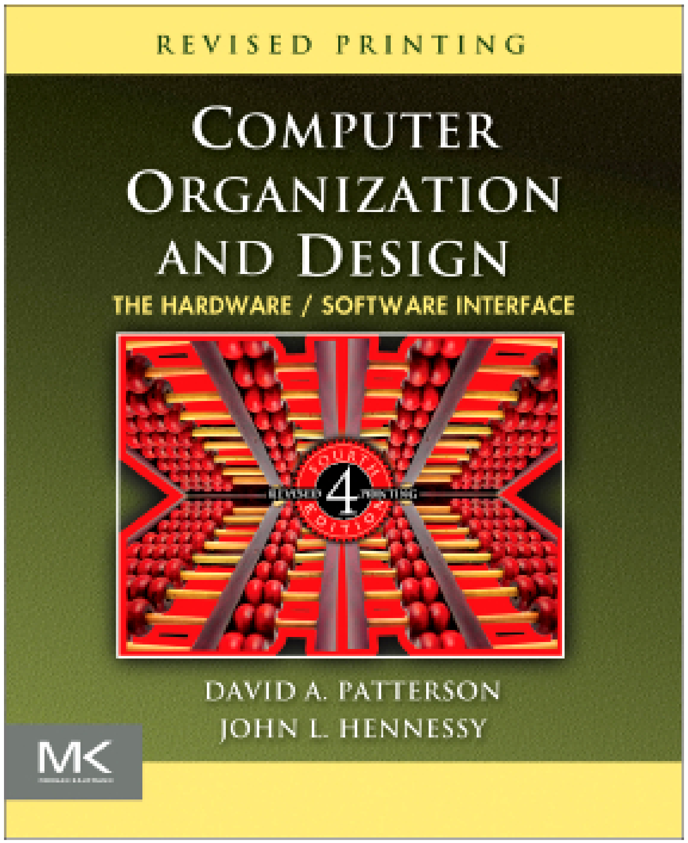
182.690 RECHNERSTRUKTUREN - INTRODUCTION

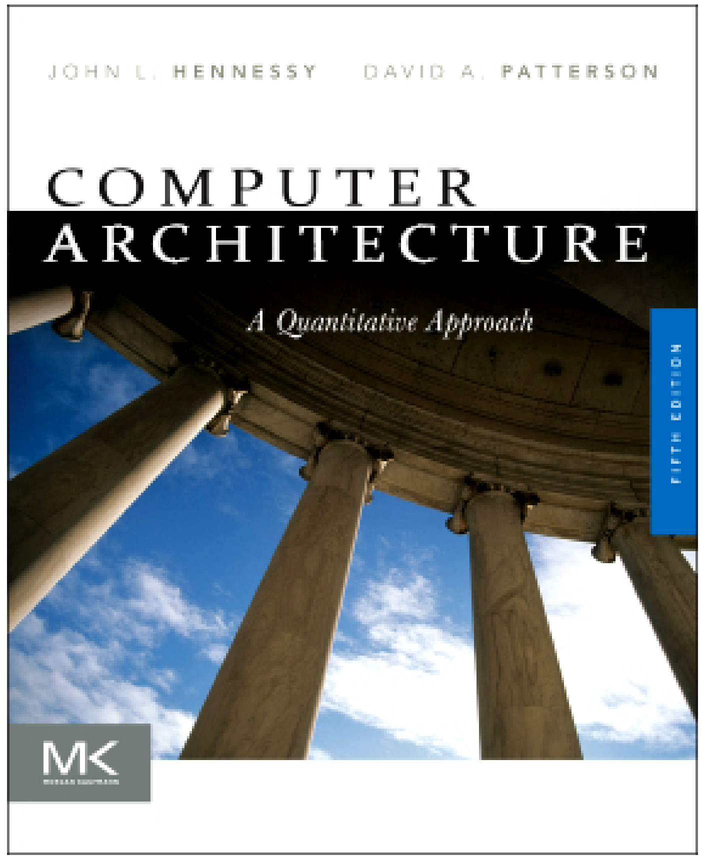
Thomas Polzer
tpolzer@ecs.tuwien.ac.at
Institut für Technische Informatik

Textbook



- Computer Organization and Design
  The Hardware / Software Interface
- David A. Patterson and John L. Hennessy
- Course based on the 4th, revised edition (ISBN: 978-0-12-374750-1)

Further Reading …



Lecture Organization

- Time and Location:
  - Blocked lecture
  - Thursday: 12:00 (c.t.), lecture hall: EI4
  - Friday: 11:30 (s.t.), lecture hall: EI10
  - Starts today

- Registration in TISS mandatory

Curriculum

- Recommended in the 3rd term
- STEOP mandatory!

- Preceding lectures:
  - VU Grundlagen Digitaler Systeme (1st term)
- Follow up lectures:
  - VO Digital Design (4th term)
  - VO Hardware Modeling (5th term)
  - LU Digital Design and Computer Architecture (5th term)

What You Will Learn

- How programs are translated into the machine language
  - And how the hardware executes them
- The hardware/software interface
- What determines program performance
  - And how it can be improved
- How hardware designers improve performance
  - Pipelining
  - Caching
  - Superscalar
- What is parallel processing

Exam

- Written exam (on paper)
  - Theoretical questions
  - Calculations
  - Practical tasks

- Registration will be in TISS (after Christmas break)

- First exam: End of January
- Afterwards: Three exams each term

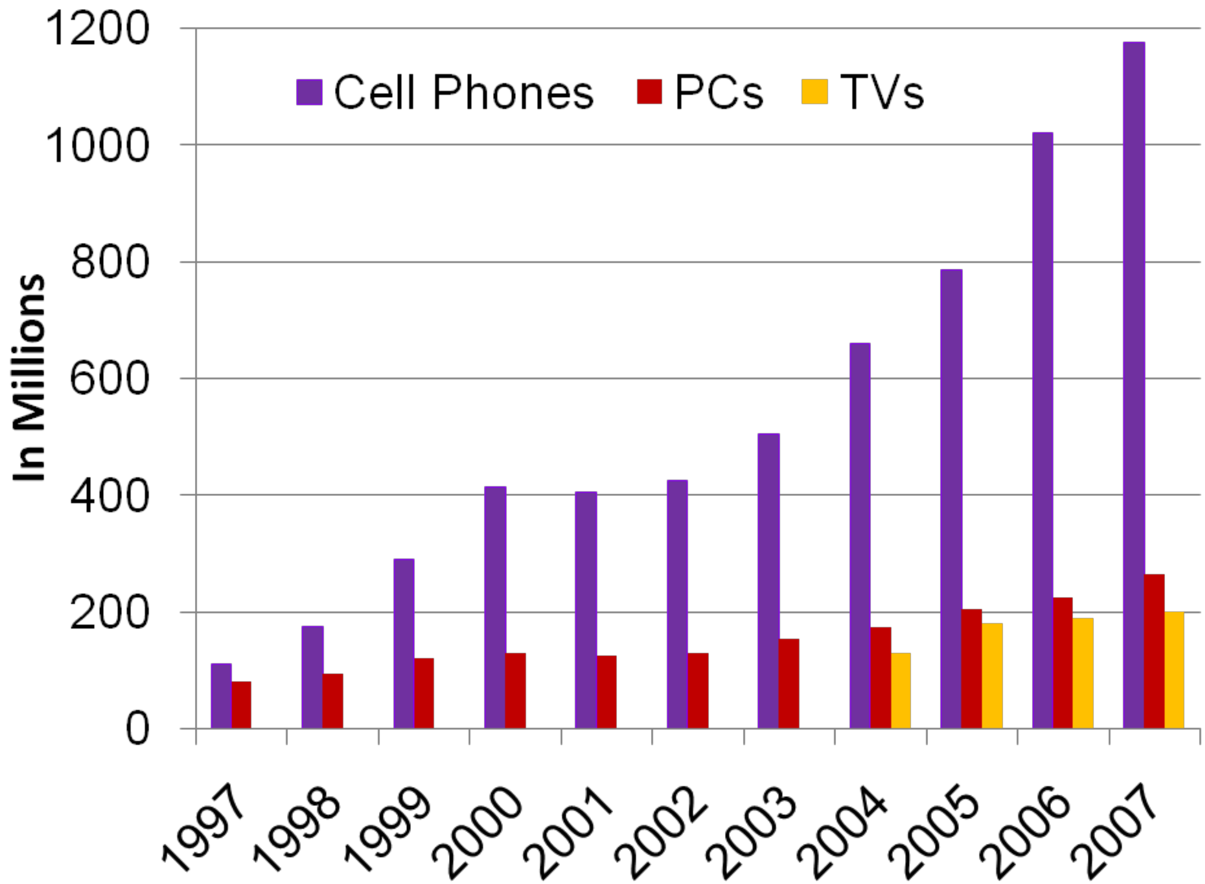
Why Computer Architecture?

- Progress in computer technology
  - Underpinned by Moore's Law
- Makes novel applications feasible
  - Computers in automobiles
  - Cell phones
  - Human genome project
  - World Wide Web
  - Search Engines
- Computers are pervasive

Classes of Computers

- Desktop computers
  - General purpose, variety of software
  - Subject to cost/performance tradeoff
- Server computers, Supercomputers
  - Network based
  - High capacity, performance, reliability
  - Range from small servers to building sized
- Embedded computers (processors)
  - Hidden as components of systems
  - Stringent power/performance/cost constraints

The Processor Market (Manufactured / Year)



Embedded Processor Characteristics

- Largest class of computers
  - Cars, Planes, Trains
  - Cellphones
  - Internet of things (Smart sensors, …)

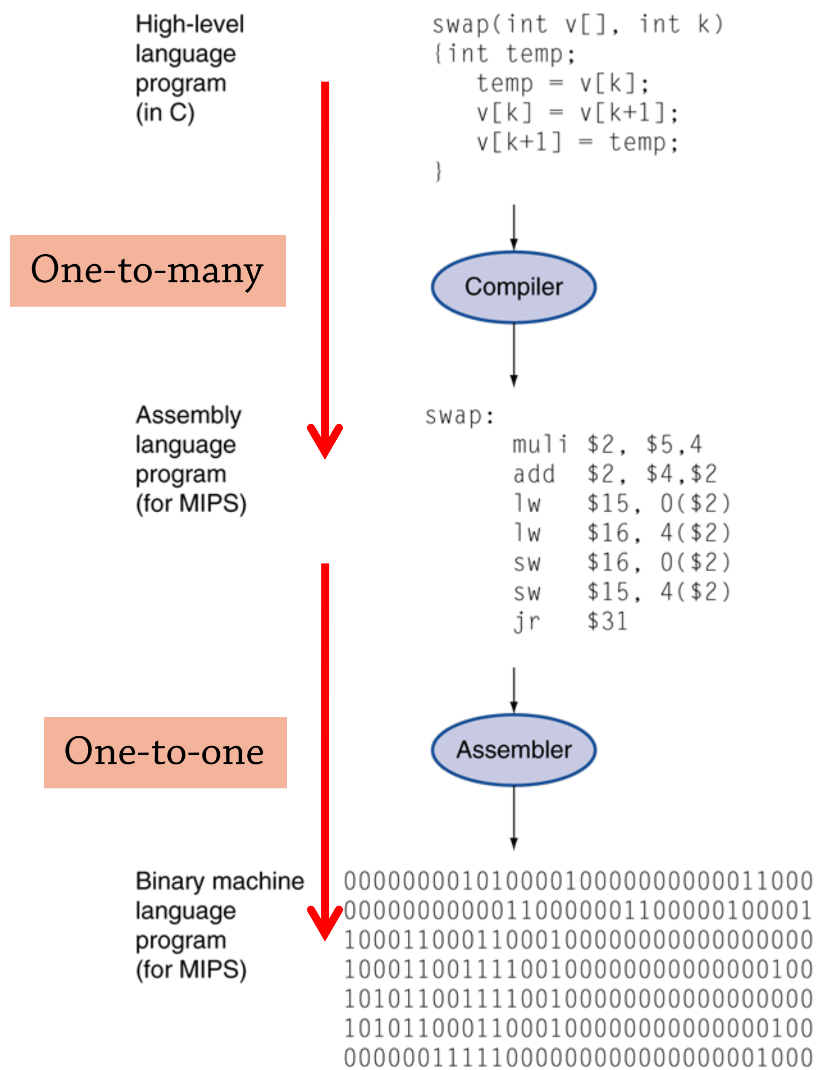
- Widest range of applications and performance
  - Often minimum performance requirements
  - Often stringent limitations on cost
  - Often stringent limitations on power consumption
  - Often high fault tolerance requirements

Performance of a Computer Program

- Algorithm
  - Determines number of operations executed
- Programming language, compiler, architecture
  - Determine number of machine instructions executed per operation
- Processor and memory system
  - Determine how fast instructions are executed
- I/O system (including OS)
  - Determines how fast I/O operations are executed

Levels of Program Code

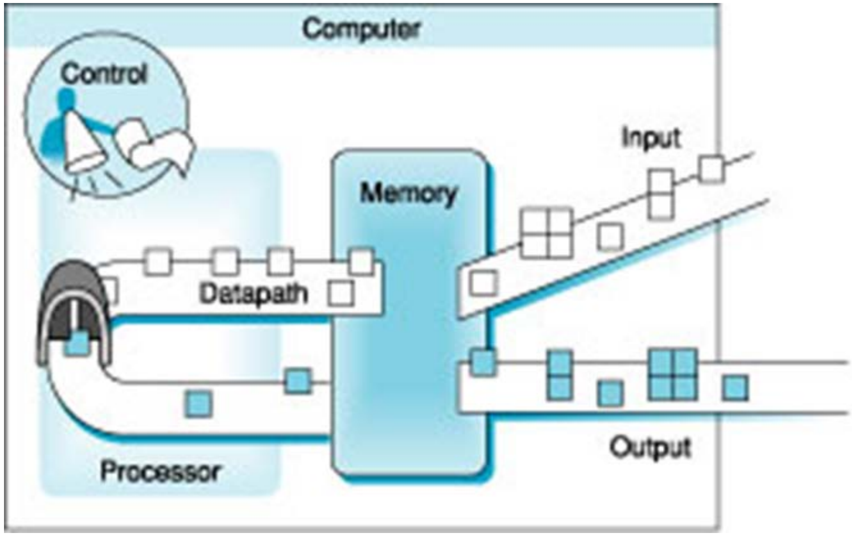
- High-level language
  - Closer to problem domain
  - Uses variables, loops, classes, …
- Assembly language
  - Textual (intermediate-) representation of instructions
- Hardware representation
  - Binary digits (bits)
  - Encoded instructions and data



Focus of the Lecture

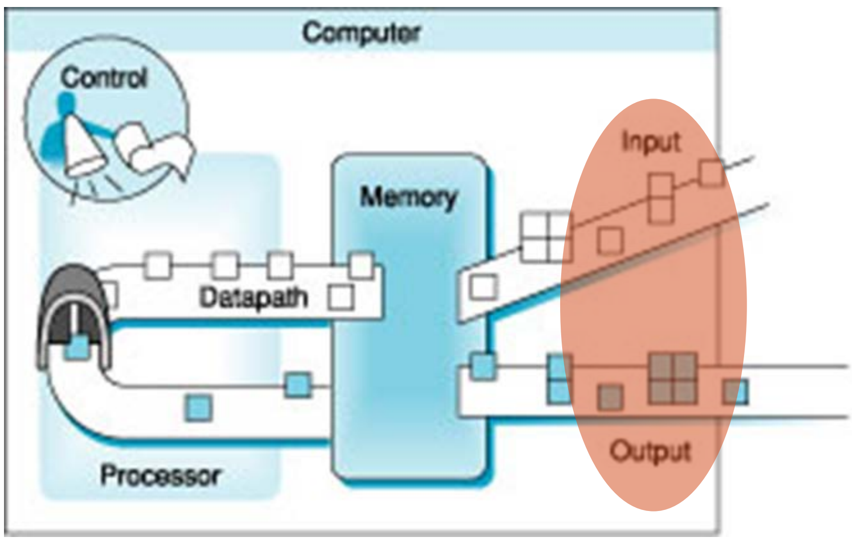
User
Application
SW — Operating system
Instruction set → Computer Architecture
HW — Organization → Computer Architecture
Logical Design
Electronics
Physics

Components of a Computer



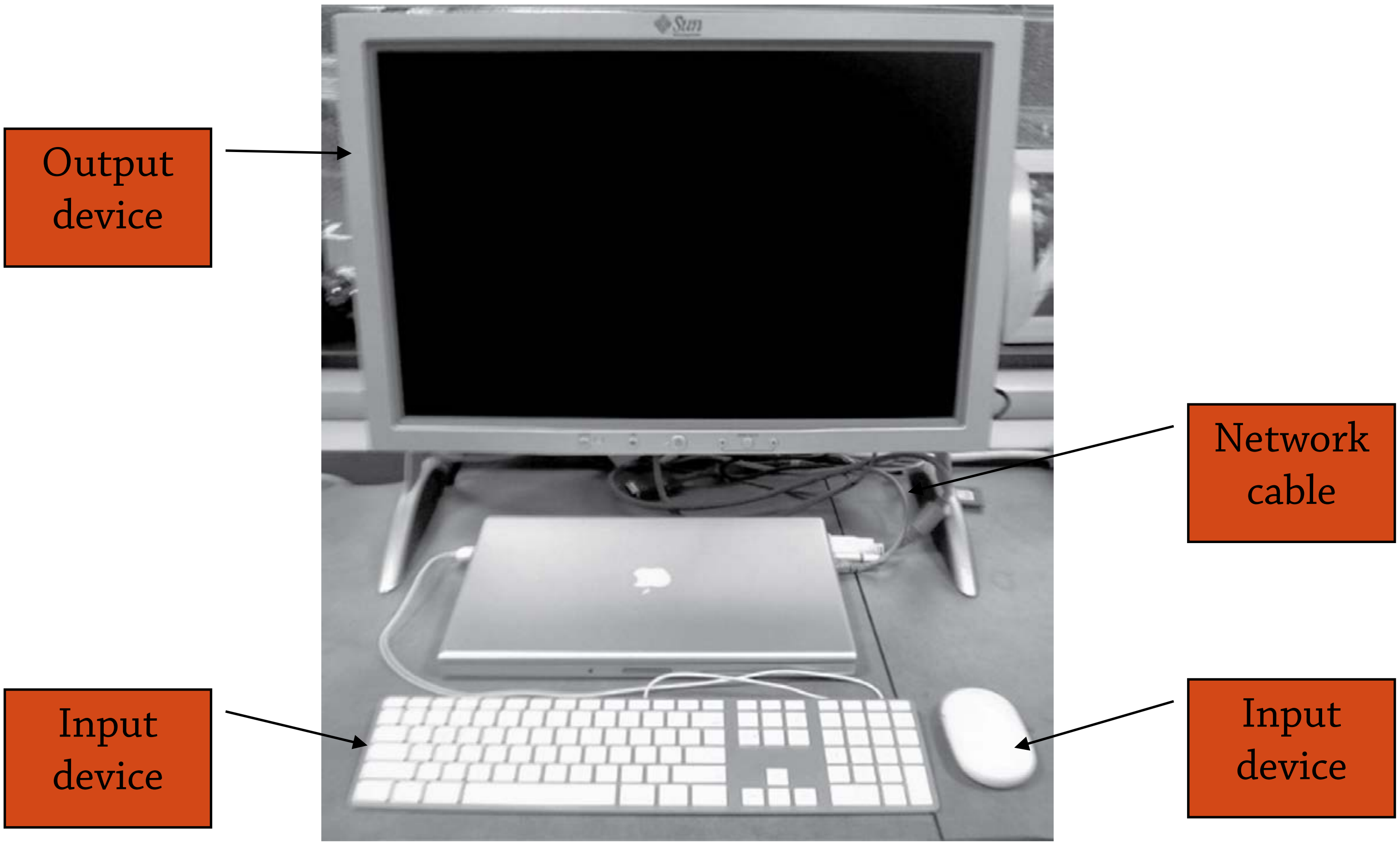
- Same components for all kinds of computers
  - Input
  - Output
  - Control
  - Datapath
  - Memory

Input / Output



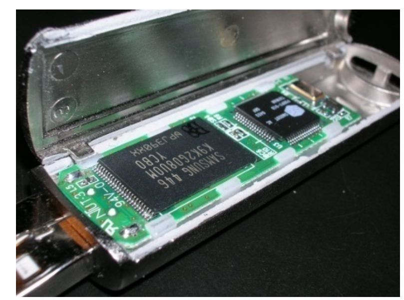
- User-interface devices
  - Display
  - Keyboard
  - Mouse
- Storage devices
  - Hard disk
  - CD/DVD
  - Flash
- Network adapters
  - Communicating with other computers

Input / Output (HID)



Input / Output (Storage)

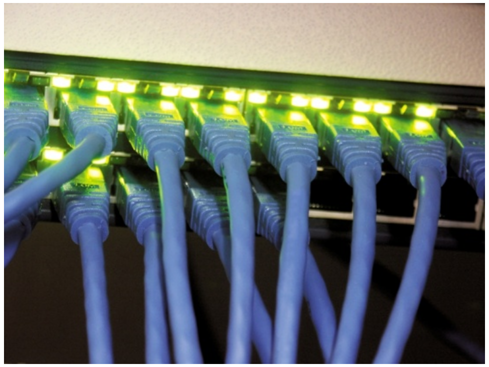
- Volatile main memory
  - Loses instructions and data when power off
- Non-volatile secondary memory
  - Magnetic disk (↑ 1 TB)
  - Flash memory (↑ 256 GB)
  - Optical disk (CDROM, DVD, Blu-ray)



Input / Output (Networks)

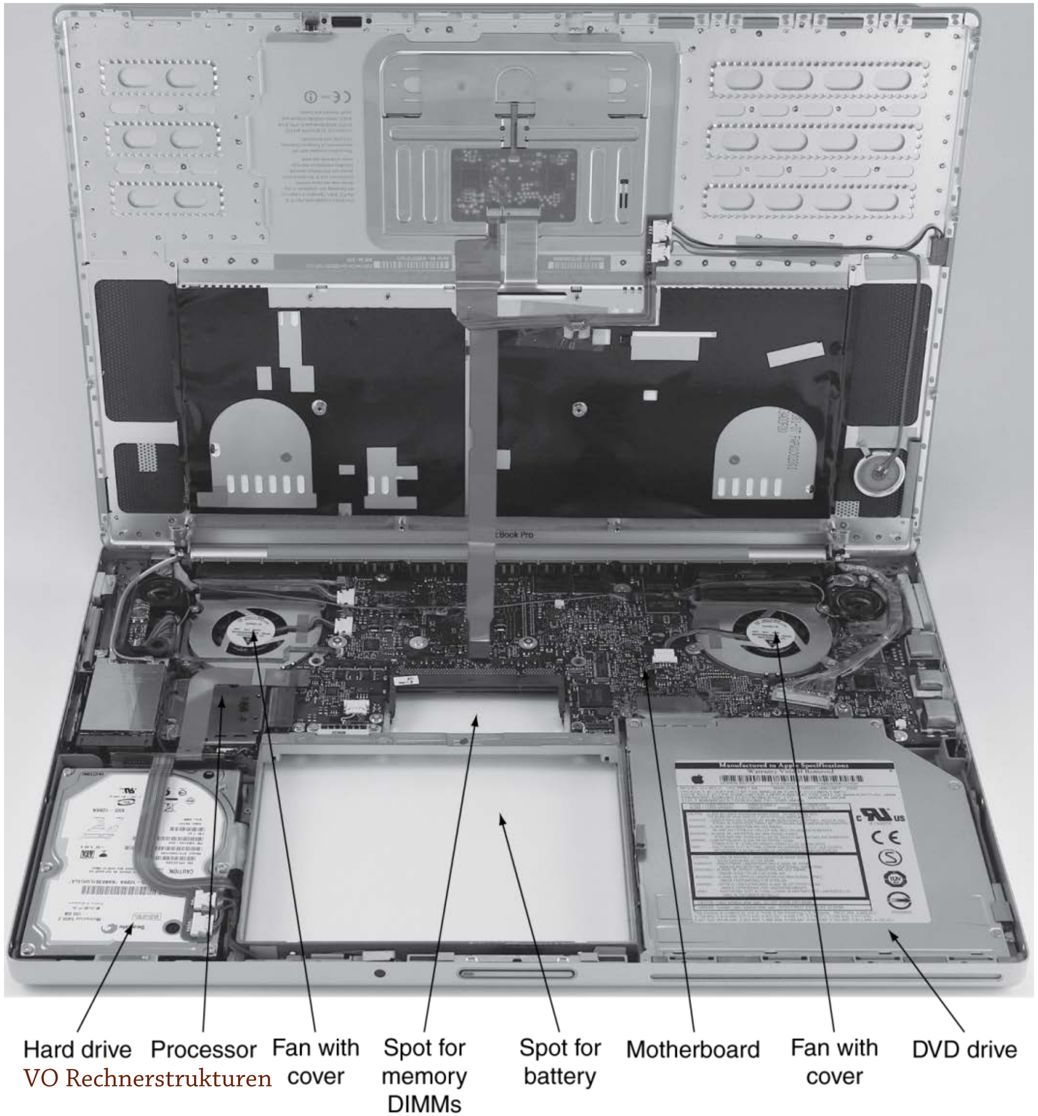
- Communication and resource sharing

- Local area network (LAN): Ethernet
  - Within a building
- Wide area network (WAN): the Internet



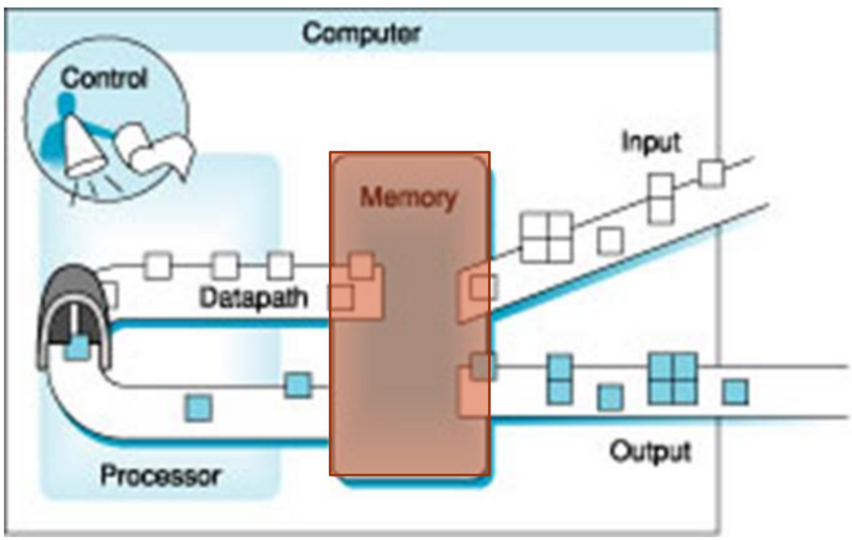
- Wireless network: Wi-Fi, Bluetooth

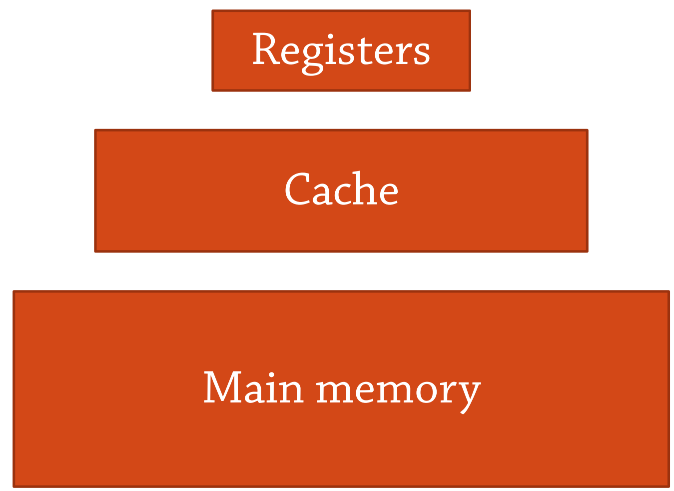
Anatomy of a Computer …



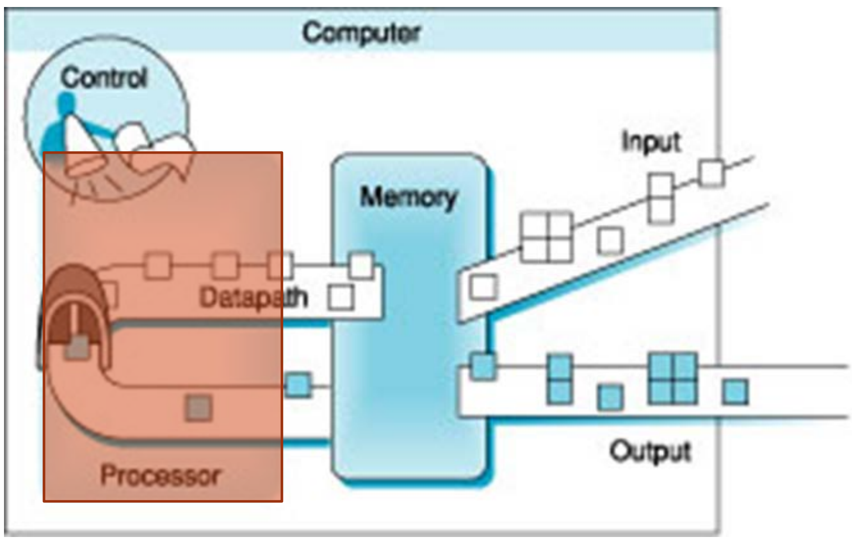
Memory Hierarchy

- Too few internal registers
- Main memory too slow
- Caching to increase access time for often used data





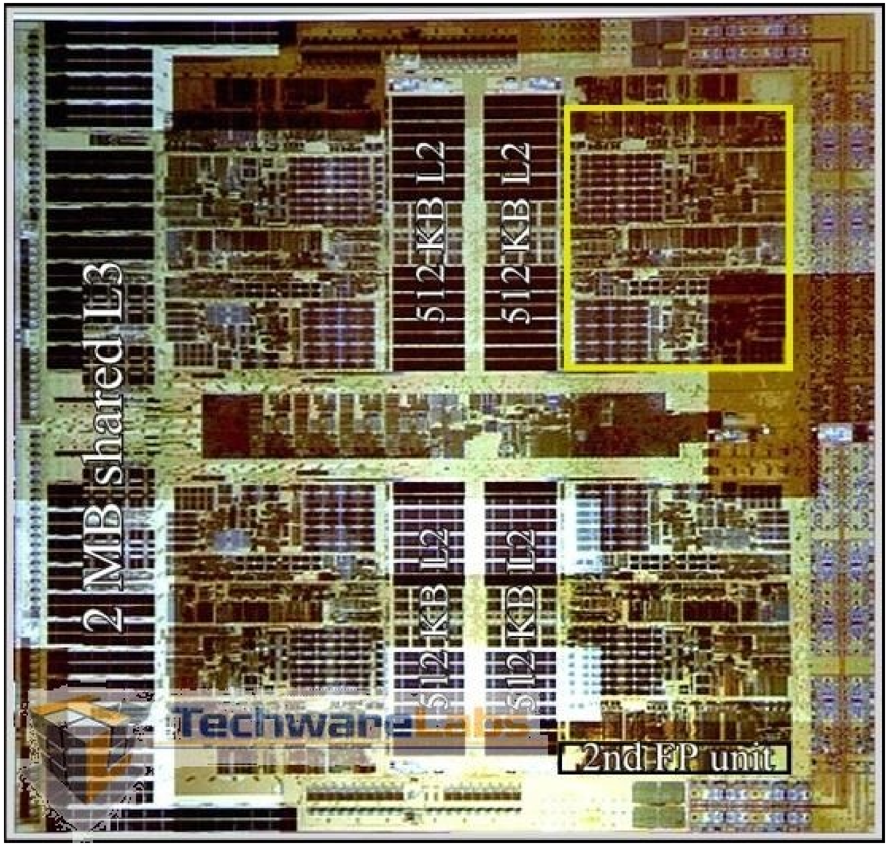
Components of a Computer



- Same components for all kinds of computers
  - Input
  - Output
  - Control
  - Datapath
  - Memory

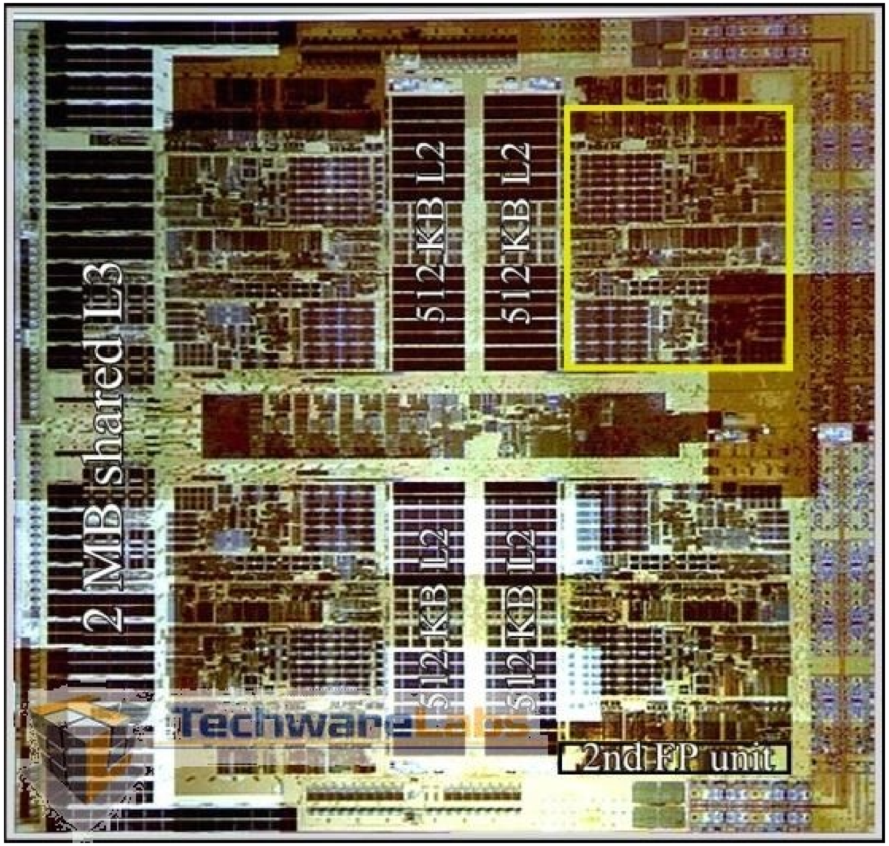
Processor (CPU)

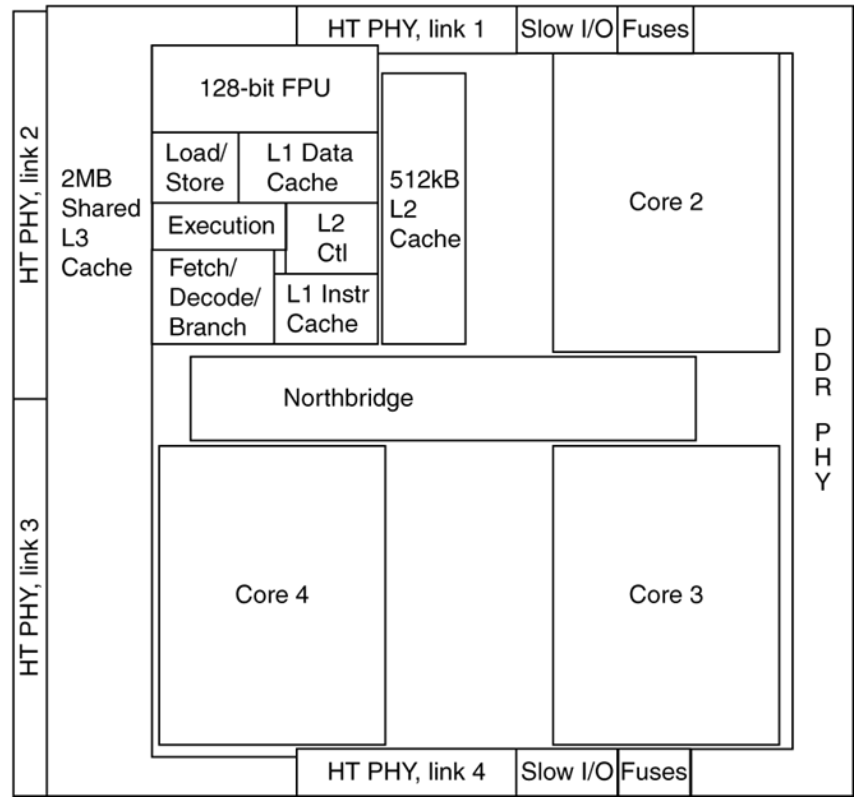
... and of a CPU



- Datapath: performs operations on data
- Control: sequences datapath, memory, ...
- Cache memory
  - Small fast SRAM memory for immediate access to data

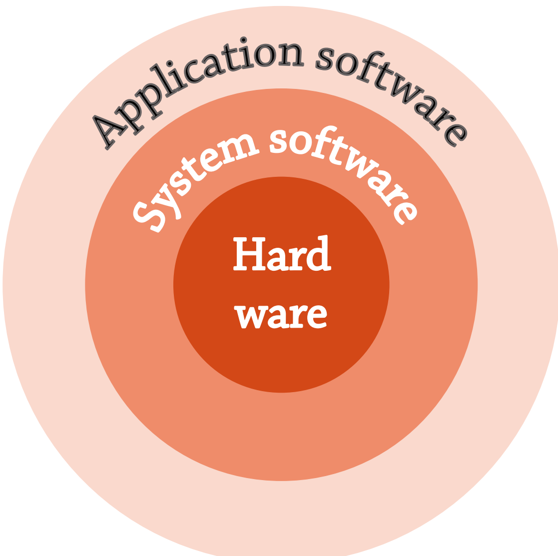
… and of a CPU





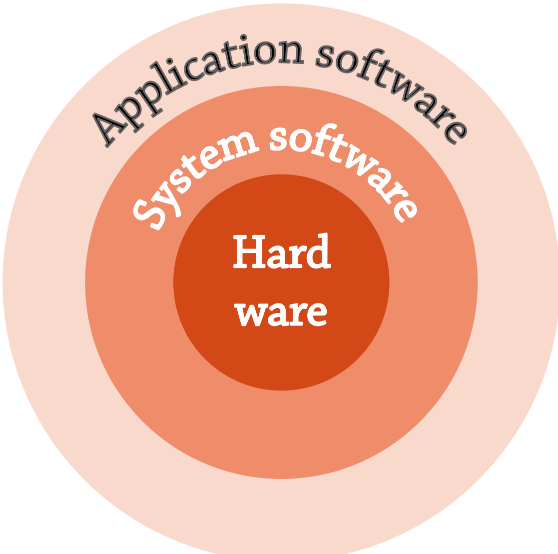
AMD Barcelona: 4 processor cores (Intel Nehalem μ-Architecture)
1.9 GHz, 65 nm technology, L1, L2, L3, integrated Northbridge, 4 OoO cores

Abstraction



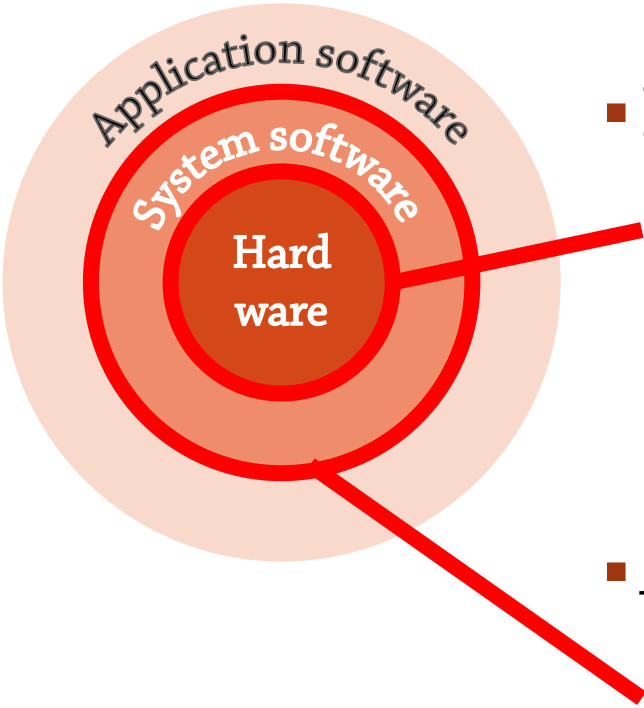
- Hiding of unnecessary lower level details
- Enables us to cope with complex systems
- *How productive would a software developer be, if he would need to calculate the required electric field to control the diffusion process in a MOS-FET???*

Abstraction



- Application software
  - Written in high-level language
- System software
  - **Compiler**: translates high level code to machine code
  - **Operating system**: service code
    - Handling input/output
    - Managing memory and storage
    - Scheduling tasks & sharing resources
- Hardware
  - Processor, memory, I/O controllers

Interfaces



- Abstraction requires defined interfaces between layers
- Instruction Set Architecture (ISA)
  - Interface between hardware and low-level system software
  - Consists of instructions, registers, memory access, I/O, …
- Application Binary Interface (ABI)
  - Combination of (user accessible) ISA and the OS interface
  - Standard interface for portable programs

Innovation

Year	Technology	Relative performance / cost
1951	Vacuum tube	1
1965	Transistor	35
1975	Integrated circuit (IC)	900
1995	Very large scale IC (VLSI)	$2,4x10^6$
2005	Ultra large scale IC	$6.2x10^9$

- Governed by Moore's law
  - 1965 by Intel's G. Moore
  - Number of transistors double every two years
- DRAM grows
  - Quadruples every three years
- Continued increase in capacity and performance
- Decreased cost!

Technology Scaling Road Map (ITRS)

Year	2004	2006	2008	2010	2012
Feature size (nm)	90	65	45	32	22
Integration Capacity (10^9 Transistors)	2	4	6	16	32

- 45 nm technology
  - 30 million devices fit on the head of a pin
  - > 2,000 across the width of a human hair

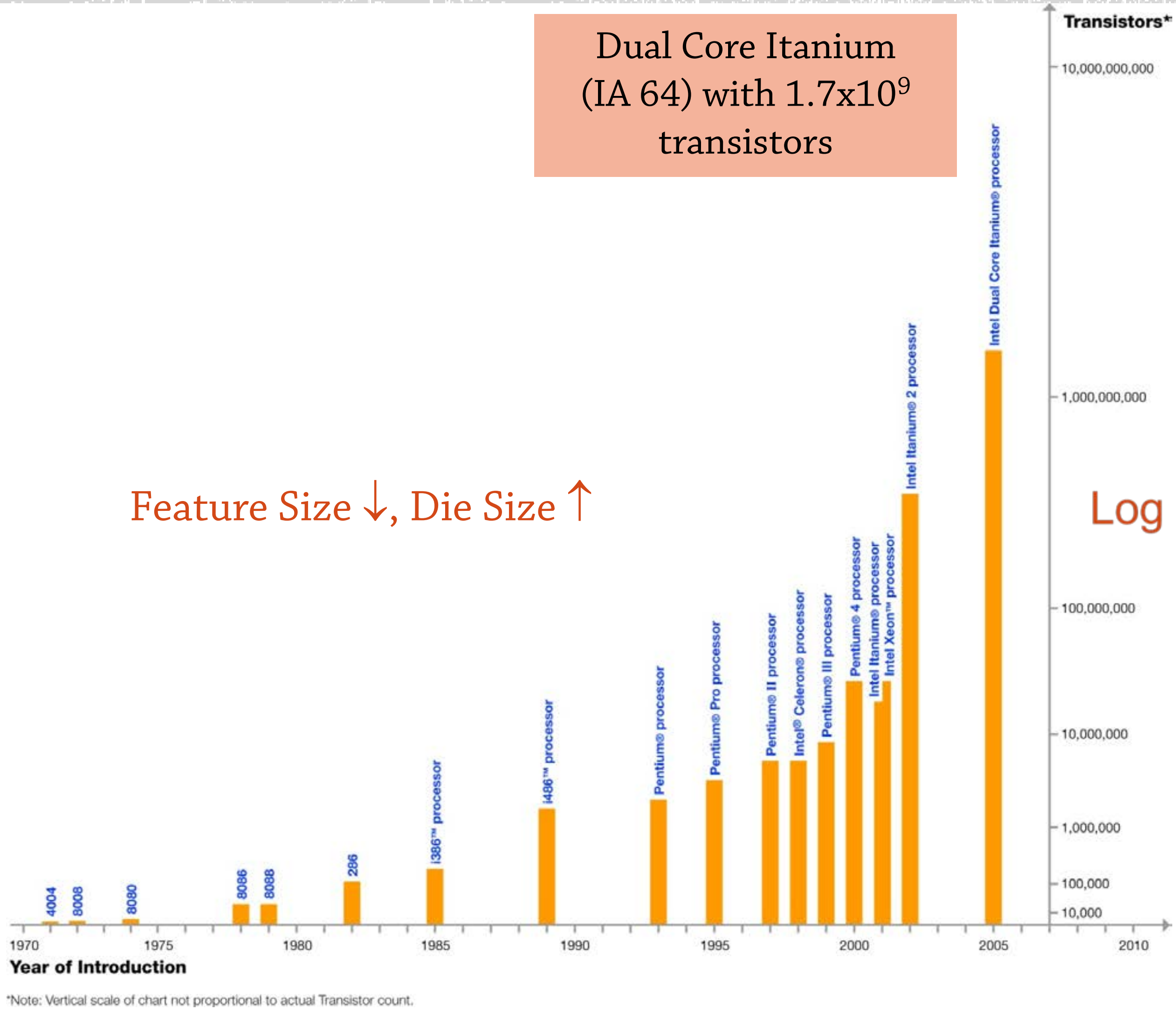
  - If car prices had fallen at the same rate as the price of a single transistor has since 1968, a new car today would cost about 1 cent.

Moore's Law

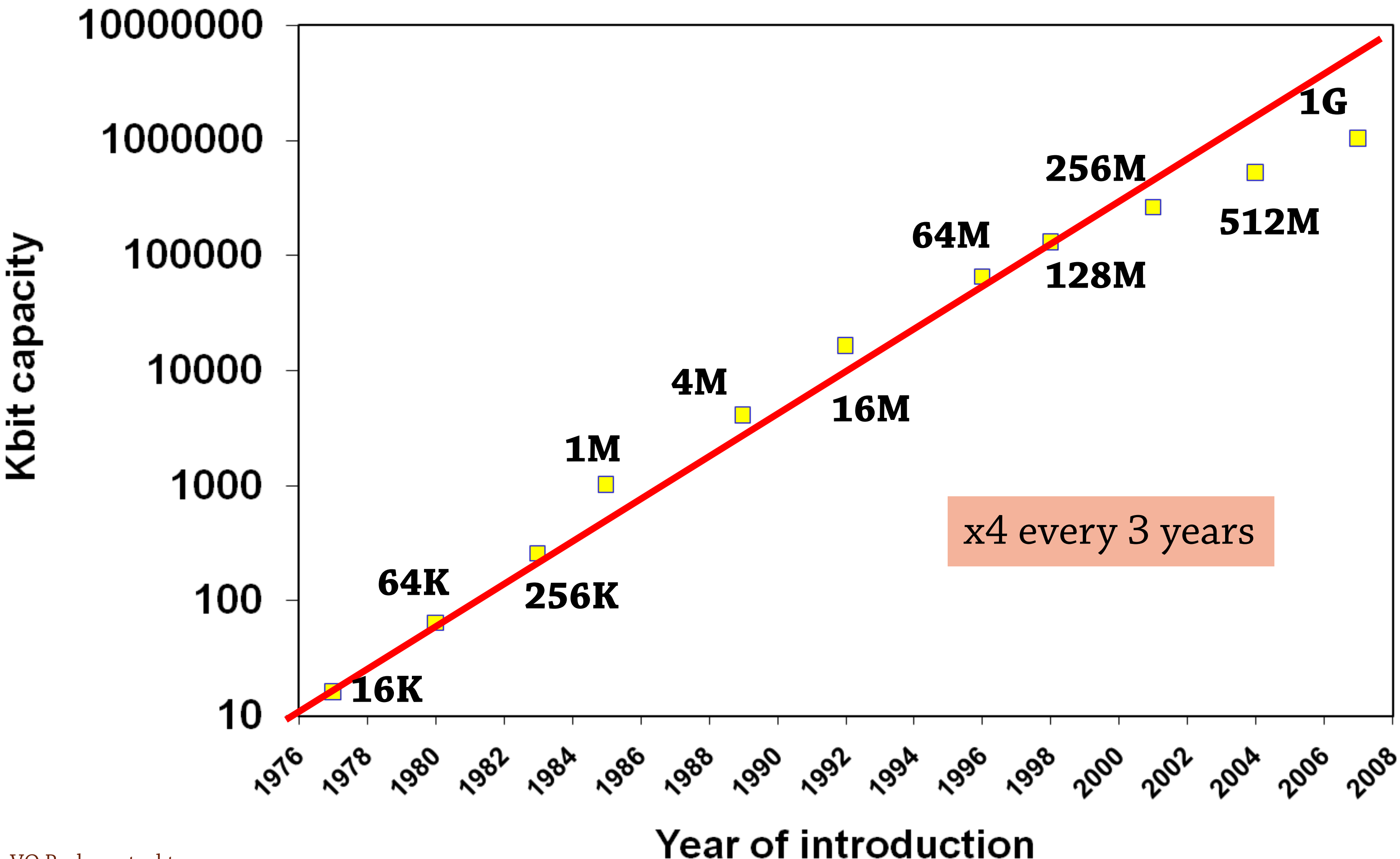
Dual Core Itanium (IA 64) with $1.7x10^9$ transistors

Feature Size ↓, Die Size ↑

Log Scale !

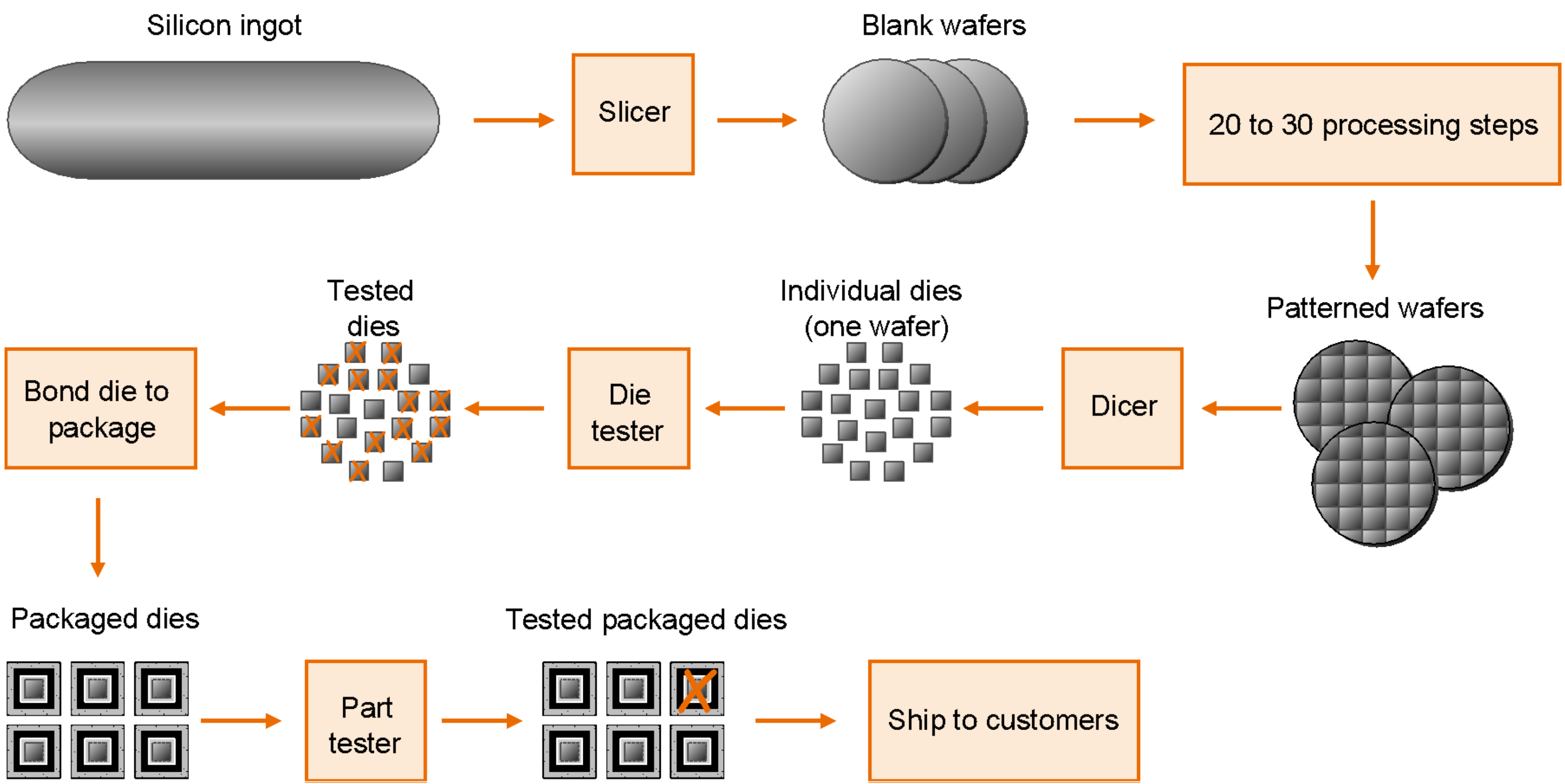


DRAM Capacity Growth

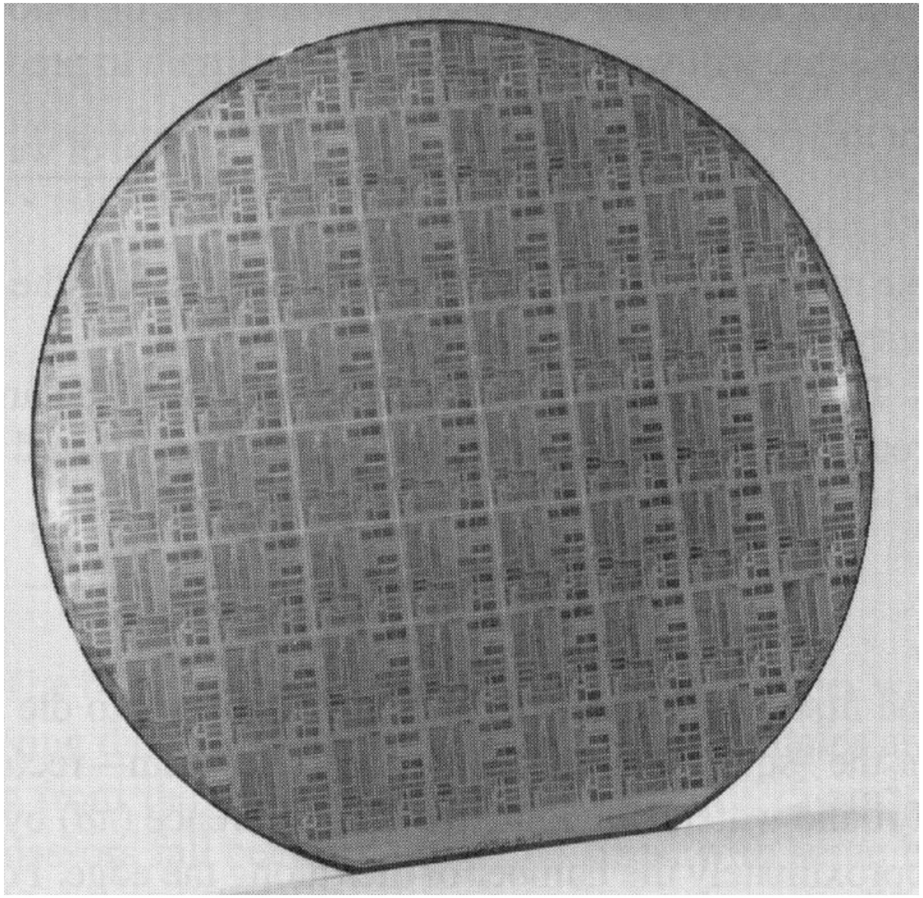


The Base Material ....

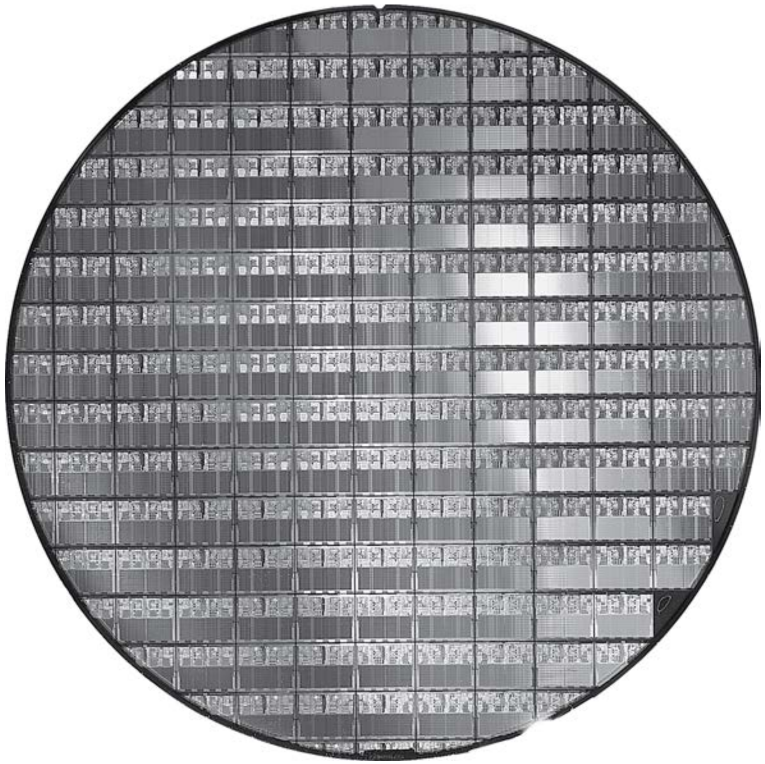
Manufacturing Process



Wafer Examples



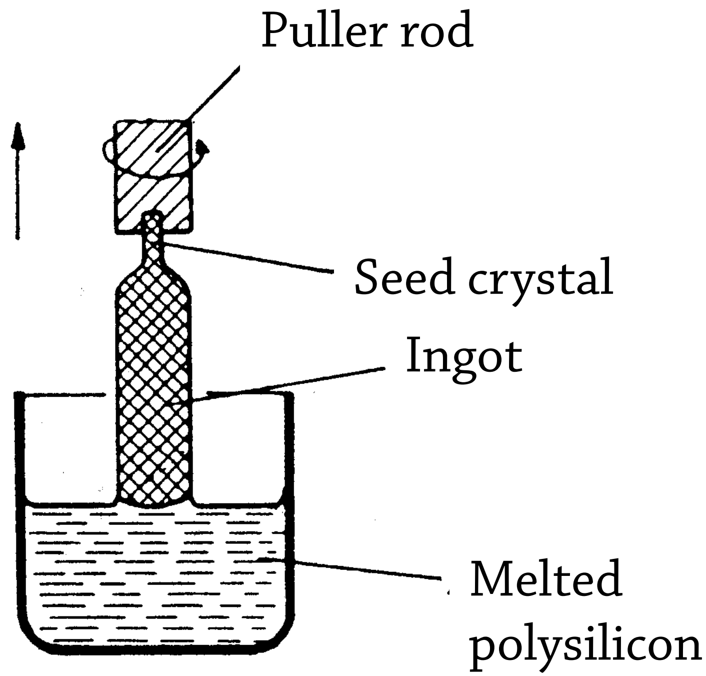
AMD Opteron X2 Wafer



- X2: 300 mm wafer, 117 chips, 90 nm technology
- X4: 45 nm technology

Why are Wafers Circular?

"Czochralski Process"



Integrated Circuit Cost

$$\text{Cost per die} = \frac{\text{Cost per wafer}}{\text{Dies per wafer} \times \text{Yield}}$$

$$\text{Dies per wafer} \approx \text{Wafer area} / \text{Die area}$$

$$\text{Yield} = \frac{1}{(1 + \text{Defects per area} \cdot \text{Die area/2})^2}$$

- Nonlinear relation to **area and defect rate**
  - Wafer cost and area are fixed
  - Defect rate determined by manufacturing process
  - Die area determined by architecture and circuit design

Costs / Die Example

- Typical values of an 8” wafer (20 cm)
  - Costs around $2500
  - Dies / wafer:
    - 269 1cm^2 Dies (Power PC class)
    - 79 3cm^2 Dies (Pentium class)
- If there is 1 defect/cm^2 which costs can we expect?
  - 44% yield for Power PC=> 118 Chips, $21
  - 16% yield for Pentium => 12 Chips, $198
- Die costs increases with the **third** power of its area!

Summary

- Different types of computers
  - Desktops, servers, supercomputers, embedded systems
- Main components of a computer
  - I/O, memory, processor
- Hierarchical layers of abstraction
  - In both hardware and software
- Instruction set architecture
  - The hardware/software interface
- Manufacturing process and costs/die

182.690 RECHNERSTRUKTUREN - INTRODUCTION

Thomas Polzer
tpolzer@ecs.tuwien.ac.at
Institut für Technische Informatik