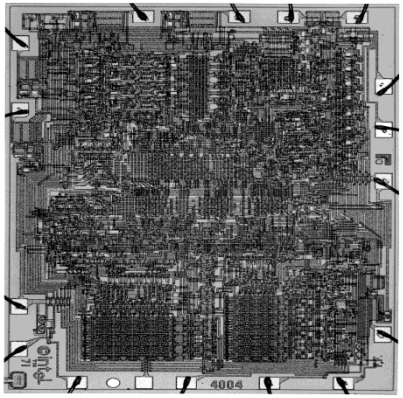
Improvement in Computer

- Radical progress in computers due to:
  - Technological improvements
    - steady
  - Better computer architectures
    - less consistent

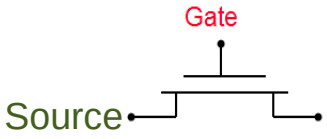
Technology: Transistor Revolution

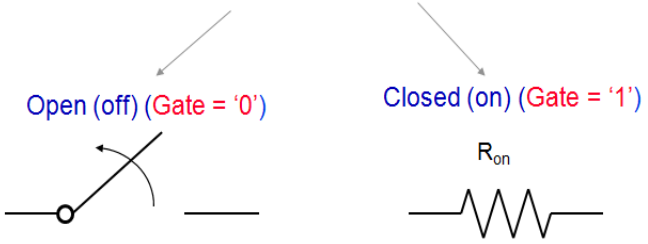


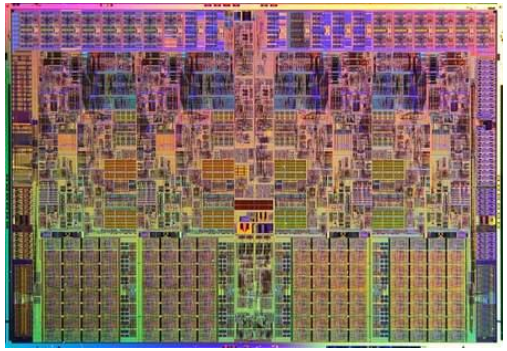
Bell Labs, 1948
First Transistor



Intel 4004, 1971 (Moore, Noyce)
2,300 transistors
740KHz operation
10µm (=10000nm) PMOS technology







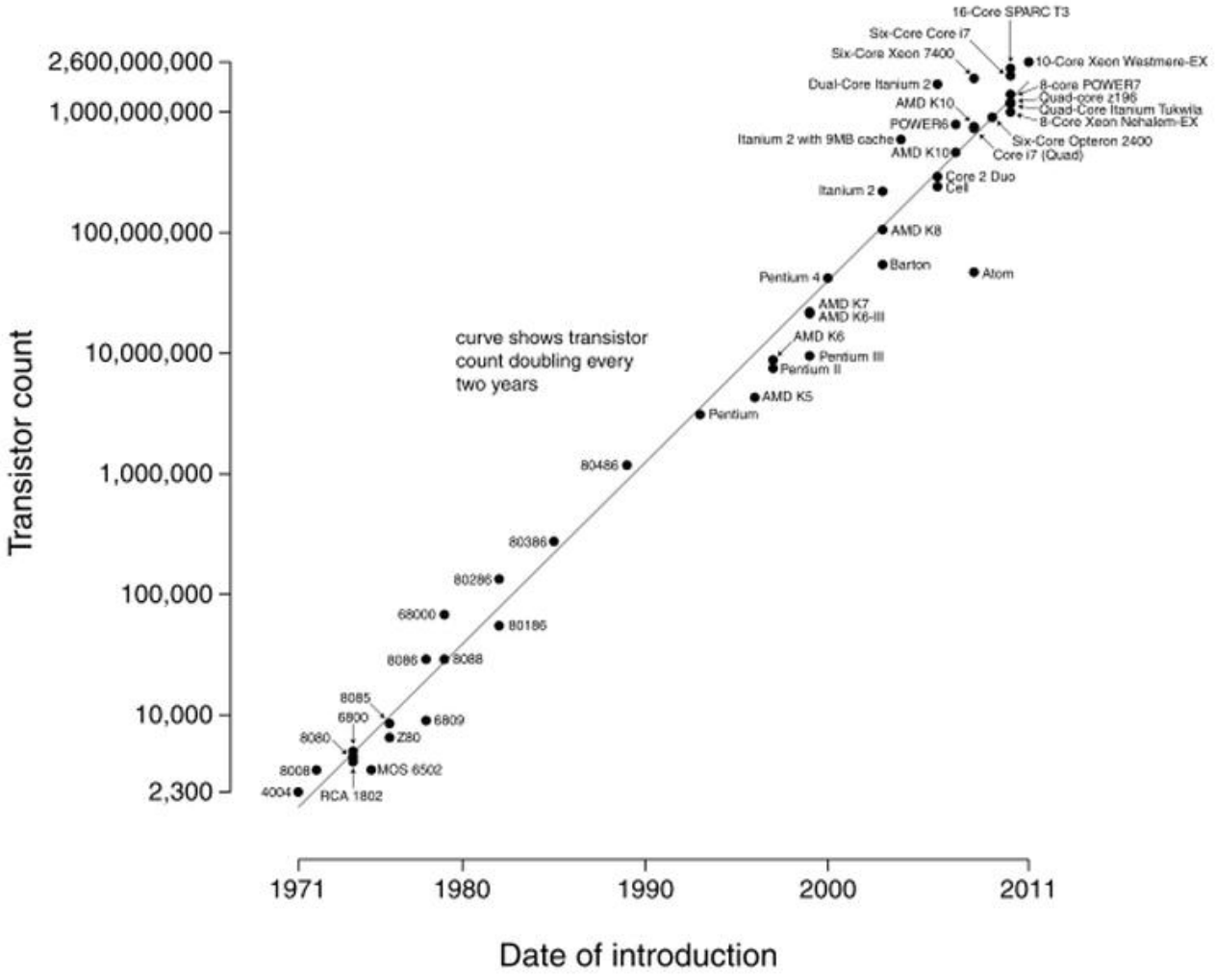
Intel Core i7, 2011
2,600,000 transistors
3.4GHz
32nm

Technology: Moore's Law

- In 1965, Gordon Moore predicted that the number of transistors that can be integrated on a die would double every 18 months (i.e., grow exponentially with time).

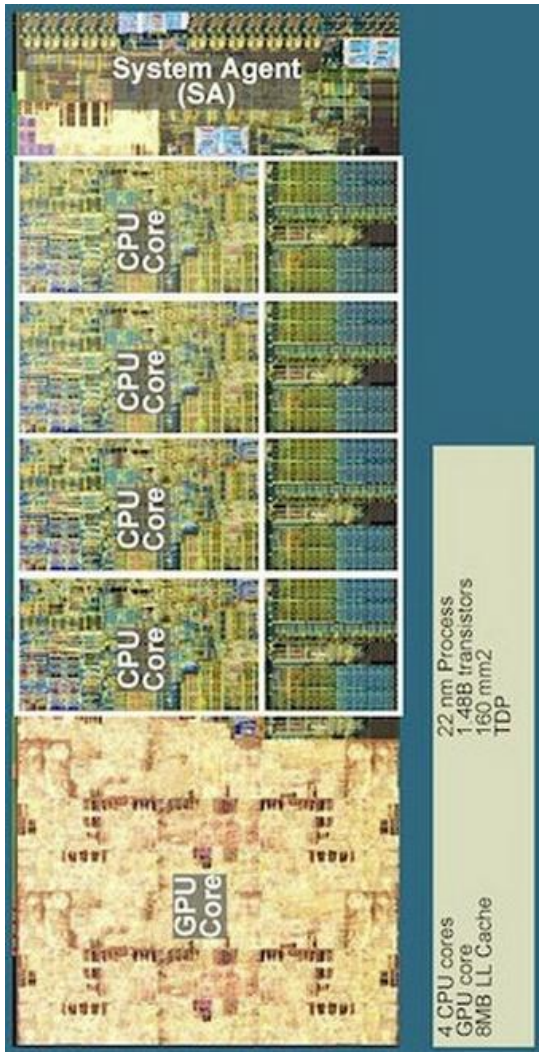
- He made a prediction that semiconductor technology will double its effectiveness every 18 months.

Technology: Transistor Count



Technology constantly on the move

- Num of transistors not limiting factor
  - Currently ~ 1+ billion transistors/chip
  - Problems:
    - Too much Power, Heat, Latency
    - Not enough Parallelism
- 3-dimensional chip technology?
  - Sandwiches of silicon
  - “Through-Vias” for communication
- On-chip optical connections?
  - Power savings for large packets
- The Intel® Core™ i7 microprocessor (“Ivy Bridge”)
  - 4 cores + GPU
  - 22 nm, tri-gate (“3D”) transistors
  - 1.4B Transistors
  - Shared L3 Cache - 8MB
  - L2 Cache - 1MB (256K x 4)



Transistor size trends and questions

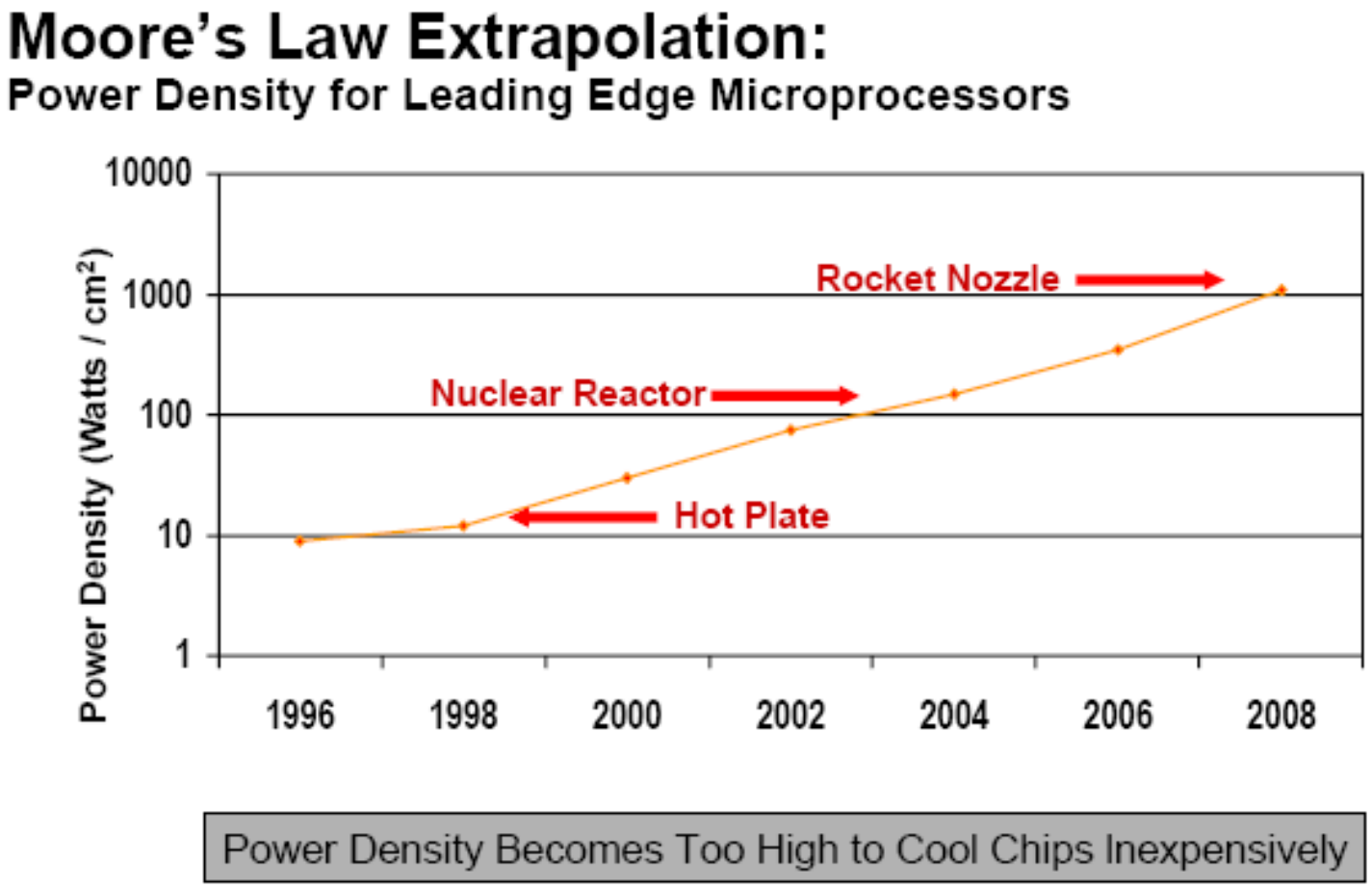
Feature sizes, higher performance?

- Transistor size went down from 10 micros to 45 nanometers
- Quadratic increase in density, linear drop in feature size
- Linear increase in transistor performance

Where is the catch?

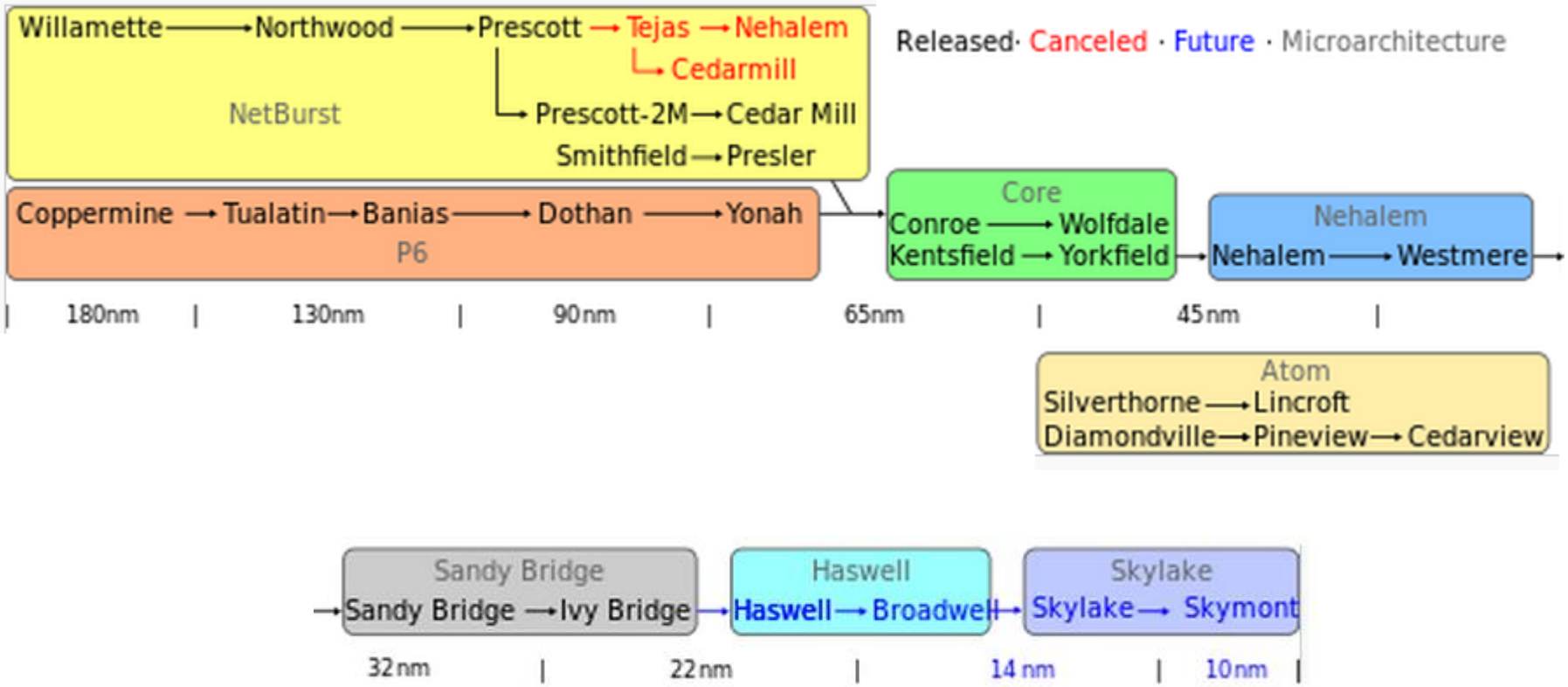
- Lower voltage to maintain safe operation
- Higher resistance and capacitance per unit of length
- Shorter wires but with higher resistance/capacitance
- Wire delays improving poorly compared to transistors

Limiting Force: Power Density

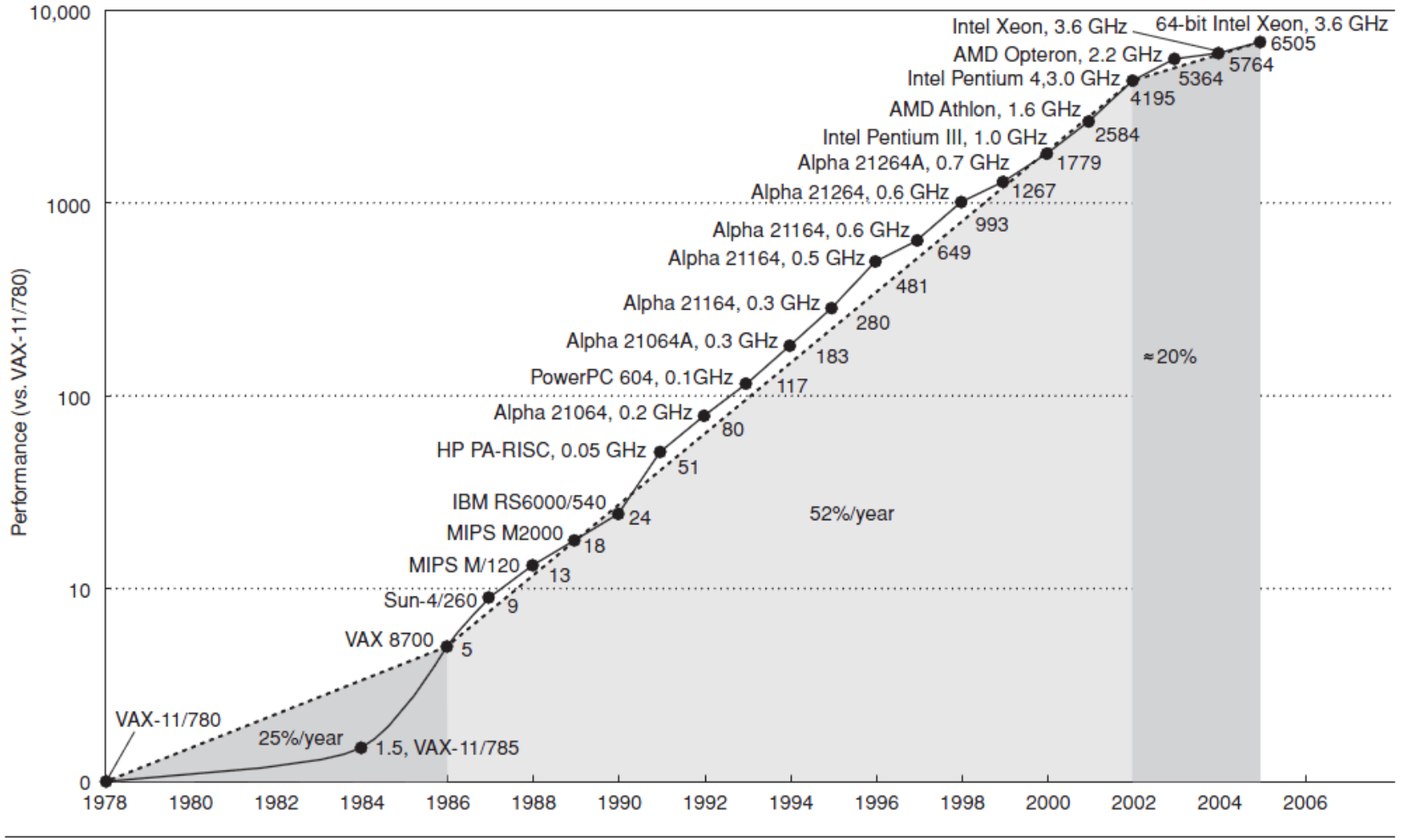


Source: Shekhar Borkar, Intel Corp

Technology vs microarchitecture: Intel CPU evolution

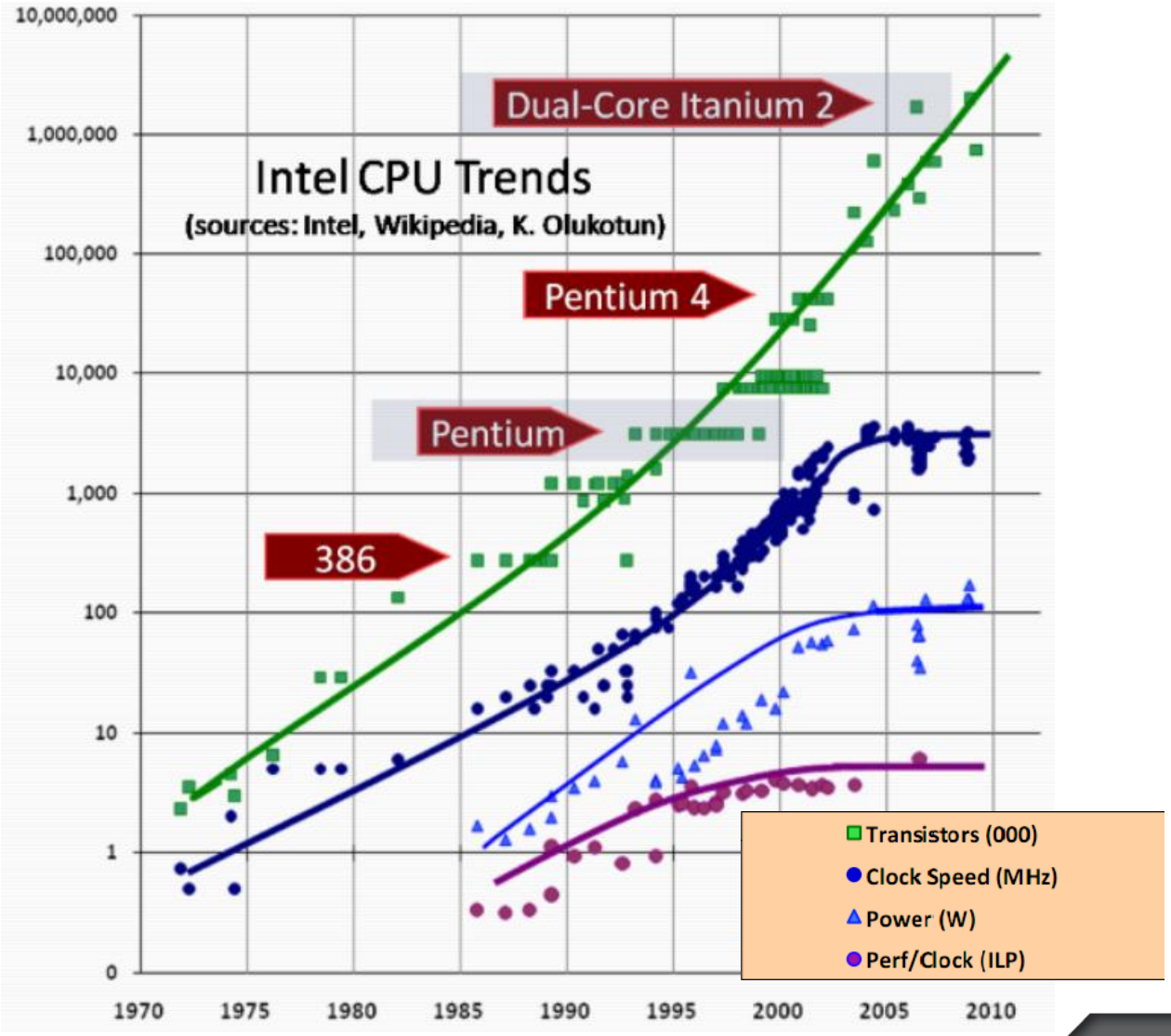


Crossroads: Uniprocessor Performance



- VAX : 25%/year 1978 to 1986
- RISC + x86: 52%/year 1986 to 2002
- RISC + x86: ??%/year 2002 to present

Trends – All in one

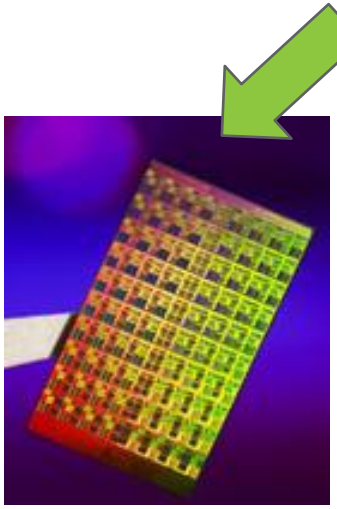


The End of the Uniprocessor Era

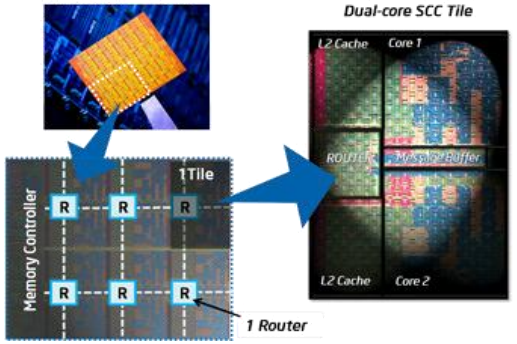
Single biggest change in the history of computing systems

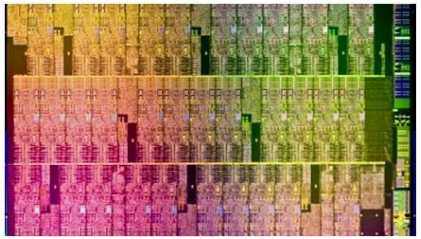
- "Power wall" Power expensive, Xtors free
  (Can put more on chip than can afford to turn on)
- "ILP wall" law of diminishing returns on more HW for ILP
- "Memory wall" Memory slow, multiplies fast
  (200 clock cycles to DRAM memory, 4 clocks for multiply)
- Power Wall + ILP Wall + Memory Wall = Brick Wall
  - Uniprocessor performance now 2X / 5(?) yrs

ManyCore Chips: The future is here



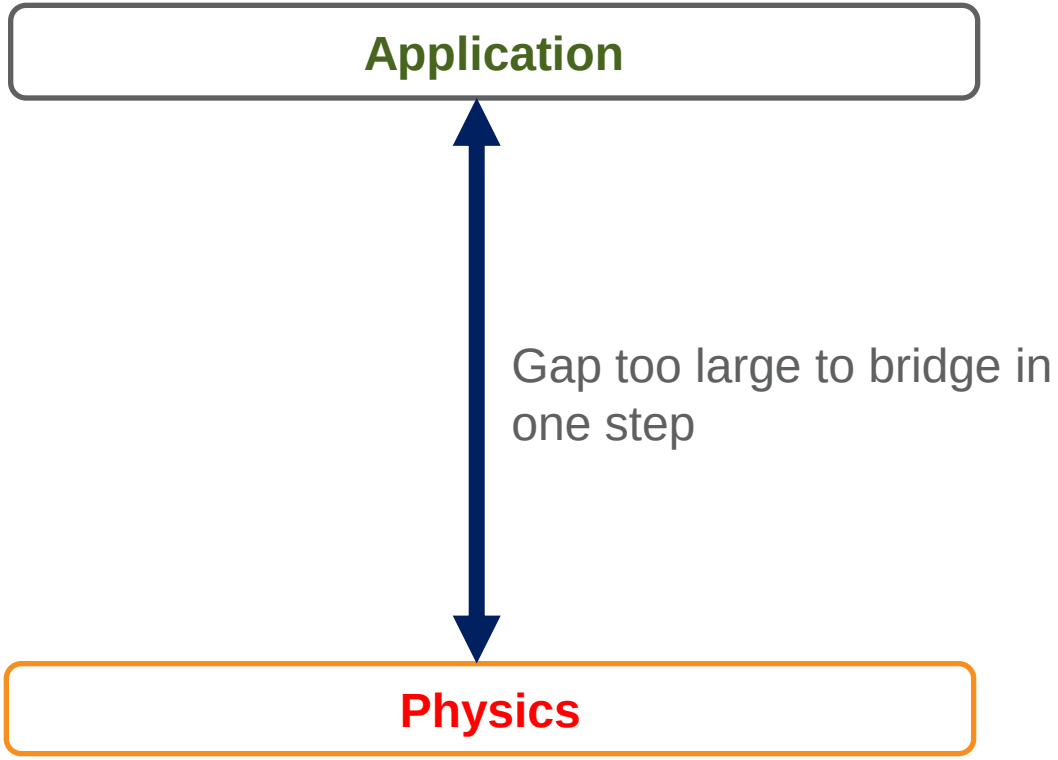
- Intel 80-core multicore chip, 2007
- Intel Single-Chip Cloud Computer (SCC), 48-cores, 2010
- Intel Many Integrated Core Architecture (MIC), 50-cores, 2012





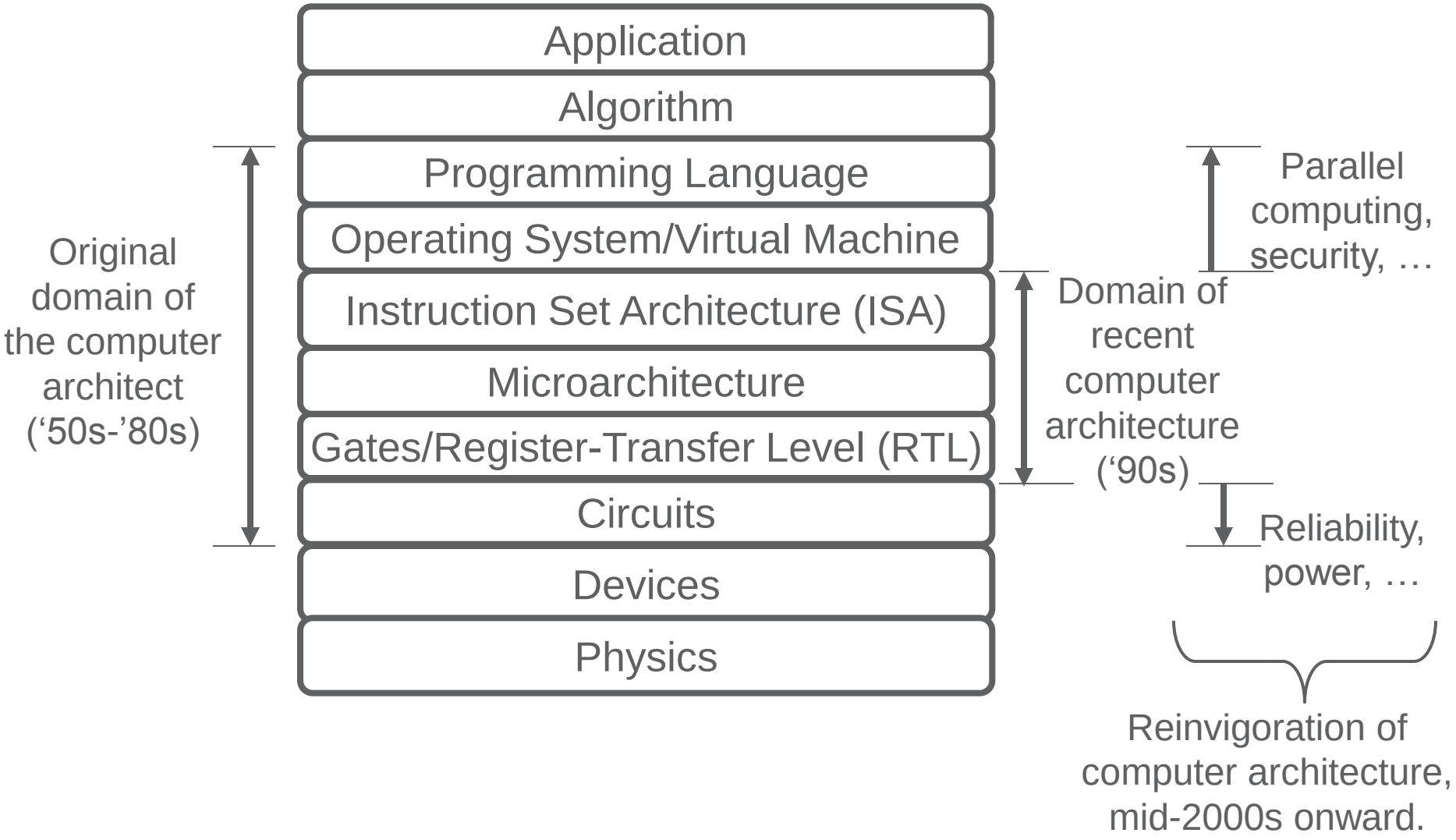
- "ManyCore" refers to many processors/chip
  - 64? 128? Hard to say exact boundary
- How to program these?
  - Use 2 CPUs for video/audio
  - Use 1 for word processor, 1 for browser
  - 76 for virus checking???
- Something new is clearly needed here…

What is Computer Architecture



In its broadest definition, computer architecture is the ***design of the abstraction layers*** that allow us to implement information processing applications efficiently using available manufacturing technologies.

Abstraction Layers in Modern Systems



Computer Architecture is an Integrated Approach

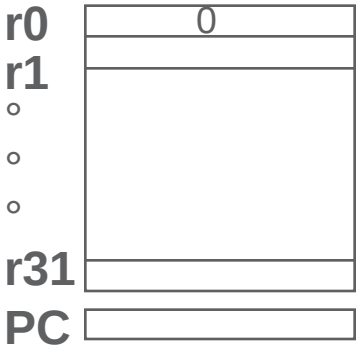
- What really matters is the functioning of the complete system
  - hardware, runtime system, compiler, operating system, and application
  - In networking, this is called the "End to End argument"
- Computer architecture is not just about transistors, individual instructions, or particular implementations
  - E.g., Original RISC projects replaced complex instructions with a compiler + simple instructions
- It is very important to think across all hardware/software boundaries
  - New technology ⇒ New Capabilities ⇒ New Architectures ⇒ New Tradeoffs
  - Delicate balance between backward compatibility and efficiency

Defining Computer Architecture (ISA)

Instruction Set Architecture

- ISAs converged to a common RISC paradigm
  - CISC ISAs implemented on RISC pipelines
- Load-store architectures, general-purpose registers
- Aligned memory addressing, simple addressing modes
- Byte, word, double-word operands
- Arithmetic, logic, control operations
- Fixed-length encoding

Example: MIPS R3000



Programmable storage

2^32 x bytes

31 x 32-bit GPRs (R0=0)

32 x 32-bit FP regs (paired DP)

PC

Data types ?

Format ?

Addressing Modes?

Arithmetic logical

Add, AddU, Sub, SubU, And, Or, Xor, Nor, SLT, SLTU,
AddI, AddIU, SLTI, SLTIU, AndI, OrI, XorI, *LUI*
SLL, SRL, SRA, SLLV, SRLV, SRAV

Memory Access

LB, LBU, LH, LHU, LW, LWL,LWR
SB, SH, SW, SWL, SWR

Control

J, JAL, JR, JALR
BEq, BNE, BLEZ,BGTZ,BLTZ,BGEZ,BLTZAL,BGEZAL

32-bit instructions on word boundary

ISA vs Computer Architecture

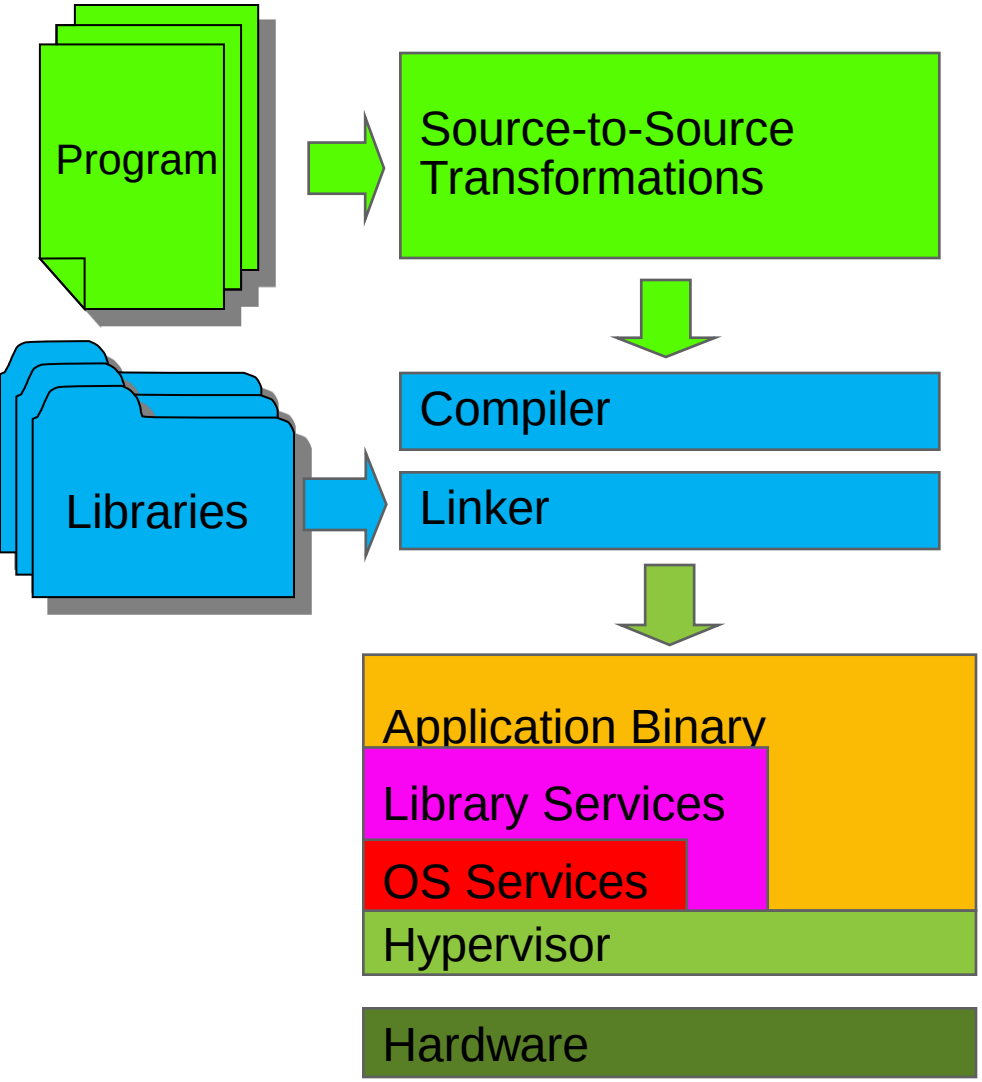
- Old definition of computer architecture = instruction set design
  - Other aspects of computer design called implementation
  - Insinuates implementation is uninteresting or less challenging
- Our view is computer architecture >> ISA
- Architect's job much more than instruction set design; technical hurdles today *more* challenging than those in instruction set design
- Since instruction set design not where action is, some conclude computer architecture (using old definition) is not where action is
  - We disagree on conclusion
  - Agree that ISA not where action is

Rest in Computer Architecture

Implementation of a computer = Organization + Hardware

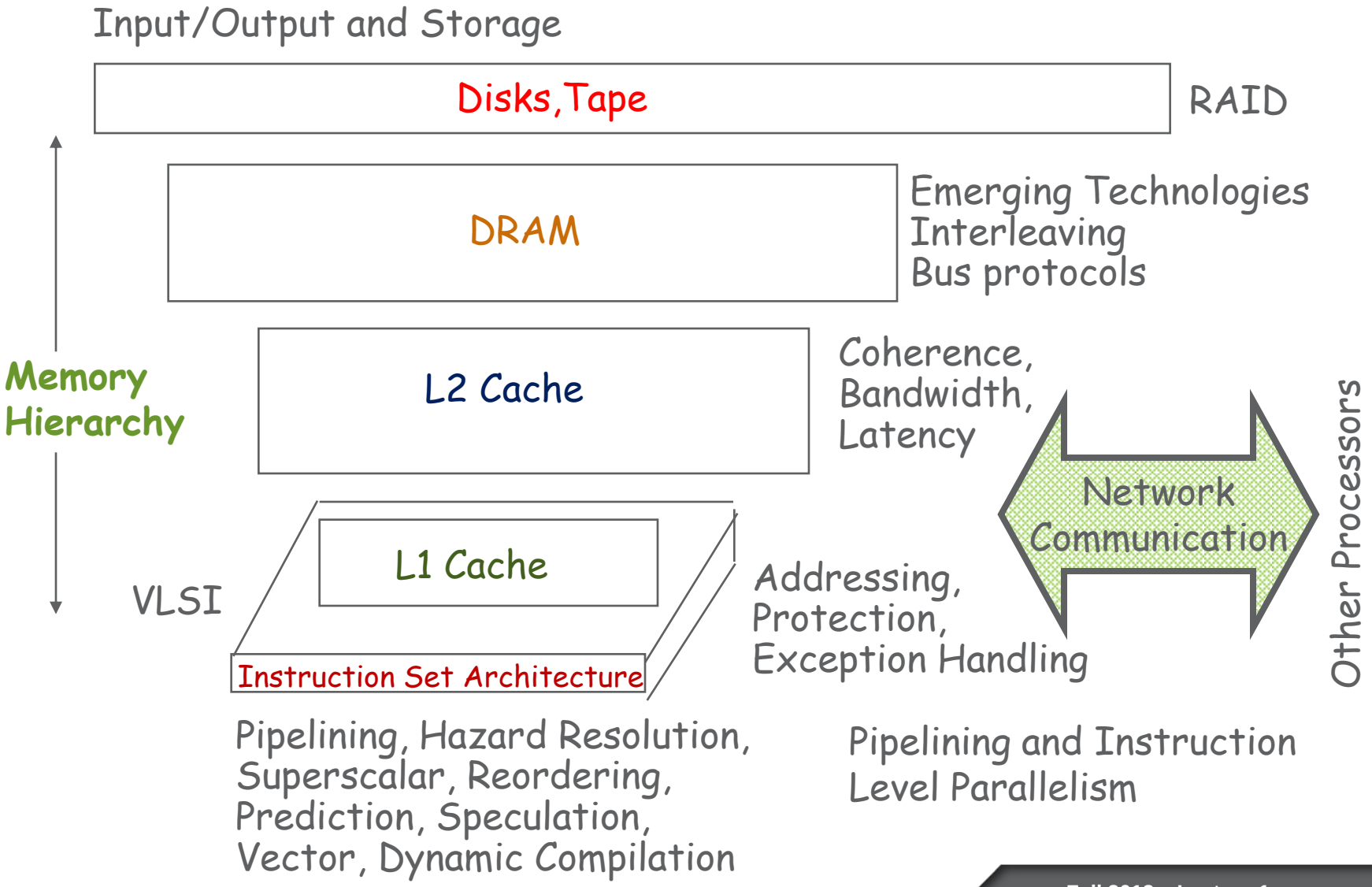
- Processor architecture
  - Pipelining, hazards, ILP, HW/SW interface
- Memory hierarchies
- Interconnects
- I/O systems
- Hardware technology used (e.g. component size)
- Computer architecture focuses on organization and quantitative principles of design

Execution is not just about HW and ISA



- The VAX fallacy
  - Produce one instruction for every high-level concept
  - Absurdity: Polynomial Multiply
    - Single hardware instruction
    - But Why? Is this really faster???
- RISC Philosophy
  - Full System Design
  - Hardware mechanisms viewed in *context* of complete system
  - Cross-boundary optimization
- Modern programmer does not see assembly language
  - Many do not even see “low-level” languages like “C”.

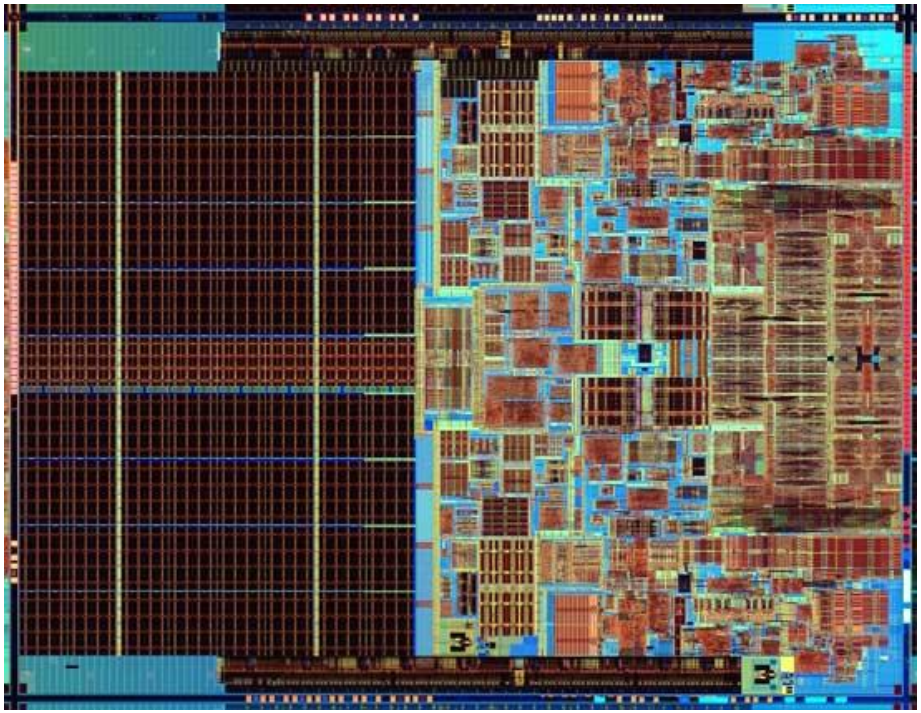
Computer Architecture Topics

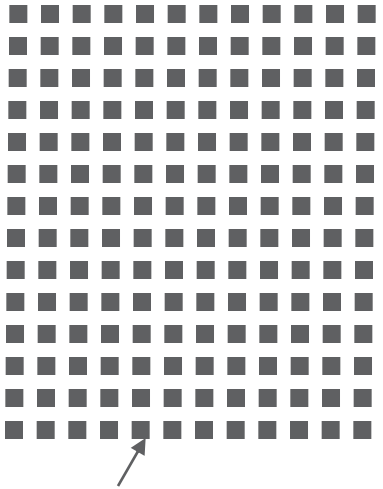


Executive Summary

The processor you built in HY225

What you'll understand after taking HY425

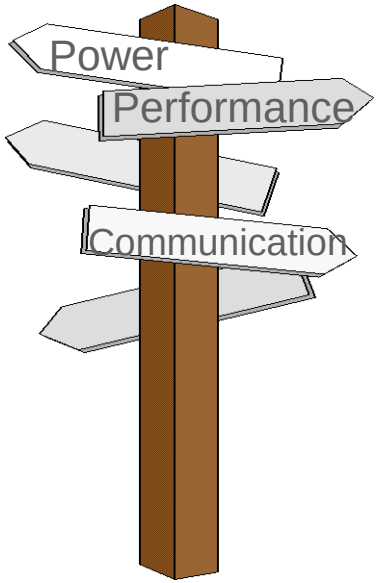




Also, the technology behind chip-scale multiprocessors

Next Lecture: Major Design Challenges

- Power
- CPU time
- Memory latency/bandwidth
- Storage latency/bandwidth
- Transactions per second
- Intercommunication
- Dependability



Everything Looks a Little Different