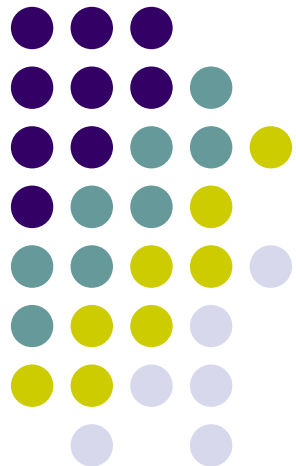
Enabling Discovery and Innovation: Role of NSF/OCI

46th AIAA Aerospace Sciences Meeting and Exhibit
Reno, NV
January 7-10, 2008

Abani K. Patra
Program Director, Office of Cyberinfrastructure
apatra@nsf.gov

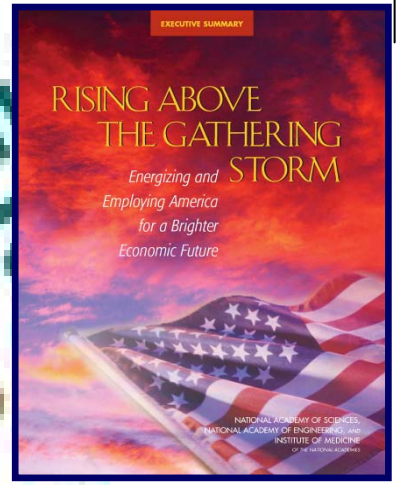


OCI People

- D. Atkins, *Director*
- J. Munoz, *Deputy Director*
- S. Meacham, *HPC, TeraGrid*
- K. Thompson, *Networking, Middleware*
- L. Nowell, *Data, HCI*
- D. Rhoten, *VO, LWD*
- A. Patra, *HPC, TeraGrid*

The World We Live in

- Friedman's Flat World vs Florida's Spiky World
  - "Industrial production" is increasingly global
  - Scientific Discovery and Innovation remain geographically local

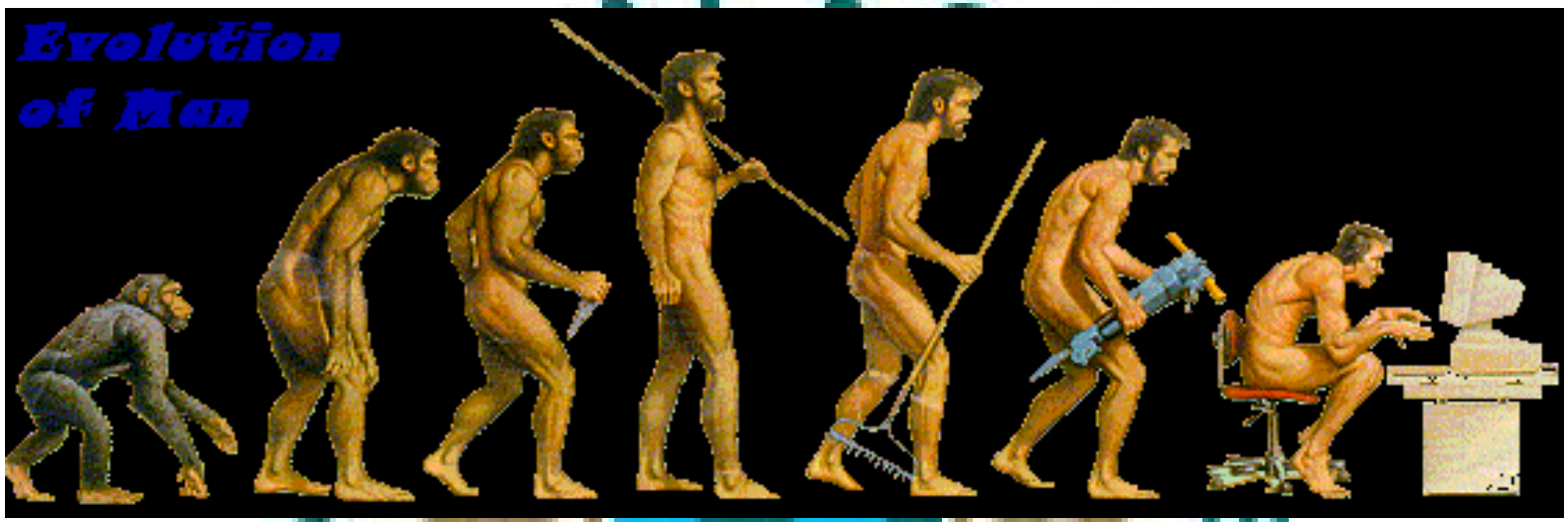


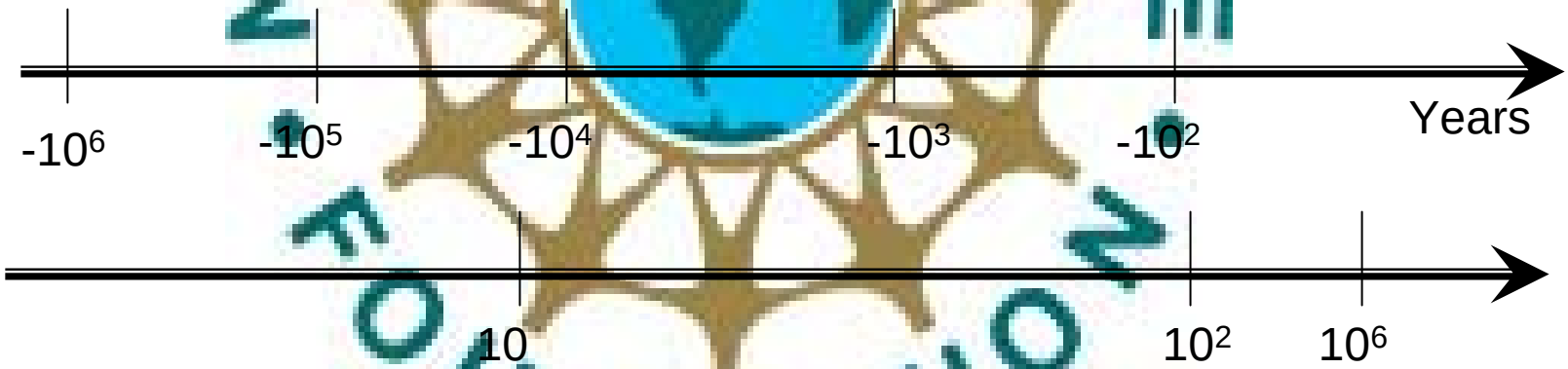
QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

Atlantic Monthly, October 2005

The World We Live in





Size of personal network / power of computing and communication

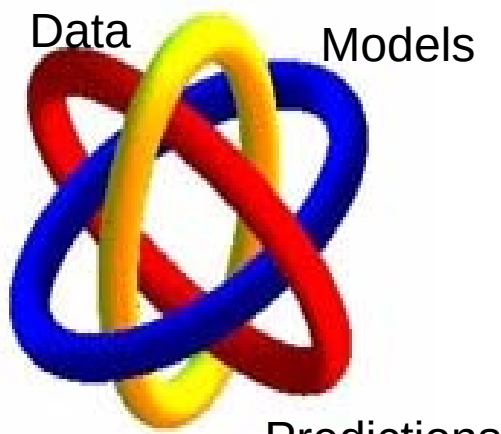
Data + Models + People = Predictions

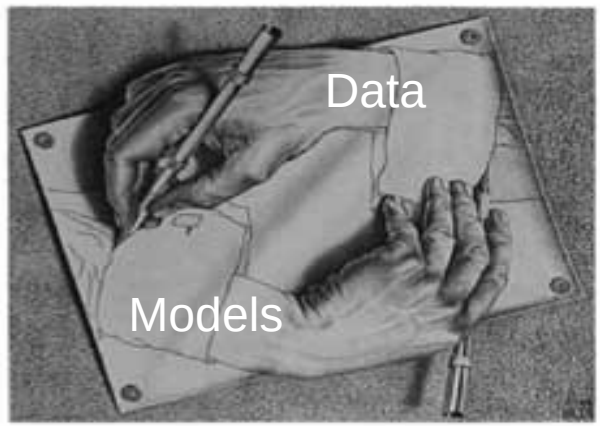
- Predictive simulations integrate the best models and **all available data** in a rigorous framework that enables meaningful extrapolation.
  - *Shark behavior in gulf predicted Katrina effects well*!

  → PREDICTIVE SCIENCE requires that all DATA and MODELS be seamlessly accessible to ALL with minimal barriers – technological and intellectual.

  → Synthesize diverse ideas and data

  → Bayesian Uncertainty Quantification vs. Newtonian Determinism

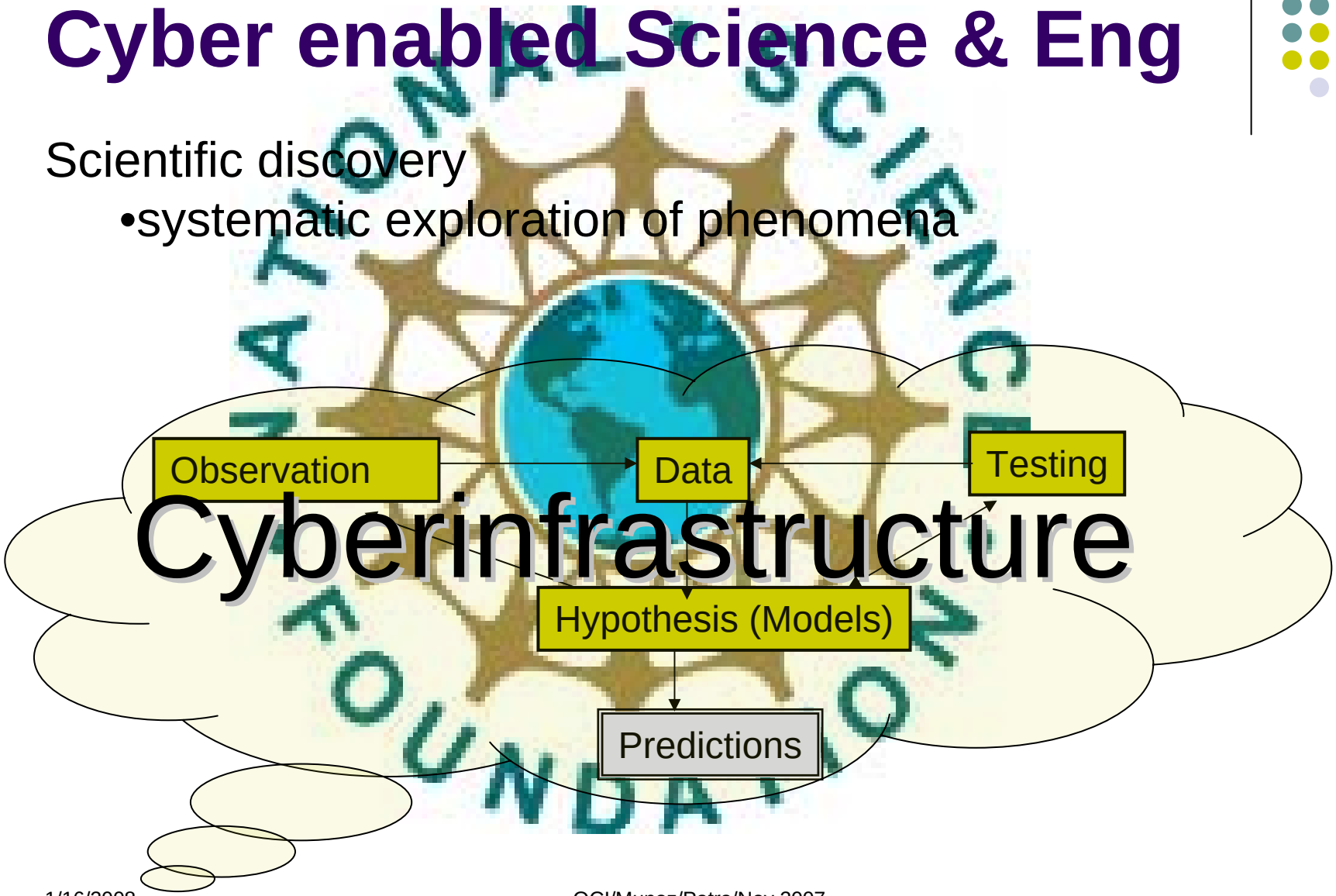




Cyber enabled Science & Eng

Scientific discovery

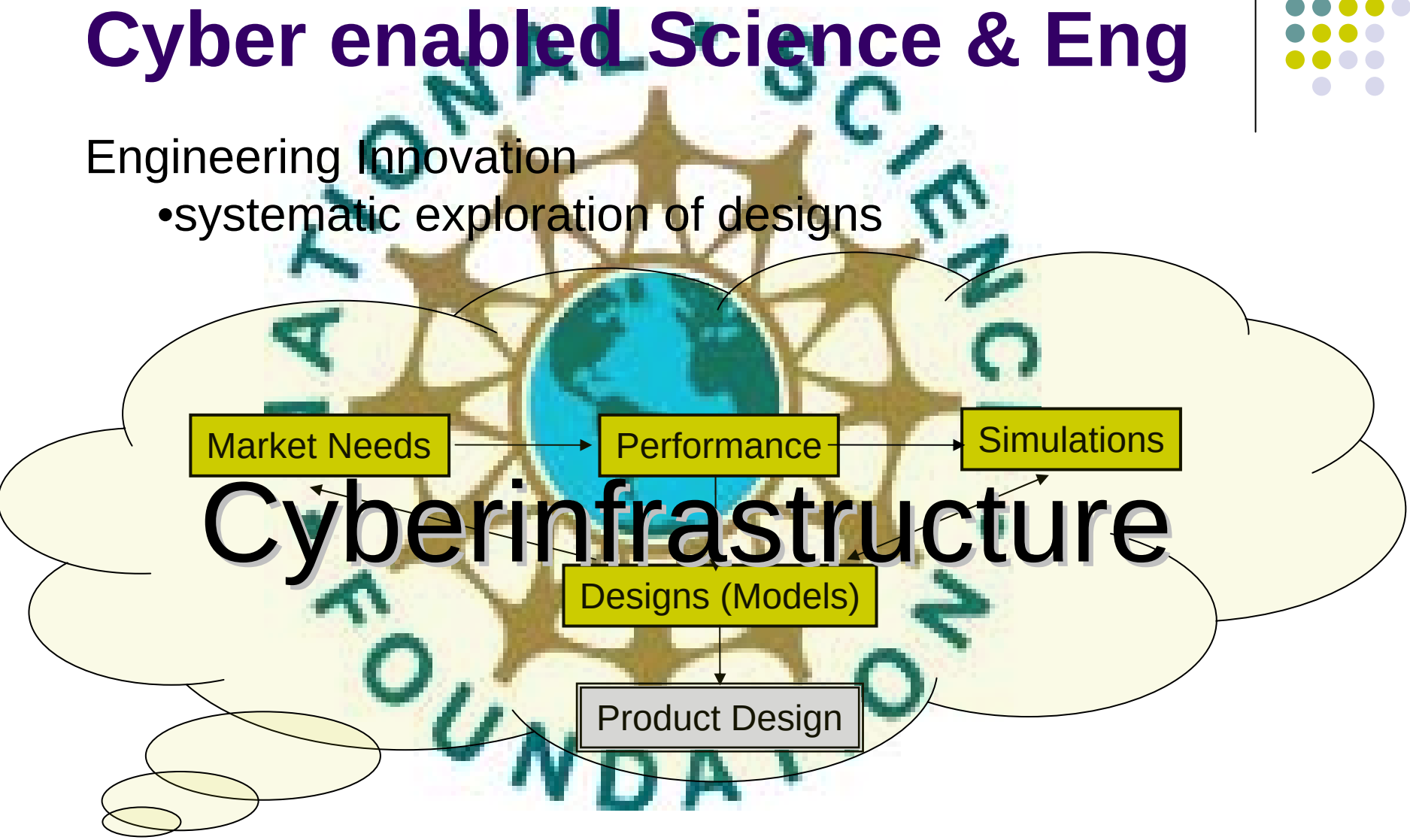
- systematic exploration of phenomena



Cyber enabled Science & Eng

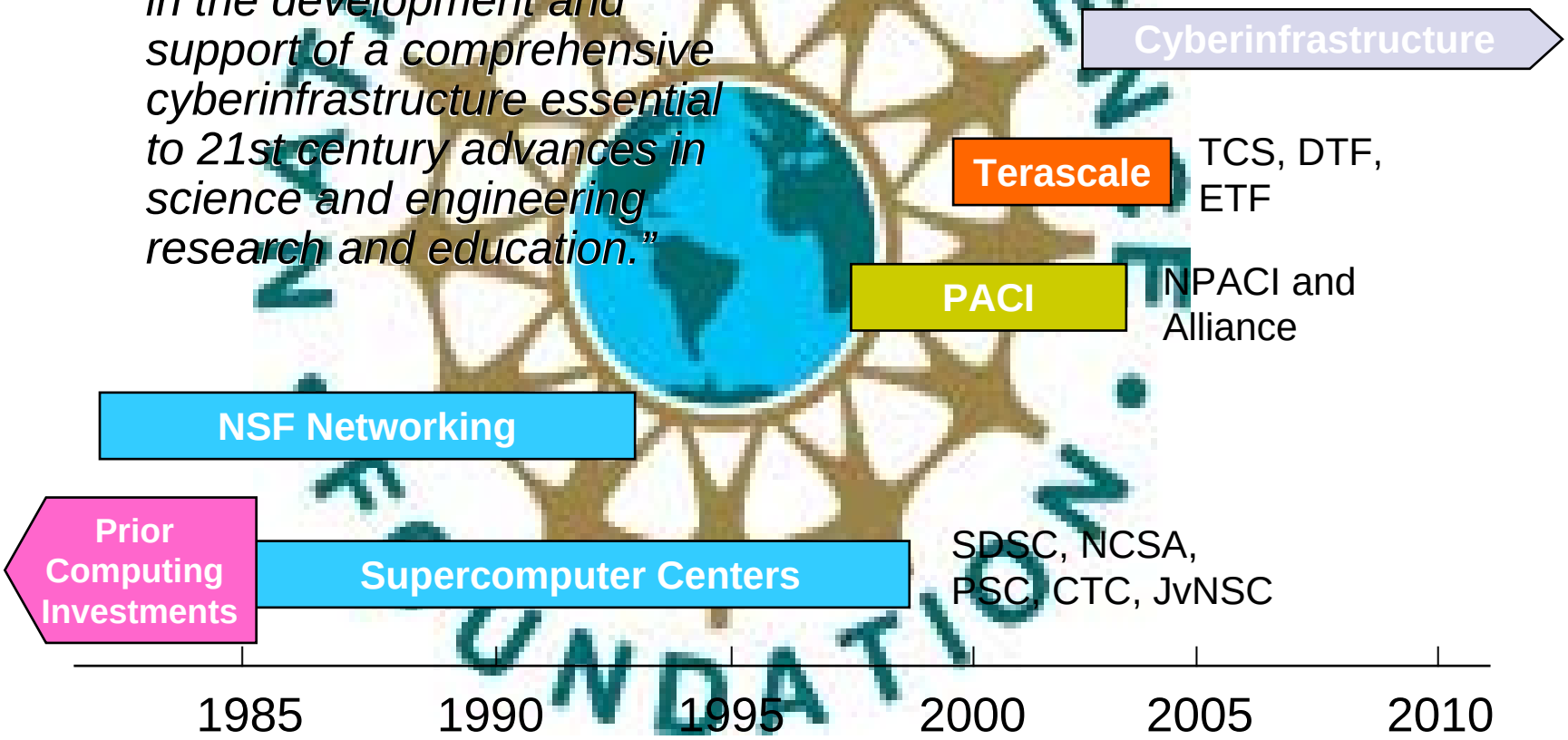
Engineering Innovation

- systematic exploration of designs



What is CI: The History

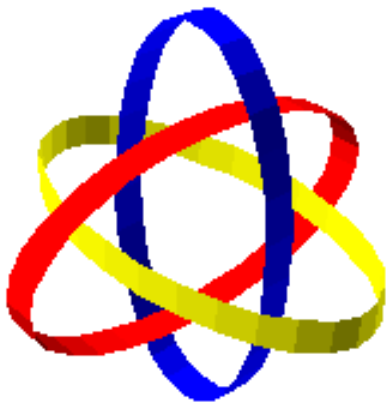
"NSF will play a leadership role in the development and support of a comprehensive cyberinfrastructure essential to 21st century advances in science and engineering research and education."



Slide from S. Iacono

Achieving the NSF CI Vision requires synergy between 3 types of activities

Transformative Application - to enhance discovery & learning



catalyzes

Provisioning -Creation, deployment and operation of advanced CI

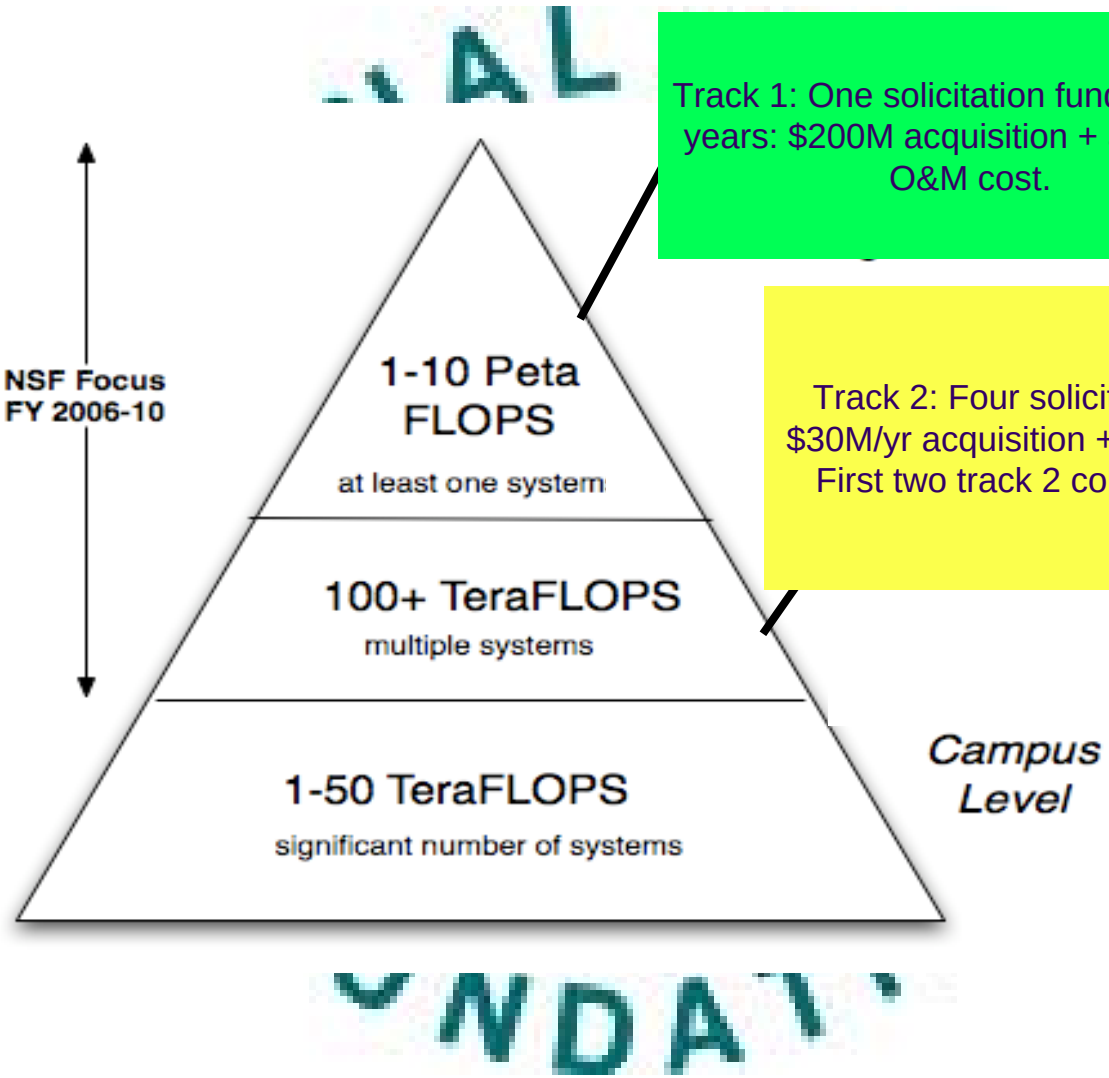
provides shared and connecting CI

Office of Cyberinfrastructure

Borromean Ring: The three rings taken together are inseparable, but remove any one ring and the other two fall apart. See www.liv.ac.uk/~spmr02/rings/

R&D to enhance technical and *social* effectiveness of future CI environments

Slide from D. Atkins



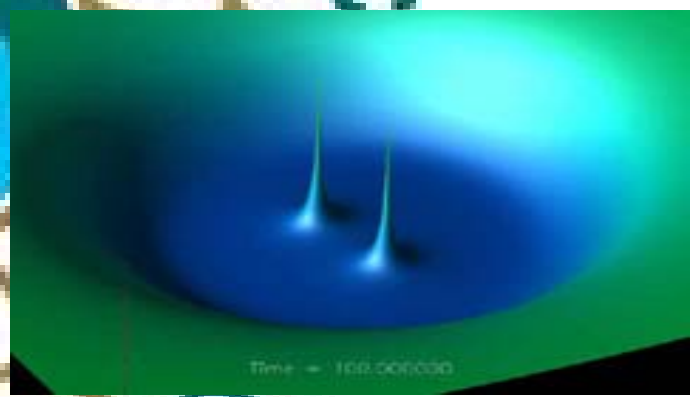
Track 1: One solicitation funded over 4 years: $200M acquisition + additional O&M cost.

Track 2: Four solicitations over 4 years: $30M/yr acquisition + additional O&M cost. First two track 2 competitions complete

Hardware Acquisitions

NSF building a portfolio of high end systems
- Equipment and 4/5 years of operations
- First Acquisition at Texas Advanced Computing Center → over 500 TF, 100 TB memory, 1.7PB disk, using 50K+ cores, Infiniband interconnect;

Proposed Ranger System at TACC



Campanelli et. al. produced the first simulations of generic, highly-spinning black holes with unequal masses by solving Einstein field equations using TeraGrid supercomputers and a specialized AMR technique.

Hardware Acquisitions

Two further acquisitions in progress

- "Track I" Leadership class machine - capable of sustaining PF/s on a range of problems with commensurate memory and communication frameworks -- awarded to University of Illinois for system named "Blue Waters" to be operational in 2011

- "Track II" – second competition awarded to Univ. of Tennessee JICS -- two more are planned

- "Track II" -- third competition underway

- NCAR has aggressive acquisition plan for support of atmospheric research!

Barriers

- *"But HPC and other high end CI tools are still only accessible to the few computational scientists[1] who can master a domain science, program parallel, distributed algorithms, and use/manage a supercomputer[2]."* – R. Waite on "Why MicroSoft has gotten into HPC?"

- *Petascale Systems will transform computational discovery but such systems will have deep hierarchies of functional units, cores, processors, cache, memory, and interconnect*, offering significant reward, but requiring significant innovation
  -- N. Nystrom, PSC

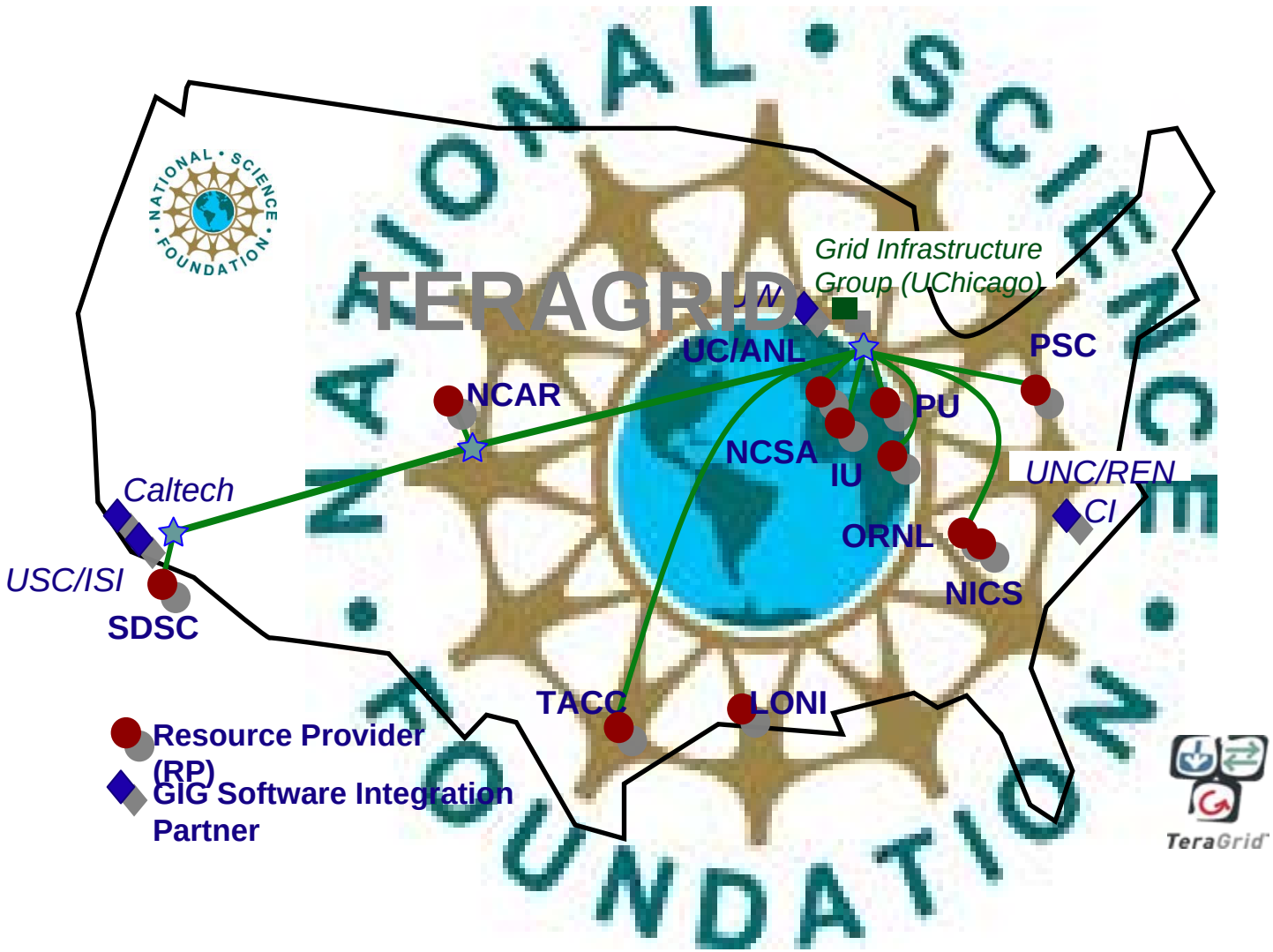
[1]*at select well provisioned sites*
[1]*whose architecture and programming change every 18 months!*

Barriers

- High Performance Computing, simple access to ALL RELATED data from observation, experiment and earlier simulations and good visualization must become **routine** – not heroic efforts!

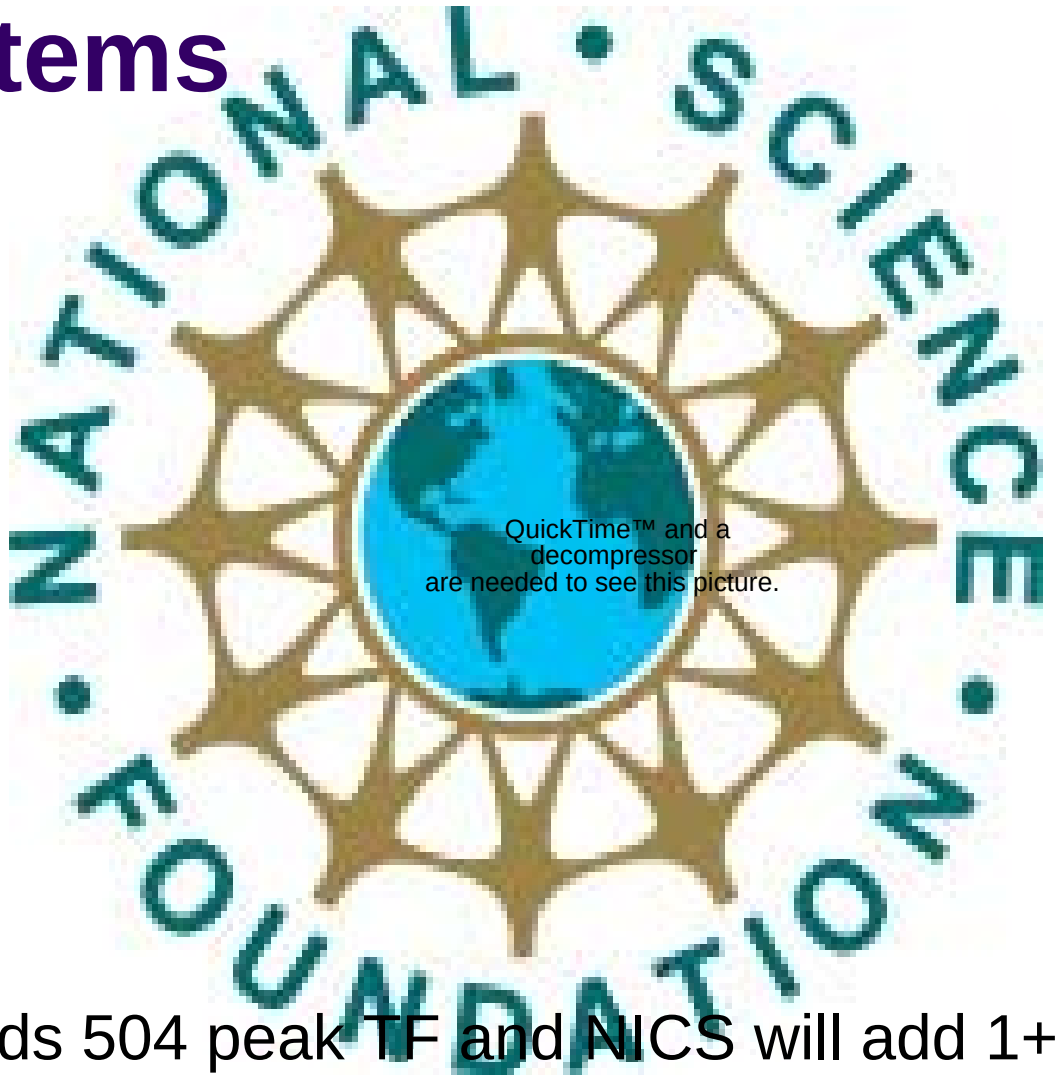
- Technological, social and economic barriers need to be surmounted.
  - Partnerships between all stakeholders will be needed.
  - Need the power of many!

11 Resource Providers, One Facility



- DEEP
- WIDE
- OPEN

Current & FutureTeraGrid HPC Systems



- Ranger adds 504 peak TF and NICS will add 1+ peak TF!
- Blue Water >> 1 PF ??

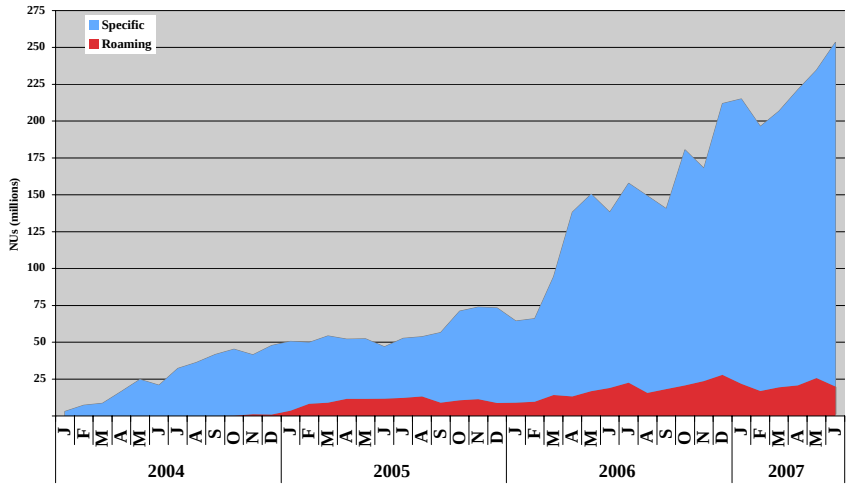
TeraGrid (TG)

- Eleven resource providers (RPs), four software integration partners, and the Grid Infrastructure Group (GIG), LSU, NICS joining in FY'08

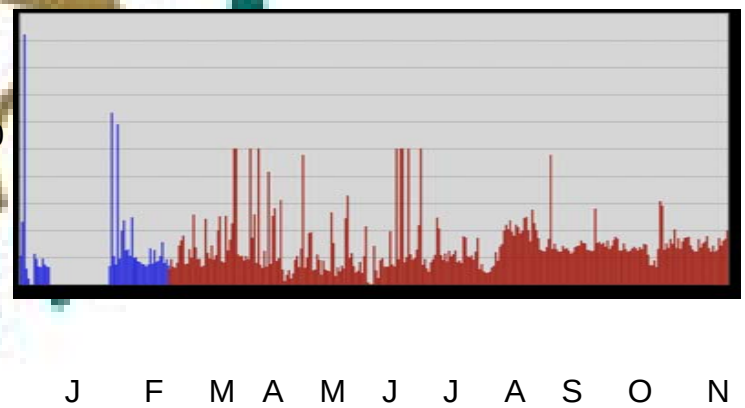
•Common user environments, global file systems, cross-site workflows, coordinated support …

- Peer reviewed merit based common allocation process for Computing, Data and as of this year sustained user support.

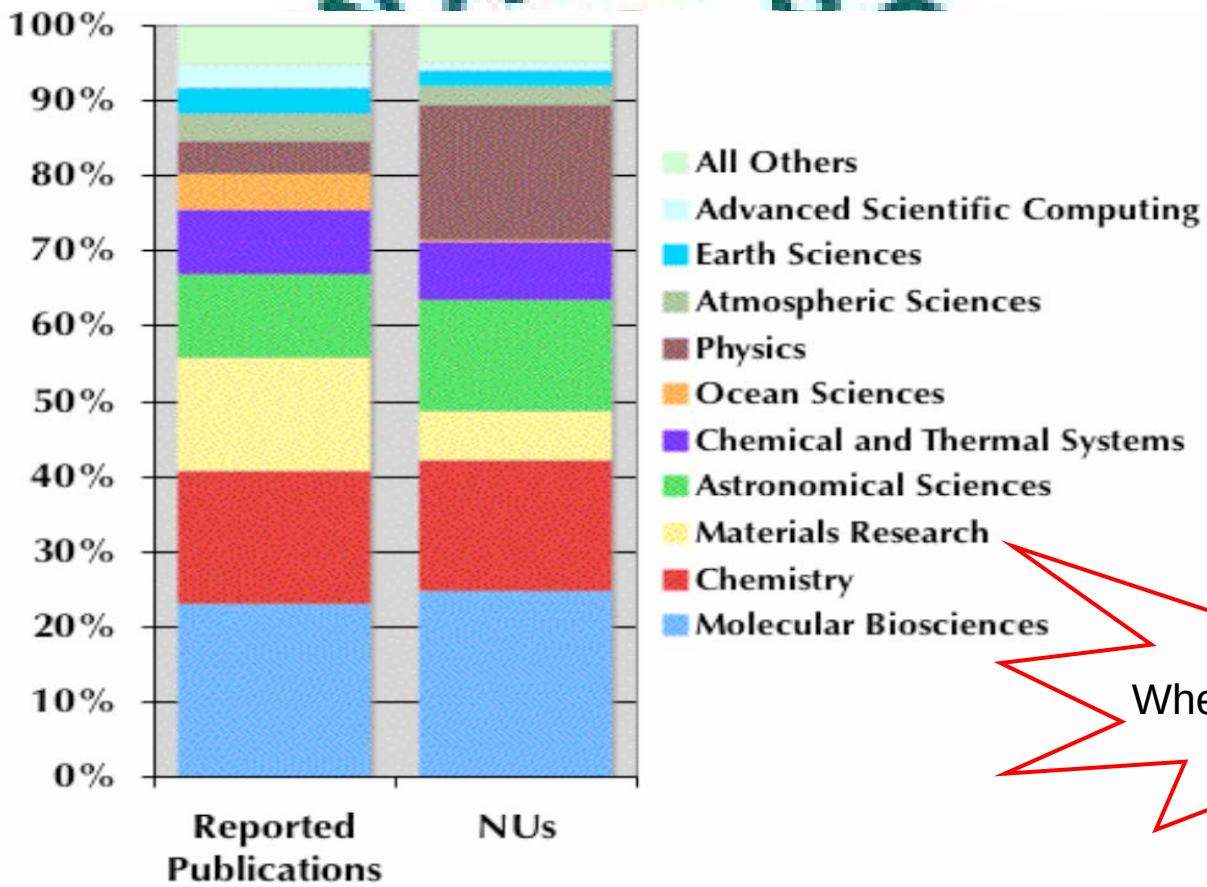
- 875* user publications acknowledged TG in CY2006 up from 485 in CY2005



[1] 1 NU is the computational equivalent of 1 CPU-hour on the Cray X-MP.



Usage and Science Impact



Where is Aerospace?

Users are from 265 institutions and 47 states.

TeraGrid User Portal

portal.teragrid.org

- TeraGrid is the largest, most powerful open science cyberinfrastructure in the world
- TeraGrid is about to become vastly more powerful with the integration of petascale computing systems
- TeraGrid offers resources, allocations, applications, and support to enable terascale/petascale science
- TeraGrid project expires in 2010. Plans for follow on are currently seeking community input. http://teragridfuture.org/

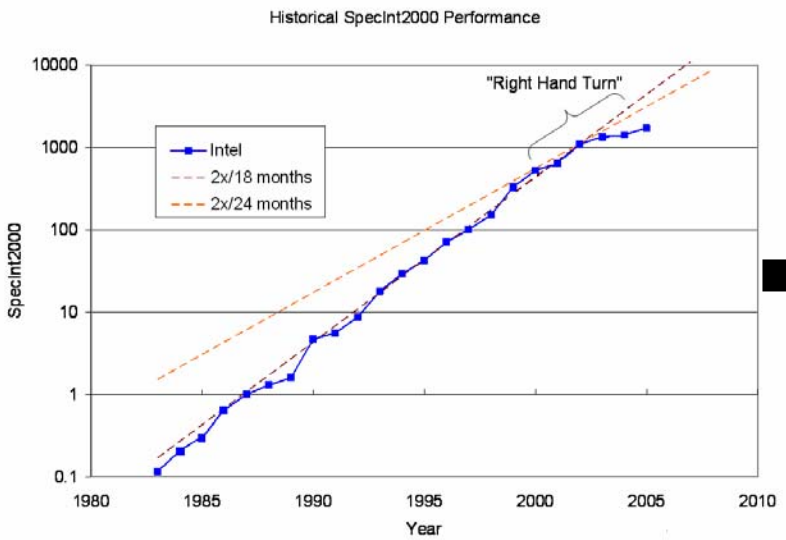
Teragrid Future

Contribute a Position Paper

The Steering Committee is soliciting position papers up to 3 pages in length. Position papers should describe how TeraGrid can help you, your project, community, or organization to address science or engineering research challenges and should consider what this might mean for future TeraGrid capabilities, governance, resources, services, etc. In keeping with the goal of the planning process, position papers should focus on needs and requirements in the timeframe 2010-2015. Some questions to help guide your thinking are given below.

- What is the vision for TeraGrid-II? Is it a continuation of the existing program with some enhancement, or is it something much different? What do we call it?
- Who are the users and what is TeraGrid-II's responsibility for large and small problems?
- What is the role of TG-II as an NSF-funded cyberinfrastructure?
- What is the role of TG-II in the larger cyberinfrastructure for research both nationally and internationally
- Continuity and its role in TG-II: Should there be institutional continuity or funding continuity over longer periods of time? Should the NSF consider new ways to review TeraGrid periodically without forcing a new competition for the infrastructure?
- Send position papers to Ann Zimmerman at asz at umich.edu by Friday, January 11, 2008. Please note that position papers will be posted on the planning web site and made available for public review and comment.

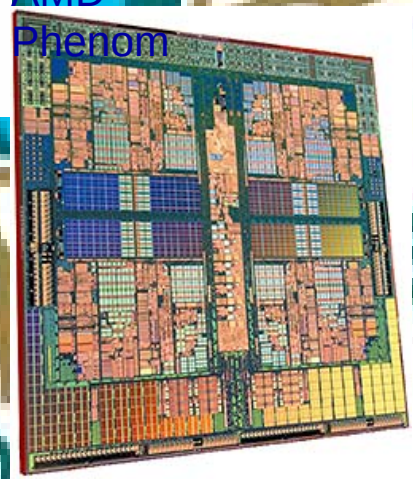
Barriers



Single Thread Performance

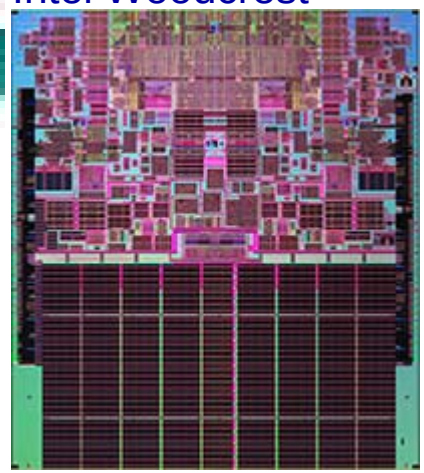
- Preserve system performance curve by increasing # of threads(cores)
- Reduce frequency/voltage => reduce power (15% => 48% power reduction)

AMD Phenom



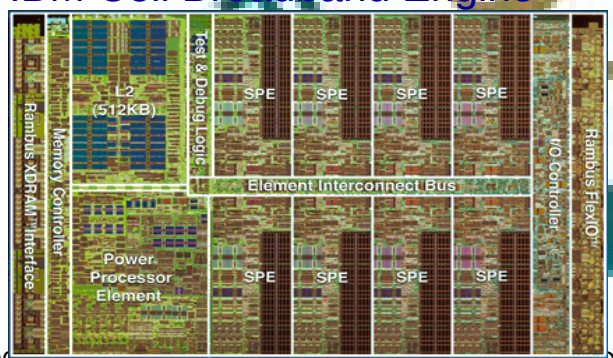
http://www.amd.com/us-en/assets/content_type/DigitalMedia/43264A_hi_res.jpg

Intel Woodcrest



http://www.intelstartyourengines.com/images/Woodcrest%20Die%20Shot%202.jpg

IBM Cell Broadband Engine



http://domino.research.ibm.com/comm/research.nsf/pages/r.arch.innovation.html?open

Barriers

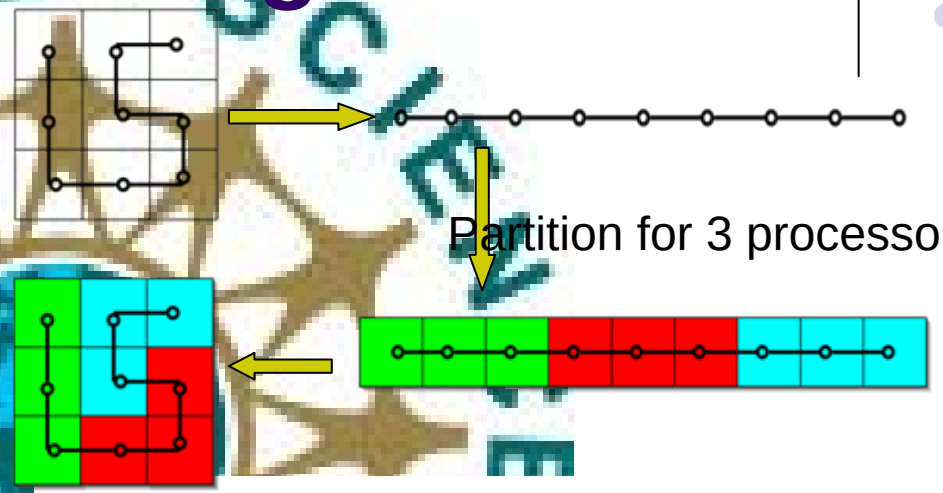
- Applications must be optimized across a hierarchy of levels: within cores, between cores on a socket, between direct-connected sockets, between sockets linked by an interconnect (possibly with rich topology), and across multiple levels of cache and memory.
  - Hierarchical Parallelism is key -- contrary to current flat data parallel models
- Support
  - *effective programming models & implementations* (software and hardware). Hybrid Programming(MPI+???), PGAS languages …
  - *effective performance tools*, 2 NSF funded groups last year Snavely UCSD et. al. and Malony et. al U. Oregon
  - *libraries and frameworks(?)*
  - *Fault Tolerance*.(?)

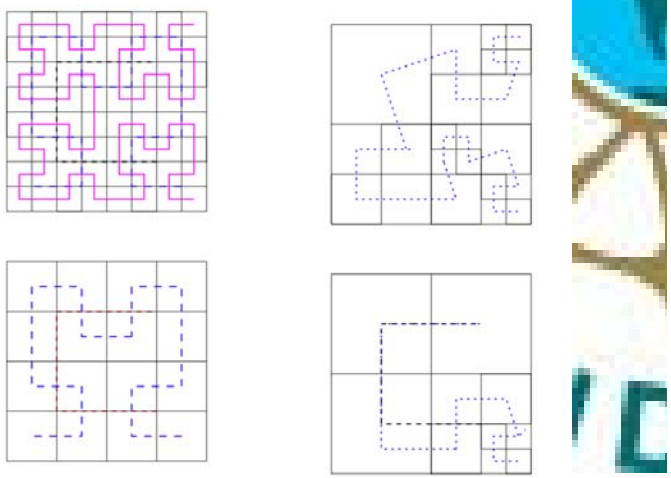
Ideas

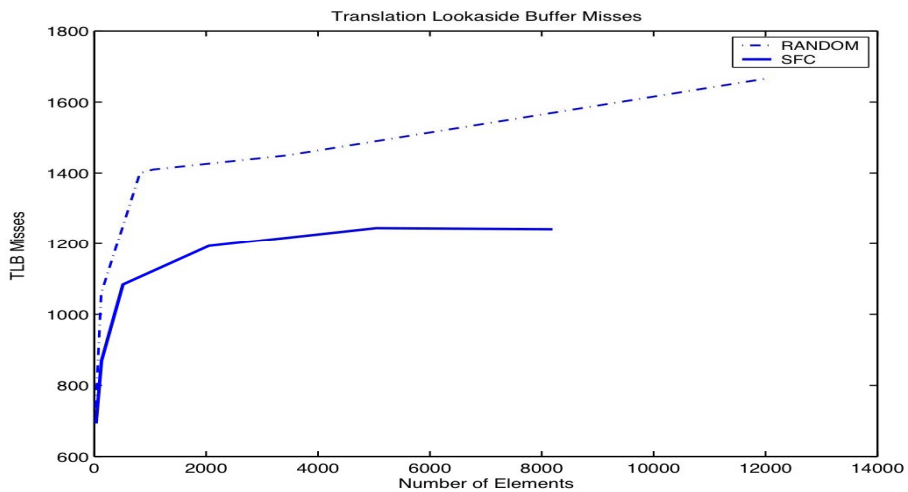
- Irregular Applications and Irregular Architectures
  - Adaptive Mesh Refinement, Multi-pole type schemes …
- Data Driven Dynamic Load Balancing --
  - Application steering using CPU, network, memory performance data?
- “Self Organization of Data, Computations and Communications?”
- Scale the science not the code
- ….

SFC Based Data Organization

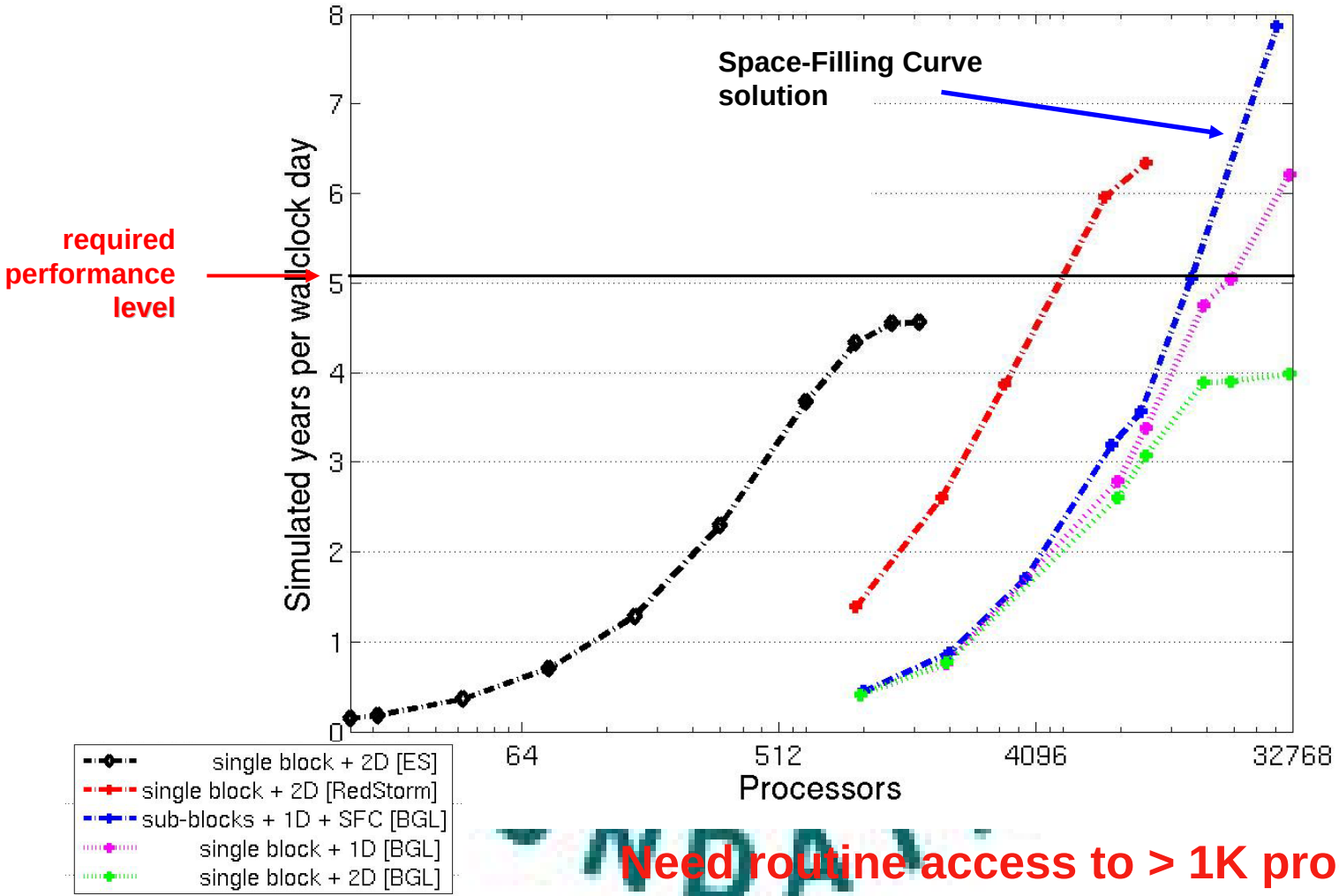
Use SFC based indexing
For data storage/access and partitioning







POP 0.1º (~10 km) Performance



Need routine access to > 1K processors to see true scaling behavior!

Slide from J. Kinter et. al.

Software Development for CyberInfrastructure

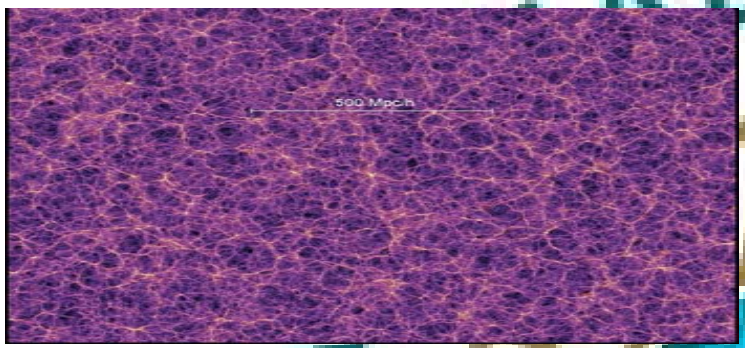
- **NSF 07-503 (closed), FY'08 deadline in April**
- Develop, deploy and sustain a set of reusable and expandable software components and systems that benefit a broad set of science and engineering applications
  - **software activities for** enhancing scientific productivity and for facilitating research and **education collaborations** through sharing of data, instruments, and computing and storage resources. The program requires open source software development
- Three focus areas in '07: HPC, Middleware, Digital Data
- Pending funding will be re-issued in 2008

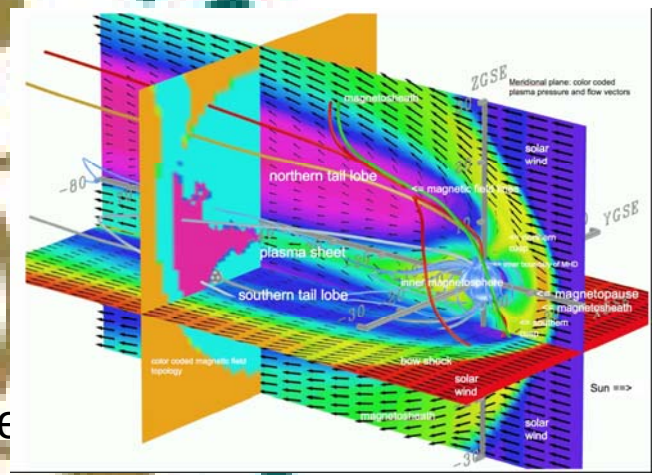
PetaScale Software Development

NSF 07-559 "PetaApps" –

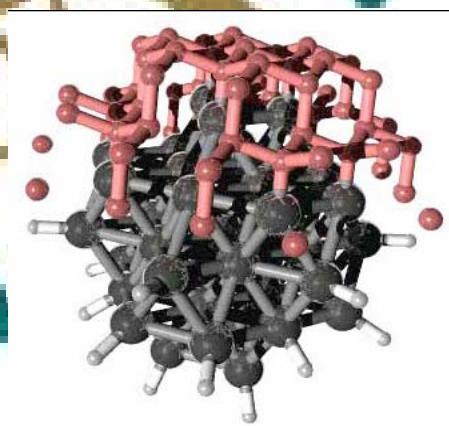
- "to develop the future simulation, optimization and analysis tools that can use petascale computing" … "implementation and exploitation of forefront techniques"
- Cross Foundational program with participation from OCI, MPS, CISE, Engineering and GEO
- $26M in 18 awards each of less than $2M over 3-5 years. Wide range of science/engineering Applications funded
- Subsequent years in discussion currently

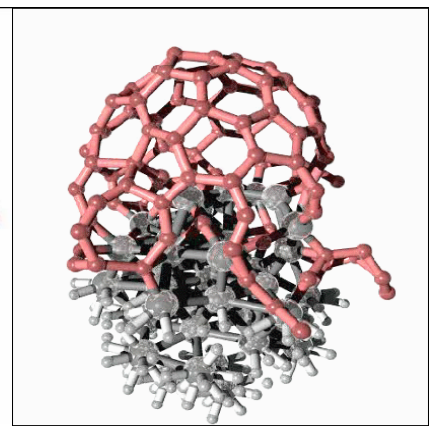
Significant Science Applications Funded





- Climate Change
- Earthquake Dynamics and structural response
- Nanoscale Transistor Models
- Supernovae Simulations
- High Reynolds Number Turbulent Flows
- Particulated Flows
- First principle molecular dynamics
- Quantum Chromo Dynamics …





Strategic Technologies for Cyberinfrastructure

PD 06-7231 Standing program Strategic Technologies for CI – Core OCI program

- support work leading to the development and/or demonstration of *innovative* cyberinfrastructure services for science and engineering research and education that fill gaps left by more targeted funding opportunities
- consider highly innovative cyberinfrastructure education, outreach and training proposals that lie outside the scope of targeted solicitations -- *("the crazies").*

Two dates each year

August and February

$2.5M

Cyber-Enabled Discovery and Innovation (CDI)

Objective of CDI:

Enhance American competitiveness by enabling innovation through the use of computational thinking.

- Multi-disciplinary research seeking contributions to more than one area of science or engineering, by innovation in, or innovative use of computational thinking
- Computational thinking refers to computational…
  - …Concepts, Methods, Models, Algorithms & Tools

CDI is Unique within NSF

- five-year initiative;
- minimum of $26M in FY 2008 in a common solicitation with up to an additional $26M available to the divisions
- Budgets projected to increase each year to a total investment of $750 M
- to create *revolutionary* science and engineering research outcomes made possible by innovations and advances in computational thinking
- emphasis on bold, multidisciplinary activities

Key Dates and URL:

- Letters of Intent (required) due: Nov 30, 07
- Preliminary Proposals due: Jan 8, 08
- Full proposals due: April 29, 08
  - Full proposals by invitation only!
- Awards: no later than October 2008

- For more information:
  - Solicitation: http://www.nsf.gov/pubs/2007/nsf07603/nsf07603.htm
  - FAQ, examples, resources: http://www.nsf.gov/crssprgm/cdi .

Transformative Research

- NEW in NSF Review Criteria:
  - To what extent does the proposed activity suggest and explore creative, original, or potentially transformative concepts?
- Additional Review Criteria in solicitation
  - Special emphasis will be placed on proposals that promise to enhance competitiveness, innovation, or safety and security in the United States.
- "Science/Innovation First"
- High Risk and High Promised Impact

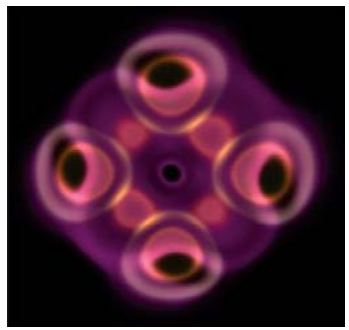
CI enabled Science & Eng

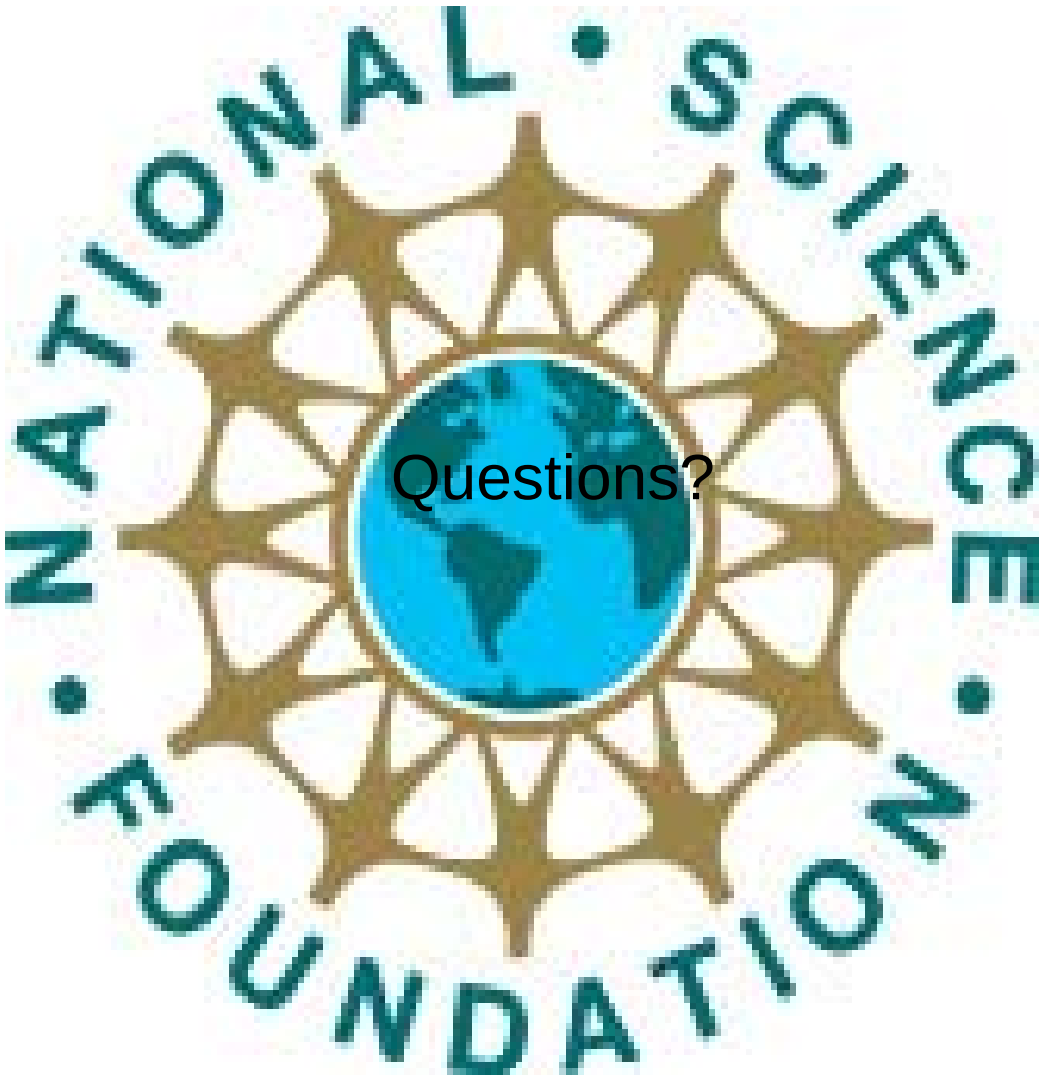
- "Discover, codify, disseminate" – classical paradigm for science comfortably done by simple rigidly structured organizations.

→ Easily Available High End Data, Computing and Analysis Resources + Community Driven Collaborative Strucures (e.g. Science Gateways, … like nanoHUB, LEAD …)

- "Engage, Explore, Apply, Share"– new paradigm for scientific efforts.

  -- McKinsey & Co. Strategic Management of Enterprise Knowledge





Questions?

Three CDI Themes

CDI seeks transformative research in the following general themes, via innovations in, and/or innovative use of, computational thinking:

- **From Data to Knowledge:** *enhancing human cognition and generating new knowledge from a wealth of heterogeneous digital data;*
- **Understanding Complexity in Natural, Built, and Social Systems:** *deriving fundamental insights on systems comprising multiple interacting elements;* and
- **Building Virtual Organizations:** *enhancing discovery and innovation by bringing people and resources together across institutional, geographical and cultural boundaries*.

Types of Projects

- Project size not measured by $$ but by magnitude of effort
- Three types are defined: Types I (~2 PI, 2 GRA), II (~3 PI, 3 GRA, 1 post-doc), and III (center scale).
- Type III, center-scale efforts, will not be supported in the first year of CDI