

Wayne State University
Dept of Electrical & Computer Engineering
Brammer Lecture Series, 3 Oct. 2007

Towards Speech Recognition in Silicon: The Carnegie Mellon *In Silico* Vox Project

Rob A. Rutenbar

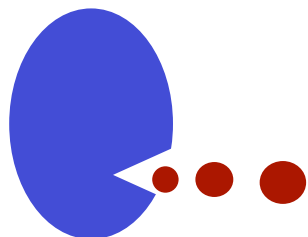
Professor, Electrical & Computer Engineering

rutenbar@ece.cmu.edu

Speech Recognition Today: *Software*

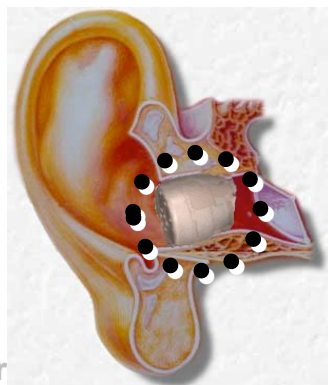


- **Quality = OK** **Vocabulary = *large***



- **Quality = *poor*** **Vocab = *small***

- ▼ The Toshiba UT103, 4 languages, ~3000 phrases, 35 hours on 2AA batteries, runs on 75MHz Toshiba processor



- ***No way...***

Today's Best Software Speech Recognizers

- **Best-quality recognition is computationally *hard***
 - ▼ For speaker-independent, large-vocabulary, continuous speech

- **1-10-100-1000 rule**
 - ▼ For **~1X** real-time recognition rate
 - ▼ For **~10%** word error rate (90% accuracy)
 - ▼ Need **~100 MB** memory footprint
 - ▼ Need **~100 W** power
 - ▼ Need **~1000 MHz** CPU

- **But, this is ~1000X away from what we need**

About This Talk

- **Some philosophy**
 - ▼ Why silicon? Why now? Why us (CMU)?

- **A quick tour: How speech recognition works**
 - ▼ What happens in a recognizer

- **A silicon architecture**
 - ▼ Stripping away all CPU stuff we don't need, focus on essentials

- **Results**
 - ▼ Silicon version: Simulation results
 - ▼ FPGA version: Live, running hardware-based recognizer

About This Talk

- **Some philosophy**
 - ▼ Why silicon? Why now? Why us (CMU)?

- **A quick tour: How speech recognition works**
 - ▼ What happens in a recognizer

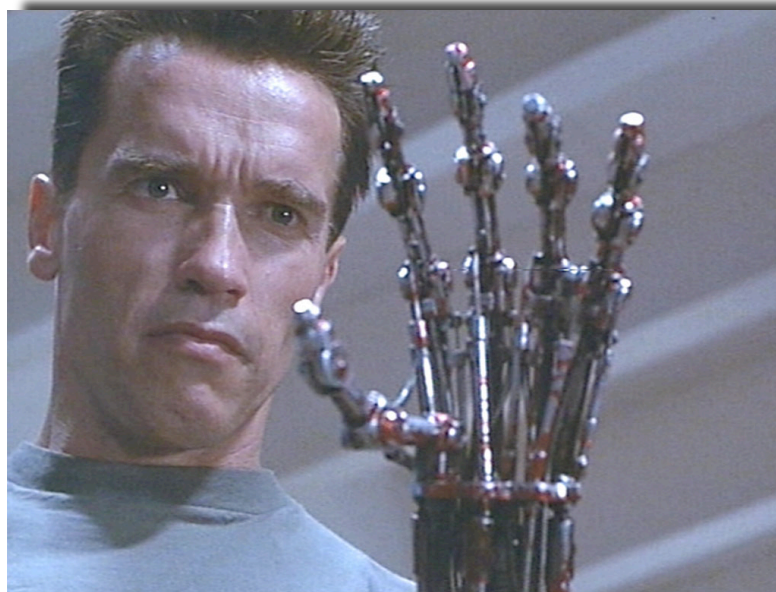
- **A silicon architecture**
 - ▼ Stripping away all CPU stuff we don't need, focus on essentials

- **Results**
 - ▼ Silicon version: Simulation results
 - ▼ FPGA version: Live, running hardware-based recognizer

Compelling Next-Generation Applications...

- ...want to go **fast**. Very fast. *Faster* than realtime.

Example: Audio Mining



Fast forward your DVD
FIND: "Hasta la vista, baby!"

Compelling Next-Generation Applications...

BBC NEWS

About the versions | Low Graphics | Help | Contact us
Last Updated: Tuesday, 28 September, 2004, 10:07 GMT 11:07 UK

E-mail this to a friend | Printable version

Backlog of terror tapes dogs FBI

The FBI has a backlog of hundreds of thousands of hours of untranslated audio recordings from possible terror suspects, a federal audit has found.

Three years after the 11 September attacks, the FBI has more than 123,000 hours of audio intercepts that it has not translated, the report said.

The report is an edited summary of a classified audit completed in July for the Justice Department.

The FBI is recruiting more linguists for Arabic, Farsi, Urdu and Pashto.

FBI director Robert Mueller says improvements are being made

3 years after 9/11, FBI still had 123,000 hrs of untranslated foreign audio

- ...could use speech→text @ **100X -1000X** realtime
- (Text→text translation also exists—diff problem)
- Could we **triage** these huge media streams to allocate scarce human intel assets?

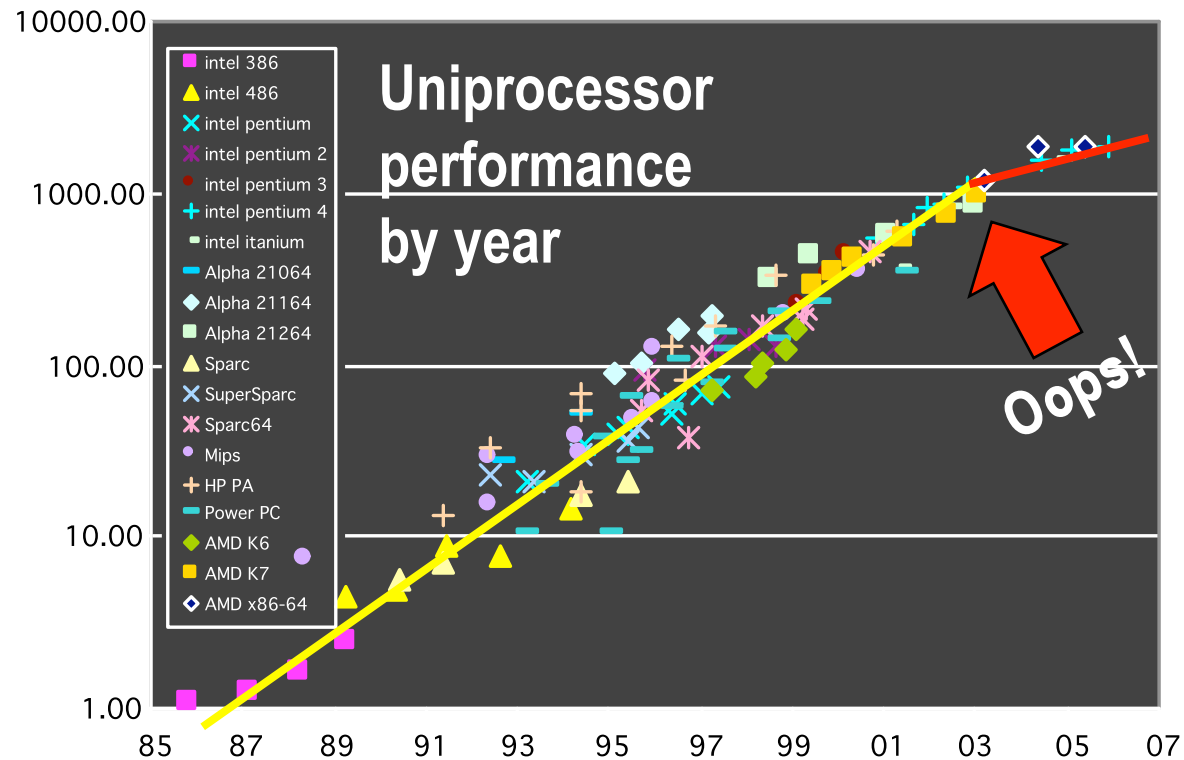
Compelling Next-Gen Applications...



- Want **low power**
- **Very low**
 - ▼ Cell phone has **3W** total power budget
 - ▼ You get **~300mW** for a new feature
- **1st- gen solns still software...**

Doesn't Moore's Law Just Save Us (*Eventually*)?

- Yes (sort of...): Transistors keep getting smaller
- No (uh oh...): Moore's Law is running out of gas

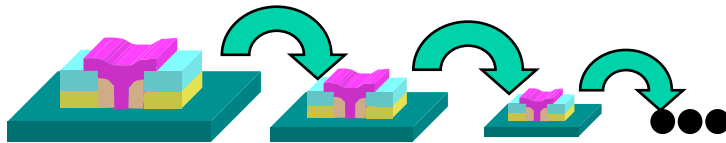


Power!

Problems with Moore's Law Scaling

■ Limits on device **size**

- ▼ Already at atomistic dimensions
- ▼ Can't scale forever when devices are already ~100 atoms wide



■ Limits on device **speed**

- ▼ Small devices leak (switches draw a little current when off)
- ▼ Power dissipation \propto clk frequency
- Can't get more performance by just upping GHz on next chip

■ Emphasis now on **design**

- ▼ More parallel architectures...
- ▼ ...with clocks running slower
- ▼ ...to get performance, but not melt

World's first quad-core processors for desktop and mainstream servers

Home · World's first quad-core processors for desktop and mainstream servers

Intel brings you the world's first quad-core processors for desktop and mainstream servers leading the industry in multi-core technology. Find out about the blistering performance of desktop or servers. See how Intel® quad-core technology delivers unprecedented performance to high-end computers.

Now introducing the Intel® Core™2 Quad processor

Intel's most advanced processor just got an upgrade to the power of four—four processing cores for the ultimate in demanding entertainment

- Up to 54% better performance for intense multimedia applications, streaming movies, and more with powerful Intel quad-core technology¹
- Up to 53% better performance when enjoying immersive 3-D gaming²
- Up to 79% faster performance for highly-threaded applications when creating multiple 3-D content³
- Up to 8MB of L2 cache and 1066 MHz Front Side Bus for an unrivaled multitasking experience


Intel® Quad-Core. Now available.

Still: Lots of Software-Based Next-Gen Work

■ Video indexing

The New York Times

SLIPSTREAM
Millions of Videos, and Now a Way to Search Inside Them



Peter DaSilva for The New York Times

Suranga Chandratillake, a co-founder of Blinkx, which offers a way to search the contents of Web videos. Already, more than 60 percent of the traffic on the Internet is video.

By JASON PONTIN
 Published: February 25, 2007

THE World Wide Web is awash in digital video, but too often we can't find the videos we want or browse for what we might like.

E-MAIL
 PRINT
 REPRINTS
 SAVE
 SHARE

■ Speech on cellphone



VoiceSignal®
 Home Company Products Gallery Customers News Support

Voice Recognition on the iPhone

VoiceSignal - the global leader in mobile voice technology. More than 150 million copies of VoiceSignal software in 21 languages have shipped embedded in devices from the world's top handset manufacturers.

VoiceMode VSpeak VSuite VSearch

VoiceMode™ 2.0 Available for Purchase! ENGLISH EDITION BUY NOW

News Partners Articles

Aug 24, 2007 - Nuance Closes Acquisition of VoiceSignal PDF www.nuance.com
 Aug 24, 2007 - VoiceSignal Voice Enables iPhone in Proof of Concept Development PDF
 Aug 22, 2007 - Trade Secret Misappropriation Case Against VoiceSignal Dismissed PDF
 more...

New Mobile Phones Supporting Voice and Speech Recognition Technology

MOTOROLA SAMSUNG NOKIA BlackBerry Treo

StreetInsider.com
 if you're not inside... you're outside

Portal Solution Study
 How Did A Large Cable Company Reduce Transfers? Read How Here!

VoiceXML Solutions
 Proven VXML Platforms & Systems Request More Information Online Now

Ads by Google

Home About StreetInsider Links FAQ Contact Us Search News Custom Search

Sun, Sep 30, 2007 12:22 PM Enter a Stock Symbol

Join Street Insider
 Member's Home
 Premium Content
 Basic Content
 My Portfolio Headlines
 Press Releases

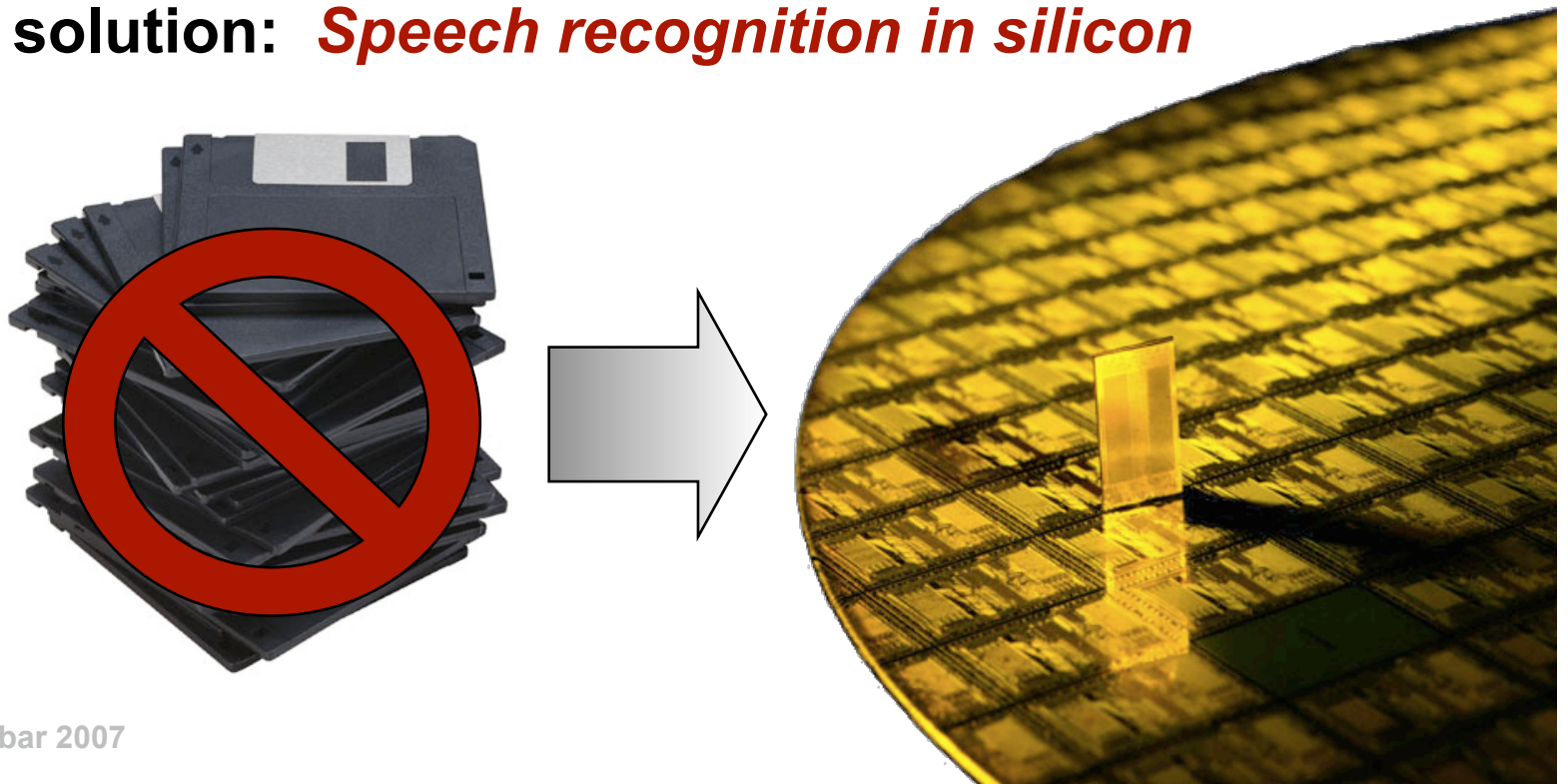
Font Size:

Basic Content
 Nuance Communications (NUAN) Acquires VoiceSignal Technologies for \$293 Million; Updates Outlook
 05-15-2007 07:45:19 AM

Market Snapshot
 28 Sep 2007 NYSE:NUAN
 2720
 2710
 2700
 2690
 1 10 11 12 1 2 3 4
 FinancialContext.com

The Carnegie Mellon *In Silico* Vox Project

- Our thesis: It's time to liberate speech recognition from the current limitations of software, because we can *always do it better in custom silicon*
- Our solution: *Speech recognition in silicon*



Aside: About the Name “*In Silico Vox*”

■ *In Vivo*

- ▼ Latin: an experiment done in a living organism.....



■ *In Vitro*

- ▼ Latin: an experiment done in an artificial lab environment.....



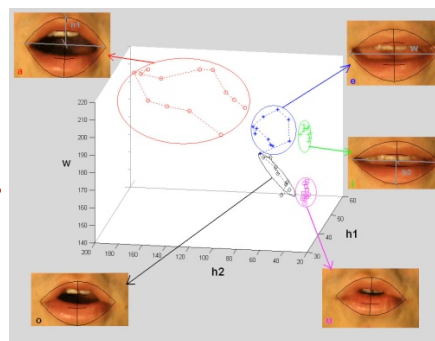
■ *In Silico*

- ▼ (Not real Latin): an experiment done via computation only.....



■ *Vox*

- ▼ Latin: voice, or word.....



Why Silicon? Why Now?

Why? Two reasons:

■ History

- ▼ We have some successful **historical** examples of this migration

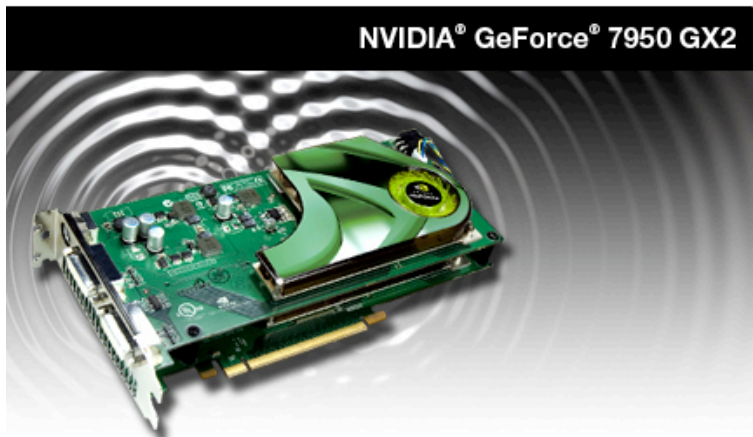
■ Performance

- ▼ Compelling apps need **100X – 1000X** more performance, **now**
- ▼ Silicon always better than software on **speed/power**

History: Graphics Engines

- **Nobody paints pixels in software anymore!**
 - ▼ Too limiting in max performance. Too inefficient in power.

True on the desktop (& laptop)



<http://www.nvidia.com>

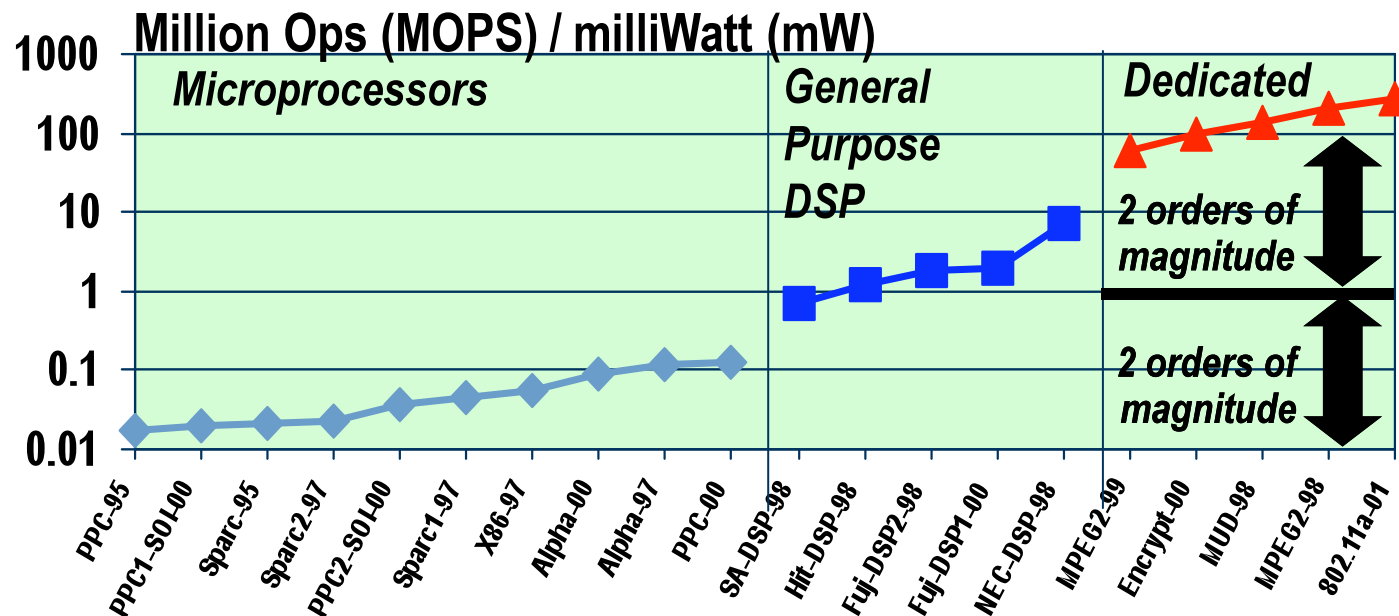
...and on your cellphone too



<http://www.mtekvision.com>

Silicon Solution: Speed *and* Power Wins

- A famous graph from Prof. Bob Brodersen of Berkeley
 - ▼ Study looked at 20 designs published at ISSCC, from 1997-2002
 - ▼ In slightly older technologies, relative to today: 180nm – 250nm
 - ▼ Dedicated designs up to **10,000X better** energy efficiency (MOPS/mW)



Silicon Speed/Power Win: Why?

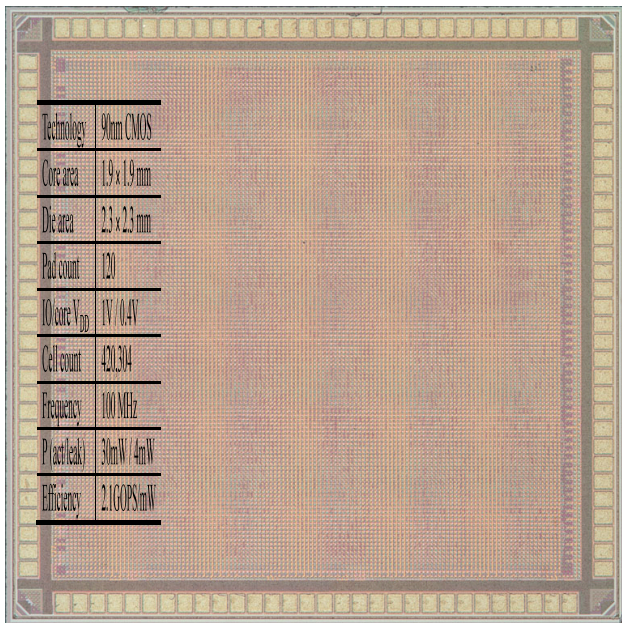
- **Programmability (flexibility) is not free**
 - ▼ Lots of extra overhead for hardware you don't need for every app
 - ▼ Baggage to fetch, decode, run instructions, one (or a few) at a time

- **Functional units not well customized to your app**
 - ▼ If you can use, say, 75 floating point units, or 36 FFT units – too bad
 - ▼ You still get 8 arithmetic units...

- **MHz/GHz to deliver speed to all users not optimal**
 - ▼ Microprocessors run fast clocks so all apps see good performance
 - ▼ Your app may be able to run a much slower clock → much less power

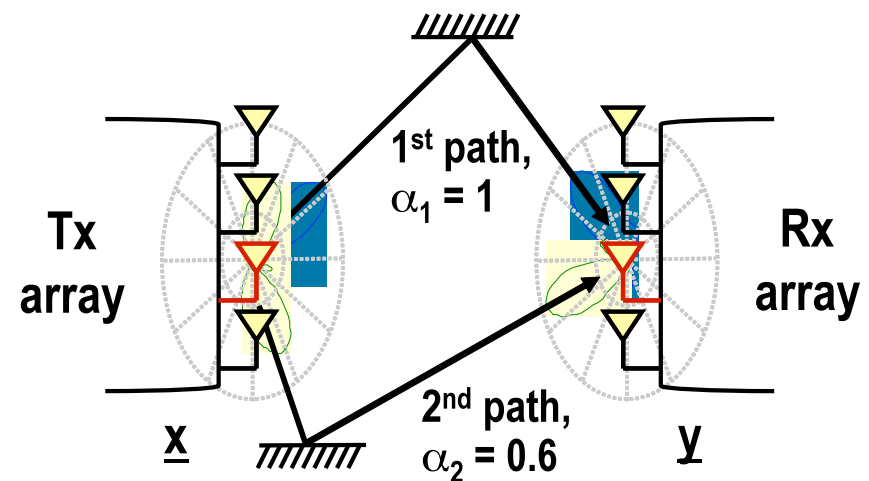
Recent Example: Parallel Radio Baseband DSP

- **90nm CMOS: adaptive DSP for multipath MIMO channel**
 - ▼ Power efficiency = **2.1GOPS/mW**
 - ▼ Area efficiency = **20GOPS/mm²**



Technology	90nm CMOS
Core area	1.9 x 1.9 mm
Die area	2.3 x 2.3 mm
Pad count	120
IO core V _{DD}	1V / 0.4V
Cell count	420,304
Frequency	100 MHz
P _{act/leak}	30mW / 4mW
Efficiency	2.1GOPS/mW

(Source: Prof. Dejan Markovitz, UCLA)



Data rate up to 250Mbps over 16 sub-carriers
Measured 34mW @ VDD=385mV

Why Us...?

- **1 site (Carnegie Mellon), 3 areas of deep expertise**
 - ▼ Impossible to do projects like this without **cross-area** linkages

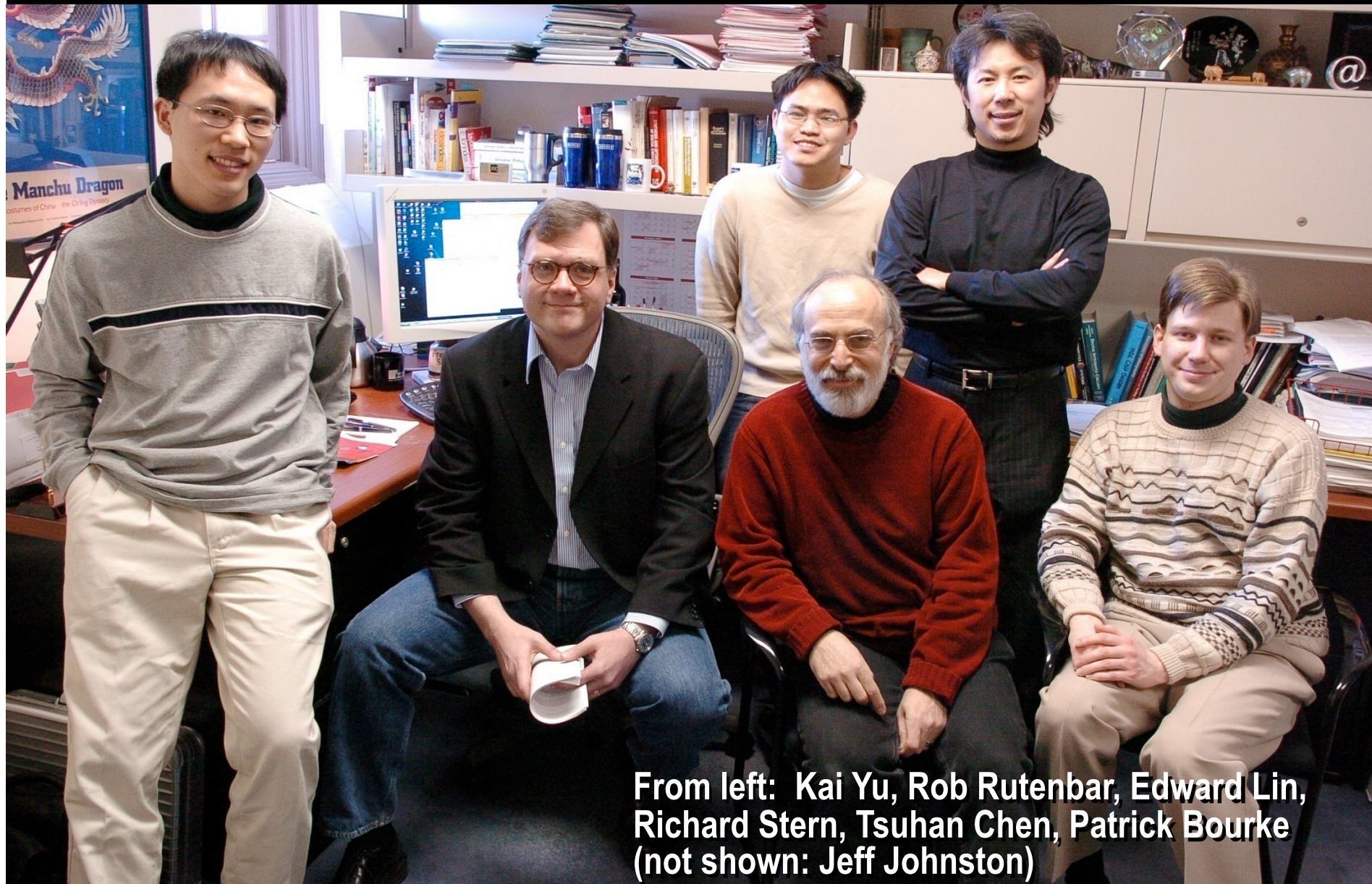


Computer Science
SPHINX Speech recognition group

Electrical & Computer Engineering
Silicon system implementation group

Electrical & Computer Engineering
Media / DSP group

Us: the CMU *In Silico* Vox Team

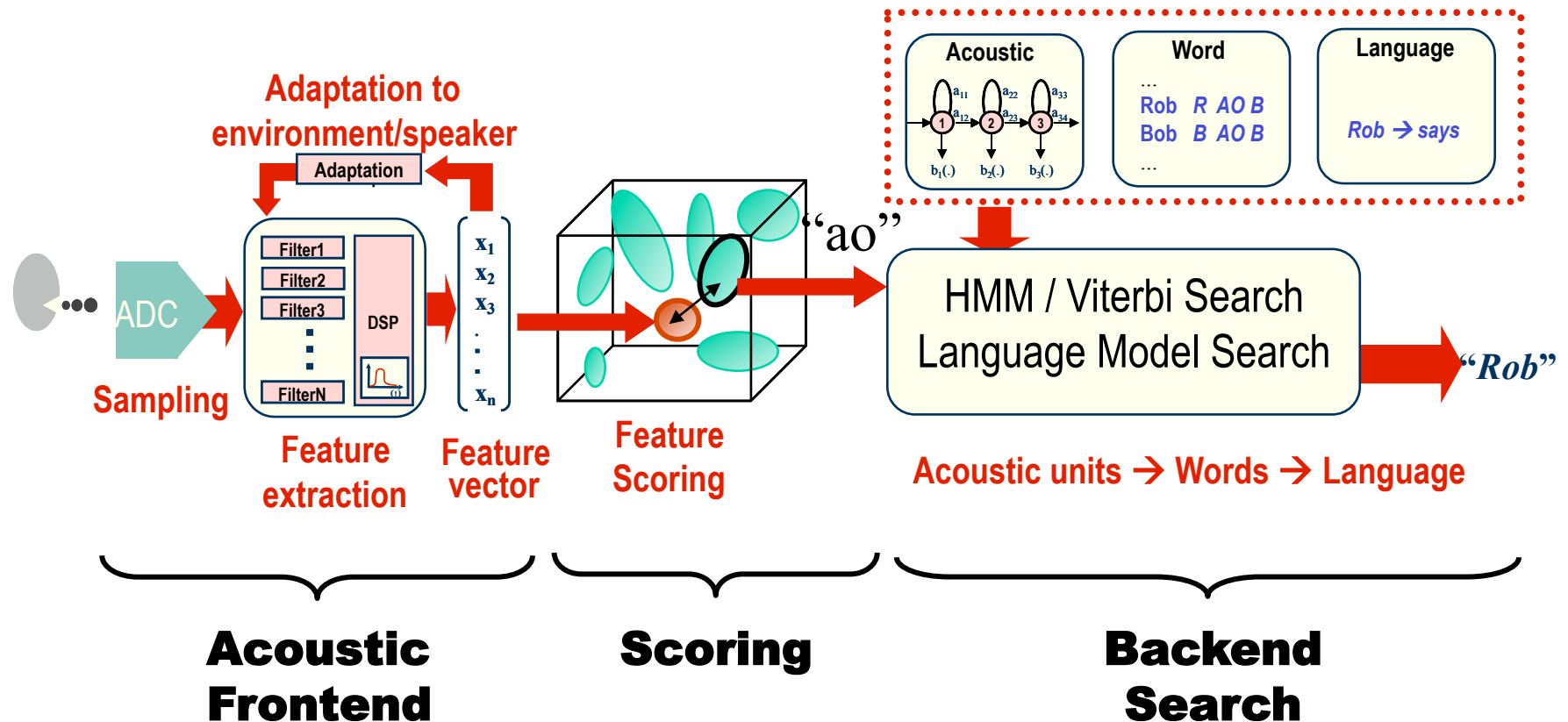


From left: Kai Yu, Rob Rutenbar, Edward Lin,
Richard Stern, Tsuhan Chen, Patrick Bourke
(not shown: Jeff Johnston)

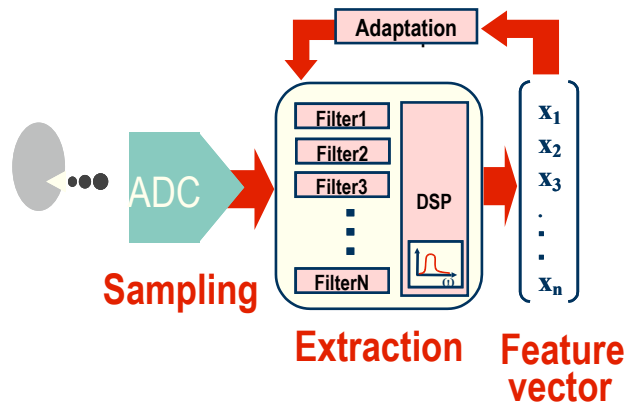
About This Talk

- **Some philosophy**
 - ▼ Why silicon? Why now? Why us (CMU)?
- **A quick tour: How speech recognition works**
 - ▼ **What happens in a recognizer**
- **A silicon architecture**
 - ▼ Stripping away all CPU stuff we don't need, focus on essentials
- **Results**
 - ▼ Silicon version: Simulation results
 - ▼ FPGA version: Live, running hardware-based recognizer

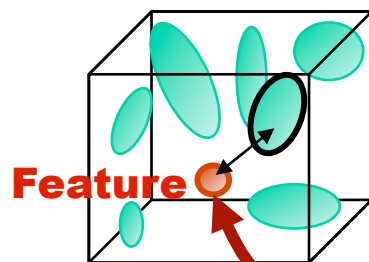
How Speech Recognition Works



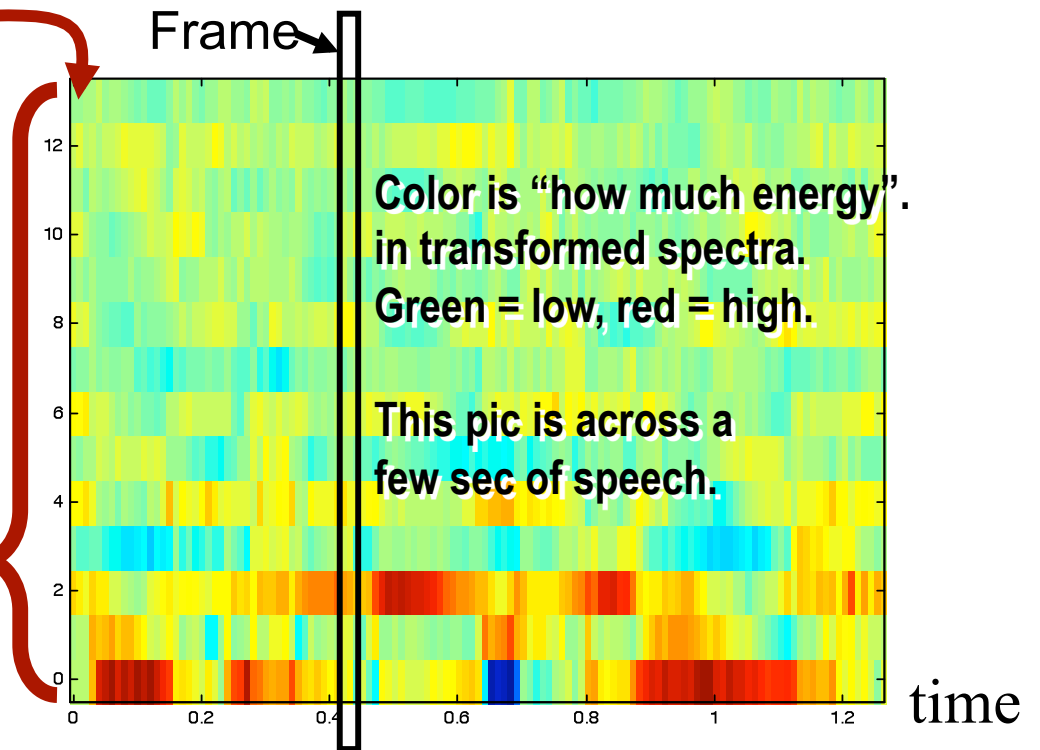
(1) Acoustic Frontend



The frontend is all **DSP**. A discrete Fourier transform (DFT) gives us the spectra. We combine and logarithmically transform spectra in ways motivated by physiology of human ears.



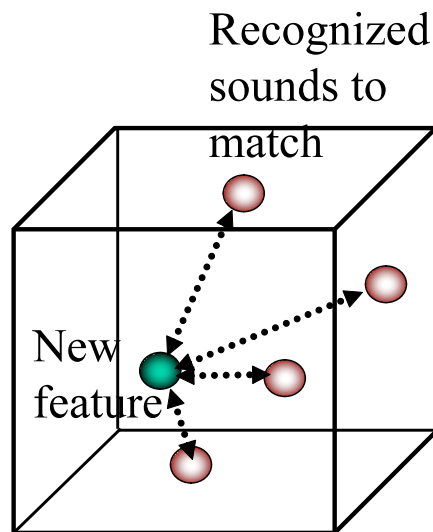
Combine these with estimates of 1st and 2nd time derivatives



(2) The Scoring Stage

■ Feature vec is a point in high-dimensional space

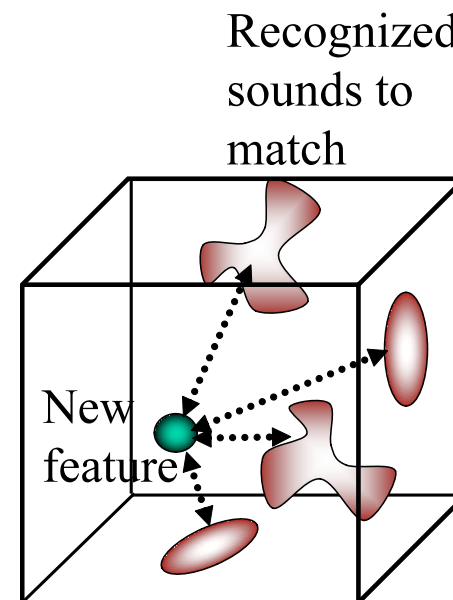
- ▼ Assume each atomic sound we can recognize is also characterized as one a “perfect” point in high-dim ($n=39$) space



- ▼ We used to do this using normalized **distance** as the metric for “likelihood”

■ Problem with using distance

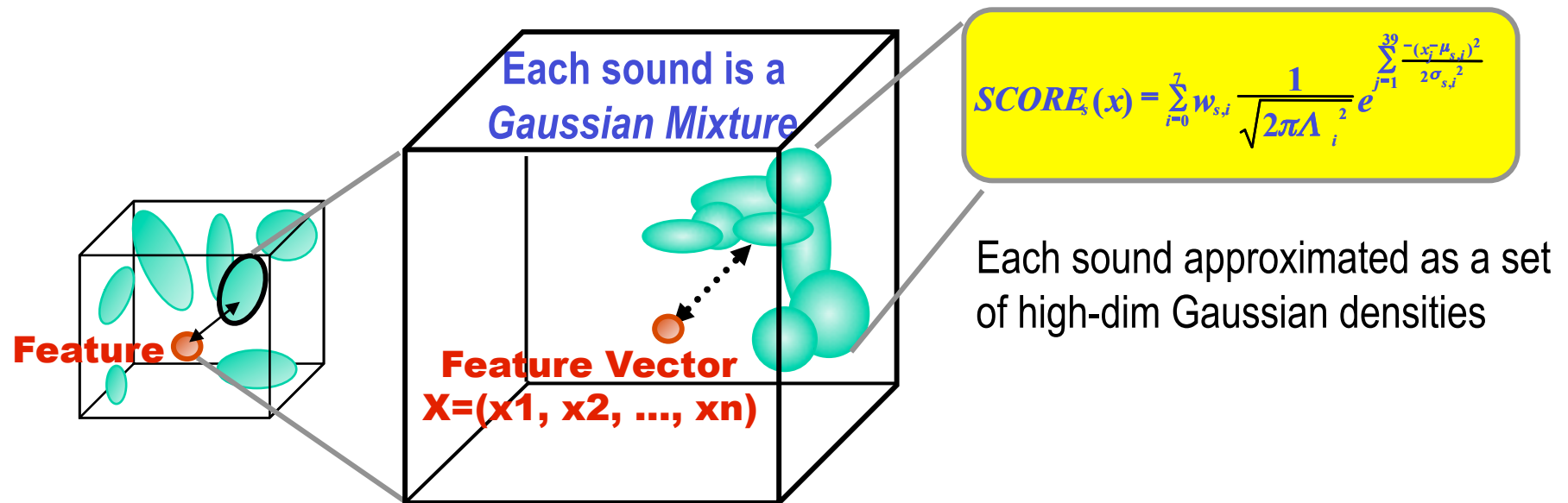
- ▼ Space “occupied” by each atomic sound **not** well modeled as a **point**



- ▼ We need to model the *shape* of the region that defines each sound

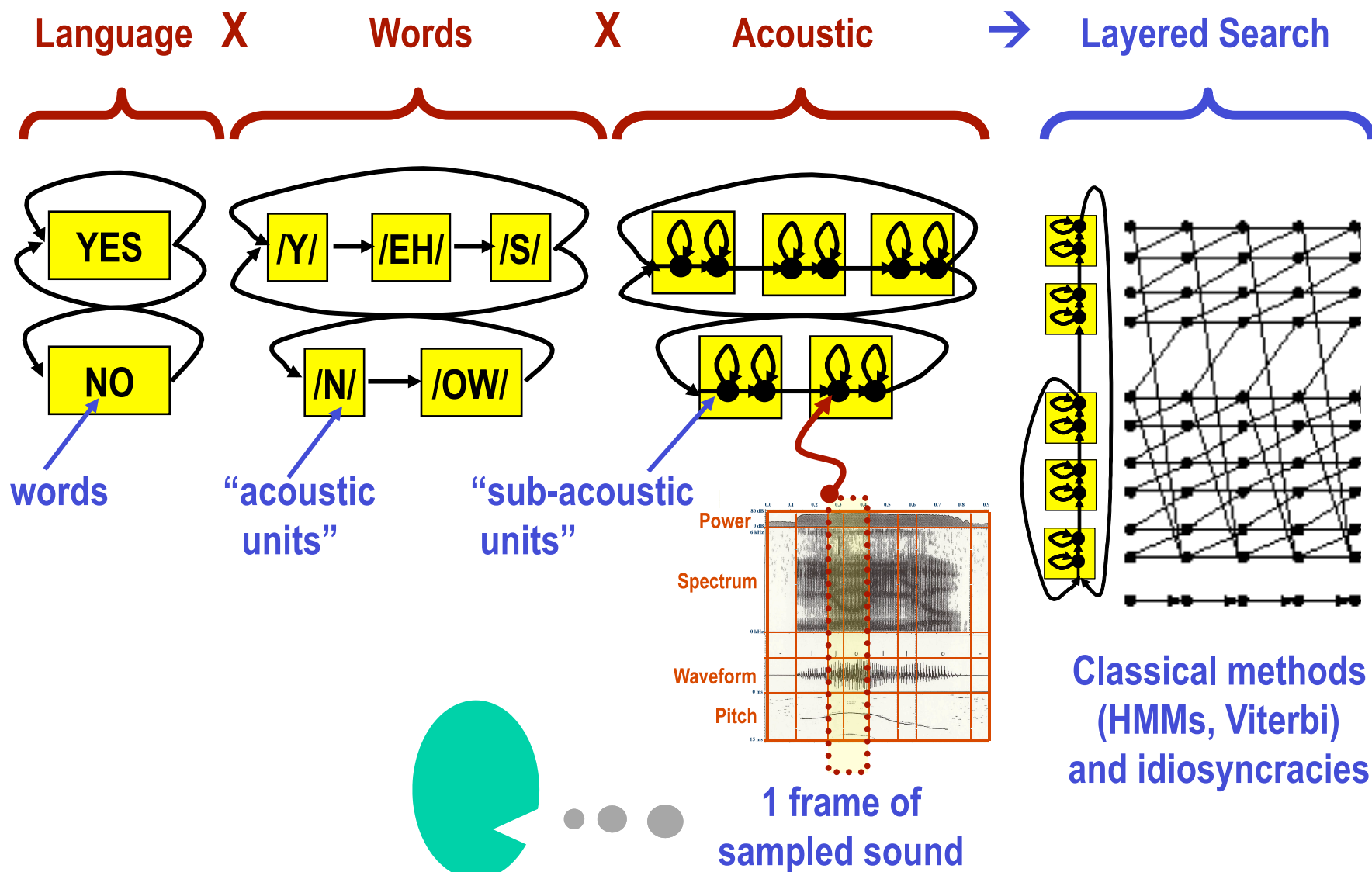
(2) Scoring Stage

- Each feature still a **point** in high-dimensional space
 - ▼ But each “atomic sound” is a **region** of this space
 - ▼ **Score** each atomic sound with $\text{Probability}(\text{sound matches feature})$



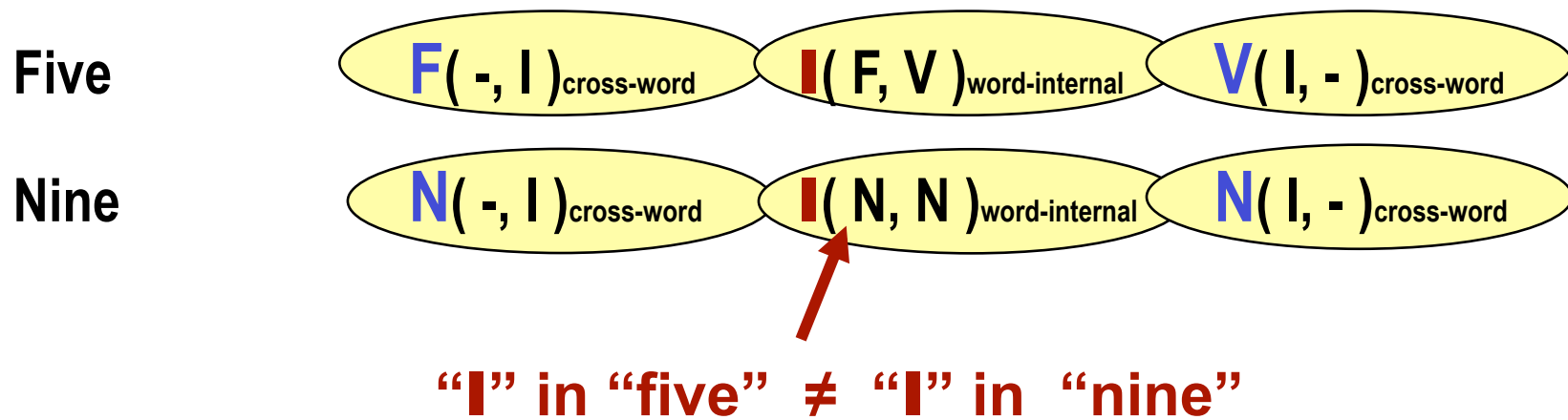
- Note: (sounds) X (dimensions) X (Gaussians) = **BIG**

(3) Search: Speech Models are *Layered* Models



Context Matters: At Bottom -- *Triphones*

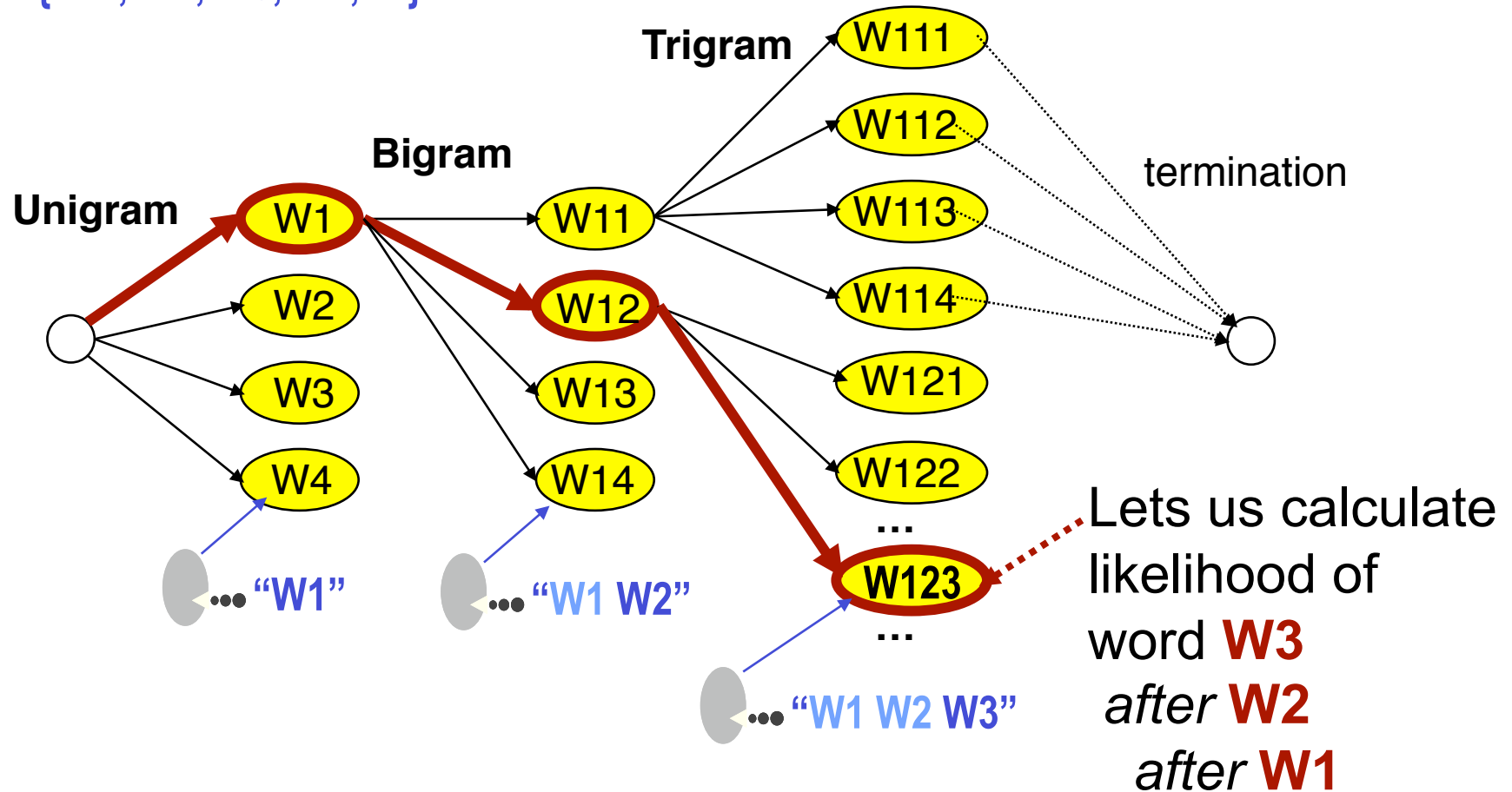
- English has ~50 atomic sounds (**phones**) but we recognize ~50x50x50 context-dependent **triphones**
 - ▼ Because “**l**” sound in “five” is different than the “**l**” in “nine”



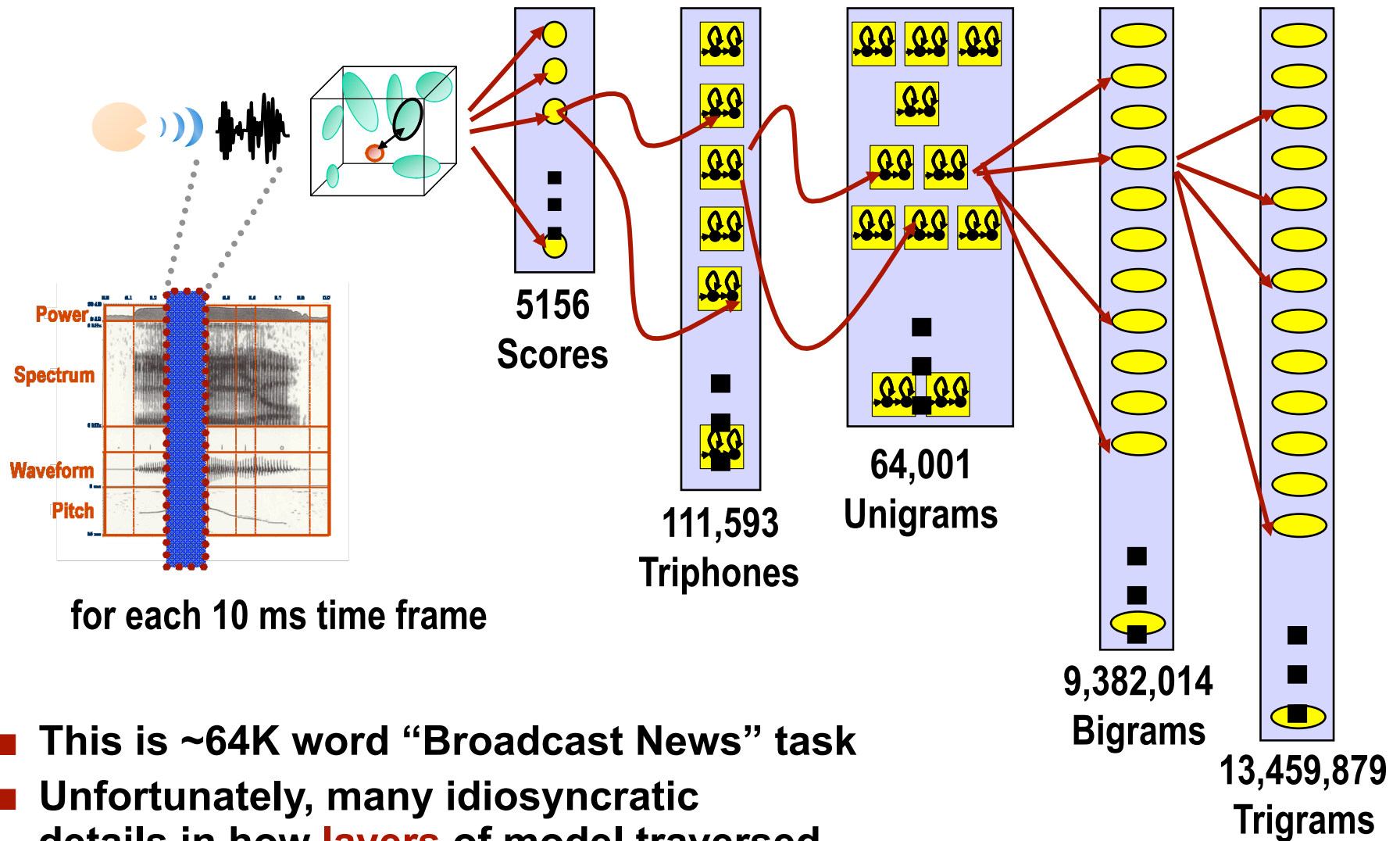
Also Context at Top: *N*-gram Language Model

Suppose we have vocabulary

{ $W_1, W_2, W_3, W_4, \dots$ }



Good Speech Models are BIG

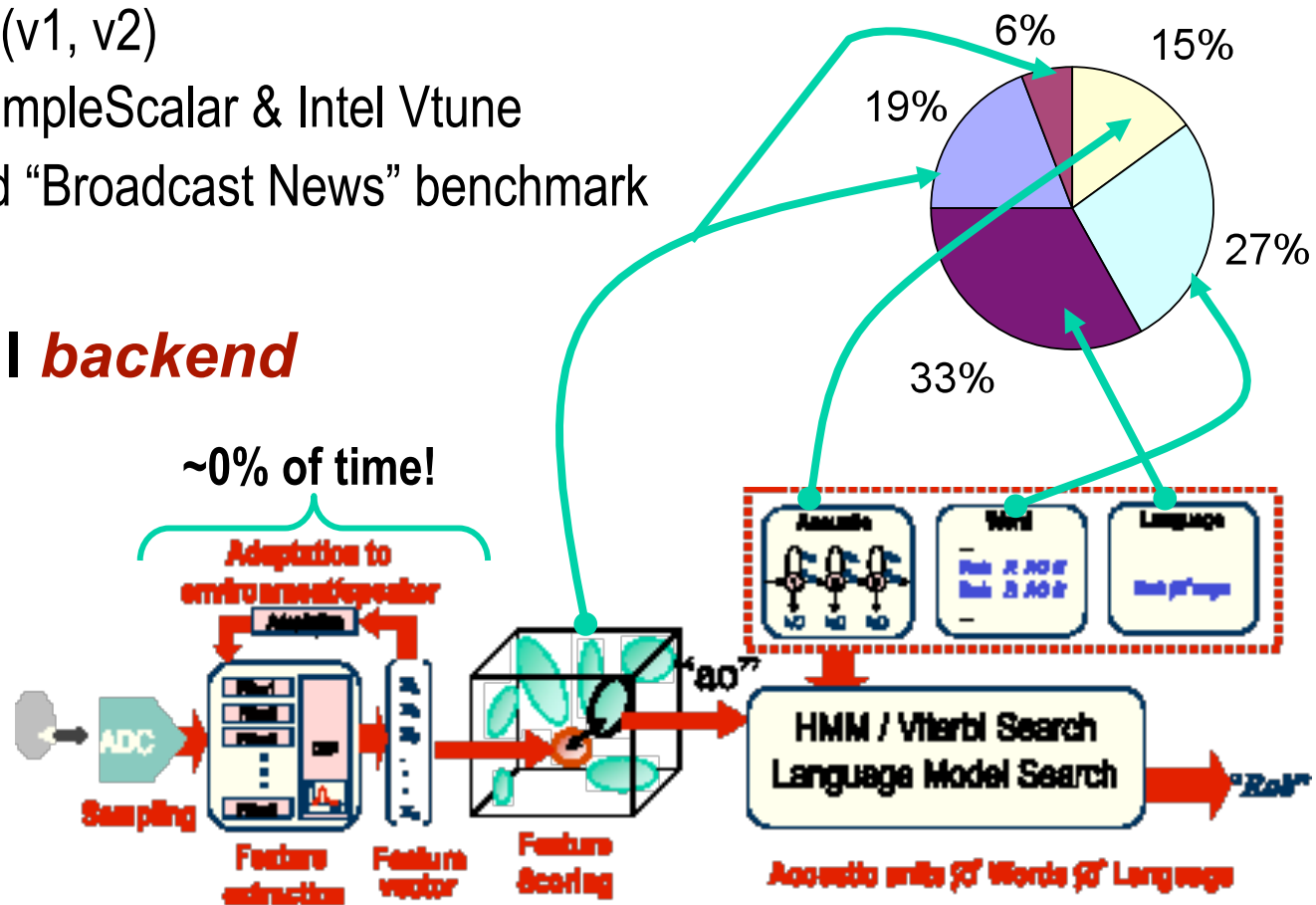


Where Does Software Spend its Time?

■ CPU time for CMU Sphinx 3.0

- ▼ Prior studies targeted less capable versions (v1, v2)
- ▼ Tools: SimpleScalar & Intel Vtune
- ▼ 64K-word "Broadcast News" benchmark

■ So: It's all *backend*



Memory Usage? SPHINX 3.0 vs Spec CPU2000

- **Cache sizes**
 - ▼ L1: 64 KB, direct mapped
 - ▼ DL1: 64 KB, direct mapped
 - ▼ UL2: 512 KB, 4-way set assoc

- **So...**
 - ▼ **Terrible locality** (no surprise, graph search + huge datasets)
 - ▼ **Load dominated** (no surprise, reads a lot, computes a little)
 - ▼ Not an insignificant **footprint**

	SPHINX 3.0	Gcc	Gzip	Equake
Cycles	53 B	55B	15 B	23 B
IPC	0.69	0.29	1.05	0.7
Instruction Mixes				
Loads	0.27	0.25	0.2	0.27
Stores	0.05	0.15	0.09	0.08
Branch's	0.14	0.2	0.17	0.12
Branch Misprediction Rates				
	0.025	0.07	0.08	0.02
Cache Miss Rates				
DL1	0.04	0.02	0.02	0.03
L2	0.48	0.06	0.03	0.30
Memory Footprint				
	64 MB	24 MB	186 MB	42 MB

About This Talk

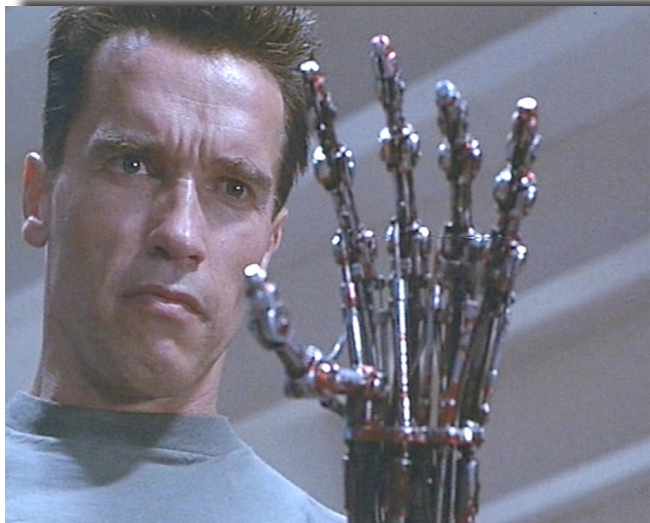
- **Some philosophy**
 - ▼ Why silicon? Why now? Why us (CMU)?
- **A quick tour: How speech recognition works**
 - ▼ What happens in a recognizer
- **A silicon architecture**
 - ▼ **Stripping away all CPU stuff we don't need, focus on essentials**
- **Results**
 - ▼ Silicon version: Simulation results
 - ▼ FPGA version: Live, running hardware-based recognizer

This Talk: How to Get to *Fast*...

Audio-mining

- Very **fast** recognizers – much faster than realtime
- App: search large media streams (DVD) quickly

FIND: “Hasta la vista, baby!”



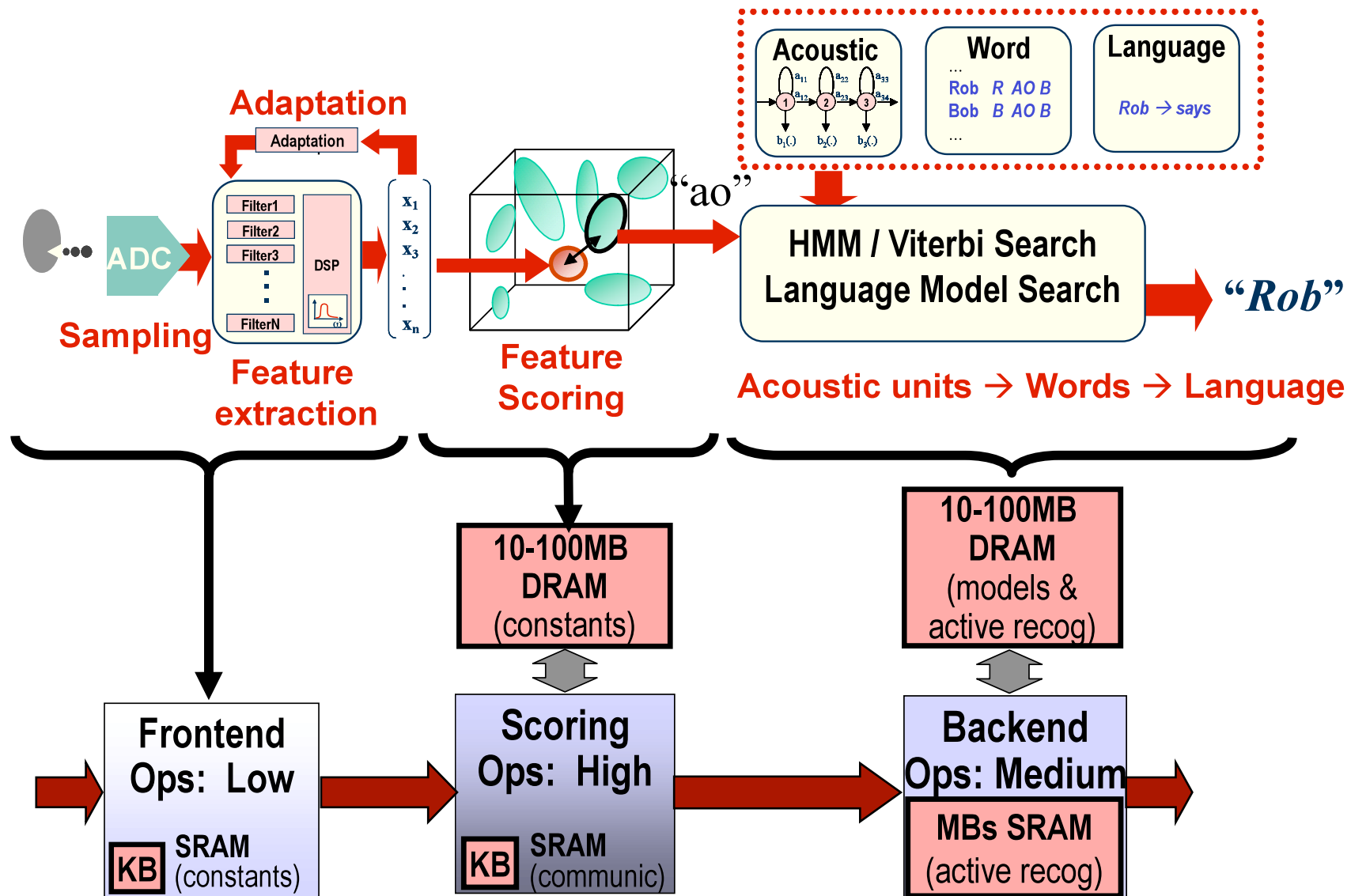
Hands-free appliances

- Very portable recognizers – high quality result on $\ll 1$ watt
- App: interfaces to small devices, cellphone dictation

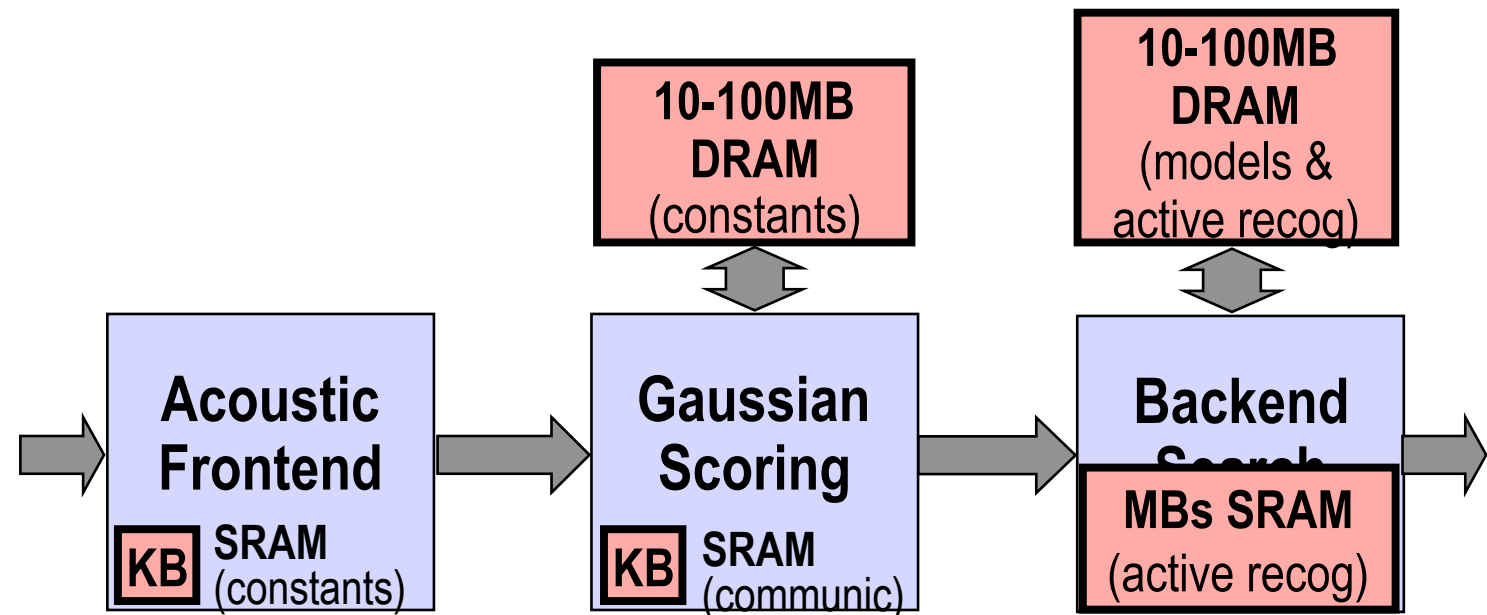


“send email to arnold - let's do lunch...”

Speech: Complex Task to do in Silicon



A Silicon Architecture: Breakdowns



Computations (Ops)	Low	High	Medium
SRAM (size)	Small	Small	Large
DRAM (size)	--	Medium/Large	Large
DRAM (bandwidth)	--	High	High

Essential Implementation Ideas

- **Custom precision, everywhere**
 - ▼ Every bit counts, no extras, no floating point – all fixed point

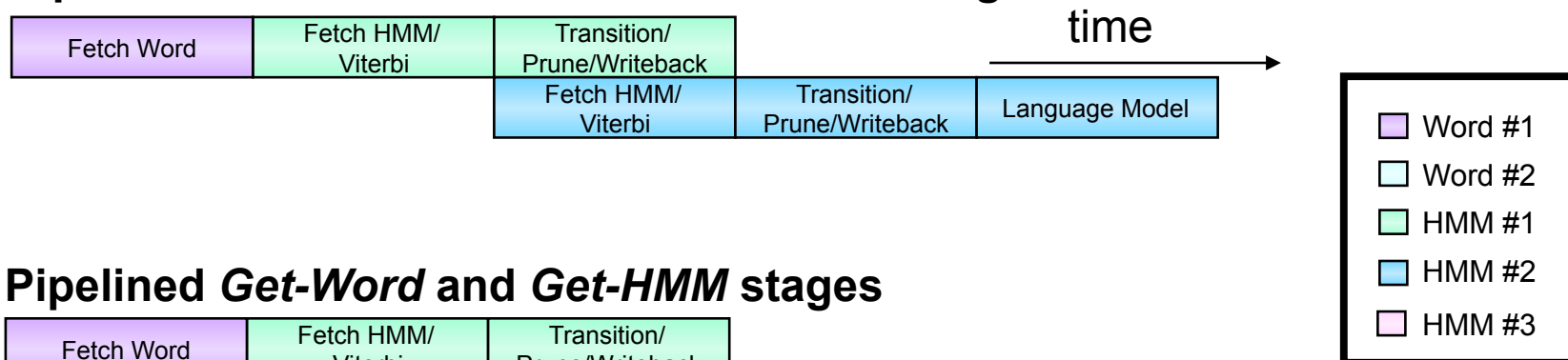
- **(Almost) no caching**
 - ▼ Like graphics chips: fetch from SDRAM, do careful data placement
 - ▼ (Little bit of caching for bandwidth filtering on big language models)

- **Aggressive pipelining**
 - ▼ If we can possibly overlap computations – we try to do so

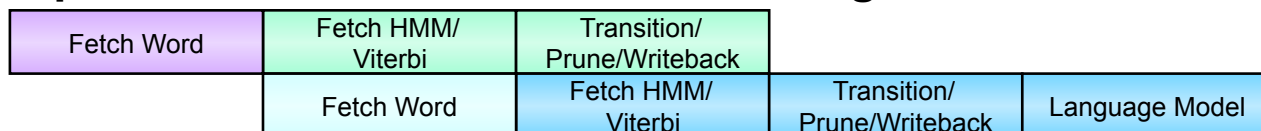
- **Algorithm transformation**
 - ▼ Some software computations are just bad news for hardware
 - ▼ Substitute some “deep computation” with hardware-friendly versions

Example: Aggressive Pipelining

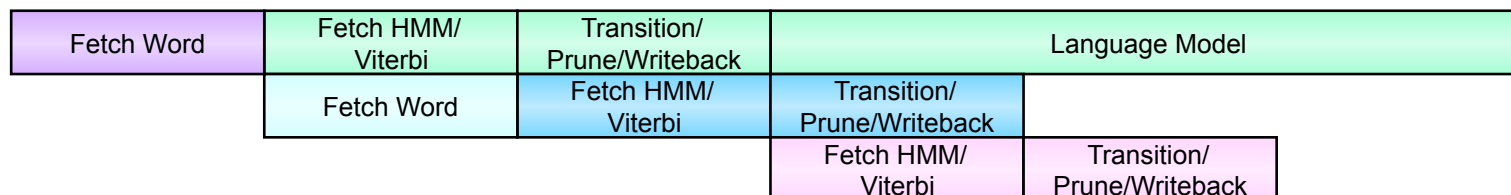
Pipelined *Get-HMM/Viterbi* and *Transition* stages



Pipelined *Get-Word* and *Get-HMM* stages



Pipelined *non-LanguageModel* and *LanguageModel* stages



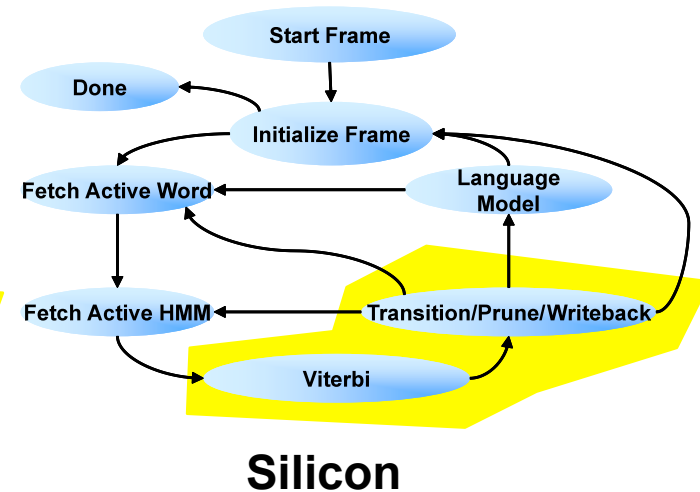
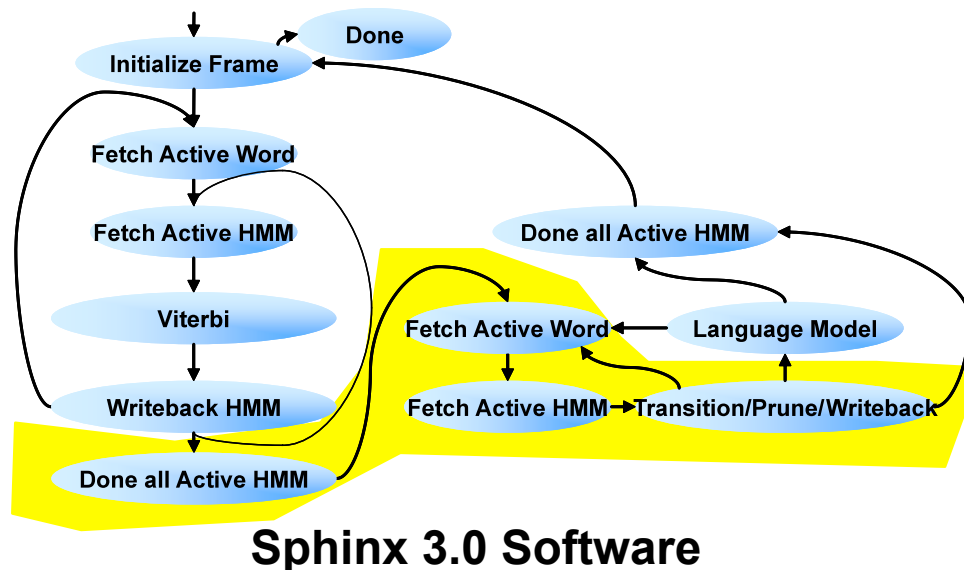
Example: Algorithmic Changes

■ Acoustic-level pruning threshold

- ▼ **Software:** Use best score of *current* frame (after Viterbi on Active HMMs)
- ▼ **Silicon:** Use best score of *previous* frame (nixes big temporal bottleneck)

■ Tradeoffs

- ▼ Less memory bandwidth, can pipeline, little pessimistic on scores



About This Talk

- **Some philosophy**
 - ▼ Why silicon? Why now? Why us (CMU)?

- **A quick tour: How speech recognition works**
 - ▼ What happens in a recognizer

- **A silicon architecture**
 - ▼ Stripping away all CPU stuff we don't need, focus on essentials

- **Results**
 - ▼ **Silicon version: Simulation results**
 - ▼ **FPGA version: Live, running hardware-based recognizer**

Design Flow: C++ Cycle Simulator → Verilog

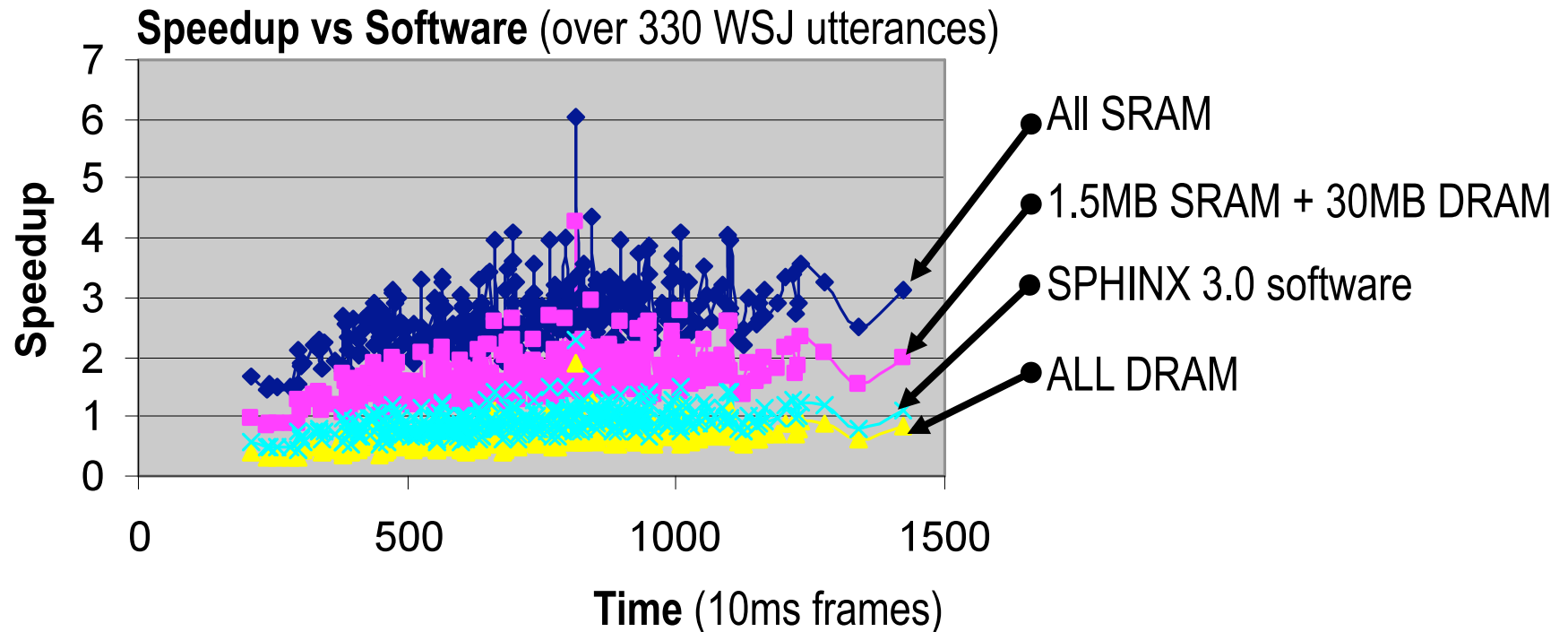
- Our benchmark: 5K-word “Wall Street Journal” task
- Cycle sim results:
 - ▼ No accuracy loss; not quite **2X @ 125MHz** ASIC clock
 - ▼ Backend search needs: ~1.5MB SRAM, ~30MB DRAM



Recognizer Engine	Word Error Rate (%)	Clock (GHz)	Speedup Over Real Time (bigger is better)
Software: Sphinx 3.3 (fast decoder)	7.32%	1 GHz	0.74X
Software: Sphinx 4 (single CPU)	6.97%	1 GHz	0.82X
Software: Sphinx 4 (dual CPU)	6.97%	1 GHz	1.05X
Software: Sphinx 3.0 (single CPU)	6.707%	2.8 GHz	0.59X
Hardware: Our Proposed Recognizer	6.725%	0.125 GHz	1.67X

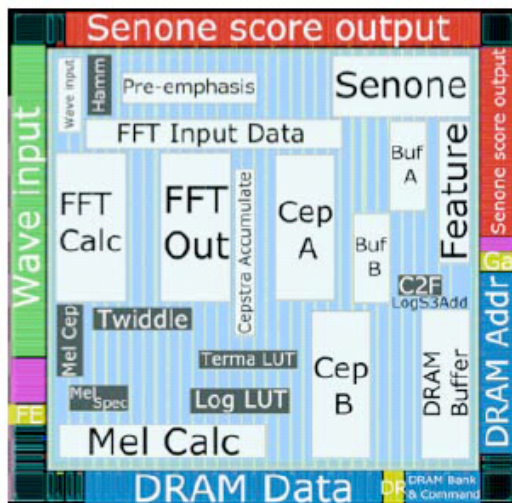
Aside: Bit-Level Verification Hurts (A Lot)

- **Common source of designer headache for silicon designs that handle large media streams**
 - ▼ Generating these sort of tradeoff curves: CPU days → weeks

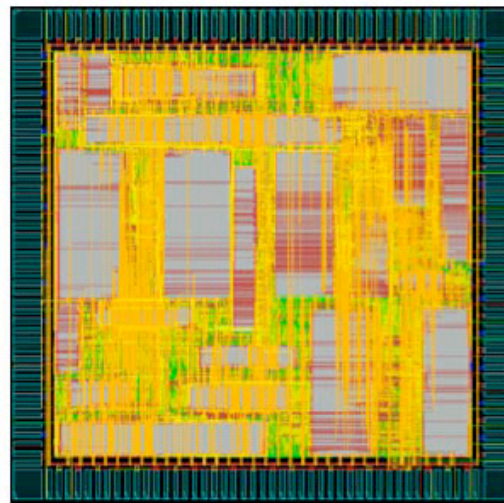


Aside: Pieces of Design = Great Class Projects

- CMU student team: Patrick Chiu, David Fu, Mark McCartney, Ajay Panagariya, Chris Thomas



Floorplan

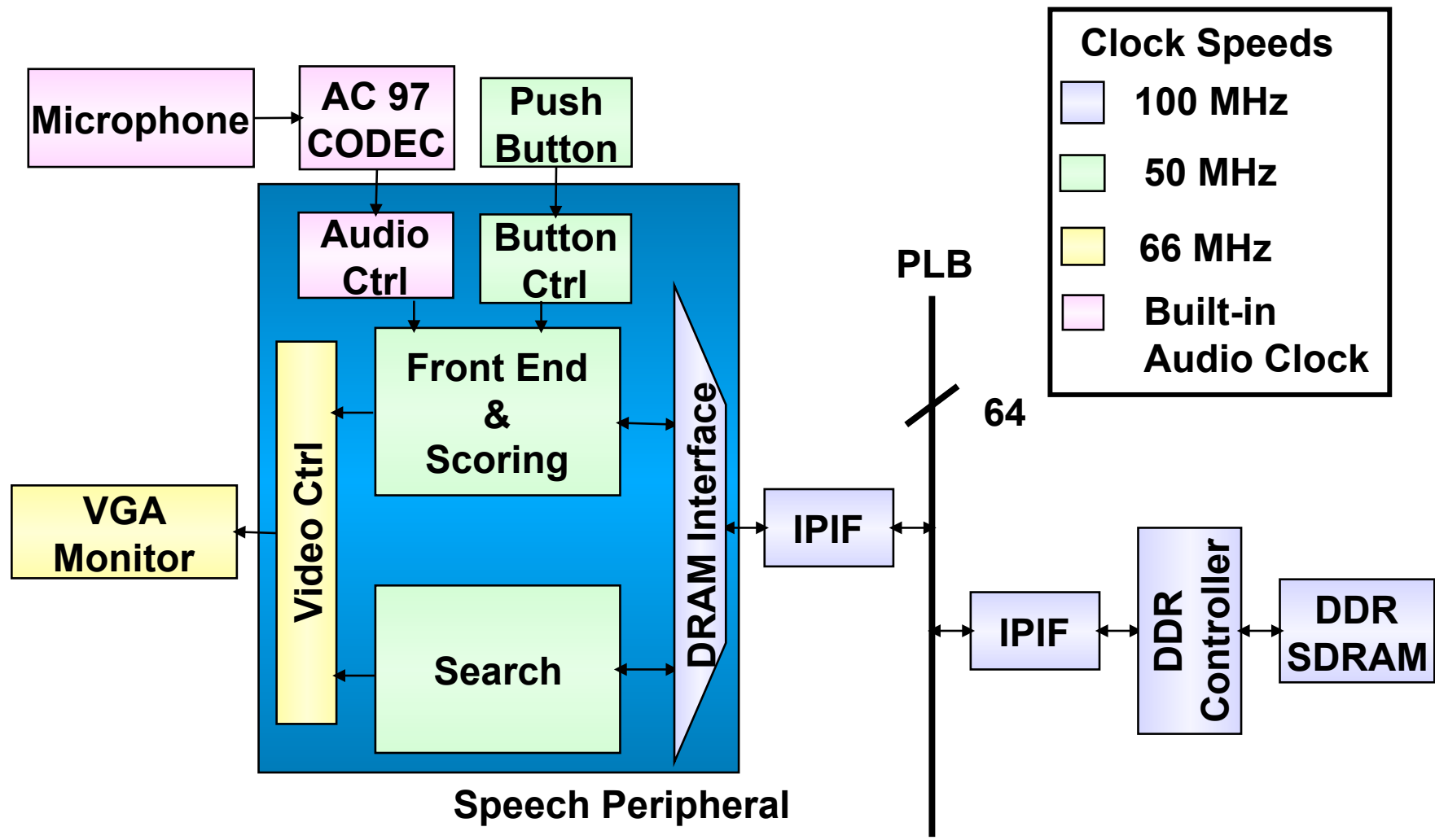


Final Layout

Area	11.16 mm ² core / 16.09 mm ² chip
Effective Utilization	53.32%
Cell Rows	657
Cells	67354
Pins	225358
IO Pins	94
Nets	79382
Avg. Pins/Net	2.84
Nets	
(Internal)	77977
(External)	94
Connections	
(Internal)	146621
(External)	188
Total net length	6.00 m
(X)	2.59 m
(Y)	3.40 m
Power Supply	1.98 V
Average Power	19.8 mW
(switching)	11.78 mW
(internal)	7.98 mW
(leakage)	0.036 mW
Power by clock domain	
Frontend	2.018 mW
Gaussian	14.25 mW
DRAM	2.57 mW
Unlocked	0.96 mW
Power by cell category	
Core	19.5 mW
Block	0.29 mW
IO	0 mW
Worst IR drop	0.012 V

Final Stats

System Block Diagram



FPGA Experimental Results

- **Aside:** as far as we know, this is the *most complex* recognizer architecture ever fully mapped into a running hardware-only form



Performance

- **Benchmark: 1K-word “Resource Management” task**
- **Results:**
 - ▼ **No** accuracy loss
 - ▼ **~ 2x slower than real-time**, but **~30X slower clock frequency**
 - ▼ Limited by DRAM access time and available FPGA resources.

Recognizer Engine	Word Error Rate (%)	Clock (GHz)	Speedup Over Real Time (bigger is better)	Efficiency (Speedup/GHz)
Software: CMU Sphinx 3.0 (single CPU)	10.88%	2.8 GHz	3.7X	1.32
Hardware: Our FPGA Recognizer	10.9%	0.05 GHz	0.5X	10

Aside: How To Tell You're Doing Something Cool...



The talking cure

Speech technology: Good speech recognition requires a fast PC. A chip-based implementation could make the technology more portable

IF YOU have sausage-sized fingers, find a pen-driven handheld computer a fiddle or have never got the hang of predictive text on your mobile phone, a new chip might provide a sympathetic ear. It's being devised by a team of researchers from Carnegie Mellon University and the University of California at Berkeley to do one thing, and one thing only: speech recognition. Using a new, hardware-based approach to the problem, the researchers hope to create a chip that performs speech recognition much more efficiently than is currently possible using software-based recognition systems. If they are right, it might soon become possible to dictate an e-mail into your BlackBerry, or edit your mobile phone's address book using voice commands alone.

Speech-recognition software has been on the market for over a decade, and in the past five years it has become advanced enough to displace keyboard entry, for some users at least. But speech-recognition packages such as IBM's ViaVoice and ScanSoft's Dragon NaturallySpeaking require a powerful desktop computer. Ask a portable device to do the same kind of computational heavy lifting, however, and its battery will be flat within minutes. Why would a chip-only solution be any better?

The reason is simple: doing something in software is more flexible, but doing the



same thing with a dedicated chip consumes far less power. Computationally difficult tasks often start out in software, and are implemented in hardware later. "You do them in software first, because it's easier," says Koh Rutenbar, professor of electrical and computer engineering at Carnegie Mellon and the lead engineer on the "In Silico Vox" speech-chip project. "You redo them in hardware later to maximize their performance."

Computer graphics, for example, have already been through this transition from software to hardware. A few years ago, PCs would grind to a halt as they tried to render complicated graphics. This no longer happens today, because specialized graphics chips—from companies such as ATI and Nvidia—do the hard work. Bob Broderen of the University of California, in Berkeley, has calculated that moving an application from a general-purpose software implementation to a specialized chip can improve efficiency by a factor of 10,000 (the efficiency metric being millions of calculations per milliwatt of power consumed).

The researchers were recently awarded a \$m grant by America's National Science Foundation to develop their speech chip. The grant was made on the basis that a speech-recognition chip would have applications in homeland security. But what starts out as government or military technology often ends up in commercial applications, as packet-switched networks and the global-positioning system demonstrate.

Besides doing away with the need to use fiddly controls on handheld computers, mobile phones and music players, a speech-recognition chip would have other uses too: it could form the basis of a powerful, portable interpreting device, or, for example, allow car drivers to change radio stations or operate navigation systems by speech alone.

Encapsulating the latest speech-recognition in hardware will not be easy, but the Carnegie Mellon researchers have the appropriate experience. They helped to develop much of today's successful speech-recognition technology, including the "Sphinx" software that forms the basis for many commercial speech-recognition systems and was developed with funding from the Defence Advanced Research Projects Agency (DARPA). The researchers hope to have a working prototype within two years.

Even so, fiddly keypads are not going away any time soon. You cannot use a speech-driven device to make a note while talking on the phone, for example, or to send a surreptitious text message during a boring meeting. Despite being small and annoying, keypads will persist. But for the less dextrous, the new chips cannot come soon enough. ■

A wider choice of software

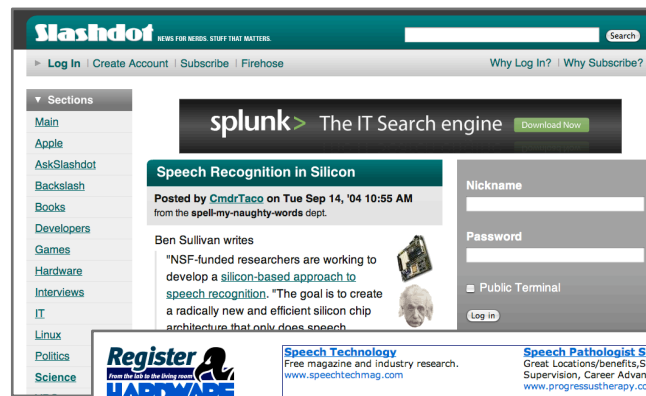
Software: A new kiosk-based approach to selling software on the high street makes obscure but useful titles available to a far larger market

A CHAIN of high-street shops that sell software isn't that a step backwards? Doesn't the future lie in high-speed software downloads over broadband lines, or the replacement of packaged software with constantly updated, web-based alternatives? Not according to Daniel Doll-Steinberg. He is the founder of SoftWide, a company that is promoting a new way to sell software: from kiosks that produce disks and packaging, on demand, inside high-street shops. Is he mad—or might he have discovered a vast untapped market?

Mr Doll-Steinberg originally set out to increase the range of software that could be sold in existing shops. With limited shelf space, most shops stock only a few dozen titles. So his firm, SoftWide, devised a kiosk-based system that can store thousands of pieces of software on a hard disk, burning a disk only when a customer wants to buy a particular title. The first kiosks were tested in vzw Smith, a British retail chain, and in Mac stores in France. But having some titles in boxes on the shelves, and others provided by kiosks, was confusing. So Mr Doll-Steinberg decided that SoftWide should open its own chain of shops. The first two opened in London in 2002, and have since been followed by three more.

The kiosk-based approach means that compared with other software retailers, SoftWide's shops sell an entirely different range of software to an entirely different type of customer. In computer superstores, 60% of software sold is games, and about 35% is business software. In SoftWide's stores, in contrast, 50% of the software sold is educational and reference software (much of which is otherwise only available by mail order), 30% is business software and only 20% games. More than half of SoftWide's customers are women, and many are pensioners.

SoftWide's unusual model has a number of benefits over superstores and online downloads. Mr Doll-Steinberg insists. It makes it possible to offer customers advice, which is hard to come by in big superstores where shelves must be kept stocked. Users feel more comfortable with physical disks than with downloads, which can go wrong, cannot be re-sold, and are unsuitable as gifts. And



Speech Technology
Free magazine and industry research.
www.speechtechmag.com

Speech Pathologist SLP/CF
Great Locations/benefits, Schools, EI CF Supervision, Career Advancement
www.progressustherapy.com



Reg Hardware » News » Bits 'n' Chips

CMU promises to fix speech recognition with a chip

By Ashlee Vance in Palo Alto | [More by this author](#)
22nd August 2006 23:32 GMT

Get the latest from us in your inbox

Hot Chips Speech technology ranks right down the list with flying cars, robots and Windows as the grandest of disappointments in geekdom. Thankfully, the horrors of the technology haven't broken the will of all researchers in the speech field.

In fact, one team at Carnegie Mellon University optimistically thinks they may have solved the speech recognition conundrum with a new chip.

post-gazette **NOW** Business
Pittsburgh Post-Gazette

NOW | NEWS | NEIGHBORHOODS | SPORTS | BUSINESS | LIVING | A & E | M
Dateline Top 50 Markets Personal Business Cars Consumer Technology Commercial Rates Consumer Rates Email Print

Technology

Sound Advice: Teleconverter can't replace telephoto lens

Connected: Getting connected to the Internet will cost you time and money

Tech Briefly: Digg adds filtering options

Product Review: Sony stuffs HD video into small camcorder

Users Guide: From the lab to

High-speed speech calls for hardware

Wednesday, September 20, 2006
By David Templeton, Pittsburgh Post-Gazette

Imagine a computer understanding everything you say, regardless of how fast you speak or the words you use.

And while you're talking to that computer, it's also turning your words immediately into type.

Then imagine technology that can do this a thousand times faster than real time as a means of processing thousands of hours of recorded speech in a fraction of the time.

Lake Fong, Post-Gazette

Rob A. Rutenbar, professor of electrical and computer engineering at Carnegie Mellon University, aims to create a computer chip that understands speech and processes it more quickly than current software can.

Summary

■ Software is too constraining for speech recognition

- ▼ Evolution of graphics chips suggests alternative: **Do it in silicon**
- ▼ Compelling performance and power reasons for silicon speech recog

■ Several “*in silico vox*” architectures in design

- ▼ Custom silicon and FPGA versions
- ▼ ~10X realtime and low-power mobile architectures in progress at CMU

■ Reflections

- ▼ Some of the most interesting experiences happen when you get people from very different backgrounds – **silicon + speech** – on same team

Acknowledgements

■ **Work supported by**

- ▼ US National Science Foundation (www.nsf.gov)
- ▼ Semiconductor Research Corporation (www.src.org)
- ▼ FCRP Focus Research Center for Circuit & System Solutions (www.fcrp.org, www.c2s2.org), one of five centers supported by the FCRP, an SRC program.

■ **We are grateful for the advice and speech recognition expertise shared with us by**

- ▼ Richard M. Stern, CMU
- ▼ Arthur Chan, CMU
- ▼ Mosur K. Ravishankar, CMU