U.S. Department of Energy's

***Office of Science***

# Perspectives on SciDAC

## (Scientific Discovery through Advanced Computing)

**Kenneth John Roche**

Computer Science and Mathematics Division

Center for Computational Sciences

Oak Ridge National Laboratory

Oak Ridge, Tennessee

**Lattice QCD AHM, Brookhaven National Laboratory**

**26 March, 2004**

# (soft) Outline:

- Context / perspective

---

- Computational science
- Metrics
- Systems and Example
- Close
- QCD project

U.S. Department of Energy's

# *Office of Science*

**U.S. scientific community suffers from a lack of computational:**

- **Capacity** **(weak / strong scaling)**
- **Capability**

**The top two priorities of the Office of Science:**

1. **(FES) ITER**
2. **(ASCR) UltraScale Scientific Computing Capacity**

**The Office of Science must:**

**Prepare the critical applications and software for the future computing capacity**

# Perspectives …

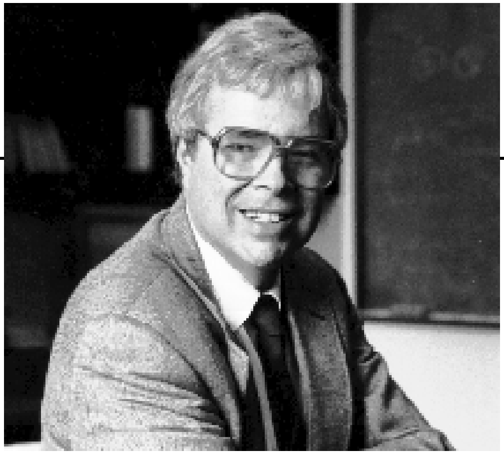
- **It remains unclear how to produce reliable and usable petascale systems from any perspective even though experts may project by Moore's Law that such systems may be delivered in 2010-2015.**

- **The enterprise of building large-scale systems has been left primarily to the computer scientists and engineers. Building faster and larger systems because it can be done does not necessarily advance science.**

- **The behavior of most of the nation's critical or fundamental computational science projects on existing DOE state-of-the-art systems is largely unknown. (where it is known –it is mostly bad)**

- **How progress should be measured and the identification of the barriers inherent in the process of making progress in the science missions needs to be considered.**

**Nobel Laureate Kenneth G. Wilson**

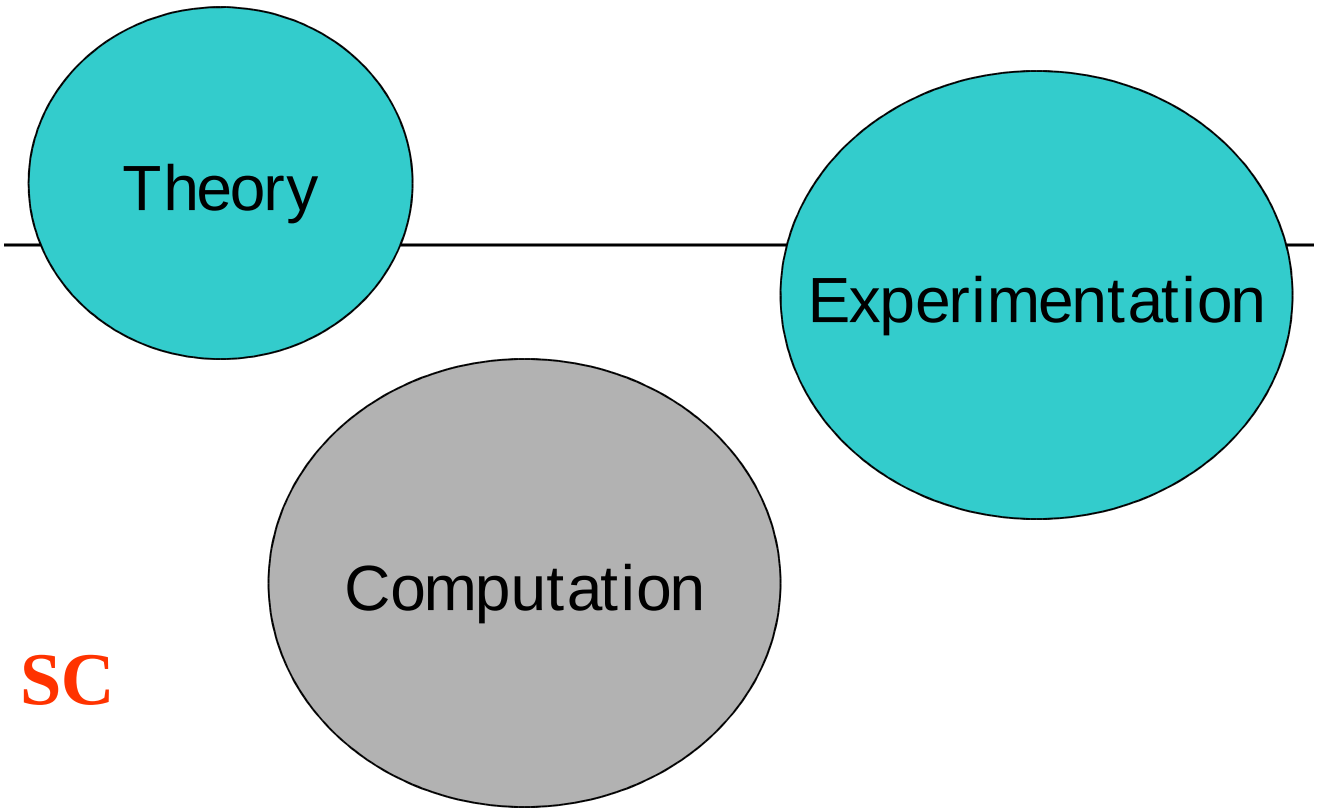


**Grand Challenge Problems (1980s - 1990s)**:

- questions to which many scientists and engineers would like to know answers
- the questions are difficult; we don't know how to do them right now
- the questions can likely be done by computer, but current computers are not "powerful" enough
- the questions have a solution (how is it known when a solution is in hand? metrics were not well defined and still are not today)

U.S. Department of Energy's

# *Office of Science*




## The 3 Pillars of DOE SC

# U.S. Department of Energy's

# *Office of Science*

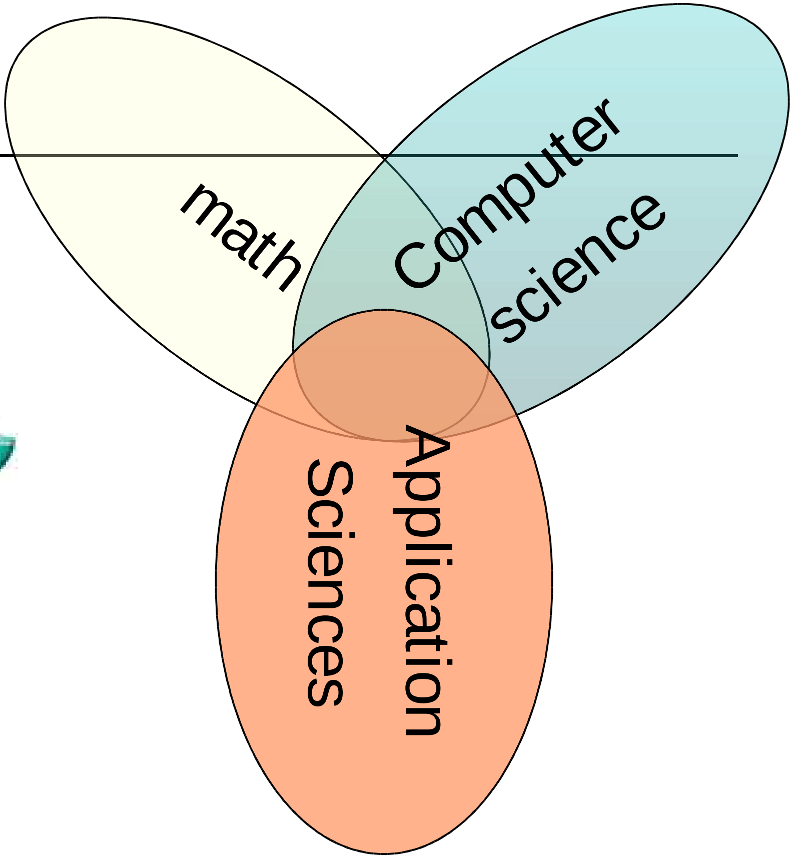
- **Build on past successes**
- **Learn from past failures**

**SciDAC**

Scientific Discovery through Advanced Computing

May 2000 ++

- **Science and research**
- **Education** **(where does this fit?)**
- **Outreach**
- **Measures and metrics**



# Define and Measure Science Based Metrics

**Applications:**

DOE SC missions

**Systems:**

DOE SC Systems
ASCI Q
Red Storm
Earth Simulator
Blue Gene L

Our world, its inhabitants

Scientific / security applications

Algorithms / simulations

(numerical) software

"standard" languages

compilers

Isa, architectures

(SMP, vector, cluster)

hardware / technology

Observational

Verification / testing

Can we measure time? Are the clocks reliable? Can we spy on the hardware? Do we know what to expect from the applications? Space ^ Time complexity?

## Application Performance vs. Machine fluctuations

-Naïve cross-sections of large multi-dimensional parameter spaces

-Computational noise

-Memory issues, instruction mix, floating point ops

## Models

-it is necessary to capture the architecture / application mapping and to provide predictive power

U.S. Department of Energy's

# Office of Science

SciDAC
Scientific Discovery through Advanced Computing

**~24 funded projects from each SC program that are best addressed through advances in computational science:**

**$20M** {
- **Biological and Environmental Research** (climate)
- **Basic Energy Sciences** (combustion, chemistry projects)
- **Fusion Energy Sciences** (Tokamak, plasmas)

**$40M** {
- **High Energy and Nuclear Physics** (astrophysics, qcd)
- **Advanced Scientific Computing Research** (*)

(*)

**Computer Science *Integrated Software Infrastructure Centers* (4)**

**Mathematics *Integrated Software Infrastructure Centers* (3)**

***Collaboratories* –virtual / distributed laboratories for**

# Office of Science Computing Today

NERSC: IBM/RS6000 SP Power 3

CRAY-X1 at ORNL

ORNL:
IBM p690 Power 4

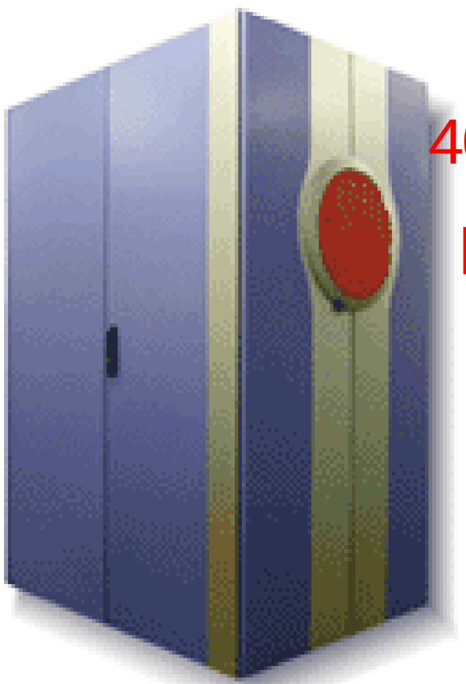
PNNL: HP Integrity

ORNL: SGI Altix

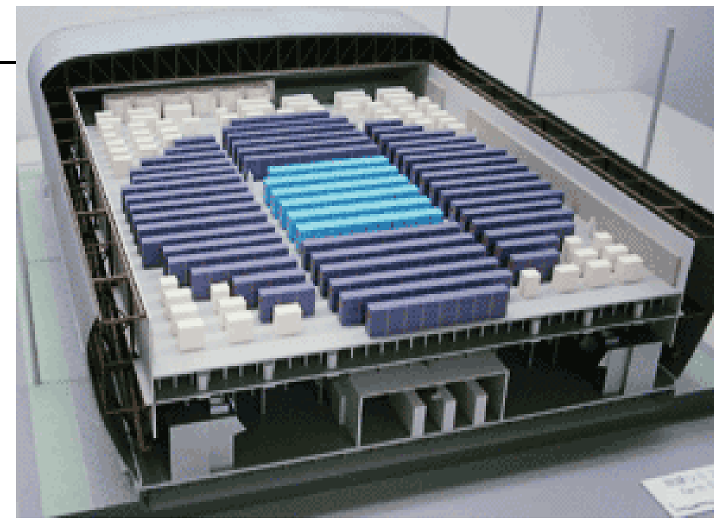
## Combination of all DOE SC Systems on previous slide:

~30TF peak performance today

## NEC Earth Simulator in Japan:



40TF peak performance today



NEC Earth Simulator Model

**N.B. -Performance alone does not account for science output.**

U.S. Department of Energy's

# Office of Science

## (BER) Climate Modeling

- 10 years ago –atmospheric models only

- today -ocean, sea ice, fixed land have been added
  w/ microscience: chemistry, geology, biology

- tomorrow –fully coupled models with full chemistry
  and dynamic vegetation land models,
  better resolution, higher productivity

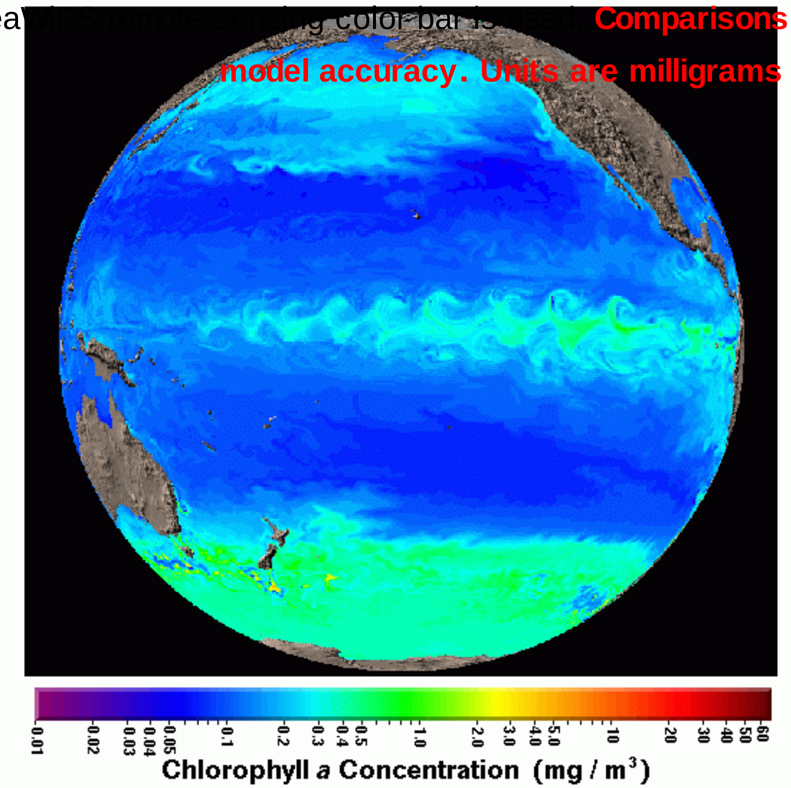
**Science / computation metric**

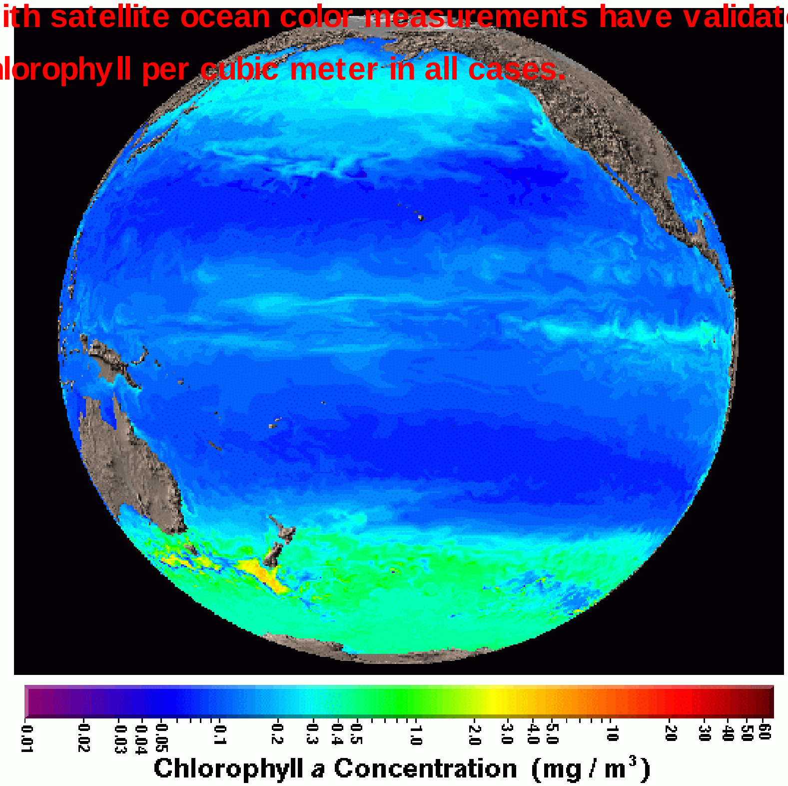
- simulated years per day of computation
  over ensemble input instances (* km^2)

# SciDAC: Ocean Biogeochemistry in the Parallel Ocean Program

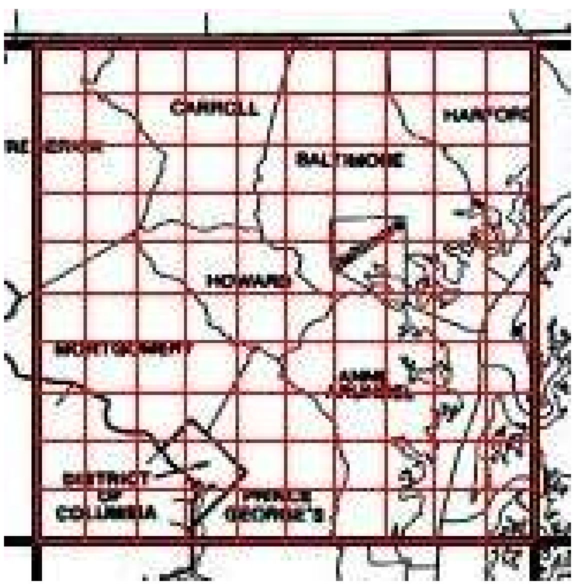
Shaoping Chu, Los Alamos National Laboratory

Two snapshots show surface chlorophyll distributions simulated with the biogeochemical version of the Parallel Ocean Program (POP) for conditions in late 1996 (a La Niña year) and late 1997 (El Niño conditions). Biological activity is intense across the equatorial Pacific during La Niña (Dec 96), extending almost from the primary Peruvian upwelling zone to New Guinea. The warm pool then shifts eastward, shutting off the upwelling of nutrient-rich cold water. Plankton growth slows and the chlorophyll peak gradually recedes towards the east (Dec 97). The standard SeaWiFS remote-sensing color bar is used. **Comparisons with satellite ocean color measurements have validated model accuracy. Units are milligrams chlorophyll per cubic meter in all cases.**



- The Earth Simulator works on a grid 10 km on a side for climate models, while U.S. computers do no better than 100 km on a side for comparable runs.




- U.S. simulations average over microclimates —mountains and coastal effects, river flow, cloud and storm systems, or hurricane storms. (adaptive mesh refinement, AMR)

- Averaging means that our models cannot credibly predict large scale fluctuations in climate change, critical for long-term drought and flood predictions.

U.S. Department of Energy's

# *Office of Science*

**Computing at the ultrascale**

**Demand driven by sciences:**

- accelerator science
- fusion science
- molecular and biological science
- climate
- combustion processes
- nuclear physics

**We are not prepared today. There is a great need for**:

- novel architectures and studies of existing ones
- program and application synthesis
- execution environments, software, compilers …etc

# Lattice QCD

- 3 quark propagators and their associated calculations

---

- compute qq matrix elements – test the Standard Model
  -ensembles of fields as inputs to build error bars

- two major codes (C / C++)
  - Can qcd run / optimize on other architectures?

- as cost per megaflop goes down and speed goes up, (when) will clusters (other) be better than QCDOC?

Lattice spacing / discrete continuum …