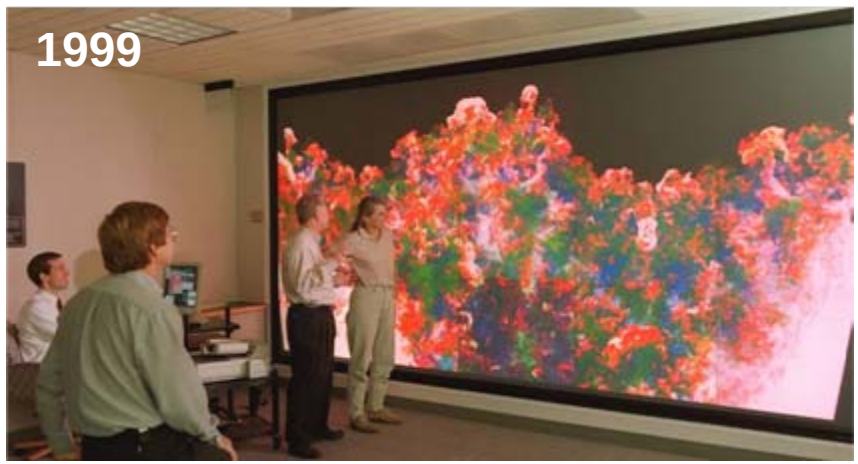
Tom DeFanti
Research Scientist
California Institute for Telecommunications and Information Technology
University of California, San Diego
Distinguished Professor Emeritus of Computer Science
University of Illinois at Chicago

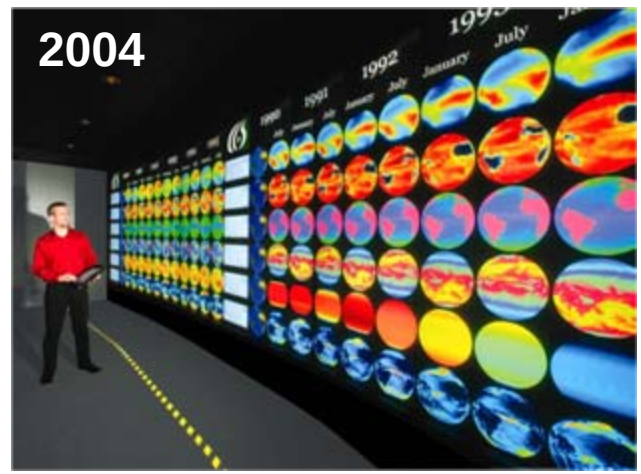
Green Power

10 years of Bringing Scalable Visualization to the Users

NCSA 4 MPixel
NSF Alliance PowerWall



LLNL 20 Mpixel Wall



ORNL 35Mpixel EVEREST

EVL 100 Mpixel LambdaVision
NSF MRI

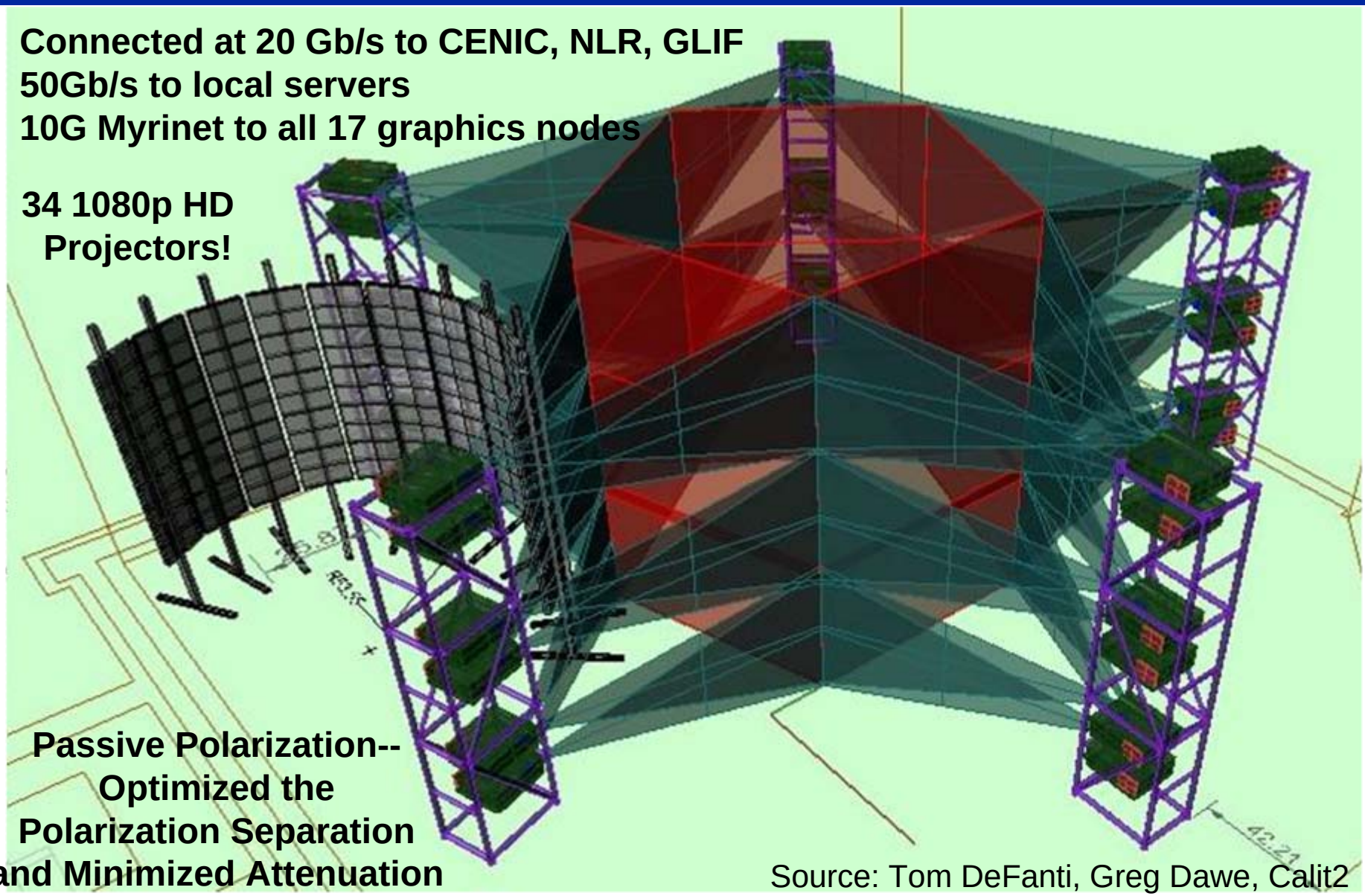
Calit2@UCI 200 Mpixel HiPerWall
NSF MRI

TACC 307 Mpixel Stallion
NSF TeraGrid

Two Orders of Magnitude Pixel Growth
(Same as 100mb/s-->10Gb/s!)

Power Consumption: Calit2's 300+ Megapixel 25kW? HiPerSpace OptIPortal

Calit2's StarCAVE
Stereo 34 Megapixel per eye 25kW? OptIPortal

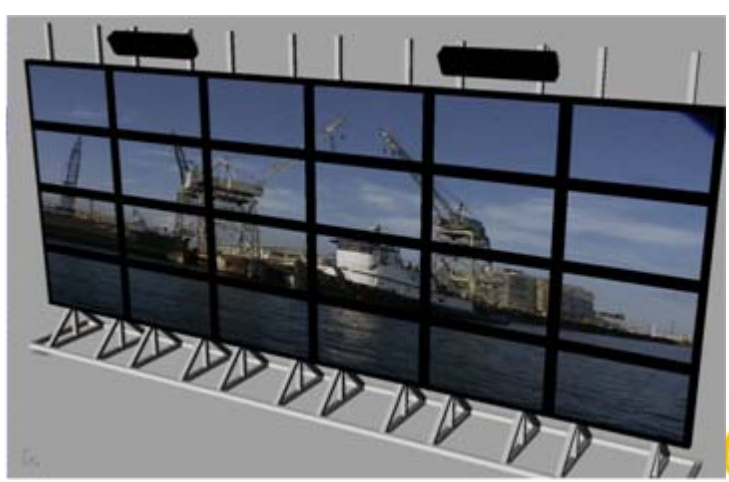


Cluster with 34 Nvidia 5600 cards--51 GB GPU Memory
(25kW is $2.5/hr, $60/day, less than renting a car)

Lenticular 3D Autostereo Alioscopy Display 400w per tile?

New work on Tiled Autostereo displays made by Alioscopy, Inc.

Dan Sandin, Bob Kooima, Andrew Prudhomme, Tom DeFanti



12kW?

evl electronic visualization laboratory

it²

Tiled Micropolarized (Xpol) 3D JVC Panels (w/Glasses)

Image for **left eye** (**1080i/60** interlace scan)

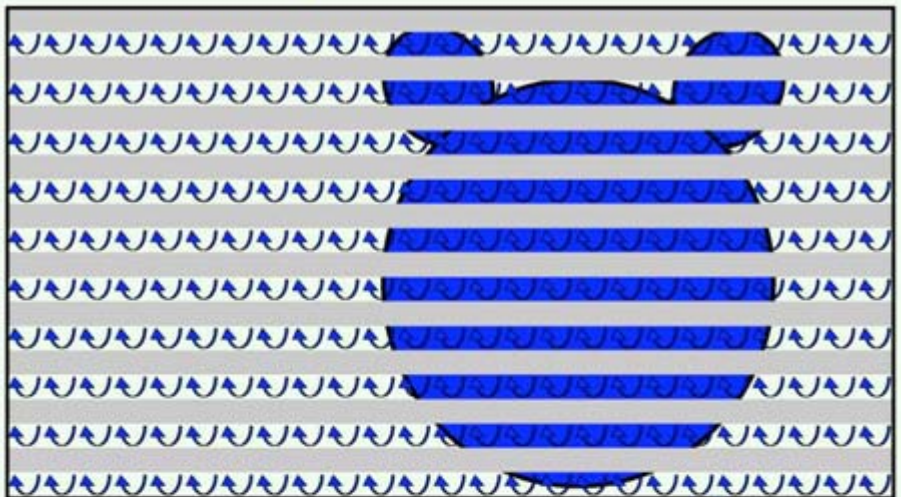
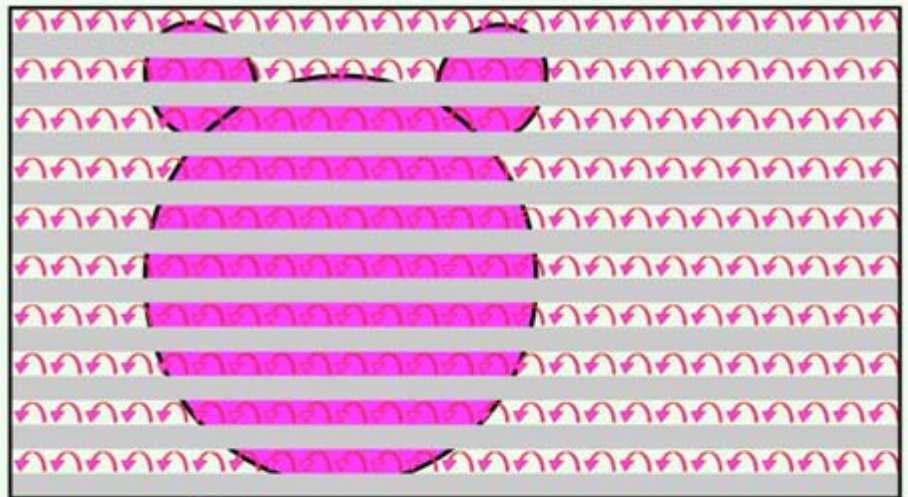
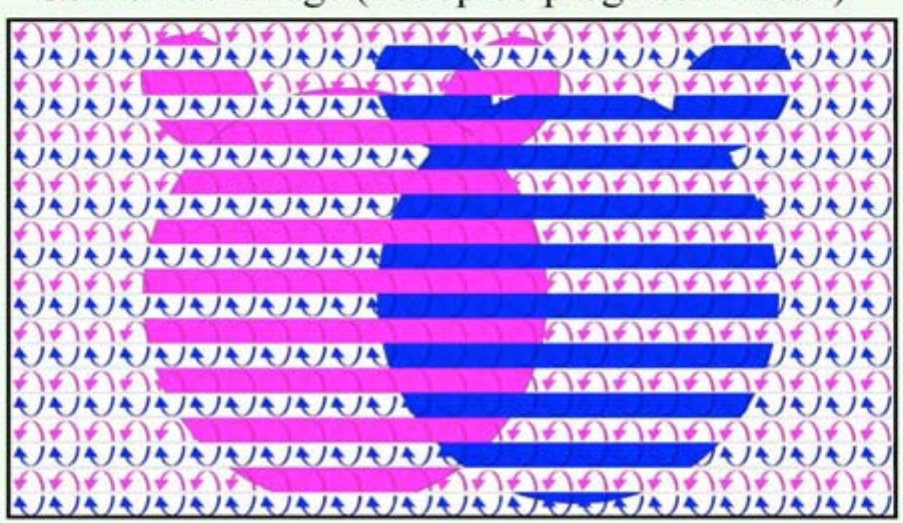


Image for **right eye** (**1080i/60** interlace scan)



Combined Image (**1080p/60** progressive scan)

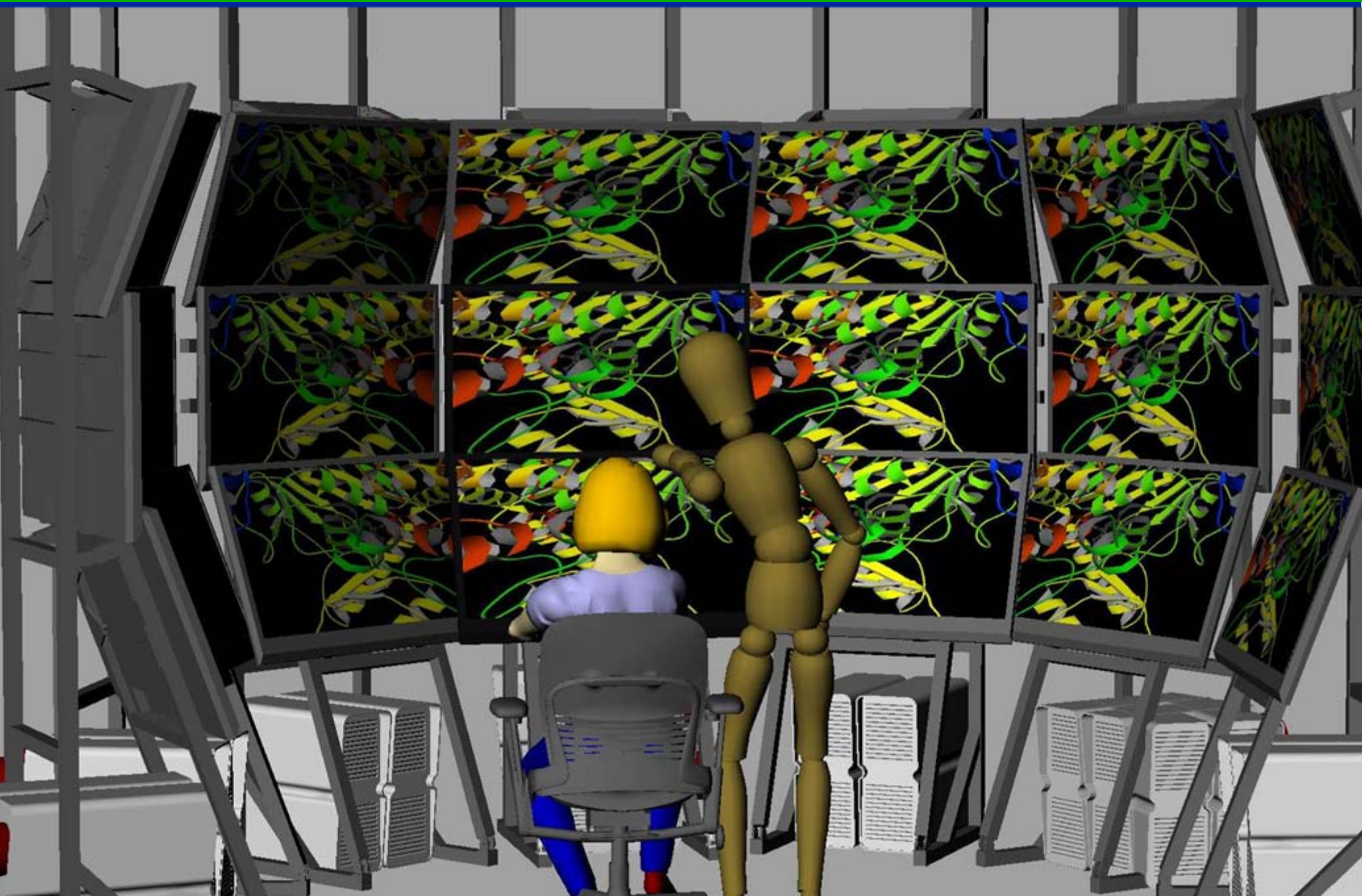


Polarizing Glasses filter the light so the viewer can perceive a correct 3D image

The **Xpol™** is attached on the surface of the LCD HDTV Display alternately

Source: NHK Media Technology, Inc.

21-Panel Xpol LCD Stereo CAVE 9kW?



GreenLight Project: Focus on University Closet Clusters

- Compute energy/rack : 2 kW (2000) to 30kW (in 2010)
- Cooling and power issues now a major factor in campus clusters
- But academic clusters are often small: departmental closets
- **Energy use of departmental computers is increasing fast creating crises of space, power, and cooling**
- **Unfortunately, almost nothing is known about how to make these shared virtual clusters energy efficient, since there has been no financial motivation to do so**

The NSF-Funded GreenLight Project
Giving Users Energy-Known Compute and Storage Options

7 Racks plus Network

Takes up 2 Parking Spaces

Data Power Cooling

- **UCSD Structural Engineering Dept.**
  Conducted Sun MDC Tests May 2007
- **UCSD Bought Two Sun MDCs in May 2008**
- **Now populated with PCs, FPGAs, PUs, 250TB disk**

$2M NSF-Funded GreenLight Project

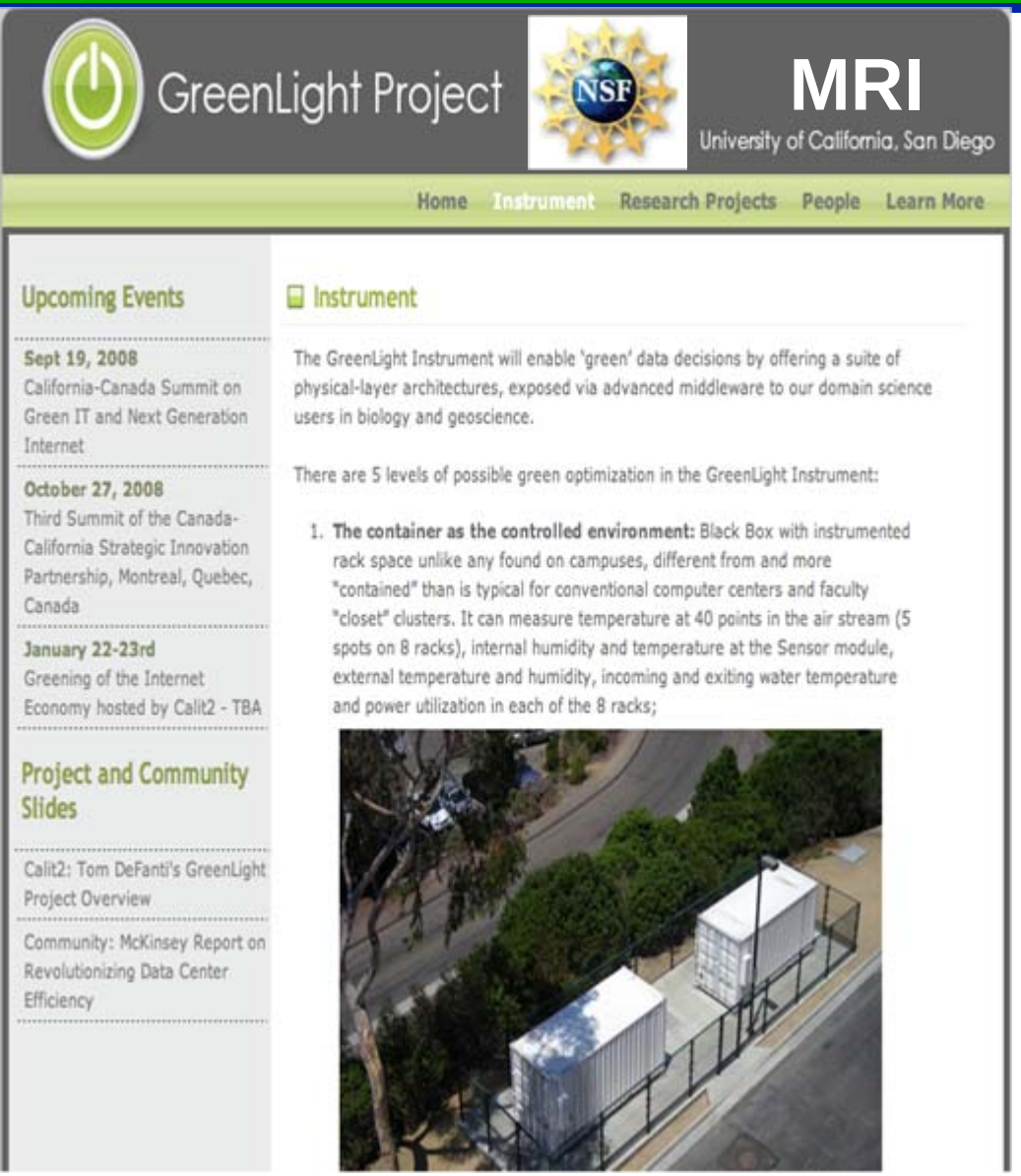
Earthquake Testing

The GreenLight Project: Instrumenting the Energy Cost of Cluster Computing

- **Focus on 5 Communities with At-Scale Computing Needs:**
  - **Metagenomics**
  - **Ocean Observing**
  - **Microscopy**
  - **Bioinformatics**
  - **Digital Media**

- **Measure, Monitor, & Web Publish Real-Time Sensor Outputs**
  - **Via Service-oriented Architectures**
  - **Allow Researchers Anywhere To Study Computing Energy Cost**
  - **Enable Scientists To Explore Tactics For Maximizing Work/Watt**
- **Develop Middleware that Automates Optimal Choice of Compute/RAM Power Strategies for Desired Greenness**

Answer the Threat to Cluster Deployment with Facts on Work/Watt



http://greenlight.calit2.net

- Computer Architecture
  - Rajesh Gupta/CSE
- Software Architecture
  - Amin Vahdat, Ingolf Kruger/CSE
- CineGrid Exchange
  - Tom DeFanti/Calit2
- Visualization
  - Falko Kuster/Structural Engineering
- Power and Thermal Management
  - Tajana Rosing/CSE
- Analyzing Power Consumption Data
  - Jim Hollan/Cog Sci
- Direct DC Datacenters
  - Tom DeFanti, Greg Hidley


PDUUsage
Phase I
Phase II
Phase III
PDU10.10.10.15-Phase1
100
50
0
4/1 17:24
4/1 17:25
4/1 17:26
4/1 17:27
4/1 17:28
4/1 17:29
4/1 17:30
4/1 17:31
4/1 17:32
4/1 17:33

Example: Improve Mass Spectrometry's Green Efficiency By Matching Algorithms to Specialized Processors

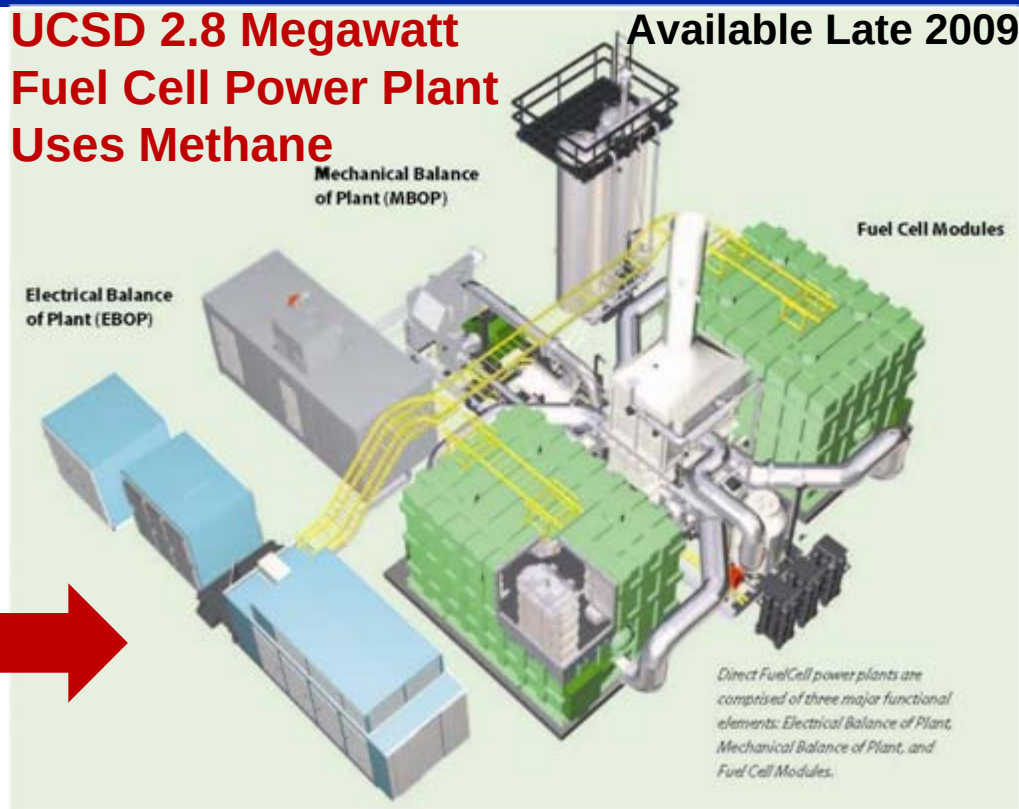
- **Inspect application implements the very computationally intense MS-Alignment algorithm for discovery of unanticipated rare or uncharacterized Post-Translational Modifications**
- **Solution: Hardware acceleration with a FPGA-based co-Processor (Convey Architecture)**
- **Results:**
  - **300x Speedup with hand FPGA coding**
  - **Increase in work/watt?**
  - **Increase in work/$ (purchase, life-cycle)?**

DC Power--UCSD is Installing Zero Carbon Emission Solar and Fuel Cell DC Electricity Generators

San Diego's Point Loma Wastewater Treatment Plant Produces Waste Methane

UCSD 2.8 Megawatt Fuel Cell Power Plant Uses Methane

Available Late 2009



2 Megawatts of Solar Power Cells Being Installed

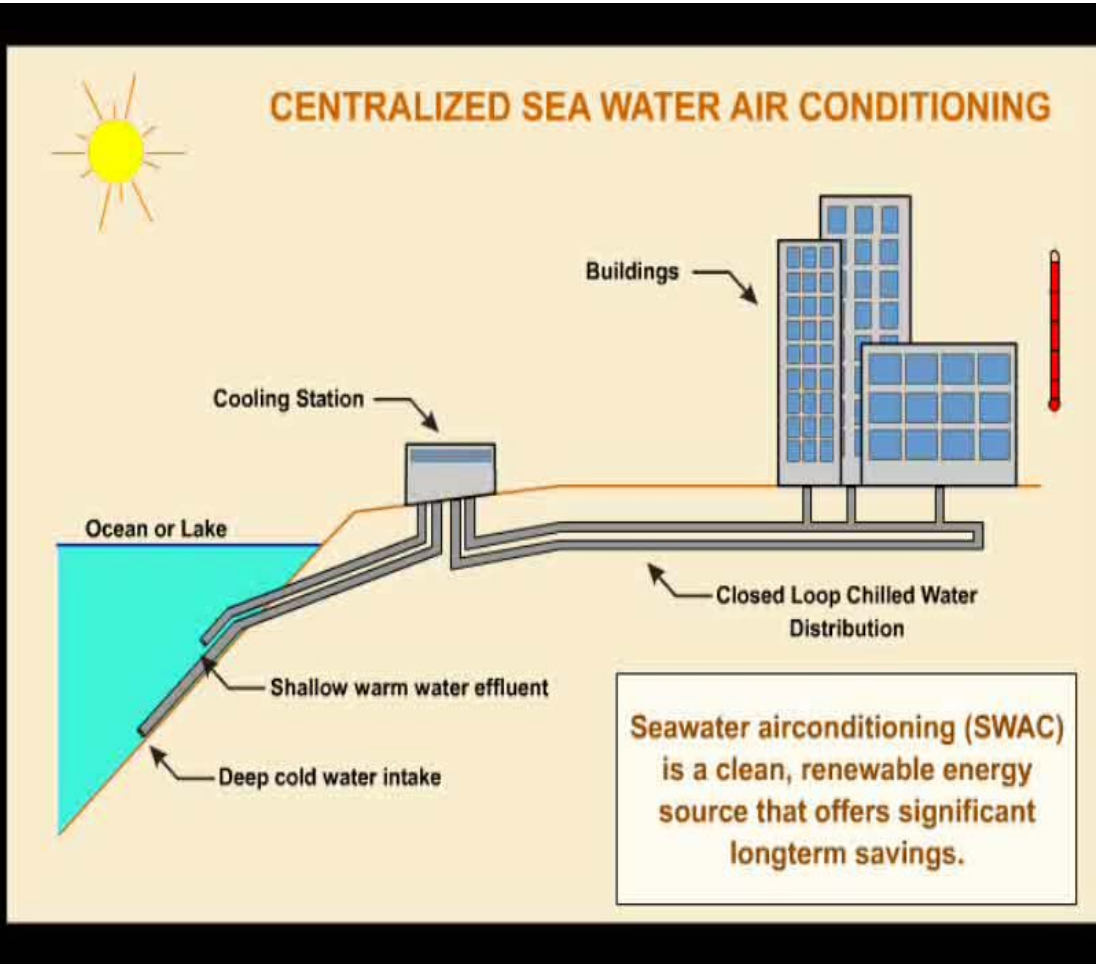
GreenLight Experiment: Direct 400v DC-Powered Modular Data Center

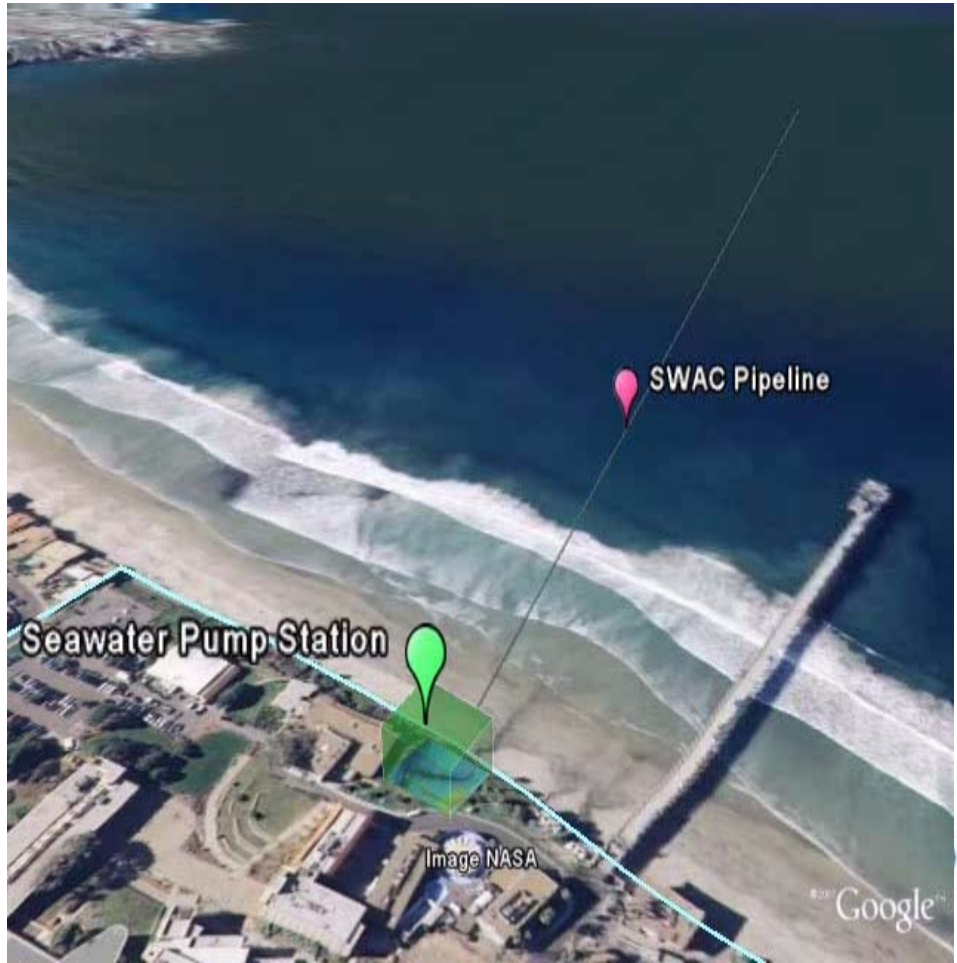
- **Concept—Avoid DC to AC to DC Conversion Losses**
  - Computers use DC power internally
  - Solar and fuel cells produce DC
  - Both plug into the AC power grid
  - Can we use DC directly?
  - Scalable/Distributable?
- **DC Generation Can Be Intermittent**
  - Depends on source
    - Solar, Wind, Fuel Cell, Hydro
  - Can use sensors to shut down or sleep computers
  - Can use virtualization to halt/shift Jobs
  - Can switch to AC as backup

UCSD DC Fuel Cell 2800kW
Sun MDC <100-200kW

UCSD is Studying a Global Demonstration Project for Sea Water Cooling

- **UCSD is uniquely located to use cold seawater from one of only 40 deep shoreline sites in the world. It can supply cold water essential for air conditioning laboratories and computer rooms**
- **Initial study of La Jolla underwater trench suggests a seawater cooling system could produce savings of $4M/yr and 100 million gallons of fresh water per year.**





More: Application of ICT Can Lead to a 5-Fold Greater Decrease in GHGs Than its Own Carbon Footprint

While the ICT sector plans to significantly step up the energy efficiency of its products and services, ICT's largest influence will be by enabling energy efficiencies in other sectors, an opportunity that could deliver carbon savings five times larger than the total emissions from the entire ICT sector in 2020.

--Smart 2020 Report

Major Opportunities for the United States*

- Smart Electrical Grids
- Smart Transportation Systems
- Smart Buildings
- Virtual Meetings

*** Smart 2020 United States Report Addendum**

www.smart2020.org

To be Energy Efficient, We Must Think about Koala-Style Computing and other "Smart" Ways

Size Your Brain Power, Visualization, Storage, Sleep Cycles, and Communications to Your Problem

HPDC: Thank You Very Much!

- **My planning, research, and education efforts are made possible, in major part, by funding from:**
  - US National Science Foundation (NSF) awards ANI-0225642, EIA-0115809, SCI-0441094, and CNS 0821155
  - State of California, Calit2 UCSD Division
  - State of Illinois I-WIRE Program, and major UIC cost sharing
- **University of Illinois at Chicago, Argonne National Laboratory, and Northwestern University for StarLight networking and management**
- **National Lambda Rail, Pacific Wave and CENIC**
- **NTT Network Innovations Lab**
- **Cisco Systems, Inc.**
- **Pacific Interface, Inc.**
- **Darkstrand, Inc.**
- **KAUST-US**
- **Sharp Labs of America**

How to Solve the Power Wall Problem of Supercomputing (in 2015)

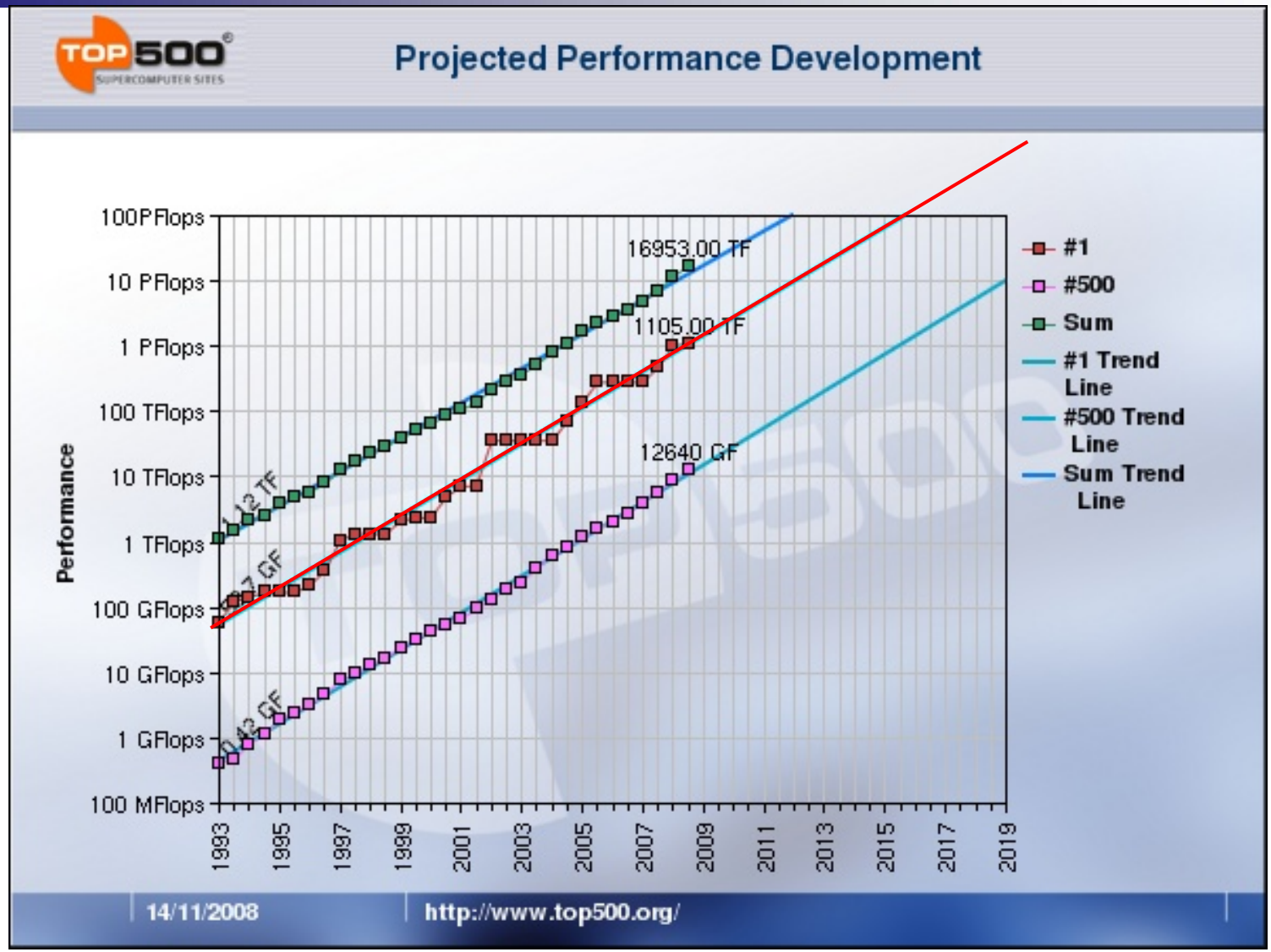
Hiroshi Nakamura

nakamura@hal.rcast.u-tokyo.ac.jp

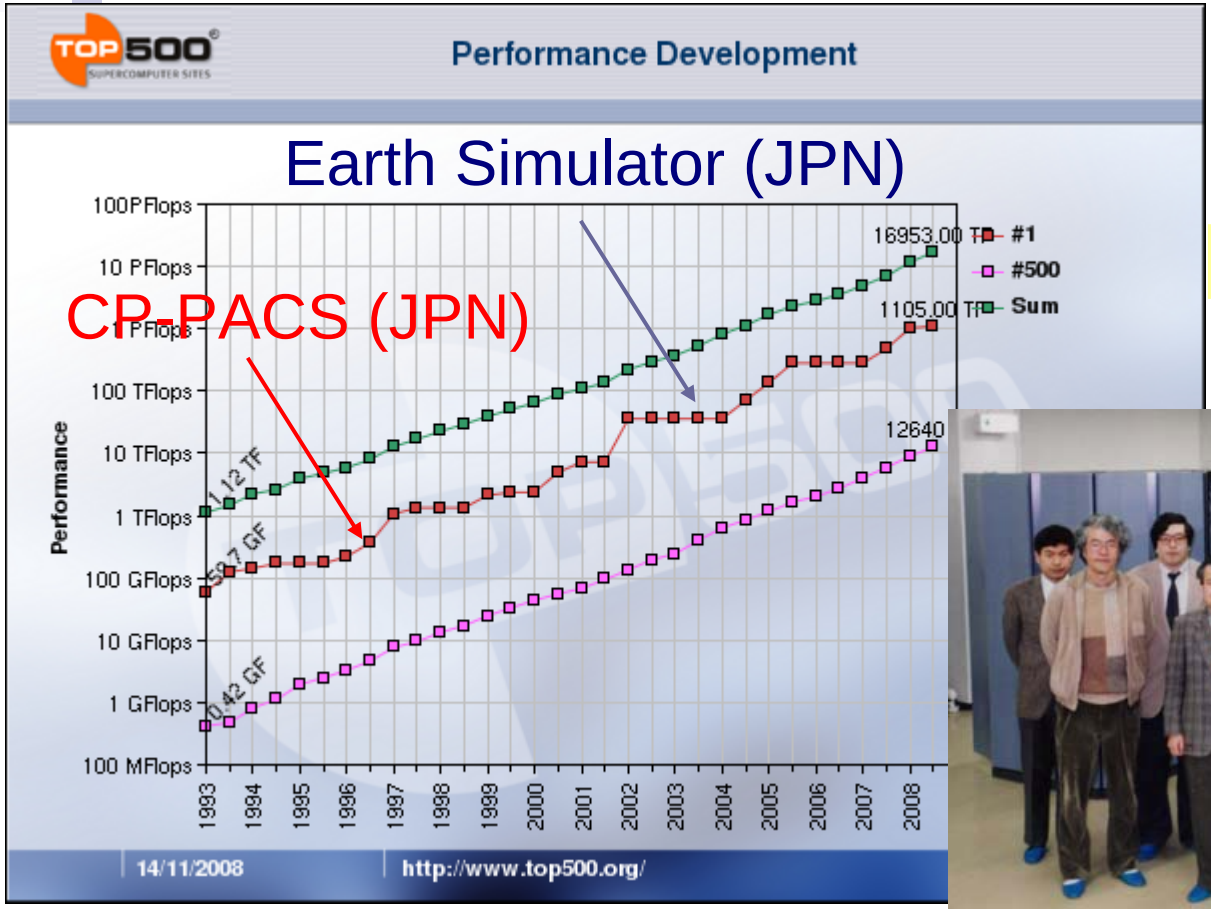
(nakamura@acm.org, hiroshi@computer.org)

The University of Tokyo

Can we achieve 100PFLOPS by 2015? or 1 ExaFLOPS by 2019?



My Standpoint: Computer Architecture Researcher



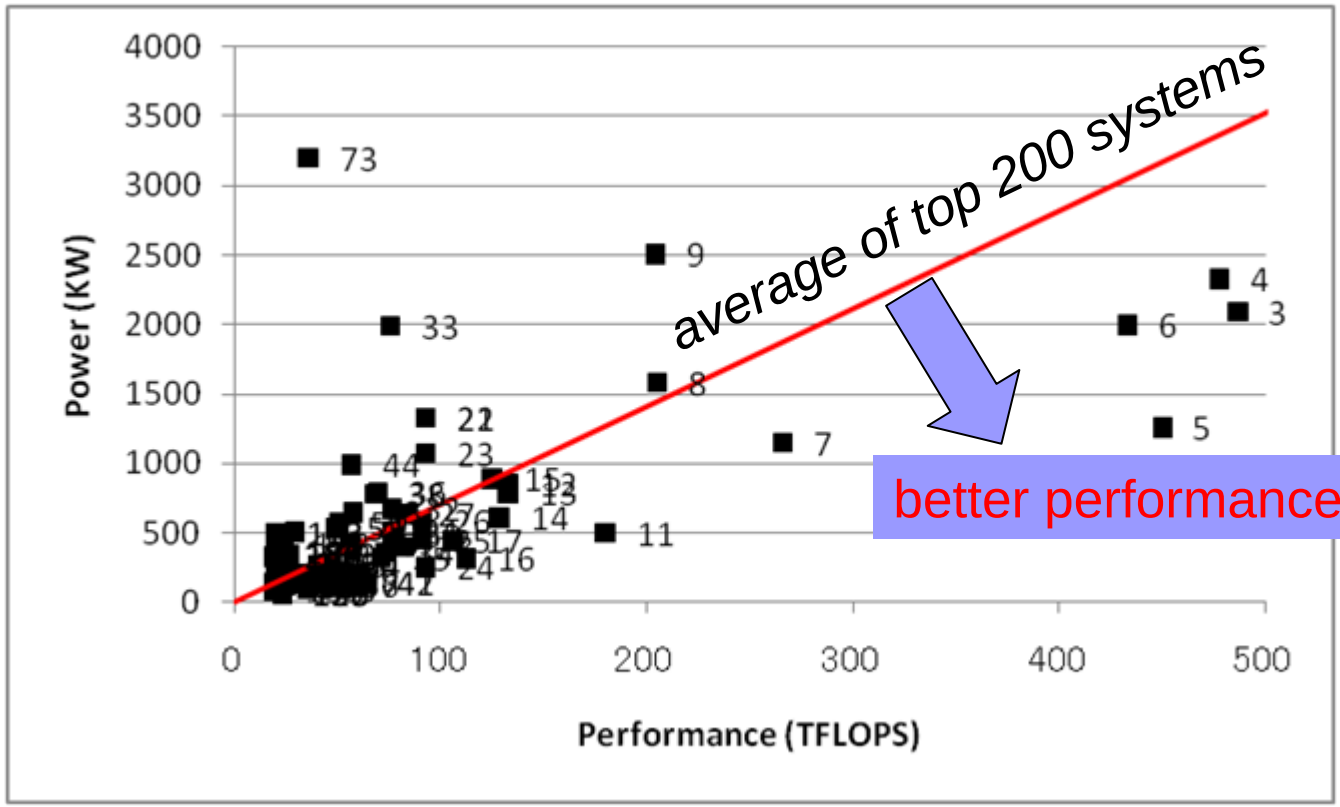
I am here !

not scaled X10 every 4 yrs ☹

- I am a “high-performance and low-power computer” architect

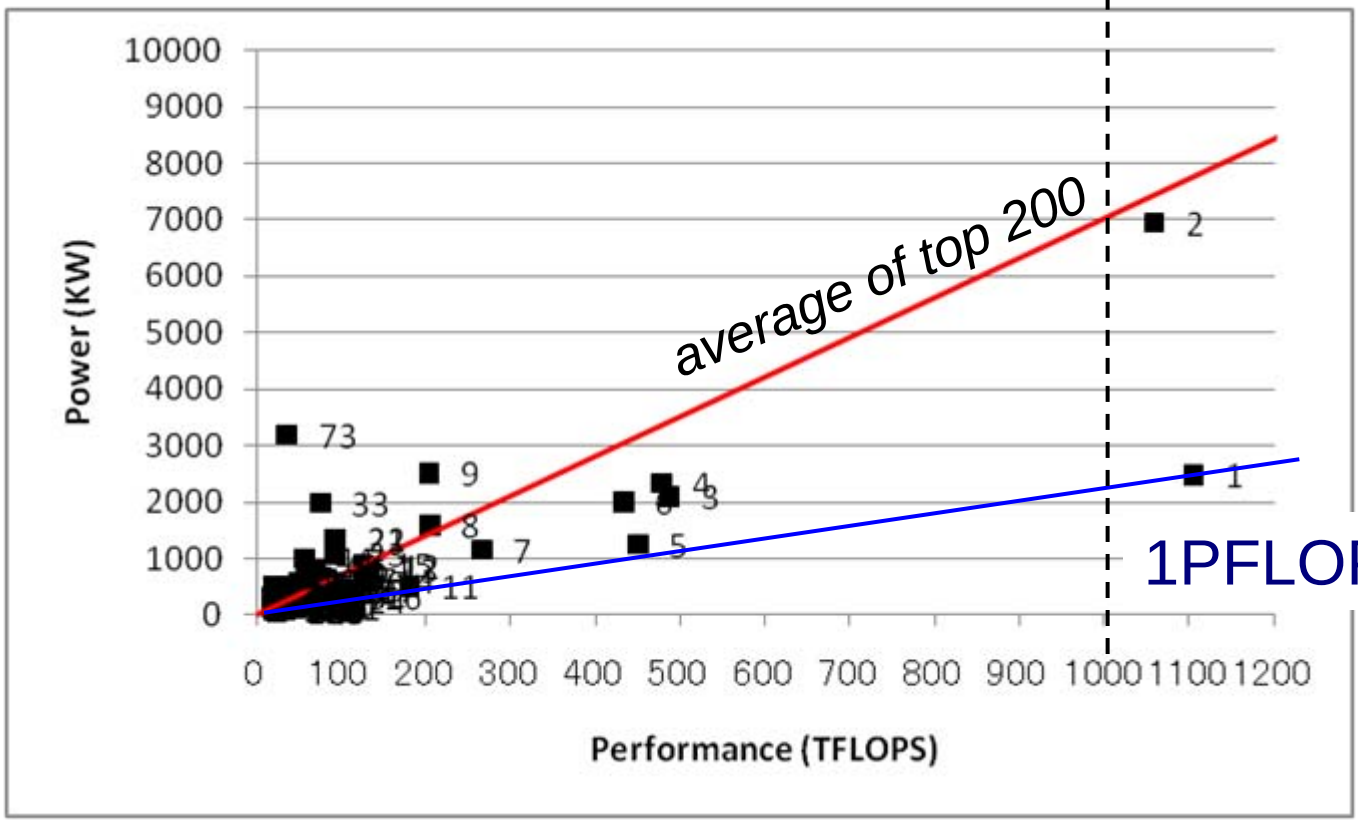
Performance/Power of Top200 Machines

www.top500.org



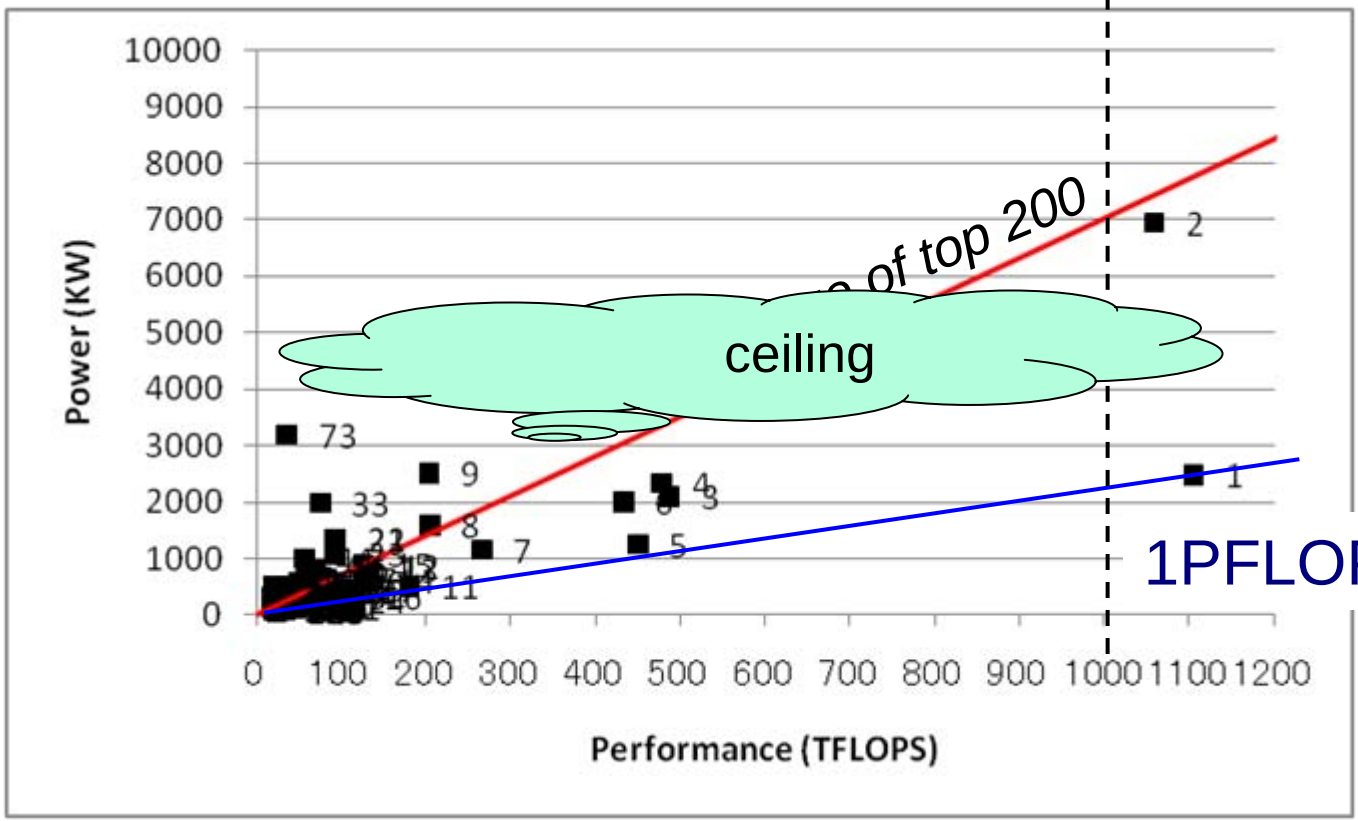
- Faster systems exhibit good performance/power ratios

Performance/Power of Top Machines



- Rank 1 and 5 systems have very good ratios
  - roadrunner (rank1), BlueGene/P (rank5)
- There seems to be a ceiling on the total power consumption = cooling limitation

Performance/Power of Top Machines



- Rank 1 and 5 systems have very good ratios
  - roadrunner (rank1), BlueGene/P (rank5)
- There seems to be a ceiling on the total power consumption = cooling limitation

Then, the Question is …

- How can we improve performance/power ratio 100X till then ?

Recap: How performance/power ratio has improved recently ?

- Voltage & Parallelism Scaling
  - for a transistor,

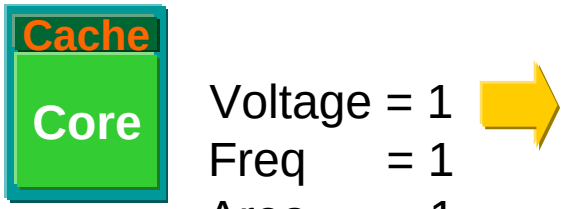
Switching Delay $t_{delay} \propto \frac{CV_{DD}}{(V_{DD} - V_{th})^{\alpha}}$

Dynamic Power $P_{dyn} = C\, V_{DD}^{2}\, f\ \alpha$

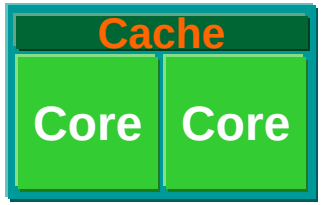
  - for a processor core

Rule of thumb

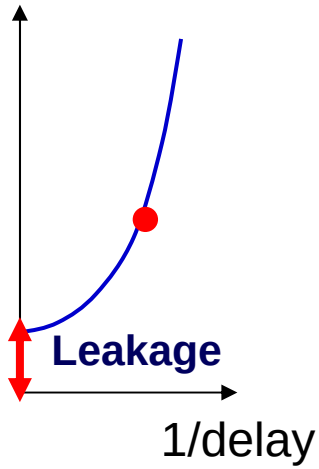
Voltage	Frequency	Power	Performance
1%	1%	3%	0.66%



Voltage = 1
Freq = 1
Area = 1
Power = 1
Performance = 1

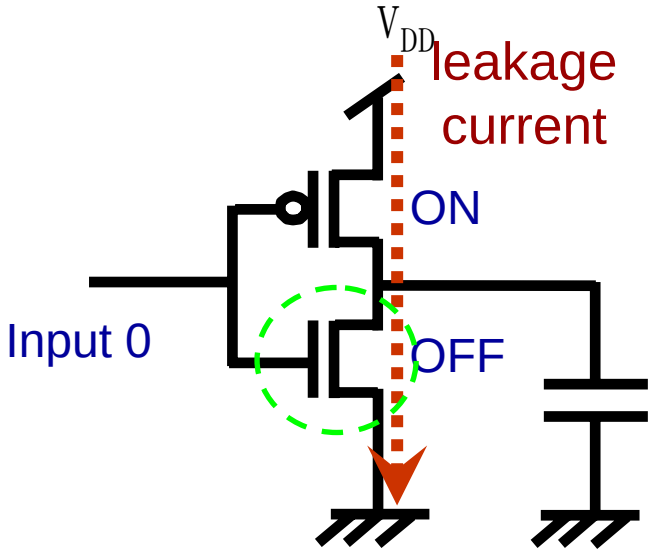


Voltage = -15%
Freq = -15%
Area = 2
Power = 1
Performance = ~1.8

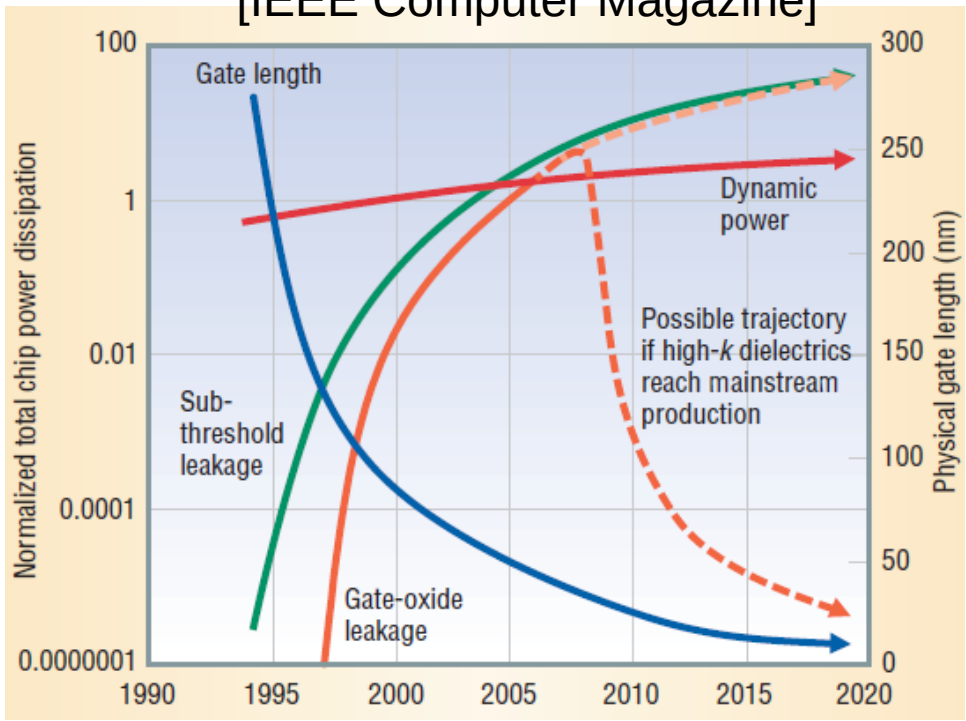


Voltage Scaling is no longer a Promising Way

- Limitations on Semiconductor Technology
  - Vth limitation, soft error, process variation
    → hard to implement reliable systems
- Leakage Problem
  - power consumption regardless of switching

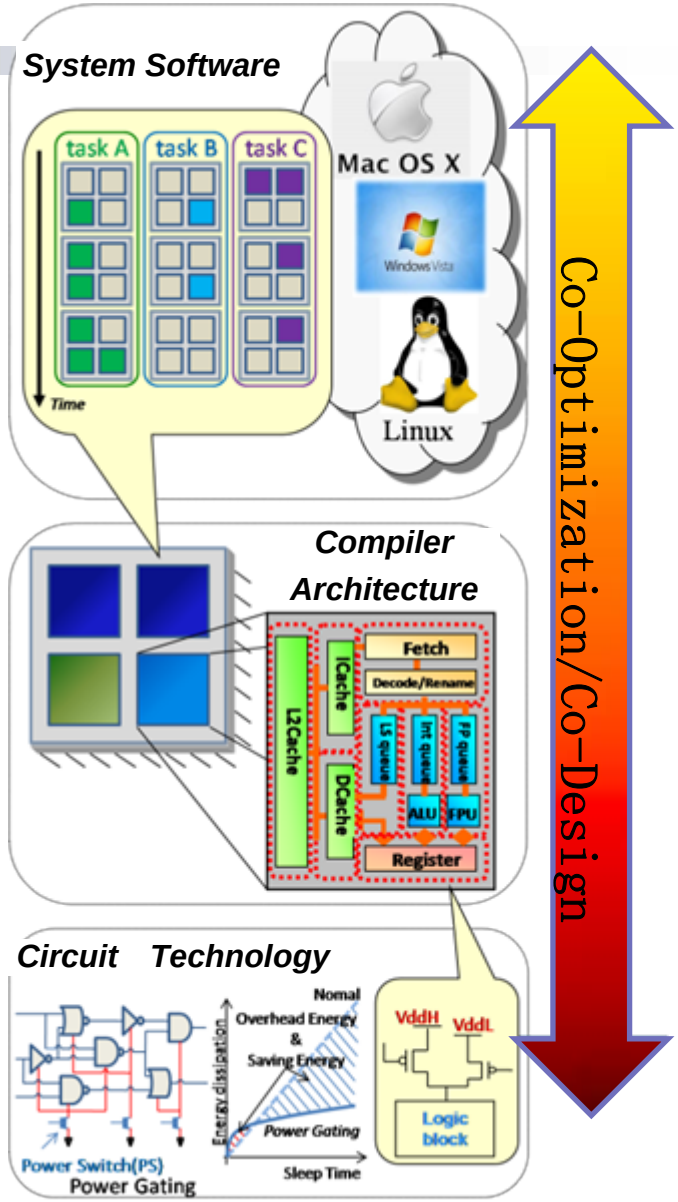


[IEEE Computer Magazine]



My Research Project

- Innovative Power Control for Ultra Low-Power and High-Performance System LSIs
  - 5 years project started from Oct, 2006
  - Supported by JST (Japan Science and Technology Agency)
  - CREST (Core Research for Evolutional Science and Technology) Program
    - Technology Innovation and Integration for Information Systems with Ultra Low Power
    - 12 projects are under operation
- Strategy: innovative power control through tight Co-Optimization / Co-Design of system software, architecture, and circuit technology.



Suggestion 1: ASAP Architecture

- Make Architecture As Simple As Possible
  - Why Simple Cores (BlueGene) or Accelerators (Roadrunner) contribute to Performance/Power Ratio ?
- NOT, Required Energy for Arithmetic Computations
  - they are the same, depend only on the technology
  - fortunately continue to decrease by enjoying the benefits of technology scaling
- BUT, Required Energy for Supplying Data to Arithmetic Units
  - Simple Control (SIMD: Instruction Handling)
  - Simple Data Movement (less flexible but simple Register/Memory Access)

→ Effective Data-Bandwidth/Power is important

Suggestion 2: Memory Hierarchy Optimization

- Memory Hierarchy has been optimized for performance
- Are they also the best for power?
- Example: Cache Decay [Kaxiras@ISCA01]
  - target: leakage power reduction
- dead time: time interval between the last reference and its eviction
- fraction of dead time : 65% integer, 80% floating point (Spec2000)

→ simple LRU replacement holds unnecessary data

→ waste of power

→ significant leakage reduction by power gating those lines

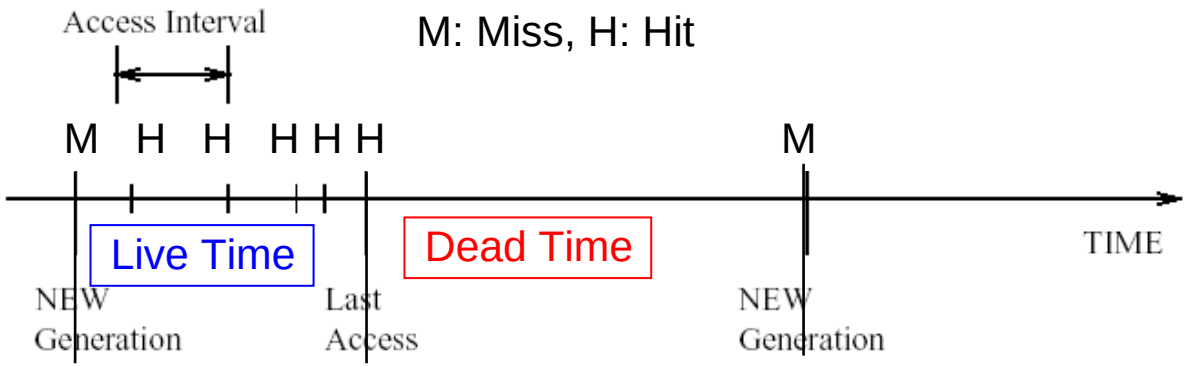


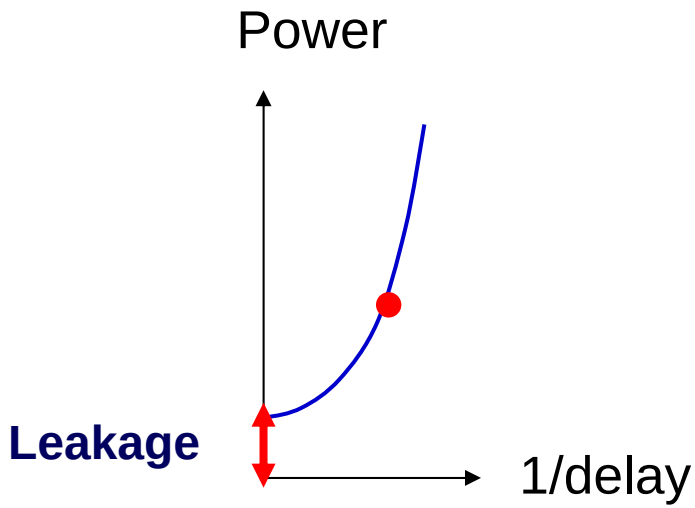
Figure 1. Cache generations in a reference stream.

Suggestion 3: Innovative Power Control
~Innovative Power/Performance Throttling~

- For a transistor,

$$\text{Switching Delay} \propto \frac{CV_{DD}}{(V_{DD}-V_{th})^{\alpha}}$$

$$\text{Dynamic Power } P_{dyn} = C\,V_{DD}^{2}\,f\,\alpha$$



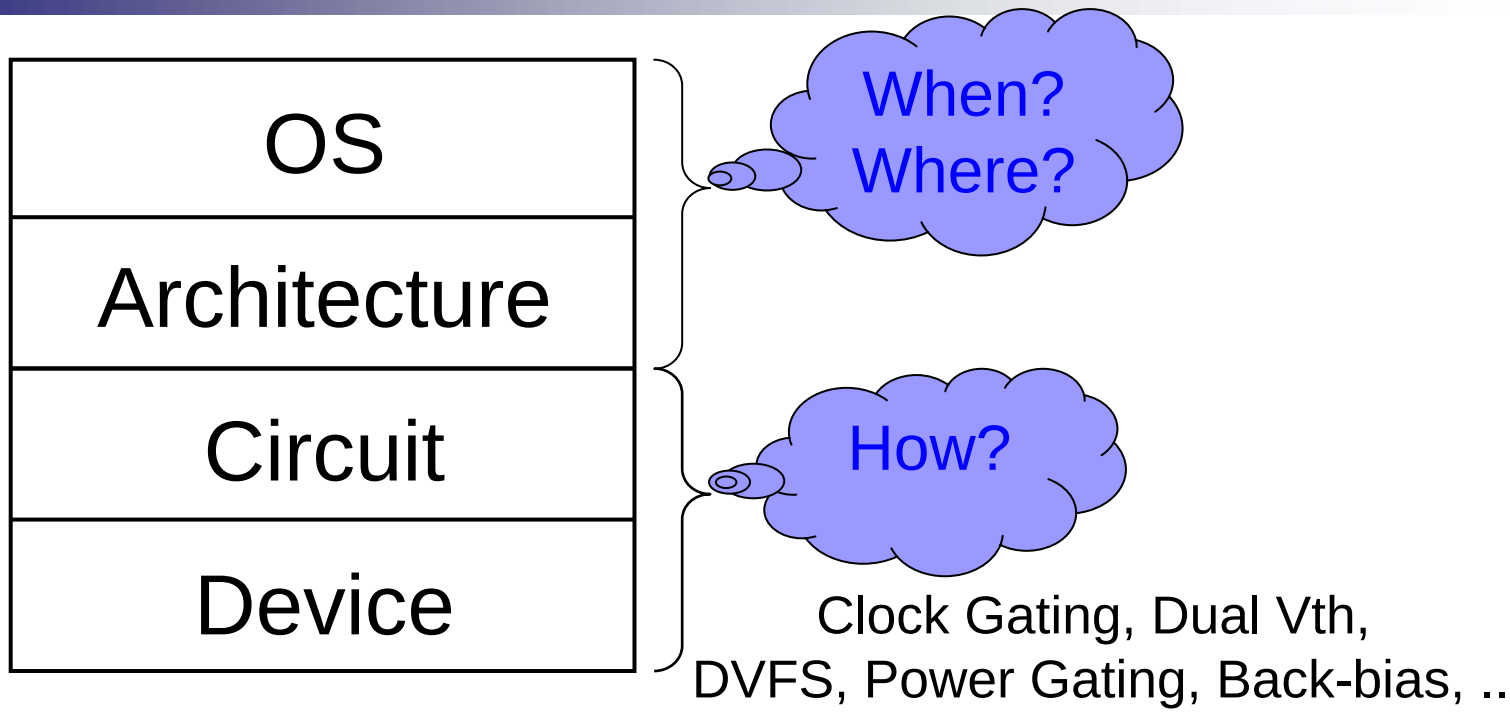
- For any complete system

  not simple because

  - Performance is limited by a bottleneck
    Power is summation of the whole system
  - Low power and slow operation for unhurried/idle parts
    → Low power consumption without performance penalty

→ Innovative Power/Performance Throttling is required

Low Power Technique so far,

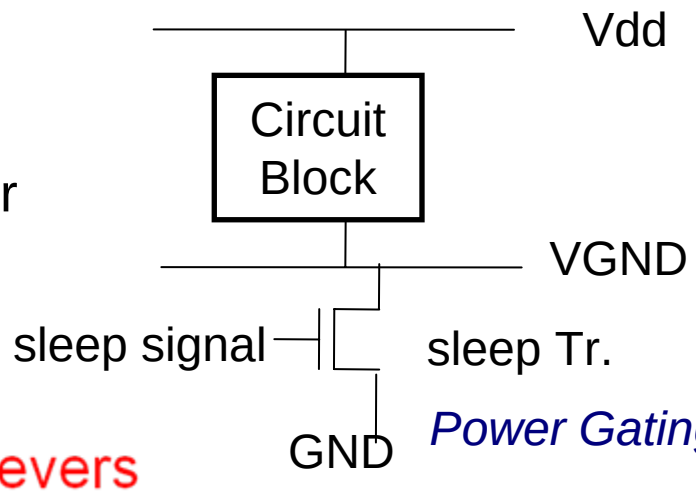


- Circuit Level: Providing levers to throttle performance/power
- Architecture, OS Level:
  find a chance to set levers, when and where ??

Example: throttling processor (=scaling down Vdd/frequency)
when memory is busier than processor (typical cases: cache miss)

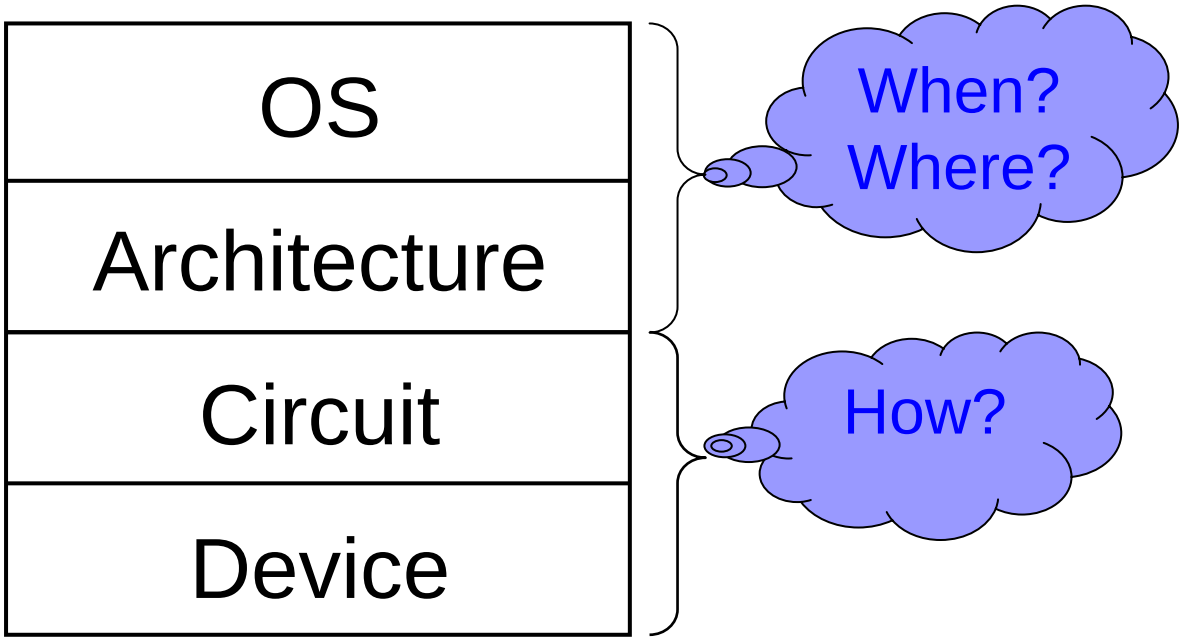
Example of Throttle Levers

- for dynamic power: Clock Gating, DVFS
  - both effective, particularly DVFS (Power $\propto$ Vdd^2)
  - Clock Gating: very fine grain control with little overhead
    → easily utilized within circuit level design
  - DVFS: tens of μ s to change Vdd through regulator
    → moderate granularity
- for leakage power: Power Gating, Body Biasing
  - both effective, but large overhead in power and performance
  - too frequent control may waste power
  - not easy for fine grain control

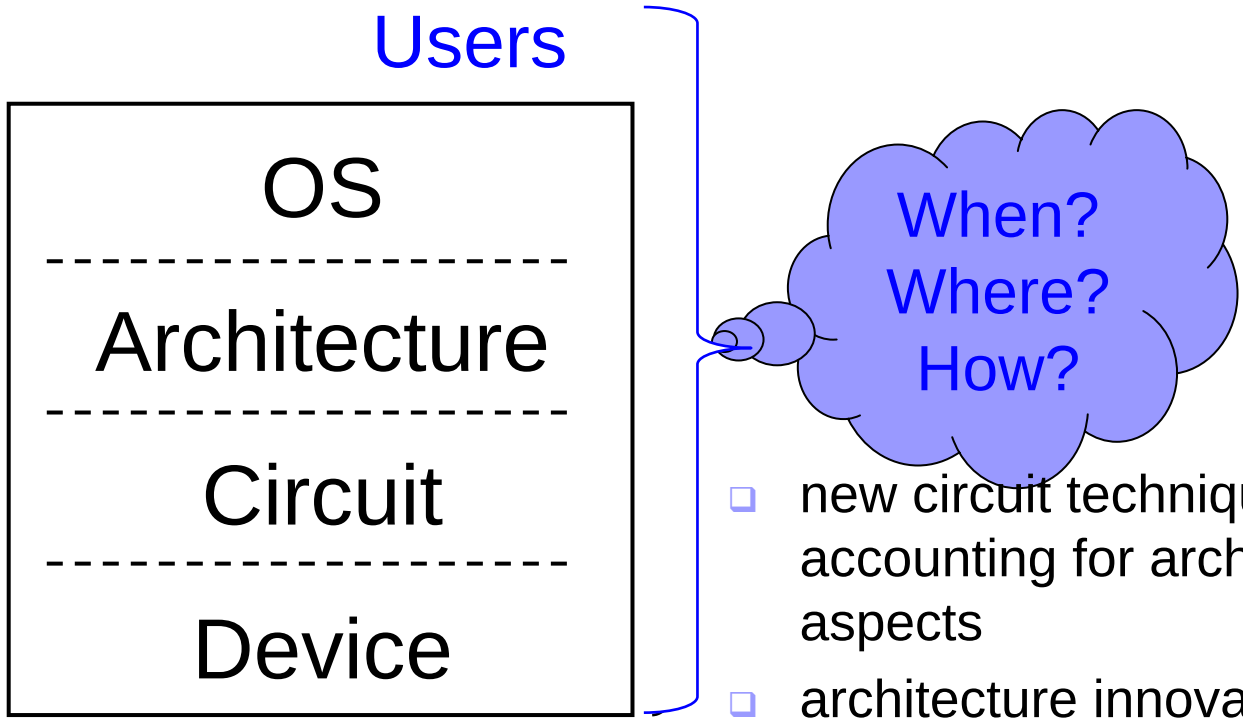


Activity Localization is important
to make good use of these Throttle Levers

Challenges for Power Wall Problem



Challenges for Power Wall Problem



- new circuit technique accounting for architecture aspects
- architecture innovation dedicated for circuit technique

- Co-Design of Throttle Levers and their Control
- Users should join the Co-Optimization
  - good abstraction of Throttle Levers for users
  - users should know the details of the Throttle Levers

Summary

- Performance/Power ratio is crucial to solve the power wall problem of supercomputing
- Suggestions:
  1. As Simple As Possible Architecture
  2. Memory Hierarchy Optimization
  3. Innovative Power Control by Co-Design of various system levels
     - Low power and slow operation for plenty of unhurried/idle parts
     - activity localization ← closely related to suggestion 2
     - Users should know the details of the throttling levers
- Good News: most HPC applications have regularity and locality
  - preferable for all these suggestions

→ Unyielding effort to make good use of regularity and locality will solve the problem.

Yes, We Can !!