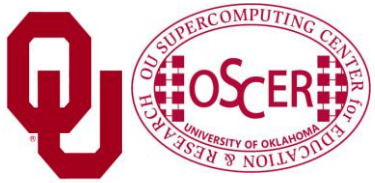
Institutional and Statewide Cyberinfrastructure: Dollars and Sense



North Dakota EPSCoR State Cyberinfrastructure Strategic Planning Workshop

Henry Neeman, Director
OU Supercomputing Center for Education & Research (OSCER)
University of Oklahoma
Thursday March 3 2011

PEOPLE


EXIT
Fall Semester 2003
Integration of HPC in
Nanotechnology
National Science Foundation

THINGS

What is Cyberinfrastructure?

What is a Ship?

"... [W]hat a ship is ... It's not just a keel and hull and a deck and sails. That's what a ship needs. But what a ship is ... is freedom."

– "Pirates of the Caribbean"

What Cyberinfrastructure NEEDS

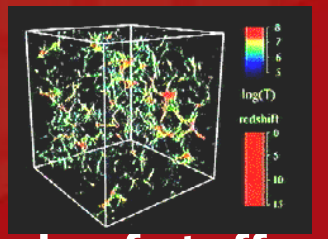
- **High Performance Computing** (supercomputing)
  - Simulation
  - Optimization
  - Data Mining
- **High Performance Networking**: moving data fast
- **High Throughput Computing**: harnessing idle PCs to do number crunching
- **Grid/Cloud/Utility Computing**: linking geographically dispersed systems to tackle larger problems
- **Scientific Visualization**: turning a vast sea of data into pictures and movies that a person can understand
- **Shared Resources**: sensor networks, instruments, data collections

What Cyberinfrastructure IS

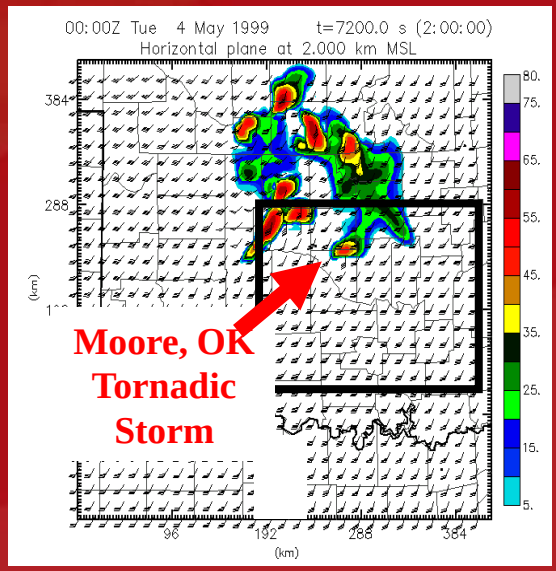
Information-intensive techniques and technologies that enable our best thinkers, inventors and implementers – academic, government, industry and non-profit – to discover and design new generations of knowledge, products, capabilities and opportunities in science, technology, engineering and mathematics.

What are HPC & HTC Used For?

- ***Simulation*** of physical phenomena – things that are too big, too small, too slow, too fast, too expensive or too dangerous to study in real life
  - Weather forecasting
  - Protein folding
  - Energy management

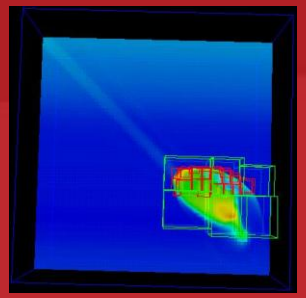


Courtesy Greg Bryan & Mike Norman



Courtesy Kelvin Droegemeier

- ***Optimization:*** picking the best mix of stuff
- ***Data mining:*** finding needles of information in a haystack of data
  - Gene sequencing
  - Detecting storms that might produce tornados



Courtesy Mordecai-Mark Mac Low

- ***Visualization:*** turning a vast sea of data into pictures that a scientist can understand

Cyberinfrastructure in Oklahoma

OU: Dell Intel Xeon Linux Cluster

1,076 Intel Xeon CPU chips/4304 cores

- 529 dual socket/quad core Harpertown 2.0 GHz, 16 GB each
- 3 dual socket/quad core Harpertown 2.66 GHz, 16 GB each
- 3 dual socket/quad core Clovertown 2.33 GHz, 16 GB each
- 2 x quad socket/quad core Tigerton, 2.4 GHz, 128 GB each

8,800 GB RAM

~100 TB globally accessible disk

QLogic Infiniband

Force10 Networks Gigabit Ethernet

Red Hat Enterprise Linux 5

Peak speed: 34.6 TFLOPs*

*TFLOPs: trillion calculations per second

sooner.oscer.ou.edu

OU: Dell Intel Xeon Linux Cluster

DEBUTED NOVEMBER 2008 AT:

- #90 worldwide
- #47 in the US
- #14 among US academic
- #10 among US academic excluding TeraGrid

sooner.oscer.ou.edu

OU: Dell Intel Xeon Linux Cluster

Purchased mid-July 2008

First friendly user Aug 15 2008

Full production Oct 3 2008

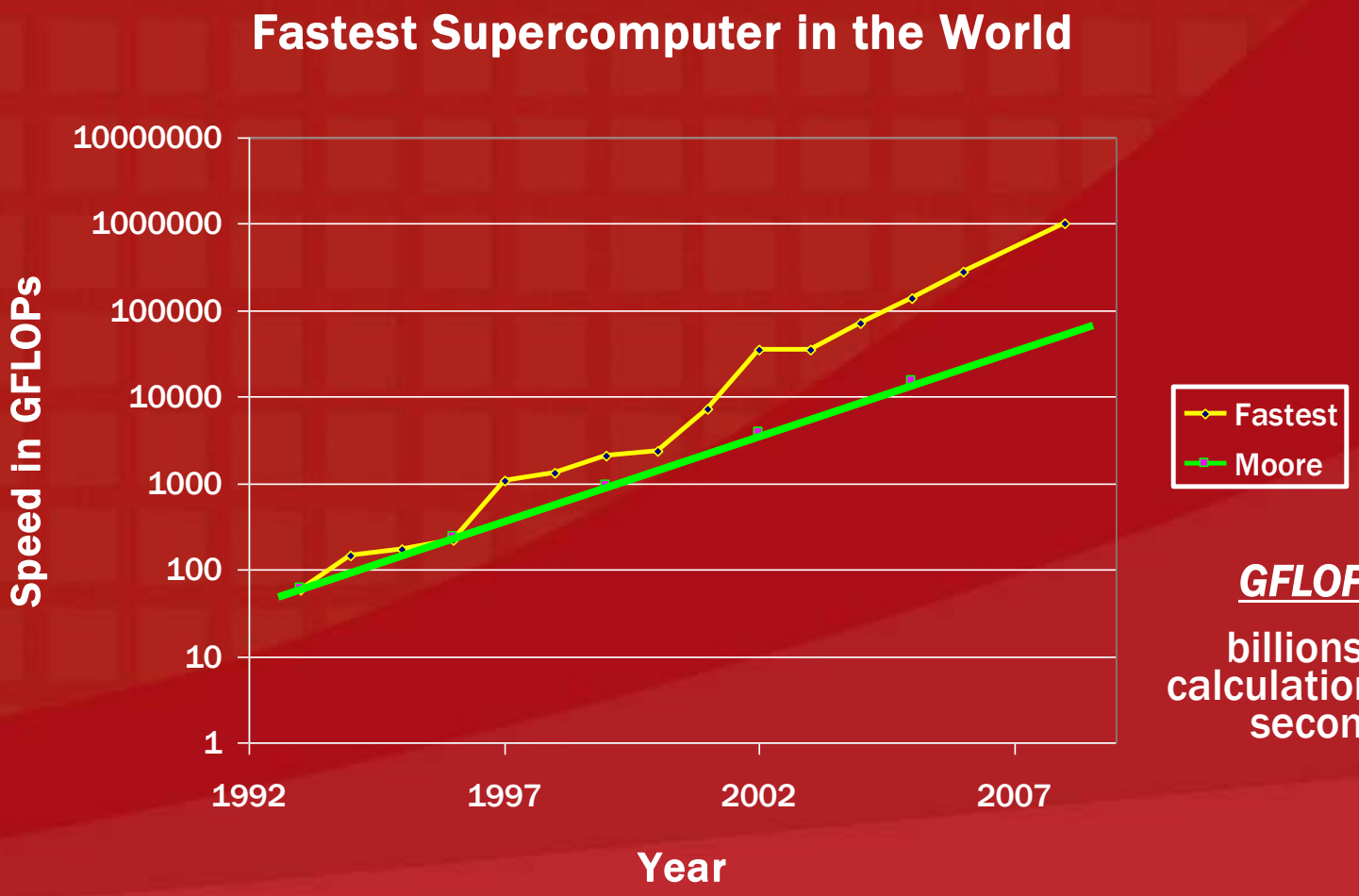
Christmas Day 2008: >~75% of nodes and ~66% of cores were in use.

sooner.oscer.ou.edu

OSU: Several Medium-sized Supercomputers

- **Pistol Pete**: Linux cluster, 64 compute nodes, 512 CPU cores, 1024 GB RAM, 25 TB disk, Infiniband, 5.4 TFLOPs, being deployed now.
- **Cimarron**: Linux cluster, 14 compute nodes, 112 CPU cores, 224 GB RAM, 3.5 TB disk, GigE, 896 GFLOPs, deployed summer 2008.
- **Spur**: Shared memory machine with 4 quad core 2.4 GHz, 128 GB RAM, 1.5 TB disk, deployed spring 2008.
- **Bullet**: Linux cluster, 64 compute nodes, 128 CPU cores, 256 GB RAM, 5 TB disk, Infiniband, 820 GFLOPs, deployed fall 2006.

Fastest Supercomputer vs. Moore



GFLOPs:

billions of calculations per second

OU: Condor Pool

Condor is a software technology that allows idle desktop PCs to be used for number crunching.

OU IT has deployed a large Condor pool (795 desktop PCs in IT student labs all over campus).

It provides a huge amount of additional computing power – more than was available in all of OSCER in 2005.

20+ TFLOPs peak compute speed.

And, the cost is very very low – almost literally free.

Also, we've been seeing empirically that Condor gets about 80% of each PC's time.

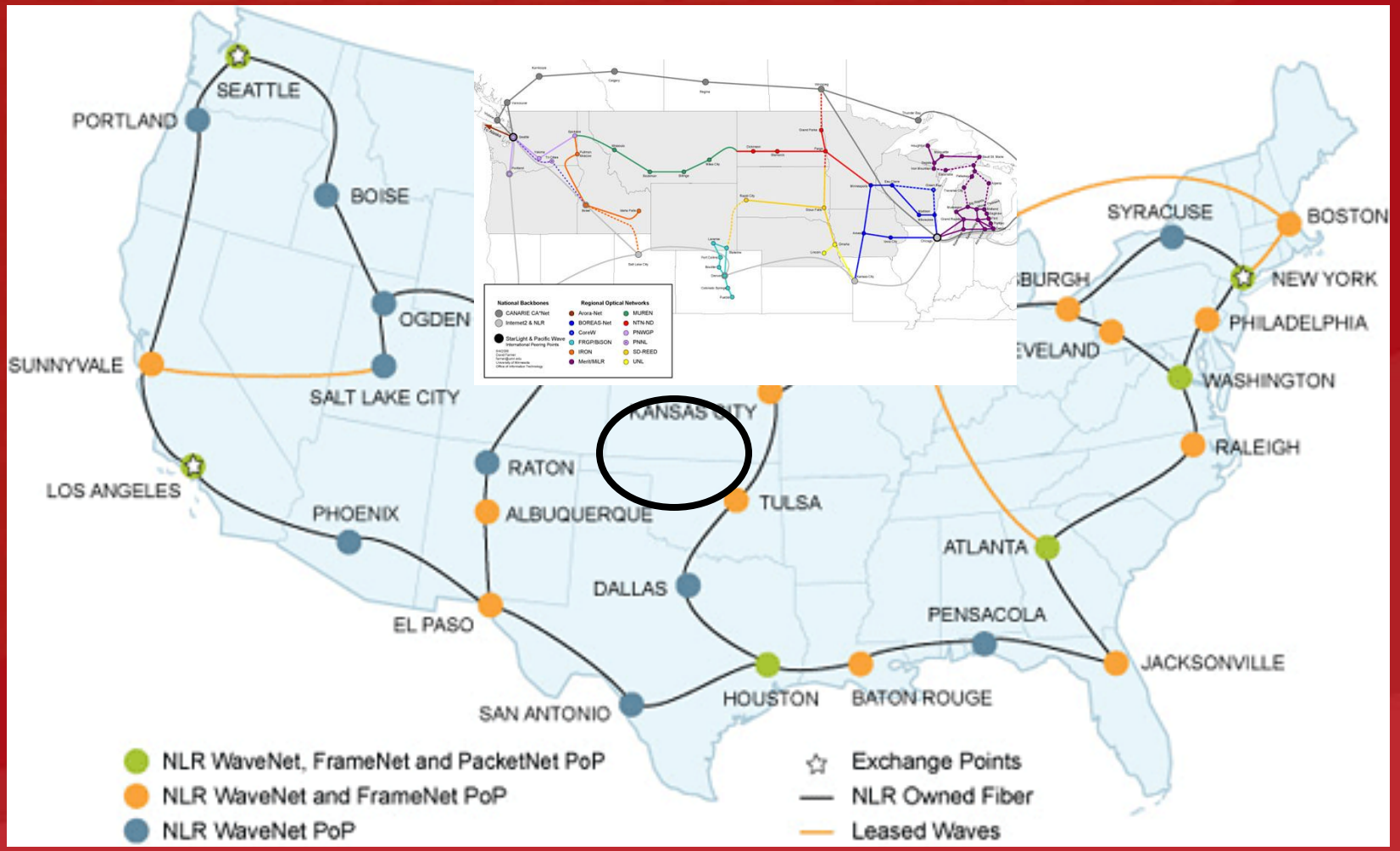
OU and OSU have worked together to get a comparable Condor pool set up at OSU.

OSU: Condor Pool

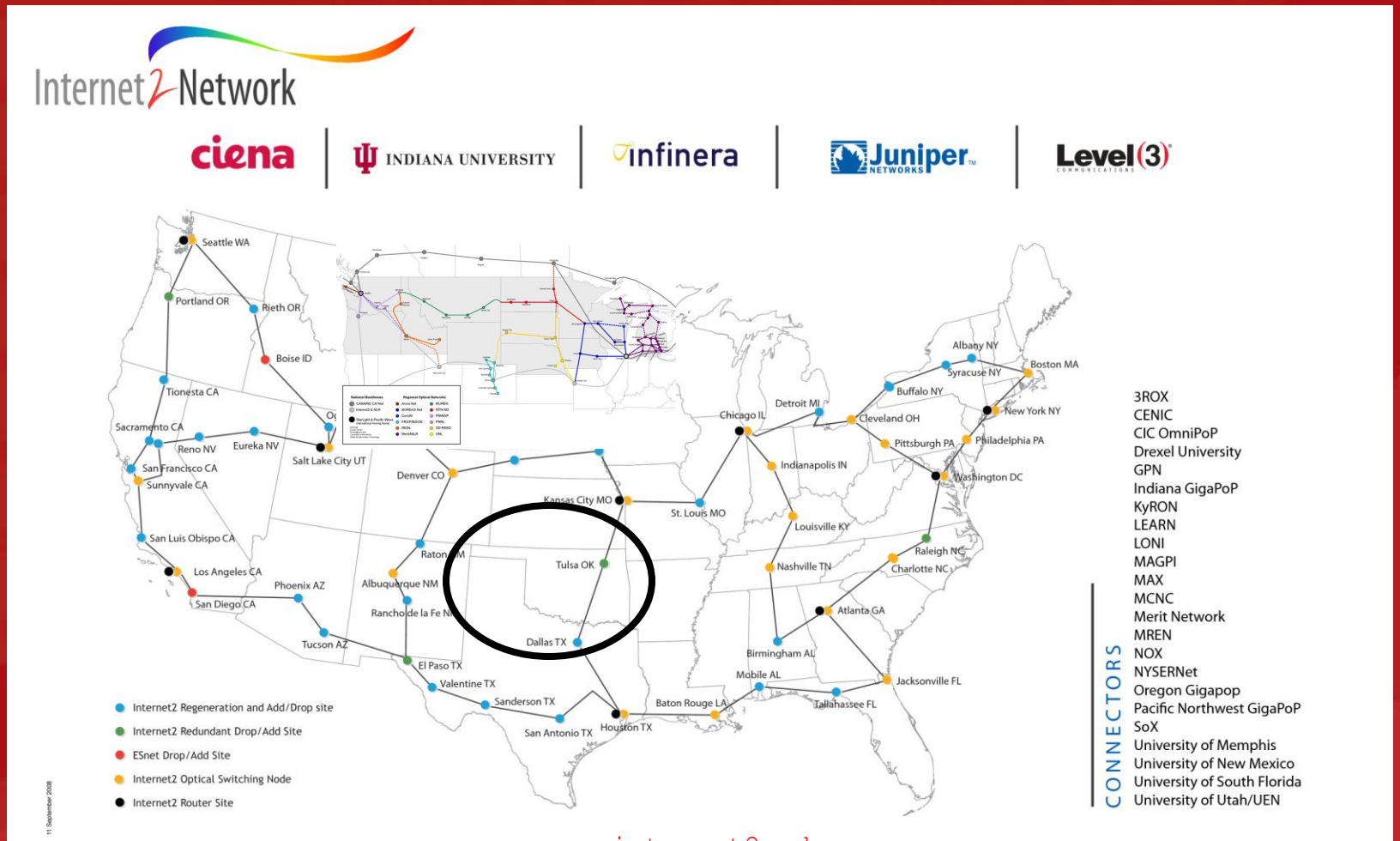
In collaboration with OU IT, OSU IT is deploying a large Condor pool (approx 300 desktop PCs in IT student labs all over campus).

Once deployed, OU's and OSU's Condor pools will "flock" together, allowing jobs to migrate invisibly between institutions, but giving each institution's jobs priority on their own PCs.

Networking: National LambdaRail



Networking: Internet2



www.internet2.edu

Why High Performance Networking?

- Ability to move large datasets rapidly
  - Natural hazards prediction & mitigation –OK: tornadoes
  - Large scale data-intensive projects – OK: high energy physics
  - 10 Gigabits per second = 100,000 Gigabytes per day
- Telepresence (including, potentially, telemedicine)
- A Seat At The Table
  - Without high performance networking, we wouldn't be in the running for major national initiatives.

Oklahoma Cyberinfrastructure Initiative

OK Cyberinfrastructure Initiative (OCII)

- All academic institutions in Oklahoma are eligible to sign up for free use of OU's and OSU's centrally-owned CI resources.
- Other kinds of institutions (government, NGO, commercial) are eligible to use, though not necessarily for free.
- Everyone can participate in our CI education initiative (not just in OK).
  - UND, Meridian Environmental Technology Inc
- Oklahoma Supercomputing Symposium open to all (10th Symposium Oct 11-12 2011).

OCII History

- NSF EPSCoR Research Infrastructure Improvement (Track 1)
  - Call for Proposals: “Cyberinfrastructure” appears 18 times.
  - OK EPSCoR asked for statewide CI plan
  - Meeting: OU & OSU CI leads, OU & OSU CIOs, OneNet director, OK EPSCoR director & associate director, science theme leaders
  - Sketched out OCII plan
- OU helped OSU set up Condor pool.
- Providing “Supercomputing in Plain English” overview talk statewide (have visited 11 of 13 public universities, 7 of 12 private universities).

OCII Accomplishments

- Other NSF EPSCoR Grants
  - Track 2: Ecoinformatics (ecology) with Kansas: $6M ($3M to OK)
  - C2: Networking: $1.17M to OK – includes OK's only HBCU, Tribal Colleges
- NSF MRI Grant: Oklahoma PetaStore ($793K to OU)
- OCII provides supercomputer accounts to many external users (outside of OU and OSU): 75+ users at 24 academic, government, industry and nonprofit organizations (and counting) – up from 34 users at 15 institutions since Dec 2008.
- OCII is directly responsible for $20M in NSF grants to OK since our first meeting in 2006 (Track1, Track2, C2, MRI).

OCII Outcomes

External Funding Outcomes (OU)

- External research funding facilitated by OSCER (Fall 2001- Fall 2009): **$186M+ total, $100M+ to OU**
- Funded projects: **162+**
- **102** OU faculty and staff in **19** academic departments and 2 other campus organizations (research centers etc)
- Comparison: Fiscal Year 2002-10 (July 2001 – June 2010): OU Norman externally funded research expenditure: $611M

Since being founded in fall of 2001, OSCER has enabled research projects comprising more than **1 / 7 of OU Norman's total externally funded research expenditure**, with more than a **7-to-1 return on investment**.

Publications Facilitated by OSCER

- 600+ publications facilitated by OSCER
  - 2010: 128+ papers
  - 2009: 105 papers
  - 2008: 100
  - 2007: 72
  - 2006: 85
  - 2005: 63
  - 2004: 27
  - 2003: 9
  - 2002: 8
  - 2001: 3

These papers would have been impossible, or much more difficult, or would have taken much longer, without OSCER's direct, hands-on help.

Education Outcomes

- "Supercomputing in Plain English"
  - 1000+ participants at 167 academic, government, industry and nonprofit groups, including:
    - OU, OSU + 12 other OK academic institutions
    - 5 OK govt agencies (state, federal, military)
    - 4 OK companies
    - 1 OK non-governmental agency
    - 36 other academic institutions in 18 other EPSCoR jurisdictions (including ND)
    - Research universities, masters-granting, bachelors-granting, community colleges, high school, middle school; minority-serving
- "Supercomputing in Plain English" overview talk
  - 20 OK academic institutions, several other OK organizations

Lessons Learned

- Cooperation between institutions is essential:
  - among PhD-granting institutions;
  - between PhD-granting and masters- and bachelors-granting;
  - between Minority Serving Institutions and others;
  - between academic and non-academic.
- A CI win for any institution in the state is a win for the whole state.
- Providing users across the state with access to centrally-owned resources at PhD-granting institutions costs little but returns a lot.
- Help the institutions that you think are your rivals; it turns out that they're your best advocates.
- CI has a huge Return on Investment.

Cyberinfrastructure and Students

Moore's Law

In 1965, Gordon Moore was an engineer at Fairchild Semiconductor.

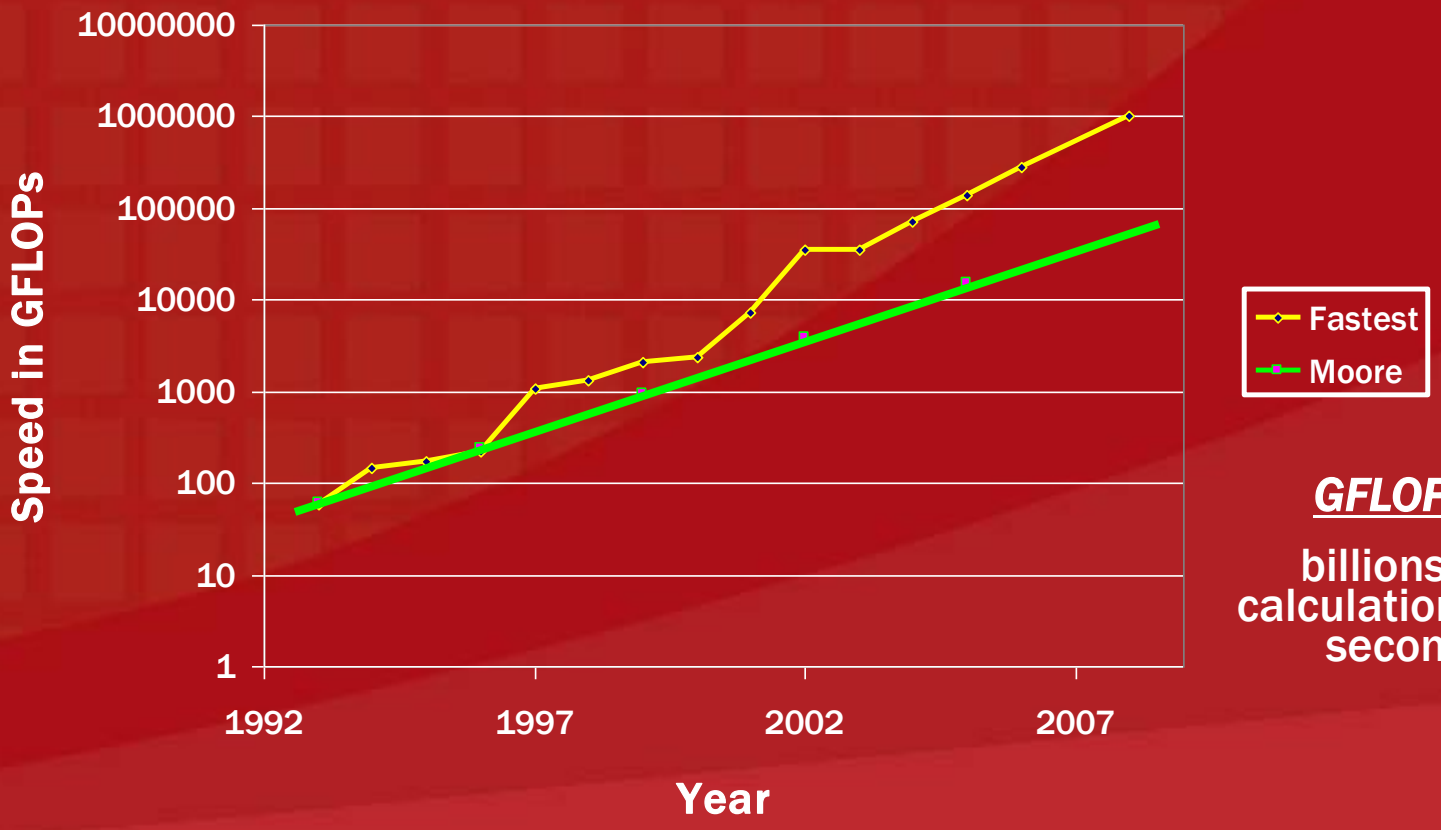
He noticed that the number of transistors that could be squeezed onto a chip was doubling about every 18 months.

It turns out that computer speed is roughly proportional to the number of transistors per unit area.

Moore wrote a paper about this concept, which became known as ***"Moore's Law."***

Moore's Law Observed

Fastest Supercomputer in the World



GFLOPs:

billions of calculations per second

CI and Young North Dakotans

- Historically we've seen that whatever happens at the high end of computing today will be on your desktop in 10 to 15 years.
- So, for North Dakota undergrads to learn about high end Cyberinfrastructure puts them ahead of the curve:
  - they'll be more competitive now;
  - they'll know something important about the future.

The Future is Now

- Computing power doubles every 18 months.
- Undergrads are about 20 years old.
- Life expectancy in the US is about 80 years.
- So they have about 60 years left.
- How much more powerful will computers be in 60 years?
  - 15 years = 10 doublings = 1024x ~ approx 1000 times better
  - 30 years = 20 doublings ~ approx a million times better
  - 45 years = 30 doublings ~ approx a billion times better
  - 60 years = 40 doublings ~ approx a trillion times better
- The problem with the future: “The future always comes too fast, and in the wrong order.” (Alvin Toffler)

Thanks for your attention!

Questions?

www.oscer.ou.edu

hneeman@ou.edu