How to Write Fast Numerical Code

Spring 2011
Lecture 3

Instructor: Markus Püschel

TA: Georg Ofenbeck

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Organizational

- **Class Monday 14.3. → Friday 18.3**

- **Mailing list**
  - Everybody subscribed
  - Emails sent will go to everybody
  - Use to find project partners

- **Email me if you know your partner**

Cost Analysis

- **Exact solution of recurrences**
  - Last time: first order
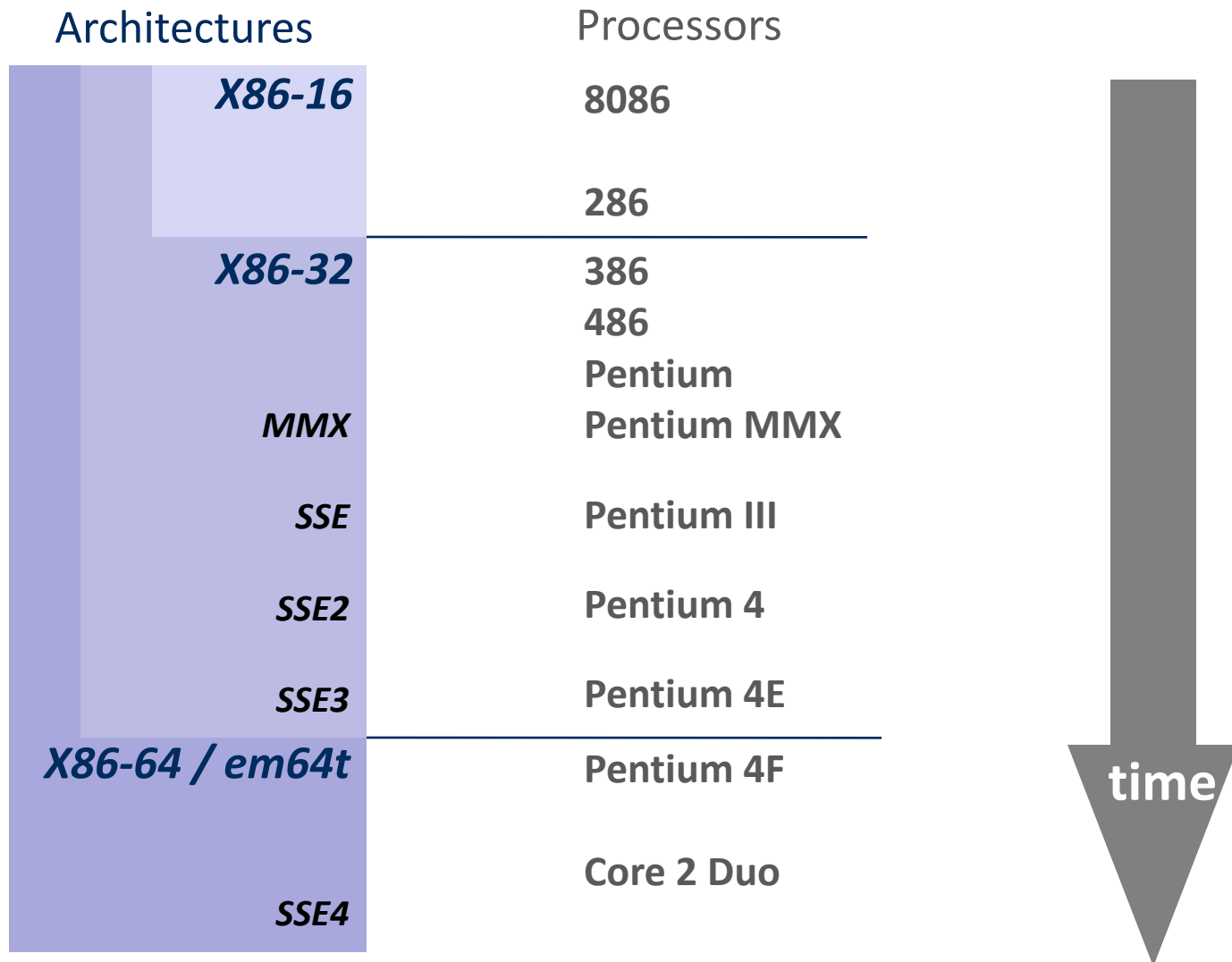 - Today: second (and higher) order (blackboard)

Today

- **Architecture**

- **Microarchitecture (numerical software point of view)**

- **First thoughts on fast code**

Definitions

- *Architecture:* (also instruction set architecture = ISA) The parts of a processor design that one needs to understand to write assembly code.

- *Examples:* instruction set specification, registers

- *Counterexamples:* cache sizes and core frequency

- Example ISAs
    - x86
    - ia
    - MIPS
    - POWER
   - SPARC

Intel Architectures (Focus Floating Point)

Architectures	Processors
X86-16	8086
	286
X86-32	386
	486
	Pentium
MMX	Pentium MMX
SSE	Pentium III
SSE2	Pentium 4
SSE3	Pentium 4E
X86-64 / em64t	Pentium 4F
	Core 2 Duo
SSE4	

time

ia: often redefined as latest Intel architecture

ISA SIMD (Single Instruction Multiple Data) Vector Extensions

- **What is it?**
    - Extension of the ISA. Data types and instructions for the parallel computation on short (length 2-8) vectors of integers or floats

4-way
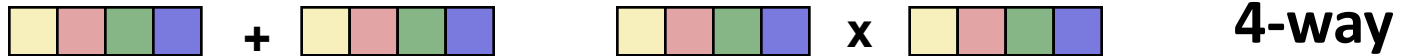
    - Names: MMX, SSE, SSE2, …

- **Why do they exist?**
    - **Useful:** Many applications have the necessary fine-grain parallelism Then: speedup by a factor close to vector length
    - **Doable:** Chip designers have enough transistors to play with

- **We will have an extra lecture on vector instructions**
    - What are the problems?
    - How to use them efficiently

Definitions

■ *Microarchitecture:* Implementation of the architecture.

■ Includes caches, cache structure, ...

■ Examples
   ■ Intel processors (Wikipedia)
   ■ Intel microarchitectures

Microarchitecture:
The View of the Computer Architect
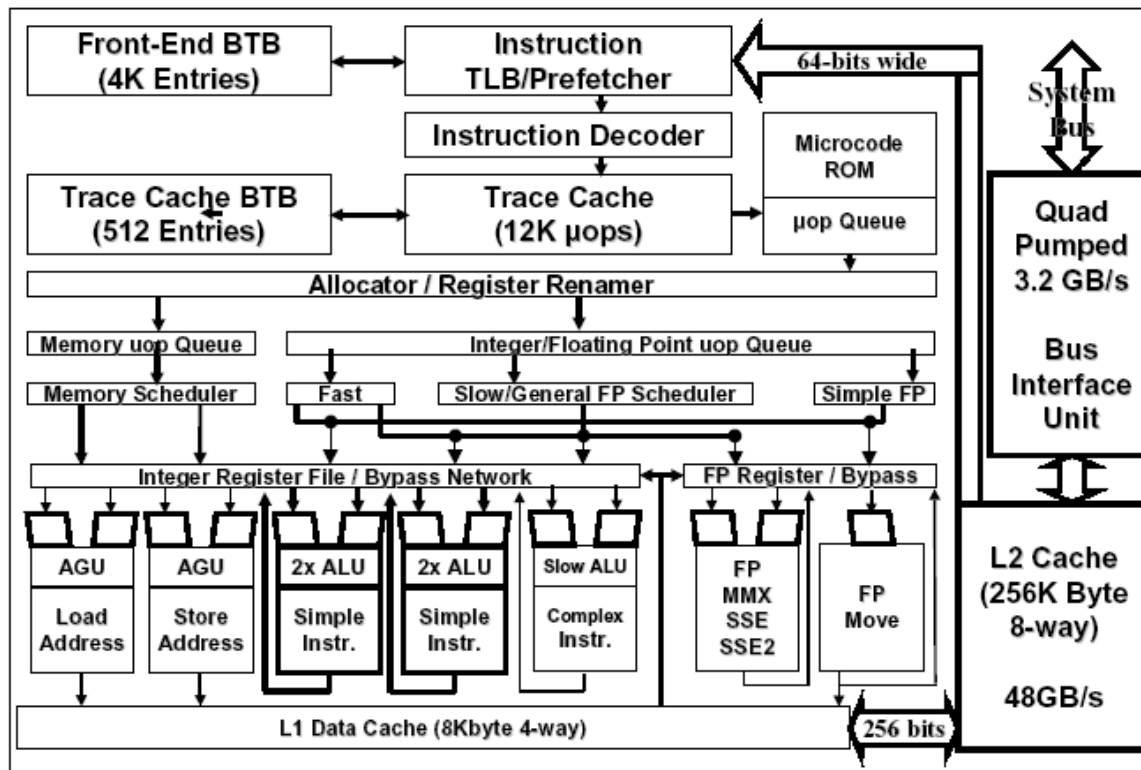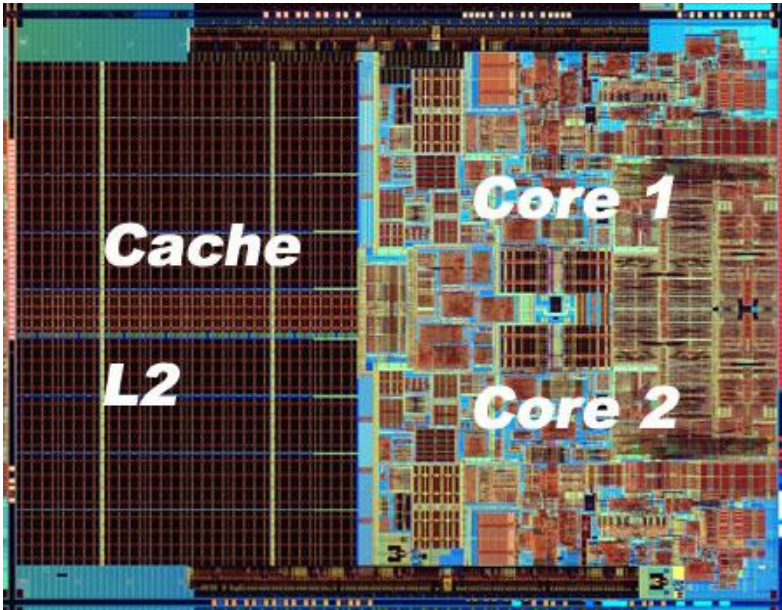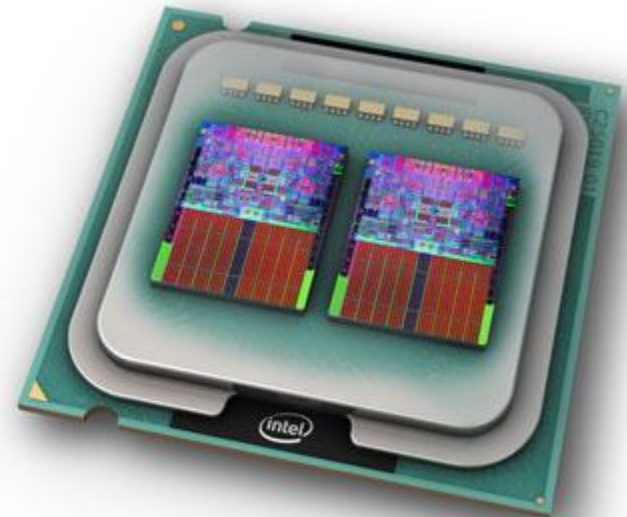

Figure 4: Pentium® 4 processor microarchitecture

we take the software developers view … (blackboard)

Core 2 Duo

2 x Core 2 Duo packaged

Detailed information about Core 2 Duo
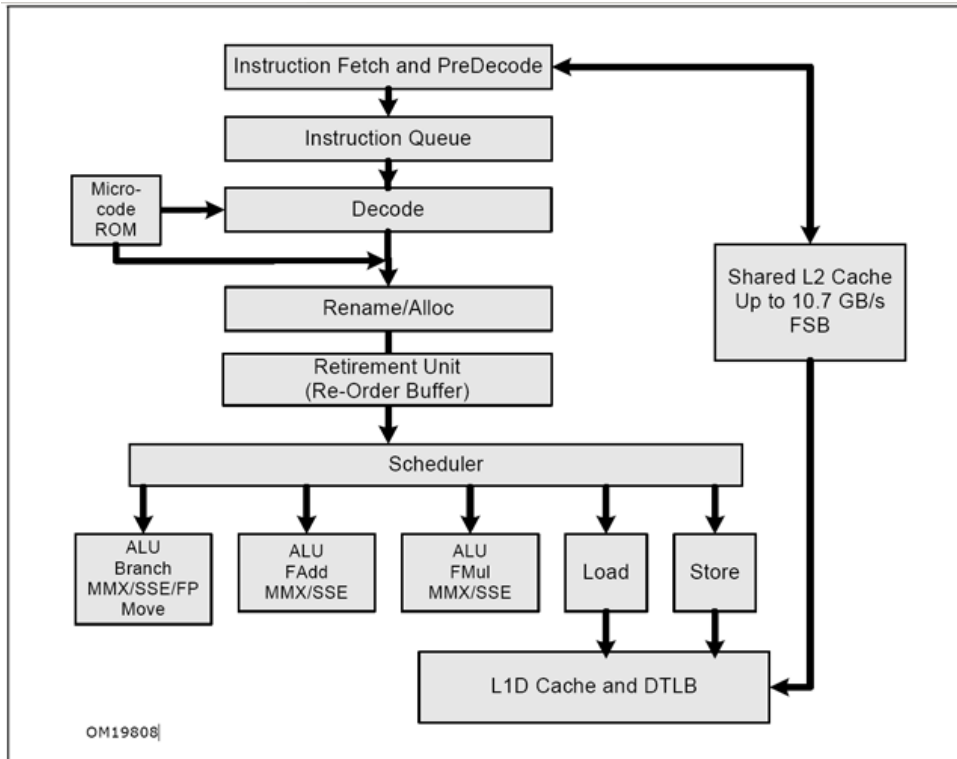
Floating Point Peak Performance

Figure 2-1. Intel Core Microarchitecture Pipeline Functionality

Latency/throughput (double)
FP Add: 3, 1
FP Mult: 5, 1

Theoretical peak performance (3 GHz, 1 core, no SIMD, double precision): *6 Gflop/s*
SIMD, 1 core, double precision: *12 Gflop/s*
SIMD, 1 core single precision: *24 Gflop/s*
2 or 4 cores: *multiply by 2 or 4*
Requires: computation has 50% adds and 50% mults

Performance: First Thought

- **It is all about keeping the floating point units busy**
  - Instruction level parallelism
  - Locality