Do More, Faster: Leveraging Computational Resources in Your Research

Dr. Dirk Colbry
Institute for Cyber Enabled Research
Adjunct Faculty, Electrical and Computer Engineering
Michigan State University

© 2014 Michigan State University Board of Trustees

Agenda

- Who am I
- Pillars of Science
- What is Advanced Computing Hardware?
- Common classes of problems
- Steps to High Performance
I am an engineer

- Undergraduate Mechanical Engineering degree from Georgia Tech
- 3 years as a Mechanical Engineer for Delta Airlines in Atlanta
- 2 years as a Robotics Engineer for FANUC Robotics in Auburn Hills

I am a Student

- M.S.E. in Computer Science and Engineering, University of Michigan
  - Artificial Intelligence
  - Thesis on Temporal Bayesian Networks
- Ph.D. in Computer Science and Engineering, Michigan State University
  - Pattern Recognition and Image Processing
  - 3D Face Recognition
I am a Researcher

- Image Analysis in Research
- High Performance Computing
- Assistive Technologies

I am a computational consultant

- One-on-one consulting
- HPC Programming
- Proposal Writing
- Training and Education
- Outreach

- Reduce the “Mean time to Science”
Agenda

- Who am I
- **Pillars of Science**
- What is Advanced Computing Hardware?
- Common classes of problems
- Steps to High Performance
Agenda

- Who am I
- Pillars of Science
- **What is Advanced Computing Hardware?**
- Common classes of problems
- Steps to High Performance

What is Advanced Computing Hardware?

- Anything more advanced than your desktop
- Local resources
  - Lab, Department, Institution (iCER)
- National resources
  - NSF (XSEDE), DOE (Jaguar), Others
- Commercial Resources (cloud computing)
  - Amazon, Azure, Liquid Web, Others
Why use Advanced Computing Hardware?

- Science takes too long
- Computation runs out of memory
- Run out of disk space
- Need licensed software
- Need advanced interface (visualization)

HPC Systems

- Large Memory Nodes (up to 6TB!)
- GPU Accelerated cluster (K20, M1060)
- PHI Accelerated cluster (5110p)
- Over 600 nodes, 7000 computing cores
- Access to high throughput condor cluster
- 363TB high speed parallel scratch file space
- 50GB replicated file spaces
- Access to large open-source software stack and specialized bioinformatics VMs
Available Software

- Center Supported Development Software
  - Intel compilers, openmp, openmpi, mvapich, totalview, mkl, pathscale, gnu...
- Center Supported Research Software
  - MATLAB, R, fluent, abaqus, HEEDS, amber, blast, ls-dyna, starp...
- Customer Software
  - gromacs, cmake, cuda, imagemagick, java, openmm, siesta...

- For a more up to date list, see the documentation wiki:
  - http://wiki.hpcc.msu.edu/

What if I need help?

- Ask us!
- Local Workshops
  - Software carpentry
  - Introduction to Linux and HPCC
  - Advanced HPCC
- Remote Training
  - VSCSE – Virtual School for Computer Science Education
  - XSEDE training Workshops
XSEDE OpenMP Workshop
October 7, 2014

• **Who?** C and Fortran programmers

• **What?** Leave with a working knowledge of how to write scalable codes using OpenMP

www.icer.msu.edu/events

CYBERINFRASTRUCTURE DAYS 2014
OCTOBER 23 & 24

• Special guest speaker presentations
• Interactive workshops on a variety of topics
• Poster session showcasing CI-enabled research at MSU
• Resource fair featuring various CI resources available to MSU researchers
• Networking opportunities

http://www.vprgs.msu.edu/ci-days
CI Days Posters
Due, October 10, 2014

- Demonstrate a Direct connection to using Cyber-Infrastructure in your research
- In progress or Completed Research
- Cash Prizes!
- Networking opportunities

CI Days Training
Thursday, October 23, 2014

- **10:00am-1:00pm**
  - Introduction to Python
  - Introduction to HPCC at MSU
  - Programming with MATLAB
- **2:00pm – 5:00pm**
  - Advanced Python
  - Advanced Topics in HPCC
  - Optimizing your MATLAB Code

www.icer.msu.edu/events
Computational Chemoinformatics & Docking Studies

November 3-6, 2014

• Who? Interested in molecular design and macromolecule-small molecule interactions

• Hands on computational exercises
  Topics include:
  – Protein-ligand interactions
  – Protein structure prediction
  – Virtual screening using molecular docking
  – Common pitfalls in protein-ligand docking

Check our Website

++More Training Coming
  – Open MPI
  – Open ACC
  – Big Data

www.icer.msu.edu/events
What if I want more?

Agenda

- Who am I
- Pillars of Science
- What is Advanced Computing Hardware?
- **Common classes of problems**
- Steps to High Performance
What problems are we solving?

- Boundary Simulations
- Data Analysis
- Search

Boundary Simulations

- Typically System of PDE (Partial Differential equations)
  - Fluid dynamics
  - Finite element analysis
  - Molecular dynamics
  - Weather
  - Etc.
- Mathematically equivalent to inverse of a matrix
Data Analysis

• Computer vision tasks
• Some Bioinformatics
• Astrophysics
• Etc.

Video Provided by Dr. Fred Dyer

Search

• Genome sequencing
• Analytics
• Optimization
• Etc.

Evolution of an artificial organism that can move and forage for food, Dr. Nicolas Chaumont
 Agenda

• Who am I
• Pillars of Science
• What is Advanced Computing Hardware?
• Common classes problems
• Steps to High Performance

Steps in Using the HPCC

1. Get an account (ask your advisor)
   https://contact.icer.msu.edu/account
2. Install needed software (SSH, SCP, X11)
3. Transfer input files and source code
4. Compile/Test programs on a developer node
5. Write a submission script
6. Submit the job
7. Get your results and write a paper!!
Single Thread Jobs

- One CPU can only run one thing at a time. (sort of)
- CPUs are not getting that much faster.

Communication

- Shared Memory
  - Ex. OpenMP
- Shared Network
  - Ex. MPI
- Distributed Network
  - Ex. Map-Reduce
- Dedicated Accelerators
  - Ex. GPGPU and Phi
- Hybrid Systems
Pleasingly Parallel

MSU HTCondor Cluster

- Runs like a screen saver and scavenges CPU cycles:
  - Approximately 400 nodes
  - Approximately 800 cores
  - Windows XP
Loosely Coupled

Commodity Cluster

Rack
High Speed Network
Chassis
Nodes
Processors / Sockets
Cores
MSU Cluster 14

- General purpose base nodes
  - 20 cores, 64 gb
- 256gb nodes
- 256gb and 2xK20 GPU nodes
- 256gb and 2xPhi Card nodes

Shared Network Communication

- Commodity Cluster
- High speed network
- MPI
  - Message Passing Interface
  - Programming standard
  - Ex Libraries: OpenMPI, MPICH/MVAPICH
- Parallel File systems
  - Lustre

InfiniBand inside
Tightly Coupled

• Cores on a processor share the same memory
• OpenMP
• Fat nodes
  – 96 cores
  – 6TB of memory
GPU

- Cards used to render graphics on a computer
- Hundreds of cores
- Not very smart cores
- But, if you can make your research look like graphics rendering you may be able to run really fast!

Intel Xeon Phi

- Cross between CPU and GPU
- About 61 Pentium III cores
  - Less cores/slower than GPU
  - Easier to use than GPU
Summary of Hardware

• Pleasantly parallel
  – HTCondor
• Loosely Coupled
  – Commodity cluster
• Tightly Coupled
  – Fat Nodes
  – GPUs
  – Phi

Which approach is the best?

• Depends on what you are doing?
• Depends on how much communication you need.
• Depends on what hardware you have.
• Depends on how much time you have.
We are here to help

- [www.hpcc.msu.edu/contact](http://www.hpcc.msu.edu/contact)
  - Questions
  - Schedule Consultations
  - Code Reviews
  - Programming help
  - Hardware Purchasing
  - Help with Grants
  - Support for Grants