CS 347: Introduction

Todd C. Mowry

Jan 13, 1998

Topics:

- **Theme**
- **Technology**
- **Historical perspective**

Course Theme

Every computer scientist/engineer should know what's under the hood

Why?

- **Some will build them**
- **Everyone will use them**
  - The more you know, the more effective you will be

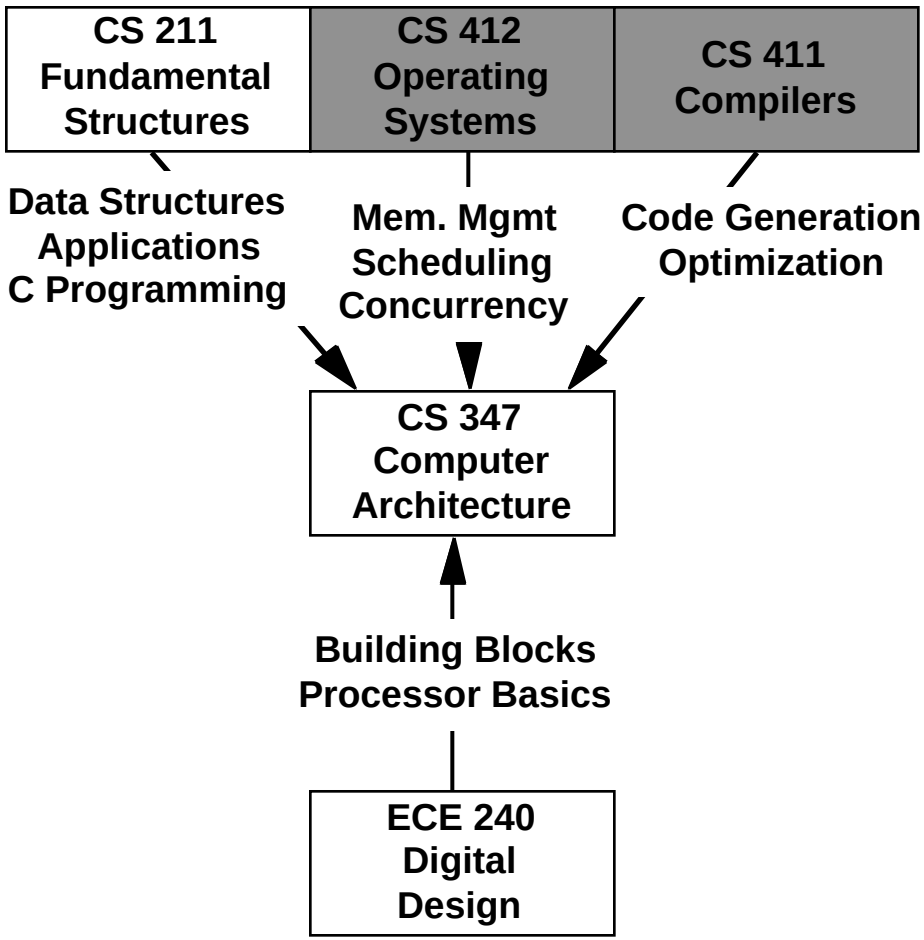
"User-Centric" Computer Systems

- **Our course is designed to make you "enlightened users"**
- **Consider entire system: processor, memory, I/O, storage, network**

Useful Outcomes

- **How to make programs run faster**
- **What kind of hardware does an application require**
- **How will things change over time**

Role within Curriculum



Architecture Bridges from EE to CS

- What are the application requirements?
- What features are required to support O.S.?
- What features can optimizing compilers use?
- What hardware can I build?
- What hardware can I build next year?

Computer architect must be broad!

Technology Improvements

Measure	*Evolution*
Integrated Circuit Logic	
• Transistors / chip	4X every 3 years
• Device speed	4X every 3 years
Dynamic Random Access Memory (DRAM)	
• Bits / chip	4X every 3 years
• Access Time	2/3X in 10 years
Magnetic Disks	
• Capacity	4X every 3 years
• Access Time	2/3X in 10 years
• Cost / byte	1/2X per year

12 Year Comparison

Parameter	*1985*	*1997*
Typical System		
Processor	**PC-AT (80286)**	**Pentium II**
RAM	**256K**	**64MB**
Disk	**20MB**	**4GB**
Clock Rate	**4MHz**	**300MHz**
Spec. Int ‘95	**~0.01**	**11.7**
Cost	**$5000**	**$2200**
Per Megabyte Memory Costs		
DRAM	**$350**	**$2**
Disk	**$47**	**$0.08**

Impact of Technology

It's the Technology, Stupid!

- Computer science has ridden the wave

Things Aren't Over Yet

- Technology will continue to progress along current growth curves
- For at least 10 more years
- Difficult technical challenges in doing so

Even Technologists Can't Beat Laws of Physics

- Noticeable effects when devices below 0.1μm (~2007)
- Quantum effects create fundamental limits as approach atomic scale
- Opportunities for new devices

Risk of Predicting the Future

Incremental Improvements Exceed Wildest Dreams

- **Silicon CMOS**
- **Magnetic disks**
- **DRAM**

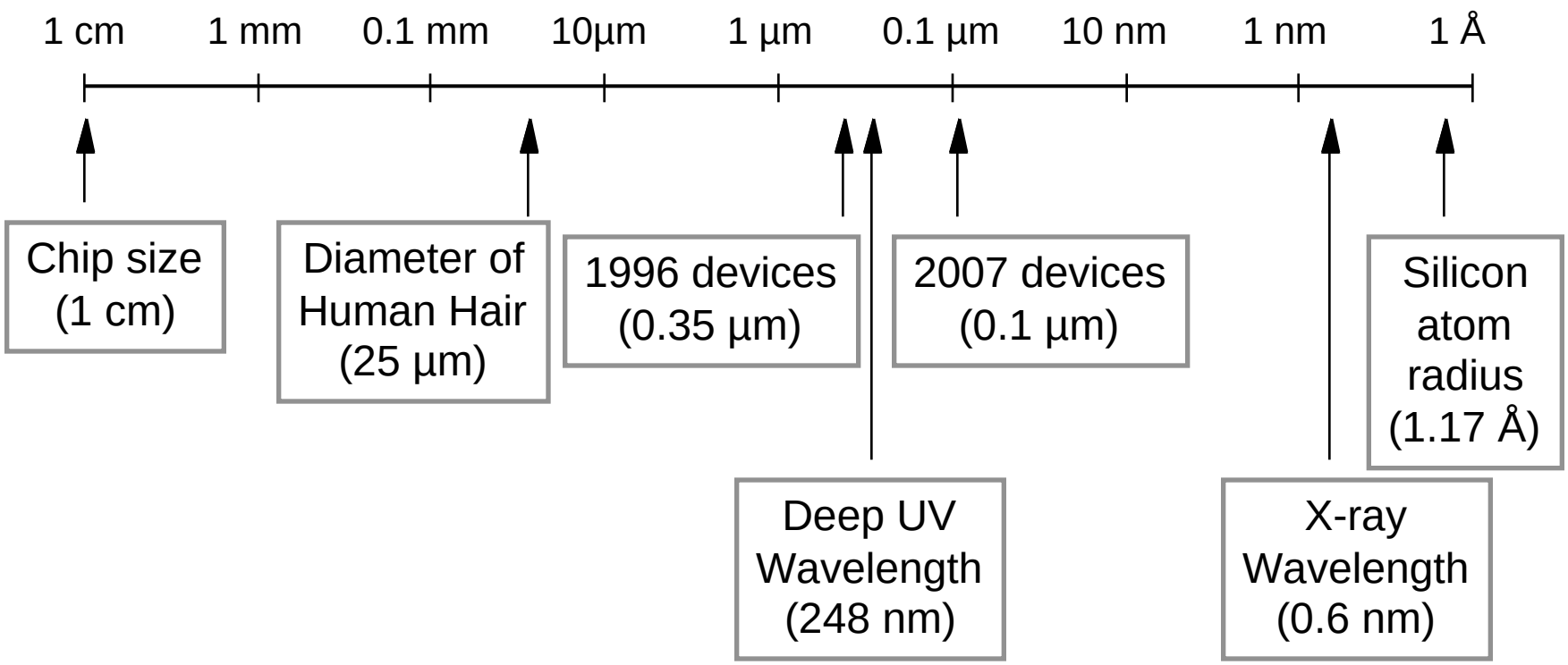
Hopes for Future Technology Never Materialize

- **Magnetic bubble memory**
- **CCD memory**
- **Gallium Arsenide**

Observations

- **In this business, “incrementing” is by multiplicative factor**
- **Economies of scale favor existing technology**
- **Shifts occur due to new market forces**
  - Drive for low power due to desire for portability
  - Emphasis on networking due to WWW

Dimensions


1 cm
1 mm
0.1 mm
10µm
1 µm
0.1 µm
10 nm
1 nm
1 Å
Chip size (1 cm)
Diameter of Human Hair (25 µm)
1996 devices (0.35 µm)
2007 devices (0.1 µm)
Deep UV Wavelength (248 nm)
X-ray Wavelength (0.6 nm)
Silicon atom radius (1.17 Å)

Scaling to 0.1µm

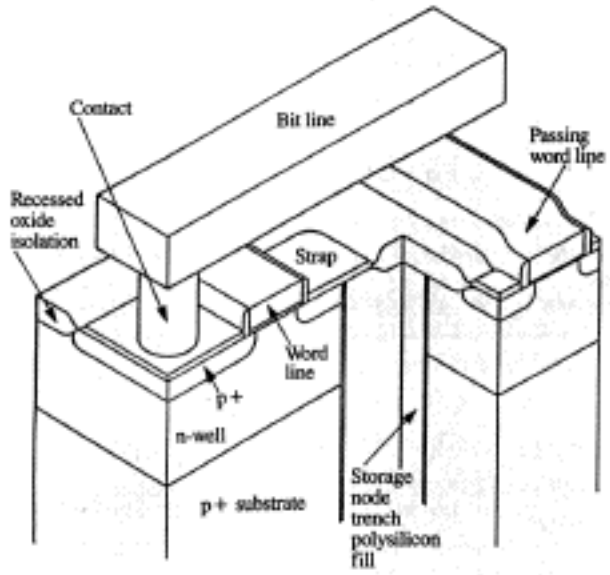
- **Semiconductor Industry Association, 1992 Technology Workshop**

Year	1992	1995	1998	2001	2004	2007
Feature size	0.5	0.35	0.25	0.18	0.12	0.10
DRAM cap	16M	64M	256M	1G	4G	16G
Gates/chip	300K	800K	2M	5M	10M	20M
Chip cm^2	2.5	4.0	6.0	8.0	10.0	12.5
Intercn. levels	3	4–5	5	5–6	6	6–7
Supply Volts	5.0	3.3	2.2	2.2	1.5	1.5
I/Os	500	750	1500	2000	3500	5000
off chip MHz	60	100	175	250	350	500
on chip MHz	120	200	350	500	700	1000

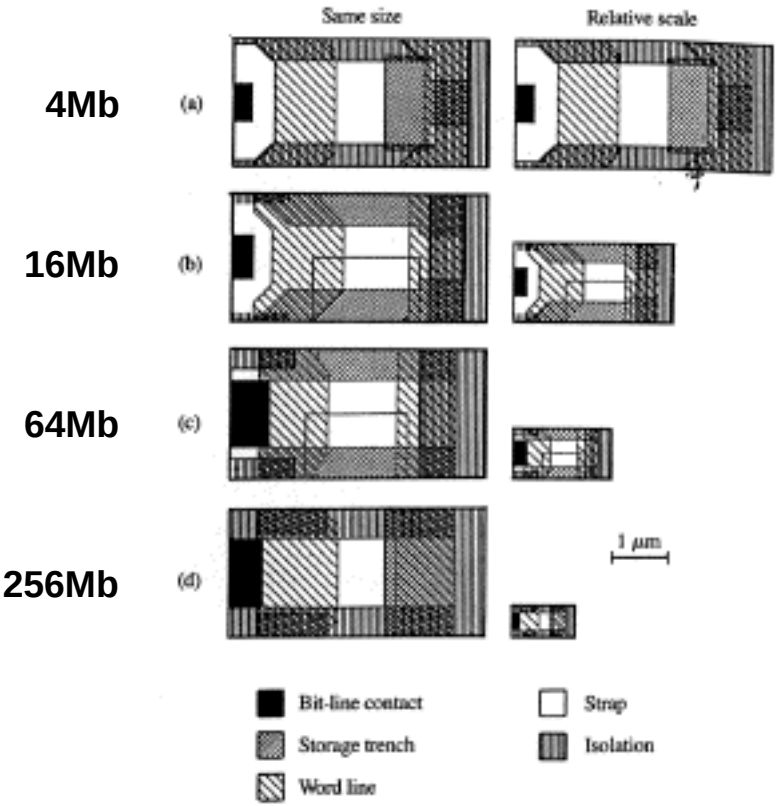
IBM DRAM Evolution

- **IBM J. R&D, Jan/Mar '95**
- **Evolution from 4 – 256 Mb**
- **256 Mb uses cell with area 0.6 μm²**

4 Mb Cell Structure



Cell Layouts



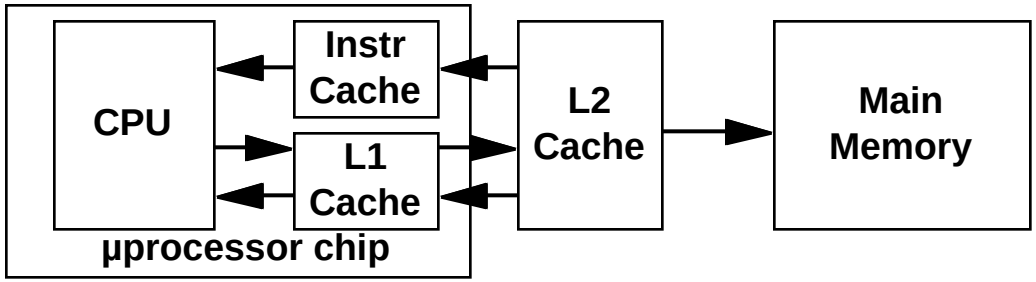
Nonuniform Progress

Processor-Memory Performance Gap

- **Processor speed improving at 60% / year**
  - 2000X improvement since 1980
- **DRAM latency improving at 7 % / year**
  - < 10X improvement since 1980
- **Accessing memory incurs higher performance penalty**

Solutions

- **Caches: Small, fast memories holding most recent data/code**



- **Wider memory busses: More bytes / access**
- **Code Optimization: Make more use of registers**

Some History

Who Cares?

- To understand anything that changes quickly, must look to the past

Ancient

- Big mainframes
- Von Neumann saw a need for at most 10 machines in the U.S.
- Groundwork laid for much of today's architecture

Microcode is King

- IBM 360 – VAX
- CPU occupied multiple logic boards
- Implement instruction set by interpretation with microengine
- Easy to add (lots of) features
- Performance not major issue (proprietary platforms)

More History

RISC Revolution

- **~1983, IBM 801, UCB RISC project, Stanford MIPS project**
- **Make the microengine be the CPU!**
  - Avoid inefficiency of interpretation layer
  - Let compilers do the optimizing
  - Implement on single chip
- **Generic Unix Box**

Superscalar Processing

- **~1990 (IBM Power-1) to present (almost all processors today)**
- **Basic idea: issue *multiple* instructions simultaneously**
  - exploit fine-grained parallelism within the instruction stream
- **In-order (Alpha 21164) vs. out-of-order (MIPS R10K, Pentium Pro)**
- **Aggressive, pipelined cache and memory subsystems**

Recent Trends

How to push Superscalar further?

- **Hardware complexity (rather than chip real estate) is the limitation**
- **One direction: VLIW (Intel "EPIC" instruction set in IA-64)**

Parallel processing

- **supercomputers -> computer servers -> desktop**

Personal computers remain a key driving force

- **Huge economies of scale**

Customers want more than raw computing power

- **Fast WWW access, multimedia (3D graphics, video, etc.)**
  - Difficult real-time performance constraints
  - ISA extensions (e.g., Intel MMX) + balanced system-level design

Mobile Computing

- **Low power consumption is important**