ECE 486/586
Computer Architecture

Prof. Mark G. Faust

Maseeh College of Engineering
and Computer Science

PORTLAND STATE
UNIVERSITY

Outline

- Motivation: why study computer architecture
- Levels of abstraction
- Classes of computers (historical taxonomy)
- Impact of microprocessor
- Classes of computers (present)
  - Desktops & laptops, Servers, Embedded
  - Personal mobile devices, clusters/warehouse scale computers
- Parallelism
- Flynn's taxonomy
- Defining computer architecture: ISA vs. "Organization"

Why Study Computer Architecture

- Computer Architects
- Embedded System Designers
- Systems Designers
- Compiler Designers
- Operating System Designers
- Programmers

- Widely applicable key concepts
  - Caching
  - Pipelining
  - Parallelism
- Ability to analyze, quantify, and make intelligent trade-offs/design decisions
  - Quantitative performance measures
- Understand interactions between code and computer organization
- Impact on OS and compilers of architectural decisions
- Framework for maintaining currency in a rapidly evolving field
- It's interesting and dynamic
  - Technology, business, marketing, religion

How Its Missteps Let Rival AMD Catch Up

By Don Clark

FEW COMPANIES have closer ties to **Intel** Corp. than **Hewlett-Packard** Co. After all, the two Silicon Valley giants collaborated for a decade on chip technology that H-P plans to use in three major computer lines.

Yet H-P also has become a vocal fan of Opteron, a competing microprocessor from **Advanced Micro Devices** Inc. H-P executives concluded in the summer of 2003—after developing a prototype computer under a secret effort called Project X—that AMD's chip was the best for some server systems, and Intel didn't have a plausible strategy to close the gap.

"We kept asking Intel, 'What are you going to do for us to be competitive?' " says Paul Miller, H-P's vice president of marketing for industry-standard servers. "The plan kept on falling short."

Many plans have fallen short lately for In-

Big Chip Dip

Source: Thomson Datastream

How did AMD, an erratic performer through much of its 35-year history, make such big waves? Part of the answer can be

the chip technology Int... stock plunged yesterday... attack in chips called fl...

Transmeta leaving CPU business?

1/5/2005 11:08:17 AM, by Eric Bangeman

CPU manufacturer Transmeta, known for their low-power processors, is evaluating an exit from the CPU market. Instead of manufacturing chips themselves, their business focus would shift towards buzzwords: licensing their intellectual property and the formation of strategic alliances to utilize their processor design as well as their research and development skills.

Transmeta burst on the scene in 2000 with the Crusoe processor. Designed with the primary goals of full x86 compatibility and the lowest possible power consumption, the Crusoe featured Transmeta's Code Morphing software, which translated x86 instructions into an internal instruction format. When compared to similarly-clocked chips from Intel or AMD, the performance of Transmeta's CPUs suffered. On the other hand, there was no comparison when it came to power consumption. A 700MHz Crusoe fabbed at the 180nm process consumed *one measly watt* of power compared to a 7...

Sony, IBM, Toshiba To Offer First Peek Of 'Cell' Chip Design

By Don Clark
And ...

A N AMBITI... puter and ... is taking a... At a technical ... cisco, researchers ... and **International** ... to release details ... chip, dubbed Cell, ... digital media and computer graphics—and to s... **Intel** Corp. and **Microsoft** Corp. from extend... their hegemony into the living room.

The chip could help send multiple stream... digital video around a home, even tailoring image resolution to fit each TV or other dev... receiving it. Cell is being targeted for use... Sony's next videogame machine—expected early as 2006—and digital TVs and DVD play... But the technology is also expected to be adap... for products ranging from high-performance c... puters to cellphones.

"We think that Cell is going to forever cha... consumer electronics," says Rick Doherty, an a... lyst at the Envisioneering Group, a market...

Please Turn to Page B4, Column 1

AMD, Via Battle Intel's Atom

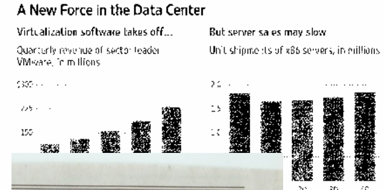
Competition in the mobile-CPU market will feed the growth of a new class of affordable but powerful ultraportable laptops.

Darren Gladstone, PC World
January 8, 2009

BUSINESS TECHNOLOGY

'Virtualization' Is Pumping Up Servers

A New Force in the Data Center



...ad of x86 servers, ...in one primary appli... and an operating sys... , many machines op... % to 25% of their ca...

Palo Alto, Calif., re... veloping a layer of ... ts between the hard... erating system. It al... o run "virtual ma... bination of applica... an operating system ... urces that act like a ... ultiple virtual ma... n each server, simul... different versions of ... dows or variants of ...

...al machine crashes. ...em keep going. Com... ...e virtualization to

roughly 300 servers scattered across 45 departments too many for his three-person staff to handle. Things got so bad, he says, that people were running servers in broom closets and under desks.

When it came time to retire those machines, Mr. Siles used VMware's technology to consolidate the county's computing operations on just 35 new physical servers. He estimates the move will save $30,000 to $40,000 a year in electricity bills. The new machines are running at roughly 65% of their capacity, versus no more than 10% before the shift.

"We hardly ever buy a new server," Mr. Siles says. "The last server I bought was almost a fiscal year ago."

Gentry Ganote, chief information officer for **Golf & Tennis Pro Shop** Inc., which owns PGA Tour superstores, reduced his need for new machines by using Virtual Iron's software and compact Hewlett-Packard computers called blade servers. Another plus: His small staff needs only about a half-hour to set up a new virtual machine—versus a week or more to install a standard server.

"Virtualization is less about cost savings than you might think, though that is an important factor," says Frank Gillett, an analyst at Forrester Research. "It's more about flexibility."

The news isn't all negative for hardware makers. Some customers who need the very highest computing performance—such as for Web search or financial transactions—may be slow to use virtual machines.

PCWorld

With Intel Inside Apple, Macs May Be Faster, Smaller

By Nick Wingfield
And Don Clark

San Francisco

NEARLY TEN YEARS ago, the rivalry between **Apple Computer** Inc. and **Intel** Corp. was so intense that Apple made a TV commercial featuring an Intel "bunny person"—a hooded character in a chip-plant cleanroom suit from Intel's own commercials—charred with fire. The point was that Apple's personal computers were "smoking" the performance of PCs using Intel microprocessors.

Yesterday, though, Apple Chief Executive Steve Jobs and Intel CEO Paul Otellini hugged onstage at Apple's annual gathering for software developers here, as the two announced a historic agreement for Apple to use Intel microprocessors.

The deal paves the way for Apple to introduce new, more powerful Mac computers, including laptops, in increasingly small, thin forms. Apple said it will introduce some Mac models running on Intel chips by June 2006, and complete the transition of its entire line of computers by the end of 2007, ending its use of PowerPC chips from **International Business Machines** Corp. and **Freescale Semiconductor** Inc.

Mr. Jobs described the move to Intel chips as the third major technological transition in the company's history, with the first being Apple's move to PowerPC microprocessors in the mid-1990s from an earlier generation of chips made by Motorola. The second transition was Apple's switch to Mac OS X, the umbrella name for its current generation Mac operating system, which is far more stable and flexible than prior versions. The current operating system is far less tightly wedded to the microprocessor chip than was the prior version, called OS 9.

Yet the switch to Intel for Apple is a major undertaking that carries all sorts of risks. Apple must not only ensure that the Mac operating system runs well on the new hardware, but it has to persuade its thousands of software developers to adapt their application programs, such as word processors and graphics programs, to run on the new systems.

Mr. Jobs said the company will ship the future machines with a utility program called Rosetta that automatically adapts PowerPC programs so that Intel-based Macs can run them. He demonstrated several programs running with the technology, though users would likely eventually want to shift to new software to exploit the full performance of the new hardware.

Analysts said demonstrating Rosetta was important for reassuring users. The technology change "is huge," said Michael Gartenberg, an analyst at Jupiter Research, "but it should have very little impact on Apple's customers."

While Apple only recently struck the deal with Intel, Apple yesterday confirmed longstanding rumors that it has for years been laying the technological foundation for a possible switch to chips made by the Santa Clara, Calif., semiconductor giant.

In a speech, Mr. Jobs described how Apple's OS X has been leading a "secret double life" on the Apple campus in Cupertino, Calif., where a squad of engineers has for five years stealthily maintained versions of the Mac operating system that run on Intel chips, even as Apple was selling only

Please Turn to Page B7, Column 1

Chip Switch

Apple will introduce Macs with Intel chips next June. What it means:

- With Intel, Macs can be more powerful and, because the chips emit less heat, smaller
- Software developers must adapt programs using tools from Apple
- Apple will help existing programs run on Intel-powered Macs, though some performance penalty is likely

Sources: Apple Computer; WSJ research

Journal Link: *Apple CEO Steve Jobs discusses the transition to using Intel processors in a video report, at* **WSJ.com/Free**.

How do computers work?

- Need to understand abstractions such as:
  - Applications software
  - Systems software
  - Assembly Language
  - Machine Language
  - Architectural Issues: i.e., Caches, Virtual Memory, Pipelining
  - Sequential logic, finite state machines
  - Combinational logic, arithmetic circuits
  - Boolean logic, 1s and 0s
  - Transistors used to build logic gates (CMOS)
  - Semiconductors/Silicon used to build transistors
  - Properties of atoms, electrons, and quantum dynamics

Computer Architecture

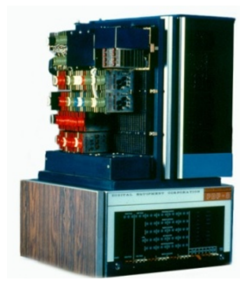
- Technology changes require continuous optimization of cost/performance
  - Moore's Law ("performance" doubling every 18 months)
  - Design decisions
- Innovation needed to satisfy market trends
  - Larger memory
  - Greater interactivity
- Changing user requirements make demands on technology
  - Portability → Lower Power
  - Graphics → Higher Performance

Computer Architecture – Historical Evolution

Mainframe

Supercomputer

Massively Parallel Computer



Mini-computer

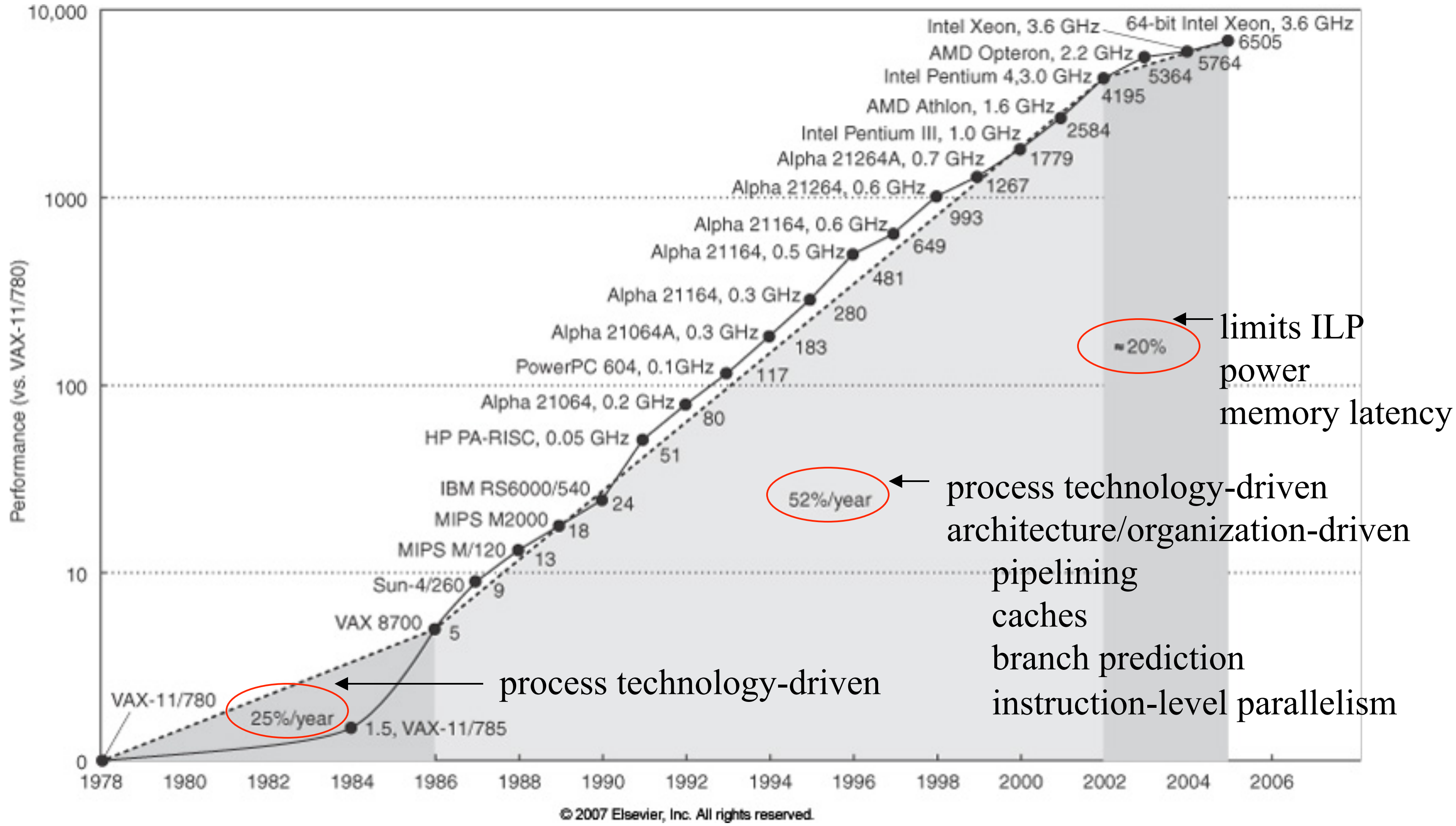
Super mini-computer

Impact of microprocessor on computer architecture

PC

Workstation

Growth in Microprocessor Performance



Growth in Microprocessor Performance

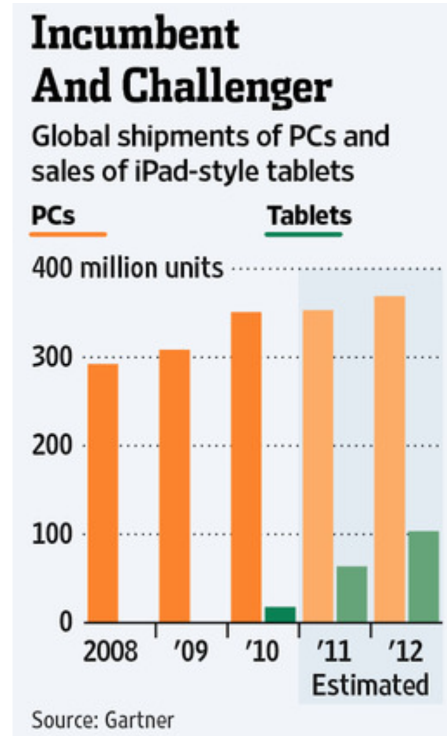
- Rate of improvement in microprocessors since early 1980s has led to major changes and the dominance of microprocessor-based computers
  - Entire new categories have emerged
    - Workstations, desktop PCs, laptops
    - Servers, embedded computers, handheld computers
    - PDAs, Smartphones
    - Netbooks
  - Minicomputers gone – replaced by servers using microprocessors
    - Digital (DEC), Data General, Prime – gone or absorbed
  - Mainframes virtually gone – replaced by multiprocessors using microprocessors
    - IBM, Dell, SUN, HP

Disruptive technology

Primary Segments Today

- Desktop and laptop computing
  - PCs (Windows, Linux)
    - Desktop, Laptop/Notebook, Netbook
  - Workstations (higher performance)
    - e.g. running CAD/EDA applications
    - SUN Sparc/Unix (Windows/NT)
- Servers
- Embedded Systems
- Personal Mobile Devices (PMD)
- Clusters/Warehouse-scale computers (WSC)

Desktop and laptop computing



- Market
  - $300 to $2,500 range (approximate)
    - Processor cost: $50 - $500
  - ~50% of PCs sold are laptops
  - ~300M annual unit volume
  - Largest segment in dollar terms
- Criteria
  - Price/performance dominates desktop
  - Graphics performance very important gaming/video markets
  - Power consumption critical for laptops
    - Can command a premium
    - Battery life, no fans
  - Well characterized applications and benchmarks
    - Guide design decisions

Servers

- Market
  - $5,000 to $10,000,000 (approximate)
    - Processor cost: $200 - $2,000 per processor
  - ~4M annual unit volume
  - Traditional "back office" applications
    - Airline reservations, banking and financial services
- Criteria
  - Scalability
    - Accommodate growth (low initial capital, incremental growth)
  - Throughput
    - Latency vs. throughput
  - Availability
    - Paramount importance (not like your PC…)
  - Energy

Cost of an Unavailable System

Application	Cost of downtime per hour (thousands of $)	Annual losses (millions of $) with downtime of 1% (87.6 hrs/yr)	0.5% (43.8 hrs/yr)	0.1% (8.8 hrs/yr)
Brokerage operations	$6450	$565	$283	$56.5
Credit card authorization	$2600	$228	$114	$22.8
Package shipping services	$150	$13	$6.6	$1.3
Home shopping channel	$113	$9.9	$4.9	$1.0
Catalog sales center	$90	$7.9	$3.9	$0.8
Airline reservation center	$89	$7.9	$3.9	$0.8
Cellular service activation	$41	$3.6	$1.8	$0.4
Online network fees	$25	$2.2	$1.1	$0.2
ATM service fees	$14	$1.2	$0.6	$0.1

Figure 1.2 The cost of an unavailable system is shown by analyzing the cost of downtime (in terms of immediately lost revenue), assuming three different levels of availability and that downtime is distributed uniformly. These data are from Kembel [2000] and were collected and analyzed by Contingency Planning Research.

Amazon.com: $12.95B in sales Q4 2010 = $5.9M/hour

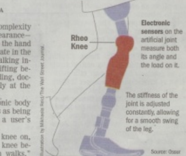
Embedded Computers

- Market
  - \$10 to \$100,000 price range (approximate)
    - Processor cost \$0.01 – \$100
  - ~500M 32-bit and 64-bit processors (2x PC market) ~Billions of 8-bit and 16-bit MCUs!
  - Largest segment by <u>unit</u> volume
  - Ubiquitous
    - Oldest microprocessor application
    - Automobiles, appliances, printers, PDAs, cell phones, set-top boxes, industrial controls, networking switches
- Criteria
  - Price
  - Real-time performance
  - Power consumption
  - Memory (code size)

Wide Applications

- Automotive
- Appliances
- Medical instrumentation
- Industrial Controls
- Networking

- >150 motors to control
- >15 microcontrollers
- >100 sensors to be checked
- Airbag and seat belt pre-tensionsrs
- 60 kW of electrical power to control



Bionic Knee 'Learns' How to Walk

New Prosthesis Is Designed To Adapt to the Movements Of Users at Varying Speeds

By Nicholas Zamiska

The human knee—whose complexity is belied by its hinge-like appearance—has proved to be second only to the hand in difficulty for doctors to replicate in the effort to help out amputees. Walking involves a delicate pattern of shifting between bearing weight and bending, doctors say, with the knee literally at the center.

Enter the Rheo Knee, a bionic body part that its maker, Ossur, touts as being able to "learn" and adapt to a user's idiosyncratic movements.

"We can basically turn the knee on, have the person walk, and the knee begins to learn how that person walks," said Scott B. Elliott, a prosthetist with Ossur who has been testing the knee for the past three years. "If they start to change their walking speed, Rheo will be

Technical Strides

A new prosthetic knee adapts to its user's gait automatically and can also be controlled by a handheld computer.

$18,000, although the final price tag for a knee, foot, socket and labor of a prosthe-

C-Leg, but he is eager to see the Rheo Knee come out of testing. "We're still a long way off from recreating a leg as good as the one you're born with," he says.

But Ossur's knee, which uses the same magnetic technology found in the shock absorption systems installed in some new Corvettes, may have an Achilles' heel, so to speak, according to competitors familiar with the design. When its battery dies, the knee becomes a "free swinging joint," leaving the user prone to a fall.

'Reviewed and Abandoned'

Indeed, Otto Bock is dismissive of Rheo. "The Rheo Knee, at its heart, uses a very sexy technology, which we reviewed and abandoned nine years ago," said Patrick Chelf, vice president of marketing and business development for Otto Bock in the Americas.

Otto Bock says that its C-Leg is the most reliable knee in the industry, helping users avoid falls.

Technicians at Ossur counter that

Embedded

Smart shoe debuts

Athletic-shoe manufacturer **Adidas** is touting a smart running shoe that uses a microprocessor to adapt the shoe's cushioning to a runner's size and stride. The processor, embedded in the arch of the Adidas 1, drives a tiny screw and cable system that adjusts the heel cushion based on signals sent back by a sensor attached to a magnet. The system is powered by a battery that conserves power by adjusting the shoe while the foot is elevated mid-stride. The scheme is designed to reduce drag and avoid resistance from the ground. The kit weighs less than 40 grams, or about 10 percent of the weight of an average running shoe. The Adidas 1 will retail for $250 a pair—with no guarantee you'll break four minutes in the mile.

Foot race: Adidas says the world's first smart shoe will give runners that extra edge down the stretch.

- Digital satellite radio
- GPS guidance systems
- ABS brakes
- Fuel injection
- Anti-collision systems
- Video camera
- Entertainment systems (DVD, CD, MP3)
- Anti-theft system

Embedded Computers

- Issues
  - Real-time performance
    - Consequences of missing an interrupt
      - Soft real-time (fail to meet requirement)
        - » Video-conferencing: jerking image from missing frame
      - Hard real-time (life/death, property damage)
        - » Airline crash, patient death, divert tons of pollutant in river
  - Software
    - Often a single program for life of system
      - Exceptions
        - » program revisions for your BMW 7 series
        - » Downloadable updates (cell phones, PDAs)
    - Ability to optimize application
      - Unlike PC or (usually) server market [NFS]
  - Optimize for memory usage
    - Why? Often not using DRAM but small flash and on-board
    - How? Instruction Set Architecture (ISA) impact…

Embedded Computers

- Issues
  - Power consumption/dissipation (heat)
    - Battery life
    - May not be able to use fans
    - Less expensive package
      - Plastic vs. Ceramic
  - May not use discrete microprocessor
    - “Core” with application-specific custom hardware
    - e.g. Xilinx FPGA with PowerPC core (ARM , MCU)
  - Another approach: Digital Signal Processors (DSPs)
    - Optimized architectures for repetitive operations
      - Dot product, FFTs (multiply/accumulate)

Personal Mobile Devices (PMD)

- Market
  - $100 - $100 Price range
    - Processor cost $10 - $100
    - Wireless devices with multimedia interfaces
      - Cell phones, tablets
- Criteria
  - Cost
  - Energy Efficiency (battery operated)
  - Real-time performance
    - Video playback, responsiveness
  - Memory

Clusters/Warehouse Scale Computers (WSC)

- Market
  - $100,000 - $200,000,000 Price range
    - Processor cost $50 - $250
    - Collections of desktop or servers connected via LANs
    - Acting as single computer
    - Web services (Amazon, hosting, etc)
- Criteria
  - Price-performance
  - Throughput
  - Energy

Summary of Classes of Computers

Features	Personal Mobile Device (PMD)	Desktop	Server	Cluster/ Warehouse-Scale Computer (CWS)	Embedded
System Price	\$100 - \$1000	\$300 - \$2500	\$5K -\$10M	\$100K - \$200M	\$10 - \$100K
μprocessor Price	\$10 - \$100	\$50 - \$500	\$200 - \$2000	\$50 - \$250	\$0.01 - \$100
Critical Design Issues	Cost, energy, media performance, responsiveness	Price-performance, energy, graphics performance	Throughput, availability, scalability, energy	Price-performance, throughput, energy proportionality	Price, energy, application-specific, memory

Parallelism

- Data-level parallelism
  - Many data items to be operated on
- Task-level parallelism
  - Multiple tasks that can operate independently in parallel

- Instruction-level parallelism
  - Pipelining, speculative execution
- Vector architectures and Graphic Processor Units (GPUs)
  - Apply single instruction to collection of data in parallel
- Thread-level parallelism
- Request-level parallelism

Flynn's Taxonomy

- Single instruction, single data stream (SISD)
  - Uniprocessor (though may exploit instruction level parallelism
  - Superscalar, speculative execution
- Single instruction, multiple data stream (SIMD)
  - Same instruction executed on multiple processors with different data streams (exploit data level parallelism)
  - Vector architectures, GPUs, multimedia extensions to instructions sets
- Multiple instruction, single data stream (MISD)
  - No example exists
- Multiple instruction, multiple data stream (MIMD)
  - Each processors executes own instructions on own data
  - Thread-level (multicore) and task-level (clusters/WSC)

Definition of Computer Architecture

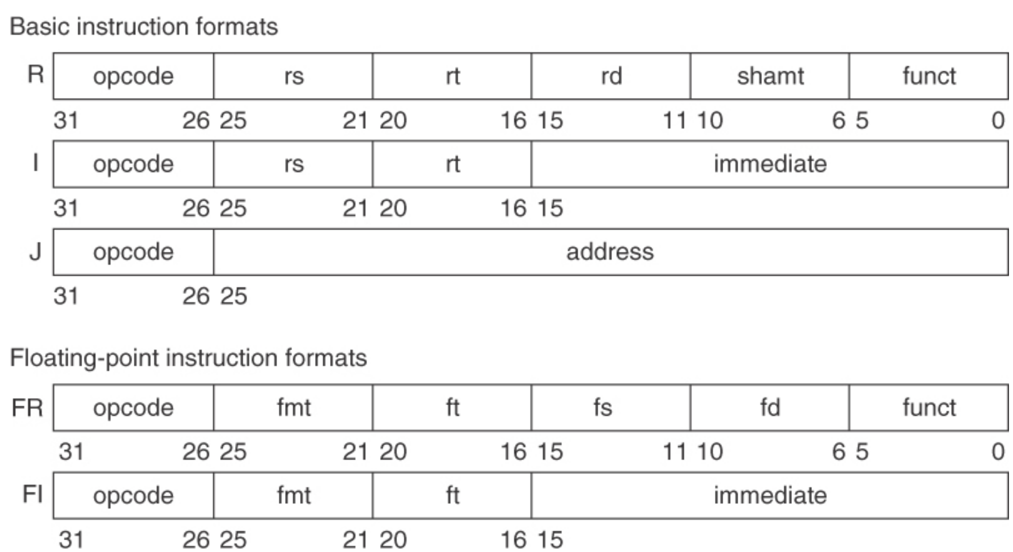
- Historically, computer architecture was concerned with instruction set architecture (rest was “implementation” or “organization”). AMD and Intel processors both implementing same x86 instruction set have same architecture despite different pipelines, caches, branch prediction algorithms
- Our text uses architecture to refer to:
  - Instruction Set Architecture (ISA): H/W & S/W boundary
    - Idea is that compiler, OS, programmer indifferent if ISA the same. True except for performance!
  - Organization: high-level aspects and structure of the computer
    - Caches, pipelines, branch prediction
  - Hardware: detailed logic design, packaging, process technology
  - Successful computer architect must consider all these aspects as well as functional requirements of the segment, price, power, performance objectives

Instruction Set Architecture

- Class of ISA
  - Register/memory, Load/store
  - Registers, accumulator, stack
- Memory addressing
  - Byte/word addressing
  - Alignment
- Addressing modes
  - Register, immediate, displacement, indirect, auto-increment
- Types and size of operands
  - 8-bit (ASCII character), 16-bit (Unicode character, halfword)
  - 32-bit (integer, word), 64-bit (long integer, double word)
  - Floating point (32-bit single precision, 64-bit double precision)
  - Strings, multimedia support
- Operations
  - Data transfer, arithmetic, logic, control, floating point

Instruction Set Architecture

- Control flow
  - Conditional, unconditional branches, subroutine calls/returns
  - Condition codes vs. register contents
  - Return address via register or stack
- Encoding
  - Variable or fixed length instructions
  - Field lengths, position



Examples

- Same ISA, different organization
  - NEC VR 5432 and VR 4122 embedded MIPS 64 processors
  - VR 5432
    - Pipeline
    - Cache
    - Floating point hardware
  - VP 4122
    - No pipeline
    - No Cache
    - Floating point implemented in S/W
- Same ISA, different hardware
  - Memory controller off-chip or on-chip
  - Integrated graphics controller or separate chip
  - Achieve different price/performance point

Important Functional Requirements

- Application area
  - General purpose desktop
  - Scientific workstations
  - Servers
  - Embedded systems
- Software compatibility
  - Source or binary
  - Why? Where most important? Where least important?
  - Intel mantra, but at what cost?
- Operating system requirements
  - Address space, memory management, protection
- Standards
    - IEEE 754 floating point, I/O bus, networks, OS, programming languages