eScience: Computational Science for the 21st Century

Ed Lazowska

Bill & Melinda Gates Chair in
Computer Science & Engineering
University of Washington

President, PNWGP

8th Annual GLIF Workshop
October 2008

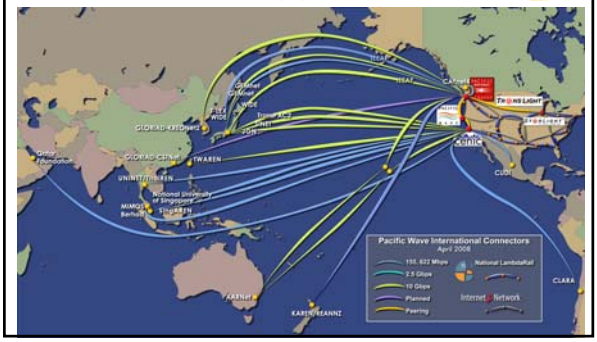
http://lazowska.cs.washington.edu/GLIF.pdf

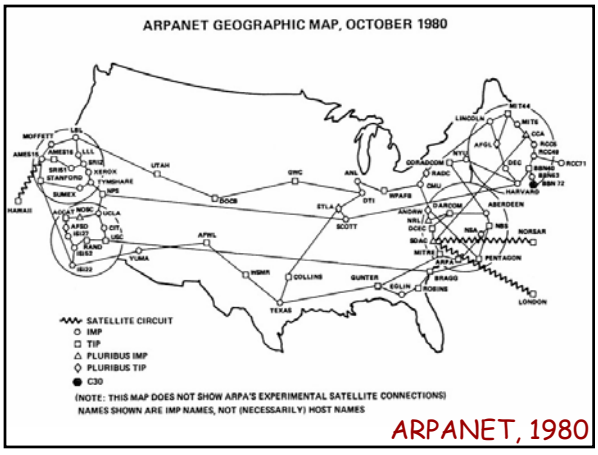
This morning

- Why your work matters
- eScience
- The Computing Community Consortium, and Grand Challenges for our field

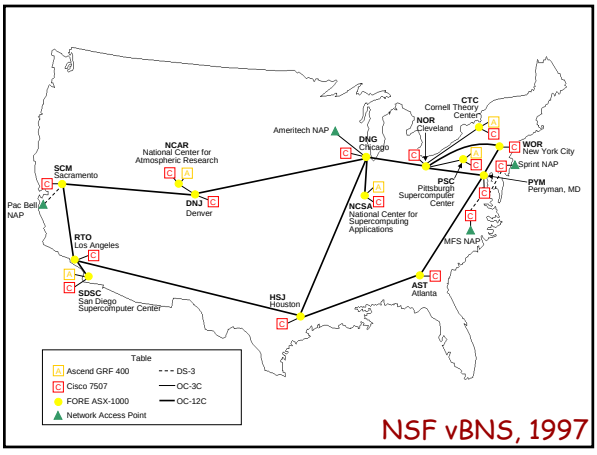
A connected region – then

A connected region – now

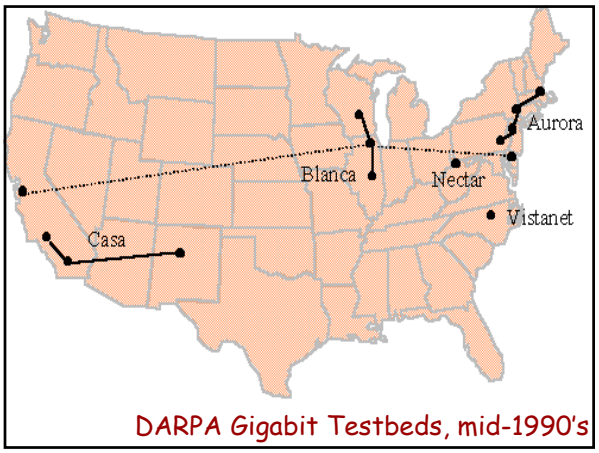




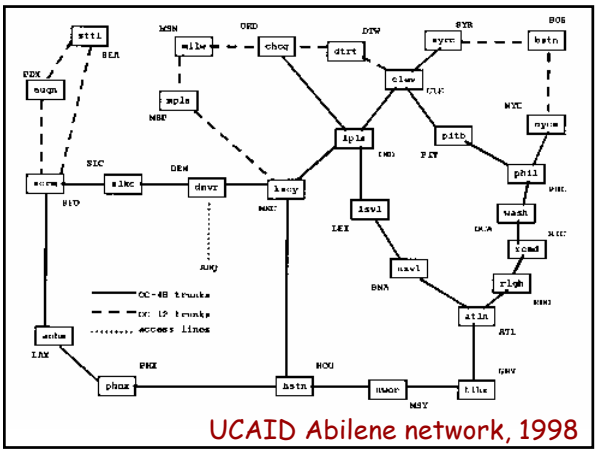
ARPANET, 1980



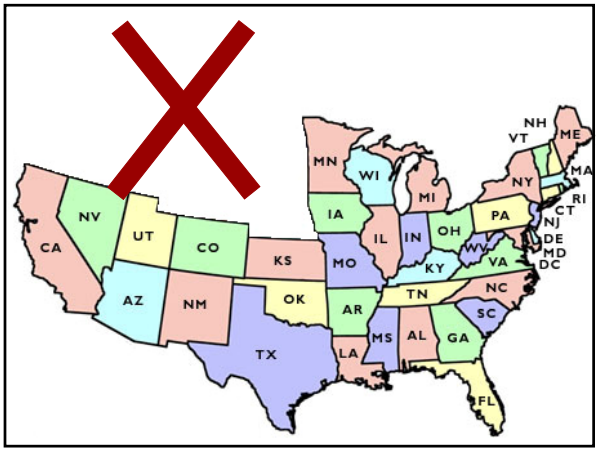
NSF vBNS, 1997

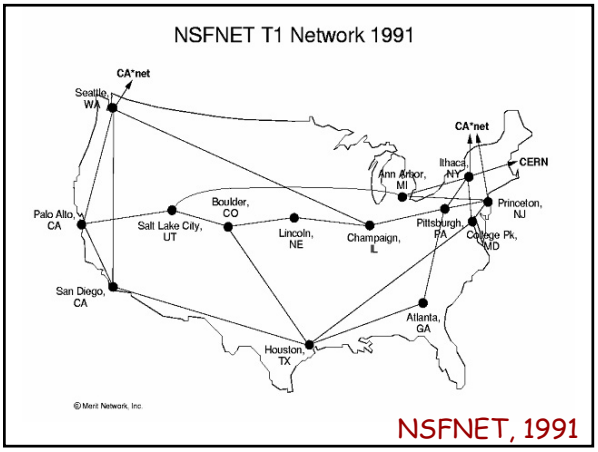


DARPA Gigabit Testbeds, mid-1990's

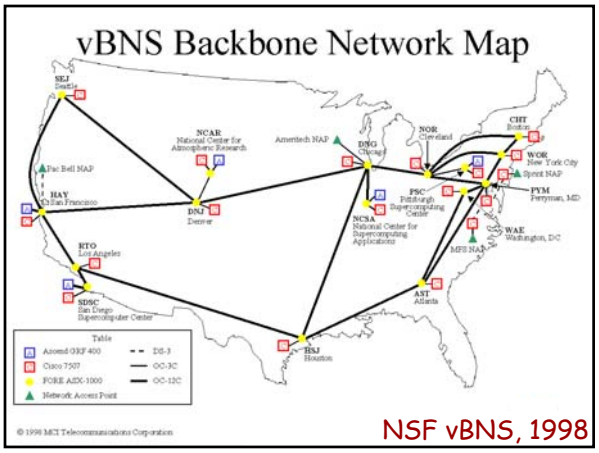


UCAID Abilene network, 1998

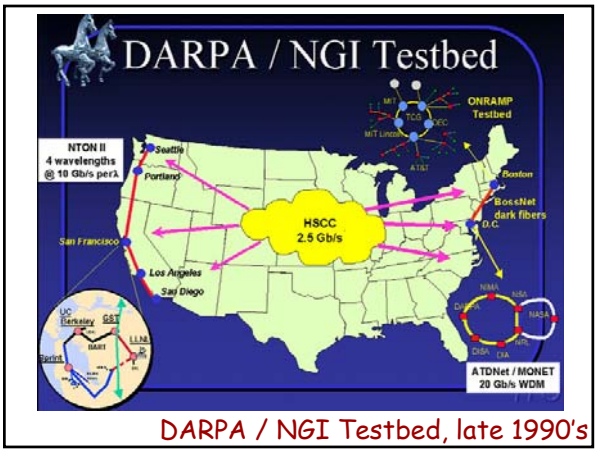




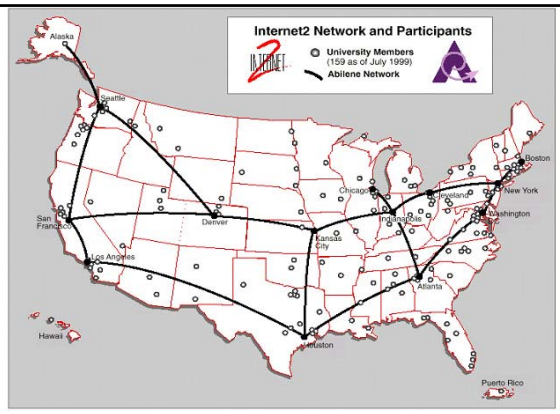
NSFNET, 1991



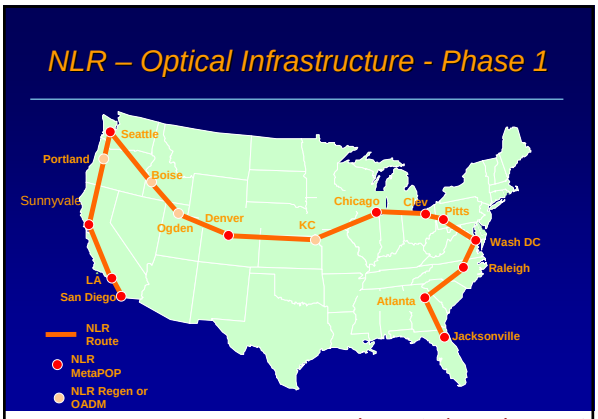
NSF vBNS, 1998



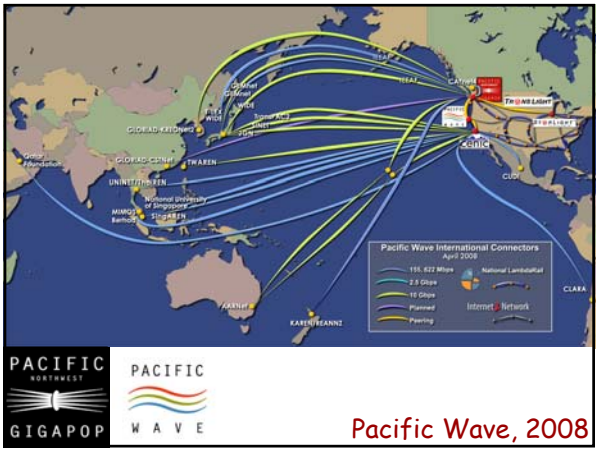
DARPA / NGI Testbed, late 1990's



Internet2, 1999

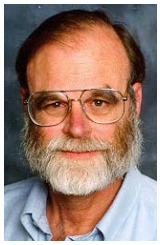


National LambdaRail, 2004

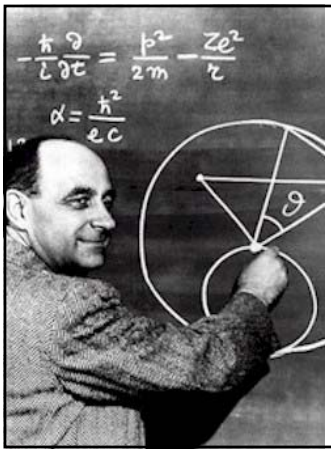


Pacific Wave, 2008

eScience: Sensor-driven (data-driven) science and engineering



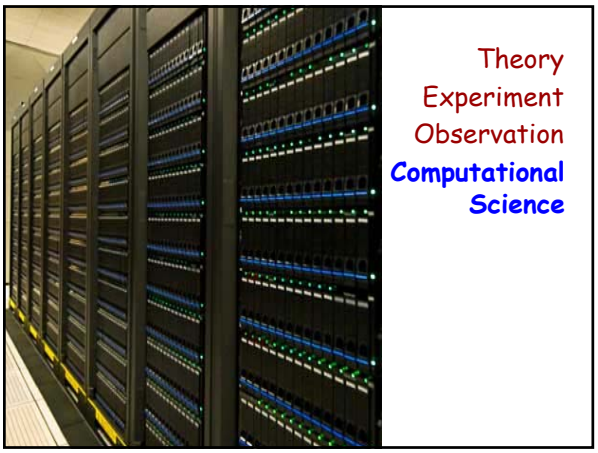
Transforming science (again!)

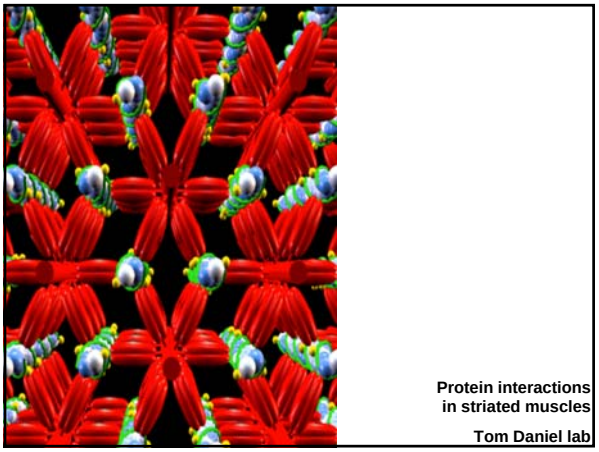


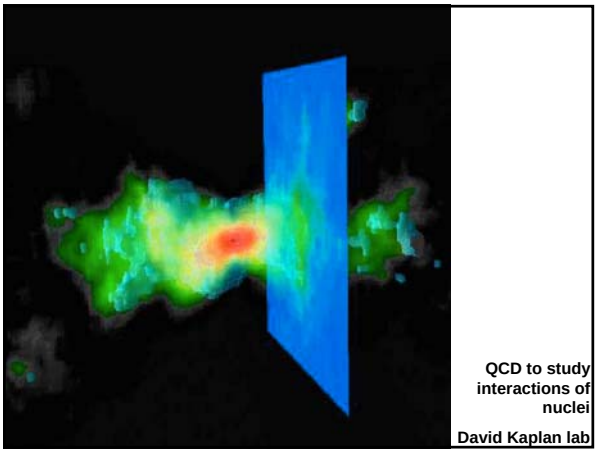
Theory
Experiment
Observation

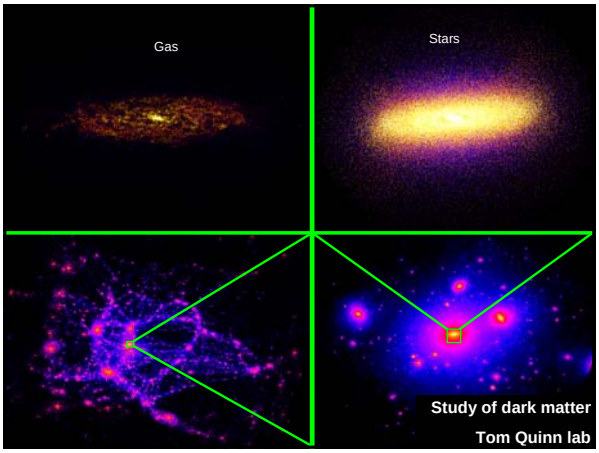
Theory
Experiment
Observation

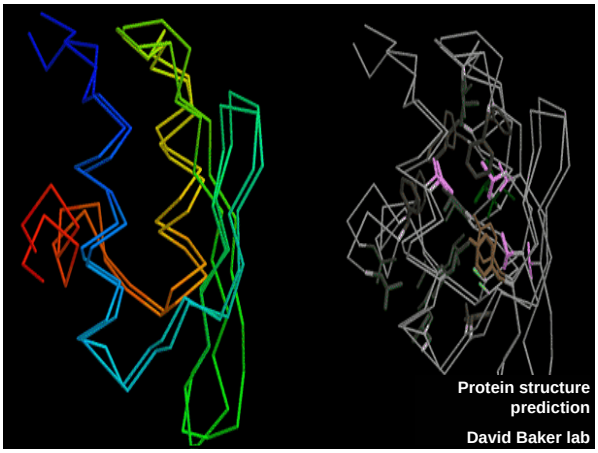

Theory
Experiment
Observation


Theory
Experiment
Observation
Computational
Science


Protein interactions
in striated muscles
Tom Daniel lab


QCD to study
interactions of
nuclei
David Kaplan lab

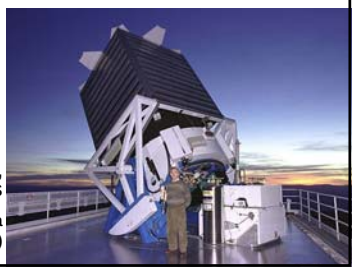

Gas
Stars
Study of dark matter
Tom Quinn lab


Protein structure
prediction
David Baker lab

Theory
Experiment
Observation
Computational Science
eScience

eScience is about the data

- Massive volumes of data from sensors and networks of sensors

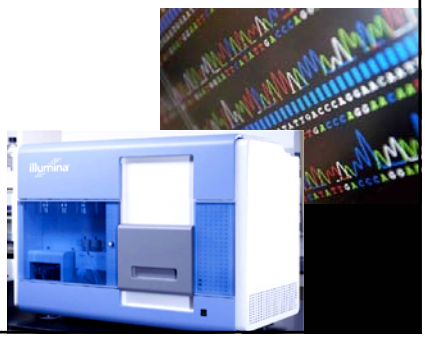


Apache Point telescope, SDSS

15TB of data (15,000,000,000,000 bytes)

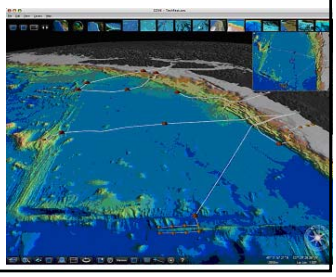
Large Hadron Collider

700MB of data per second, 60TB/day, 20PB/year



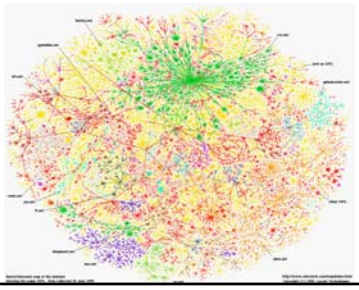
Illumina Genome Analyzer

~1TB/day



Regional Scale Nodes of the NSF Ocean Observatories Initiative

2000 km of fiber optic cable on the seafloor, connecting thousands of chemical, physical, and biological sensors



The Web

20+ billion web pages x 20KB = 400+TB

One computer can read 30-35 MB/sec from disk => 4 months just to read the web

Point-of-sale terminals

eScience is about the *analysis* of data

- The automated or semi-automated extraction of knowledge from massive volumes of data
  - There's simply too much of it to look at!

The technologies of eScience

- Sensors and sensor networks
- Databases
- Data mining
- Machine learning
- Data visualization

eScience will be pervasive

- Computational science was a niche
  - As an institution (e.g., a university), you didn't need to excel in order to be competitive
- eScience capabilities must be broadly available
  - If not, you'll simply cease to be competitive

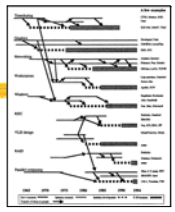
The Computing Community Consortium

The Computing Community Consortium supports the computing research community in creating compelling research visions and the mechanisms to realize these visions.

Computing has changed the world

- Advances in computing change the way we live, work, learn, and communicate
- Advances in computing drive advances in nearly all other fields
- Advances in computing power our economy
  - Not just through the growth of the IT industry – through productivity growth across the entire economy

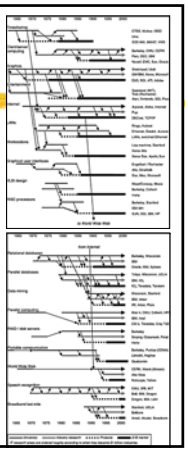
Research has built the foundation



- Timesharing
- Computer graphics
- Networking (LANs and the Internet)
- Personal workstation computing
- Windows and the graphical user interface
- RISC architectures
- Modern integrated circuit design
- RAID storage
- Parallel computing

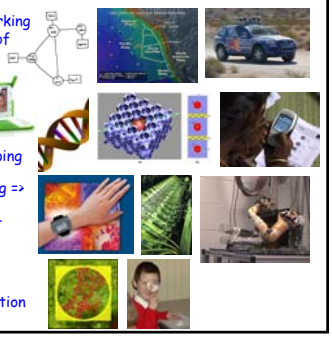
Much of the impact is recent

- Entertainment technology
- Data mining
- Portable communication
- The World Wide Web
- Speech recognition
- Broadband last mile



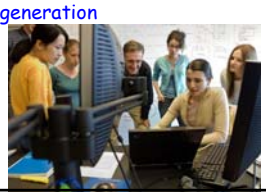
The future is full of opportunity

- Creating the future of networking
- Driving advances in all fields of science and engineering
- Wreckless driving
- Personalized education
- Predictive, preventive, personalized medicine
- Quantum computing
- Empowerment for the developing world
- Personalized health monitoring => quality of life
- Harnessing parallelism: many-core and DISC
- Neurobotics
- Synthetic biology
- The algorithmic lens: Cyber-enabled Discovery and Innovation



We must work together to establish, articulate, and pursue visions for the field

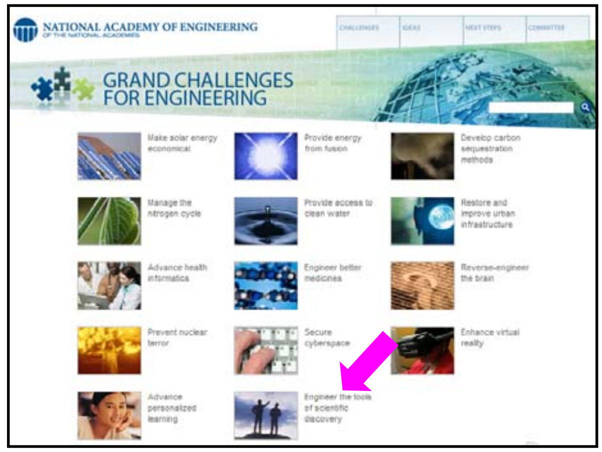
- The challenges that will shape the intellectual future of the field
- The challenges that will catalyze research investment and public support
- The challenges that will attract the best and brightest minds of a new generation

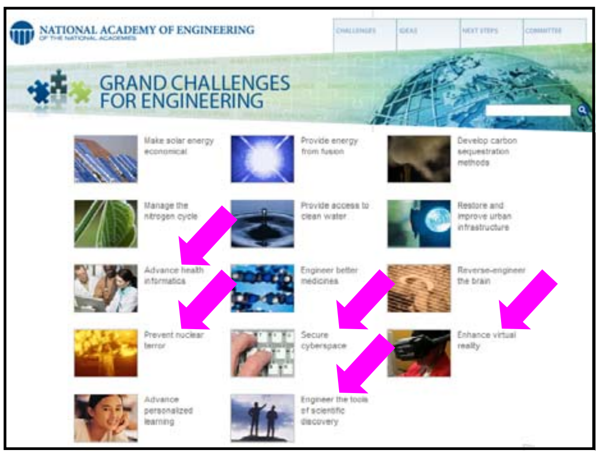


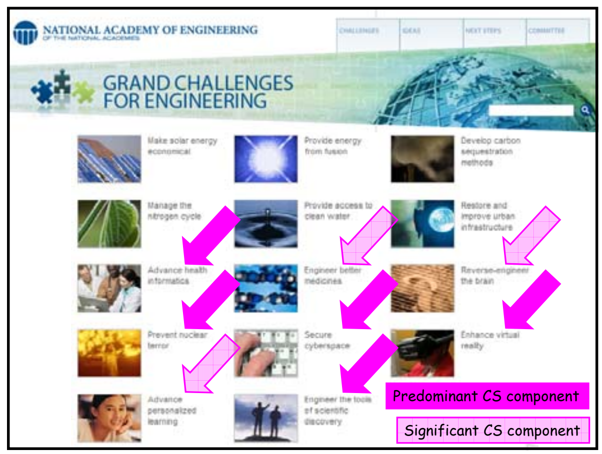
To this end, NSF asked CRA to create the Computing Community Consortium

- To catalyze the computing research community to consider such questions
  - To envision long-range, more audacious research challenges
  - To build momentum around such visions
  - To state them in compelling ways
  - To move them towards funded initiatives
  - To ensure "science oversight" of large-scale initiatives
- A "cooperative agreement" with NSF
  - Close coordination

The next ten years …







The bottom line …

- The future really couldn't be brighter
  - Well, ignoring Iraq, the economy, the election, and the failure of our education system