

Lecture 23:
Goodbye to Computer Architecture,
Future Predictions,
and Your Cal Cultural Heritage

Professor David A. Patterson
Computer Science 252
Spring 1998

Final Lecture

- **Review and Goodbye to Computer Architecture, topic by topic + follow-on courses**
- **Final Administritivia, include slide total**
- **Future Directions for Computer Archtitecture?**
- **Learning about your heritage as Cal students/ future alumni**
- **Course evaluation by HKN**
- **Drinks at LaVal's**

Chapter 1: Performance and Cost

- Amdahl's Law:

$$\text{Speedup}_{\text{overall}} = \frac{\text{ExTime}_{\text{old}}}{\text{ExTime}_{\text{new}}} = \frac{1}{(1 - \text{Fraction}_{\text{enhanced}}) + \frac{\text{Fraction}_{\text{enhanced}}}{\text{Speedup}_{\text{enhanced}}}}$$

- CPI Law:

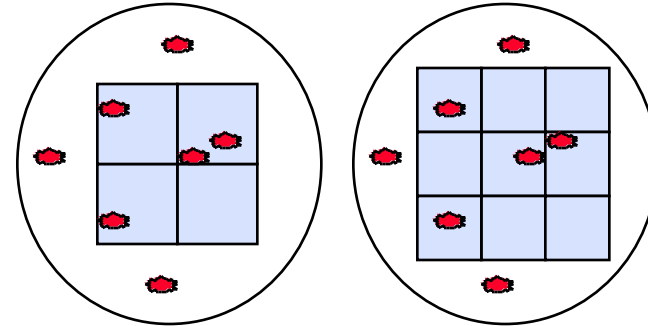
$\text{CPU time} = \frac{\text{Seconds}}{\text{Program}} = \frac{\text{Instructions}}{\text{Program}} \times \frac{\text{Cycles}}{\text{Instruction}} \times \frac{\text{Seconds}}{\text{Cycle}}$

- Designing to Last through Trends

	<u>Capacity</u>	<u>Speed</u>
Logic	2x in 3 years	2x in 3 years
DRAM	4x in 3 years	2x in 10 years
Disk	4x in 3 years	2x in 5 years
Processor 2x every 1.5 years?		

Chapter 1: Performance and Cost

- **Die Cost goes roughly with die area⁴**
 - Microprocessor with 100M transistors in 2000?



- **Cost vs. Price**
 - Can PC industry support engineering/research investment? (e.g., DEC laying off 15,000)
- **For better or worse, benchmarks shape a field**
- **Interested in learning more on integrated circuits?
EE 241 “Advanced Digital Integrated Circuits”**
- **Interested in learning more on performance?
CS 266 “Introduction to Systems Performance”**

Goodbye to Performance and Cost

- **Will sustain 2X every 1.5 years?**
 - Can integrated circuits improve below 1.8 micron in speed as well as capacity?
- **5-6 yrs to PhD =>**
16X CPU speed, DRAM Capacity, Disk capacity?
(1500 MHz CPU, 1GB DRAM, 100 GB disk?)

Chapter 2: Instruction Set Architecture

- **What ISA looks like to pipeline?**
 - Cray: load/store machine; registers; simple instr. format
- **RISC: Making an ISA that supports pipelined execution**
- **80x86: importance of being their first**
- **VLIW/EPIC: compiler controls Instruction Level Parallelism (ILP)**
- **Interested in learning more on compilers and ISA?
CS 264/5 “Advanced Programming Language Design and Optimization”**

Goodbye to Instruction Set Architecture

- **What did IA-64/EPIC do well besides floating point programs?**
- **What happened on EPIC code size vs. x86?**
- **Was Intel Oregon increase x86 performance so as to make Intel Santa Clara EPIC performance similar?**
- **Did reconfigurable processors (e.g., BRASS, RAW) prove useful? On what class of applications?**

Chapters 3/4: Pipelined Implementation

- **Miracle of Pipelining: Bandwidth vs. latency**
- **Superscalar breaks single instruction/clock cycle limit**
 - Hazards/Dependencies limit: HW & SW techniques to overcome limits
 - Conditional Branches as one Limit: branch prediction
 - Memory system as another limit
- **SW Pipelining: Symbolic Loop Unrolling to get most from pipeline with little code expansion, little overhead**
- **Scoreboard: Allow instructions behind stall to proceed**
- **Out-of-order execution: Helps cache misses as well**
- **Reservations stations: renaming to larger set of registers + buffering source operands**
 - Prevents registers as bottleneck
 - Avoids WAR, WAW hazards of Scoreboard
 - Beyond basic block

Goodbye to Pipelined Implementation

- Did wider superscalar, more out-of-order machines work well, or were they beyond the point of diminishing returns?
- What about more exotic ideas?
 - **Value prediction**: predict the next value of a variable (e.g., loop counter) to get by dependencies?
 - **Simultaneous Multithreading**: since getting little benefit from most programs of wide superscalar, out-of-order machines, schedule multiple threads to get most of hardware

Appendix B: Vector Processors

- **Vector is alternative model for exploiting ILP**
- **Accommodates long memory latency, doesn't rely on caches as does Out-Of-Order, superscalar/VLIW designs**
- **If code is vectorizable, then simpler hardware, more energy efficient, and better real-time model than Out-of-order machines**
- **Design issues include number of lanes, number of functional units, number of vector registers, length of vector registers, exception handling, conditional operations**
- **What % of computation is vectorizable? What % do compilers deliver? For new apps?**

DSP architectures

- **Continuous I/O stream, real time requirements**
- **Multiple memory accesses**
- **Datapath: Multiply width, Wide accumulator, Guard bits/shifting rounding, Saturation**
- **Autoinc/autodec addressing**
- **Weird things: Circular & Reverse addressing**
- **Special instructions**
 - **shift left and saturate (arithmetic left-shift)**
 - **zero overhead loops**

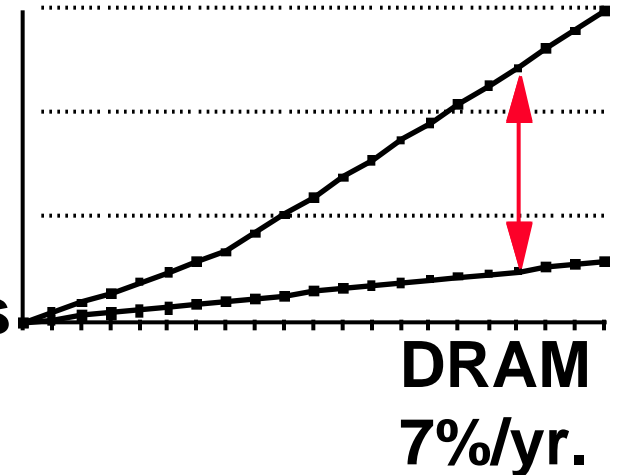
Goodbye to Vectors, DSPs

- **Multimedia instructions (Intel MMX, HP MMX, SPARC VIS, Motorola AltiVec) represent a resurgence of vector-like instructions: where they hype, or did they really help performance of multimedia apps?**
- **Did vector prove to be a better match to new apps such as multimedia & DSP, programming in HLL?**
- **Did DSPs survive distinct from microprocessors?**

Chapter 5: Memory Hierarchy

MPU
60%/yr.

- Processor-DRAM Performance gap
- 1/3 to 2/3 die area for caches, TLB
- Alpha 21264: 108 clock to memory
648 instruction issues during miss
- 3 Cs: Compulsory, Capacity, Conflict
- 4 Questions: where, who, which, write
- Applied recursively to create multilevel caches
- Performance = $f(\text{hit time, miss rate, miss penalty})$
 - danger of concentrating on just one when evaluating performance
- Integration of Processors into Memory, into Disks?
CS 294-2 (Patterson) Fall 1998, Control No.: 25160



Cache Optimization Summary

$$CPUtime = IC \times CPI_{Execution} + \frac{Memory\ accesses}{Instruction} \times Miss\ rate \times Miss\ penalty \times Clock\ cycle\ time$$

	Technique	MR	MP	HT	Complexity
miss rate	Larger Block Size	+	-		0
	Higher Associativity	+		-	1
	Victim Caches	+			2
	Pseudo-Associative Caches	+			2
	HW Prefetching of Instr/Data	+			2
	Compiler Controlled Prefetching	+			3
	Compiler Reduce Misses	+			0
miss penalty	Priority to Read Misses		+		1
	Subblock Placement		+	+	1
	Early Restart & Critical Word 1st		+		2
	Non-Blocking Caches		+		3
	Second Level Caches		+		2
hit time	Small & Simple Caches	-		+	0
	Avoiding Address Translation			+	2
	Pipelining Writes			+	1

memory hierarchy art: taste in selecting between alternatives to find combination that fits well together

Goodbye to Memory Hierarchy

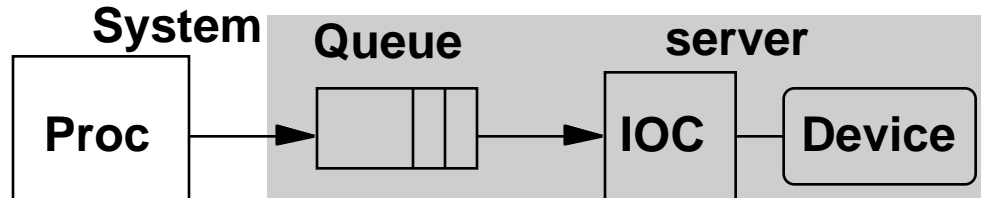
- Will L2 cache keep growing? (e.g, 64 MB L2 cache?)
- Will multilevel hierarchy get deeper? (e.g, l4 cache?)
- Will DRAM capacity/chip keep going at 4X / 3 years? (e.g., 16 Gbit chip?)
- Will processor and DRAM/Disk be unified?
For which apps?
- Out-of-order CPU hides L1 data cache miss (3–5 clocks), but hide L2 miss? (>100 clocks)
- Memory hierarchy likely overriding issue in algorithm performance: do algorithms and data structures of 1960s work with machines of 2000s?

CS 252 Administrivia

- **Projects by 5PM Mon. May 11; send email when done**
- **Many, many interesting projects**
- **Several students and faculty said they enjoyed poster session and mentioned what great jobs you did**
 - Former CS252 TA, now a professor, remarked at how much better
- **You have seen the full conference cycle: topic selection, investigation, real deadlines, oral presentation, poster session, written presentation**
- **Many capable of being turned into published papers**
- **Hope you noticed that feedback was important for both your ideas and your presentation of ideas, and the benefit of presenting preliminary results before the final deadline**
- **252 lectures slides on line: 1062 more slides than pages of textbook!**

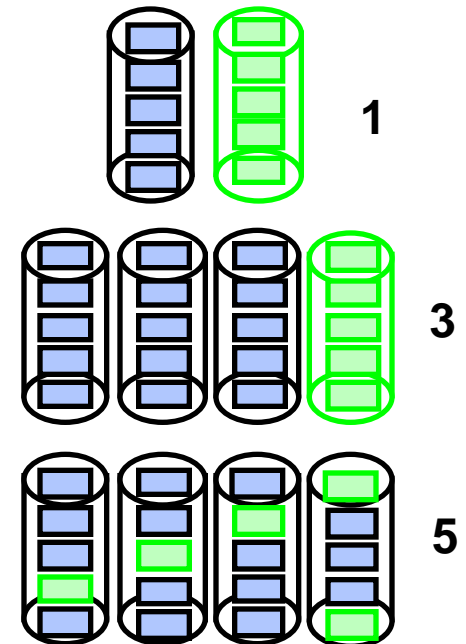
Chapter 6: Storage I/O

- Disk BW 40%/yr, areal density 60%/ yr, \$/MB faster?
- Little's Law: $Length_{system} = rate \times Time_{system}$
(Mean number customers = arrival rate x mean service time)



- Throughput vs. response time
- Value of faster response time on productivity

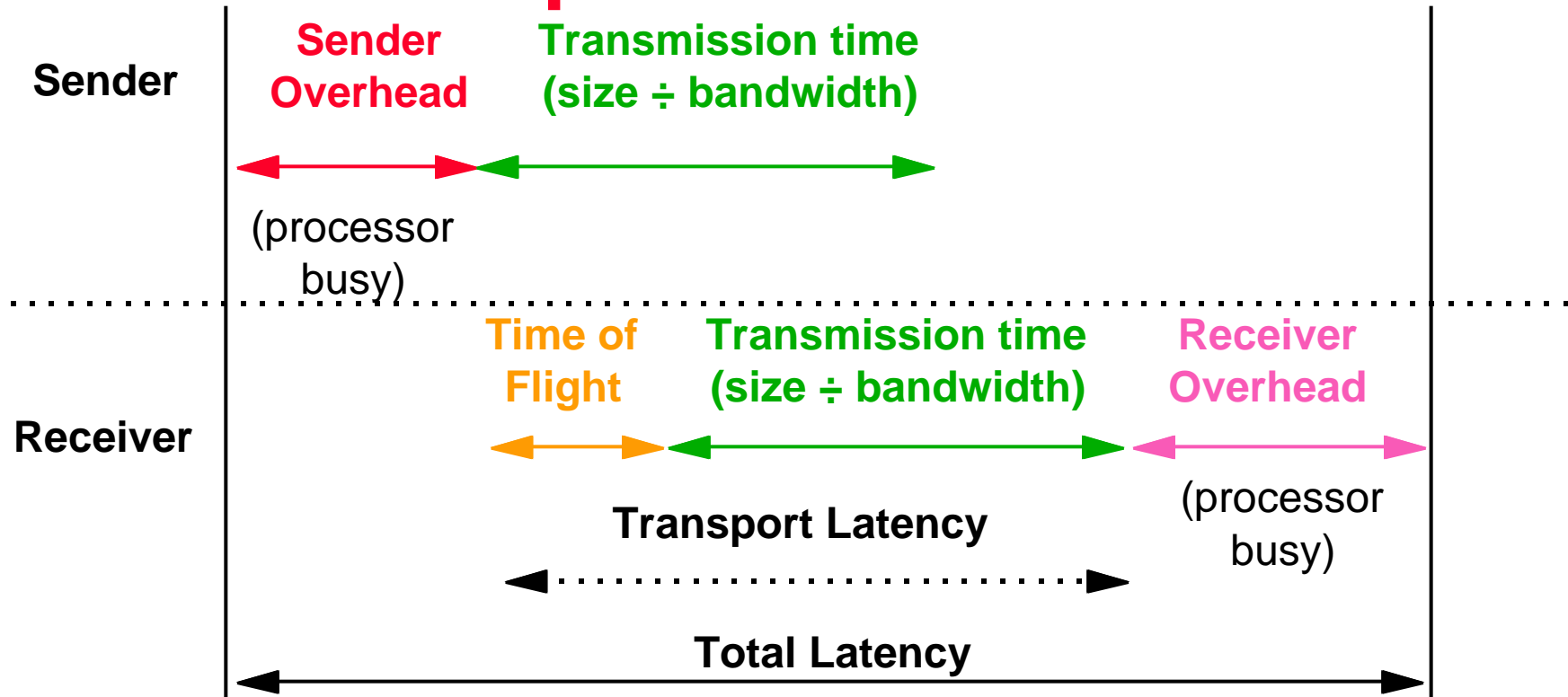
- Benchmarks: scaling, cost, auditing, response time limits
- RAID: performance and reliability
- Queueing theory? IEOR 161, 267, 268
- SW storage systems? CS 286
“Implementation of Data Base Systems”



Goodbye to Storage I/O

- **Disks growing at 4X/ 3 years more recently:
Will I continue get email messages to reduce file storage for the rest of my career?**
- **Heading towards a personal terabyte:
hierarchical file systems vs. database to organize personal storage?**
- **Disks attached directly to networks, avoiding the file server? (“Network Attached Storage Devices”)**
- **What going to do when can have video record of entire life on line?**

Chapter 7: Networks



$$\text{Total Latency} = \text{Sender Overhead} + \text{Time of Flight} + \text{Message Size} \div \text{BW} + \text{Receiver Overhead}$$

High BW networks + high overheads violate of Amdahl's Law

Chapter 7: Networks

- Similarities of MPP interconnects, LANs, WANs
- Integrated circuit revolutionizing networks as well as processors
- Switch is a specialized computer
- Protocols allow heterogeneous networking , handle normal **and** abnormal events
- Interested in learning more on networks?
EE 122 “Introduction to Computer Networks” (McCanne)
CS 268 “Computer Networks” (McCanne)

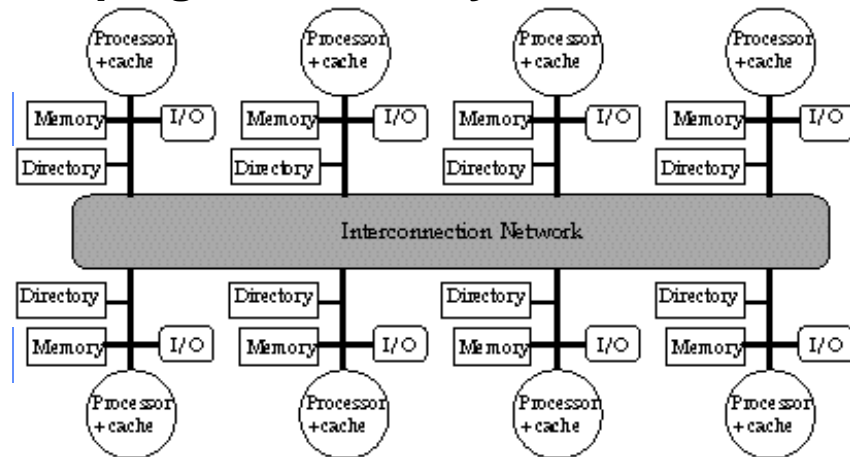
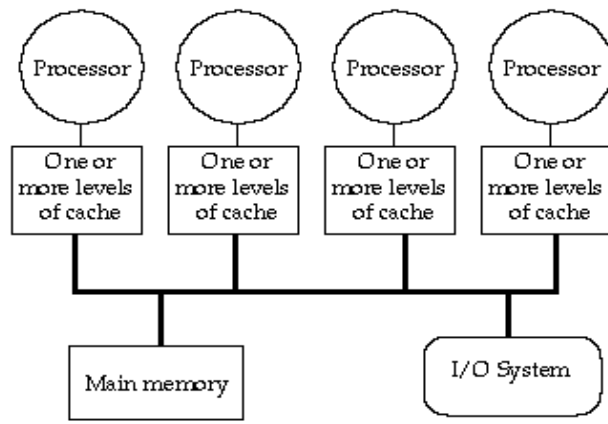
Goodbye to Networks

- **Will network interfaces follow example of graphics interfaces and become first class citizens in microprocessors, thereby avoiding the I/O bus?**
- **Will Ethernet standard keep winning the LAN wars? e.g., 100 Mbit/sec, 1 Gbit/sec, 10 Gbit/sec, ... ?**
- **Who will win the WAN wars long term: telephony vs TCP/IP bigots?**

Chapter 8: Multiprocessors



- **Shared, uniform memory access vs. Shared non-uniform memory access vs. Message Passing**
 - Cache coherency protocols: Snooping vs. directory



- Interested in learning more on multiprocessors:
CS 258 “Parallel Computer Architecture” (Spring 99)
E 267 “Programming Parallel Computers”

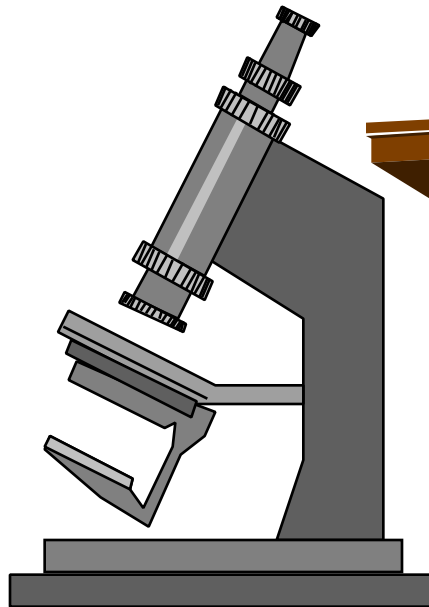
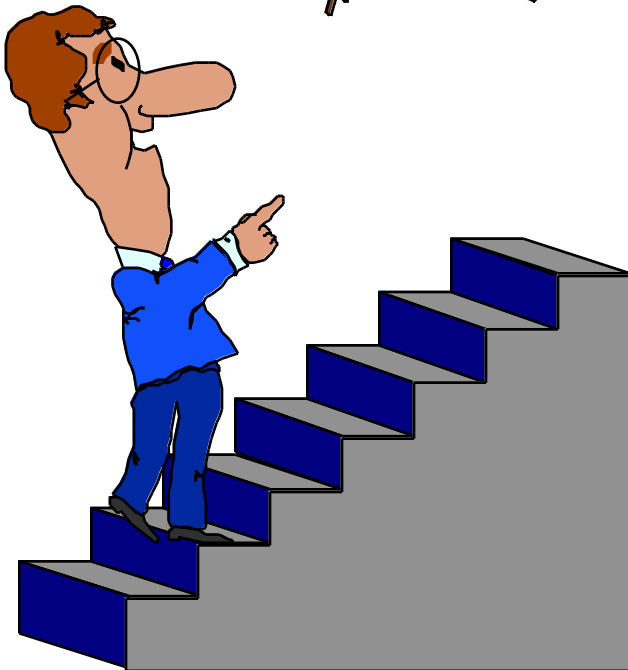
Goodbye to Multiprocessors

- **Successful today for file servers, time sharing, databases, graphics; will parallel programming become standard for production programs?
If so, what enabled it: new programming languages, new data structures, new hardware, new courses, ...?**
- **Which won large scale number crunching, databases: Clusters of independent computers connected via switched LAN vs. large shared NUMA machines?
Why?**

How to be a Success in Graduate School

- **1) “Swim or Sink”**
 - “Success is determined by me (student) primarily”
 - Faculty will set up the opportunity, but its up to me leverage it
- **2) “Read/learn on your own”**
 - “Related to 1), I think you told me this as you handed me a stack of about 20 papers”
- **3) “Teach your advisor”**
 - “I really liked this concept; go out and learn about something and then teach the professor”
 - Fast moving field, don’t expect prof to be at forefront everywhere

Role Changes during Project



Alternatives to a Bad Career

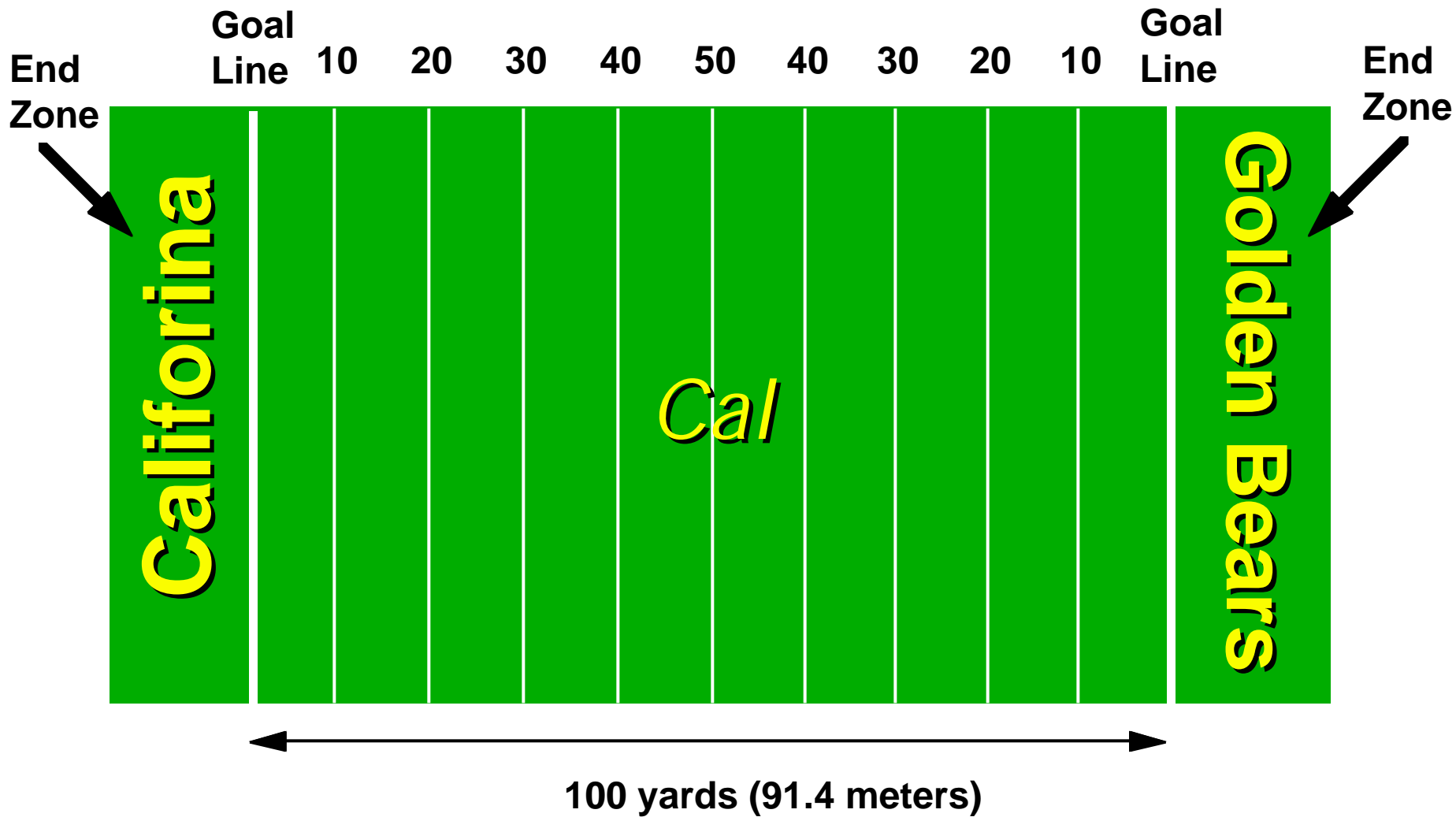
- **Goal is to have impact:**
Change way people do Computer Science & Engineering
 - Evaluation of academic research uses bad benchmarks
=> skews academic behavior
- **Many 3 - 5 year projects gives more chances for impact**
- **Feedback is key: seek out & value critics**
- **Do “Real Stuff”:** make sure you are solving some problem that someone cares about
- **Taste is critical in selecting research problems, solutions, experiments, & communicating results; taste is acquired and improved by feedback**
- *Students* are the coin of the academic realm

Future Directions in Computer Architecture

- **1000X performance increase in “stationary” computers, consolidation of industry
=> time for architecture/OS/compiler researchers declare victory, search for new horizons?**
- **Apps/metrics of future to design computer of future!**
- **Mobile Multimedia (PDA, wearable) offer many new challenges: energy efficiency, size, real time performance**
 - PDA of future + VIRAM-1 one example, hope others will follow
- **3D Telepresense: being there digitally (and virtually) will be as good as being there physically**
 - “be there” at the opening ceremonies for the next Olympics: parade around the track with the Olympians and join the final torch runner for her dash up the stairs to light the Olympic torch!

Cal Cultural History: ABCs of American Football

- Started with “soccer”; still 11 on a team, 2 teams, 1 ball, on a field; object is to move ball into “[goal](#)”; most goals wins
- New World changes the 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 & forwards (“[pass](#)”) sometimes
- How to stop players carrying the ball? Grab them & knock them down by making knee hit the ground (“[tackle](#)”)
 - if drop ball (“[fumble](#)”), other players can pick it up and score
- Score by moving ball into goal (“[cross the goal line](#)” or “[into the end zone](#)”) scoring a “[touchdown](#)” (6 points), or kicking ball between 2 poles (“[goal posts](#)”) scoring a “[field goal](#)” (3, unless after touchdown = 1: “[extra point](#)”)
- Kick ball to other team after score (“[kickoff](#)”); laterals OK
- Game ends when no time left (4 15 min quarters) & person with ball is stopped (Soccer time only: 2 45 min halves, time stops play)



The Spectacle of Football

- **Rose Bowl**: Prestigious bonus game played January 1 if have a great year (“playoffs”)
 - preceded by parade
 - national TV coverage
- **1929 Rose Bowl Game**
 - Cal vs. Georgia Tech
 - Cal going left to right (\Rightarrow),
GeorgiaTech right to left (\Leftarrow)
 - Georgia Tech player fumbles football
 - Cal player, Roy Reigel, picks up football
and tries to avoid Georgia Tech players
- **Let’s see what happens on video**

- Play nearby archrival for last game of season
- Cal's archrival is Stanford; stereotype is Private, Elitist, Snobs
- **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**; Stanford: **Red** & **White**)
 - School nicknames (Cal: Golden Bear; Stanford: Cardinal)
 - School mascot (Cal: Oski the bear; Stanford: a tree(!))
 - Leaders of cheers ("cheerleaders")
- "Bands" (orchestras that march) from both schools at games; before game, at halftime, after game
 - Stanford Band more like a drinking club; "Animal House"
 - Plays one song: "All Right Now"
 - Stanford used to yell "boring" at band during Cal's performance

1982 Big Game

- “There has never been anything in the history of college football to equal it for sheer madness.” *Sports Illustrated*
- Cal coach is Joe Kapp, former Cal player; tells team to play 100% for 60 minutes (“40 for 60”; “Bear will not die”); 1st year as coach; lasts 5 years (“Never give up”)
- Stanford coach is Paul Wiggin, former Stanford player, lots of coaching experience; fired from job next year
- Stanford Quarterback is John Elway, who goes on to be a professional All Star football player (still playing today, won 1st Superbowl in 1998)
- Cal Quarterback is Gail Gilbert, who goes on to be a non-starting professional football player (stopped playing 1996)
- Stanford lost 4 games in last few minutes of game
- Let’s see what happens on video

- **Cal only had 10 men on the field; last second another came on (170 pound Steve Dunn #3) & makes key 1st block**
- **Kevin Moen #26: 6'1" 190 lb. safety, never scored in 4 years at Cal**
 - laterals to Rodgers (and doesn't give up)
- **Richard Rodgers #5: 6' 200 lb. safety, "Don't fall with the ball."**
 - laterals to Garner
- **Dwight Garner #43: 5'9" 185 pound running back**
 - almost tackled, 2 legs & 1 arm pinned, laterals to Rodgers
- **Richard Rodgers #5 (again): "Give me the ball, Dwight."**
 - laterals to Ford
- **Mariet Ford #1: 5'9", 165 pound wide receiver**
 - leg cramps, overhead blind lateral to Moen & blocks 3 players
- **Moen (again) cuts through Stanford band into end zone**
- **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."**
- **4 Cal players + Stanford Trombonist (Gary Tyrrell) hold reunion every year at Big Game; Stanford revises history (20-19 on Axe)**

Your Cal Cultural History

- **Cal students/alumni heritage is the greatest college football play in > 100 years**
- **Cal students/alumni work hard and play hard**
- **Cal students/alumni handle adversity**
- **Cal students/alumni never give up!**
- **Cal students/alumni triumph over great odds!**