What does the Future Hold?

Prof. Hakim Weatherspoon
CS 3410, Spring 2015
Computer Science
Cornell University
Announcements

Final Project

Demo Sign-Up via CMS.

sign up Tuesday, May 12\textsuperscript{th}
or Wednesday, May 13\textsuperscript{th}

CMS submission due:

• Due 6:30pm Wednesday, May 13\textsuperscript{th}
Announcements

Prelim2 Results

• Mean 61.5 ± 17.3 (median 62), Max 95.5
• Pickup in Homework Passback Room (216 Gates)
Prelim2 Results

$2^{64} = 16EB$

8 byte = 64-bit

<table>
<thead>
<tr>
<th>V</th>
<th>R</th>
<th>W</th>
<th>X</th>
<th>Physical Page Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td>0x10045</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>0xC20A3</td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td>0x4123B</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>0x10044</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>0x10044</td>
</tr>
</tbody>
</table>

Virtual Memory

16kB

Physical Memory

$2^{48}$ or 256TB

16kB

34-bit = 48-bit – 14 bit Physical Memory
### Prelim2 Results

2^{64} = 16EB

8 bytes = 64-bit

<table>
<thead>
<tr>
<th>V</th>
<th>R</th>
<th>W</th>
<th>X</th>
<th>Physical Page Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td>0x10045</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>0xC20A3</td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td>0x4123B</td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td>0x10044</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Virtual Memory

2^{64} = 16EB

2^{14} = 2^{50}

2^{50} x 8 = 2^{53}

8PB

34-bit = 48-bit – 14 bit

Physical Memory

Virtual Memory

Physical Memory

2^{48} or 256TB

16kE
### Prelim2 Results

- $2^{21} = 2\text{MB}$

### 8 byte = 32-bit

#### Virtual Memory

- $2^{64}$
- $2^{14}$
- $= 2^{50}$
- $2^{50} \times 8 = 2^{53}$
- $8\text{PB} + 2\text{MB}$

#### Physical Memory

<table>
<thead>
<tr>
<th>Physical Page Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>0</td>
</tr>
</tbody>
</table>

- 0x10045
- 0xC20A3
- 0x4123B
- 0x10044

#### Announcements

- $16\text{kE}$
Announcements

Multi-level Page Table

- 25 bits
- 25 bits
- 14 bits
- vaddr

Page Directory

- Page Table

- PDEntry

- PTEntry

Page

- Word

2^{25} \times 8B = 2^{28}B = 256MB = 1 page

Page table = 256MB + 256MB + 2MB
Announcements

Multi-level Page Table

- `vaddr`
- `Page Table`
- `Page`
- `Page Table = 512MB + 2MB = 514MB`

Page Directory
- `PDEntry`

Page Table
- `PTEntry`

Page
- `Word`

- `PTBR`

- `2^{25} \times 8B = 2^{28}B = 256MB = 1\text{ page}`

Diagram:

- 25 bits
- 25 bits
- 14 bits

Announcements

How to improve your grade?

*Submit a course evaluation and drop lowest in-class lab score*

- To receive credit, Submit before Monday, May 11th
Announcements
Lord of the Cache Games Night was great!
Announcements
Lord of the Cache Games Night was great!

• Winner: Team *xyzzy*

Andrew Matsumoto and Ian Leeming
Announcements
Lord of the Cache Games Night was great!

- Winner: Team *xyzzy*

Andrew Matsumoto and Ian Leeming
Announcements
Lord of the Cache Games Night was great!

- Champion of Champions: 2015 vs 2011

Big Picture about the Future
“Sometimes it is the people that no one imagines anything of who do the things that no one can imagine”

--quote from the movie The Imitation Game
“Can machines think?”

-- Alan Turing, 1950

Computing Machinery and Intelligence
The Bombe
used by the Allies to
break the German
Enigma machine during
World War II
Turing Machine
1936

Alan Turing
Big Picture

How a processor works? How a computer is organized?
What’s next?

More of Moore
Moore’s Law

Moore’s Law introduced in 1965

- Number of transistors that can be integrated on a single die would double every 18 to 24 months (i.e., grow exponentially with time).

Amazingly visionary

- 2300 transistors, 1 MHz clock (Intel 4004) - 1971
- 16 Million transistors (Ultra Sparc III)
- 42 Million transistors, 2 GHz clock (Intel Xeon) – 2001
- 55 Million transistors, 3 GHz, 130nm technology, 250mm2 die (Intel Pentium 4) – 2004
- 290+ Million transistors, 3 GHz (Intel Core 2 Duo) – 2007
- 731 Million transistors, 2-3Ghz (Intel Nehalem) – 2009
- 1.4 Billion transistors, 2-3Ghz (Intel Ivy Bridge) – 2012
Curve shows 'Moore's Law': transistor count doubling every two years.
Why Multicore?

Moore’s law

• A law about transistors
• Smaller means more transistors per die
• And smaller means faster too

But: Power consumption growing too…
What to do with all these transistors?

Multi-core
Multi-core

The first transistor
- on a workbench at AT&T Bell Labs in 1947
- Bardeen, Brattain, and Shockley

An Intel Westmere
- 1.17 billion transistors
- 240 square millimeters
- 32 nanometer: transistor gate width
- Six processing cores
- Release date: January 2010

http://www.theregister.co.uk/2010/02/03/intel_westmere_ep_preview/
The first transistor
  • on a workbench at AT&T Bell Labs in 1947
  • Bardeen, Brattain, and Shockley

• An Intel Ivy Bridge
  – 1.4 billion transistors
  – 160 square millimeters
  – 22 nanometer: transistor gate width
  – Up to eight processing cores
  – Release date: April 2012
What to do with all these transistors?

Cloud Computing
Cloud Computing

The promise of the Cloud

- ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.

NIST Cloud Definition
Cloud Computing

The promise of the Cloud

- ubiquitous, convenient, on-demand access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.

NIST Cloud Definition
Cloud Computing

The promise of the Cloud

- ubiquitous, convenient, on-demand access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.  

Requires fundamentals in systems

- Computation
- Networking
- Storage
Cloud Computing

Large organizations considering using the cloud

- New York Times
- Netflix
- Nintendo
- Cornell
- Library of Congress

The more data you have, the harder it is to move

- Switching providers entails paying for bandwidth twice
- Inhibits opportunistic migration
Cloud Computing

How hard is to program with a Exabyte of data?

Titan tech boom, randy katz, 2008
Cloud Computing

Datacenters are becoming a commodity
Order online and have it delivered

• Datacenter in a box: already set up with commodity hardware & software (Intel, Linux, petabyte of storage)

• Plug data, power & cooling and turn on
  – typically connected via optical fiber
  such datacenters
Cloud Computing = Network of Datacenters
Cloud Computing

• How to optimize a global network of data centers?
Cloud Computing = Network of Datacenters
Cloud Computing

Vision

The promise of the Cloud

• A computer utility; a commodity
• Catalyst for technology economy
• Revolutionizing for health care, financial systems, scientific research, and society

However, cloud platforms today

• Entail significant risk: vendor lock-in vs control
• Entail inefficient processes: energy vs performance
• Entail poor communication: fiber optics vs COTS endpoint
Example: Energy and Performance

Why don’t we save more energy in the cloud?

No one deletes data anymore!
  • Huge amounts of seldom-accessed data

Data deluge
  • Google (YouTube, Picasa, Gmail, Docs), Facebook, Flickr
  • 100 GB per second is faster than hard disk capacity growth!
  • Max amount of data accessible at one time $\ll$ Total data

New scalable approach needed to store this data
  • Energy footprint proportional to number of HDDs is not sustainable
What to do with all these transistors?

Embedded Processors
Where is the Market?

- **Embedded Desktop Servers**
  - 1998: 290
  - 1999: 488
  - 2000: 892
  - 2001: 862
  - 2002: 1122

- Millions of Computers
Where is the Market?

- **Embedded**
- **Desktop Servers**

<table>
<thead>
<tr>
<th>Year</th>
<th>Embedded</th>
<th>Desktop Servers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1998</td>
<td>290</td>
<td>93</td>
</tr>
<tr>
<td>1999</td>
<td>488</td>
<td>114</td>
</tr>
<tr>
<td>2000</td>
<td>892</td>
<td>135</td>
</tr>
<tr>
<td>2001</td>
<td>862</td>
<td>129</td>
</tr>
<tr>
<td>2002</td>
<td>1122</td>
<td>131</td>
</tr>
</tbody>
</table>

Millions of Computers
Where is the Market?

- **1998**
  - Embedded: 290
  - Desktop: 93
  - Servers: 3

- **1999**
  - Embedded: 488
  - Desktop: 114
  - Servers: 3

- **2000**
  - Embedded: 892
  - Desktop: 135
  - Servers: 4

- **2001**
  - Embedded: 862
  - Desktop: 129
  - Servers: 4

- **2002**
  - Embedded: 1122
  - Desktop: 131
  - Servers: 5

(Note: Numbers represent millions of computers.)
Where is the Market?

- Cell Phones
- PCs
- TVs

Millions of Computers

<table>
<thead>
<tr>
<th>Year</th>
<th>Cell Phones</th>
<th>PCs</th>
<th>TVs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1997</td>
<td>11093</td>
<td>135</td>
<td>114</td>
</tr>
<tr>
<td>1999</td>
<td>295</td>
<td>135</td>
<td>114</td>
</tr>
<tr>
<td>2001</td>
<td>405</td>
<td>136</td>
<td></td>
</tr>
<tr>
<td>2003</td>
<td>502</td>
<td>136</td>
<td></td>
</tr>
<tr>
<td>2005</td>
<td>785</td>
<td>202</td>
<td>202</td>
</tr>
<tr>
<td>2007</td>
<td>1182</td>
<td>265</td>
<td>200</td>
</tr>
</tbody>
</table>
Where to?

Smart Dust…
Security?

Cryptography and security...

TPM 1.2

IBM 4758

Secure Cryptoprocessor
Security?

Stack Smashing...

Before

- buffer[1024]
- ret address of CalcAverage()
- ...
- rest of the stack
- ...

After

- "Success ;)
- nothing meaningful here
- address of printf
- return address of main()
- address of buffer[0]
- ...
- rest of the stack
- ...
What’s next?
Moore’s Law

Moore’s Law introduced in 1965

- Number of transistors that can be integrated on a single die would double every 18 to 24 months (i.e., grow exponentially with time)

Amazingly visionary

- 2300 transistors, 1 MHz clock (Intel 4004) - 1971
- 16 Million transistors (Ultra Sparc III)
- 42 Million transistors, 2 GHz clock (Intel Xeon) – 2001
- 55 Million transistors, 3 GHz, 130nm technology, 250mm2 die (Intel Pentium 4) – 2004
- 290+ Million transistors, 3 GHz (Intel Core 2 Duo) – 2007
- 731 Million transistors, 2-3Ghz (Intel Nehalem) – 2009
- 1.4 Billion transistors, 2-3Ghz (Intel Ivy Bridge) – 2012
Moore’s Law

![Graph showing Moore’s Law over time with data points for 1962, 1965, and 1970. The x-axis represents the number of transistors per integrated circuit, and the y-axis represents the number of transistors per chip.]

Number of Components per Integrated Circuit

Relative Manufacturing Cost per Component
Parallelism

Dennard scaling: power

Must exploit parallelism for performance

MIMD: multiple instruction, multiple data
  • Multicore

SIMD: single instruction, multiple data
  • GPUs
My slide from 2008

Do you believe?

I WANT TO BELIEVE

© Kavita Bala 2008 Computer Science, Cornell University
Is Moore’s law dead?
Some thoughts

Bob Colwell
Chief Architect Pentium
DARPA

Introduction
Bill Dally, Nvidia CTO

Talk
The Chip Design Game at the End of Moore's Law
Hot Chips, Aug 2013
Singularity

Approximate Computing

Better interfaces
  Brain interfaces

Specialized chips
  Make it programmable

More
Supercomputers

Petaflops: GPUs/multicore/100s-1000s cores
Japan and the rest of the world are faced with various problems that are hard to solve. The challenge for us to tackle is how to solve these issues promptly without further delay. To do this, we need to gather wisdom from around the world and accelerate our cutting-edge research in a variety of fields. Supercomputers will be crucial in achieving these goals. Fujitsu is striving to enable a prosperous future for the Earth and its peoples through the development of supercomputers.

One Fujitsu aim is to complete the development of the K computer by 2012 together with RIKEN, in accordance with the High
Petaflops

Tianhe-2 is the fastest computer in the world! It is a 33.86 petaflop supercomputer
GPUs for Scientific Computing

- Medical Imaging, U of Utah (146X)
- Molecular Dynamics, U of Illinois (36X)
- Video Transcoding, Elemental Tech (18X)
- Matlab Computing, AccelerEyes (50X)
- Astrophysics, RIKEN (100X)
- Financial simulation, Oxford (149X)
- Linear Algebra, Universidad Jaime (47X)
- 3D Ultrasound, Techniscan (20X)
- Quantum Chemistry, U of Illinois (130X)
- Gene Sequencing, U of Maryland (30X)
GPUs for Neural Nets
GPUs for Graphics, of course
What to do with all these transistors?

You could save the world one day?
Alan Turing’s Bombe
Used to crack Germany’s enigma machine

ENIAC - 1946
First general purpose electronic computer. Designed to calculate ballistic trajectories
Games

Graphics

Scientific Computing

Save the world?

Embedded Computing

Security

Cloud Computing

Cryptography

Smart Dust & Sensor Networks
Survey Questions

Are you a better computer scientist and software engineering knowing “the low-level stuff”? 

How much of computer architecture do software engineers actually have to deal with? 

What are the most important aspects of computer architecture that a software engineer should keep in mind while programming?
Why?
These days, programs run on hardware...
... more than ever before

Google Chrome
→ Operating Systems
→ Multi-Core & Hyper-Threading
→ Datapath Pipelines, Caches, MMUs, I/O & DMA
→ Busses, Logic, & State machines
→ Gates
→ Transistors
→ Silicon
→ Electrons
Where to?

CS 3110: Better concurrent programming

**CS 4410/4411: The Operating System!**

CS 4420/ECE 4750: Computer Architecture

CS 4450: Networking

CS 4620: Graphics

MEng

5412—Cloud Computing,  5414—Distr Computing
5430—Systems Secuirty, 5413 – high perf systems and netowrking
5300—Arch of Larg scale Info Systems
6644 – Modeling the world

And many more...
Why?
Your job as a computer scientist will require knowledge about the computer.

Research/University
Cornell University
Faculty of Computing and Information Science

Industry
IBM
Google
Amazon

Government
Department of Energy
Department of Defense
National Security Agency

United States of America
Thank you!

If you want to make an apple pie from scratch, you must first create the universe.

– Carl Sagan