CS590C: Introduction to Computational Science
(a.k.a CS501)
http://www.cs/homes/ayg/CS590C/www/cs590c.html

Ananth Grama
164D, Computer Science Building, 494 6964
ayg@cs.purdue.edu
http://www.cs.purdue.edu/people/ayg
Office Hours: Monday 11:30 AM - 12:30 PM,
Wednesday 11:30 AM - 12:30 PM,
(By appointment)

What is Computational Science?

- ❏ The science of using computers for problem solving.
- ❏ Diverse problems such as protein synthesis, lattice structures, weather modeling, geology.
- ❏ Extend traditional definition to include problems such as stock market analysis, economics, psychology.

How is it different from computer science?

- ❏ The focus in computer science is the computer itself. The issues dealt with are related to speeding up the computers;
- ❏ Making them easier to use;
- ❏ Finding fast algorithms for core problems (sorting, searching, etc.).

What makes computational science special, difficult?

- ❏ Interdisciplinary nature: people from computer science, mathematics, pure sciences, and others have to work together.

- ❏ The core techniques in problem solving draw heavily on mathematical sciences.

- ❏ The computational requirements in many cases require the use of High Performance Computing
(Parallel processing poses an added dimension of difficulty)

- ❏ The diverse background of the community makes it difficult to disseminate results.

Scientific Computing, High Performance Computing, Grand Challenge Problems

- ❏ Many problems in computational science require the use of high performance computing

- ❏ Grand challenge problems represent a subset of problems in scientific computing.

What is the HPCC initiative?

- ❏ HPCC (High PErformance Computing and Communications) initiative is a national initiative to develop fast computers and communications.

The goals of the Federal HPCC Program are to:

- ❏ Extend U.S. technological leadership in high performance computing and computer communications

- ❏ Provide wide dissemination and application of the technologies to speed the pace of innovation and to improve the national economic competitiveness, national security, education, health care, and the global environment

- ❏ Provide key enabling technologies for the National Information Infrastructure (NII) and demonstrate select NII application

In FY 1996 the HPCC Program is a $1.1 billion effort!

What are Grand Challenge Problems?

“.. fundamental problems in science or engineering with potentially broad social, political, and/or scientific impact, that could be advanced by applying high performance computing resources.”

Identified thrust areas for grand challenge problems:

- electronic structure of materials
- turbulence
- genome sequencing and structural biology
- global weather and climate
- speech and language
- drug design
- vision and cognition
- fundamental properties of matter (Quantum Chromodynamics)
- pollution dispersion
- .. and others.

Evolution of hardware.

❑ Parallel computers offer magnitudes higher performance enabling solution of hitherto unrealizable problems.

❑ The microprocessor revolution puts tremendous performance on the desktop. In addition to problem solving, this facilitates visualization and representation of complex data.

❑ Advances in network technology enable the use of clusters of workstations as parallel platforms.

❑ Advances in symmetric multiprocessing put the power of expensive vector processors (the Crays) in cheap high end workstations (e.g. the SGI Power Challenge)

Evolution of Hardware (the flip side)

- ❏ Memory speeds have not kept up with processor speeds (deep cache hierarchies). This causes a large gap in peak and realizable performance of conventional computers. Hand-tuning necessary even for uniprocessors.

- ❏ Parallel computers are still hard to program (push for standardization)

- ❏ Complex architectures require more complicated algorithms and programs.

- ❏ Solution: Identify key computational kernels, use handcoded libraries for these.

- ❏ Removes the burden of tuning programs from application engineers.

Evolution of Software.

- ❏ Problem solving environments: Integrated visualization, and problem solving (EllPACK).
- ❏ Tools for mathematical (numeric and symbolic manipulation) Matlab, Mathematica, Maxima.
- ❏ Availability of standardized libraries for a variety of problems (BLAS, LAPACK, LINPACK, OSL etc.).
- ❏ Availability of a variety of visualization tools (PBM, Utah Raster Toolkit, ..)
- ❏ Parallel libraries also under development (ScaLAPACK).
- ❏ Standardization of lower level constructs:
- ❏ For parallel machines: Parallel Virtual Machine, Message Passing Interface (for message passing parallel computers), High Performance Fortran, Fortran90 (data parallel programming), Pthreads (shared memory programming)
- ❏ The revolution of open systems.

Evolution of Algorithms

❏ Faster machines pose additional challenges for existing algorithms (finer finite element meshes, impact of finite machine precision on convergence, etc.).

❏ New problems drive development of newer algorithms (hierarchical methods for n-body methods, linear system solvers such as GMRES, Fast Fourier Transforms...)

Focus on

- ❏ error analysis
- ❏ exact solutions and expansions
- ❏ uniqueness proofs
- ❏ theorems

A fast algorithm lets you do on a uniprocessor what a dumb algorithm cannot even do, or takes a very fast and expensive computer to do. Algorithm development is of Prime importance. Perhaps a lot more so than any other area.

Role of Visualization.

- ❏ Typical problems generate large amounts of data. Effective visualization tools are the only means of interpreting this data.

Need for 3-D visualization tools

- ❏ High speed communication from the computing device (typically between the supercomputing platform and the visualization device).

- ❏ Problems in virtual reality and simulation have critical reliance on visualization technology.

Role of networking media:

- ❏ Communication is critical!
- ❏ I/O devices such as disks, visualization devices must have very fast access to the computing platform.
- ❏ Often, computers cooperate to solve a single problem. These computers must be able to communicate among themselves extremely fast.
- ❏ Since supercomputers may be geographically removed from the actual user, there is a need for wide area networking.
- ❏ Fast networks can be used to interconnect workstations and use them in clusters for fast problem solving.

Emerging Trends in Computational Science

❏ Development of Problem Solving Environments (PSEs)

❏ A PSE is an integrated hardware-software system providing all the computational facilities required for solving a target class of applications.

❏ It solves the problem by communicating in the users terms.

❏ Greatly increased hardware performance (TeraFLOP achievable, current goal is a PetaFLOPS class computer).

❏ Greatly increased software quality (integrated libraries and standardized interfaces allow the programmer to write code at a much higher level).

❏ Integrating knowledge into PSEs.

❏ Scientific Visualization emerges as a new technology (higher resolution, dimensionality, multimedia support)

Design cycle of an application program