“The first great scientific breakthrough of the 21st century – the decoding of the human genome announced in February 2001 – was a triumph of large-scale computational science. When the Department of Energy (DOE) and the National Institutes of Health (NIH) launched the Human Genome Project in 1990, the most powerful computers were 100,000 times slower than today’s high-end machines; private citizens using networks could send data at only 9600 baud; and many geneticists performed their calculations by hand….it was expected to take decades.”

---Report to the President, June 2005, “Computational Science: Ensuring America’s Competitiveness”

This validates an additional way of “doing science”
A New Way of Discovery

**Discovery of the SASI**
*(Standing Accretion Shock Instability)*

- Wholly computational discovery
- Instability serves to move the shock outward explosion?
- Provides ‘natural’ explanation for neutron star spin up


How To “Do Science”

The four methods of “doing modern science”

- Observational Science
- Experimental Science
- Theoretical Science
- Computational Science
Scientific Method Process (1/2)

- Research
- Working Hypothesis
- Design
- Experiment
- Interpretable
- Data and Results
- Collect
- Conduct
- Real World Question

Scientific Method Process (2/2)

- First described ~400 years ago
- Is not constant – evolves as a result of technology
  - Peer review is a result of print
  - Repeatability of experiments is a result of peer review and collaboration (societies, not just letters)
  - Statistical sampling is due to advancements in mathematics
  - Etc.
- The impact of computing is only now being realized
“The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be solvable.”
--Paul Dirac, Royal Academy, London, 1929

“It is nice to know the computer understands the problem, but I would like to understand it too.”
--Eugene Wigner (when confronted with the computer generated results of a quantum mechanics calculation)

What is Computational Science? (1/5)

- Computational science is the integration of computing technology into scientific research
- It is the application of computer simulation and other computational methods to the solution of scientific problems and the understanding of scientific phenomenon
- Computing becomes a “full partner” in scientific discovery
- It is not to be confused with computer science which is the study of topics related to computers and information processing
What is Computational Science? (3/5)

Computational science seeks to gain an understanding of scientific processes through the use of mathematical methods on computers.
What is Computational Science? (4/5)

- Used to:
  - Perform experiments that might be too dangerous to perform in a lab
  - Perform experiments that happen too quickly or too slowly
  - Perform experiments that might be too expensive
  - Perform experiments that are only solvable using computational approaches
  - Visualize phenomenon in the past, present, or future
  - Perform “what-if” experiments
  - Data mine through huge datasets
  - Etc., etc.

What is Computational Science? (5/5)

“Computational Science was built on the vision that computers would represent a virtual laboratory where one could explore new concepts from simulations and comparison of these with experimental data.”

---Geoffrey Fox, Indiana University

Analyze - Predict
Data Assimilation
Information Simulation
Information Technology
Datamining
Computational Science

**Our Vision:**
*The Three Pillars of Scientific Discovery*

- **Theory**
- **Experiment**
- **Simulation**

*Std Model = Standard Model of Particles and Interactions*
*FSP = Fusion Simulation Project*
*LOCD = Lattice Quantum Chromodynamics*

(US Dept. of Energy, Office of Science)
A Computational science investigation should include:

- An application - a scientific problem of interest and the components of that problem that we wish to study and/or include.
- Algorithm - the numerical/mathematical representation of that problem, including any numerical methods or recipes used to solve the algorithm.
- Architecture – a computing platform and software tool(s) used to compute a solution set for the algorithm.

---

**Computational Science Process**

- **Simplify** 
- **Represent** 
- **Interpret** 
- **Translate** 
- **Simulate** 
- **Results and Conclusions** 
- **Computational Model** 
- **Mathematical Model** 
- **Working Model** 
- **Real World Model**
The Modeling Process

- **Modeling** is the application of methods to analyze complex real-world problems in order to make predictions about what might happen with various actions.

- A system exhibits *probabilistic* or *stochastic behavior* if an element of chance exists. Otherwise, it exhibits *deterministic behavior*. A **probabilistic or stochastic model** exhibits random effects, while a **deterministic model** does not.

- A **static model** does not consider time, while a **dynamic model** changes with time.

- In a **continuous model**, time changes continuously, while in a **discrete model** time changes in incremental steps.

(Ref: Shiflet & Shiflet)

Major Approaches to Computational Science Problems

- **System dynamics models** provide global views of major systems that change with time (e.g., equation-based physics problems).

- **Cellular automaton simulations (finite element)** provide local views of individuals affecting individuals. The world under consideration consists of a rectangular grid of cells, and each cell has a state that can change with time according to rules (e.g., visualization of lattice gauge QCD).
Real World Problem

Identify *Real-World Problem*:
- Perform background research, workable problem
- Conduct investigations (Labs)
- Select computational tool
- *Understand current activity and predict future behavior*

Working Model

Simplify → *Working Model*:
Identify and select factors to describe important aspects of *Real World Problem*; determine those factors that can be neglected.

- State simplifying assumptions
- Determine governing principles, physical laws
- Identify model variables and inter-relationships
Mathematical Model

Represent → Mathematical Model:
Express the Working Model in mathematical terms; write down mathematical equations or an algorithm whose solution describes the Working Model.

In general, the success of a mathematical model depends on how easy it is to use and how accurately it predicts.

Computational Model

Translate → Computational Model:
- Change Mathematical Model into a form suitable for computational solution.
- Computational models include tool-specifics.
Results/Conclusions

Simulate → Results/Conclusions:
Run “Computational Model” to obtain Results; draw Conclusions.

- Verify your computer program; use check cases; explore ranges of validity.
- Graphs, charts, and other visualization tools are useful in summarizing results and drawing conclusions.

Real World Problem

Interpret Conclusions:
Compare with Real World Problem behavior.

- If model results do not “agree” with physical reality or experimental data, reexamine the Working Model (relax assumptions) and repeat modeling steps.
- Often, the modeling process proceeds through several “cycles” until model is “acceptable.”
Example: Projectile Motion

Determine the motion of a projectile starting from a height $H$, above the ground with initial velocity $V$.

Working Model (1/2)

- Governing principles: $d = v*t$ and $v = a*t$
- Simplifying assumptions:
  - Gravity is the only force acting on the body
  - Flat earth
  - No drag (air resistance)
  - Model variables are $H$, $V$, $g$, $t$, $s$, and $v$
Working Model (2/2)

Form a discrete-in-time model to determine the position and velocity of the projectile above the ground at equally spaced times, \( t_0, t_2, t_4, \ldots, t_n \); e.g. \( t_0 = 0 \) sec, \( t_2 = 1 \) sec; \( t_2 = 2 \) sec, etc.

\[
\begin{array}{cccc}
  v_0 & v_1 & v_2 & \ldots & v_n \\
  s_0 & s_1 & s_2 & \ldots & s_n \\
  t_0 & t_1 & t_2 & \ldots & t_n \\
\end{array}
\]

Mathematical Model

Given an initial time, \( t_0 \), an initial height, \( H \), and an initial velocity, \( V \), generate the time history of heights, \( s_n \), and velocities, \( v \), by the formulas

\[
\begin{align*}
s_0 &= H, \quad v_0 = V \\
\Delta s_n &= v_{n-1} \Delta t, \quad s_n = s_{n-1} + \Delta s_n \\
\Delta v_n &= -g \Delta t, \quad v_n = v_{n-1} + \Delta v_n \\
\end{align*}
\]

\( n = 1, 2, \ldots \)
Computational Model

Pseudo-Code

Input
- $t_0$, initial time; $V$, initial velocity; $H$, initial height
- $g$, acceleration due to gravity; $\Delta t$, time step;
- $i_{\text{max}}$, maximum number of steps

Output
- $t_i$, t-value at time step $i$
- $s_i$, height at time $t_i$
- $v_i$, velocity at time $t_i$

Example: Projectile Motion

Initialize
- set $t_i = t_o = 0$; $v_i = v_o = V$; $s_i = s_o = H$
- print $t_i$, $s_i$, $v_i$

Time stepping: $i = 1, i_{\text{max}}$
- set $t_i = t_i + \Delta t$
- set $s_i = s_i + v_i \cdot \Delta t$
- set $v_i = v_i - g \cdot \Delta t$
- print $t_i$, $s_i$, $v_i$
- if ($s_i \leq 0$), quit
To create a more realistic model of a moving projectile, some of the simplifying assumptions could be dropped; e.g., incorporate drag - depends on shape of the projectile, is proportional to velocity.

Improve discrete model:
- Approximate velocities in the midpoint of time intervals instead of the beginning.
- Reduce the size of $\Delta t$. 
Most Scientific Computing Problems are More Complex Than Moving Projectiles

Scientific Simulation
Example – Electron-Gamma Showers (EGS)

- To simulate the interaction of particle beams of varying energies on fixed targets of various materials and geometries
- To study the resulting particle showers
- Simulations based upon known laws of physics and observed interactions (cross sections) between particles
- Allows “what-ifs” not possible or feasible in the laboratory

1 GeV Photon Shower

1 GeV photon shower in 10 r.l. of Pb (Ecut=0.1, Pcut=0.01 Mev)
EGS Applications

- Materials physics
- Radiation/health physics
- Radiation medicine
- Education
- Etc.

Finite Element and Lattice Methods
Finite Element Method (FEM)

- Many problems in engineering and applied science are governed by differential or integral equations
- The solutions to these equations would provide an exact, closed-form solution to the particular problem being studied
- However, complexities in the geometry, properties and in the boundary conditions that are seen in most real-world problems usually means that an exact solution cannot be obtained or obtained in a reasonable amount of time

Finite Element Method (2/2)

- In the FEM, a complex region defining a continuum is discretized into simple geometric shapes called elements
- The properties and the governing relationships are assumed over these elements and expressed mathematically in terms of unknown values at specific points in the elements called nodes
- An assembly process is used to link the individual elements to the given system. When the effects of loads and boundary conditions are considered, a set of linear or nonlinear algebraic equations is usually obtained
- Solution of these equations gives the approximate behavior of the continuum or system
Example – Lattice QCD Simulation (1/3)

- In quantum theories such as QCD, particles are represented by fields.
- To simulate the quark and gluon activities inside matter on a computer, physicists calculate the evolution of the fields on a four-dimensional lattice representing space and time.
- A typical lattice simulation that approximates a volume containing a proton might use a grid of 24x24x24 points in space evaluated over a sequence of 48 points in time.
- The values at the intersections of the lattice approximate the local strength of quark fields.
- The links between the points simulate the rubber bands—the strength of the gluon fields that carry energy and other properties of the strong force through space and time, manipulating the quark fields.

Example – Lattice QCD Simulation (2/3)

- At each step in time, the computer recalculates the field strengths at each point and link in space.
- The algorithm for a single point takes into account the changing fields at the eight nearest-neighbor points, representing the exchange of gluons in three directions of space—up and down; left and right; front and back—and the change of the fields over time—past and future.
Example – Lattice QCD Simulation (3/3)

Nuclear Fuel Rod Degradation
Advanced Test Reactor Simulation at INL (Idaho National Laboratory)

Simulation vs. CGI?

- [http://www.youtube.com/watch?v=_FIKonHQF8Y](http://www.youtube.com/watch?v=_FIKonHQF8Y)
Topics in Computational Science and Engineering

- High Performance Computing
- Data Mining
- Simulation
- Scientific Visualization
- Programming (Traditional and Symbolic Manipulation Tools)
- Collaboration systems/E-Science
- Analysis Packages
- Display and text processing systems

Data Mining
Data Mining

- Modern science is driven by data analysis like never before. We have an ability to collect and process data that is increasing exponentially!
- “…the analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner.”
- The extraction of useful patterns from data sources, e.g., databases, texts, web, image.

- Sequential pattern mining:
  A sequential rule: $A \rightarrow B$, says that event $A$ will be immediately followed by event $B$ with a certain confidence

- Deviation/anomaly/exception detection:
  discovering the most significant changes in data

- Data visualization: using graphical methods to show patterns in data

- High performance computing

- Bioinformatics
Why Data Mining

- Rapid computerization produces huge amounts of data
- How to make best use of data?
- A growing realization: knowledge discovered from data can be used for competitive advantage and to increase intelligence

Purpose of Data Mining

- Locating phenomenon from spatially, temporally, or logically related factors, each of which is defined at different levels of abstraction
- Content based searching and browsing
- Feature extraction
- Reduction in data volume
- Scientific analysis
- Searching for anomalies
Data Mining Fields

- Data mining is an emerging multi-disciplinary field:
  - Statistics
  - Machine learning
  - Databases
  - Visualization
  - Data warehousing
  - High-performance computing

Typical Data Mining Tasks

- Classification:
  - mining patterns that can classify future data into *known* classes

- Association rule mining:
  - mining any rule of the form \( X \rightarrow Y \), where \( X \) and \( Y \) are sets of data items

- Clustering:
  - identifying a set of *similar* groups in the data
Data Mining

- Define problem
- Data collection
- Data preparation
- Data modelling
- Interpretation/Evaluation
- Implement/Deploy model

Machine Learning

- “...the study of computer algorithms capable of learning to improve their performance on a task on the basis of their own experience.”
- Often this is “learning from data”.
- A sub-discipline of artificial intelligence, with large overlaps into statistics, pattern recognition, visualization, robotics, control, ...
Patterns (1/2)

- Patterns are the relationships and summaries derived through a data mining exercise.

- Patterns must be:
  - valid
  - novel
  - potentially useful
  - understandable
Patterns (2/2)

- Patterns are used for
  - prediction or classification
  - describing the existing data
  - segmenting the data (e.g., the market)
  - profiling the data (e.g., your customers)
  - Detection (e.g., intrusion, fault, anomaly)

Data (1/2)

- Data mining typically deals with data that have already been collected for some purpose other than data mining
- Data miners usually have no influence on data collection strategies
- Large bodies of data cause new problems: representation, storage, retrieval, analysis, ...
Data (2/2)

- Even with a very large data set, we are usually faced with just a sample from the population.
- Data exist in many types (continuous, nominal) and forms (credit card usage records, supermarket transactions, government statistics, text, images, medical records, human genome databases, molecular databases).

Data Modelling and the Scientific Method

- Data modelling plays an important role at several stages in the scientific process:
  1. Observe and explore interesting phenomena
  2. Generate hypotheses
  3. Formulate model to explain phenomena
  4. Test predictions made by the theory
  5. Modify theory and repeat (at 2 or 3)

- The explosion of data suggests that we need to (partially) automate numerous aspects of the scientific method.
Pattern recognition is a research area in which patterns in data are found, measured, and used to recognize, classify, and discover objects.

This is a catchall phrase that includes:
- Classification
- Clustering
- Data mining
- etc
Pattern Recognition Approaches

- **Statistical Pattern Recognition**
  - The data is reduced to vectors of real numbers that measure objects features. Statistical modeling is then used for recognition, classification, etc.

- **Structural Pattern Recognition**
  - The data is converted to a discrete and structured form such as trees, graphs, grammars, etc. Techniques related to computer science subjects such as graph matching and parsing are used.

Scientific Visualization
The Challenge

- Transform the *data* into *information* (understanding, insight) thus making it useful to people.
- Support specific tasks
- Improve performance as compared to existing mechanisms

Information Visualization

- Provide tools that present data in a way to help people understand and gain insight from it
- Cliches
  - “Seeing is believing”
  - “A picture is worth a thousand words”

“The use of computer-supported, interactive, visual representations of abstract data to amplify cognition.”
**Main Idea**

- **Visuals help us think**
  - Provide a frame of reference, a temporary storage area

- **External cognition**
  - Role of external world in thinking and reason
  - Multiplication exercise

**Information Visualization**

- **What is “information”?**
  - Items, entities, things which do not have a direct physical correspondence
  - Examples: baseball statistics, stock trends, connections between criminals, car attributes...

- **Scientific Visualization**
  - Primarily relates to and represents something physical or geometric
  - Examples
    - Air flow over a wing
    - Stresses on a girder
    - Weather over Pennsylvania
Key Attributes

- **Scale**
  - Challenge often arises when data sets become very large

- **Interactivity**
  - Want to show multiple different perspectives on the data

- **Tasks**
  - Want to support specific tasks – not just to create a cool demo
  - Support discovery, decision making, explanation

Visualization Success Stories

Illustration of John Snow’s deduction that a cholera epidemic was caused by a bad water pump, circa 1854.

Dots indicate location of deaths.

Napoleon’s March

Minard graphic

size of army direction
direction

latitude longitude
temperature date

What is Scientific Visualization?

- It is a transformation of abstract data into readily-comprehensible images
- It relies on human cognitive processes
Geometric and Visual Computing Areas

- Computer Aided Geometric Design (CAGD): Curves/surfaces
- Solid Modeling: Representations and Algorithms for solids
- Computational Geometry: Provably efficient algorithms
- Computer-Aided Design (CAD): Automation of Shape Design
- Computer-Aided Manufacturing (CAM): NC Machining
- Finite Element Meshing (FEM): Construction and simulation

New Topics in Computational Science and Engineering

- “Collaboratories” and scientific workspaces of the future
- Scientific research in virtual worlds
- Exascale Science
- Web Science
Programming and Mathematical Solvers

Popular Symbolic/Mathematical Software Packages (1/2)

- **Mathematica**
  - Advantages - premier all-purpose mathematical software package; It integrates swift and accurate symbolic and numerical calculation, all-purpose graphics, and a powerful programming language
  - Disadvantages – Steep learning curve, expensive; premier all-purpose mathematical software package. It integrates swift and accurate symbolic and numerical calculation, all-purpose graphics, and a powerful programming language

- **Matlab**
  - Advantages - combines efficient computation, visualization and programming for linear-algebraic technical work and other mathematical areas
  - Disadvantages - Not for analytical/symbolic math
Popular Symbolic/Mathematical Software Packages (2/2)

- **Maple**
  - Advantages - powerful analytical and mathematical software which does the same sorts of things that Mathematica does, with similar high quality; programming language is procedural -- like C or Fortran or Basic -- although it has a few functional programming constructs.
  - Disadvantages - Worksheet interface/typesetting not as developed as Mathematica's, but it is less expensive

- **IDL (Interactive Data Language)**
  - Advantages - excels at processing real-world data, especially graphics, and has a reasonably simple syntax, especially for those familiar with Fortran or C; makes it as easy as possible to read in data from files of numerous scientific data formats
  - Disadvantages - Does not do symbolic math

Scientific Workspace of the Future (SWOF)
Ad Hoc Collaboration

“Sue, there’s a meeting with CERN you might want to participate in…”

Distance Learning

“...if over time, the galaxy stretches like...”
Distributed Exploratory Analysis

"so in effect, as these neuron pathways form..."

Interactive Scientific Computing

"let's re-run the simulation with this molecule over here..."
Online virtual worlds have great potential as sites for research in the social, behavioral, and economic sciences, as well as in human-centered computer science. A number of research methodologies are being explored, including formal experimentation, observational ethnography, and quantitative analysis of economic markets or social networks.
What is a Computational Scientist?

Computational Scientist Degree: Learning Outcomes and Objectives

July 21, 1995

Students completing this program will:

- Be prepared to join and contribute to interdisciplinary teams in scientific research or industry.
- Have a firm understanding of the physical theory and computational strategies used in the student's major discipline. This includes the ability to select the most appropriate algorithm for the problem.
- Understand the computational methods used in industry standard software packages in the student's major discipline.
- Have a general knowledge of the computational methods used in other areas and their potential application to the student's major discipline.
- Have a solid foundation in numerical methods and a general understanding of statistical data analysis.
- Understand computer architectures used in modern scientific computation.
- Be familiar with currently used programming languages for scientific computing (C, FORTRAN 90 and FORTRAN 77) including optimization techniques.
- Have an in-depth understanding of parallel algorithm development and other methods and tools required for high performance computing.
- Be able to apply the tools of visualization to the analysis of scientific data and produce images/animation that extract the essential information and display it so the relevant context is clear.
- Have demonstrated ability in creating useful tools for computational scientific research. The tools needed are those that are powerful and flexible enough to solve a variety of real problems, safe to run and optimized enough to handle large problems.
Engineering and Information Technology

Mathematics and Engineering Analysis

M&CT’s Mathematics and Engineering Analysis unit is Boeing’s central math resource -- about 100 mathematicians and scientists with extensive experience in industrial R&D and problem solving. The organization’s experts are skilled in applied statistics, computational science, computational linear algebra and differential equations, geometry and optimization, mathematical modeling, operations research, and modeling and simulations technology. They work with engineers and other technologists on the design and manufacturing of commercial airplanes and military and space systems. They also conduct applied R&D and do consulting and software development for non-Boeing customers.

Applied Statistics

The Applied Statistics group is devoted to a combination of R&D and consulting in support of aerospace industry needs. Their work ranges through most standard areas of industrial statistics: survey design and analysis, time series and forecasting, probability and stochastic models, statistical quality control, Design of Experiments (DOE), and reliability.

The aerospace industry, rich in diverse and complex problems, has led us to other areas including image processing, spatial statistics, industrial epidemiology, data mining, design and analysis of computer experiments, Bayesian information retrieval, financial modeling, stochastic optimization, and statistical tolerancing.

Computational Science

The Computational Science group provides software and support to Boeing engineers and mathematicians engaged in solving the mathematical problems that arise while using computers to design complex structures and aircraft and simulate airflow.

The group’s areas of expertise include:
- Application of parallel computing techniques to mathematical computation
- Web-enabled delivery and analysis tools
- Advanced engineering computing techniques
- C, C++, Fortran, Java, and Visual Basic programming languages
- Mathematical software

Thank You

Questions, Comments?

bebo@slac.stanford.edu