CS 61C:
Great Ideas in Computer Architecture
Overview and Berkeley Culture

Instructor:
David A. Patterson
http://inst.eecs.Berkeley.edu/~cs61c/sp12

Agenda

- Administrivia
- Survey Results
- Extra Credit Top Scores
- Course Overview
- Cal Culture
- HKN Course Evaluation

Administrivia

- All grades but Project 4 finalized: 4/27; Proj 4 5/1
- Go to lab to finalize any labs Th/Fr
- Final Review: Sun April 29, 2-5PM, 2050 VLSB
- Extra office hours: Thu-Fri May 3-4 1-5PM
- Final Exam: Wed May 9 11:30-2:30, 1 Pimentel
  - Designed for 90 minutes, you will have 3 hours
  - Comprehensive (particularly problem areas on midterm), but focused on course since midterm: lecture, lab, hws, and projects are fair game
  - 8 ½ inch x 11 inch crib sheet like midterm

Some Survey Results

- I felt the midterm was

3% Far too difficult
10% Somewhat harder than it should have been
64% Fair
20% A little too easy
2% Far too easy

Some Survey Results

- How much time per week to you spend on average in 61C (including lecture, discussion, and labs)?

18% <10 hours per week
35% 11-12 hours per week
16% 13-14 hours per week
20% 15-16 hours per week
8% 17-20 hours per week
4% >21 hours per week

Berkeley guidelines 3 hours/unit

(EECS classes often > guidelines)

Some Survey Results

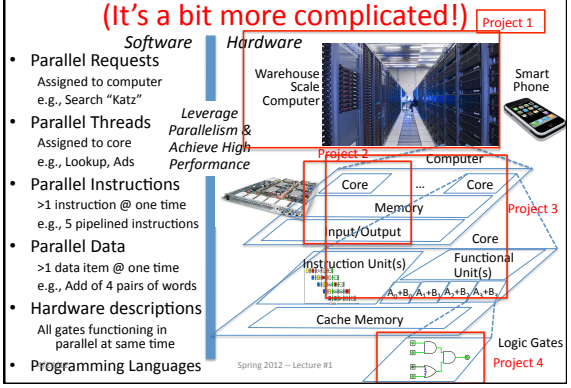
- Rated as "Enjoyed and learned a lot":

Project#4: Processor Design in Logisim (69%)
Logisim Labs: ALU (66%), Logisim (51%)
C memory management lab (42%)
Project#3: Matrix Multiply Performance Improvement, Part 1 (41%) & Part 2 (31%)

Some Survey Results

- Did 61C material come up in interviews for internships or jobs? 51% Yes
  - "MapReduce was a hit, memory management/stack/heap questions came up, optimization questions (cache blocking, etc.) came up"
  - "I was simply asked how I would change one of my responses for a system that had multiple cores."
  - "Interview with VMWare - Asked about experience writing parallel programs using Hadoop. Interview with Intel - Asked about using SSE intrinsics for software optimization"

New-School Machine Structures (It's a bit more complicated!)



6 Great Ideas in Computer Architecture

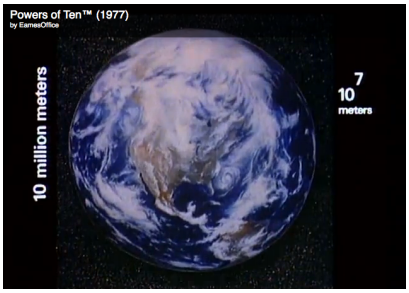
1. Layers of Representation/Interpretation
2. Moore's Law
3. Principle of Locality/Memory Hierarchy
4. Parallelism
5. Performance Measurement & Improvement
6. Dependability via Redundancy

Powers of Ten inspired 61C Overview

- Going Top Down cover 3 Views
1. Architecture (when possible)
2. Physical Implementation of that architecture
3. Programming system for that architecture and implementation (when possible)
- http://www.powersof10.com/film

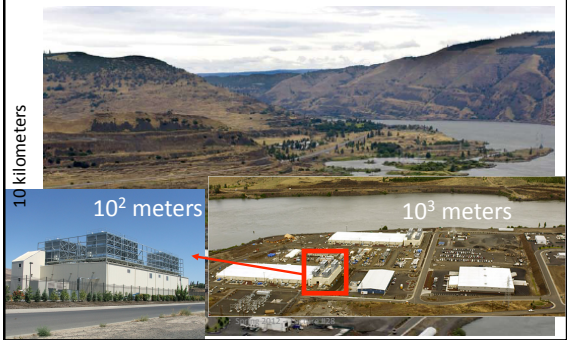
Earth

10^7 meters



Google's Oregon WSC

10^4 meters



Google Warehouse

- 90 meters by 75 meters, 10 Megawatts
- Contains 40,000 servers, 190,000 disks
- Power Utilization Effectiveness: 1.23
  - 85% of 0.23 overhead goes to cooling losses
  - 15% of 0.23 overhead goes to power losses
- Contains 45, 40-foot long containers
  - 8 feet x 9.5 feet x 40 feet
- 30 stacked as double layer, 15 as single layer

Containers in WSCs

10^2 meters

100 meters

Google Container

10^1 meters

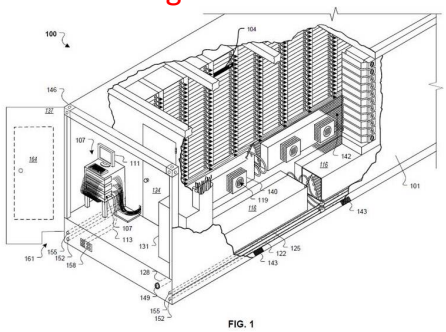
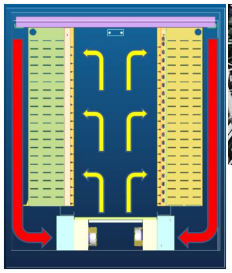


FIG. 1

Google Container

10^0 meters

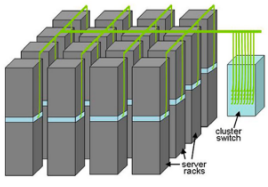
10 meters



- 2 long rows, each with 29 racks
- Cooling below raised floor
- Hot air returned behind racks

Equipment Inside a Container

Server (in rack format):



7 foot Rack: servers + Ethernet local area network switch in middle ("rack switch")

Array (aka cluster): server racks + larger local area network switch ("array switch") 10X faster => cost 100X: cost f (N^2)

10^0 meters

Google Rack

1 meter

- Google rack with 20 servers + Network Switch in the middle
- 48-port 1 Gigabit/sec Ethernet switch every other rack
- Array switches connect to racks via multiple 1 Gbit/s links
- 2 datacenter routers connect to array switches over 10 Gbit/s links

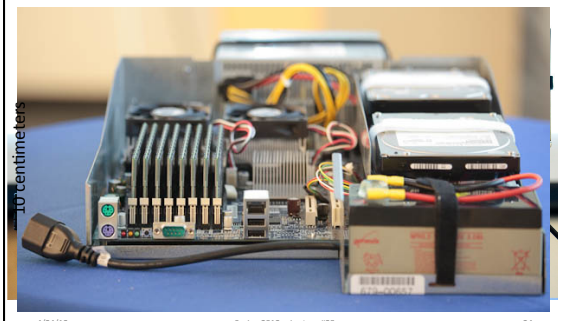
Programming WSC: MapReduce

```
public static class SumReduce extends Reducer<Text, LongWritable, Text,
   LongWritable> {
/** Actual reduce function.
* @param key Word.
*  @param values Values for this word (partial counts).
* @param context ReducerContext object for accessing output,
*            configuration information, etc.
*/
@Override
public void reduce(Text key, Iterable<LongWritable>values,
     Context context) throws IOException, InterruptedException {
   long sum = 0L;
 for (LongWritable value : values) {sum += value.get();}
    context.write(key, new LongWritable(sum));
  }
}
```


6 Great Ideas in Computer Architecture inside the Warehouse Scale Computer

1. Layers of Representation/Interpretation
   - WSC, Container, Rack
2. Moore's Law
3. Principle of Locality/Memory Hierarchy
4. Parallelism
   - Task Level Parallelism, Data Level Parallelism
5. Performance Measurement & Improvement
   - Measure PUE to improve PUE
6. Dependability via Redundancy
   - Multiple WSCs, Multiple Racks, Multiple Switches

Google Server Internals 10^{-1} meters



Google Board Details

- Supplies only 12 volts
- Battery per board vs. large battery room
  - Improves PUE: 99.99% efficient local battery vs 94% for battery room
- 2 SATA Disk Drives
  - 1 Terabyte capacity each
  - 3.5 inch disk drive
  - 7200 RPM
- 2 AMD Opteron Microprocessors
  - Dual Core, 2.2 GHz
- 8 DIMMs
  - 8 GB DDR2 DRAM
- 1 Gbit/sec Ethernet Network Interface Card

Programming Multicore Microprocessor: OpenMP

```
#include <omp.h>
#include <stdio.h>
static long num_steps = 100000;
int value[num_steps];
int reduce()
{    int i;      int sum = 0;
#pragma omp parallel for private(x) reduction(+:sum)
     for (i=1; i<= num_steps; i++){
        sum = sum + value[i];
     }
}
```


6 Great Ideas in Computer Architecture inside the Server

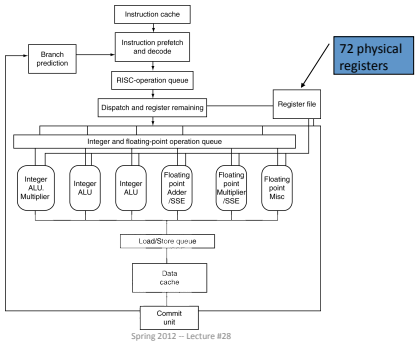
1. Layers of Representation/Interpretation
2. Moore's Law
   - More transistors => Multicore
3. Principle of Locality/Memory Hierarchy
4. Parallelism
   - Thread Level Parallelism
5. Performance Measurement & Improvement
   - Timers, counters
6. Dependability via Redundancy

10^{-2} meters

centimeters

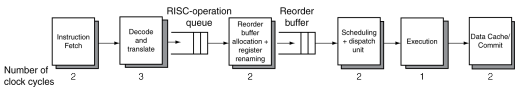
AMD Opteron Microprocessor

AMD Opteron Microarchitecture



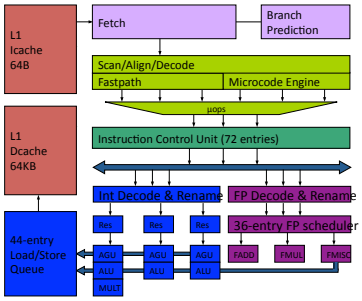
AMD Opteron Pipeline Flow

- For integer operations



- 12 stages (Floating Point is 17 stages)
- Up to 106 RISC-ops in progress

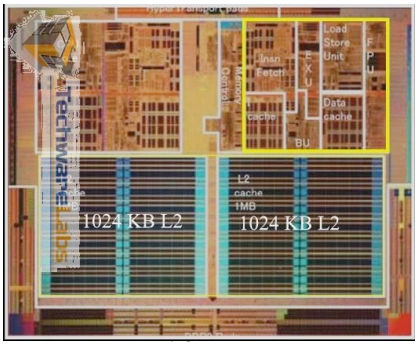
AMD Opteron Block Diagram



10^{-2} meters

centimeters

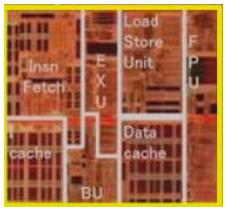
AMD Opteron Microprocessor



10^{-3} meters

millimeters

AMD Opteron Core



Programming One Core: C with Intrinsics

```
void mmult(int n, float *A, float *B, float *C)
{
  for ( int i = 0; i < n; i+=4 )
    for ( int j = 0; j < n; j++ )
    {
      __m128 c0 = _mm_load_ps(C+i+j*n);
      for( int k = 0; k < n; k++ )
        c0 = _mm_add_ps(c0, _mm_mul_ps(_mm_load_ps(A+i+k*n),
                                       _mm_load1_ps(B+k+j*n)));
      _mm_store_ps(C+i+j*n, c0);
    }
}
```

Inner loop from gcc –O -S

Assembly snippet from innermost loop:

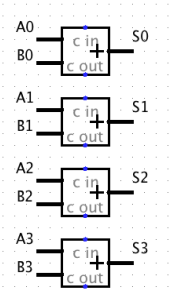
```
movaps  (%rax), %xmm9
mulps   %xmm0, %xmm9
addps   %xmm9, %xmm8
movaps  16(%rax), %xmm9
mulps   %xmm0, %xmm9
addps   %xmm9, %xmm7
movaps  32(%rax), %xmm9
mulps   %xmm0, %xmm9
addps   %xmm9, %xmm6
movaps  48(%rax), %xmm9
mulps   %xmm0, %xmm9
addps   %xmm9, %xmm5
```

6 Great Ideas in Computer Architecture inside the Microprocessor

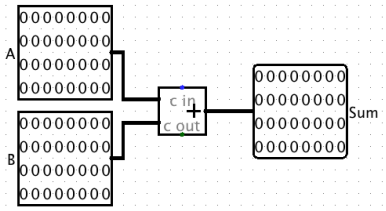
1. Layers of Representation/Interpretation
   - Instruction Set Architecture, micro operations
2. Moore's Law
3. Principle of Locality/Memory Hierarchy
4. Parallelism
   - Instruction Level Parallelism (superscalar, pipelining)
   - Data Level Parallelism
5. Performance Measurement & Improvement
6. Dependability via Redundancy

SIMD Adder

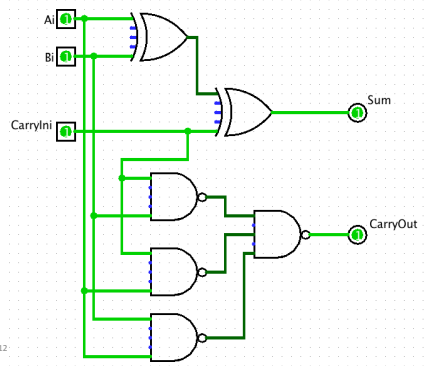
- Four 32-bit adders that operate in parallel
  - Data Level Parallelism



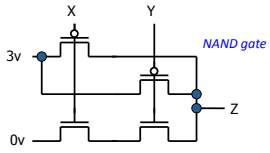
One 32-bit Adder



1 bit of 32-bit Adder



Complementary MOS Transistors (NMOS and PMOS) of NAND Gate

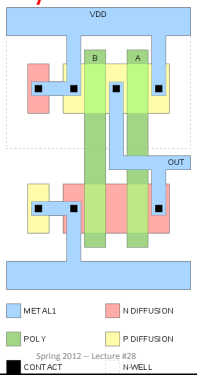


x	y	z
0 volts	0 volts	3 volts
0 volts	3 volts	3 volts
3 volts	0 volts	3 volts
3 volts	3 volts	0 volts

10^{-7} meters

Physical Layout of NAND Gate

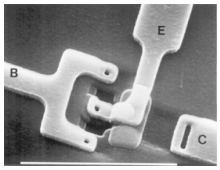
100 nanometers



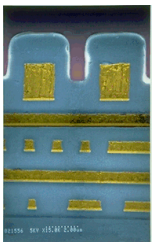
10^{-7} meters

Scanning Electron Microscope

100 nanometers



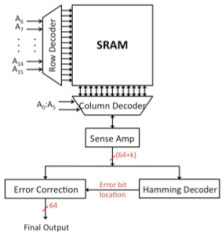
Top View



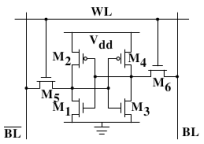
Cross Section

10^{-6} meters

Block Diagram of Static RAM



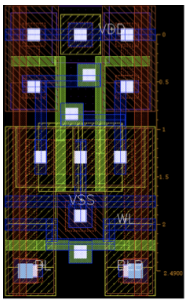
1 Bit SRAM in 6 Transistors



10^{-7} meters

Physical Layout of SRAM Bit

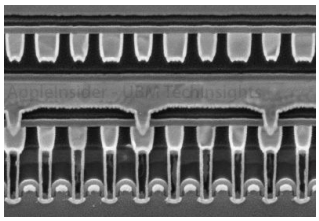
100 nanometers



10^{-7} meters

SRAM Cross Section

100 nanometers



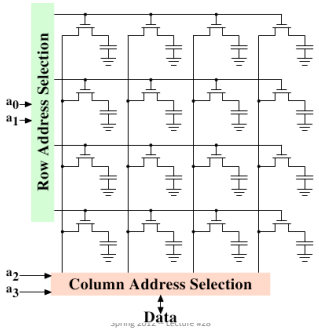
DIMM Module

- DDR = Double Data Rate
  - Transfers bits on Falling AND Rising Clock Edge
- Has Single Error Correcting, Double Error Detecting Redundancy (SEC/DED)
  - 72 bits to store 64 bits of data
  - Uses "Chip kill" organization so that if single DRAM chip fails can still detect failure
- Average server has 22,000 correctable errors and 1 uncorrectable error per year

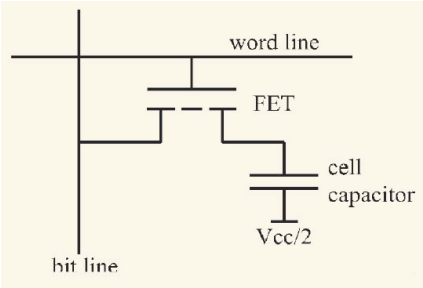
10^{-6} meters

DRAM Bits

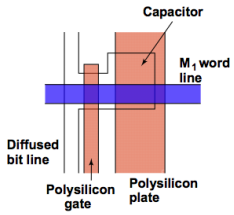
1 micron



DRAM Cell in Transistors



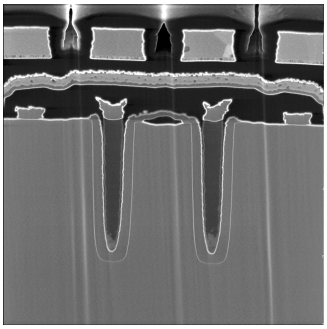
Physical Layout of DRAM Bit



10^{-7} meters

Cross Section of DRAM Bits

100 nanometers



AMD Dependability

- L1 cache data is SEC/DED protected
- L2 cache and tags are SEC/DED protected
- DRAM is SEC/DED protected with chipkill
- On-chip and off-chip ECC protected arrays include autonomous, background hardware scrubbers
- Remaining arrays are parity protected
  - Instruction cache, tags and TLBs
  - Data tags and TLBs
  - Generally read only data that can be recovered from lower levels

Programming Memory Hierarchy: Cache Blocked Algorithm

- The blocked version of the i-j-k algorithm is written simply as (A,B,C are submatricies of a, b, c)

```
for (i=0;i<N/r;i++)
  for (j=0;j<N/r;j++)
    for (k=0;k<N/r;k++)
      C[i][j] += A[i][k]*B[k][j]
```

  - r = block (sub-matrix) size (Assume r divides N)
  - X[i][j] = a sub-matrix of X, defined by block row i and block column j

6 Great Ideas in Computer Architecture inside the chips

1. Layers of Representation/Interpretation
2. Moore's Law
   - Higher capacity caches and DRAM
3. Principle of Locality/Memory Hierarchy
   - Caches, TLBs
4. Parallelism
   - Data Level Parallelism
5. Performance Measurement & Improvement
   - Memory Traffic, Cache Misses
6. Dependability via Redundancy
   - Parity, SEC/DED

Course Summary

- As the field changes, cs61c had to change too!
- It is still about the software-hardware interface
  - Programming for performance!
  - Parallelism: Task-, Thread-, Instruction-, and Data-MapReduce, OpenMP, C, SSE instrinsics
  - Understanding the memory hierarchy and its impact on application performance
- Interviewers ask what you did this semester!

Agenda

- Administrivia
- Survey Results
- Extra Credit Top Scores
- Course Overview
- Cal Culture
- HKN Course Evaluation

What to Emphasize about CS at Cal?

- *US News and World Report Rankings 2012*
  - *Source: http://grad-schools.usnews.rankingsandreviews.com/best-graduate-schools/top-science-schools/computer-science-rankings*
- Top Graduate Programs in Computer Science
  - 4 way tie for #1: Berkeley, CMU, MIT, Stanford
- 4 specialties: AI, Prog. Languages, Systems, and Theory
- Specialty Systems in Computer Science
  1: Berkeley
  2: MIT
  3: CMU
  4: Stanford

Doesn't Reputation Lag Reality?

- 2007 Intel and Microsoft announced they would fund a Universal Parallel Computing Research Center (UPCRC): $10M over 5 years
- Invited the 25 top Computer Science departments to submit proposals
- 4 finalists site visits:
  Berkeley, Illinois, MIT, Stanford
- August 2007 Intel and Microsoft announce technical committee unanimously selected Berkeley as the top choice of the competition

What to Emphasize about Cal overall?

- Top public university (US News and World Report)
- Top graduate program in the world?
  - 35/36 programs in top 10 in 1994
  - 48/52 programs in top 10 in 2010
- University doing the most public good
  - *Washington Monthly* 2009!
- Faculty Awards?
  - 9 current Nobel Prize winners (22 all time)!
  - 3 Turing Award winners ("Nobel of Computer Science")!
  - 3 Fields Medalists ("Nobel of Math")!
  - 32 "Genius" awards (MacArthur fellows)
  - 91 in National Academy of Engineering!
  - 137 in National Academy of Science!

Source: http://www.berkeley.edu/about/honors/

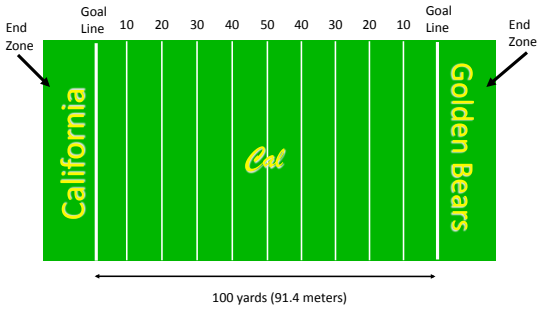
Cal Cultural History: Football!

- Started with "soccer" (aka football)
  - 11 on a team, 2 teams, 1 ball, on a field; object is to move ball into "goal"; most goals wins. No hands!
- New World changes rules to increase scoring:
  - Make goal bigger! (full width of field)
  - Carry ball with hands
  - Can toss ball to another player backwards or laterally (called a "lateral") anytime and forwards ("pass") sometimes
- How to stop players carrying the ball? Grab them & knock them down by making knee hit the ground ("tackle")
  - In soccer tackle the ball; football tackle the person

ABCs of American Football

- Score by...
  - Moving football into goal ("cross the goal line" or "into the end zone") scoring a "touchdown"
    - (6 points)
  - Kicking football between 2 poles ("goal posts") scoring a "field goal"
    - (worth 3 points, unless after touchdown, then its just 1 point: "extra point")
- Kick ball to other team after score ("kickoff")
  - laterals OK
- Game ends when no time left (four 15 min quarters) and person with ball is stopped
  - Soccer: two 45 min halves, time stops play

Football Field



Spectacle of American Football

- Cal's archrival is Stanford
  - Stereotype is Rich, Elitist Snobs
  - E.g, derby Man City vs. Manchester United
- Play nearby archrival for last game of season
  - Called "The Big Game": Cal vs. Stanford, winner gets a trophy ("The Axe")
  - Oldest rivalry west of Mississippi; 100th in 1997
- American college football is a spectacle
  - School colors (Cal **Blue** & **Gold** v. **Red** & **White**)
  - Nicknames (Golden Bears v. Stanford Cardinal)
  - School mascot (Oski the bear v. a tree(!))
  - Leaders of cheers ("cheerleaders")

Spectacle of American Football

- "Bands" (orchestras that march) from both schools at games
- March & Play
  - Before game, at halftime, after game
- Stanford Band more like a drinking club (seen the movie "Animal House"?)
  - Plays one song: "All Right Now" (1970)
  - Cannot march and play

1982 Big Game: "The Play"

"Top 20 favorite sports event in 20th century," Sports Illustrated

"Greatest Football Play of All Time," Best Damn Sports Show

"…The Play, widely considered the most dramatic ending in college football history" , AP news

"…widely considered the most famous play in college football history ," Stanford Magazine

("The Play" in Wikipedia en.wikipedia.org/wiki/The_Play)

- Stanford
  - Quarterback is John Elway, who goes on to be a professional Hall of Fame football player (retired 1999)
  - Possibly greatest quarterback in college history?
    - In 1982, they had lost 4 games in last minutes
- Stanford has just taken lead with 4 seconds left in game; Cal team captain yells in huddle "Don't fall with the ball!"; watch video

Notes About "The Play"

- "Allright here we go with the kick-off. Harmon will probably try to squib it and he does. Ball comes loose and the Bears have to get out of bounds. Rogers along the sideline, another one... they're still in deep trouble at midfield, they tried to do a couple of....the ball is still loose as they get it to Rogers. They get it back to the 30, they're down to the 20...Oh the band is out on the field!! He's gonna go into the endzone!!! He got into the endzone!! … THE BEARS HAVE WON!!! THE BEARS HAVE WON!!! Oh my God, the most amazing, sensational, dramatic, heart rending... exciting thrilling finish in the history of college football!" – KGO's Joe Starkey

Notes About "The Play" (1/2)

- Cal only had 10 men on the field; last second another came on (170 pound Steve Dunn #3) and makes key 1st block
- Kevin Moen #26: 6'1" 190 lb. safety,
  - laterals to Rodgers (and doesn't give up)
- Richard Rodgers #5: 6' 200 lb. safety, Cal captain "Don't fall with that ball."
  - laterals to Garner
- Dwight Garner #43: 5'9" 185 lb. running back
  - almost tackled, 2 legs & 1 arm pinned, laterals
- Richard Rodgers #5 (again): "Give me the ball!"
  - laterals to Ford

Notes About "The Play" (1/2)

- Mariet Ford #1: 5'9", 165 pound wide receiver
  - Smallest player, leg cramps; overhead blind lateral to Moen and blocks 3 Stanford players
- Moen (again) cuts through Stanford band into end zone (touchdown!), smashes Trombonist
- On field for Stanford: 22 football players, 3 Axe committee members, 3 cheerleaders, 144 Stanford band members (172 for Stanford v. 11 for Cal)
  - "Weakest part of the Stanford defense was the woodwinds." -- Cal Fan
- Cal players + Stanford Trombonist (Gary Tyrrell) hold reunions; Stanford revises history (Changes score on Axe to 20-19); claims Garner's knee was down – see video

Special Thanks to the TAs:
Rimas Avizienis,
Scott Beamer,
Alan Christopher,
Eric Liang,
Paul Ruan,
Ian Vonseggern

The Future for Future Cal Alumni

- What's The Future?
- New Century, Many New Opportunties: Parallelism, Cloud, Statistics + CS, Bio + CS, Society (Health Care, 3rd world) + CS
- Cal heritage as future alumni
  - Hard Working / Can do attitude
  - Never Give Up ("Don't fall with the ball!")
  - Smallest on field, 3 big guys charging you: you make a play!
- "The best way to predict the future is to invent it" – Alan Kay (inventor of personal computing vision)
- Future is up to you!