

Forecast: Cloudy, with scattered storms

Or, why the world is scary. But fun.

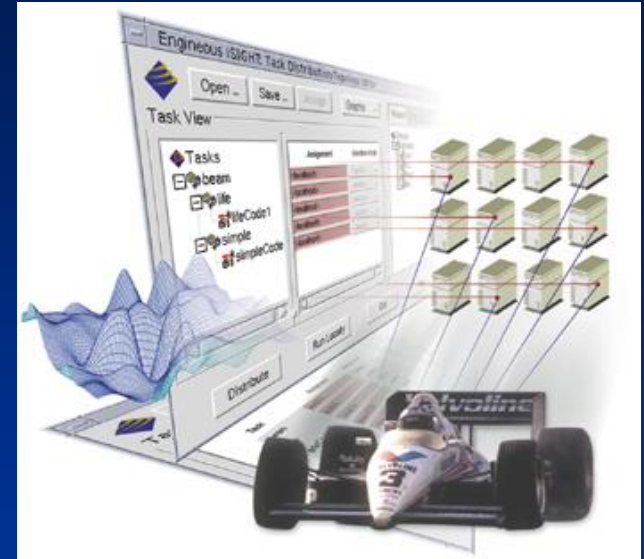
Dr. Dan Andresen (dan@k-state.edu)

CIS115, February 25, 2014



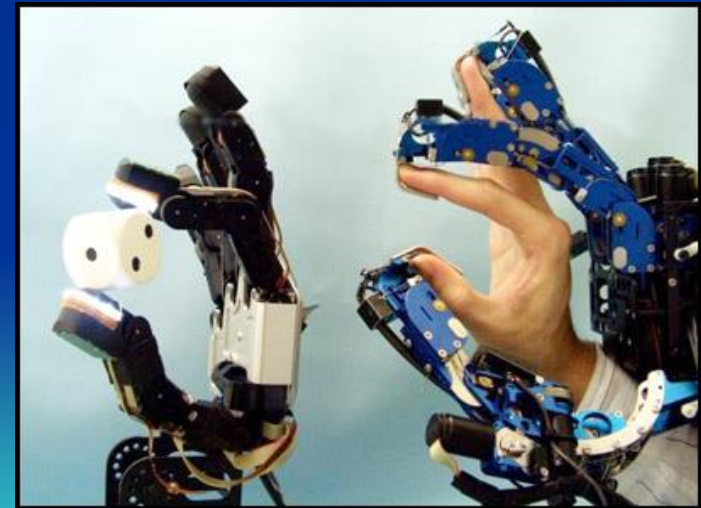
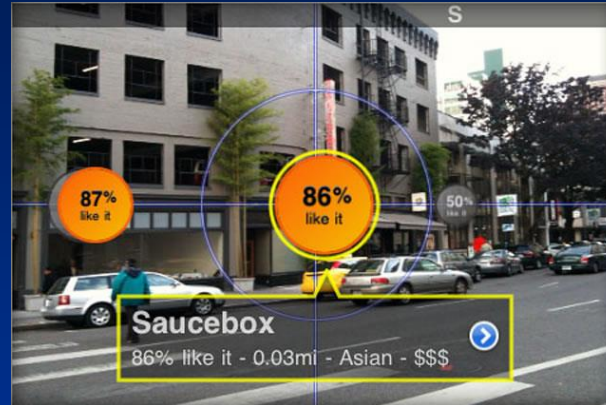
Computing gets smaller (and bigger)

- More specialized
- More embedded
- Grid/cloud computing
- Networks get FAST!



UI's become specialized

- Voice
- Mobile
- Decent AI
 - 10x speedup needed



Software becomes specialized

- Web services dominate
- Reliability over features
- Are games reality?



Microsoft Office Online

Search Microsoft.com for: Go

You've always wanted your business to be online. Microsoft just made it easier.

Microsoft Office Live

Be one of the first to experience Microsoft Office Live for free >>

Microsoft Office Live is coming.

Today, an online presence is almost a requirement for small business success. That's why Microsoft is introducing Microsoft® Office Live—a set of affordable business productivity services designed to help you grow your business more easily by establishing a professional presence online.

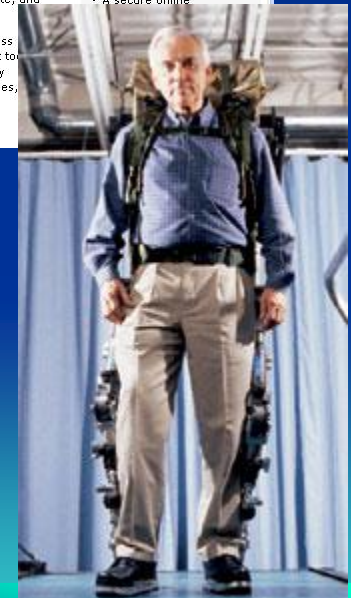
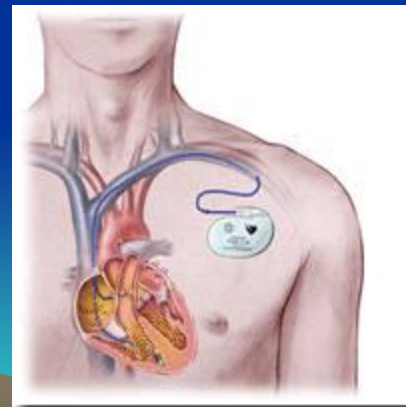
Microsoft Office Live will provide your company with its own domain name, Web site, and e-mail accounts for free.

Additionally, Microsoft Office Live will offer you and your employees expert business management applications, such as customer, project, and document management tools and a security-enhanced private Web site—affordably managed and maintained by Microsoft—where you can work together and share information with your employees, customers, suppliers, and contractors.

A beta version of Microsoft Office Live will launch in early 2006.*

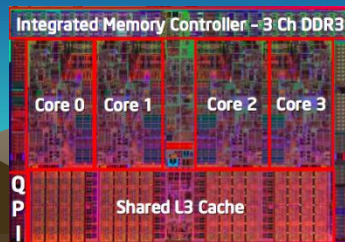
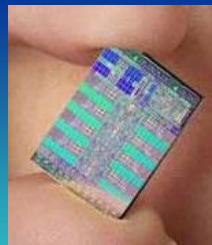
WHY MICROSOFT OFFICE LIVE?

- A professional Web site, expertly hosted by Microsoft
- A secure online



Hardware gets specialized

- Moore's law is dead
 - Killed by power demands
- Future is multi-core, “appropriate” performance
- Divisions blur
- Metcalfe's law still going



NOVEMBER 16, 2009

Supercomputers with 100 million cores coming by 2018

The push is on to build exascale supercomputers that can solve the planet's biggest problems

By Patrick Thibodeau | [Computerworld](#)

Clock speeds will only double, triple in next 15 years

International Fabless 2005

Fundamental changes in design required

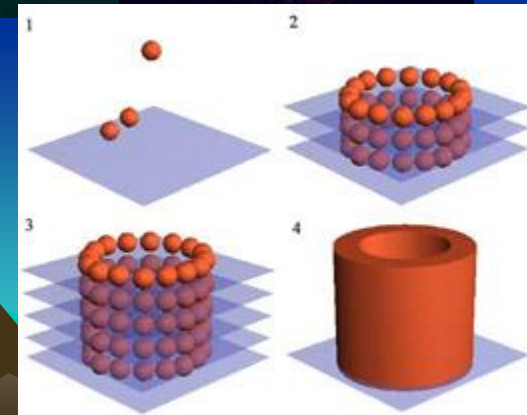
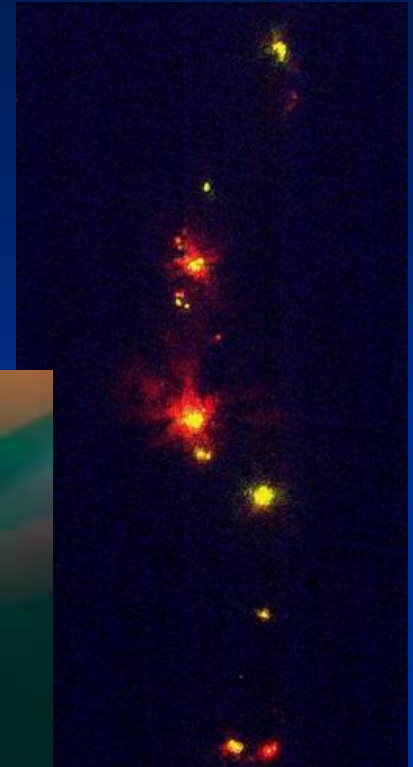
Monday 14 November 2005, 09:25

Twice the power for half the price every 18 months

Year	1979	1984	2005	2017
RAM	16k	128k	256mb	104,032 mb
Hard drive	128k	400k	60gb	12,191 gb
Speed	2mhz	10mhz	1600 mhz	650,199 mhz
Cost	\$5000	\$3900	\$900	\$9

Computing is getting a lot less “normal”

- Quantum computing
- Biological/genetic computing
- Optical computing
- Nanocomputing



The law of unintended consequences

- “Power corrupts, and absolute power corrupts absolutely.”
- Computers have power, and computer software is corrupt, so we’re doomed before we even begin.
- We’ve got to build resilient systems.



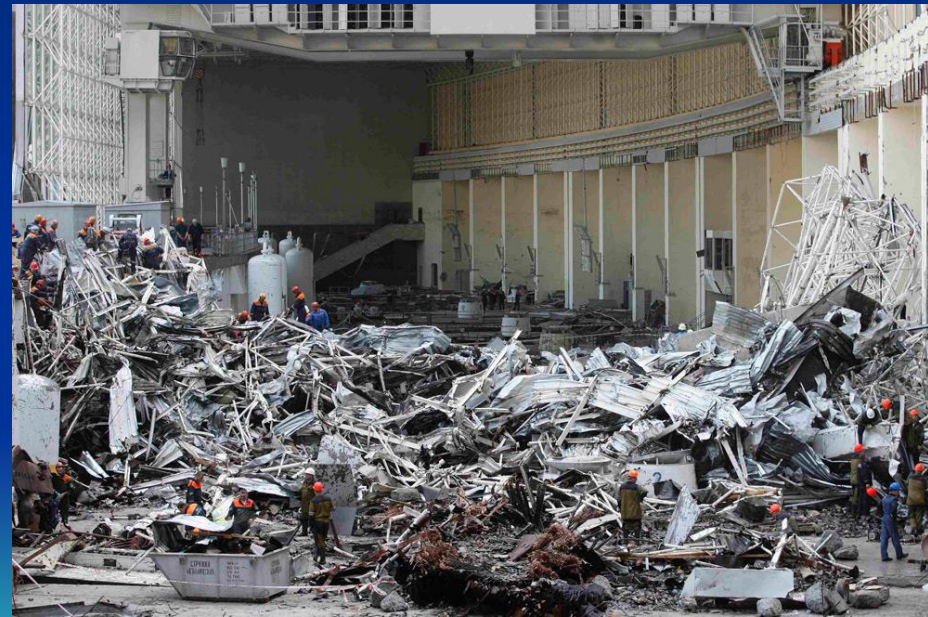
Danger, Will Robinson

Sayano-Shushenskaya power plant



Before

After – 7+ killed, dozens missing



Others: Stuxnet, Sony PSN, ...



Thinks get cloudy – in a good way



Thin clients + Internet + Supercomputing/clusters =
Cool tools + Big Data + Big Science



How much data?

- Google processes 20 PB a day (2008) – stores 10EB ('14)
- Wayback Machine has 3 PB + 100 TB/month (3/2009)
- Facebook stores 180 PB of user data/year (11/'12)
- eBay has 6.5 PB of user data + 50 TB/day (5/2009)
- CERN's LHC will generate 15 PB a year (??)



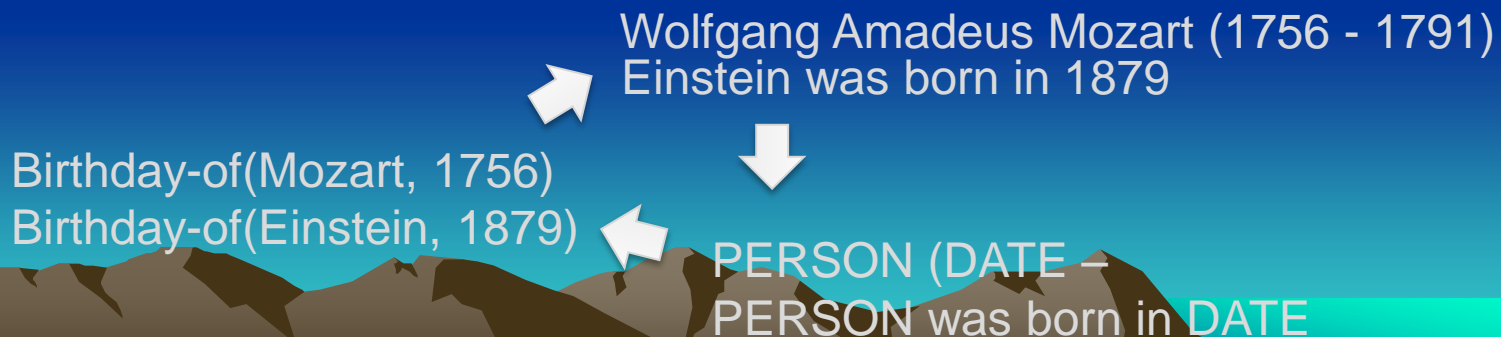
640K ought to be enough for anybody.

What to do with more data?

- Answering factoid questions
 - Pattern matching on the Web
 - Works amazingly well

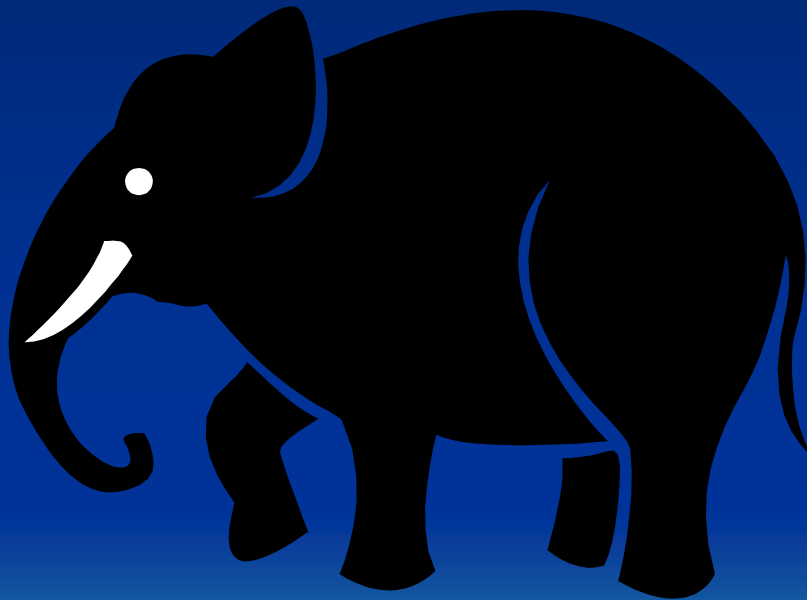
Who shot Abraham Lincoln? → X shot Abraham Lincoln

- Learning relations
 - Start with seed instances
 - Search for patterns on the Web
 - Using patterns to find more instances

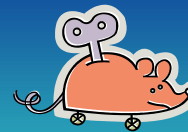
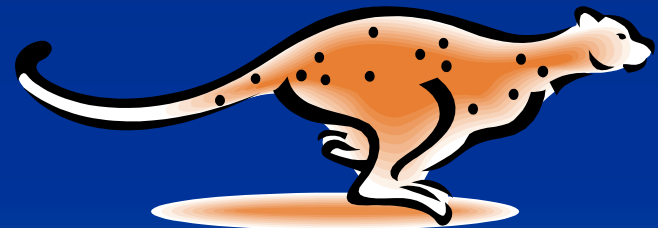


What is Supercomputing About?

Size



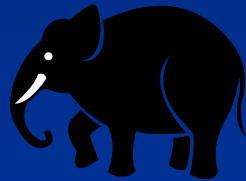
Speed



Laptop

What is Supercomputing About?

- **Size**: Many problems that are interesting to scientists and engineers **can't fit on a PC** – usually because they need more than a few GB of RAM, or more than a few 100 GB of disk.

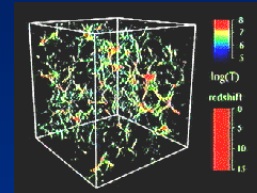


- **Speed**: Many problems that are interesting to scientists and engineers would take a very very long time to run on a PC: months or even years. But a problem that would take **a month on a PC** might take only **a few hours on a supercomputer**.



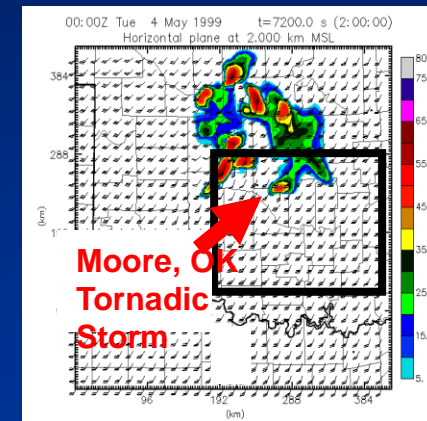
What Is HPC Used For?

- **Simulation** of physical phenomena, such as
 - Weather forecasting
 - Galaxy formation
 - Oil reservoir management



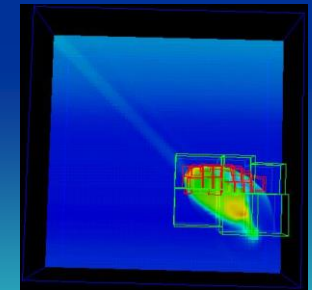
- **Data mining**: finding **needles** of information in a **haystack** of data, such as

- Gene sequencing
- Signal processing
- Detecting storms that might produce tornados

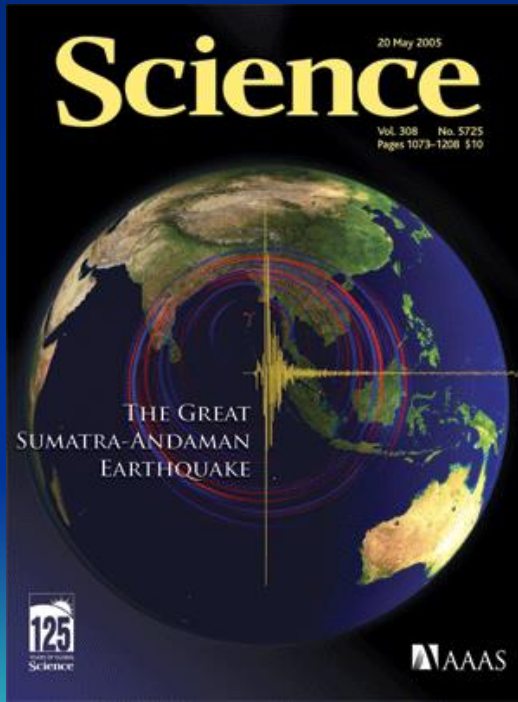
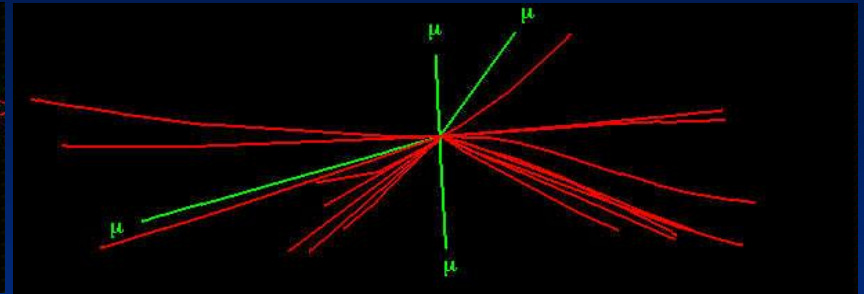
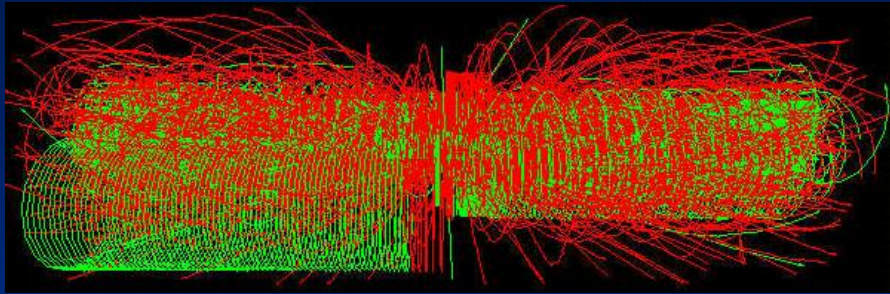


May 3 1999^[2]

- **Visualization**: turning a vast sea of **data**^[3] **pictures** that a scientist can understand



System-Level Science: We Have Much to Communicate!

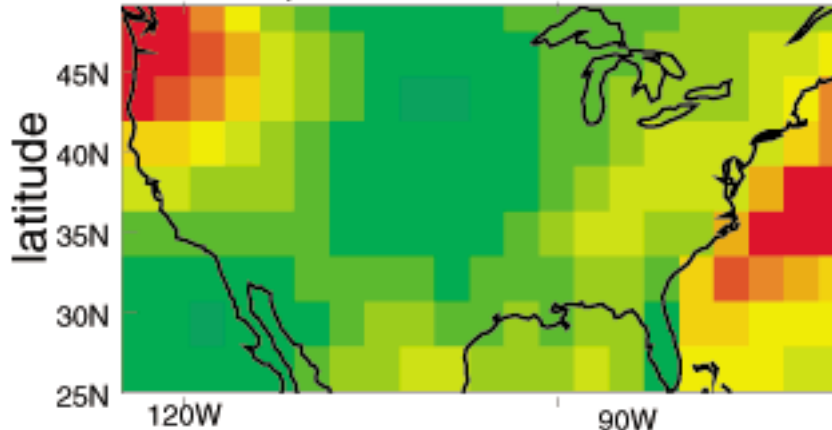


Problems too large &/or complex to tackle alone ...

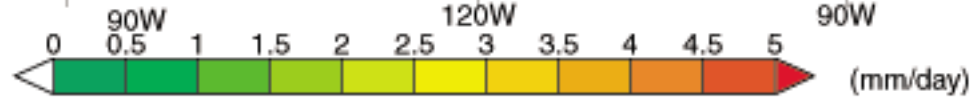
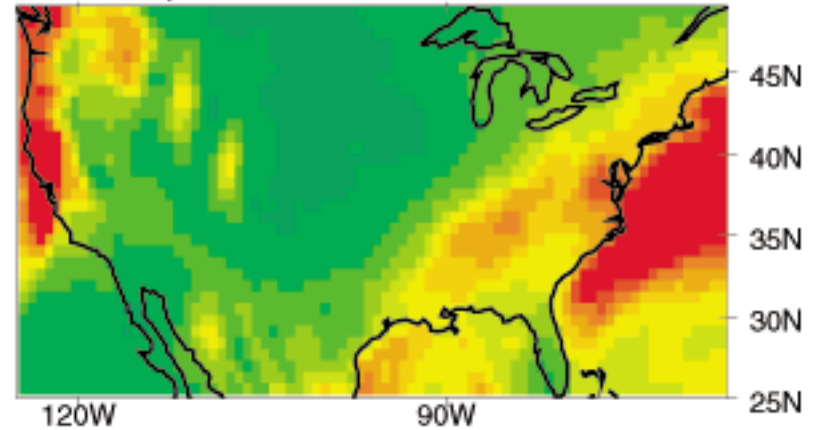
Wintertime Precipitation

As model resolution becomes finer, results converge towards observations

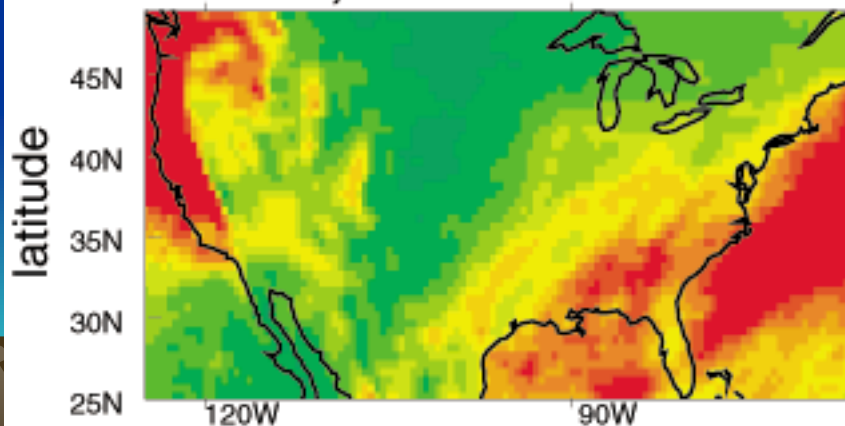
model, 300 km resolution



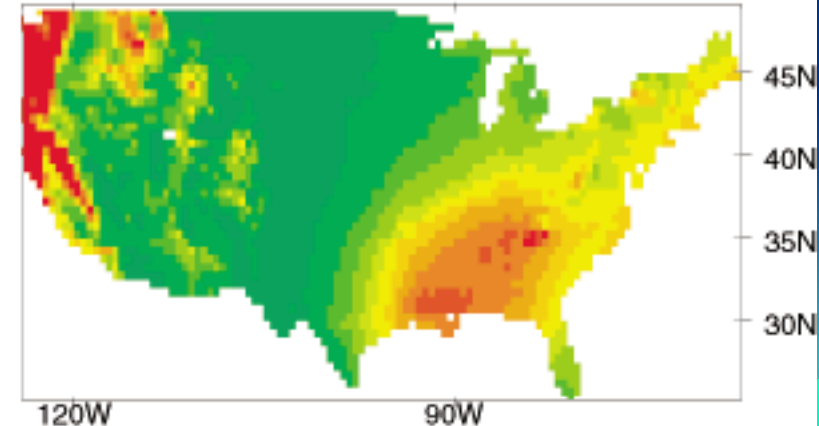
model, 75 km resolution



model, 50 km resolution



observations



Simulation: The Third Pillar of Science

- **Traditional scientific and engineering method:**

- (1) Do **theory** or paper design
- (2) Perform **experiments** or build system

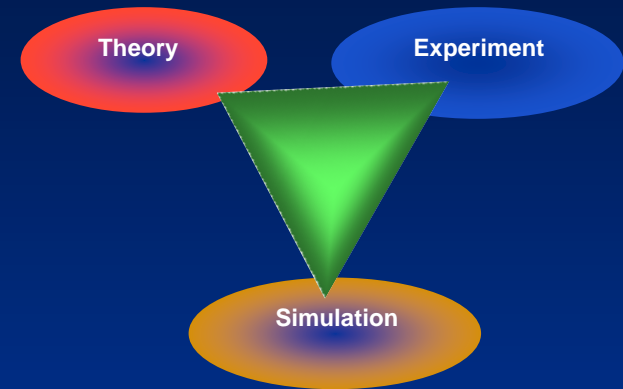
- **Limitations:**

- Too difficult—build large wind tunnels
- Too expensive—build a throw-away passenger jet
- Too slow—wait for climate or galactic evolution
- Too dangerous—weapons, drug design, climate experimentation

- **Computational science and engineering paradigm:**

- (3) Use high performance computer systems to **simulate and analyze** the phenomenon

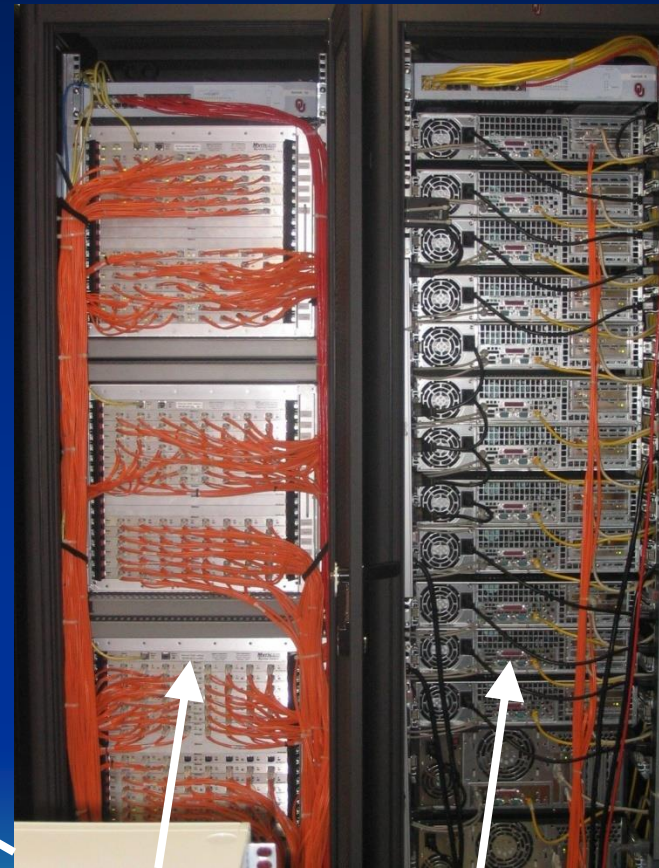
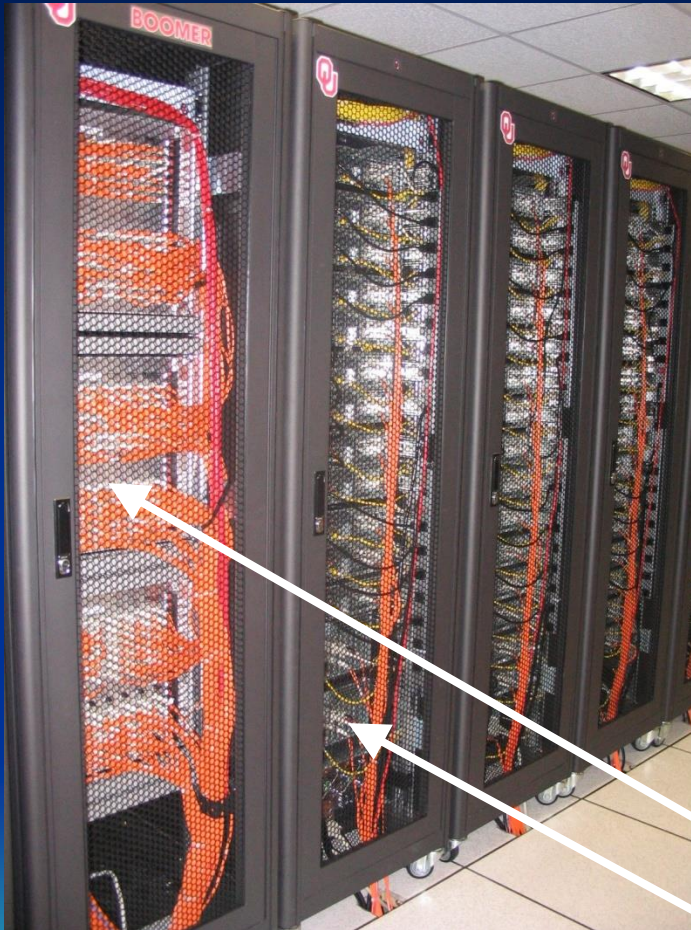
- Based on known physical laws and efficient numerical methods
- Analyze simulation results with computational tools and methods beyond what is used traditionally for experimental data analysis



Economic Impact of HPC

- **Airlines:**
 - System-wide logistics optimization systems on parallel systems.
 - Savings: approx. \$100 million per airline per year.
- **Automotive design:**
 - Major automotive companies use large systems (500+ CPUs) for:
 - CAD-CAM, crash testing, structural integrity and aerodynamics.
 - One company has 500+ CPU parallel system.
 - Savings: approx. \$1 billion per company per year.
- **Semiconductor industry:**
 - Semiconductor firms use large systems (500+ CPUs) for
 - device electronics simulation and logic validation
 - Savings: approx. \$1 billion per company per year.
- **Energy**
 - Computational modeling improved performance of current nuclear power plants, equivalent to building two new power plants.

An Actual Cluster



Interconnect

Nodes

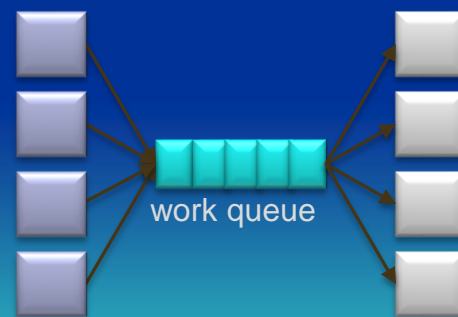
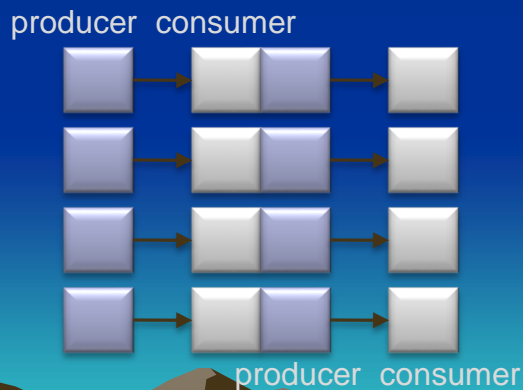
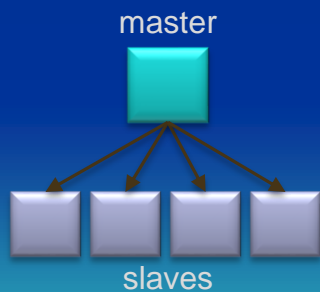
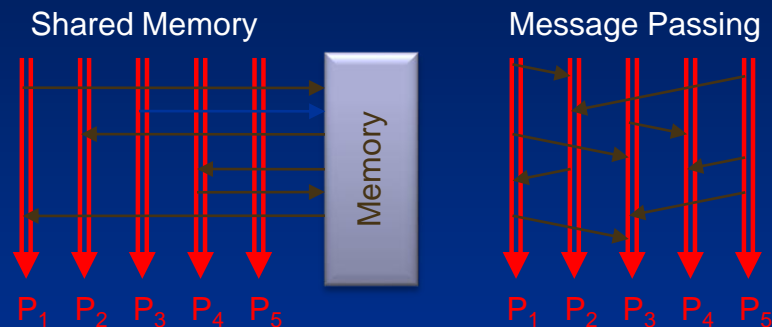
Beocat

Research Computing for
K-State & Kansas

- > Two thousand cores
- 150 compute nodes
- 800 TB storage
- 7.6TB RAM
- 2 administrators
- \$0 fee

Current Tools

- Programming models
 - Shared memory (pthreads)
 - Message passing (MPI)
- Design Patterns
 - Master-slave
 - Producer-consumer flows
 - Shared work queues



Why HPC is Worth the Bother

- What HPC gives you that you won't get elsewhere is the ability to do **bigger, better, more exciting science**. If your code can run faster, that means that you can tackle much bigger problems in the same amount of time that you used to need for smaller problems.
- HPC is important not only for its own sake, but also because what happens in HPC today will be on your desktop in about 10 to 15 years: it puts you **ahead of the curve** – as a worker or a business.



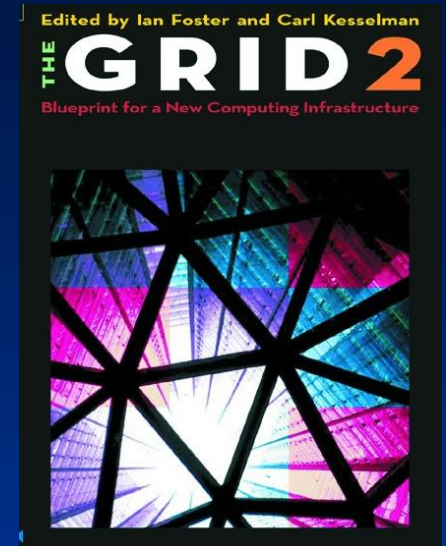
The Future: All Software is Network-Centric

- We don't build or buy “computers” anymore, we borrow or lease required resources
 - ◆ When I walk into a room, need to solve a problem, need to communicate
- A “computer” is a dynamically, often collaboratively constructed collection of processors, data sources, sensors, networks
 - ◆ Similar observations apply for software



Resources

- Supercomputing in Plain English
 - Henry Neeman, OU
 - <http://www.oscer.ou.edu/education.php>
- Globus Alliance – <http://www.globus.org>
- Open Science Grid – <http://www.opensciencegrid.org>
- Background information –
<http://www.mcs.anl.gov/~foster>



Thanks to James Demmel (Berkeley) & John Gilbert (UCSB) for the generous use of their class materials.
Thanks to Jimmy Lin & his on “Introduction to MapReduce” lecture.
Funded in part by the National Science Foundation grants 1006860 and 0919443.



Books

