ESE534:
Computer Organization

Day 1: January 13, 2010
Introduction and Overview

Today

- Matter Computes
- Architecture Matters
- This Course (short)
- Unique Nature of This Course
- Change
- More on this course

2

## Review (assert?):
Two Universality Facts

- NAND gate Universality [Day 2, ESE200/CIS240]
  - We can implement any computation by interconnecting a sufficiently large network of NAND gates
- Turing Machine is Universal [CIS262]
  - We can implement any *computable* function with a TM
  - We can build a single TM which can be programmed to implement any computable function
- Day 2 reading (on Blackboard) SciAm-level review

3

Matter Computes

- We can build NAND gates out of:
  - transistors (semiconductor devices)
    - physical laws of electron conduction
  - mechanical switches
    - basic physical mechanics
  - protein binding / promotion / inhibition
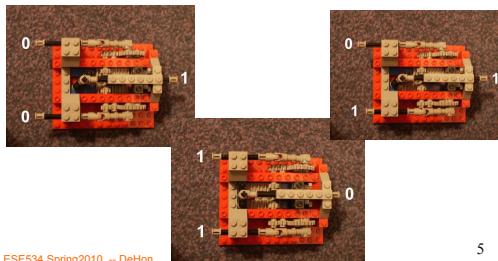   - Basic biochemical reactions
  - …many other things
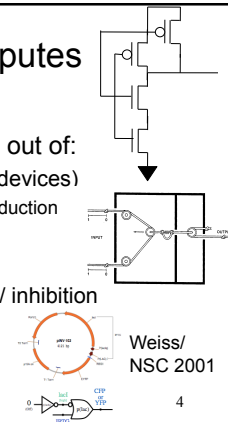
Weiss/
NSC 2001

4

LEGO™ Logic Gates

- http://goldfish.ikaruga.co.uk/logic.html

5

Starting Point

- Given sufficient raw materials:
  - can implement any computable function

- Our goal in computer architecture
  - is **not** to figure out how to compute new things
  - rather, it is an **engineering** problem

6

1

Engineering Problem

- Implement a computation:
  - with least resources (in fixed resources)
    - with least cost
  - in least time (in fixed time)
  - with least energy

- Optimization problem
  - how do we do it best?
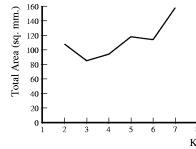
7

Quote

- "An Engineer can do for a dime what everyone else can do for a dollar."

8


Architecture Matters?

- How much difference is there between architectures?
- How badly can I be wrong in implementing/ picking the wrong architecture?
- How efficient is the IA-32, IA-64, GPGPU?
  - Is there much room to do better?
- Is architecture done?
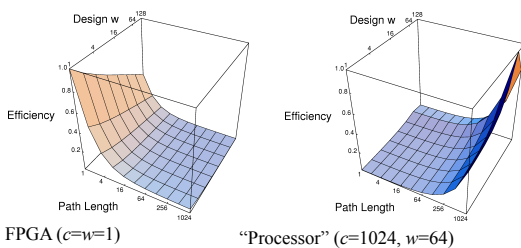 - A solved problem?

9

Peak Computational Densities from Model

- Small slice of space
  - only 2 parameters
- 100× density across

- Large difference in peak densities
  - large design space!
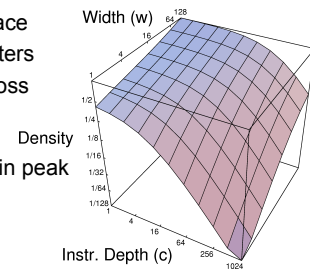
10

Yielded Efficiency

FPGA ($c=w=1$)   "Processor" ($c=1024$, $w=64$)

- Large variation in **yielded** density
  - large design space!

11

Architecture **Not** Done

- Many ways, not fully understood
  - design space
  - requirements of computation
  - limits on requirements, density...
- …and the *costs* are **changing**
  - optimal solutions change
  - creating new challenges and opportunities

12

Personal Goal?

- Develop systematic design
- Parameterize design space
  - adapt to costs
- Understand/capture req. of computing
- Efficiency metrics
  - (similar to information theory?)
- …we'll see a start at these this term

Compute
Memory
Interconnect

Architecture **Not** Done

- Not here to just teach you the forms which are already understood
  - (though, will do that and give you a strong understanding of their strengths and weaknesses)

- **Goal**: enable you to design and synthesize new and better architectures

This Course (short)

- How to organize computations
- Requirements
- Design space
- Characteristics of computations
- Building blocks
  - compute, interconnect, retiming, instructions, control
- Comparisons, limits, tradeoffs

This Course

- Sort out:
  - Custom, RISC, SIMD, Vector, VLIW, Multithreaded, Superscalar, EPIC, MIMD, FPGA, GPGPUs
- Basis for design and analysis
- Techniques

- [more detail at end]

Graduate Class

- Assume you are here to learn
  - Motivated
  - Mature
  - Not just doing minimal to get by and get a grade
    - …at the last minute
- Problems
  - May not be fully, tightly specified

Uniqueness of Class

Not a Traditional Arch. Class

- Traditional class (240, 370, 501)
  - focus RISC Processor
  - history
  - undergraduate class on μP internals
  - then graduate class on details
- This class
  - much broader in scope
  - develop design space
  - see RISC processors in context of alternatives

Authority/History

- ``Science is the belief in the **ignorance** of experts.''
      -- **Richard Feynman**
- Traditional Architecture has been too much about history and authority
- Should be more about engineering evaluation
  - **physical world** is "final authority"
- **Goal**: Teach you to think critically and independently about computer design.

Next Few Lectures

- Quick run through logic/arithmetic basics
  - make sure everyone remembers
  - (some see for first time?)
  - get us ready to start with observations about the key components of computing devices
- Trivial/old hat for many
  - But will be some observations couldn't make in ESE200/CIS370
- May be fast if seeing for first time
- Background quiz intended to help me tune

Themes

- Design Space
- Parameterization
- Costs
- Change
- Structure in Computations

Focus

- Focus on raw computing organization
- **Not** worry about nice abstractions, models
  - 501, 370, 240 provide a few good models
    - Instruction Set Architecture (ISA)
    - Shared Memory
    - Transactional…
  - …and you should know others
    - Dataflow, streaming, data parallel, …

Change

- A key feature of the computer industry has been rapid and continual change.
- We must be prepared to adapt.

- True of this course as well
  - ….things are still changing…
 - We'll try to figure it out together…

4

What has changed?

- [Discuss]
- Capacity
  - Total
  - Per die
- Size
- Applications
  - Number
  - Size/complexity of each
  - Types/variety
- Use Environment
  - Embedded
  - Mission critical

- Speed
  - Ratio of fast memory to dense memory
  - Wire delay vs. Gate delay
  - Onchip vs. inter-chip
- Joules/op
- Mfg cost
  - Per transistor
  - Per wafer
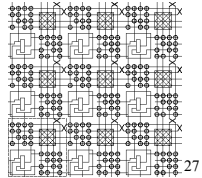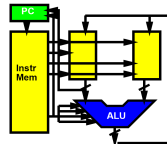 - NRE (Non-Recurring Engineering)
- Reliability
- Limited by
  - Transistors, energy…

25

Intel's Moore's Law
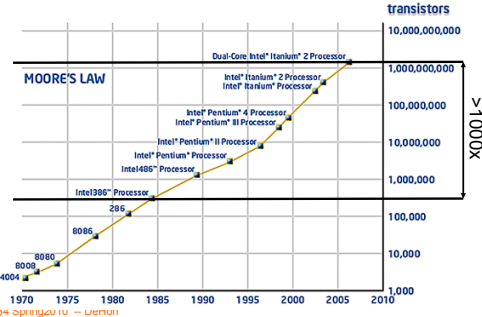
1983 (early VLSI)

- Early RISC processors
  - RISC = Reduced Instruction Set Computer
  - RISC-II, 40K transistors
  - MIPS, 24K transistors
  - ~10MHz clock cycle
- Xilinx XC2064
  - 64 4-LUTs
    - LUT = Look-Up Table
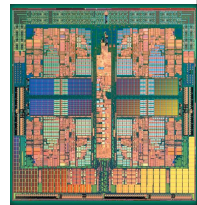   - 4-LUT – program to be any gate of 4 inputs
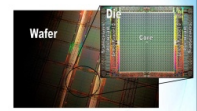
27

Today

AMD Quad Core Opteron

- CPUs
  - Billions of transistors
  - Quad CPU per die
  - Multi-issue, 64b processors
  - GHz clock cycles
  - MByte caches
- FPGAs
  - >500,000 bit processing elements
 - 10s of Mbits of on-chip RAM

Altera Stratix IV

http://www.ciol.com/Semicon/SemiPipes/News-Reports/Altera-starts-shipping-40-nm-Stratix-IV-FPGAs/161208113780/0/

More chip capacity?

Cosmic Cube / CACM 1985

- Should a 2010 single-chip multiprocessor look like a 1983 multiprocessor systems?
  - Processor→processor latency?
  - Inter-processor bandwidth costs?
  - Cost of customization?
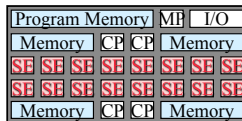
Calisto™ BCM1500
Nichols/Microprocessor Forum 2001

29

Memory Levels

- Why do we have 5+ levels of memory today?
  - Apple II, IBM PC had 2
  - MIPS-X had 3

30

MOS Transistor *Scaling*
(1974 to present)

S=0.7
[0.5x per 2 nodes]

	Pitch	Gate

[from Andrew Kahng]

Will This Last Forever?

L_g (μm)

[Moore, ISSCC2003]

	Pitch	Gate
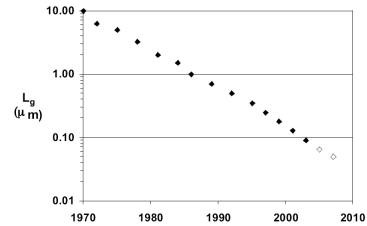
Historical Power Scaling

Feat Size (um)
Vdd
Power/10
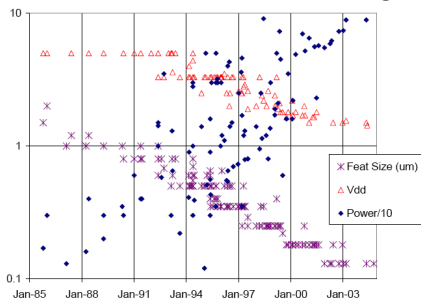
[Horowitz et al. / IEDM 2005]

Interesting Times

- Challenges to continue scaling
  - Power density
  - Reliability

- What does the end-of-scaling mean to architecture?

Class Components

Class Components

- Lecture (incl. preclass exercise)
- Reading [~1 required paper/lecture]
  - No text (mostly online: Blackboard, IEEE, ACM)
- 9 assignments
  - (roughly 1 per week)
- Final design/analysis exercise
  - (4 weeks)
- Note syllabus, course admin online
  - See URL bottom of logic assignments

Preclass Exercise

- Like Background Quiz but more focused
- Motivate the topic of the day
  - Introduce a problem
  - Introduce a design space, tradeoff, transform
- Work for 5-10 minutes before start lecturing

37

Feedback

- Will have anonymous feedback sheets for each lecture
  - Clarity?
  - Speed?
  - Vocabulary?
  - General comments

38

Fountainhead Quote

Howard Roark's Critique of the Parthenon
-- Ayn Rand

39

Fountainhead Parthenon Quote

"Look," said Roark. "The famous flutings on the famous columns---what are they there for? To hide the joints in wood---when columns were made of wood, only these aren't, they're marble. The triglyphs, what are they? Wood. Wooden beams, the way they had to be laid when people began to build wooden shacks. Your Greeks took marble and they made copies of their wooden structures out of it, because others had done it that way. Then your masters of the Renaissance came along and made copies in plaster of copies in marble of copies in wood. Now here we are making copies in steel and concrete of copies in plaster of copies in marble of copies in wood. Why?"

40

41

Computer Architecture Parallel

- Are we making:
  - copies in submicron CMOS
  - of copies in early NMOS
  - of copies in discrete TTL
  - of vacuum tube computers?

42

7

Big Ideas

- Matter Computes
- Efficiency of architectures varies widely
- Computation design is an engineering discipline
- Costs change $\Rightarrow$ Best solutions (architectures) change
- Learn to cut through hype
  - analyze, think, critique, synthesize

43

Post Class

- NRE = Non-Recurring Engineering
  - The fixed cost in contrast to the per piece cost
- Scaling – Moore's Law (26) and 31—33
  - Develop on Day 7
- Universality – read Day 2 materials on Blackboard
- Slides – note that I will post slides before lecture; print out a copy if you want them to follow along in lecture

44

8