

# CS250

## VLSI Systems Design

### Lecture 6: Project Ideas

---

Spring 2016

John Wawrzynek  
with  
Chris Yarp (GSI)

**Thanks to John Lazzaro for the slides**

# Today's lecture plan ...

---

- \* **Power and energy.** The techniques you'll be able to use in your project.
- \* **Pareto Optimality ...** and how it impacts your project.
- \* **Accelerator interface ...** and its limits.

Break

- \* **Starting points.** Brief descriptions of projects ideas you may want to pursue.

# Power techniques available for project

---

\* Parallelism and pipelining ✓

~~\* Power-down idle transistors~~

\* Slow down non-critical paths ✓

\* Clock gating ✓

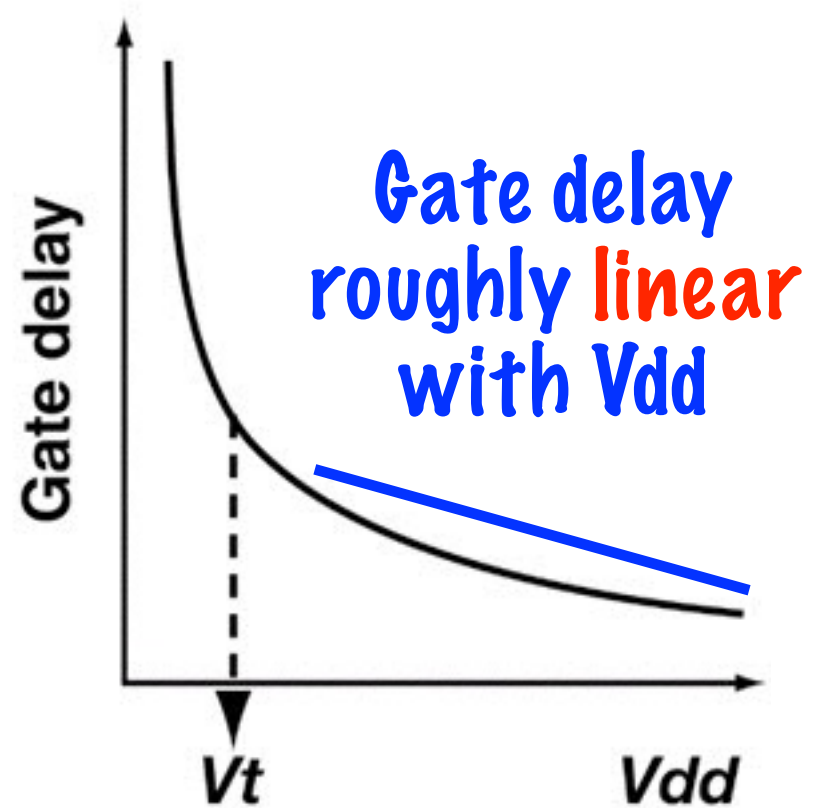
\* Data-dependent processing ✓

~~\* Thermal management~~

Cell libraries characterized at multiple  $V_{dd}$  values.

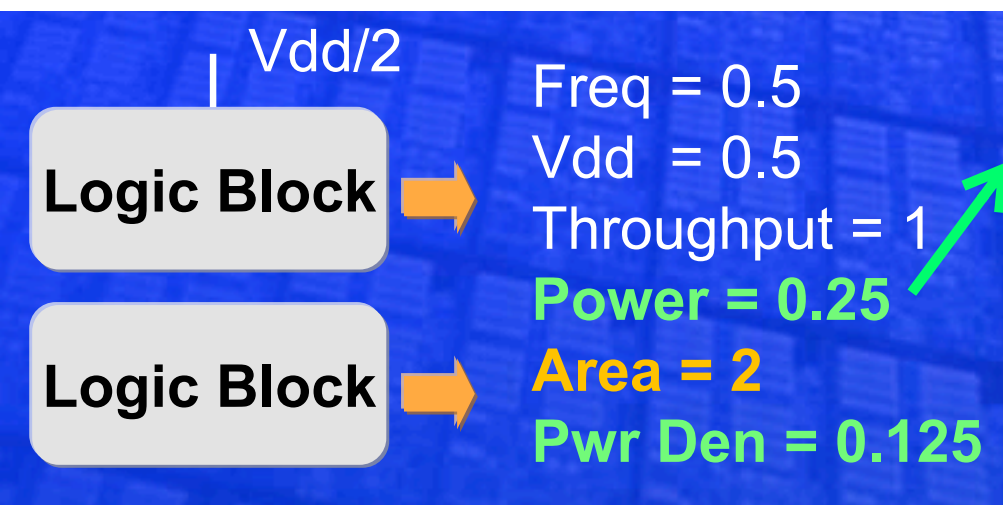
So, you can pick a different  $V_{dd}$  value for each of your implementations.

Several  $V_{dd}$  values in one implementation not supported.



## The main trick:

Top block processes "left", bottom "right".



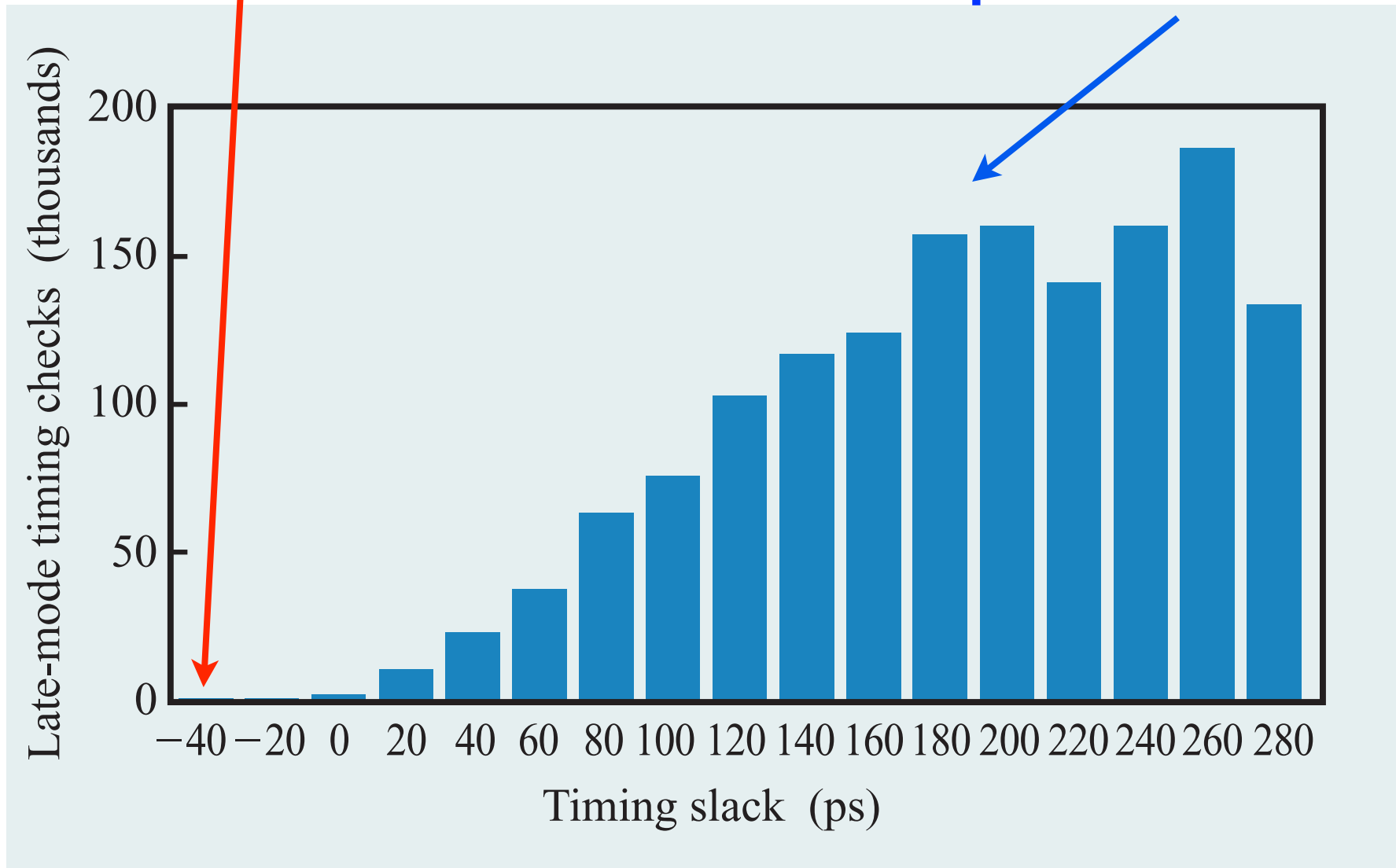
$$P \sim \#blks \times F \times V_{dd}^2$$
$$P \sim 2 \times 1/2 \times 1/4 = 1/4$$

$CV^2$  power only

# Not by varying $V_{dd}$ , but by cell choice

The critical path

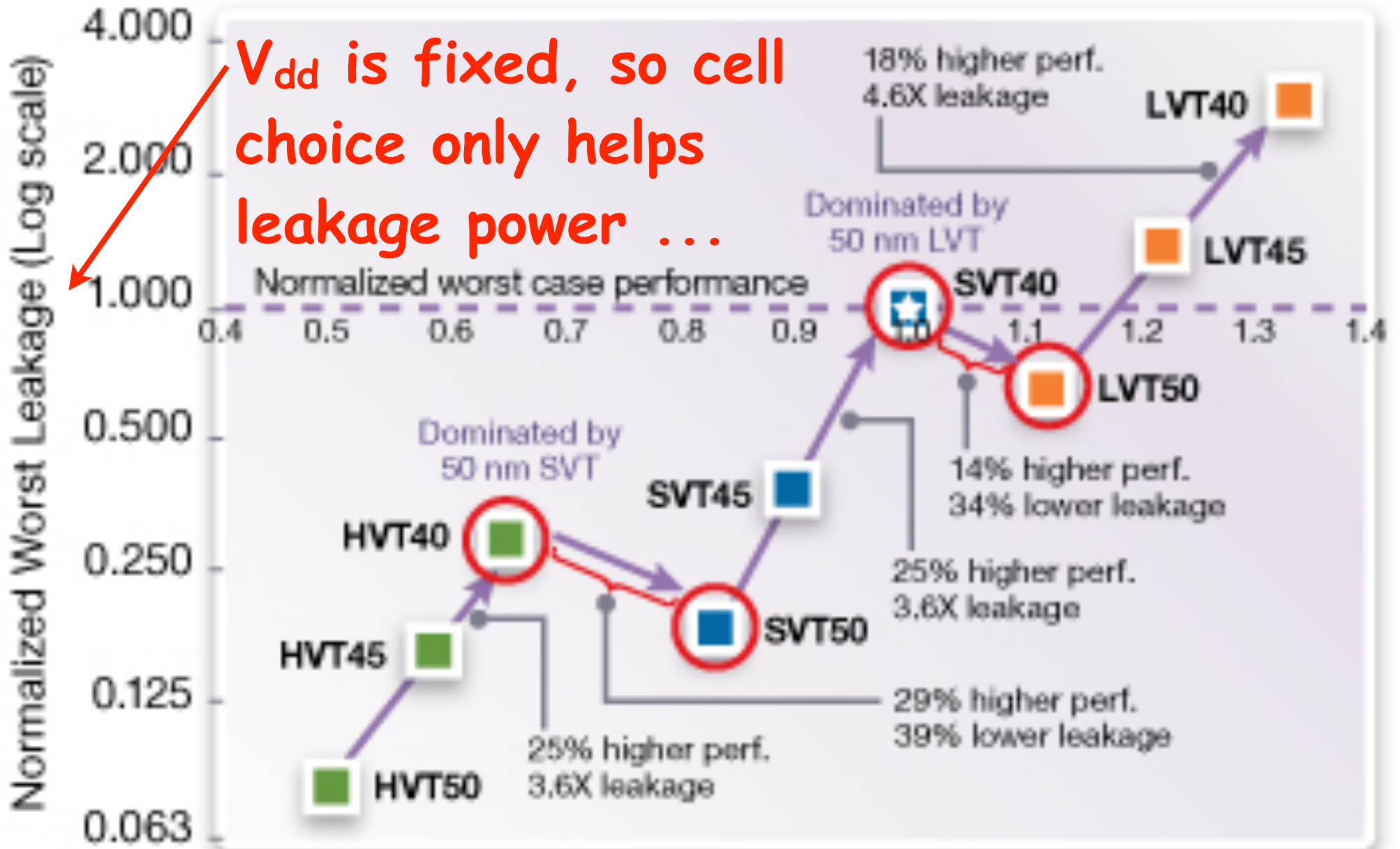
Most wires have hundreds of picoseconds to spare.



From "The circuit and physical design of the POWER4 microprocessor", IBM J Res and Dev, 46:1, Jan 2002, J.D. Warnock et al.



# (H,S,L) == High Vt, Standard Vt, Low Vt

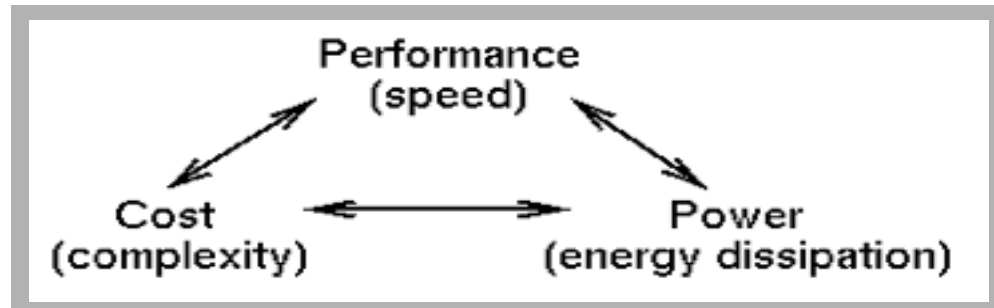


(40, 45, 50) are channel lengths (in nm)

\*Source: Synopsys (Virage Logic)

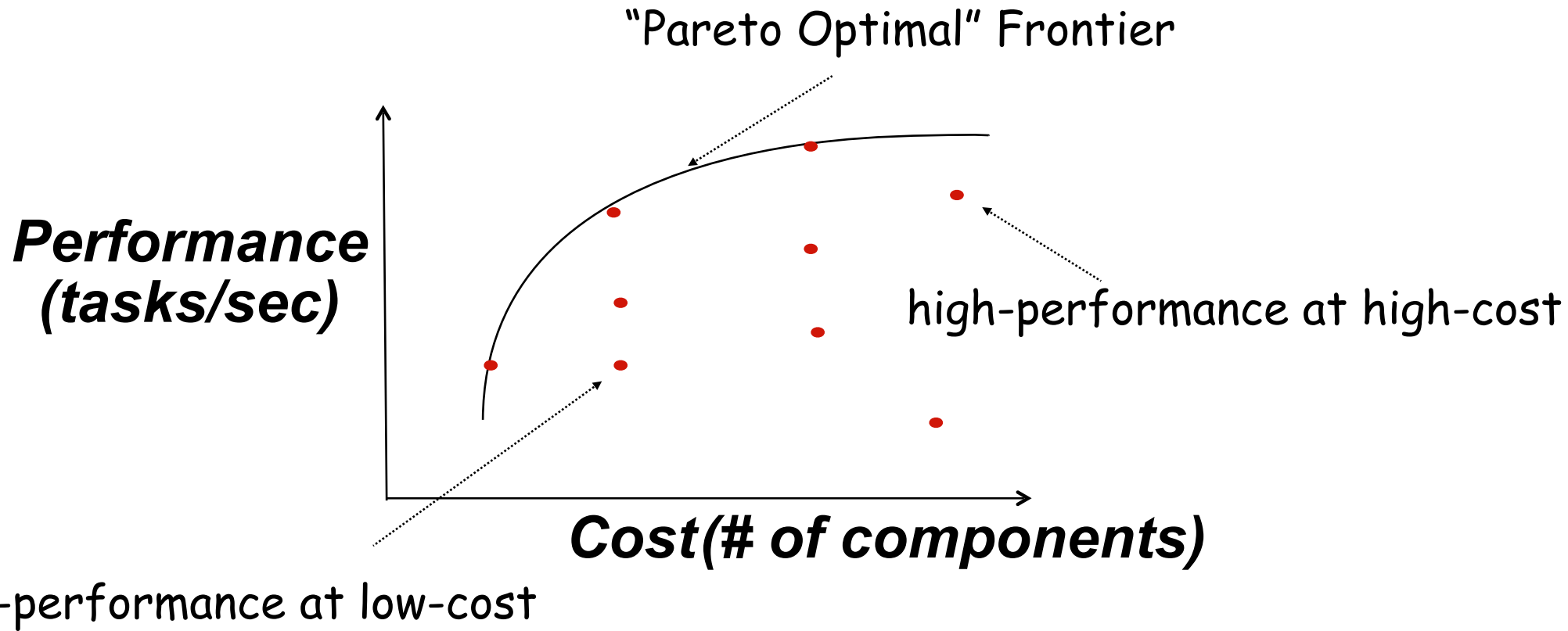


# Basic Design Tradeoffs



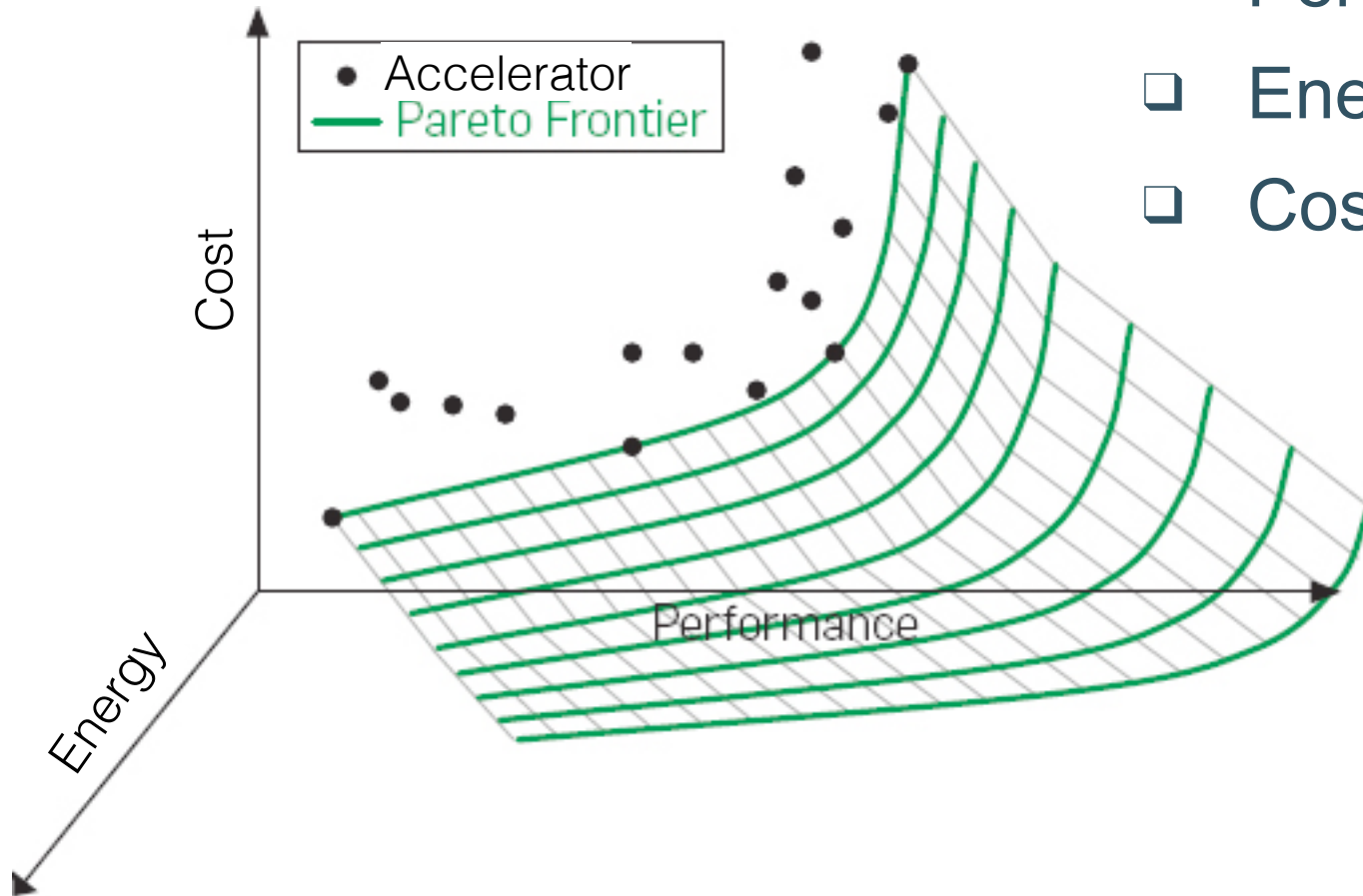
- Improve on one metric at the expense of the others
- Tradeoffs exist at every level in the system design
- Design Specification
  - Functional Description
  - Performance, cost, power constraints
- Designer must make the tradeoffs needed to achieve the function within the constraints
- The **design space** is all the feasible design points (in this 3D space)
- Examining points in that space is called “Design Space Exploration” (DSE).
- Other secondary metrics:
  - time-to-market
  - NRE
  - upgradability/flexibility

# Design Space & Optimality (perf & cost)





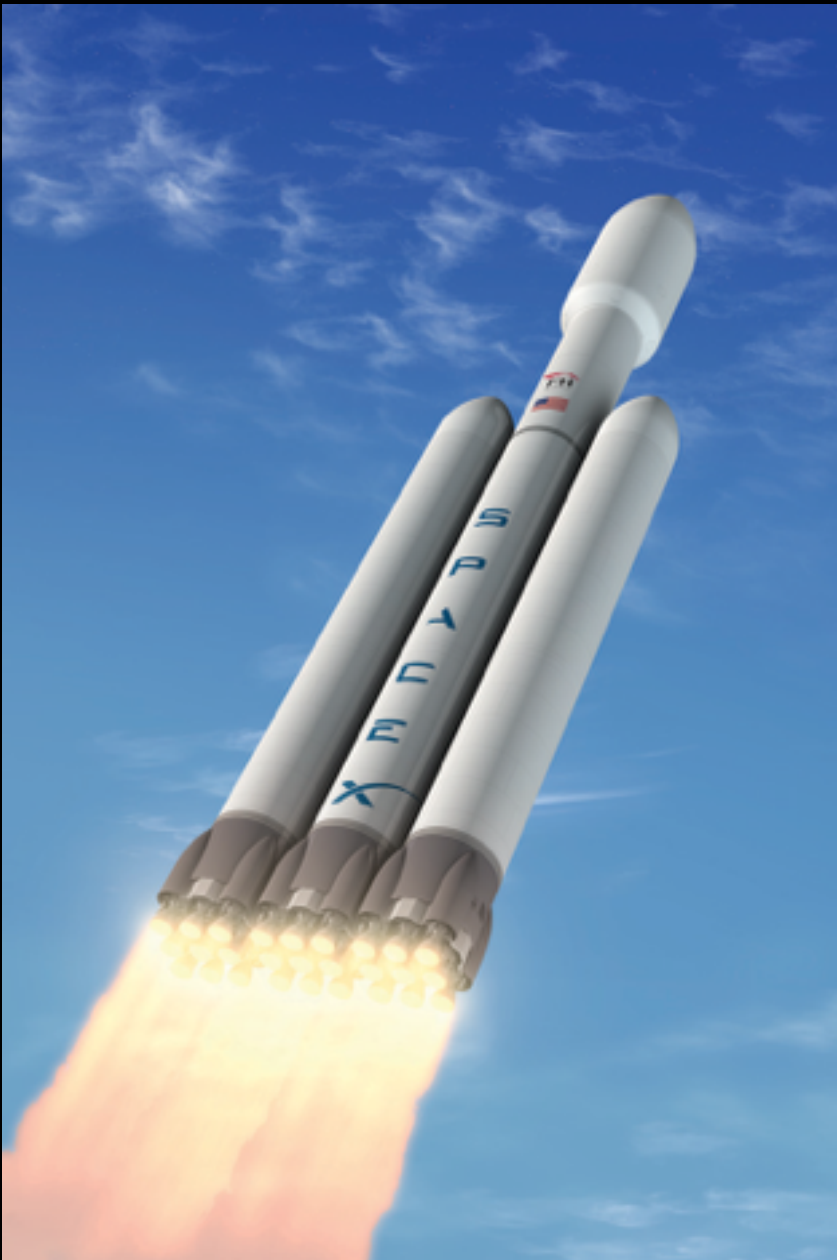
# VLSI Design Space is 3D



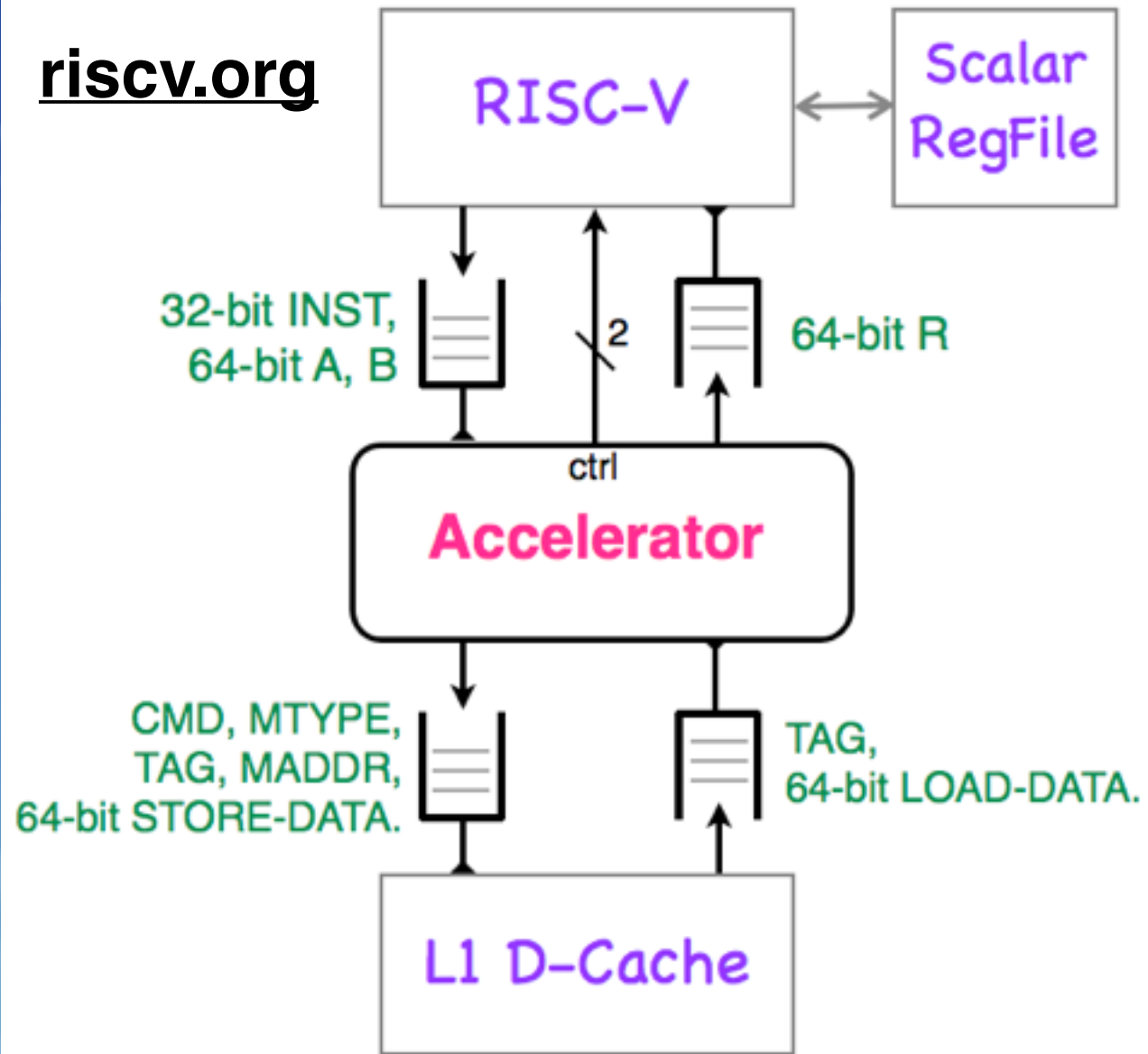
- ❑ Performance in ops/sec
- ❑ Energy is ops/J
- ❑ Cost in die area

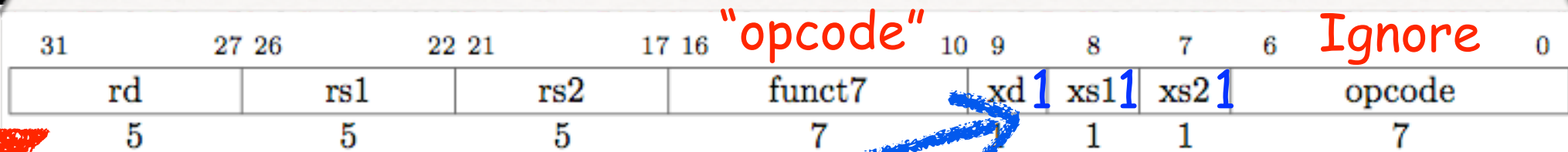
**Project Objective: Determine the Pareto Frontier for some accelerator design over *at least 2 dimensions*, using RTL and physical mapping design variations. Algorithmic design alternatives time permitting.**

# Rocket Facts

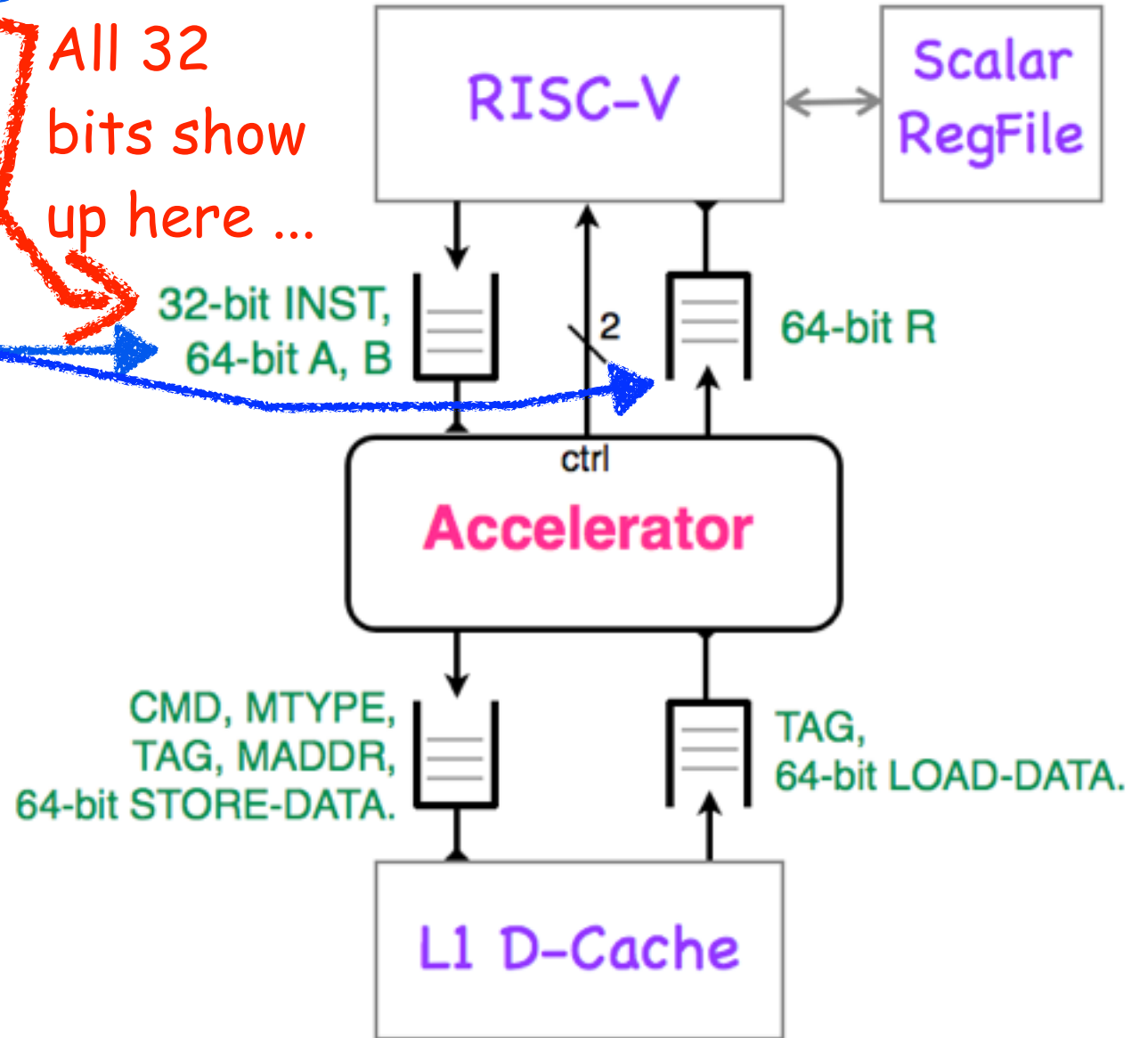


[riscv.org](http://riscv.org)





Blue "1" bits yield register queuing shown on diagram ... the 7-bit funct7 field is accelerator's 128 opcodes.



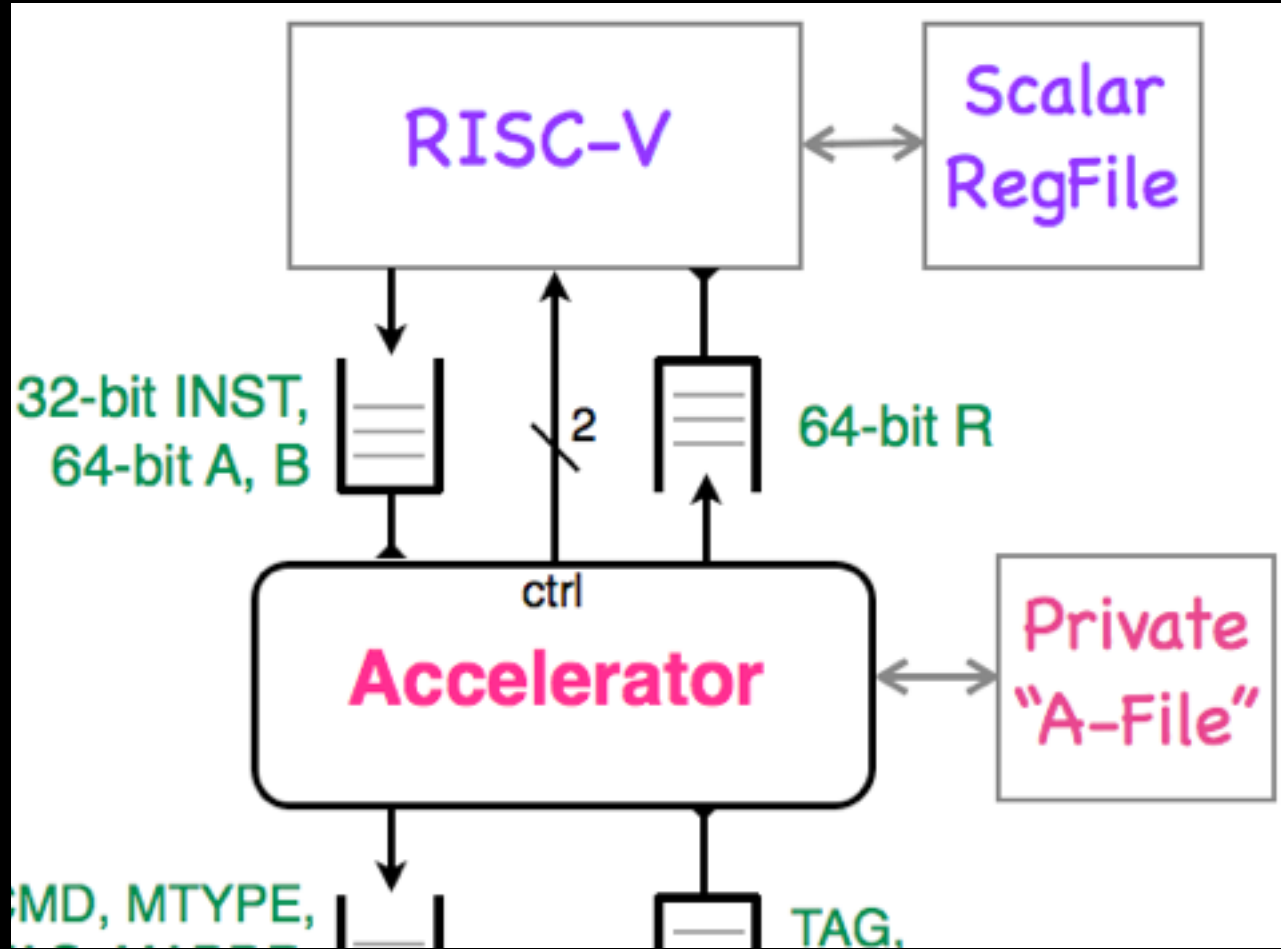
# RISC-V Regfile Ops

|        |      |      |      |
|--------|------|------|------|
| 17 16  | 10 9 | 8    | 7    |
| funct7 | 1xd  | 1xs1 | 1xs2 |
| 7      | 1    | 1    | 1    |

# Private "A-File" Ops

|        |      |      |      |
|--------|------|------|------|
| 17 16  | 10 9 | 8    | 7    |
| funct7 | 0xd  | 0xs1 | 0xs2 |
| 7      | 1    | 1    | 1    |

How to manage private 32-entry "A-File" register bank without using up opcode bits ...



MOV: A-file to Regfile

|        |      |      |      |
|--------|------|------|------|
| 17 16  | 10 9 | 8    | 7    |
| funct7 | 1xd  | 0xs1 | 0xs2 |
| 7      | 1    | 1    | 1    |

MOV: Regfile to A-File

|        |      |      |      |
|--------|------|------|------|
| 17 16  | 10 9 | 8    | 7    |
| funct7 | 0xd  | 1xs1 | 1xs2 |
| 7      | 1    | 1    | 1    |

# D-Cache Facts

Size: 64KB L1 with 64 byte lines.

CMD: Load, Store

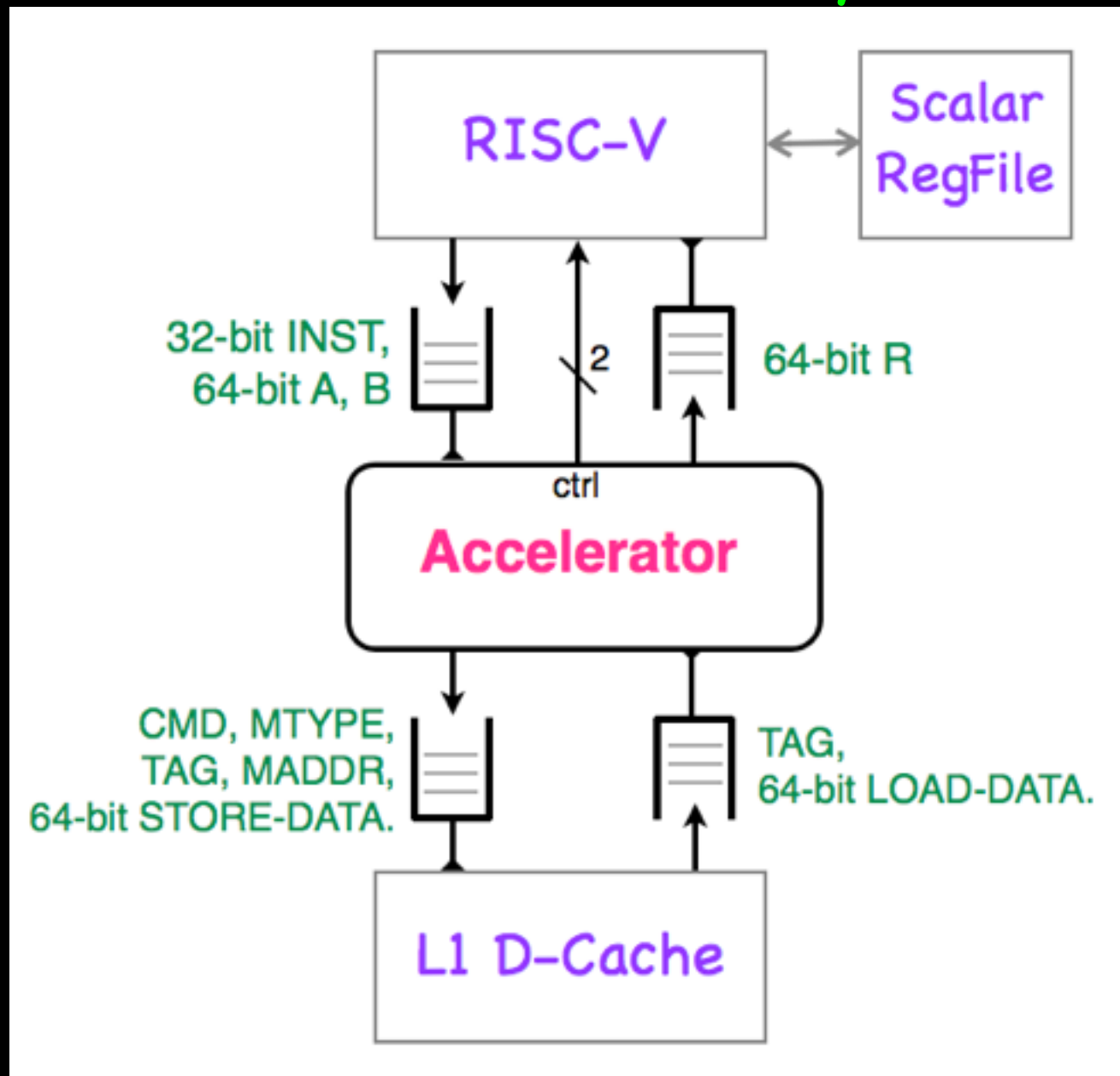
MTYPE:

8, 16, 32, 64 bits.

MADDR:

Align to MTYPE

TAG: 9-bits. Lets loads be OoO, up to 4 missed loads.



Performance: 4 cycle latency on a cache hit, 40-60 on a miss. No prefetching built in ...

# Break

---





# Projects from Fall 2014

- ▶ **Hardware Accelerator for Exact Dot Product**: a coprocessor capable of computing a dot product exactly by use of a “complete register” (CR) that encodes a fixed point representation of twice the IEEE754 double precision. Showed that the coprocessor is realizable in silicon, requiring only 11% of the parent-processor’s area. Additionally, the accelerator showed speedups of 3-6x over a conventional dot product and matrix multiplication while providing both exactness and reproducibility.
- ▶ **Hardware-Accelerated Key Value Store**: a hardware accelerator for the Memcached key-value store: shows a 10x improvement in latency for 40% of requests without adding significant overhead to the remaining requests.
- ▶ **A Compile-Time and Run-Time Configurable Viterbi Decoder in Chisel Hardware Construction Language**: accelerator outperforms pure software implementations in throughput by a factor of 500 to 10000.
- ▶ **Accelerator to Solve System of Linear Equation on A RISC-V Processor**: Algorithm based on matrix condensation and matrix mirroring is adapted from the Journal of Discrete Algorithms. Two variants of the baseline implementation based on parallelism and higher condensation are explored for performance, power, and area metric.



# Projects from Fall 2013

- ▶ Power Modeling (for power estimation)
- ▶ Elliptic Curve Cryptography
- ▶ SHA3
- ▶ Automatic Pipelining
- ▶ Correlation Engine
- ▶ Source Routing (for NN simulation)
- ▶ Memory Controller
- ▶ DREAMER
- ▶ Configurable Precision (for vector unit)
- ▶ Convolution Engine



# Click Prediction Acceleration

富嶽三十六景 甲川  
三段水面

長谷川雪村





Advertisers pay Google **\$1.47**, on average, if Google Search displays their ad in response to the search term **mt fuji vacation**.

富士山 甲川 三石水画

Google AdWords

Home

Campaigns

Opportunities

Tools and Analysis ▾

Billing ▾

My account ▾

Keyword Planner

Add ideas to your plan

Your product or service

mt fuji vacation

Go

Targeting ?

All locations

All languages

Google

Ad group ideas

Keyword ideas

Search terms

▼ Avg. monthly searches ?

Competition ?

Avg. CPC ?

mt fuji vacation

↕

10

Low

\$1.47

富嶽三十六景 甲川 三段水面

Since Google is only paid if the user clicks, they predict, in real time, which of the bidding ads is **most likely** to yield a click.

Google

mt fuji vacation



Web

Images

Maps

Shopping

Videos

More ▾

Search tools

About 250,000 results (0.18 seconds)

Ad related to mt fuji vacation ⓘ

**Mount Fuji Tours - Book 5-star rated Mt. Fuji tours - viator.com**

[www.viator.com/mt-fuji](http://www.viator.com/mt-fuji) ▾

★★★★★ 328 reviews for viator.com

With Hakone from Tokyo on Viator.

Viator.com has 91,540 followers on Google+

5-Star Rated Tokyo Tours

Bullet Train Tours

Top Mount Fuji Tours

**Fuji Tourism and Vacations: 10 Things to Do in Fuji, Japan ...**

[www.tripadvisor.com](http://www.tripadvisor.com) ▾ Asia ▾ Japan ▾ Chubu ▾ Shizuoka Prefecture ▾

Fuji Tourism: TripAdvisor has 236 reviews of Fuji Hotels, Attractions, and Restaurants making it your best Fuji Vacation resource. ... Fuji from Nagoya, and then a climb of Mt Fuji - suggestions? by Joe\_from\_Boston 10 replies; What is the best ...

**Mount Fuji - Chubu - Reviews of Mount Fuji - TripAdvisor**

[www.tripadvisor.com](http://www.tripadvisor.com) ▾ Asia ▾ Japan ▾ Chubu ▾ Things to Do in Chubu ▾

Hawaii ad penalized.

Ads ⓘ

**Fiji Vacations from \$1599**

[www.pacificislands.com/Fiji-packages](http://www.pacificislands.com/Fiji-packages) ▾

1 (800) 888 0120

Fiji vacation packages with Air Travel before Dec 11. Save \$450

**All-Inclusive Vacations**

[www.libertytravel.com/Vacations](http://www.libertytravel.com/Vacations) ▾

1 (855) 461 9661

Up to 50% Off Our Package Rates. Inquire Online Or Talk To An Agent.

**Vacation Packages Hawaii**

[www.tripadvisor.com/Hawaii](http://www.tripadvisor.com/Hawaii) ▾

★★★★★ 30 reviews for tripadvisor.com

Find Deals & Read Real Reviews. Hawaii deals on TripAdvisor!



## Ad Click Prediction: a View from the Trenches

H. Brendan McMahan, Gary Holt, D. Sculley, Michael Young,  
Dietmar Ebner, Julian Grady, Lan Nie, Todd Phillips, Eugene Davydov,  
Daniel Golovin, Sharat Chikkerur, Dan Liu, Martin Wattenberg,  
Arnar Mar Hrafnkelsson, Tom Boulos, Jeremy Kubica

Google, Inc.

Basic idea: **Billions** of “features” are developed to predict, given an ad and a search, how likely it is that the searcher will click on the ad.

$$\mathbf{a} \cdot \mathbf{b} = \sum_{i=1}^n a_i b_i = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n$$

**a** vector: **feature values** for the search.

**b** vector: **feature values** for the ad.

Example 50 features: Does geo info indicate that the searcher is in the state of (1) Alabama?  
(2) Alaska ... (50) Wyoming. Binary, sparse features.

## Ad Click Prediction: a View from the Trenches

H. Brendan McMahan, Gary Holt, D. Sculley, Michael Young,  
Dietmar Ebner, Julian Grady, Lan Nie, Todd Phillips, Eugene Davydov,  
Daniel Golovin, Sharat Chikkerur, Dan Liu, Martin Wattenberg,  
Arnar Mar Hrafnkelsson, Tom Boulos, Jeremy Kubica

Google, Inc.

To rank each ad: Take the dot product of  $\mathbf{a}$  and  $\mathbf{b}$  for each ad, give the highest-valued ads placement.

$$\mathbf{a} \cdot \mathbf{b} = \sum_{i=1}^n a_i b_i = a_1 b_1 + a_2 b_2 + \dots + a_n b_n$$

" $n$ " here is in the billions, but non-zero " $\mathbf{a}$ " and " $\mathbf{b}$ " values are in the thousands. This real-time system needs to exploit the sparsity to perform well.

A good candidate problem for an accelerator.

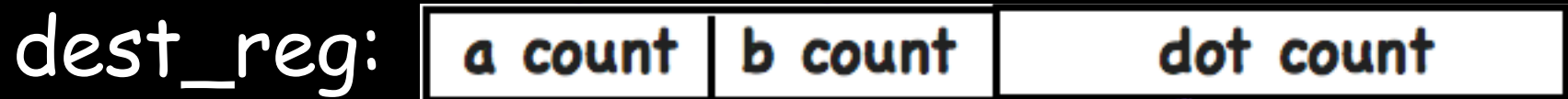
# One Instruction



SBDT dest\_reg, a\_reg, b\_reg

a\_reg: Holds 64-bit memory address pointing to the first byte of "a" list.

b\_reg: Holds 64-bit memory address pointing to the first byte of "b" list.



# of list elements.  
saturating 16-bit  
unsigned ints.

32-bit unsigned int

$$\mathbf{a \cdot b = \sum_{i=1}^n a_i b_i = a_1 b_1 + a_2 b_2 + \dots + a_n b_n}$$





Gabriel Cramer 1704-1752

# Cramer's Rule

Solve this  
matrix equation:

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

$A \quad x \quad b$

With determinants:

To solve for  $x_i$ :

Substitute  $b$  for  
column  $i$  in the  
numerator ...

$$x_i = \frac{\det(A_i)}{\det(A)} \quad i = 1, \dots, n$$

$$x_1 = \frac{\begin{vmatrix} b_1 & a_{12} \\ b_2 & a_{22} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}} = \frac{b_1 a_{22} - b_2 a_{12}}{a_{11} a_{22} - a_{12} a_{21}}$$
$$x_2 = \frac{\begin{vmatrix} a_{11} & b_1 \\ a_{21} & b_2 \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}} = \frac{b_2 a_{11} - b_1 a_{21}}{a_{11} a_{22} - a_{12} a_{21}}$$

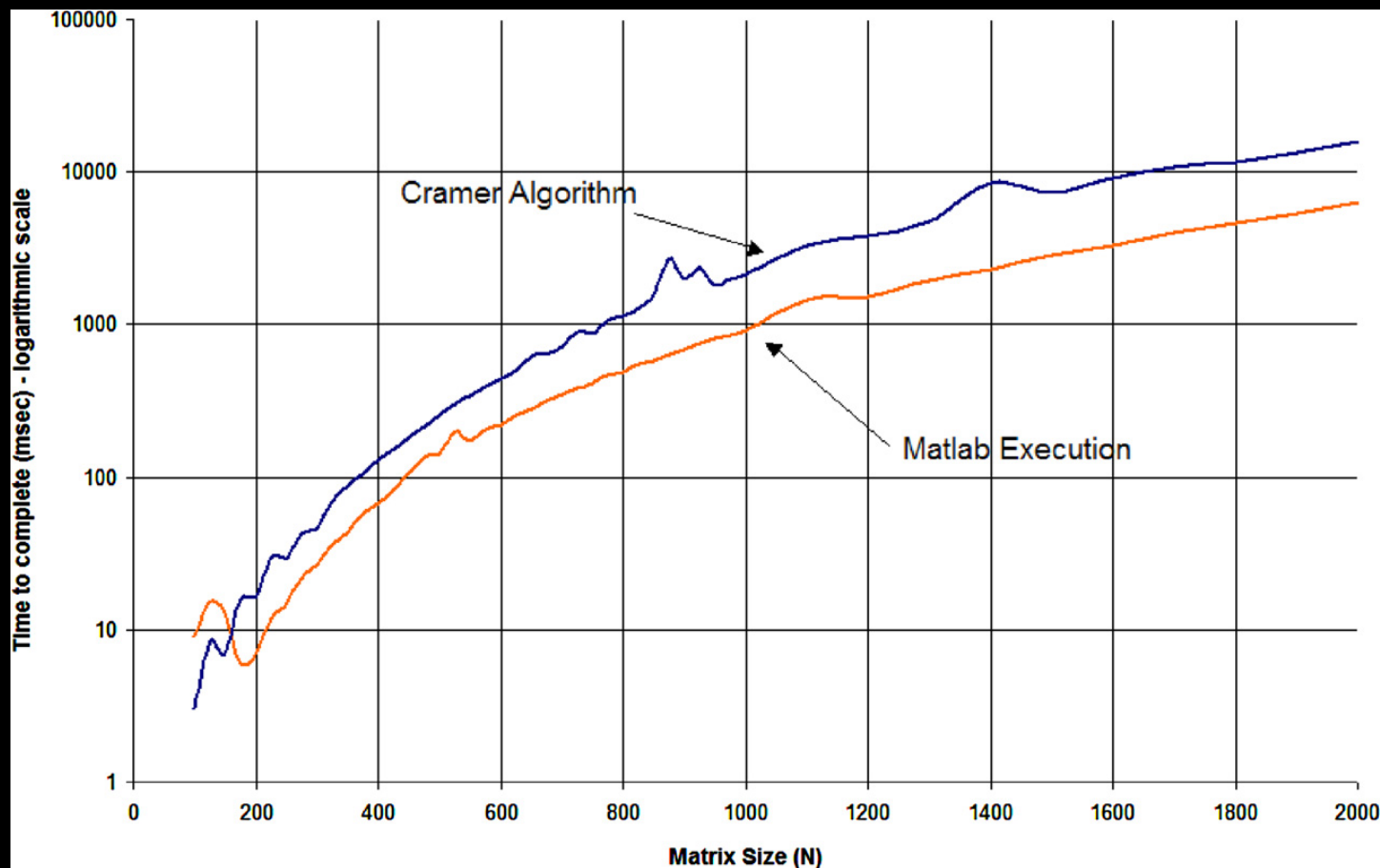
# Useful?

A condensation-based application of Cramer's rule for solving large-scale linear systems

Ken Habgood\*, Itamar Arel

*Department of Electrical Engineering and Computer Science, The University of Tennessee, Knoxville, TN, USA*

Recent work shows how to make Cramer's rule scale and be stable for large systems:



So, let's make an accelerator based on it ...

# Observations

\* Determinant works on a matrix, but returns a scalar. We use accelerator instructions to compute determinants, and let RISC-V compute the  $x$  vector by doing the divides.

$$x_i = \frac{\det(A_i)}{\det(A)} \quad i = 1, \dots, n$$

by accelerator by RISC-V

\* Determinants of matrices with integer coefficients can be computed exactly, with only integer multiplies and adds. So, we restrict our accelerator accordingly.

# Chió's Trick

For  $n = 3$ , computes determinant of a  $3 \times 3$  matrix by computing the  $2 \times 2$  determinant of four  $2 \times 2$  determinant results.

$$\det(A) = \frac{1}{a_{11}^{n-2}} \begin{vmatrix} \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} & \begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} & \dots & \begin{vmatrix} a_{11} & a_{1n} \\ a_{21} & a_{2n} \end{vmatrix} \\ \begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix} & \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} & \dots & \begin{vmatrix} a_{11} & a_{1n} \\ a_{31} & a_{3n} \end{vmatrix} \\ \vdots & \vdots & \ddots & \vdots \\ \begin{vmatrix} a_{11} & a_{12} \\ a_{n1} & a_{n2} \end{vmatrix} & \begin{vmatrix} a_{11} & a_{13} \\ a_{n1} & a_{n3} \end{vmatrix} & \dots & \begin{vmatrix} a_{11} & a_{1n} \\ a_{n1} & a_{nn} \end{vmatrix} \end{vmatrix}$$


Can be ignored (cancels out of  $Ax=b$ )

$$x_i = \frac{\det(A_i)}{\det(A)} \quad i = 1, \dots, n$$

We can reuse the "leaf"  $2 \times 2$  determinants when we compute the full set of  $\det(A_i)$  and  $\det(A)$ .

# Reuse

Color coding shows reuse.



$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{21} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$$

Implicit architected state for the accelerator can be used to store and reuse partial results ...

$$\det A = \begin{vmatrix} (a_{11} a_{22} - a_{12} a_{21}) & (a_{11} a_{23} - a_{13} a_{21}) \\ (a_{11} a_{32} - a_{12} a_{31}) & (a_{11} a_{33} - a_{13} a_{31}) \end{vmatrix}$$

$$\det A_2 = \begin{vmatrix} (a_{11} b_2 - b_1 a_{21}) & (a_{11} a_{23} - a_{13} a_{21}) \\ (a_{11} b_3 - b_1 a_{31}) & (a_{11} a_{33} - a_{13} a_{31}) \end{vmatrix}$$

$$\det A_3 = \begin{vmatrix} (a_{11} a_{22} - a_{12} a_{21}) & (a_{11} b_2 - b_1 a_{21}) \\ (a_{11} a_{32} - a_{12} a_{31}) & (a_{11} b_3 - b_1 a_{31}) \end{vmatrix}$$

## Two instructions

First, it clears all implicit state



**DETA** dest\_reg, a\_reg, len\_reg

a\_reg: 64-bit memory address of  $A$  matrix.

len\_reg: Holds the  $n$  of the  $n \times n$   $A$  matrix.

dest\_reg: Return register for  $\det(A)$ .

Retains  $n$  and partial results for  $A$ .



**DETAI** dest\_reg, b\_reg, col\_reg

b\_reg: 64-bit memory address of  $b$  vector.

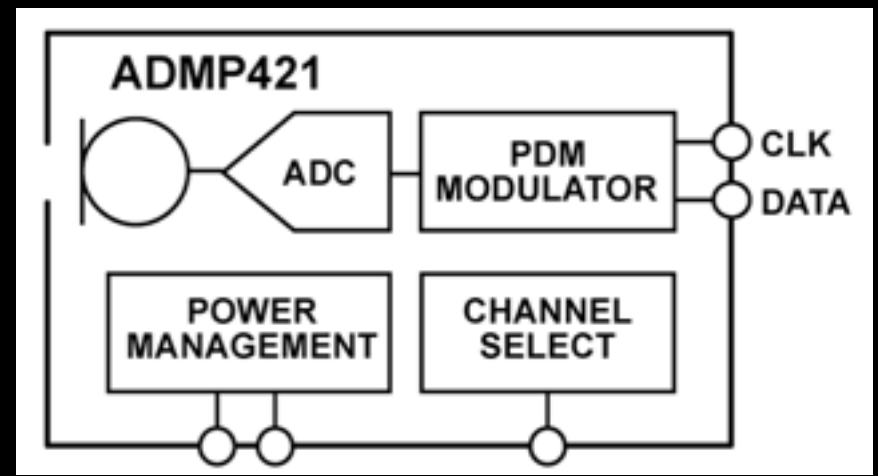
col\_reg: The " $i$ " (column) for  $\det(A_i)$

dest\_reg: Return register for  $\det(A_i)$ .

Adds to partial results (for  $A_i$ ).



# MEMS microphone post-processing accelerator



PDM\_Data\_From MEMS\_microphon

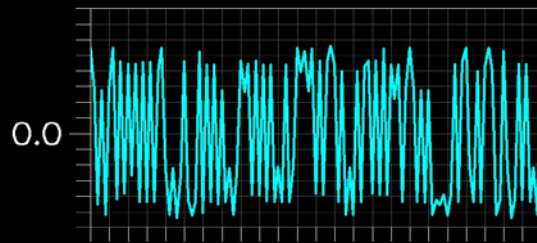


Figure 8. PDM data coming out of microphone

CIC\_0/P

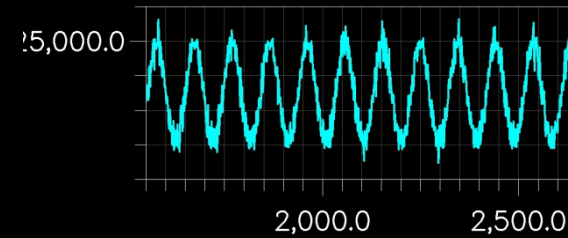
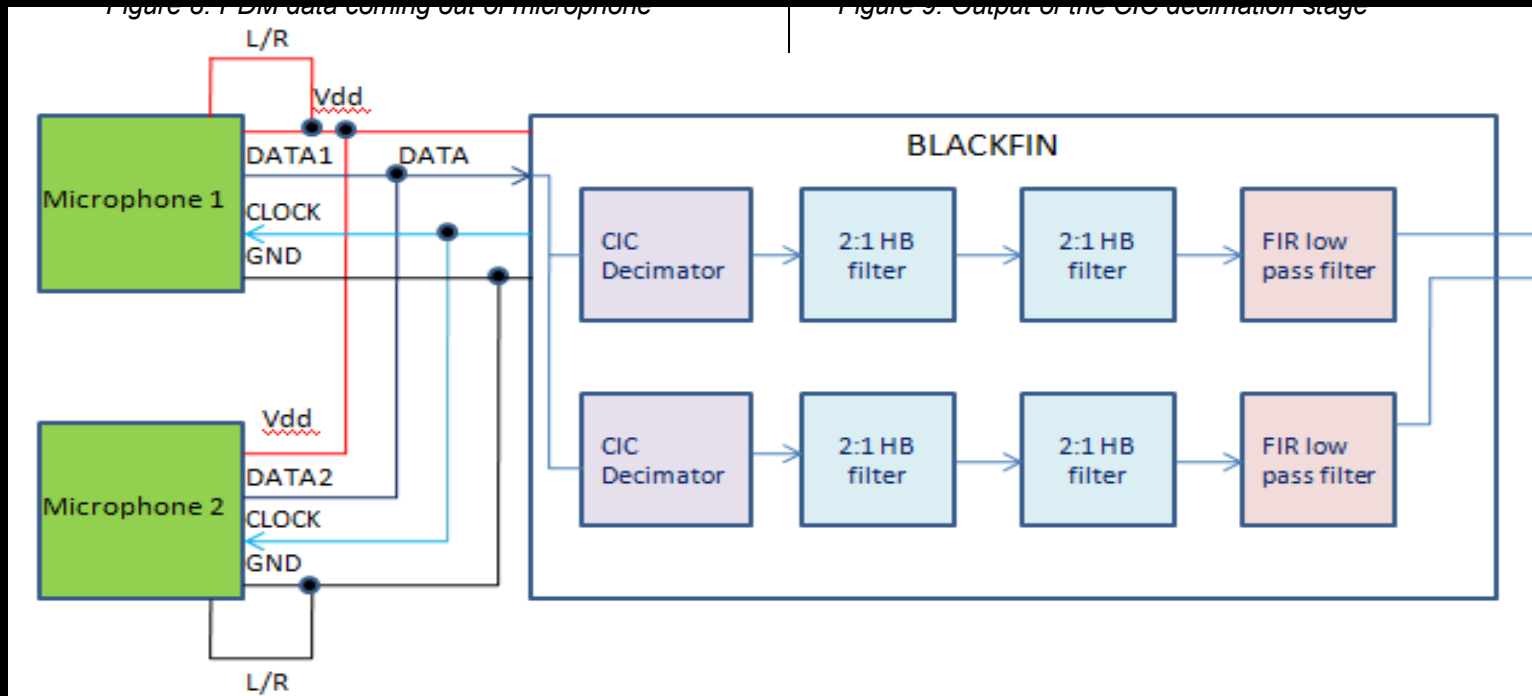


Figure 9. Output of the CIC decimation stage



# Quick Ideas

## A 280 mV-to-1.1 V 256b Reconfigurable SIMD Vector Permutation Engine With 2-Dimensional Shuffle in 22 nm Tri-Gate CMOS

Steven K. Hsu, *Member, IEEE*, Amit Agarwal, *Member, IEEE*, Mark A. Anders, *Member, IEEE*,

## Thin Servers with Smart Pipes: Designing SoC Accelerators for Memcached

Kevin Lim  
HP Labs

David Meisner  
Facebook

Ali G. Saidu  
ARM R&D

## Systolic Sorting on a Mesh-Connected Network

HANS-WERNER LANG, MANFRED SCHIMMLER,  
HARTMUT SCHMECK, AND HEIKO SCHRÖDER

## FINITE AUTOMATA BASED COMPRESSION OF BI-LEVEL AND SIMPLE COLOR IMAGES

KAREL CULIK II<sup>†</sup> and VLADIMIR VALENTA

Department of Computer Science, University of South Carolina, Columbia, SC 29208, U.S.A.  
*e-mail:* culik@cs.sc.edu

## Convolution Engine: Balancing Efficiency & Flexibility in Specialized Computing

Wajahat Qadeer, Rehan Hameed, Ofer Shacham,  
Preethi Venkatesan, Christos Kozyrakis, Mark A. Horowitz

# Project Proposal

|        |                        |                                 |  |   |
|--------|------------------------|---------------------------------|--|---|
| 11-Feb | Lecture 8              | Hardware design patterns II     | Lab 3 out : SRAMs + multi Vt + ICC + PrimeTime | Lab 2 due<br>Preliminary Project Proposal due |
| 16-Feb | Lecture 9              | Advanced Chisel                 |  |   |
| 18-Feb | Lecture 10             | Testing and Design Verification |  |   |
| 23-Feb | Oral Project Proposals |                                 |  |   |
| 25-Feb | Oral Project Proposals |                                 | Lab 4 out : Rocket Processor Interface         | Lab 3 due                                     |

- ▶ **Email me a brief description of a project idea (one paragraph):**
  - ▶ Describe the function
  - ▶ Why you are interested (does this relate to your research or another class?)
  - ▶ Any background you have related to the project
  - ▶ Partner
  - ▶ More than one idea is fine