1. Distribution and Toxicity Examples
2. Computational Infrastructures

February 28, 2006

Guha Jayachandran
(guha@stanford.edu), CS379A

# Congo Red

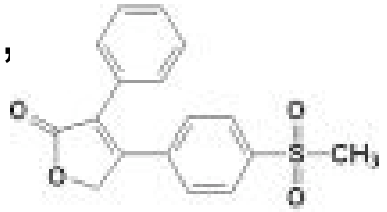
- Red dye

- Not used for cotton or paper dyeing anymore because toxic

- Known to bind to beta amyloid oligomers and fibrils in vitro, but besides being toxic, can’t cross blood-brain barrier

# “Vioxx is Here” No More

- Nonsteroidal anti-inflammatory drug for arthritis and pain
- Less gastrointestinal side effects than aspirin, ibuprofen, etc.
  - Reason: Only inhibits only COX-2 and not COX-1




- Widely marketed and prescribed
  - Sales of over $2 billion per year
- Voluntarily withdrawn in 2004 due to increased risk of heart attack and stroke
  - Allegations of scientific misconduct and much litigation ongoing. Estimated that Vioxx caused ~100 K heart attacks in 5 years.
  - Increased cardiac risk may be due to metabolites formed when compound becomes ionized
  - FDA advisory panel has recommended allowing resumption of sales saying benefit outweighs risk

# Computational Infrastructures

Guha Jayachandran
(guha@stanford.edu), CS379A

# Themes

- Moore's Law
- Parallelism
- Clusters vs. supercomputers
- Special purpose vs. general purpose

# Metrics

- Standard benchmarks are used to measure FLOPS (floating point operations per second)
  - Small differences can be misleading because people tune to benchmark, but fine for scale
- Power
- Cost per FLOP and power consumption per FLOP should also be considered
- Memory
- Networking
- Storage
- Flexibility (qualitative)

# Parallelism

- Within processor
  - Cell processor
- Within computer
  - Multiprocessor, CPU with GPU
- Clusters/grids
  - Very widely adopted over past few years
- Global distributed computing

# Processors

- Standard desktop computer CPUs (Pentium 4, AMD Athlon, etc.)
  - Few GFLOPS
  - Not increasing in speed as fast as before
  - SIMD helps (SSE, 3DNow, or AltiVec)
- GPUs
  - Couple hundred GFLOPS and speed growing faster than CPUs'
  - Small cache, stream programming
  - Floating point or less
- Cell processors just out
  - 300 GFLOPS (single precision)
  - One Power (PPE) plus 8 synergistic processing elements (SPEs)

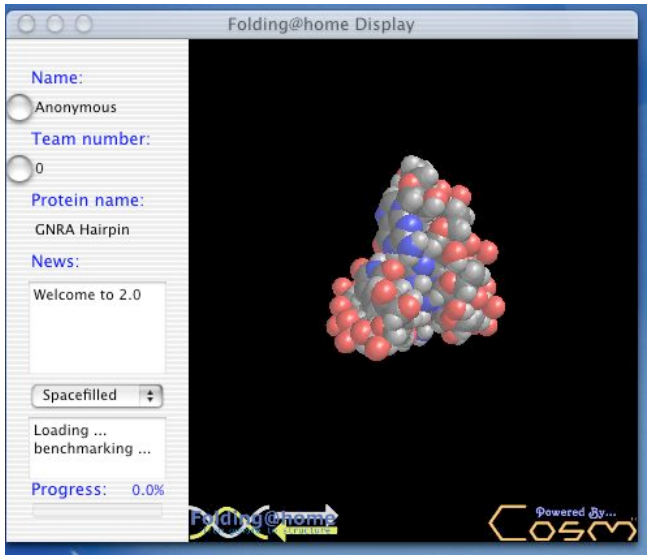
# Grid or Cluster Computing

- Very widely adopted over past few years
- Network together a number of computers
- Can operate on a single task through MPI but ideal for trivially parallelizable job
- Sources of computers
  - Dedicated computer's cluster
  - Grid of desktops (Novartis has all its employees' computers on a grid)
  - Utility computing (pay to use)

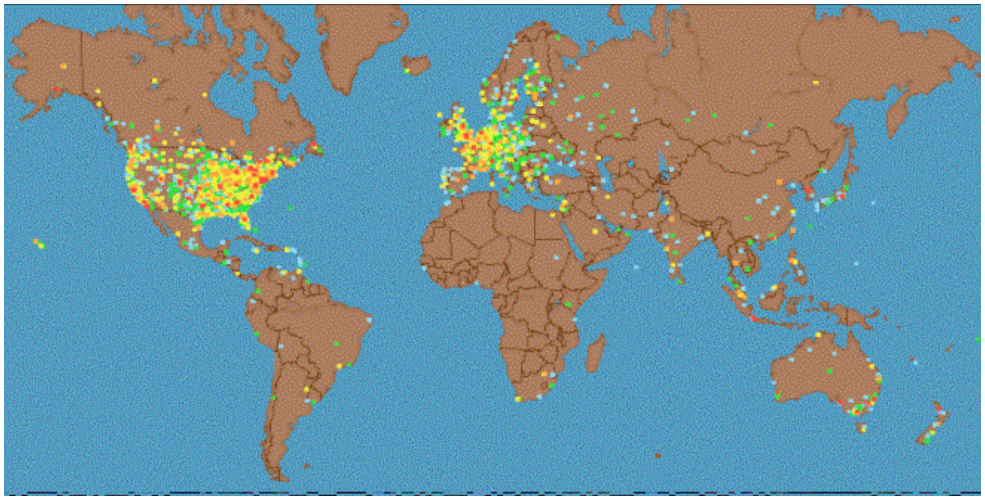
# Global Distributed Computing

- Individuals around the world download software which runs calculations assigned by central servers
- Folding@Home
  - ~200,000 active clients for 200 TFLOPS
  - Achieved first unbiased M.D. simulation of a protein folding
- BOINC tries to make setting up (or running) a project easier

# Folding@Home




Client program screenshot



~200,000 active clients worldwide

Earth's city lights from space

# Projects

(Outdated, from 2003)

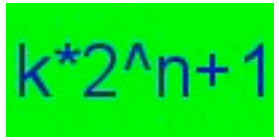















# Supercomputers

- Nice if you can get one
- Strength: Communication between processors
- IBM Blue Gene
  - Starting configuration priced over $1 M
  - BlueGene/L at Lawrence Livermore at top of Top 500 list right now
  - BlueGene/L - 280 TFLOPS
- Fujitsu BioServer
  - 1,920 low power (embedded) processors
  - Test system was used for Gromacs, CAChe, and MOPAC (typical computational chemistry programs)

# “Top 500” List

| Rank | Site | Computer | Processors | Year | $R_{max}$ | $R_{peak}$ |
|---|---|---|---|---|---|---|
| 1 | DOE/NNSA/LLNL<br>United States | BlueGene/L - eServer Blue Gene Solution<br>IBM | 131072 | 2005 | 280600 | 367000 |
| 2 | IBM Thomas J. Watson Research Center<br>United States | BGW - eServer Blue Gene Solution<br>IBM | 40960 | 2005 | 91290 | 114688 |
| 3 | DOE/NNSA/LLNL<br>United States | ASC Purple - eServer pSeries p5 575 1.9 GHz<br>IBM | 10240 | 2005 | 63390 | 77824 |
| 4 | NASA/Ames Research Center/NAS<br>United States | Columbia - SGI Altix 1.5 GHz, Voltaire Infiniband<br>SGI | 10160 | 2004 | 51870 | 60960 |
| 5 | Sandia National Laboratories<br>United States | Thunderbird - PowerEdge 1850, 3.6 GHz, Infiniband<br>Dell | 8000 | 2005 | 38270 | 64512 |
| 6 | Sandia National Laboratories<br>United States | Red Storm Cray XT3, 2.0 GHz<br>Cray Inc. | 10880 | 2005 | 36190 | 43520 |
| 7 | The Earth Simulator Center<br>Japan | Earth-Simulator<br>NEC | 5120 | 2002 | 35860 | 40960 |
| 8 | Barcelona Supercomputer Center<br>Spain | MareNostrum - JS20 Cluster, PPC 970, 2.2 GHz, Myrinet<br>IBM | 4800 | 2005 | 27910 | 42144 |
| 9 | ASTRON/University Groningen<br>Netherlands | Stella - eServer Blue Gene Solution<br>IBM | 12288 | 2005 | 27450 | 34406.4 |
| 10 | Oak Ridge National Laboratory<br>United States | Jaguar - Cray XT3, 2.4 GHz<br>Cray Inc. | 5200 | 2005 | 20527 | 24960 |
| 11 | Lawrence Livermore National Laboratory<br>United States | Thunder - Intel Itanium2 Tiger4 1.4GHz - Quadrics<br>California Digital Corporation | 4096 | 2004 | 19940 | 22938 |
| 12 | Computational Biology Research Center, AIST<br>Japan | Blue Protein - eServer Blue Gene Solution<br>IBM | 8192 | 2005 | 18200 | 22937.6 |

Nov. 2005 list, from www.top500.org

# Cluster or Supercomputer

- Supercomputer needed for tightly coupled computation
  - But often this is not needed or can be circumvented
- Clusters are more flexible and can be upgraded more easily
- CPUs in clusters can be quite good today—supercomputers without advantage in processor speed

# Special or General Purpose

- MDGRAPE an example of a specialized system for a problem domain
  - Also ASICS and certain supercomputers
- Specialized processors/computers may have immediate performance advantages
- General purpose gives more flexibility
- General will usually advance faster because many constituents for its development

# Cost and Power

| | $/GFLOP | W/GFLOP |
|---|---|---|
| MDGRAPE-3 | 15 | 0.2 |
| Pentium 4 | 400 | 14 |
| BlueGene/L | 140 | 6 |
| Earth Simulator | 8000 | 128 |

Taiji 2004

# Readings

- Building and managing production bioclusters (Dagdigian)
- Protein Explorer: A Petaflops Special-Purpose Computer System for Molecular Dynamics Simulations (Taiji, et. al.)
- Introduction to the Cell multiprocessor (Kahle, et. al.)