Towards Integrated Power-Balanced Computing

Presented by:
Allan Snavely
PMaC Lab
Performance Modeling & Characterization Lab

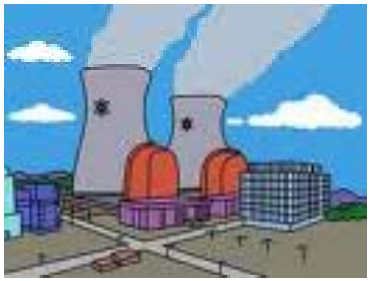
SciDAC10 July 12, 2010

SDSC
SAN DIEGO SUPERCOMPUTER CENTER

PMaC
Performance Modeling and Characterization

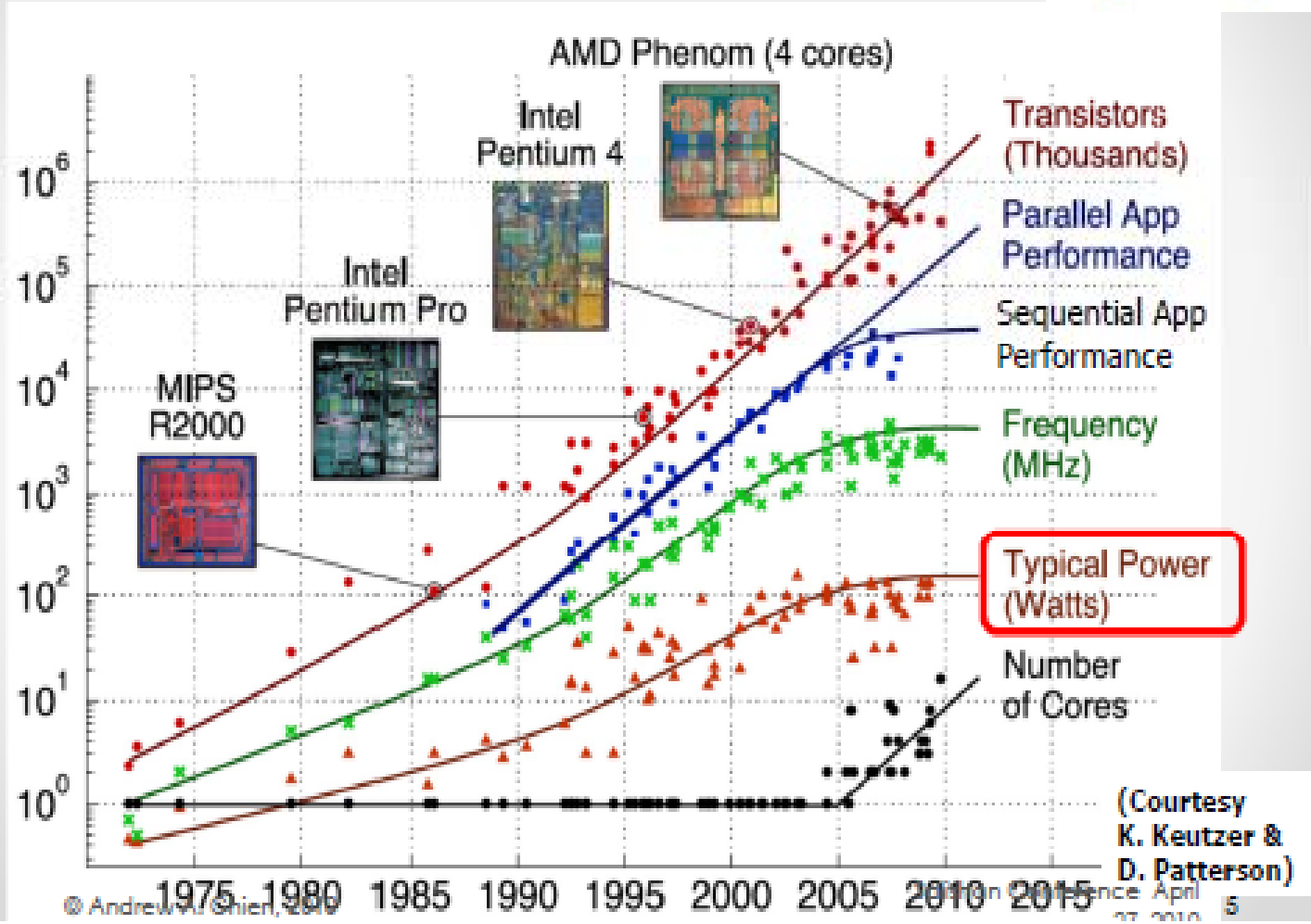
An Exa times a Nano is a Giga!

- **A future system circa (2017?) may be capable of 10^{18} flops**
- **The energy required for an operation today is O(nJ) i.e. 10^{-9} joules**
  - **I remind you that a joule a second is a watt**
- **Then 10^{9} joules/s is a Gigawatt**



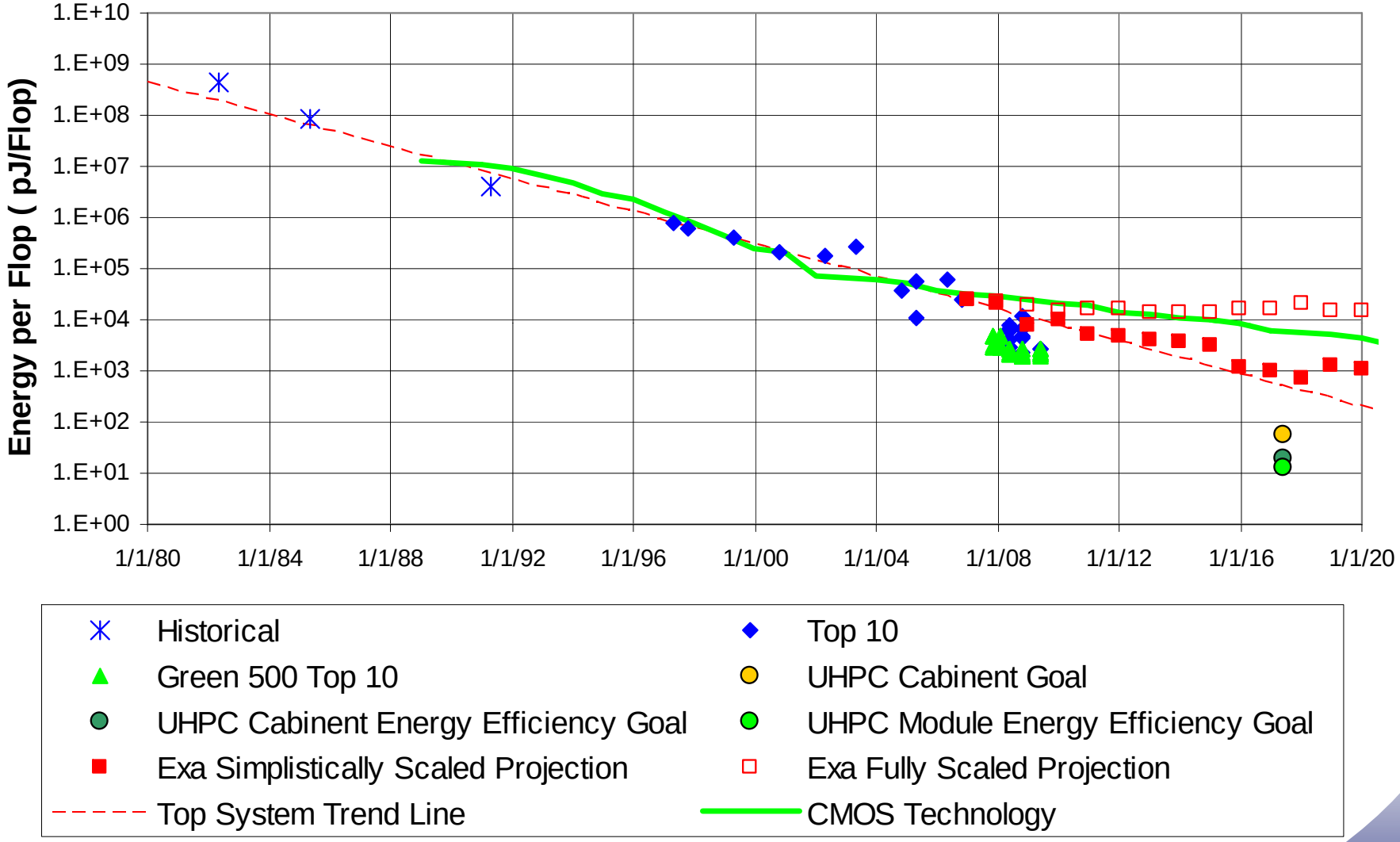
A supercomputer center of the future

Current trends are high-parallelism lower power but fundamentals constrain power



Energy roadmap and projections

courtesy Peter Kogge

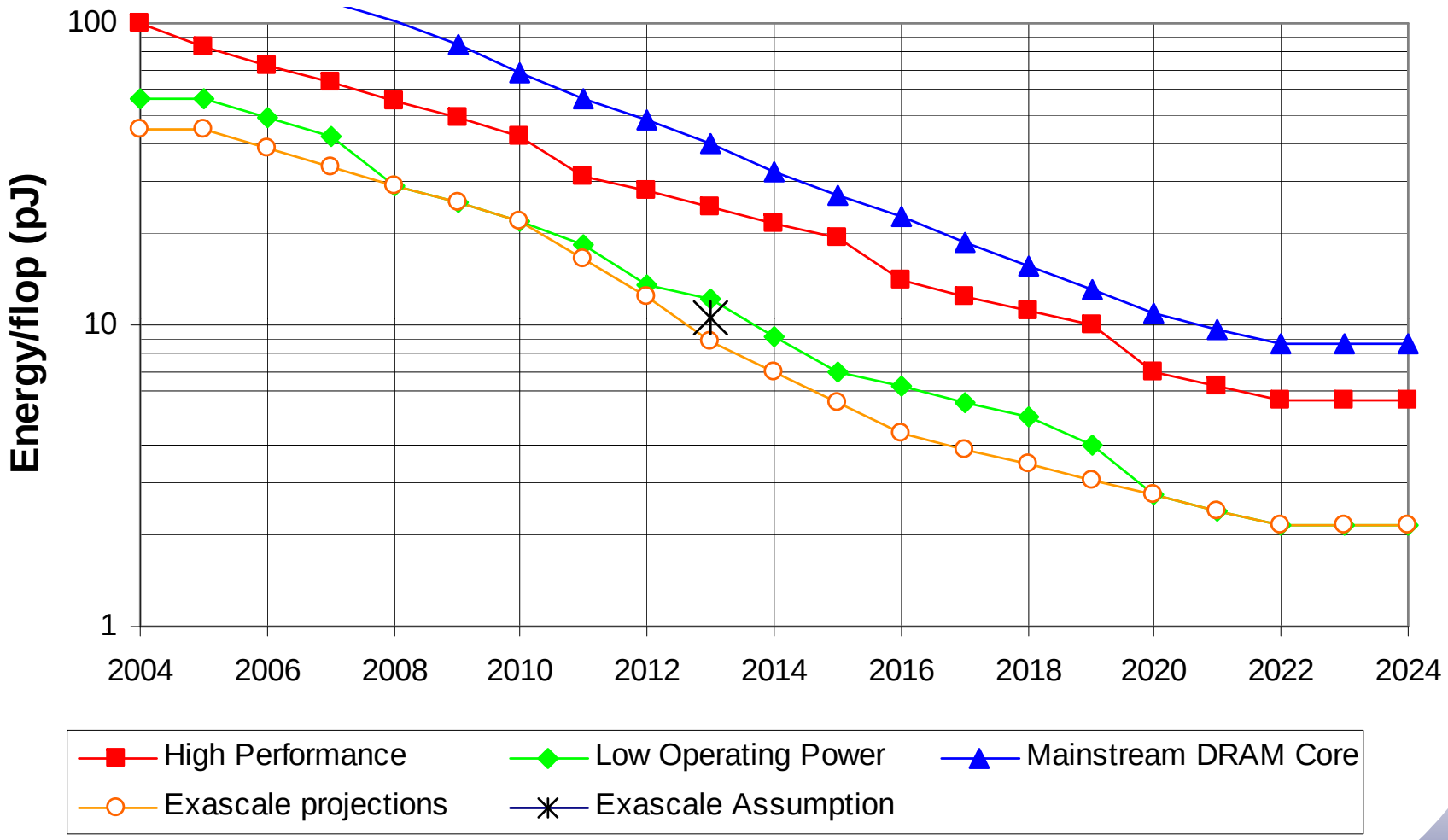


Trend Lines on Prior Chart

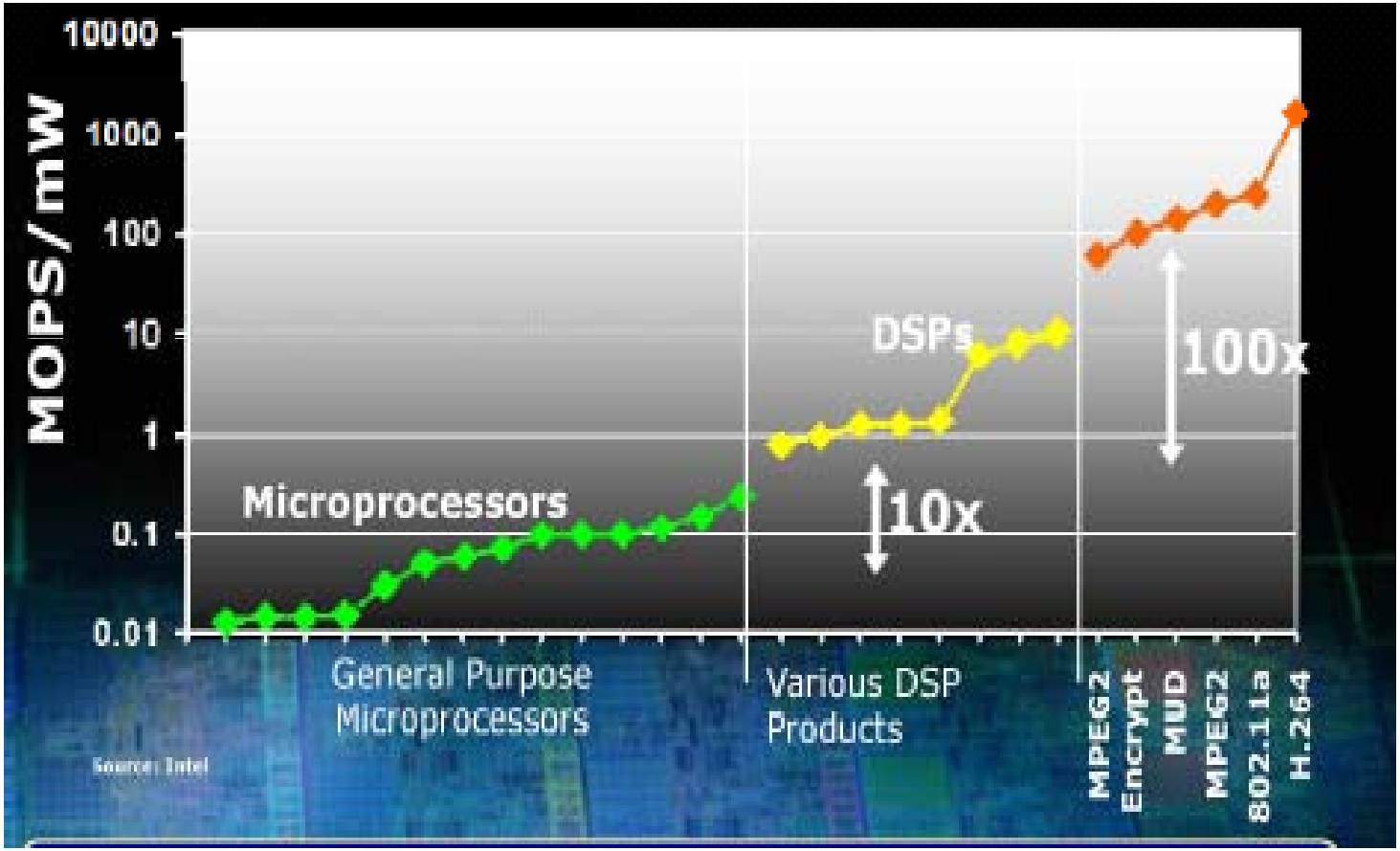
- **Red "Top System Trend Line"**
  - **Fitted through data for top system from Top500**
- **Green "CMOS Technology" computed from CV^2**
  - **Where C assumed to scale as feature size**
  - **And V scaled as V_{dd} for high power logic**
  - **And then "scaled" to allow comparison with "Top" trend**
    - **"Centered" at 1988**
- **"Exa .. Projections" use projection data made during Exascale Technology Study for "business as usual high performance microprocessor-driven architectures"**
  - **"Simplistic" = Highly Optimistic scenario**
  - **"Fully Scaled" = Highly Pessimistic scenario**

FPU Energy per Flop

from DARPA Exascale Hardware Report

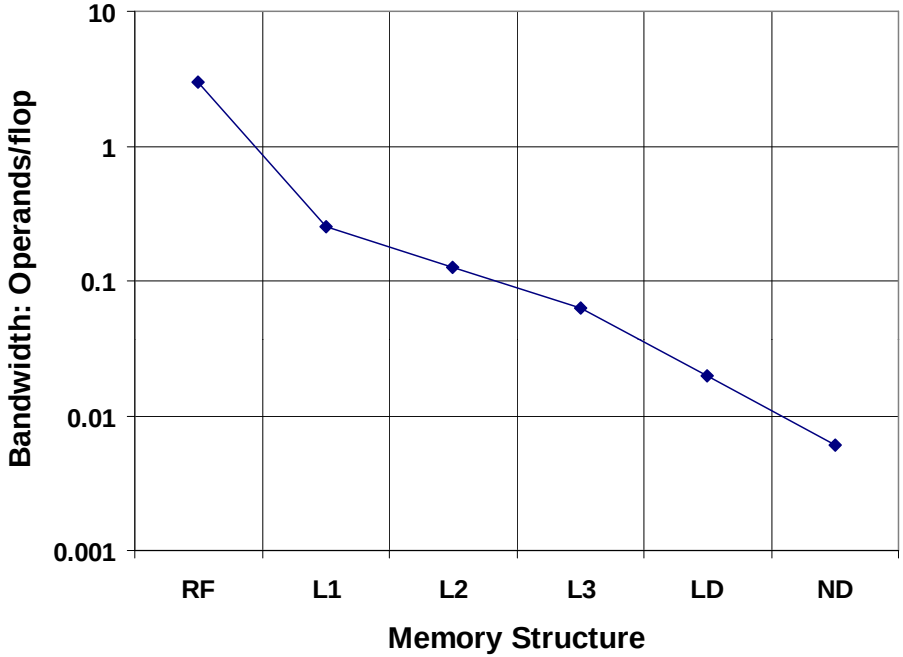


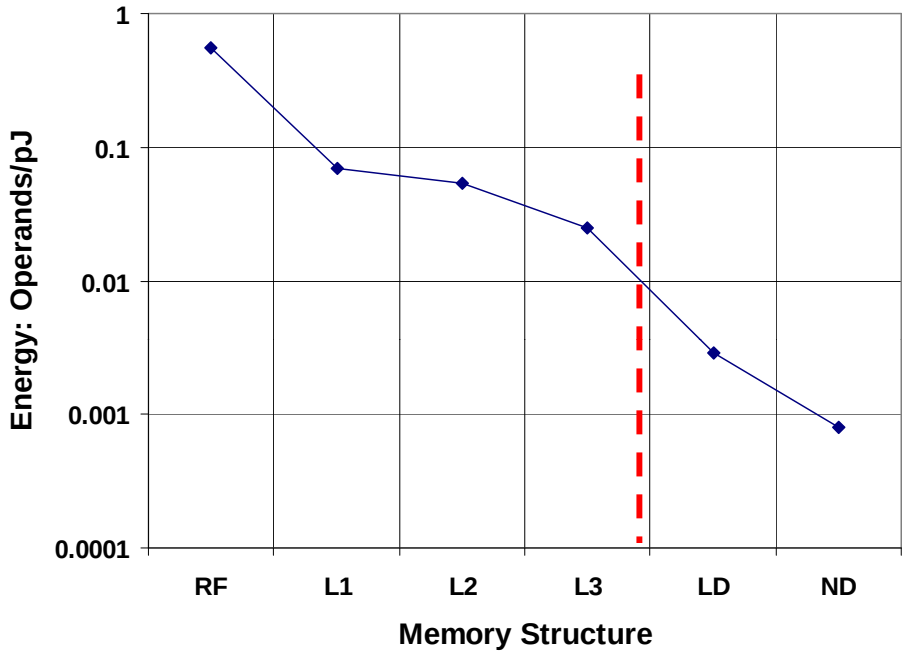
Customization Can Save Energy



Moving Data Remains Fundamental

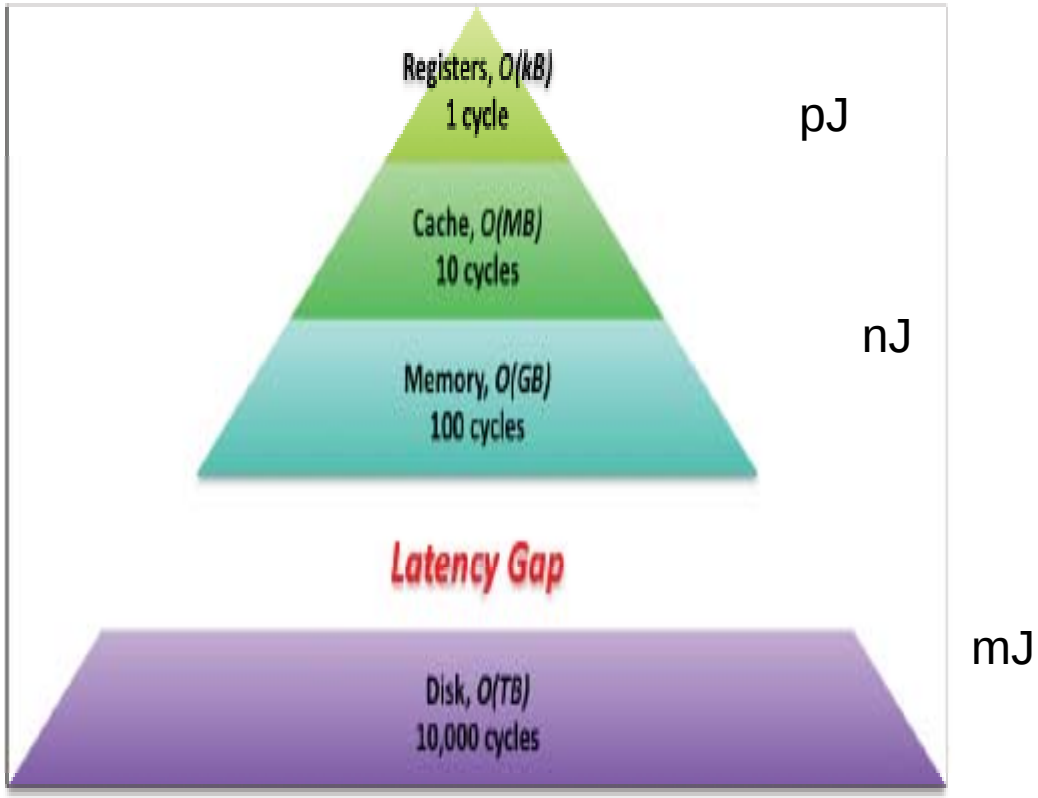
courtesy Peter Kogge





- Both bandwidth and energy tapers drop off strongly once going “off-core”
- Bandwidth tapers declines gracefully thereafter
- Energy taper declines sharply going off chip

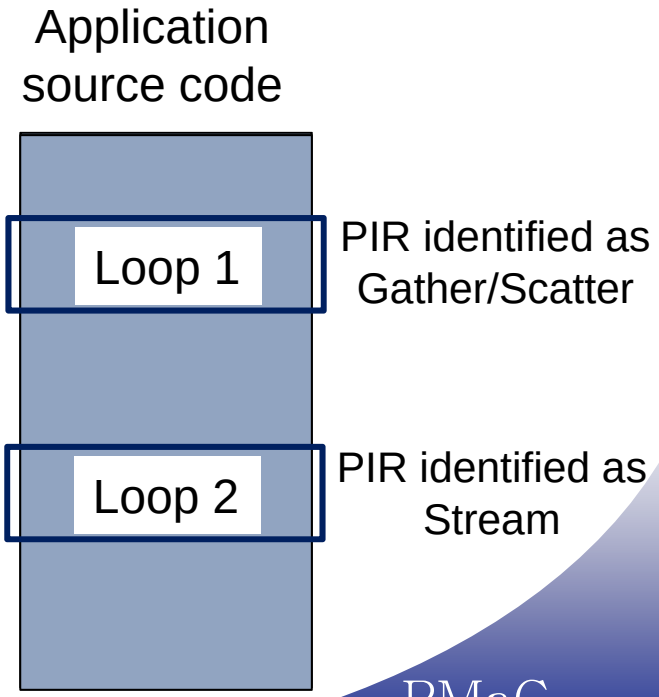
Going forward energy budget is allabout data motion



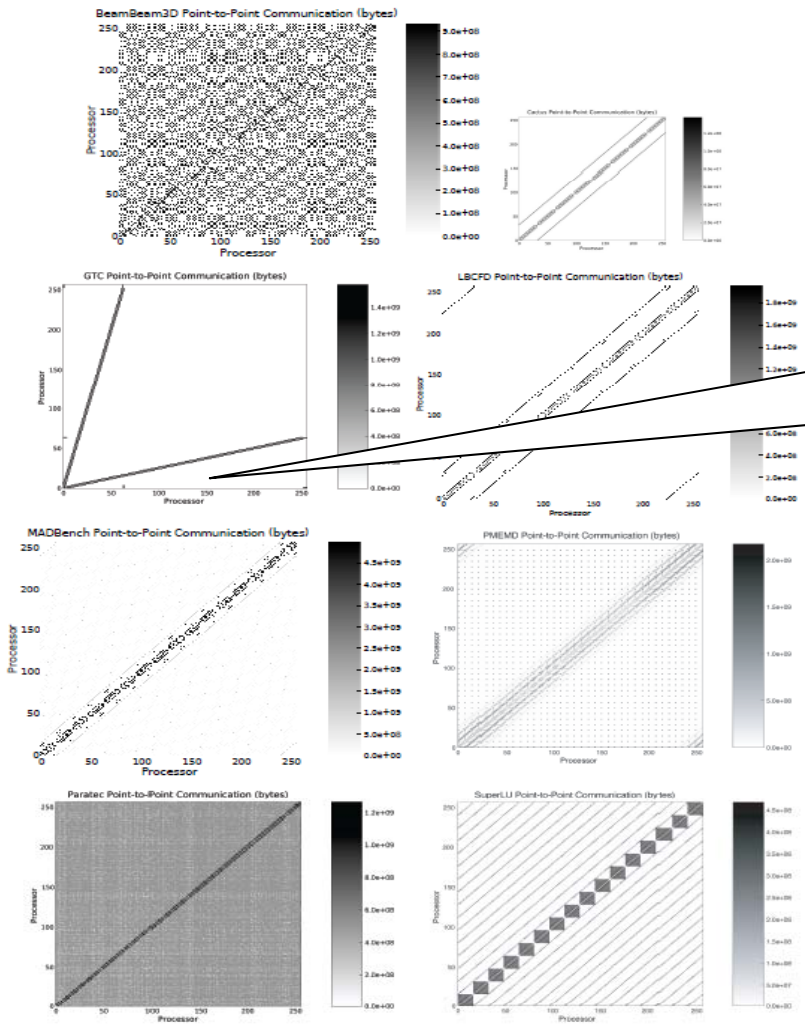
We can't afford to move an unused bit further than need be

Idioms are patterns of data access

- **PIR Goal: Identify common memory access patterns in source code**
- **Idiom: A common memory access pattern**
- **Static analysis to recognize patterns in the source code, identify loops as idioms.**
- **Built within GCC – portable to most systems**
- **An idiom is pattern used often in applications (e.g. loops) 8 common idioms:**
  - **Stream (A[i] = A[i] + B[i])**
  - **Transpose (A[i][j] = B[j][i])**
  - **Gather (A[i] = B[index[i]])**
  - **Scatter (A[index[i]] = B[i])**
  - **Random Access (A[i] = B[rand(seed)])**
  - **Reduction (s = s + B[i])**
  - **Stencil (A[i] = A[i-1] + A[i+1])**
  - **Gather reduction (s = s + B[index(i)])**



Communication Idioms



Why not turn off the part of the network you don't need?

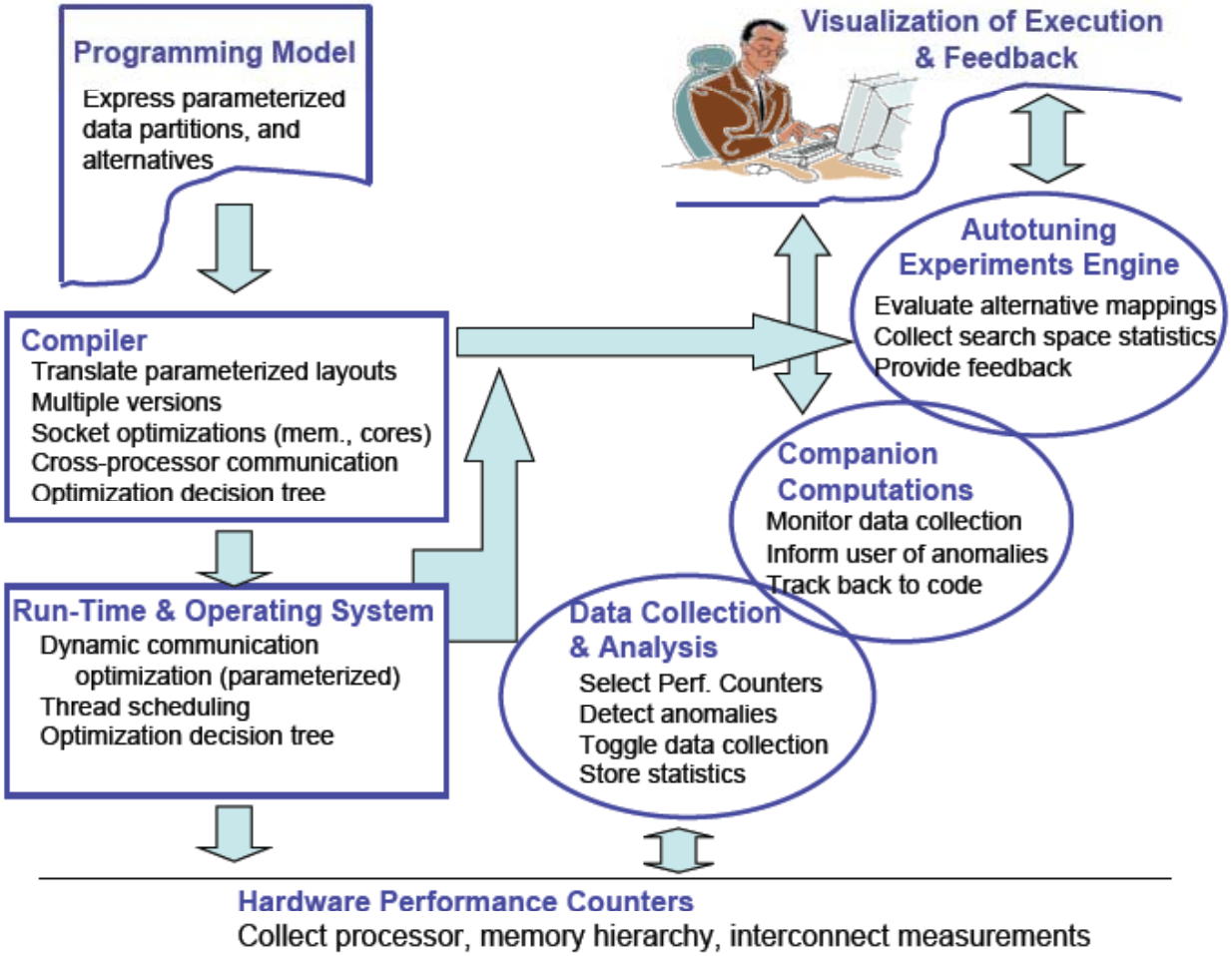
Meta-Idioms (Dwarfs, Motifs)

- **Molecular Dynamics**
  - **Anton, Grape, all-to-all, hardware for force calculations**
- **Signal processing**
  - **DSP, SIMD, specialized prefetching and transpose**
- **Graph algorithms, decision class (Chess)**
  - **Data flow, memory synch, PIM**
- **Inverse problems, Data Mining**
  - **Robust memory and I/O**
- **Dense linear algebra**
  - **Fast FPUS**

Not everything has to be at the same time

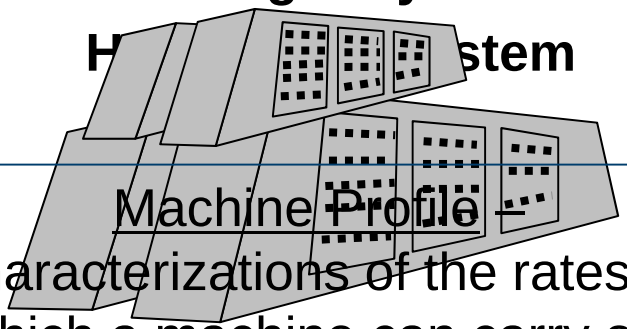
- **Voltage scaling processors**
- **"Wires are cheap turning them on is expensive" – Shekhar Borkhar**
- **10x10 architecture of Andrew Chien**
- **In HPC we run the same thing over and over again**
  - **(and so why not learn the data access patterns and benefit from them?)**

Integrated Wetware/Software/Hardware



Basic piece of a performance/power model

HPC Target System



HPC Application

HPC Application

Machine Profile – characterizations of the rates at which a machine can carry out fundamental operations

Characteristics of HPC system – Machine Profile

Application signature – detailed summaries of the fundamental operations to be carried out by the application

Requirements of HPC Application – Application Signature

Measured or projected via simple benchmarks on 1-2 nodes of the target system

Collected via trace tools on base system

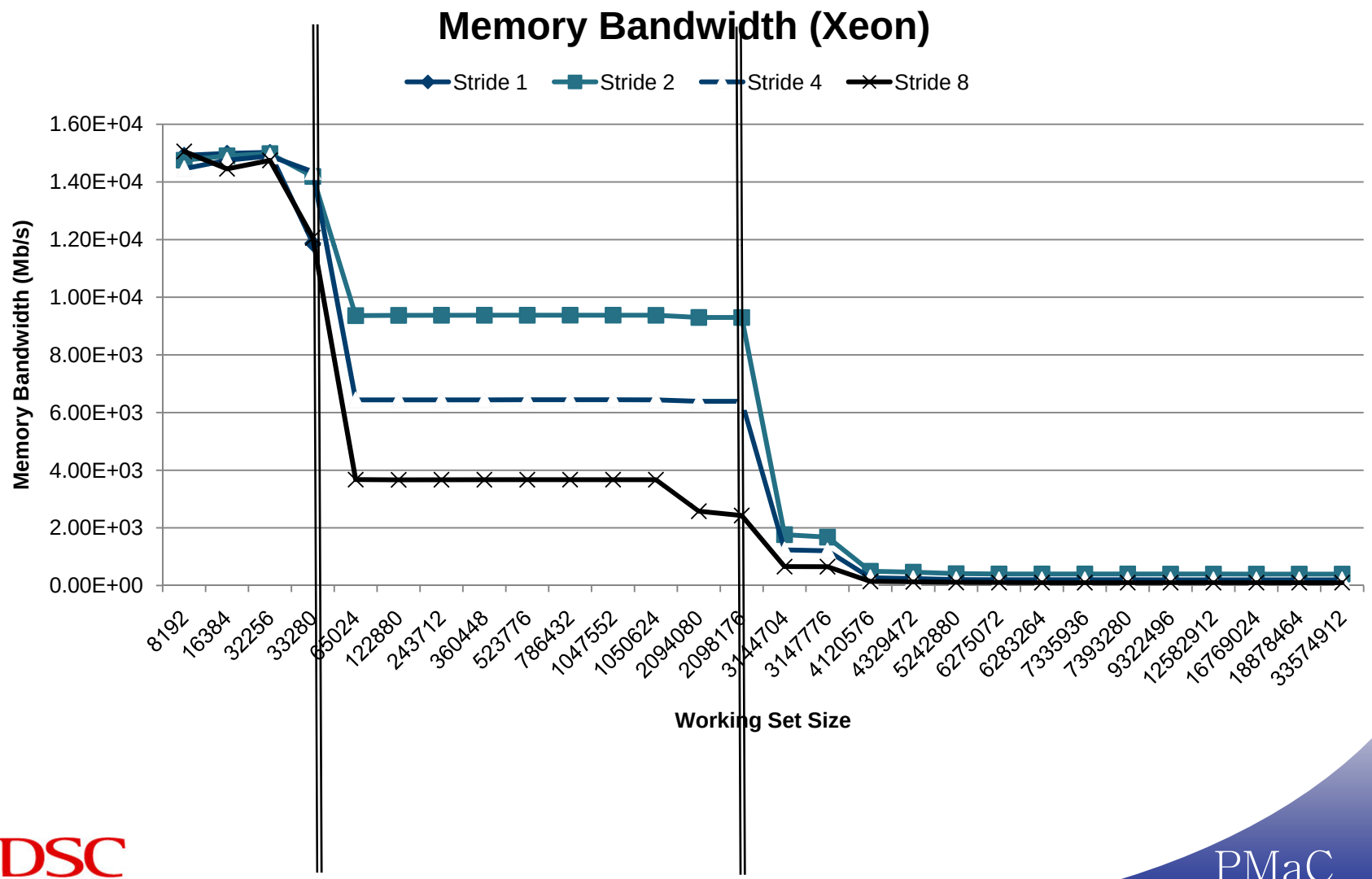
Performance of Application on Target system

Convolution Methods

map Application Signatures to Machine Profiles

produce performance prediction

MultiMAPS

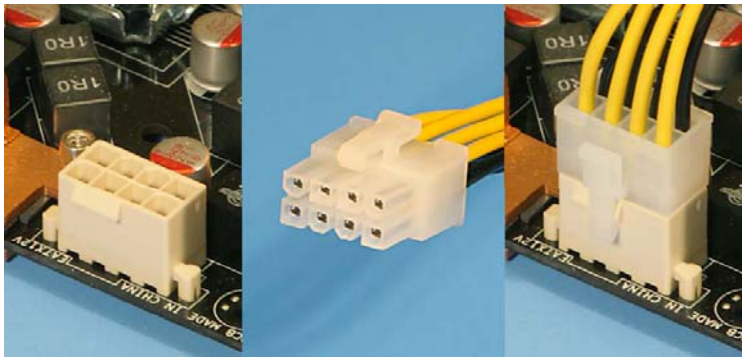


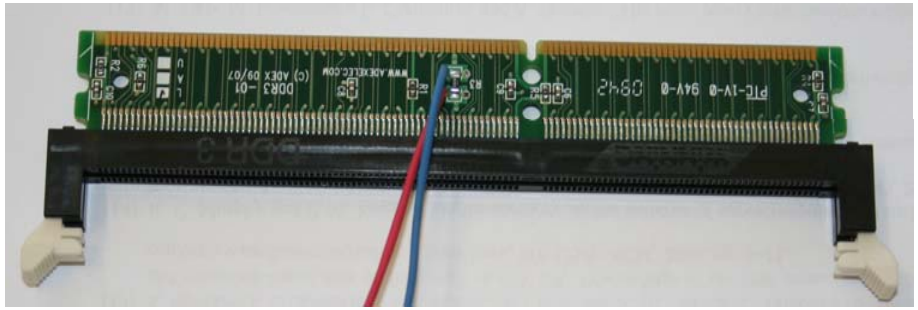
SDSC
SAN DIEGO SUPERCOMPUTER CENTER

PMaC
Performance Modeling and Characterization

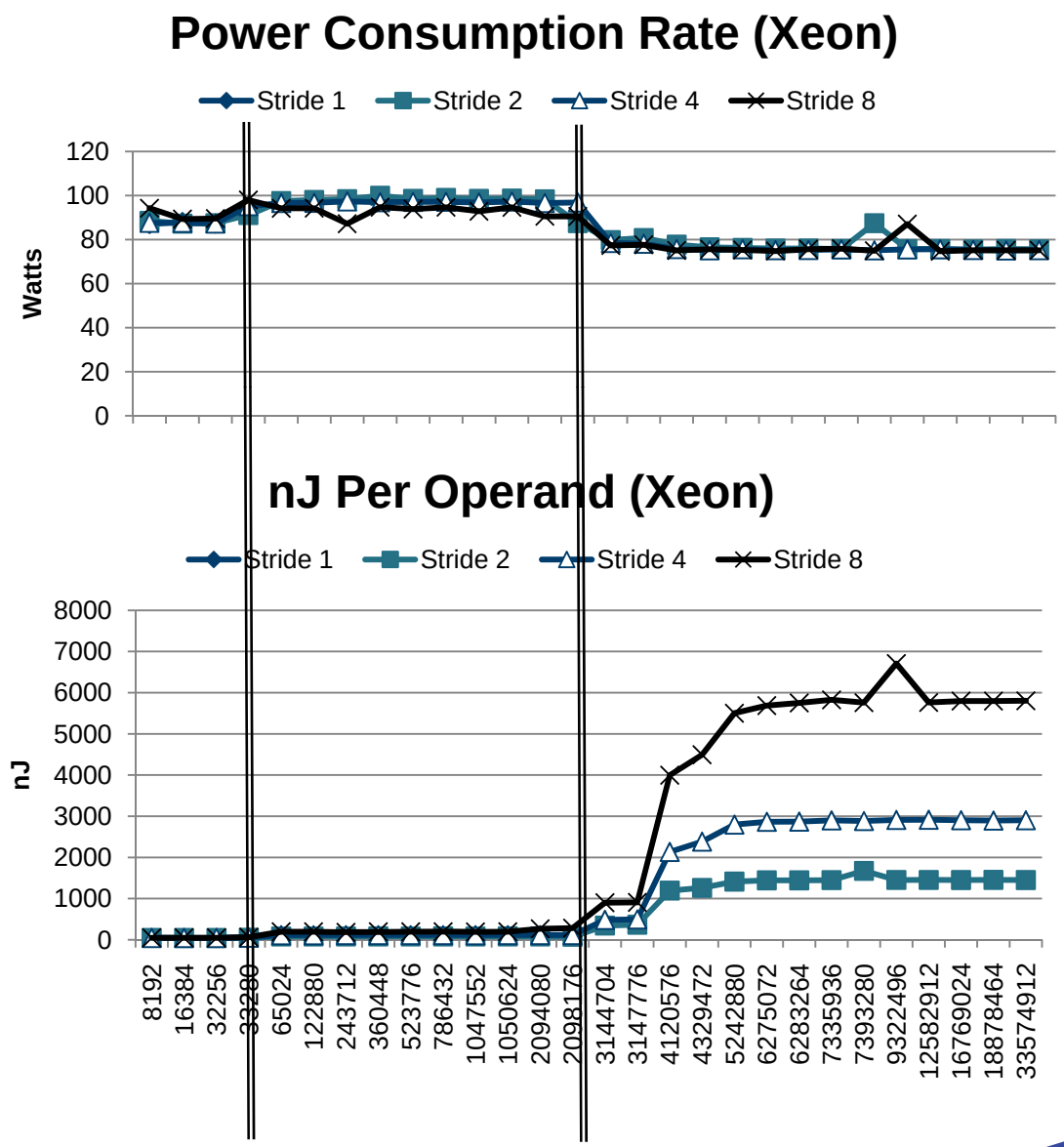
Adding Power Measurement Harness

- **Insert shunt resistor between the power supply and CPU or DIMM**
- **Measure AC power consumption of node**
- **Collect data with DAQ (NI USB-6255)**
- **Calculate Power Consumption**





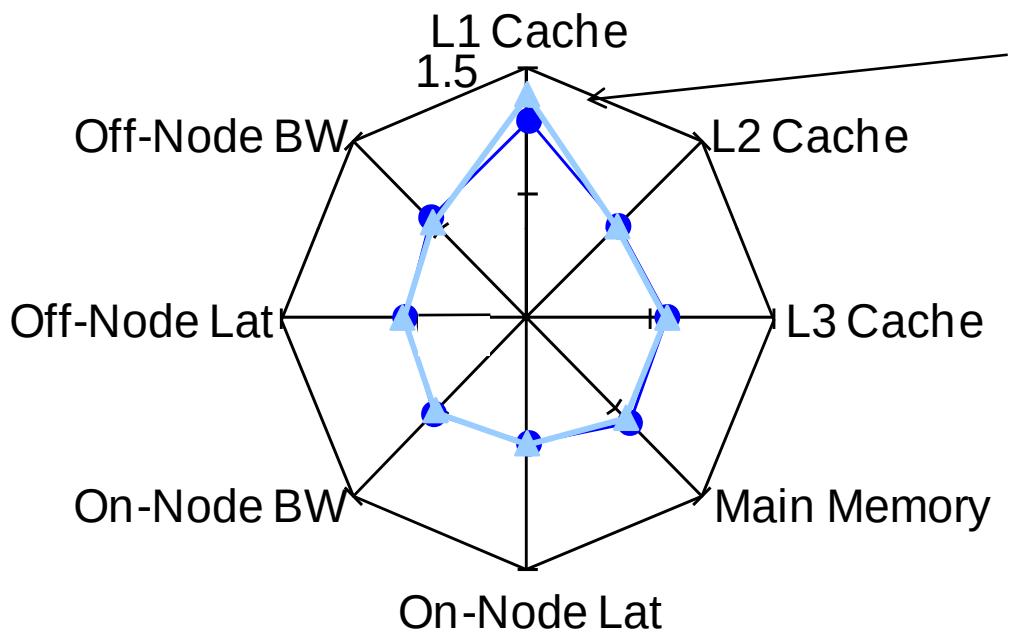
Power Profile of Xeon Dual Quad-core



SDSC
SAN DIEGO SUPERCOMPUTER CENTER

PMaC
Performance Modeling and Characterization

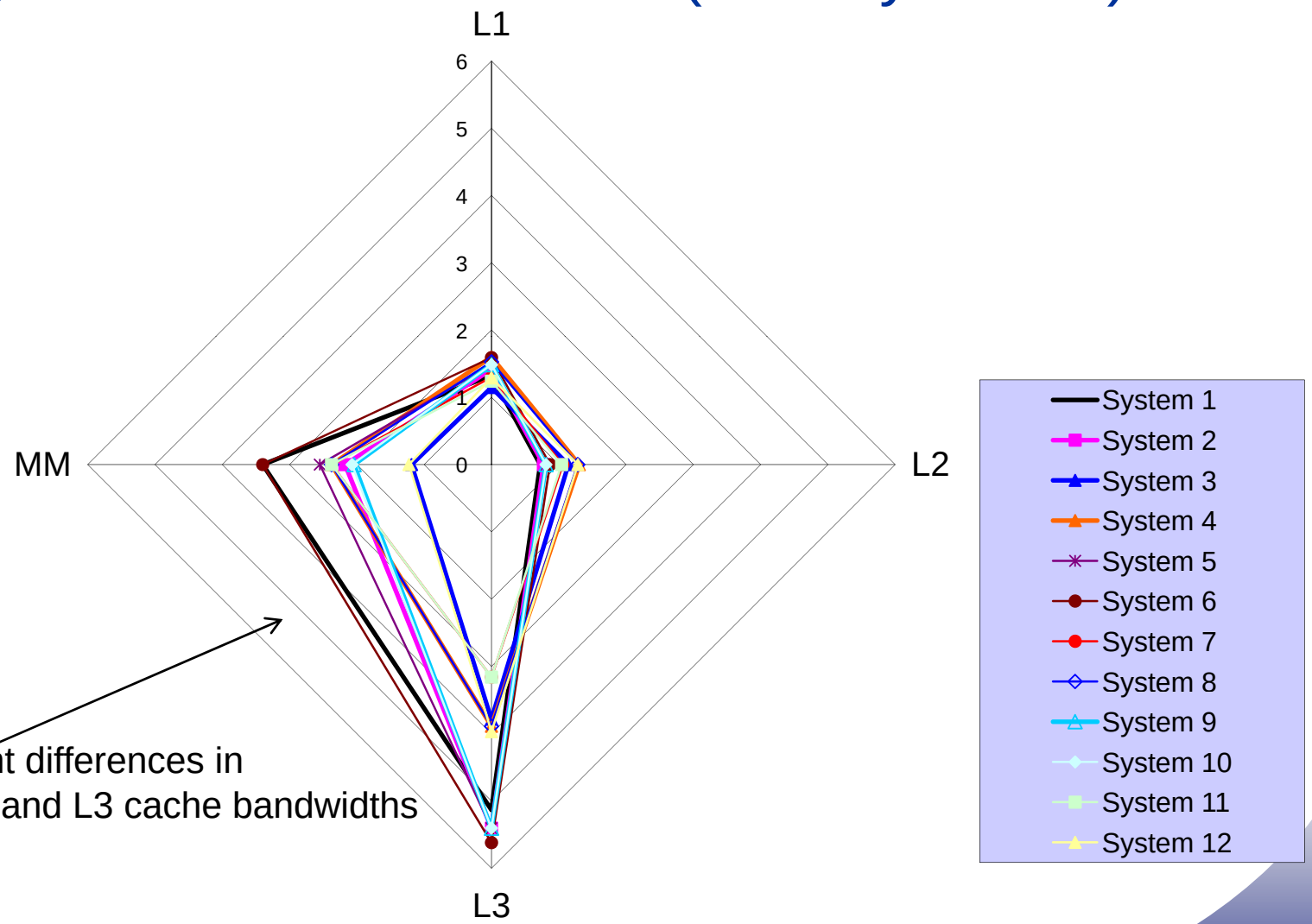
A Kiviat diagram ("radar plot") is a useful way to visualize performance responses



Think of a "bulge" as a steep performance response gradient in that dimension

Example: modeled ops/watt response of WRF to 2x improvement in each machine dimension

Memory profiles as Kiviat diagrams of systems out to 2012 (anonymous)

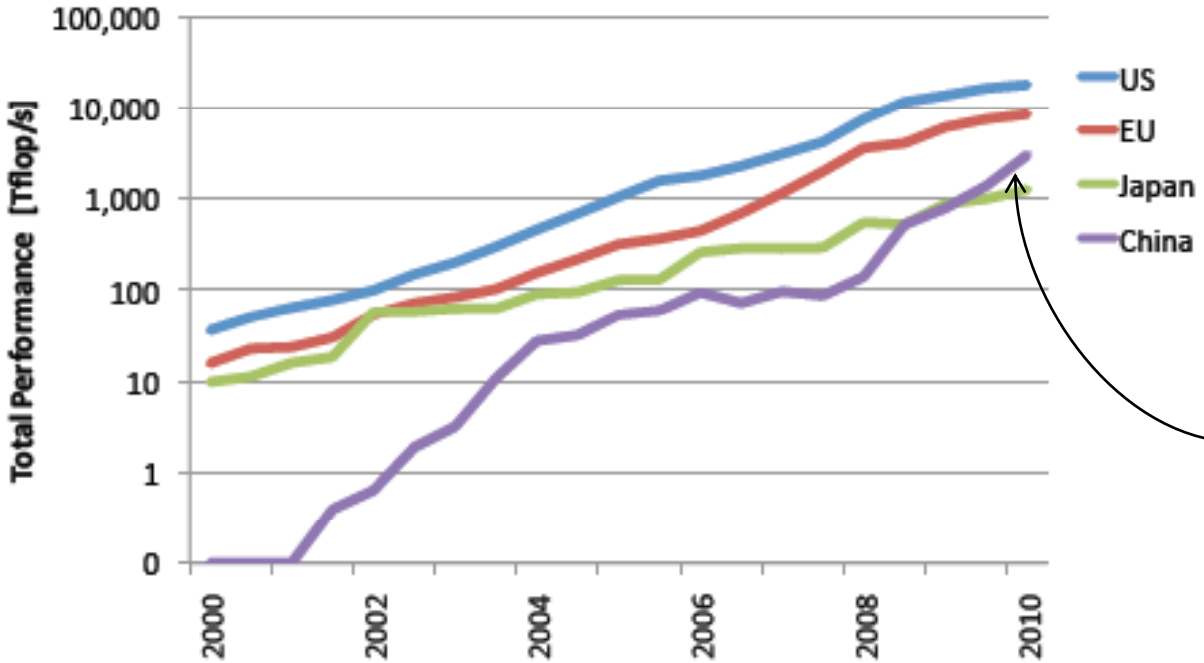


Very significant differences in
Main memory and L3 cache bandwidths

Assuming the “better” systems are more energy hungry are they worth it?

Exascale in 2017?

Data courtesy Jack Dongarra



How about Exascale in 2013?

There is not anything to prevent a nation-state from spending 1 GW on 1 Exaflop if they deem it in their national interest