Introductions: How did I get here?

Hints at where I will try to take this...

- High energy physics
- Black hole perturbation theory
- Very hard, try supercomputers
- Black hole collisions
- Very hard, try working with computer scientists
- Neutron star collisions
- Very hard, try larger scale collaborations, add in high speed nets, grids
- Build interdisciplinary structure in LA
- Very hard, try national scale!

Developing a working code was at least as hard as the theory!

Grand Challenge teams, CS experts, new tech, critical to success

Critical importance of international cooperation and networking

Software engineering, tools, reusable components

Grids: collaboration, data, computation for complex problems

Visualization! Local, remote, large data, collaboration, PR, science!

Optical networks, data deluge

All disciplines can use common components

How much this all costs!
General Points

- ACCI is for *all* of NSF, not just OCI
  - OCI Dir (Seidel) is designated Fed. Official
- Federal Advisory Committee Act (FACA)
  - Charter, rules regarding Col, minutes, etc
- Your advice is “advisory” :-)
  - You, and subcommittees you may have, are a primary mechanism for us to get advice
  - I am going to propose task forces and will request ACCI involvement
- I, OCI, and NSF need your advice!!
Gamma-ray bursts!

- All energy emitted in lifetime of sun bursts out in a few seconds: what are they? Colliding BH-NS? SN?

- GR, hydrodynamics, nuclear physics, radiation transport, neutrinos, magnetic fields: globally distributed collab!

- Scalable algorithms, complex AMR codes, viz, PFlops*week, PB output!

LHC: What is the nature of mass? Higgs particle?

- ~10K scientists, 33+ countries, 25PB data, distributed!

- Planetary lab for scientific discovery!

Now, compare with observation...LIGO! GR becomes a data-driven science.
Grand Challenge Communities Combine it All…

*Where is it going to go?*
Collaborations for Complex Problems

Great Challenge of 21st Century

- Every field of science
  - General Relativity
  - High Energy Physics
  - Geosciences, Bio, SBE...
  - And all combinations...
- Science and Society being transformed
- Cyberinfrastructure plays central role in collaborations for complex problems
  - No single community can attack challenges
  - Technical and social issues for distributed communities; may not know each other!
  - Compute & Data intensive
Conclusions

• Complex problems are a key challenge for next generation science and engineering
  • Many of them require all CI to solve, not just part (e.g., HPC)
  • All are becoming data-challenged
  • Large-scale collaborations are often needed to address these problems

• OCI has the responsibility to build out CI to support all this, for all directorates, for all campuses in the nation
NSF Vision

“national-level, integrated system of hardware, software, data resources & services... to enable new paradigms of science”

1. Virtual Organizations for Distributed Communities
2. High Performance Computing
3. Data & Visualization/Interaction
4. Learning & Work Force Needs & Opportunities
What is OCI doing to address this?
High End Computing

- "Modeling, simulation, and knowledge from data collections, [which] is increasingly essential to scientific and engineering
- Sustained petascale capable systems
- HPC software and tools
- Necessary scalable applications
- Sharing among academic institutions to optimize the accessibility and use of HPC assets deployed and supported at the campus level"
High End Computing

- “Modeling, simulation, and knowledge from data collections, [which] is increasingly essential to scientific and engineering
- Sustained petascale capable systems; going beyond HPC!
- HPC software and tools
- Necessary scalable applications
- Sharing among academic institutions to optimize the accessibility and use of HPC assets deployed and supported at the campus level”

AMR on a million processors?

Toolkit for complex CI?
Greatly expanding capacity of the TeraGrid for digital exploration with reduced oversubscription and queue wait times.

- PSC-Phase 2 & UTK Phase 3
- UTK Phase 2
- UTK Phase 1 & Ranger upgrade
- TACC Ranger
- TG Aggregate 6-07

TF/s Peak

June 07: 180
Early 08: 684
Summer 08: 925
2009: > 1,500
2010: > 2,800

Track1 brings this up by many times!
Blue Waters Project

Blue Waters Petascale Computing System (2011!)

• Blue Waters General Characteristics
  • Based on IBM PERCS
  • 1 petaflops sustained performance on real applications

• Blue Waters System Characteristics
  • > 200,000 cores using multicore POWER7 processors
  • > 32 gigabytes of main memory per SMP
  • > 10 petabytes of user disk storage
  • 100 Gbps external connectivity (initial)
  • Fortran, Co-Array Fortran, C/C++, UPC, MPI/MPI2, OpenMP, Cactus, Charm++

• Blue Waters Interim Systems at NCSA
  • POWER 5+ and POWER6 software and application development testbeds

• Blue Waters System Training and Support
Two very important programs

- **STCI: Strategic Technologies for CI ($3M only!)**
  - Development and/or demonstration of innovative CI; highly innovative CI education, outreach and training proposals outside scope of targeted solicitations
  - Two dates each year: August and February

- **PetaApps: Petascale Applications ($18M)**
  - Future simulation, optimization and analysis tools for petascale computing; implementation and exploitation of forefront techniques
  - Cross Foundational program with participation from OCI, MPS, CISE, Engineering, SBE and GEO
    - Last Year: $26M in 18 awards each of less than $2M over 3-5 years.
  - Not a regular program...something needs to be!
TeraGrid Phase III (XD)

- XD is looking for innovative proposals for a new infrastructure to deliver the next generation of high-end national digital services: anticipated $32M/yr

- Goal is to advance science and engineering by providing researchers and educators with the capability to work with extremely large amounts of digitally represented information and making it easy to move between local and national resources

- Anticipate researchers working with much larger range of digital artifacts, including digital text, digitized physical samples, real-time data streams, ...

- Status: preproposals received; panel met Dec 1-5; recommendations on going forward
Missing or Weak in HPC

- STCI only regular program, and only $3M!
- Planned Programs
  - SDCI (?)
- Needed programs
  - A BIG software initiative! D. Reed: Software/computing should be 9:1!
    - “CESM”, other activities, sustainability
  - A BIG applications initiative
- Future
  - Clouds
  - Roadmap beyond Track 1,2, XD
Data, Data Analysis, and Visualization

• “Any cogent plan must address the phenomenal growth of data in all dimensions

• Goals are to
  • Catalyze the development of a system of science and engineering data collections that is open, extensible, and evolvable
  • Support development of a new generation of tools and services for data discovery, integration, visualization, analysis and preservation

• The resulting national digital data framework will be an integral component in national CI”
Data, Data Analysis, and Visualization

• “Any cogent plan must address the phenomenal growth of data in all dimensions.”

• Goals are to
  • Catalyze the development of a system of science and engineering data collections that is open, extensible, and evolvable
  • Support development of a new generation of tools and services for data discovery, integration, visualization, analysis and preservation
  • The resulting national digital data framework will be an integral component in national CI”

Where is my simulation data? Compare with experiment?

I can’t visualize my results!
"The data deluge represents an opportunity to advance U.S. leadership in science and technology, and harnessing it has become a national priority."

Storage Networking Industry Association (SNIA) 100 Year Archive Requirements Survey Report

"there is a pending crisis in archiving... we have to create long-term methods for preserving information, for making it available for analysis in the future." 80% respondents: >50 yrs; 68% > 100 yrs
$100M DataNet Program (5 Years)
(Sustainable Digital Data Preservation & Access Network Partners)

- **Goals:**
  - Catalyze development of multi-disciplinary science & engineering data collections: open, extensible, & evolvable, sustainable over 50+ years.
  - Support development of a new generation of tools & services facilitating data acquisition, mining, integration, analysis, visualization.

- **Status:** NSB approved two awards! Round 2 pre-proposals Nov. 13, w/invited full proposals due May 15, 2009.
INTEROP and IRNC

**INTEROP**
- Re-purposing data – using it in innovative ways & combinations not envisioned by its creators – requires finding & understanding data of many types & from many sources. Community building!
- Interoperability – ability of two or more systems or components to exchange information & use exchanged information.
- Status: Highest priority of DWG, 7 awards made in first competition. Status FY09 competition is uncertain, pending funding availability.

**IRNC**
- Being readied for 2009
- Jen Schopf and Bill Chang
- 15:1 ROI from international collabs
- Science and political benefit: Pakistan

---

**Awards**
- **TransPAC2** (U.S. – Japan and beyond)
- **GLORIAD** (U.S. – China – Russia – Korea)
- **Translight/PacificWave** (U.S. – Australia)
- **TransLight/StarLight** (U.S. – Europe)
- **WHREN** (U.S. – Latin America)
Missing or Weak in Data

- INTEROP funding
- Integration of DataNet with overall architecture
- Research & development programs on metadata, search, tools, future data environments
  - Some metadata awards in 2007 SDCI program
- Visualization
  - Workshops, small STCI awards...
- Network programs
  - EIN, eScience apps, Connections program, MREFC, remote instruments, etc...
  - Software, scheduling, integration of campuses
- Cybersecurity!!
Virtual Organizations for Distributed Communities

- "A VO functions as a coherent unit...through the use of end-to-end CI systems, providing shared access to resources and services, often in realtime"

- Technological framework...experimental facilities, instruments and sensors, applications, tools, middleware, remote access

- Operational framework from campus level to international scale…"

- Specific interpretation: Next generation Grand Challenge teams for science, engineering, humanities...
Virtual Organizations for Distributed Communities

- “A VO functions as a coherent unit...through the use of end-to-end CI systems, providing shared access to resources and services, often in realtime

- Technological framework...experimental facilities, instruments and sensors, applications, tools, middleware, remote access

- Operational framework from campus level to international scale…”

- Specific interpretation: Next generation Grand Challenge teams for science, engineering, humanities...
**VOSS, VOs**

- **VOSS**

- What constitutes effective virtual organizations? How do they enhance research and education production and innovation?

- **CDI**

- **VOs**

- Science Gateways: NanoHub, NEES, LEAD...
Missing or Weak in VOs

- VOSS-2?
- VOs for K-20, citizen scientists
- Need a new Grand Challenge Communities program! Distributed communities; collaborative frameworks, tools, social networks for complex problem solving
Learning and Workforce Development

• “NSF will:
  • Identify and address the barriers to utilization of cyberinfrastructure tools, services, and resources
  • Promote the training of faculty, educators, students, researchers
  • Encourage programs that will explore and exploit cyberinfrastructure, including taking advantage of the international connectivity it provides...”
Learning and Workforce Development

• “NSF will:
  • Identify and address the barriers to utilization of cyberinfrastructure tools, services, and resources
  • Promote the training of faculty, educators, students, researchers
  • Encourage programs that will explore and exploit cyberinfrastructure, including taking advantage of the international connectivity it provides…”
CyberLearning

- New Report to NSF, sponsored by OCI, EHR
- Town hall meeting at NSF to brainstorm!
- Programs desired to
  - Bring this into classrooms
  - Create next generation of “cyber-savvy” workers, educators
Missing or Weak in LWFD

- IREC, CI-TEAM 2
- Broadening participation
- Programs for development of young scientists
  - Computational science grads, postdocs, career awards
  - Who researches, prototypes, develops, uses next generation CI
- Curriculum development, best practices for computational science & engineering & edu
  - This could be one of our biggest contributions!
  - No group doing this! NSF can have huge impact!
I need all of this to start to solve my problem!
Analysis: What is needed

- Comprehensive, balanced, integrated national high performance cyberinfrastructure!

- Beyond HPC/TG: Integrates campuses, multiple CIs & agencies, instruments, data, grids, HPC, clouds, optical nets, software...

- 88% 2007 budget -> HPC! (65% in 2005, 70% 2006)

- All components needed for large scale distributed collaborations, apps

- Support for cyber-workforce development

- Computational Science as a career path

- New Grand Challenges with new scales of interdisciplinarity, proper reward system
Remarks

• OCI big step forward: tremendous vision
  • Huge gap between vision and implementation
• Still focuses on “equipment”
  • Needs to fill out portfolio to comprehensive CI, including software, data, viz, etc...
• Hard to do more than HPC: $185M budget, Blue Waters (the machine) costs ~ $200M

• The great opportunity:
  • OCI can become a base for computational science & engineering research and education
  • As neutral partner, OCI can catalyze integrative approach to collaborative computational science for complex problems
To do all this: Staff and Budget
Staff to handle very complex awards

• 10 Program Staff
  • 5 FTE, 5 IPA
    • Ed, Jose, Steve not (supposed to be) assigned program duties
      • In reality, Steve has incredible load!
    • Currently 5 vacancies! (1 FTE, 1IPA, Greer, Rhoten, Thompson)
    • Requesting 2 FTE, 2 IPA additional

• 7 Office staff
  • 3 program support
  • Currently 2 vacancies (1 on detail from CISE)
  • Requesting 1 additional
“CI Annual Expenditures”: $865M

<table>
<thead>
<tr>
<th>Year</th>
<th>OCI Budget</th>
<th>NSF Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>1986</td>
<td>$47M</td>
<td>$1.4B</td>
</tr>
<tr>
<td>2008</td>
<td>$185</td>
<td>$6.1B</td>
</tr>
</tbody>
</table>

2008/1986: 3.9/4.4

OCI grows more slowly than NSF over 22 years!!

OCI Expenditures in Other Parts of NSF: $680M

OCI Budget: $185M
What next?

- We have an excellent high level vision document!
- We have started to address some parts of it very well...
- We have started to address other parts of it in pieces...
- Some parts are missing entirely

- All parts need to be aggressively built out, integrated, coupled to/driven by applications
  - Need *end-to-end* integration; better architecture, software, apps!
  - Need to address support of a computational science community
    - Computational Science is 3rd pillar of science & engineering?
    - Who creates, prototypes, uses next generation CI?
Some missing elements
Barely show up in Vision Doc

- High performance optical nets, bandwidth (lambdas?) to desktop
  - Campus, RONS, Internet-2, NLR, IRNC,...: end-to-end!
  - Bandwidth doubling every 9 months? Researchers are choked at their labs!
    - A. Bement: “Those massive conduits are reduced to two-lane roads at most college and university campuses”

- National CI must be well integrated with campus, local lab, desktop CI
  - Overall architecture, campus integration!
  - Campus is where all the researchers are!
  - Compute, manage, visualize, analyze the data...from lab!
“Itself a discipline, computational science serves to advance all of science...despite the fundamental contributions of computational science now constitutes the “third pillar” of scientific inquiry, enabling researchers to build and test models of complex phenomena...

OCI can be the missing base for computational science at NSF...would have tremendous impact
COV Themes

• Imbalance in portfolio due to insufficient resources
  • “a bad situation... failure to address will result in major opportunity losses”

• Critical staffing shortages
  • “OCI cannot continue through heroic efforts”

• Research activities: they make the case for us to focus on computational science
  • “OCI must be a research organization”

• OCI as a broker of community activities: OCI can play central role in catalyzing cross-cutting research programs across NSF (e.g., the new grand challenges)
  • “sustained effort must be devoted to bridging divides between the infrastructure and research programs”
Strategies for Funding

- Partner with Directorates
- Partner with Agencies in US and Abroad
  - Deep discussions with DOE, EU already
    - DOE: possible MOU, areas of synergy, complementarity
  - Future discussions with Asia
- Increase investments!
  - OCI/NSF ratio gets smaller over time, existing budget more concentrated on HPC...
  - New programs
  - Need help here!
Transition Team Update

- Obama transition team very active at NSF
- Henry Kelly visited OCI, many ADs, OD
- Very well informed, sympathetic
- Spoke to Henry and others about FY09 and FY10 funding

- One priority is to inject money rapidly into economy
- All areas of interest, including HPC, networks/campuses, storage, WF development, software, applications
Task Forces

- Timelines: 12-18 months
- We will need help to organize, energize
- Co-led by NSF PDs: OCI + other
- Membership from ACCI
- Membership from community
- Possible member from other agencies: DOE, EU, etc
- Workshop(s)
  - Recommendations
- We then go back and develop programs
Task Forces

- CI Reuse (internal)
- Software
- Networking & campus integration
- Education and workforce development
- Computing infrastructure (HPC, clouds, grids...)
- Data, viz
- VOs and next generation grand challenge applications
My Ideas for Rough Plan for Future

• Attempt to fill out holes in current CI coverage
  • Software, algorithms, networks...
  • Alternate computing modalities beyond HPC: Clouds, data, campus integration, etc...

• Workforce development!

• Reposition OCI as base for CI development, deployment & computational science research and education

• Launch major new program: VOs to solve complex problems at frontiers of science and engineering
  • CI-driven, with above pieces in place first
  • This is the heart of computational science!
Need Help!

• Budgets and Staff!
  • How to expand both?
  • CI deployment: must go beyond HPC procurement: software! apps! networks! clouds!

• What are the right messages? To whom should they be sent?
  • Begun work with NAS/NAE, SIAM, Agencies, Community, I-2, CASC, NSB, CCC, ...

• Expanding OCI to have research mission, positioning as a base for CS&E
  • No other Dir or Office can do this
  • Should also catalyze interdisciplinary app teams
  • Supporting computational scientists and curriculum
Need Help!

- Cooperation with agencies
  - Joint advisory with DOE on areas to develop together
  - EU, Asia...
  - NIH?

- Broadening participation

- Industry participation
  - Partnerships for strength, sustainability
  - Data, tools, software, WF development, clouds, HPC, etc

- Future plans
  - Task forces for programs in 18 months
  - New vision document in 3-4 years
Need Help!

- Cybersecurity and Education must be major strengths, both for us and to garner support from Congress. How to formulate programs that generate more funding to do this?
Summary of Top Messages

- We must fill out the CI Vision with broader portfolio
  - How to do this?
  - We have critical budget and staff shortage

- We want to expand the mission of OCI to also become the base for computational science and engineering research and education
  - Includes WF development as critical components

- We should be effective catalyst for new Grand Challenge Communities for complex problem solving in 21st century science & engineering

- Need help in achieving these goals