Review Approval

Prepare Request
Search Requests
Generate Reports
Approvals
Help
Wizard

Search Requests
New Search
Refine Search
Search Results

Clone Request
Edit Request
Cancel Request

Search Detail

Submittal Details

Document Info
Title : Completing the Journey of Moore's Law
Document Number : 5221651 SAND Number : 2004-1728 P
Review Type : Electronic Status : Approved
Sandia Contact : DEBENEDICTIS,ERIK P. Submittal Type : Viewgraph/Presentation
Requestor : DEBENEDICTIS,ERIK P. Submit Date : 04/23/2004

Author(s)
DEBENEDICTIS,ERIK P.

Event (Conference/Journal/Book) Info
Name : Presentation at IBM Austin Research Center
City : Austijn State : TX Country : USA
Start Date : 04/23/2004 End Date : 04/23/2004

Partnership Info
Partnership Involved : No
Partner Approval : Agreement Number :

Patent Info
Scientific or Technical in Content : Yes
Technical Advance : No TA Form Filed : No
SD Number :

Classification and Sensitivity Info
Title : Unclassified-Unlimited Abstract : Document : Unclassified-Unlimited
Additional Limited Release Info : None.
DUSA : None.

Routing Details

Role	Routed To	Approved By	Approval Date
Derivative Classifier Approver	YARRINGTON,PAUL	YARRINGTON,PAUL	04/24/2004
Conditions:			
Classification Approver	WILLIAMS,RONALD L.	WILLIAMS,RONALD L.	04/27/2004
Conditions:			
Manager Approver	PUNDIT,NEIL D.	PUNDIT,NEIL D.	04/27/2004
Conditions:			
Administrator Approver	ABEYTA,RAMONA D.	ABEYTA,RAMONA D.	06/08/2004

Please use current funding statement on future review and approval requests. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under Contract DE-AC04-94AL85000 Printed 6/8/2004 (ra)

SAND2004-1728P

Completing the Journey of Moore's Law

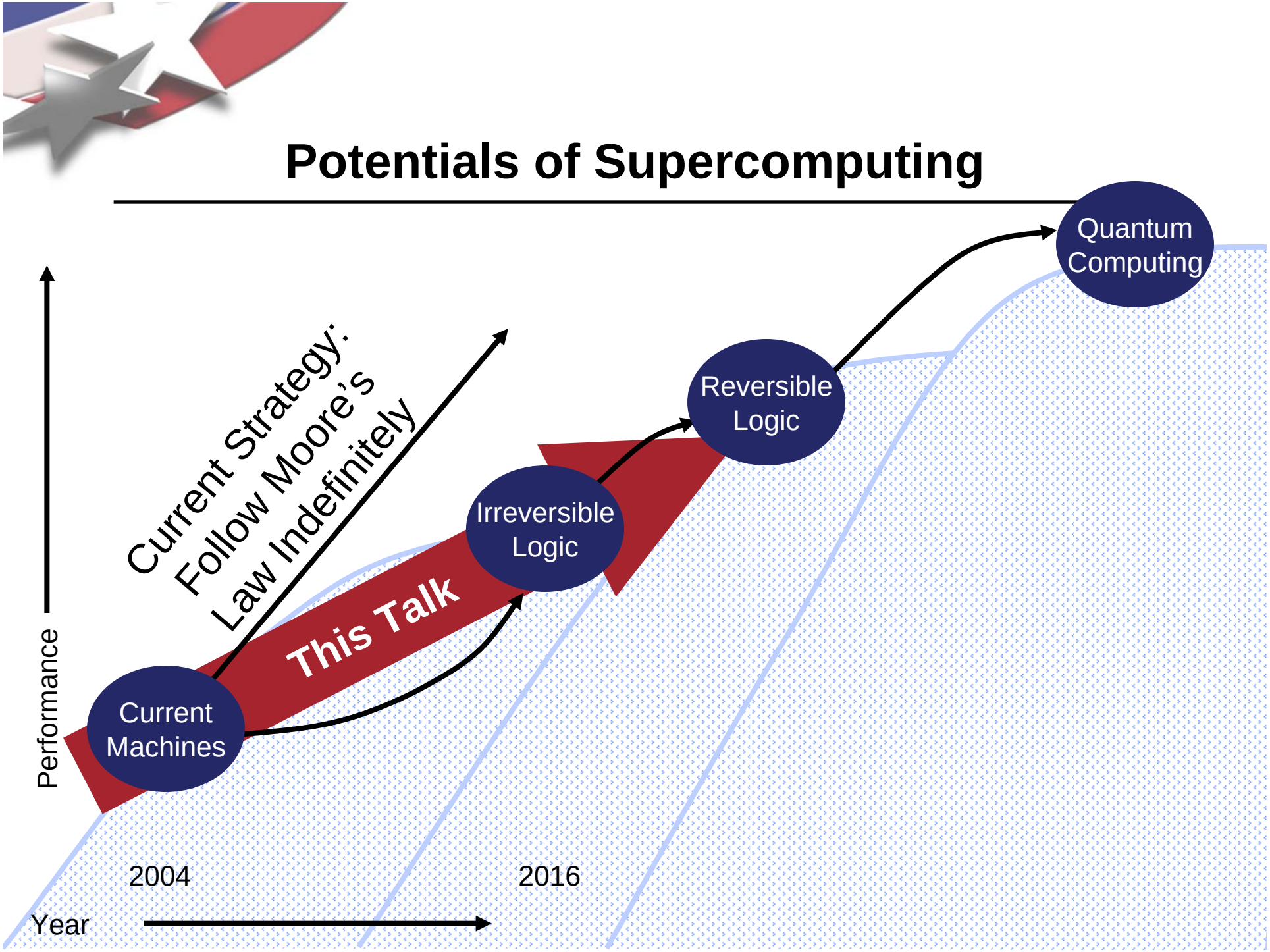
Presentation at IBM Austin

Erik P. DeBenedictis
Sandia National Laboratories

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company,
for the United States Department of Energy under contract DE-AC04-94AL85000.

Sandia National Laboratories

Potentials of Supercomputing



Outline

- Applications of the Future
- Limits of Moore's Law
- How to Reach the Limit
  - Aerogel model
  - Applications Modeling
- No Need For a Breakthrough
- Architecture
- Government-Vendor Issues

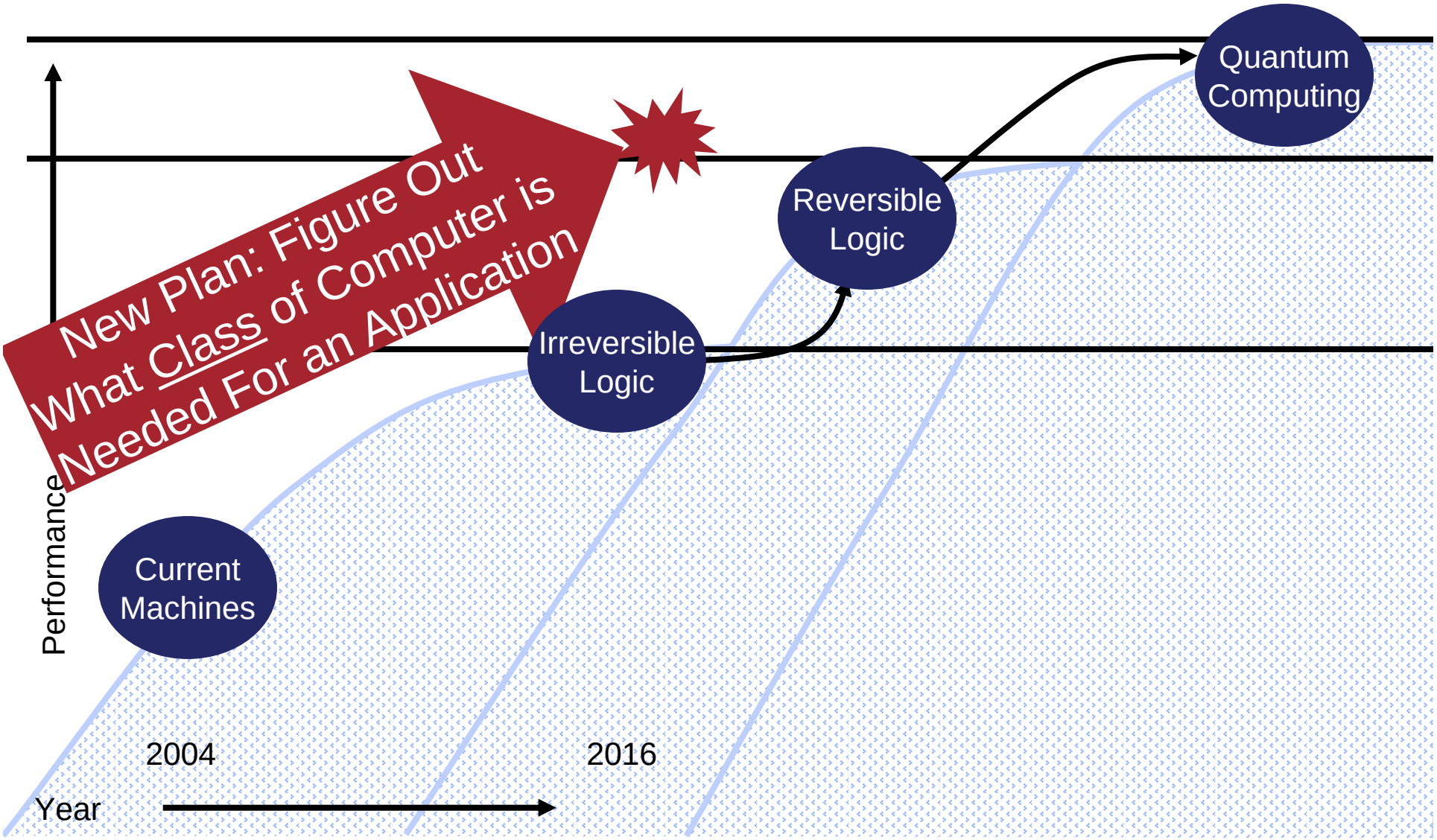
Sandia National Laboratories

Roles of Applications?

- Drives funding
  - Government funds supercomputers because applications provide benefit to society
  - "There ought to be" a document for our Government describing supercomputer benefits over the next 50 years
  - No so, see later
- Drives architecture
  - I think we should design supercomputers to run classes of applications well
    - Alternate view: computer should be easy to use
  - This talk is about the class "simulating physics on a computer"
  - We need to know about the applications to build the computer right

Sandia National Laboratories

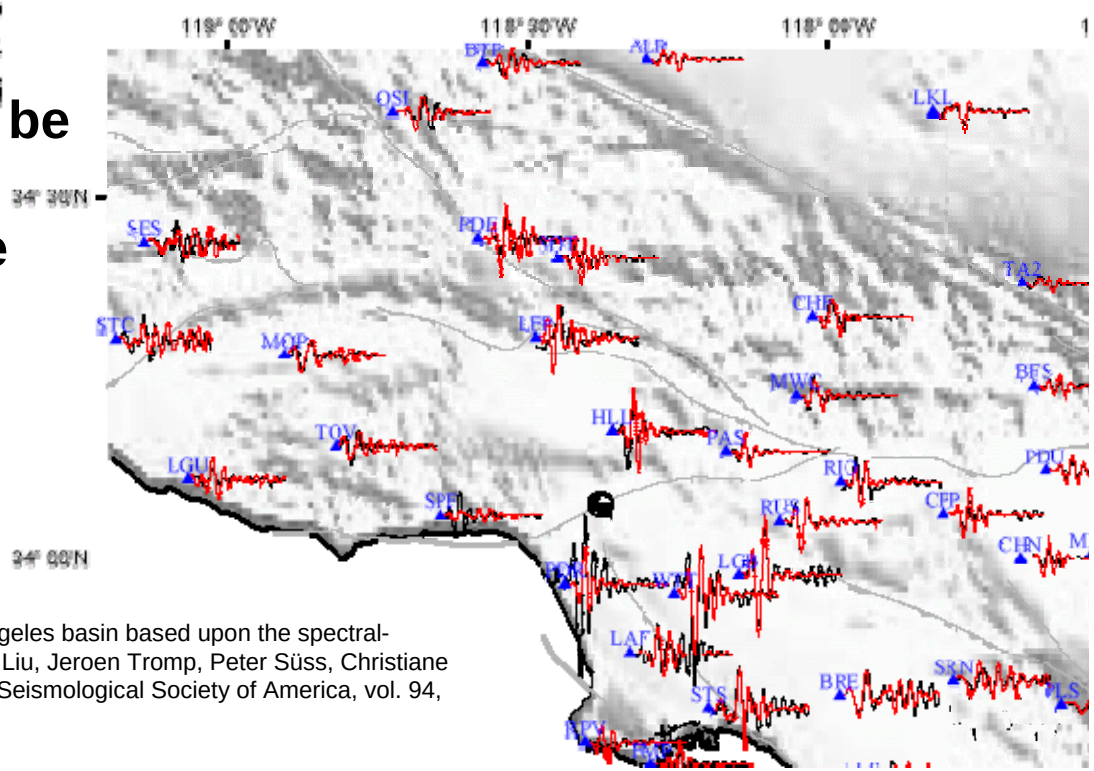
Applications



Example Application: Earthquake Mitigation

- **Simulation of ground motion due to earthquakes can be useful in deciding where it is safe to build structures**
- **Until available data can be analyzed, there will be unnecessary loss of life and property**

- **Required compute power for sufficient analysis of existing data: 1 Exaflops**



Simulations of ground motion in the Los Angeles basin based upon the spectral-element method, Dimitri Komatitsch, Qinya Liu, Jeroen Tromp, Peter Süss, Christiane Stidham and John H. Shaw, Bulletin of the Seismological Society of America, vol. 94, number 1, in press (2004)

Earthquake Risk Mitigation

- Forward Simulation
  - Spectral elements code written
  - Runs well on Earth Simulator (vectorizes)
  - Earth Simulator can do to ~.1 Hz
  - Seismographs collect data to ~100 Hz; scaling to 100 Hz would require ~3 Petaflops
- Reverse, “Imaging”
  - Adjoint method code written
  - Uses multiple instances of the forward simulation model, one for each measurement station (hundreds)
  - Scaling to image to the limit of collected data would require ~1 Exaflops

Sandia National Laboratories

Earthquake Risk Mitigation

- Algorithms: Written
- Code: Runs
- Input Data: Exists
- Consequence of Not Proceeding: People Die
- Required FLOPS: 1E = 1000P = 1,000,000T
  - 25,000 × Earth Simulator

Sandia National Laboratories

Global Climate

- There is intense political debate about whether indisputable changes in Earth's climate are natural and should be ignored or human-induced and should be a big concern
- Supercomputer simulation could be asked to
  - Answer question about what causes climate change
  - Explore mitigations

Sandia National Laboratories

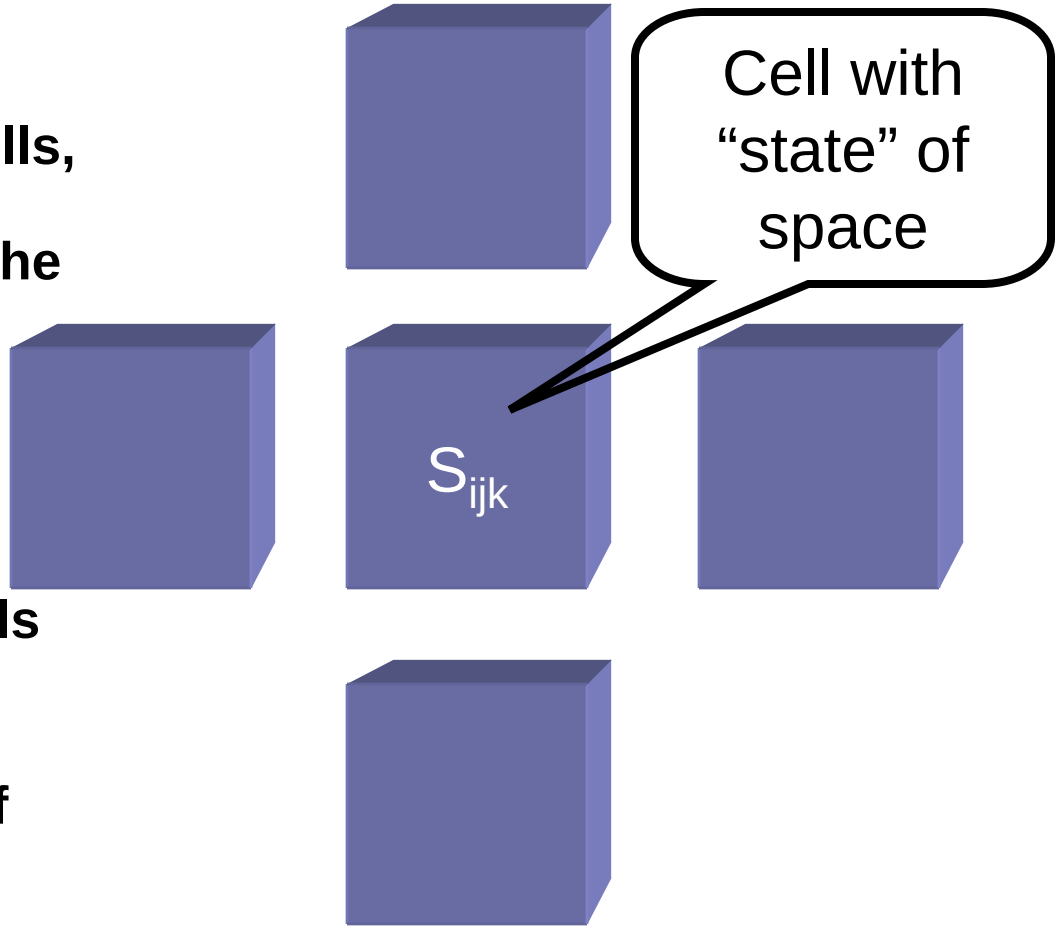
Supercomputers to Mitigate Climate Change

- Current status
  - Fairly large supercomputer codes can model climate
  - Conclusion: Results not accurate
  - Camp 1: Increase spatial resolution
  - Camp 2: Increase physics
- Possible future need for a big computer
  - Simulate climate with high spatial resolution,
  - lots of physics, and
  - repeat for 1000 possible candidate mitigations
- Expected result: something humans could do to keep climate habitable
- Required FLOPS: ?

Sandia National Laboratories

The Class of Applications I'll Talk About

- **Space is divided into cells, each with computer variables representing the physical state of the volume represented by the cell**
- **The computer updates the state of a cell for successive time intervals ΔT based on some physical laws**
- **I. e. S_{ijk}' = f(S_{ijk}, states of nearby cells)**



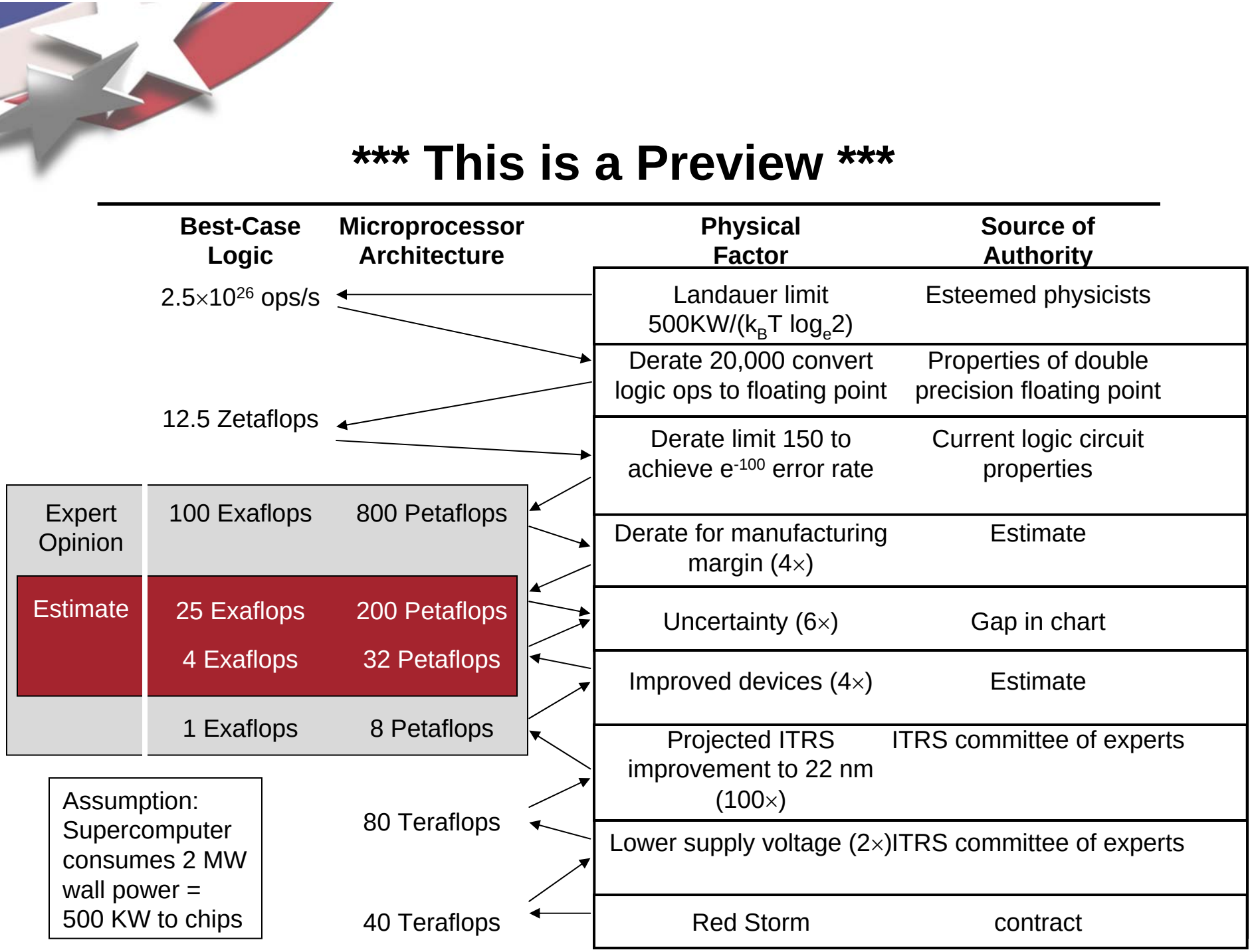
Sandia National Laboratories

Outline

- Applications of the Future
- Limits of Moore's Law
- How to Reach the Limit
  - Aerogel model
  - Applications Modeling
- No Need For a Breakthrough
- Architecture
- Government-Vendor Issues

Sandia National Laboratories

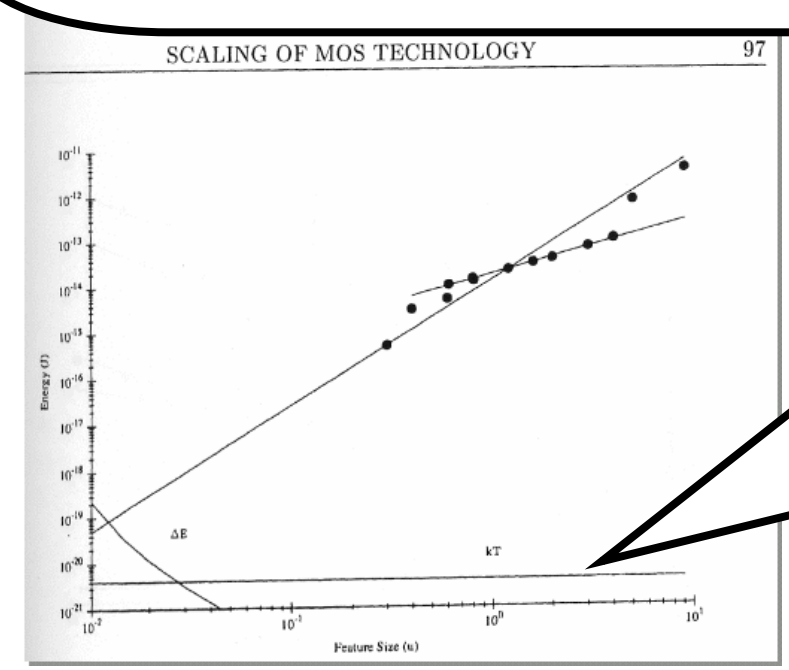
*** This is a Preview ***



Thermal Noise Limit

This logical irreversibility is associated with physical irreversibility and requires a minimal heat generation, per machine cycle, typically of the order of kT for each irreversible function.
– R. Landauer 1961


Irreversibility and Heat Generation in the Computing Process



kT "helper line," drawn out of the reader's focus because it wasn't important at the time of writing
– Carver Mead, Scaling of MOS Technology, 1994

Sandia National Laboratories

Metaphor to FM Radio on Trip to Dallas

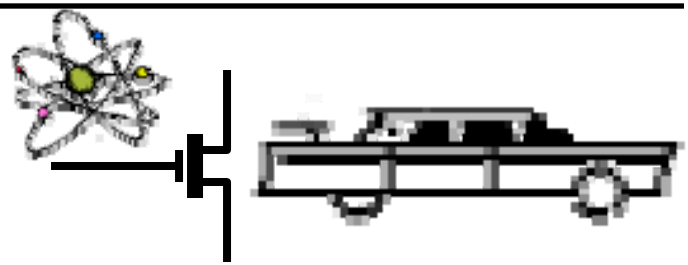
- You drive to Dallas listening to FM radio
- Music clear for a while, but noise creeps in and then overtakes music
- Why?
  - Signal at antenna weakens
  - Thermal electron noise constant at $k_B T$
- Analogy: You live out the next dozen years buying PCs every couple years
- Electrical effect
  - Moore's Law causes switching energy of gates to decrease at about 30% per year
  - Thermal electron noise constant at $k_B T$

Details: Erik DeBenedictis, "Taking ASCI Supercomputing to the End Game," SAND2004-0959

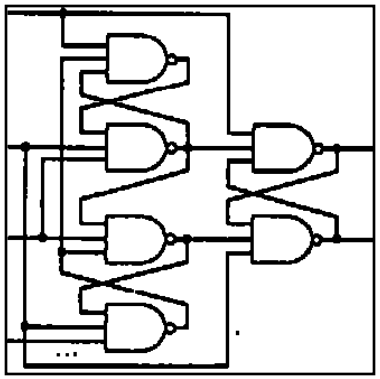
Sandia National Laboratories

FM Radio and End of Moore's Law

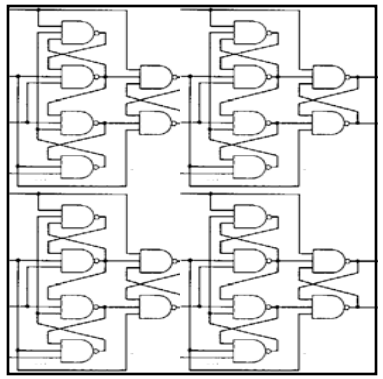
Distance



Driving away from FM transmitter→less signal
Noise from electrons → no change



Shrink



Increasing numbers of gates→less signal power
Noise from electrons → no change

Sandia National Laboratories

Our Expectations of Reliability

- What is the consequence of a computer making a spontaneous logic error?
  - We replace the computer
  - Worse than DRAM, where we would add ECC logic
  - Less severe than a heart-lung machine, where we would not build the machine in the first place
- A supercomputer operating at the physical limits a dozen years from now will perform 10^{30}-10^{40} gate operations in its lifetime
- To avoid premature replacement, the probability of a glitch in a gate should be 10^{-30}-10^{-40} per operation

Sandia National Laboratories

Impact of Power on Reliability

- According to the ITRS roadmap, gates in 2016 and based on 22 nm transistors will be at 10× the power necessary to maintain reliable operation
- However, signal energy is lost for all sorts of reasons and manufacturing tolerances make it unwise to design to the limits
- End of road is on the map!

SNR (db)	Power Ratio	P_{error}
10	10	3.9×10^{-6}
14	25	6.8×10^{-13}
18	63	1.4×10^{-29}
22	160	3.3×10^{-71}
26	400	1.8×10^{-175}
30	1,000	4.5×10^{-437}
34	2,500	7.1×10^{-1094}
38	6,300	2.2×10^{-2743}
42	16,000	1.8×10^{-6886}
46	40,000	3.8×10^{-17293}
50	100,000	3.2×10^{-43433}
54	250,000	8.1×10^{-10194}
58	630,000	1.8×10^{-274025}
62	1,500,000	9.6×10^{-688315}

Noise Limit

2016

Today

$$q := \int_t^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx; \; t \to \sqrt{2 * 10^{\frac{snr}{10}}}$$

Sandia National Laboratories

Noise Levels

- 0 db Limit of hearing
- 20 db Rustling leaves
- 40-50 db Typical neighborhood
- 60-70 db Normal conversation
- 80 db Telephone dial tone
- 85 db City traffic inside car
- 90 db Train whistle @500’
- 95 db Subway train @200’
- 90-95 db Ear damage

- Today: 50 db
  - Thermal noise:Logic:: Rustling leaves:Talking
- 2016: 30 db
  - Thermal noise:Logic:: Talking:Train Whistle
- Reliability limit 20 db
  - Thermal noise:Logic:: Outside neighborhood:Talking

Sandia National Laboratories

SIA Semiconductor Roadmap

- Generalization of Moore's Law
  - Projects many parameters
  - Years through 2016
  - Includes justification
  - Panel of experts
    - known to be wrong
  - Size between Albuquerque white and yellow pages

Sandia National Laboratories

Semiconductor Roadmap

YEAR OF PRODUCTION	2010	2013	2016
DRAM ½ PITCH (nm)	45	32	22
MPU / ASIC ½ PITCH (nm)	30	22	15
MPU PRINTED GATE LENGTH (nm)	25	18	13
MPU PHYSICAL GATE LENGTH (nm)	18	13	9
Physical gate length high-performance (HP) (nm) [1]	18	13	9
Equivalent physical oxide thickness for high-performance T_{ox} (EOT) (nm) [2]	0.5-0.8	0.4-0.6	0.4-0.5
Gate depletion and quantum effects electrical thickness adjustment factor (nm) [3]	0.5	0.5	0.5
T_{ox} electrical equivalent (nm) [4]	1.2	1.0	0.9
Nominal power supply voltage (V_{dd}) (V) [5]	0.8	0.5	0.4
Nominal high-performance NMOS sub threshold leakage current, I_{sdleak} (at 25°C) (µA/µm) [6]	3	7	10
Nominal high-performance NMOS saturation drive current, I_{on} (at V_{dd} at 25°C) (µA/µm) [7]	1200	1500	1500
Required percent current-drive "mobility/transconductance improvement" [8]	30%	70%	100%
Parasitic source/drain resistance (Rsd) (ohm-µm) [9]	110	90	80
Parasitic source/drain resistance (Rsd) pe	25%	30%	35%
Parasitic capacitance percent of ideal gat	31%	36%	42%
High-performance NMOS device τ (C_{gate} * V_{dd} / I_{on}-nMOS/µm) [11]	0.39	0.22	0.15
Relative device performance [12]	4.5	7.2	10.7
Energy per (WL$_{gate}$=F) device switching transition (C_{gate}*(F*L$_{gate}$)*V^2) (fJ/Device) [14]	0.015	0.007	0.002
Static power dissipation per (WL$_{gate}$=F) device (Watts/Device) [13]	9.7E-08	1.4E-07	1.1E-07

1,000 k_BT/transistor

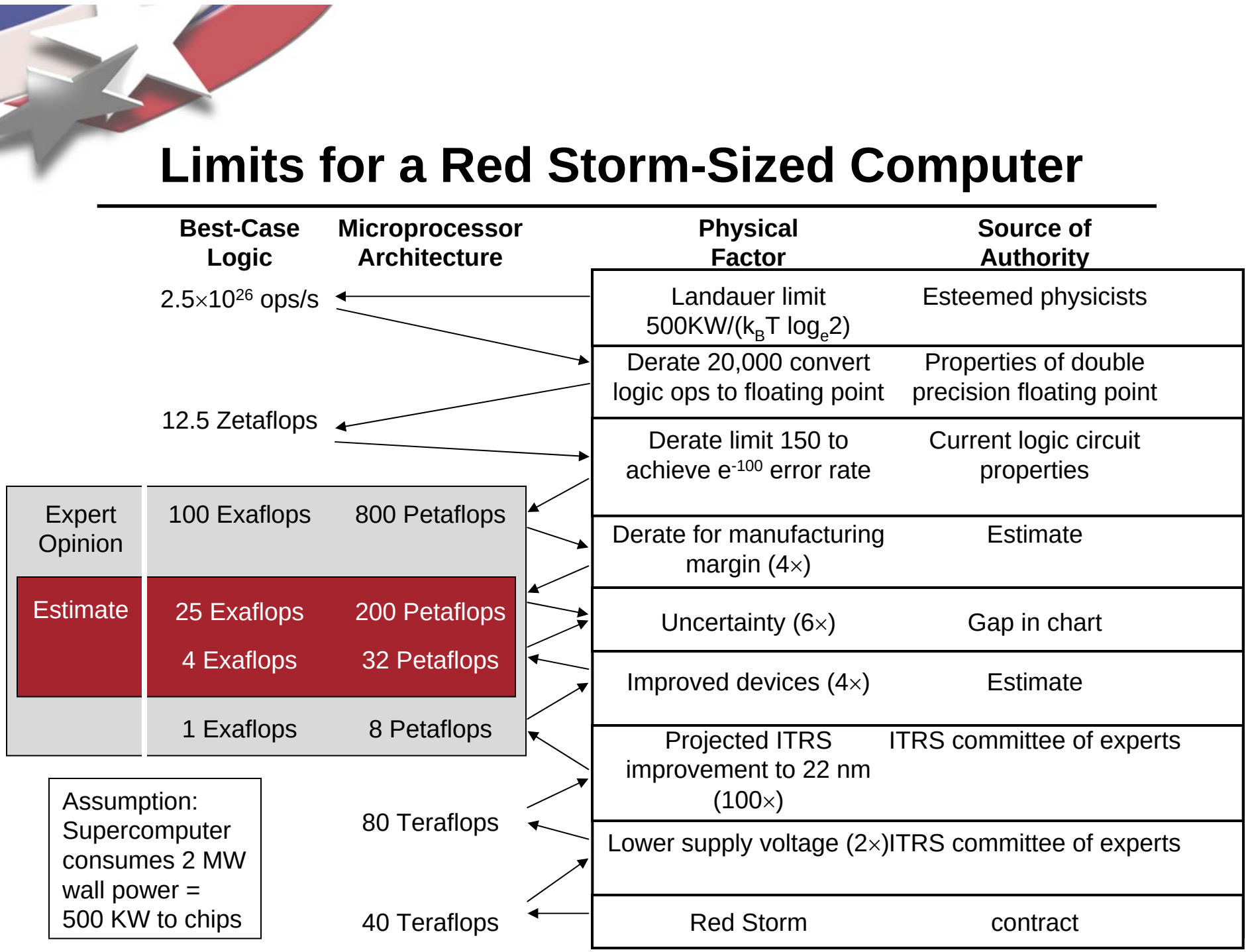
White—Manufacturable Solutions Exist, and Are Being Optimized

Yellow—Manufacturable Solutions are Known

Red—Manufacturable Solutions are NOT Known

Sandia National Laboratories

Limits for a Red Storm-Sized Computer



Outline

- Applications of the Future
- Limits of Moore's Law
- How to Reach the Limit
  - Aerogel model
  - Applications Modeling
- No Need For a Breakthrough
- Architecture
- Government-Vendor Issues

Sandia National Laboratories

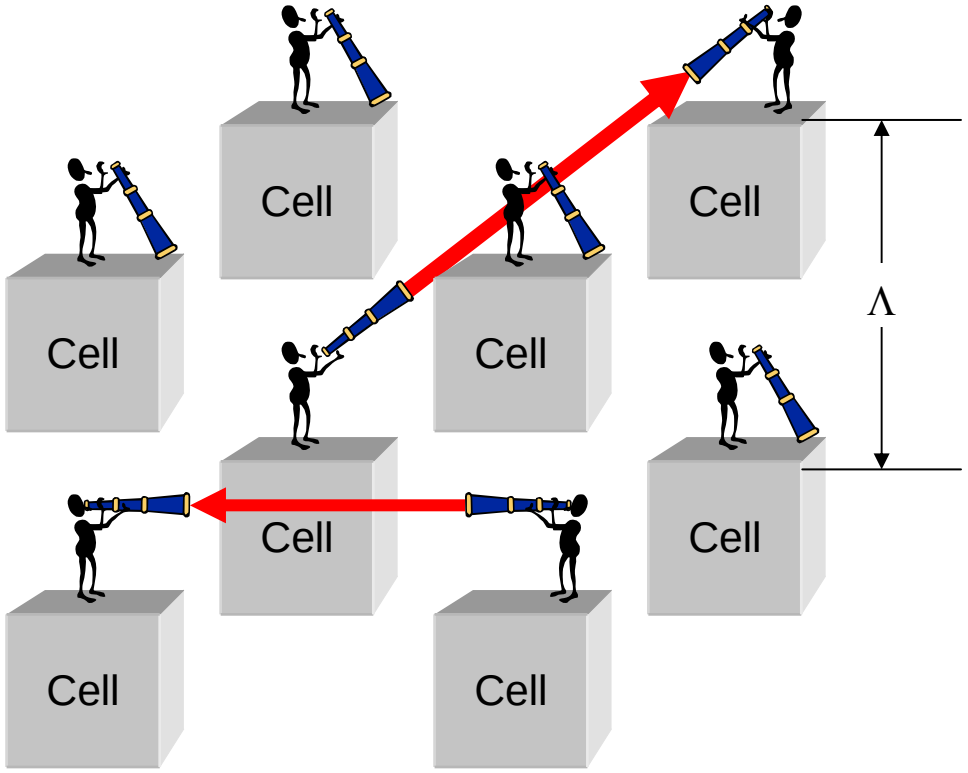
Can We Reach the Limit?

- Method: Compare modeled running time on perfect computer to real computer
- Application: Local calculations with global time step (SOR)
- Technology comparison:
  - 22 nm transistors with 3D atom-by-atom assembly
  - Our best shot at an architecture
- Definition of Success: Our best shot comes within a constant factor of the theoretical peak

Sandia National Laboratories

Aerogel Computer

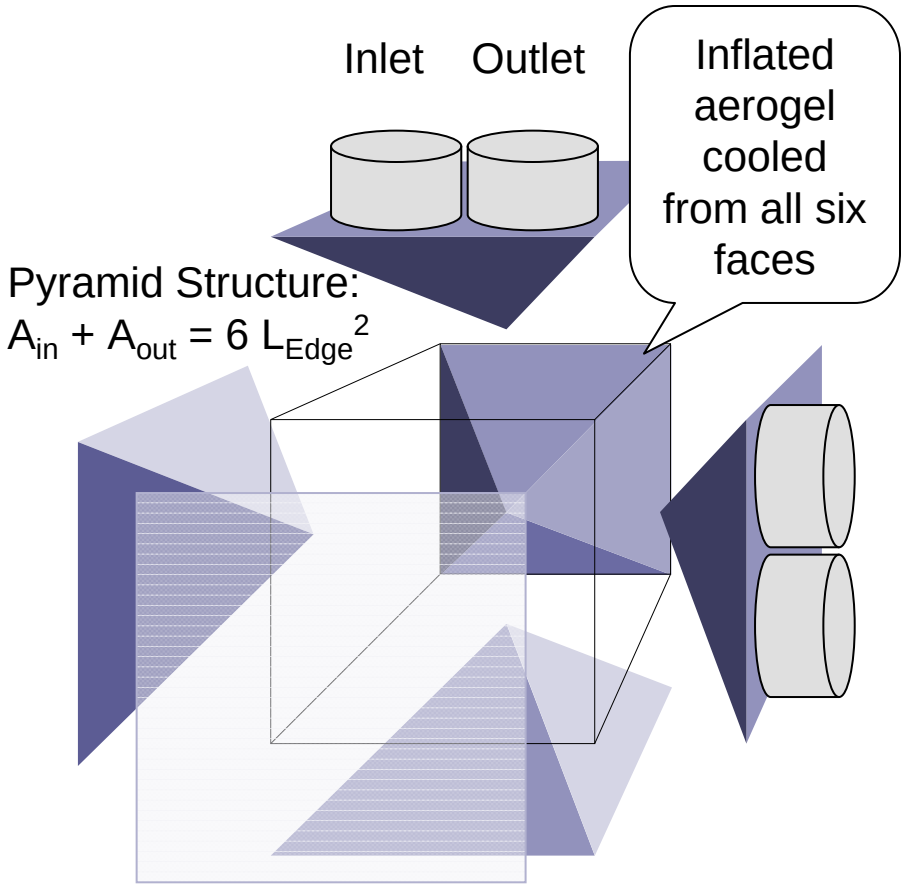
- Devise algorithm for a hypothetical aerogel computer
  - Cell may be gate
  - Cell may be memory
  - Is space for cooling, but no cooling
- Model application runtime
- Engineer real computer
- Model application runtime
- If runtimes similar, you succeeded



Sandia National Laboratories

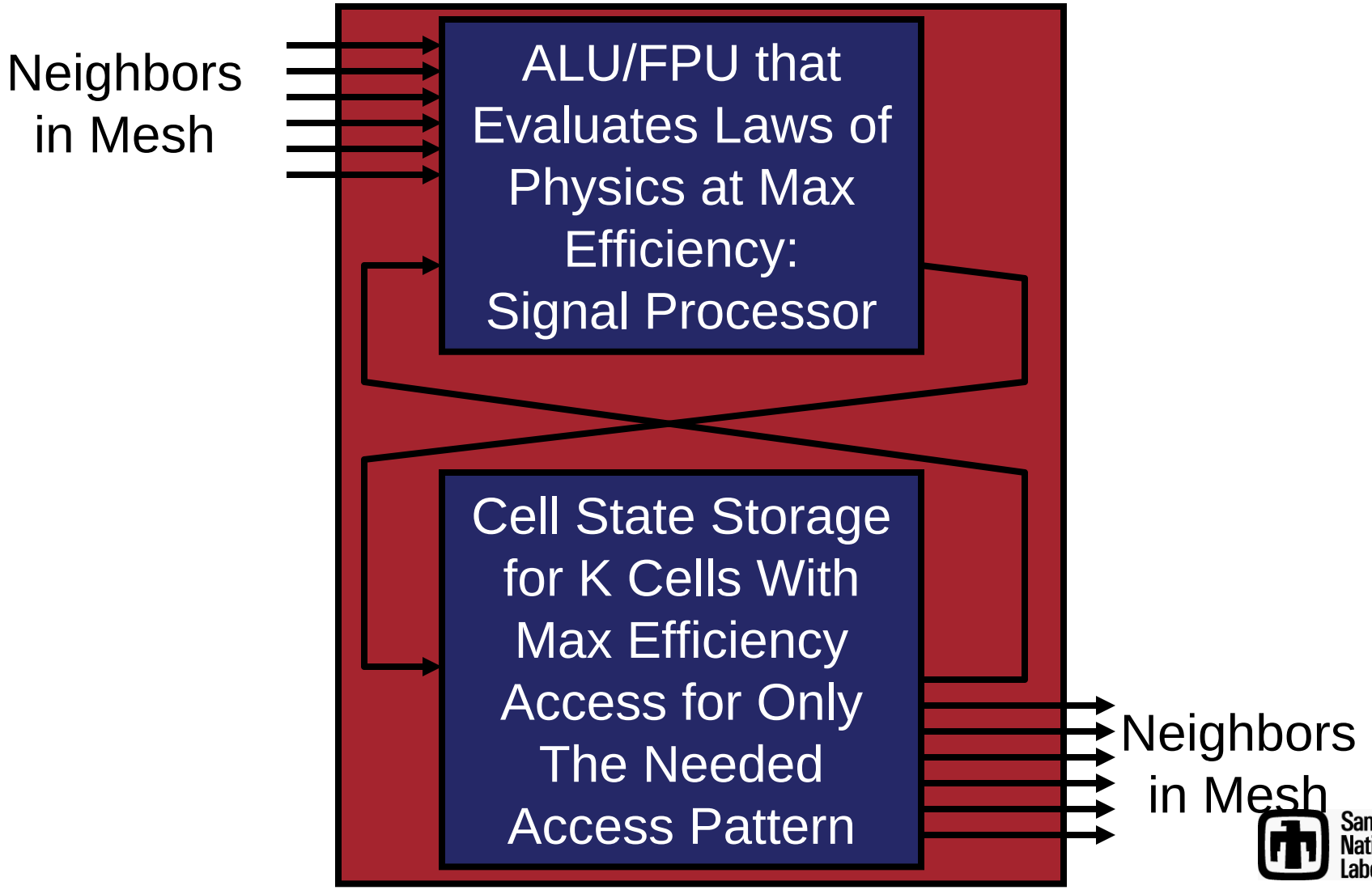
Aerogel Cooling

- **Inflate aerogel computer to point where heat emerging from faces is less than capacity of a designated cooling system**
  - **Air 45KW/m^2**
  - **Water 62MW/m^2**
  - **Pulse ∞W/m^2**



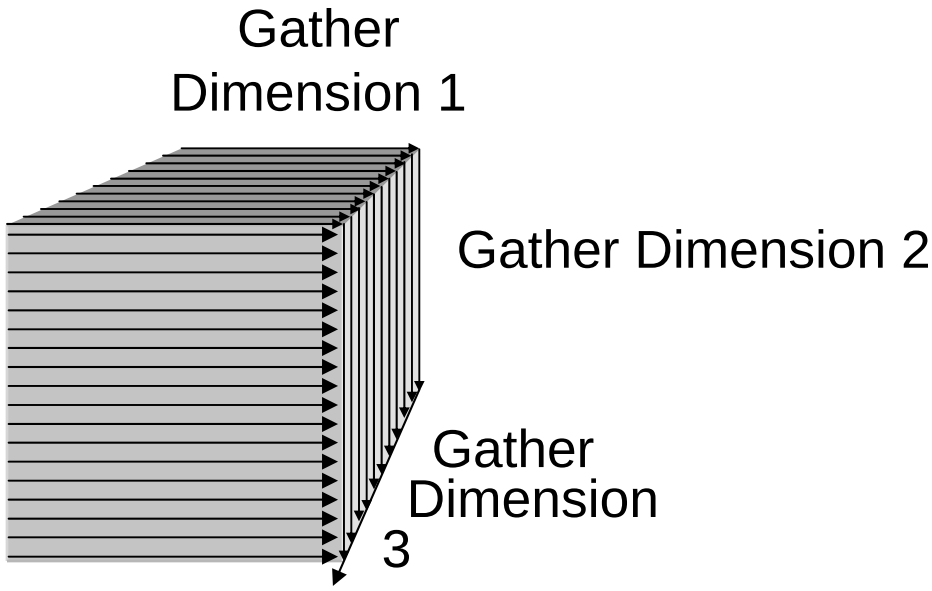
Sandia National Laboratories

Architecture Target



Sandia National Laboratories

Global Synchronization



Application Modeling

- Sample Problem
  - 3D finite difference equation with global synchronization
  - SOR method

$$T_{Step} = \frac{K \times F_{cell}}{floprate} + T_{Global}$$

  - where
    - K is memory size

- Global synchronization limited by speed of light

$$T_{Global} \geq \frac{2\sqrt{3} \times L_{Edge}}{c}$$

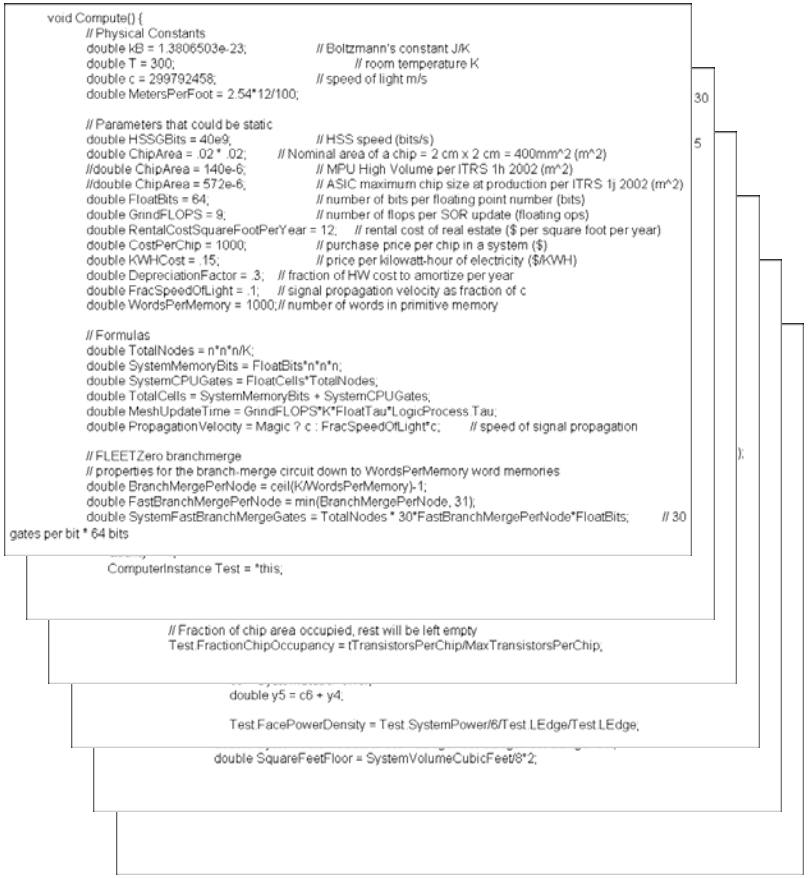
  - where
    - L_{Edge} is edge dimension of cube

$$6 \times L_{Edge}^{2} \times C_x \leq Power$$

Sandia National Laboratories

Actual Applications Modeling

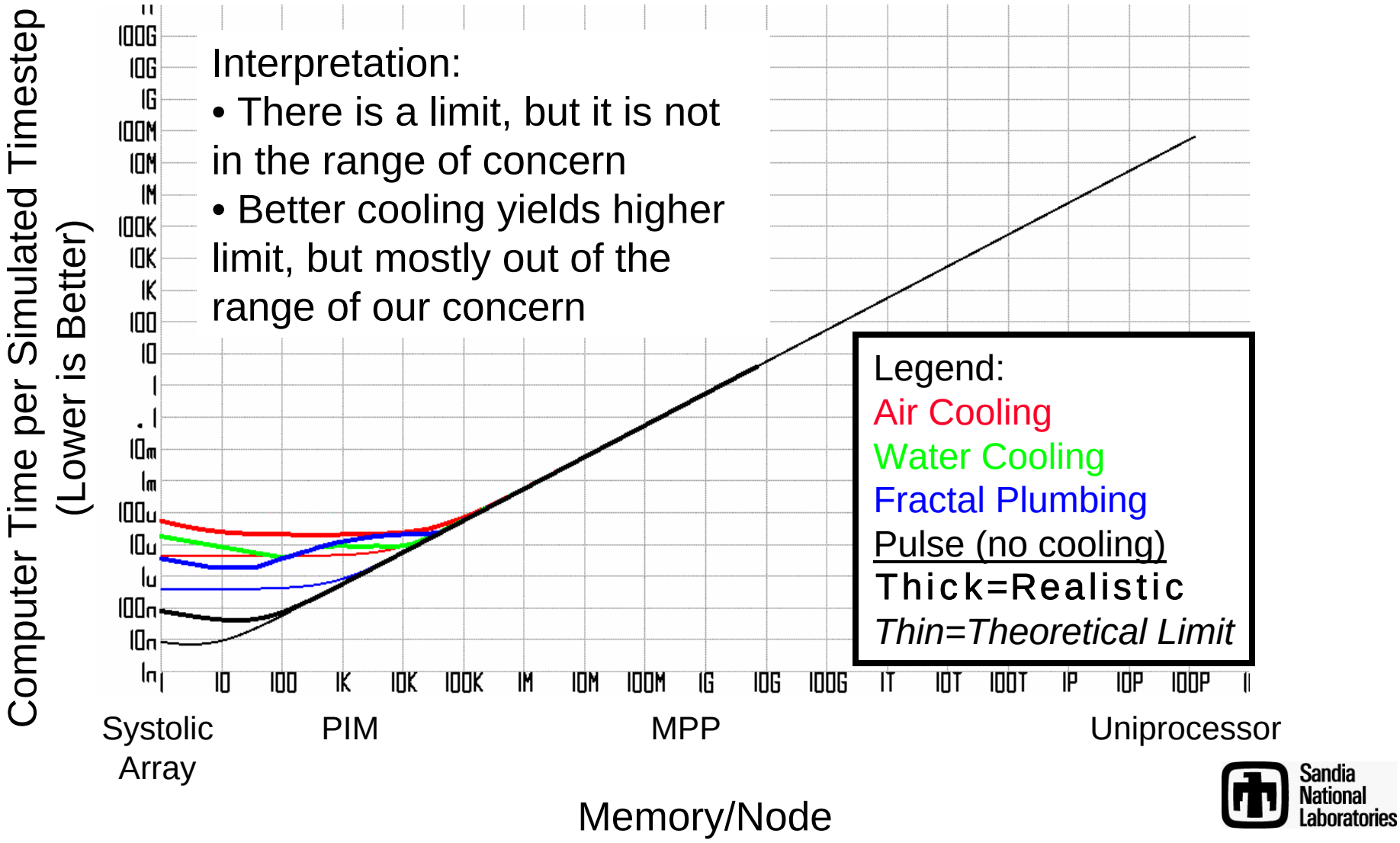
- Actual code was several hundred lines of C++
- Theoretical limit covered
  - Coolant
- Realistic covered
  - Layout on a 2D surface of a particular size
  - Heat sink limits
  - I/O bandwidth from chip
  - Coolant



```
void Compute() {
        // Physical Constants
        double kB = 1.3806503e-23;              // Boltzmann's constant J/K
        double T = 300;                                 // room temperature K
        double c = 299792458;                   // speed of light m/s
        double MetersPerFoot = 2.54*12/100;

        // Parameters that could be static
        double HSSGBits = 40e9;                 // HSS speed (bits/s)
        double ChipArea = .02 * .02;        // Nominal area of a chip = 2 cm x 2 cm = 400mm^2 (m^2)
        //double ChipArea = 140e-6;             // MPU High Volume per ITRS 1h 2002 (m^2)
        //double ChipArea = 572e-6;             // ASIC maximum chip size at production per ITRS 1j 2002 (m^2)
        double FloatBits = 64;                  // number of bits per floating point number (bits)
        double GrindFLOPS = 9;                  // number of flops per SOR update (floating ops)
        double RentalCostSquareFootPerYear = 12;    // rental cost of real estate ($ per square foot per year)
        double CostPerChip = 1000;              // purchase price per chip in a system ($)
        double KWHCost = .15;                   // price per kilowatt-hour of electricity ($/KWH)
        double DepreciationFactor = .3;   // fraction of HW cost to amortize per year
        double FracSpeedOfLight = .1;     // signal propagation velocity as fraction of c
        double WordsPerMemory = 1000;// number of words in primitive memory

        // Formulas
        double TotalNodes = n*n*n/K;
        double SystemMemoryBits = FloatBits*n*n*n;
        double SystemCPUGates = FloatCells*TotalNodes;
        double TotalCells = SystemMemoryBits + SystemCPUGates;
        double MeshUpdateTime = GrindFLOPS*K*FloatTau*LogicProcess.Tau;
        double PropagationVelocity = Magic ? c : FracSpeedOfLight*c;       // speed of signal propagation

        // FLEETZero branchmerge
        // properties for the branch-merge circuit down to WordsPerMemory word memories
        double BranchMergePerNode = ceil(K/WordsPerMemory)-1;
        double FastBranchMergePerNode = min(BranchMergePerNode, 31);
        double SystemFastBranchMergeGates = TotalNodes * 30*FastBranchMergePerNode*FloatBits;       // 30
gates per bit * 64 bits
                ComputerInstance Test = *this;

                // Fraction of chip area occupied, rest will be left empty
                Test.FractionChipOccupancy = tTransistorsPerChip/MaxTransistorsPerChip;

                double y5 = c6 + y4;

                Test.FacePowerDensity = Test.SystemPower/6/Test.LEdge/Test.LEdge;

                double SquareFeetFloor = SystemVolumeCubicFeet/8*2;
```

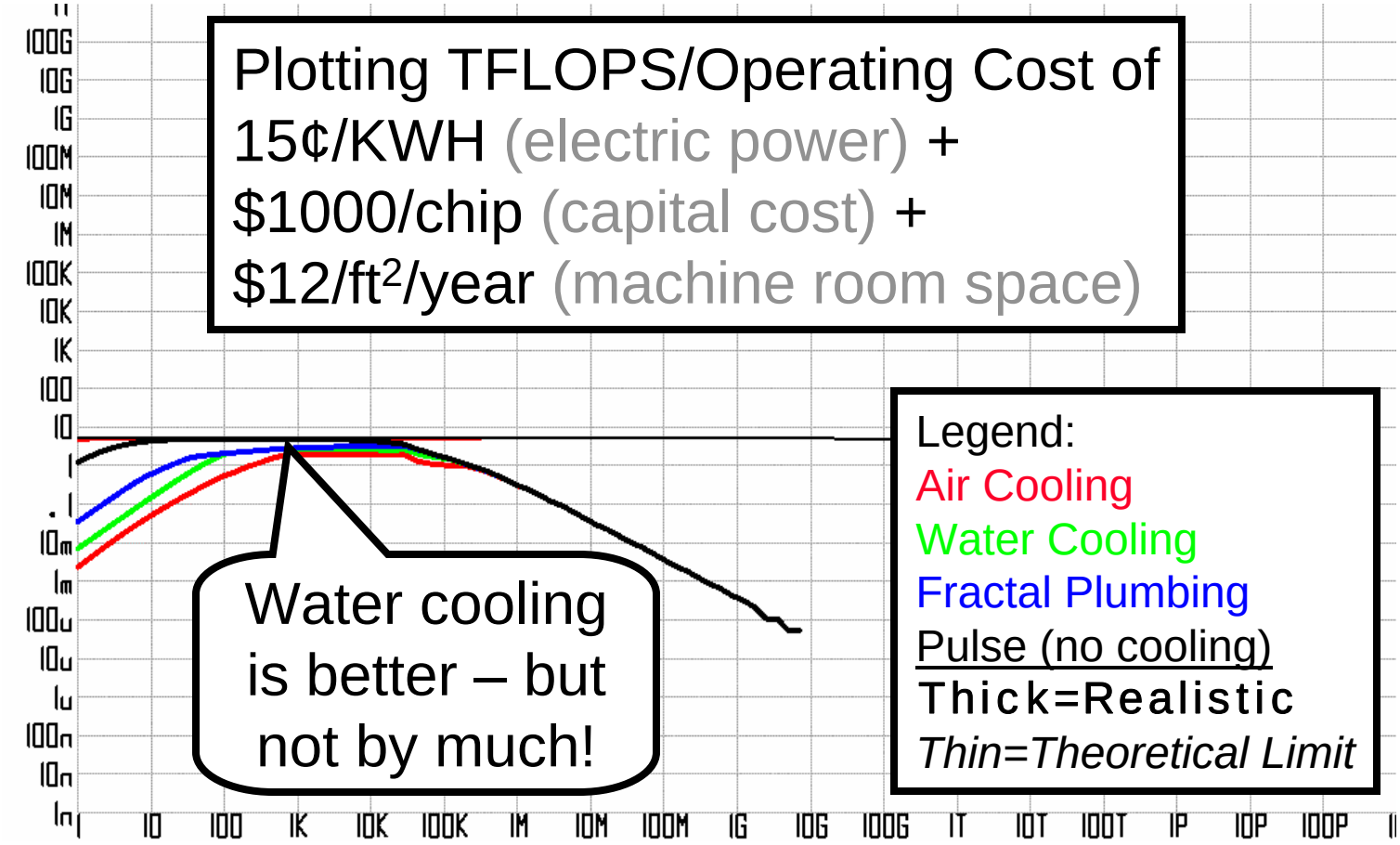
Sandia National Laboratories

Performance on Sample Problem



Sandia National Laboratories

Cost Efficiency



Sandia National Laboratories

Outline

- Applications of the Future
- Limits of Moore's Law
- How to Reach the Limit
  - Aerogel model
  - Applications Modeling
- No Need For a Breakthrough
- Architecture
- Government-Vendor Issues

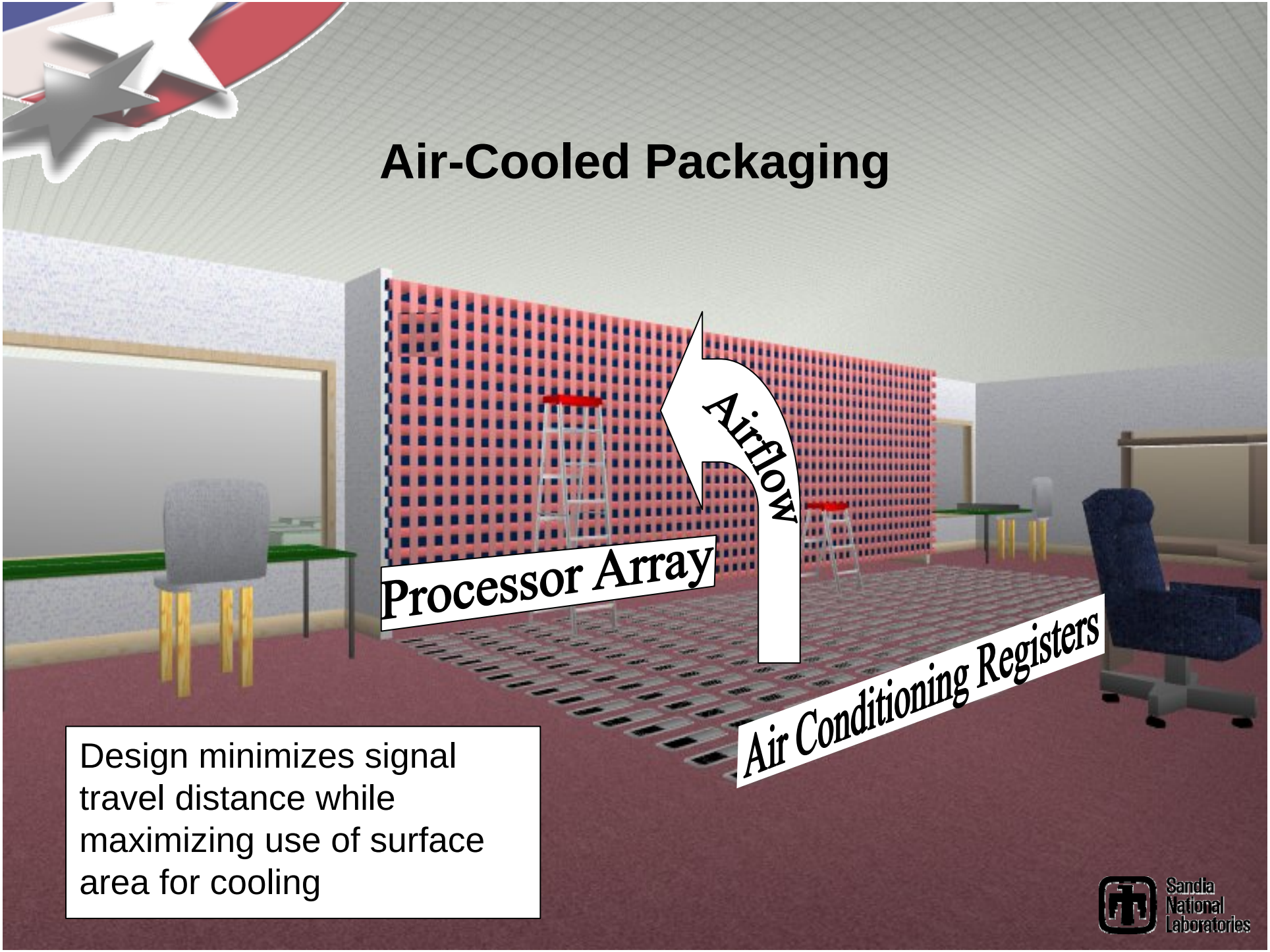
Sandia National Laboratories

Example of Computer at Physics Limit

- Sandia is often approached by people who say we need some elaborate technology in order to run our applications at the Petaflops level
  - Do we need elaborate technology?
  - Is the person just looking for research funding?
- Question: can we make a computer that runs at the limits out of inexpensive components?
  - Yes, subsequent slides are example

Sandia National Laboratories

Air-Cooled Packaging



Design minimizes signal travel distance while maximizing use of surface area for cooling

Sandia National Laboratories

Outline

- Applications of the Future
- Limits of Moore's Law
- How to Reach the Limit
  - Aerogel model
  - Applications Modeling
- No Need For a Breakthrough
- Architecture
- Government-Vendor Issues

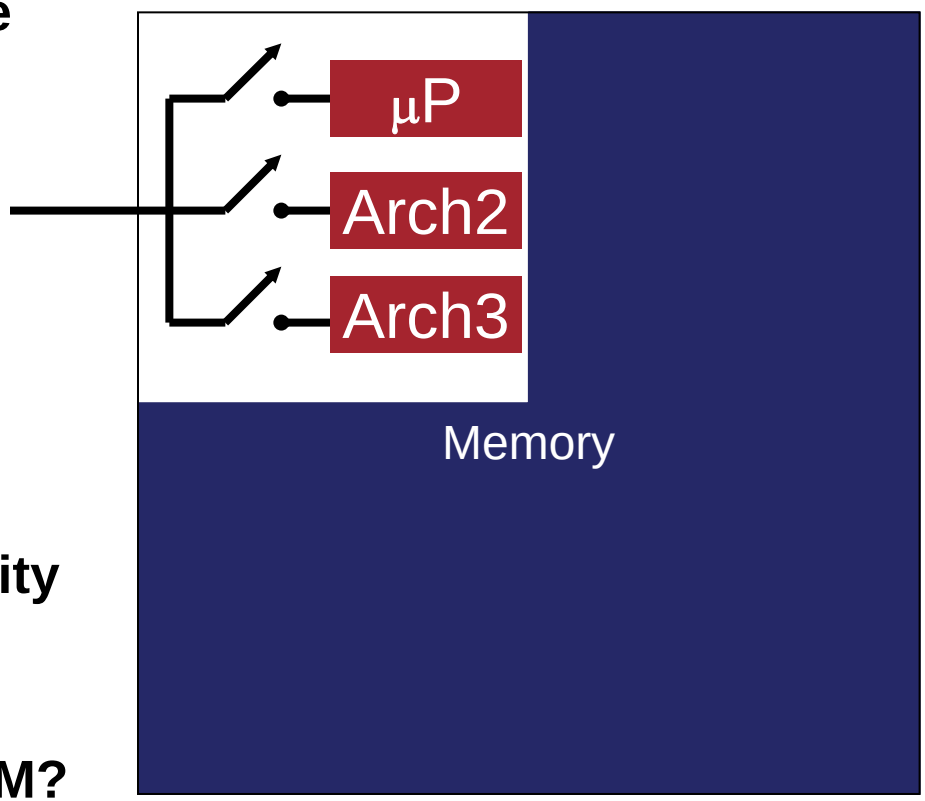
Sandia National Laboratories

Which Microarchitecture?

- Task: Pick a winner
  - Candidates μP, PIM, vector, FPGA, reconfigurable, streaming, maybe more
  - Each has advantages
  - Not clear which is best
  - Government gets bad press for picking winners too early
- Why do we pick winners
  - Logic is a scarce resource
  - But hang on a minute, don't we have more transistors than we know what to do with, and even turn some off at times?
- Can we change the rules of the game to make NOT picking a winner a virtue?

Multi-Architecture Idea

- **Architecture to comprise**
  - **μP and accelerator architectures 1 and 2**
  - **Power control circuit so only one is turned on at a time**
- **Benefit**
  - **Can expect support from cluster community and advocates of architectures 2 and 3**
- **Arch2=Vector, Arch3=PIM?**



Outline

- Applications of the Future
- Limits of Moore's Law
- How to Reach the Limit
  - Aerogel model
  - Applications Modeling
- No Need For a Breakthrough
- Architecture
- Government-Vendor Issues

Sandia National Laboratories

Ideas on Govt.-Vendor Expectations

- The will be applications requiring arbitrarily large FLOPS rates
- Some of these applications will approach the limits of CMOS microprocessors and some will reach into more powerful computing classes
  - Probably in 12-24 years
- Somebody will build these machines
- Software may have a lifespan of 20 years or more
  - Maturity of Moore's Law before software obsolescence
- Can vendors show that their architectures will be viable over the entire lifetime of the software written for it?

Sandia National Laboratories

Ideas on Govt.-Vendor Expectations

- From Govt. perspective, some architecture may be headed to a dead end, others may have a role in the end game solution
- How to Tell?
  - Roadmap exists for taking the architecture forward in time
  - Power consumption under control
  - Some plausible story exists for describing how well applications will run in the distant future

Sandia National Laboratories

Extreme Computing

- I define Moore's Law as smaller transistors applied to similar logic circuits
  - 2N transistors for a N-input gate
  - However, some people believe Moore invented the exponential (explain joke)
- Moore's Law limited to 100 $k_B T$ energy per op
  - Doesn't matter how small the line width
- Can recover and recycle energy
- Reversible logic
- Non-FLOPS computing – neural nets, molecular…

Sandia National Laboratories

Reversible Logic

- Reversible logic dissipates energy through "friction"
- If you run reversible logic at speed $\propto 1/n$, it will dissipate power $\propto 1/n^2$
- However, any design will have a parasitic power loss, so actual loss is not $\propto 1/n^2$, but

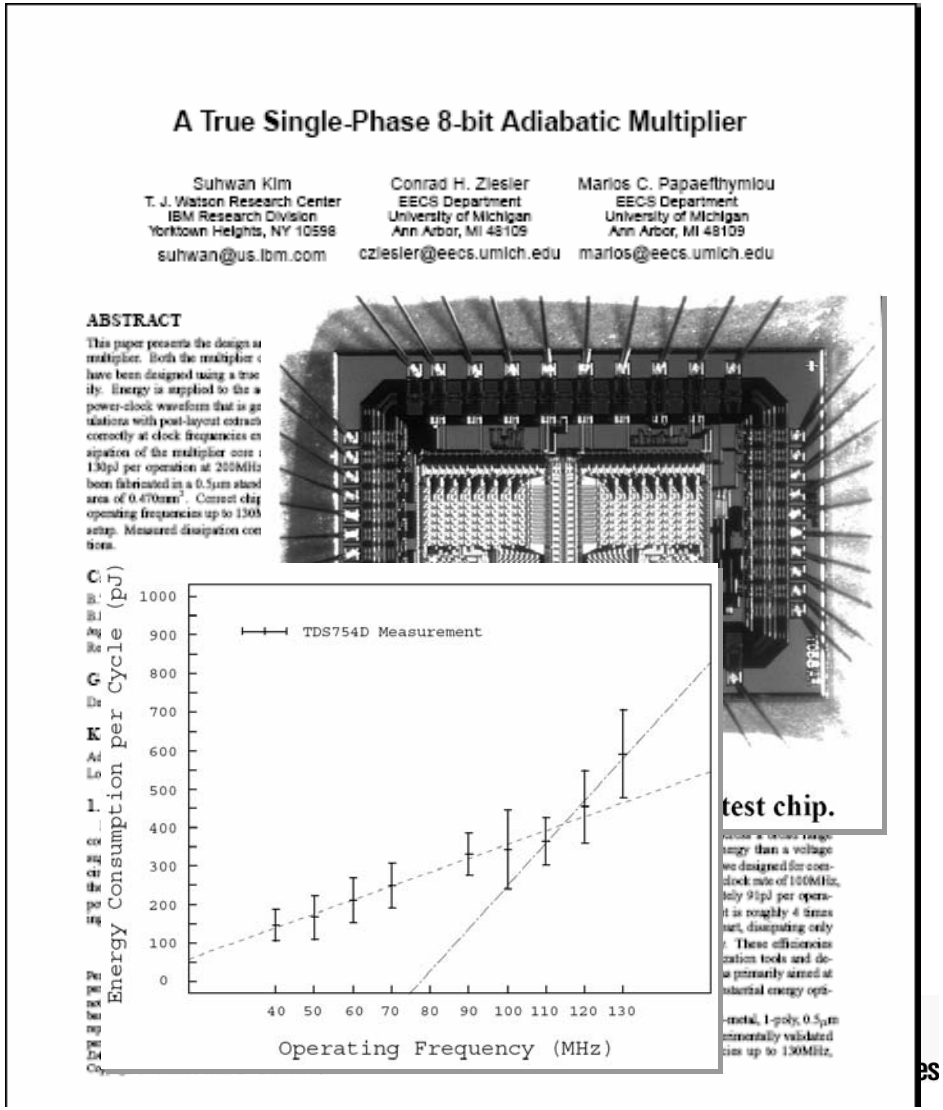
$$\text{Power} = \frac{P_0}{n^2} + P_{\text{parasitic}}$$

- Measured power down 4×, limit 2000×

Sandia National Laboratories

Reversible Multiplier Status

- **8×8 Multiplier Designed, Fabricated, and Tested by IBM & University of Michigan**
- **Power savings was up to 4:1**



A True Single-Phase 8-bit Adiabatic Multiplier

Suhwan Kim
T. J. Watson Research Center
IBM Research Division
Yorktown Heights, NY 10598
suhwan@us.ibm.com

Conrad H. Ziesler
EECS Department
University of Michigan
Ann Arbor, MI 48109
cziesler@eecs.umich.edu

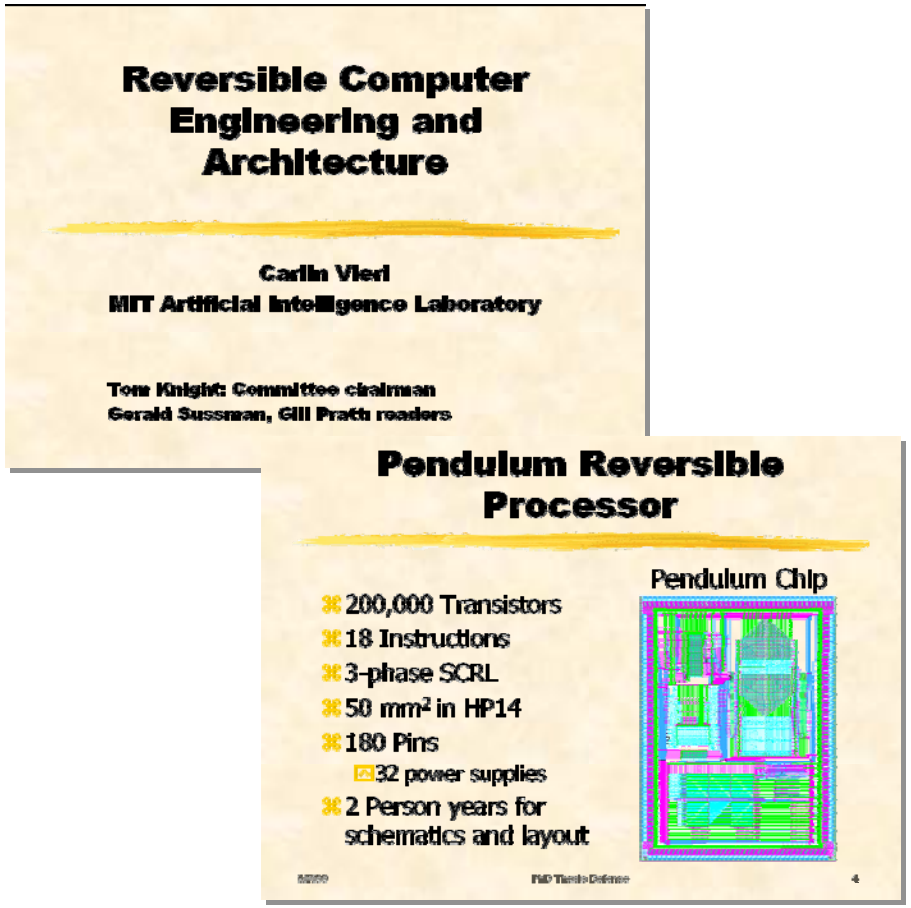
Marios C. Papaefthymiou
EECS Department
University of Michigan
Ann Arbor, MI 48109
marios@eecs.umich.edu

ABSTRACT

test chip.

Reversible Microprocessor Status

- Status
  - Subject of Ph. D. thesis
  - Chip laid out (no floating point)
  - RISC instruction set
  - C-like language
  - Compiler
  - Demonstrated on a PDE
  - However: really weird and not general to program with +=, -=, etc. rather than =

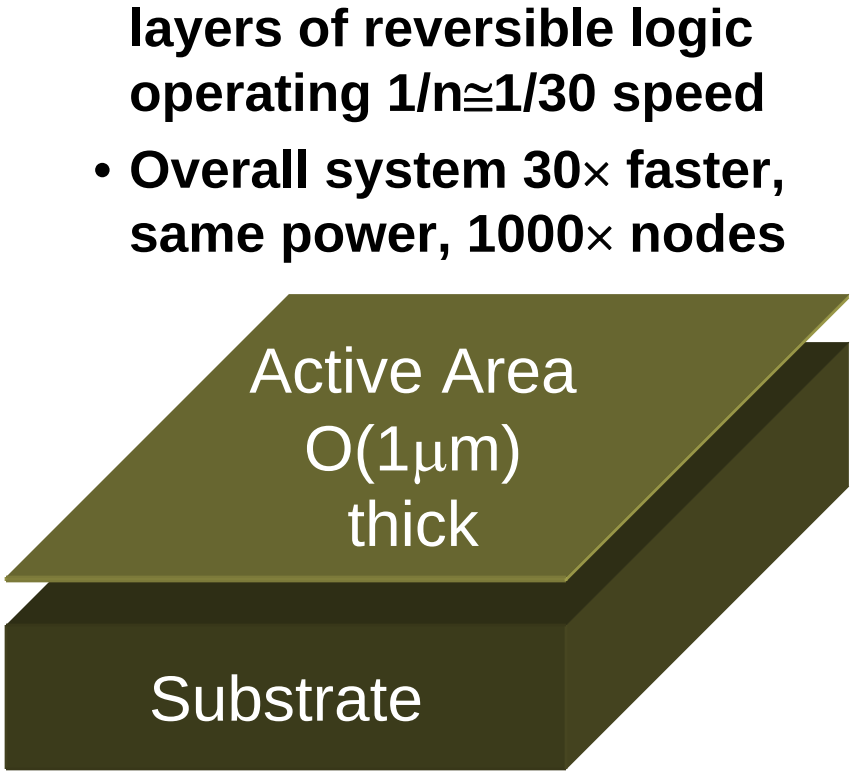


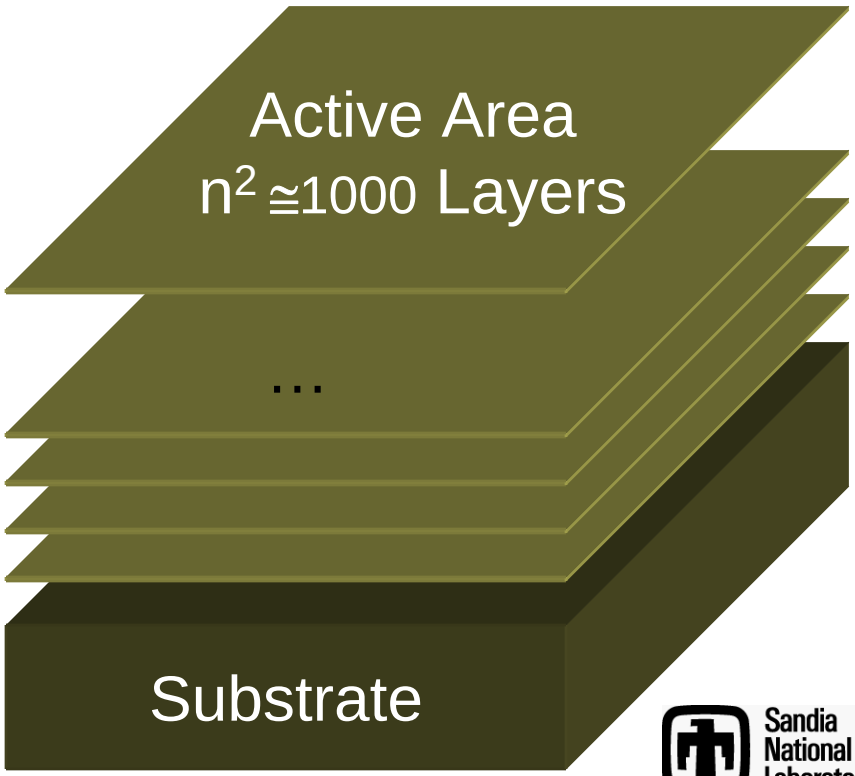
Sandia National Laboratories

Thought Model for Reversible Red Storm

- **Replace each Red Storm node with chips constructed from $n^2 \cong 1000$ layers of reversible logic operating $1/n \cong 1/30$ speed**
- **Overall system 30× faster, same power, 1000× nodes**

- **Will become feasible for small "line width"**





Sandia National Laboratories

Conclusions

- There exists at least one application that gives a valuable result to society and requires Exaflops of computing (more work needed here)
- Supercomputers may increase in power exponentially for a very long time (100 years?)
  - But CMOS driven by Moore's Law will flatline in 1-2 dozen years
- We can predict the end of Moore's Law with reasonable certainty (I gave a table with numbers)
- We can reach the limits of Moore's Law without major breakthroughs
- Applications modeling can be applied to distant future devices to yield quantitative information on computer architectures

Sandia National Laboratories

Backup

Sandia National Laboratories

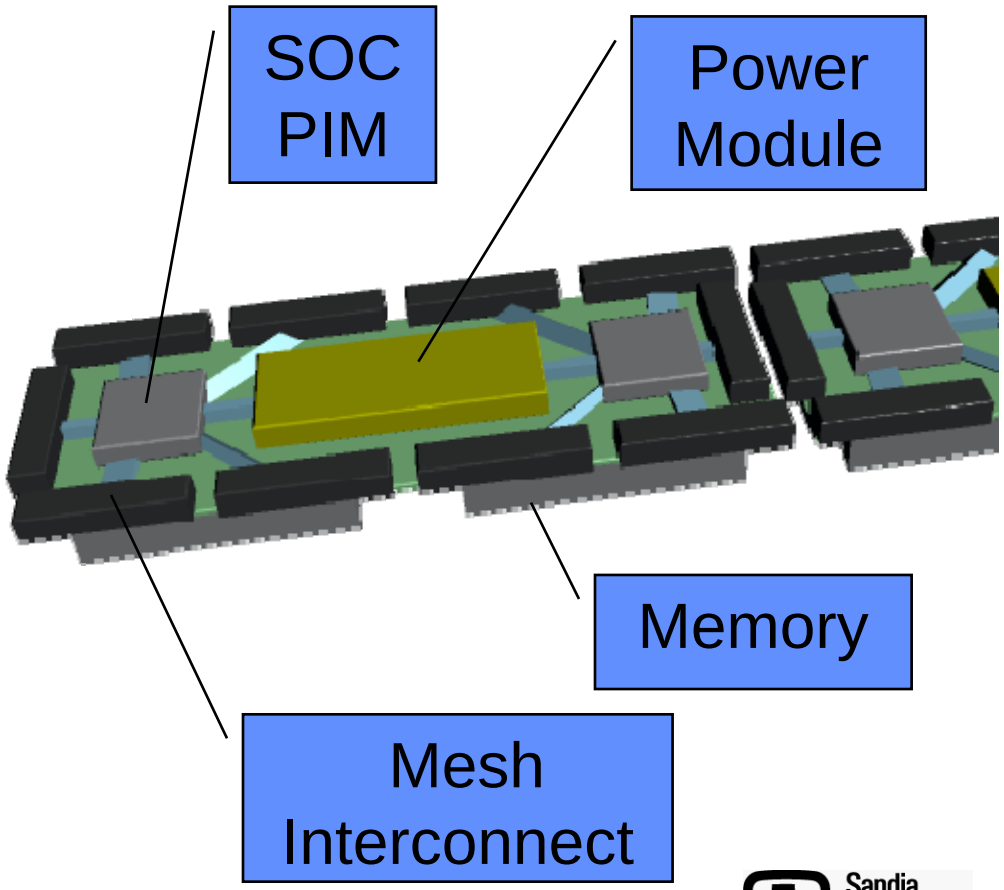
General Specifications at Physics Limit

	Red Storm	Limit μP Mode	Limit Turbo Mode
Nodes	10,000	200,000	2,000,000
Node Type	μP	μP	TBD – say 10 vector pipes
Clock	2 GHz	20 GHz	20 GHz
Flops/node	4 GFLOPS	40 GFLOPS	400 GFLOPS
Sys. Peak	40 TFLOPS	8 PFLOPS	800 PFLOPS
MPI Latency	2.5 μS	100 ns	N/A – no MPI
Power	2 MW	2 MW	2 MW

Sandia National Laboratories

Packaging for a Spatial Locality

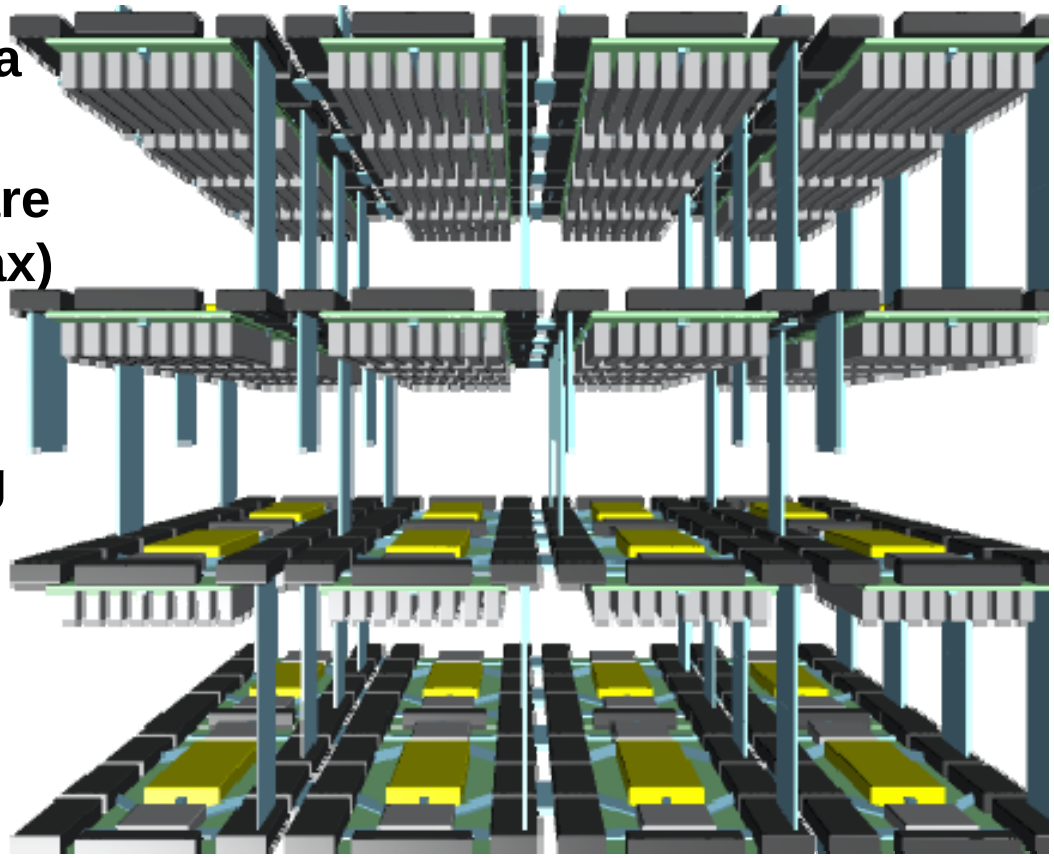
- **Basic Module**
  - **2 Nodes**
  - **Each node is an ASIC System On Chip Processor In Memory**
  - **Each node has memory under ASIC**
  - **Each module includes a power module**
  - **Six mesh Interconnects**
- **Modules connect end-to-end in “Shish Kabobs”**



Sandia National Laboratories

Packaging for a Spatial Locality

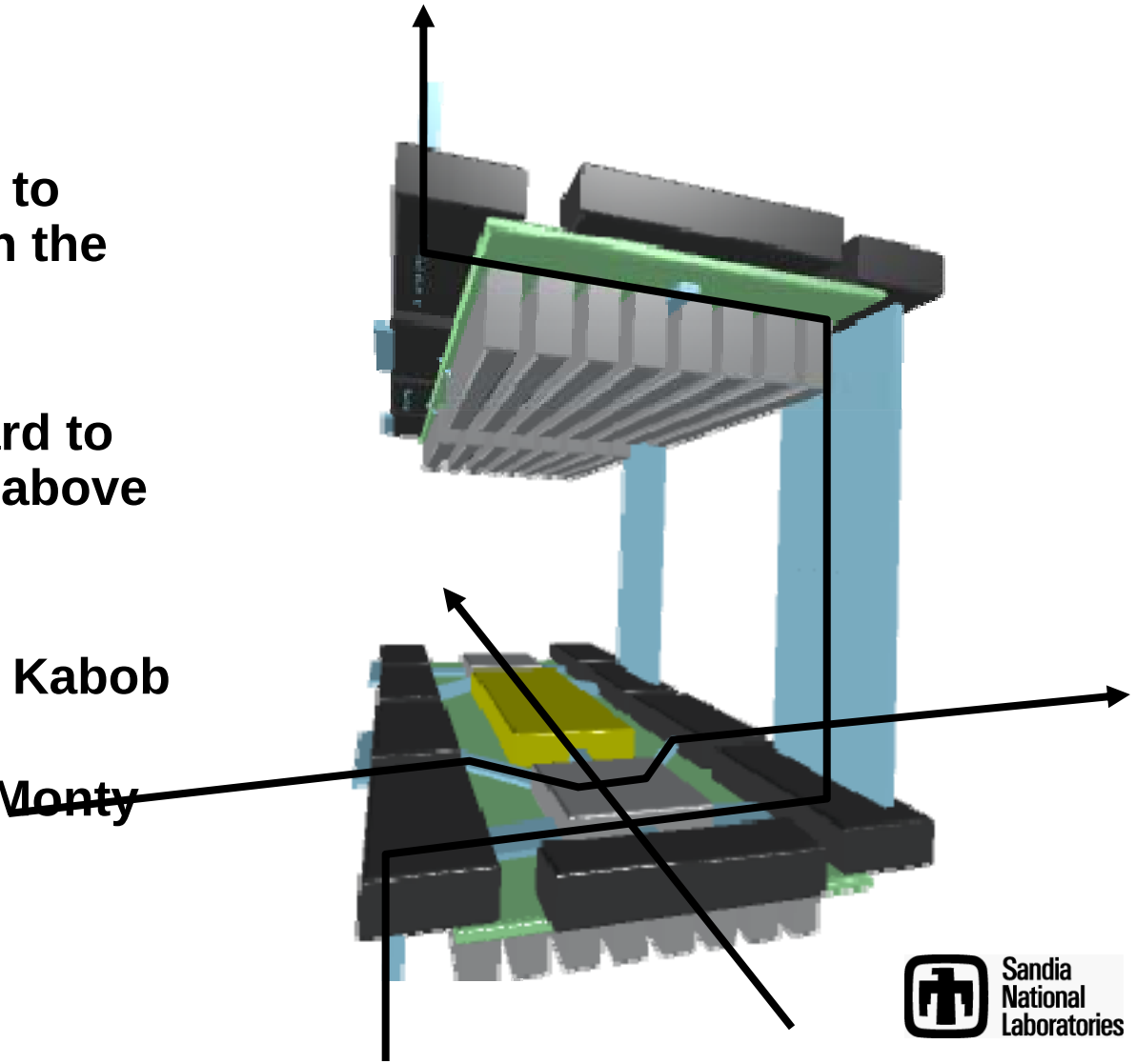
- Entire supercomputer is a single structure
- All mesh network wires are of constant length (8” max)
- Air flows front to back
  - General approach will work for liquid cooling as well



Sandia National Laboratories

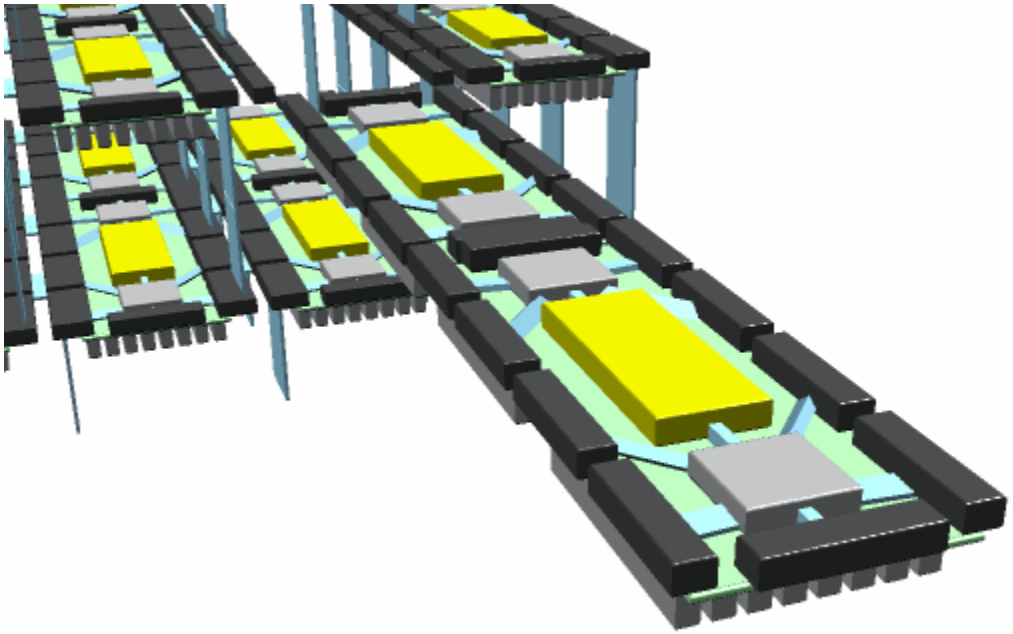
Nearest-Neighbor Interconnect

- X Dimension
  - From one board to another laying in the same plane – 2”
- Y Dimension
  - 8” from one board to another spaced above or below – 8”
- Z Dimension
  - Along the Shish Kabob – 4”
  - Name courtesy Monty Denneau IBM



Sandia National Laboratories

Maintenance



- Each “Shish Kabob” can be removed for maintenance
- Connects via side-connect technology
  - Similar to Cray shuttle connectors on T3E and X1
- Each Shish Kabob can be composed of segments to avoid limits on PC board technology
- Depth should be OK to 6’

Sandia National Laboratories

Backup: Landauer's Arguments

- Landauer makes three arguments in his 1961 paper
  - Kintetics of a bistable well
  - Entropy generation
- We review the second →

- Entropy of a system in statistical mechanics:

$$S = k_B \log_e(W)$$

W is number of states

- Entropy of a mechanical system containing a flip flop in an unknown state:

$$S = k_B \log_e(2W)$$

- After clearing the flip flop:

$$S = k_B \log_e(W)$$

- Difference $k_B \log_e(2)$

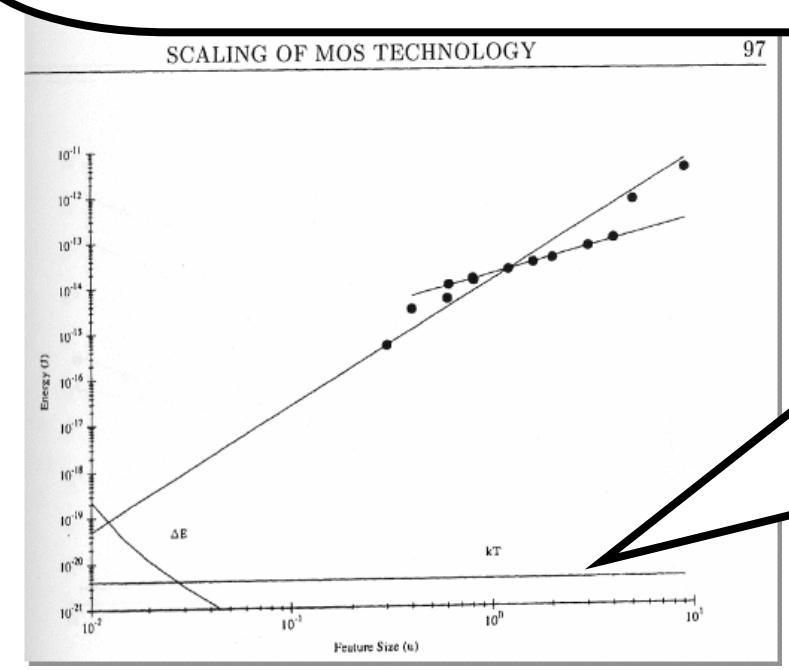
Backup: Landauer's Arguments II

- Second law of thermodynamics says entropy of universe must increase
  - Entropy is disorder
- Say you clear a computer memory of n bits. The computer's memory is initially disordered (arbitrary bits) but becomes ordered (all zero). Entropy goes down.
- However, entropy of universe must increase.
- Resolution is that the material of the memory chip becomes more disordered (hotter), offsetting the information in the memory
- A logic gate with multiple inputs but one output has fewer output states than input states: same idea

Backup: $k_B T$ Should Not Be A Surprise

This logical irreversibility is associated with physical irreversibility and requires a minimal heat generation, per machine cycle, typically of the order of kT for each irreversible function.
– R. Landauer 1961

Irreversibility and Heat Generation in the Computing Process



kT "helper line," drawn out of the reader's focus because it wasn't important at the time of writing
– Carver Mead, Scaling of MOS Technology, 1994

Sandia National Laboratories

Backup: Floating Point

- A floating point unit has about 100,000 gates
- About 20,000 gates will switch for each operation
- Therefore,

$$E_{FLOP} \approx$$
$$20{,}000\ E_{gate} \approx$$
$$2{,}000{,}000\ k_B\ T$$

- Landauer limit is:

  100 TFLOPS/watt

- Accounting for engineering losses, more realistic:

  10 TFLOPS/watt

- If a μP is 1% efficient, the probable limit for a microprocessor is:

  10 TFLOPS/watt chip

Sandia National Laboratories

Backup: What About Cryrogenics?

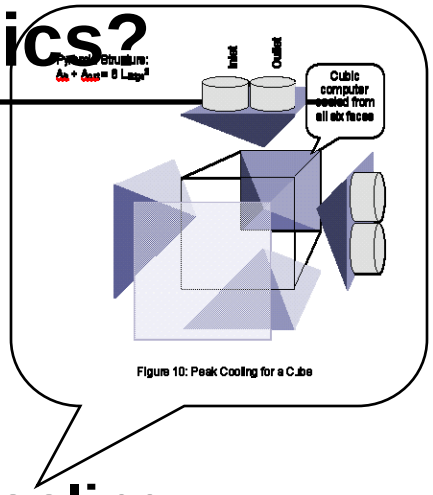
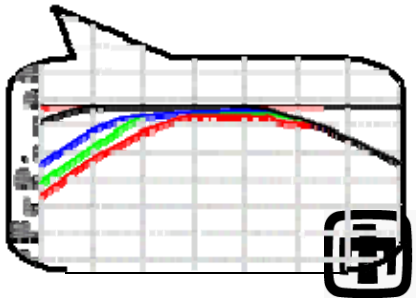


Figure 10: Peak Cooling for a Cube

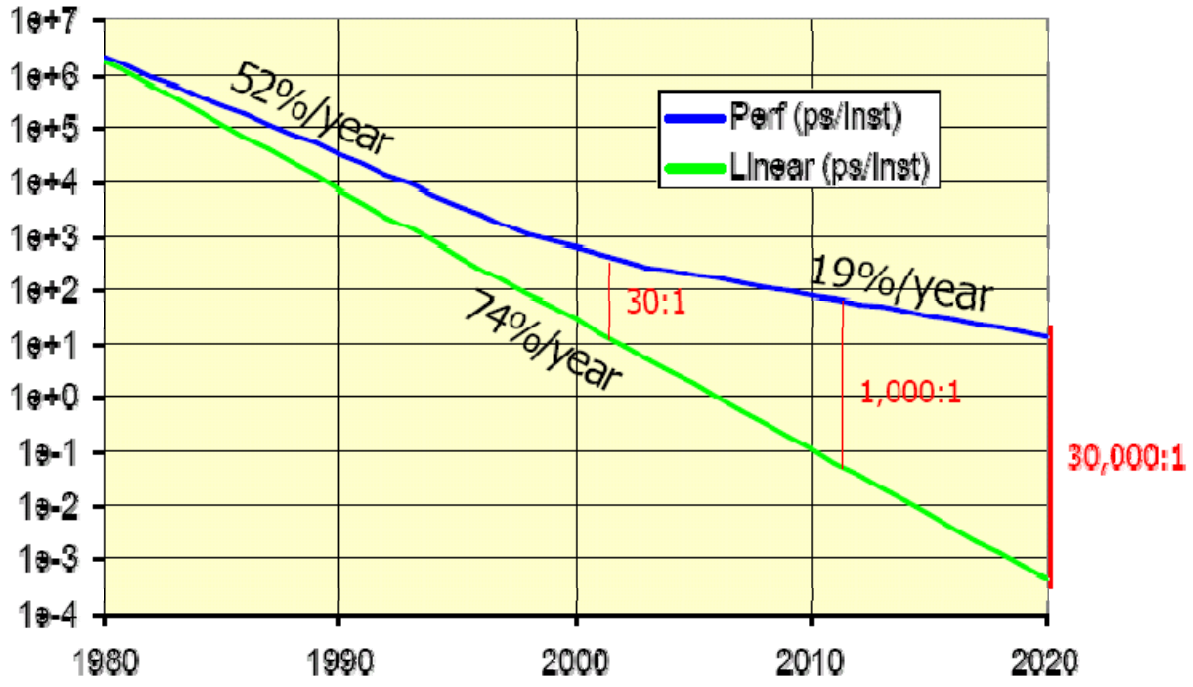
- Minimum power per logic op 100 $k_B T$
- Minimum power per FLOP 2×10^6 $k_B T$
- Analysis
  - At any T, performance may depend on cooling
  - Cutting T won't save power because of offsetting power in refrigerator, but may make cooling system more efficient
- However
  - Applications modeling indicates DOE apps aren't especially dependent on cooling
- Conclusion: Use room temperature



Sandia National Laboratories

Backup: Authority on μP Efficiency

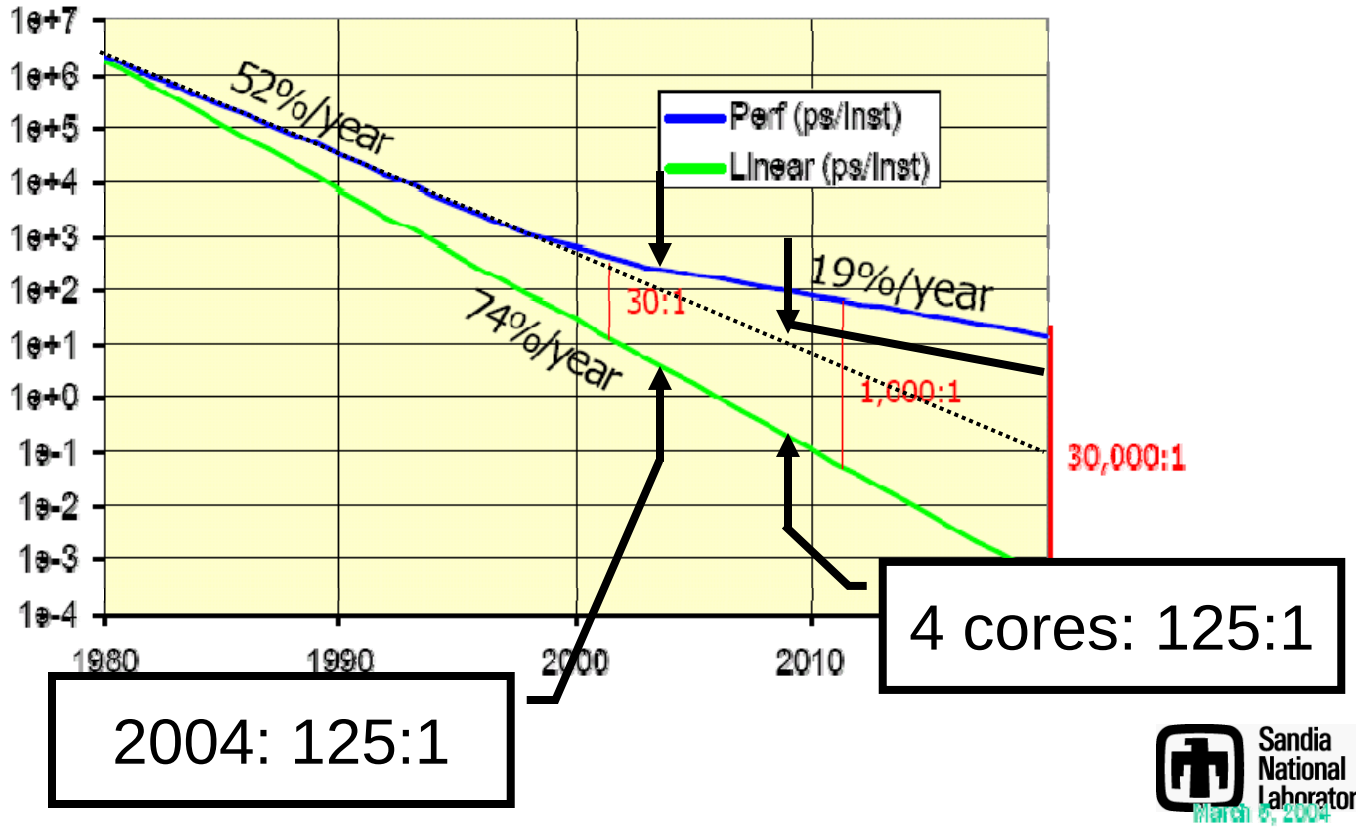
Data parallelism realizes full potential of increased transistor count



Citation:
Bill Dally,
ASCI PI
Meeting 2004

Backup: Authority on μP Efficiency

Data parallelism realizes full potential of increased transistor count



Citation:
Bill Dally,
ASCI PI
Meeting 2004

Backup: Languages

- For many years, computer languages have targeted higher programmer productivity, trading easy programming for higher resource consumption during execution. This was believed to be OK because Moore's Law would cut the excess cost over time. Not so anymore
- Need to study languages for mature "irreversible logic" computers that are both easy to use and avoid excessive use of resources

Sandia National Laboratories

Backup Slide: Analog Computing

- Floating Point Energy/Op
  - $20{,}000 \times 100 \times k_B T =$
  - $2 \times 10^6 \, k_B T$
- Analog Energy/Op
  - $k_B T \log_e(\text{“\# states”})$
  - $k_B T \log_e(2^{64})$
  - $64 \, k_B T \log_e 2$
  - $44 \, k_B T$
- Analog 45,000 more efficient

- Heisenberg Uncertainty Principle
  - $\Delta E \, \Delta t \geq h/(2\pi)$
- Waiting Time
  - $\Delta E = 2^{-64} \times 64 \, k_B T \log_e 2$
  - $\Delta t \geq \dfrac{h}{2\pi \times 2^{-64} \times 64 \, k_B T \log_e 2}$
  - $\Delta t \geq$ ~3 hours
- Analog really slow

Sandia National Laboratories

Conclusions

- When we look into the future of supercomputing
  - We see some haze
  - However, the end of the road is becoming visible through the haze
- Knowing the end of the road helps now
  - What applications should we anticipate solving?
  - Some software written today will run on end-of-road supercomputers. What architectures will/will not be around?
- Other roads follow

Sandia National Laboratories