Chemistry Packages at CHPC

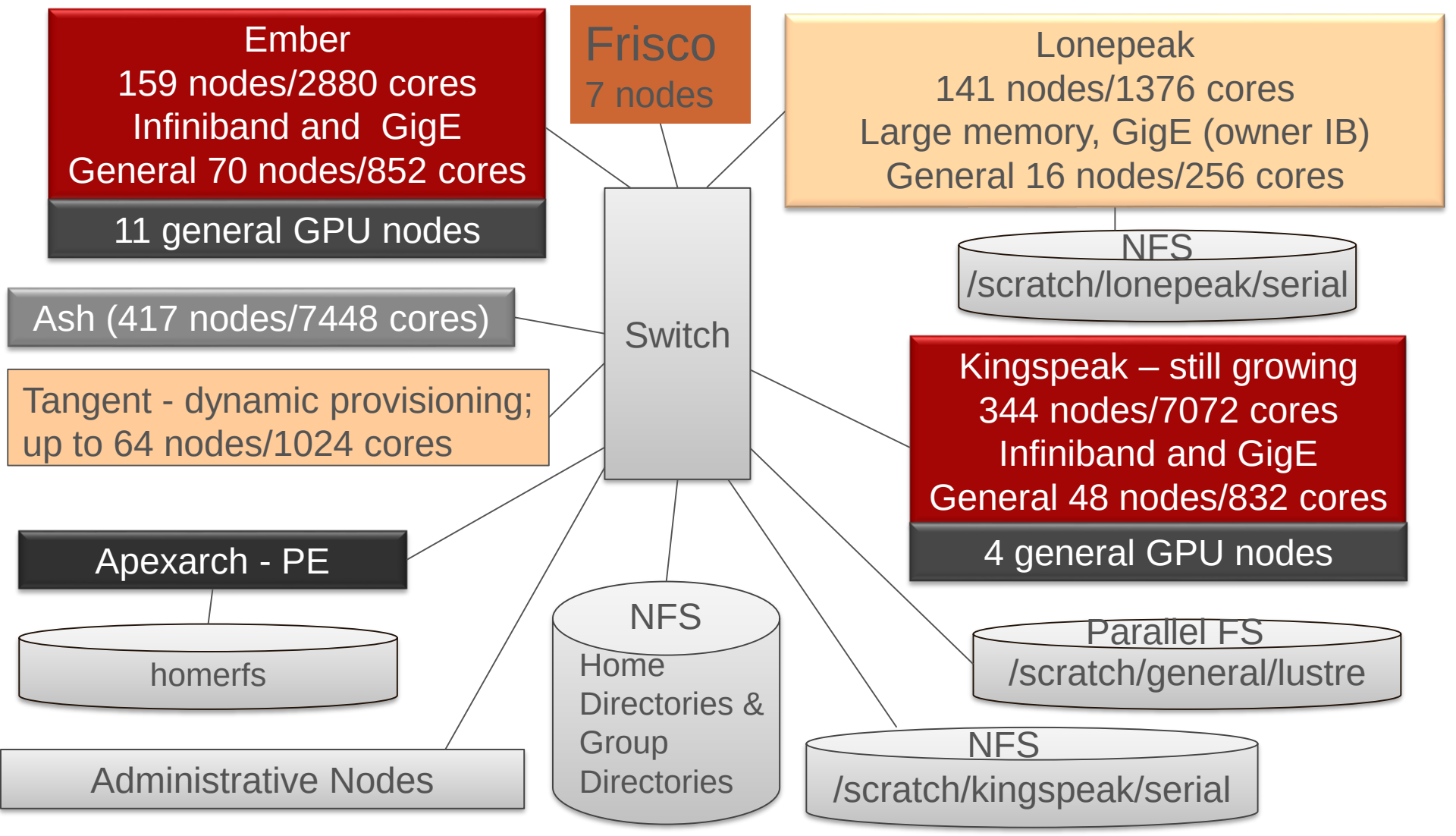
Anita M. Orendt
Center for High Performance Computing
anita.orendt@utah.edu
Fall 2016

Purpose of Presentation

- Identify the computational chemistry software and related tools currently available at CHPC
- Present brief overview of these packages
- Present how to access packages on CHPC
- Information on usage of Gaussian09

Survey about Presentation

- https://www.surveymonkey.com/r/8CKKY3C


Ember
159 nodes/2880 cores
Infiniband and GigE
General 70 nodes/852 cores
11 general GPU nodes
Frisco
7 nodes
Lonepeak
141 nodes/1376 cores
Large memory, GigE (owner IB)
General 16 nodes/256 cores
NFS
/scratch/lonepeak/serial
Ash (417 nodes/7448 cores)
Tangent - dynamic provisioning;
up to 64 nodes/1024 cores
Switch
Kingspeak – still growing
344 nodes/7072 cores
Infiniband and GigE
General 48 nodes/832 cores
4 general GPU nodes
Apexarch - PE
homerfs
NFS
Home
Directories &
Group
Directories
Parallel FS
/scratch/general/lustre
NFS
/scratch/kingspeak/serial
Administrative Nodes

Brief Overview CHPC Resources

- *Computational Clusters*
  - kingspeak, ember – general allocation and owner-guest on owner nodes
  - lonepeak – both general and owner nodes, runs without allocation
  - tangent – only general nodes, variable number of nodes, runs without allocation
  - ash– can run as smithp-guest
- *Home directory* – NFS mounted on all clusters
  - /uufs/chpc.utah.edu/common/home/<uNID>
  - Default is 50GB with no backup unless your group purchases home directory space
- *Scratch systems*
  - /scratch/general/lustre – all clusters – 700 TB
  - /scratch/kingspeak/serial – all clusters but lonepeak – 175 TB
  - /scratch/lonepeak/serial – lonepeak only – 33 TB
  - /scratch/local on compute nodes, varying sizes
- *Applications*
  - /uufs/chpc.utah.edu/sys/pkg & /uufs/chpc.utah.edu/sys/installdir
  - /uufs/***$UUFSCELL***/sys/pkg where ***$UUFSCELL*** = kingspeak.peaks, ember.arches, ash.peaks, for cluster specific builds (mostly due to infinband interconnect)

Getting Started at CHPC

- Account application – *an online process*
  - https://www.chpc.utah.edu/apps/profile/account_request.php
- Username is your unid with password administrated by campus
- Interactive nodes
  - two CHPC owned nodes per cluster (***cluster***.chpc.utah.edu) with round-robin access to divide load (on ash, use ash-guest.chpc.utah.edu)
- CHPC login scripts – in account when created –
- Now using modules
  - https://www.chpc.utah.edu/documentation/software/modules.php
- Getting started guide
  - www.chpc.utah.edu/docs/manuals/getting_started
- Problem reporting system
  - http://jira.chpc.utah.edu or email to issues@chpc.utah.edu

Sources of Useful Information

- Getting Started Guide
  - https://www.chpc.utah.edu/documentation/gettingstarted.php
- CHPC policies
  - https://www.chpc.utah.edu/documentation/policies/index.php
- Cluster Usage Guides
  - https://www.chpc.utah.edu/documentation/guides/index.php
- Application Documentation
  - https://www.chpc.utah.edu/documentation/software/index.php
- Programming Guide
  - https://www.chpc.utah.edu/documentation/ProgrammingGuide.php

Interactive Node Usage

- Interactive nodes for prepping/testing of input files, analyzing results, compilations, debugging, data transfer, etc
  - no running of jobs
  - 15 min ***MAX*** cpu
  - no jobs of ***ANY*** time length that negatively impact ability of other users to get work done (e.g., heavy cpu, memory usage and/or i/o)
- https://www.chpc.utah.edu/documentation/policies/2.1GeneralHPCClusterPolicies.php#Pol2.1.1

Batch System -- SLURM

- Used to access compute nodes
  - https://www.chpc.utah.edu/documentation/software/slurm.php
  - example scripts, basic commands, information on SLURM environmental variables and directives
- Walltime limits
  - 72 hours (long qos by request) on all clusters except ash which is 24 hours

Access to Interactive Nodes

- FastX is the best option
- Server installed on all cluster interactive nodes as well as the seven frisco nodes
- Client can be downloaded and used on any university owned OR personal machine
- Clients exist for windows, mac and linux
- See instructions on how to download at https://www.chpc.utah.edu/documentation/software/fastx2.php

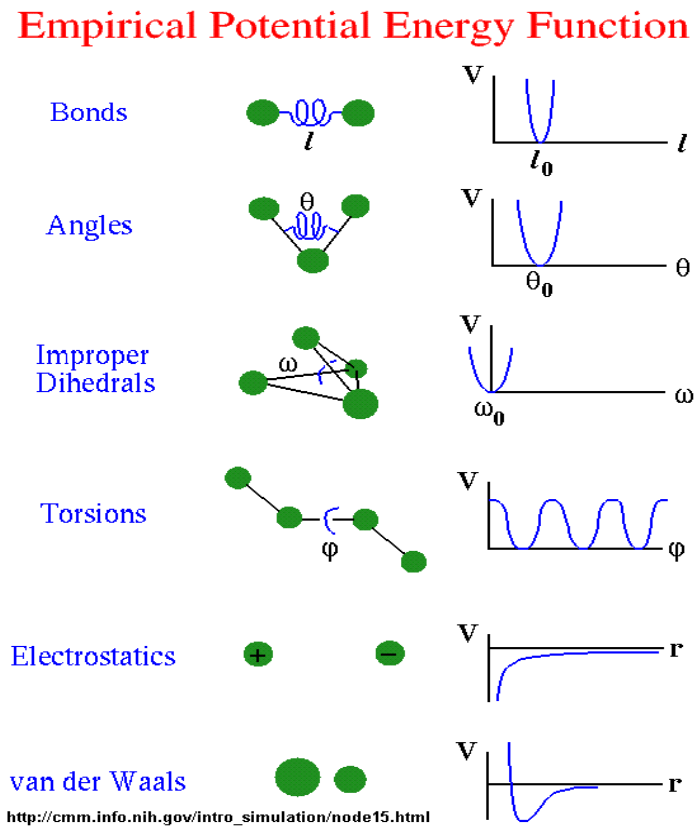
Computational Chemistry Packages

- Molecular Mechanics/Dynamics
- Semi Empirical
- Electronic Structure Calculations
  - HF, DFT, MPn, CI, CC, multi-reference methods, etc
- Solid state/Materials Modeling packages
  - Plane Wave Pseudopotential based packages

Calculation type & level of theory used depends on a number of factors – the information you are after, system being studied, size of system, computational resources. Often consider accuracy versus computational cost

Molecular Mechanics/Dynamics

- Amber
  - http://ambermd.org
- Gromacs
  - http://www.gromacs.org/
- NAMD
  - http://www.ks.uiuc.edu/Research/namd/
- LAMMPS
  - http://lammps.sandia.gov/
- Charmm
  - http://www.charmm.org/
  - licensed by group



Note – Gaussian and NWChem also have some MM capabilities

MM/MD packages

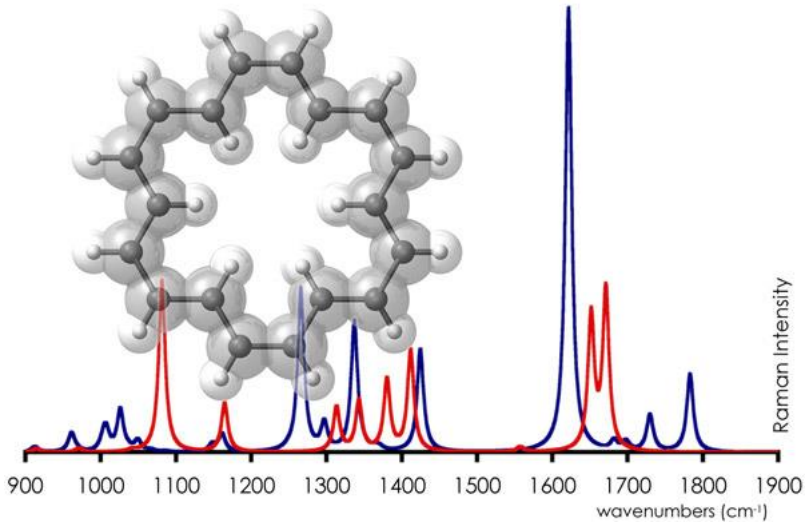
- Latest Versions:
  - Amber15 (Amber14 with AmberTools15)
    - ***module spider amber/15***
  - Gromacs 5.0.4 (single and double precision builds
    - ***module spider gromacs/5.1.1***
  - Lammps – 10Aug15 version
    - ***module spider lammps/7Dec15***
  - HOOMD
    - ***module spider hoomd/2.0.0.s***
    - ***module spider hoomd/2.0.0.d***
  - NAMD
    - ***module spider namd/2.10.ibverbs***
- Basic information on getting started:
  - https://www.chpc.utah.edu/documentation/software/amber.php
  - https://www.chpc.utah.edu/documentation/software/gromacs.php
  - https://www.chpc.utah.edu/documentation/software/lammps.php

Semi-Emprical Packages

- These are based on the Hartree-Fock formalism of electronic structure methods, but make many approximations and obtain some parameters from empirical data
- MOPAC2012 is the only stand alone semi-empirical only package installed at CHPC
- To use:
  - ***module load mopac***
  - ***mopac input.mop***
- Note – mopac only runs on a single core
- Sample scripts
  - **/uufs/chpc.utah.edu/sys/pkg/mopac/etc/test.slurm**
  - **/uufs/chpc.utah.edu/sys/pkg/mopac/etc/test.slurm and my.conf**
- Gaussian, GAMESS also has some semi-empirical capabilities

Quantum (Mainly) Packages

- Gaussian09
- Orca
- NWChem (6.3)



There are other packages not currently installed available free of charge, e.g.,

- GAMESS - http://www.msg.ameslab.gov/gamess/ (general purpose, properties)
- Psi4 – http://www.psicode.org/ (general purpose, properties)
- CFOUR – http://www.cfour.de/ (coupled cluster)
- SIESTA – http://departments.icmab.es/leem/siesta/ (DFT, linear scaling)

Gaussian09

- Commercial electronic structure package
  - http://www.gaussian.com for information and User's Guide
- Version D01 of G09 installed
  - /uufs/chpc.utah.edu/sys/pkg/gaussian09/EM64T (on all but lonepeak)
  - /uufs/chpc.utah.edu/sys/pkg/gaussian09/EM64TL (for lonepeak)
  - Also have AMD64 and legacy AMD64 builds
- For information on accessing the CHPC installation
  - https://www.chpc.utah.edu/documentation/software/gaussian09.php

NWChem

- Package developed at PNNL to work on massively parallel systems
- http://www.nwchem-sw.org
- Goal: Computational chemistry solutions that are scalable with respect to both chemical system size and MPP hardware size
- Has quantum mechanics, molecular mechanics/dynamics, and quantum molecular dynamics, plane waves for periodic systems
- Version 6.3 (***NOT*** with Python support) – working on 6.6 install
  - **/uufs/chpc.utah.edu/sys/pkg/nwchem/nwchem-6.3/bin/LINUX64**
  - **/uufs/ember.arches/sys/pkg/nwchem/nwchem-6.3/bin/LINUX64**
  - **/uufs/kingspeak.peaks/sys/pkg/nwchem/nwchem-6.3/bin/LINUX64**
  - **/uufs/chpc.utah.edu/sys/pkg/nwchem/nwchem-6.3/bin/LINUX6**4
- To run:
  - Load appropriate modules – do "***module spider nwchem***" for information
- More information and example batch script at
  - https://www.chpc.utah.edu/documentation/software/nwchem.php

Solid State/Materials Packages

Plane wave codes for the study of systems under periodic boundary conditions (PBC)

- NWChem has some functionality
- Quantum Espresso - http://www.quantum-espresso.org/
  - Understands crystal space groups
  - Has GIPAW module to do NMR calculations
  - J-ICE and XCrysDen to view
- Wien2K - http://www.wien2k.at/
  - Licensed per research group
- VASP – http://cms.mpi.univie.ac.at/vasp/
  - Ab initio QM/MD
  - Licensed per research group

Support Packages

- Molecular Viewers
  - Gaussview
  - Molden
  - VMD
  - Chimera
- Babel (Openbabel)
- Docking
  - Dock
  - AutoDock (and Autodock Vina)
- Cambridge Structural Database
- ECCE (mostly for NWChem)
  - Special case - Talk to me first if you want to use

GaussView

- Molecular builder and viewer for Gaussian input/output files
- CHPC provides campus license for linux version
- Chemistry Department has campus license for GaussView for windows; can buy into license
- Access with
  - ***module load gaussian09***
  - ***gv &***
- DO NOT submit jobs from within GaussView – instead create and save input file and use batch system
- Examples of how to use to show MO's, electrostatic potentials, NMR tensors, vibrations given on Gaussian's web page http://faculty.ycp.edu/~jforesma/educ/

Molden

- Another program for viewing molecular/electronic structures; version 4.7
- Works with Gamess, Gaussian, Molpro
- Supports plots of electronic density, MOs, etc
- More information at http://www.cmbi.ru.nl/molden/molden.html
- How to use at CHPC:
  - ***module load molden***
  - ***molden &***

Chimera

- For information on package and usage https://www.cgl.ucsf.edu/chimera/
- Version 1.10.1 installed
- /uufs/chpc.utah.edu/sys/installdir/chimera/1.10.1/bin/chimera
- Access by
  - ***module load chimera***
  - ***chimera &***

vgl
vglrun -c proxy vmd &

VMD

- Visualization, mainly for MM/MD
- latest version (1.9.2) installed
- Reads a number of different file formats
- Information at http://www.ks.uiuc.edu/Research/vmd
- Can install on own desktop (windows/mac/linux versions available)
- To use:
  - ***module load vmd***
  - ***vmd &***
  - ***vglrun –c proxy vmd & (for nodes with vgl capability (frisco6/7)***
- *Can use for 3D viewing on CHPC's vis wall*

OpenBabel

- Tool to interconvert structure files between a number of formats used in molecular modeling
- See openbabel.org for more information
- To run:
  - ***module load openbabel***
  - ***babel -i < input-type > < infile > -o < output-type > < outfile >***
  - ***babel –H*** to see format for usage, options, and input/output file-types

Dock/AutoDock

- Programs to look at binding of a small molecule within the active site of a receptor, usually a macromolecule
- Dock
  - version 6.6 installed
  - get info at: http://dock.compbio.ucsf.edu
  - ***module load gcc/4.7.2 mpich2/3.0.4.g dock***
  - ***dock6.mpi*** to start (needs arguments)
- Autodock
  - version 4.2.5.1
  - info available at http://autodock.scripps.edu
  - ***module load autodock***
  - ***autodock4*** (with proper arguments) or ***autogrid4***
  - autodocktools, a GUI interface, installed – start with ***adt &***
  - Autodock Vina – multicore performance and enhanced accuracy – start with ***vina*** (needs arguments)

Cambridge Structural Database

- Moved from library to CHPC summer 2006 – additions to the database are made every 3-4 months; package updated annually
- www.ccdc.cam.ac.uk for information
- Need CHPC account to use
- From PC need Xterm/Xwindowing software (Putty/XMing work well) to start session on any of the interactive nodes
  - ***module load csd***
  - ***cq &*** <- to start conquest (search engine)
  - ***mercury &*** <- to start crystal structure viewer
  - The first time you use it on a given computer you will be asked to confirm licensing
    - need to provide site/license codes (840/6CD8FF)

Integrated Molecular Platform Schrodinger Suite

- Commercial package geared for chemical modeling and simulation in the pharmaceutical field, specifically drug discovery
- http://www.schrodinger.com/
- Interface – Maestro (free for academia)
- Calculation code includes Jaguar (Quantum), Macromodel (MM), Qsite (QM/MM)
- Interfaces with Desmond for MD (installed)
- Tools for structure based drug design
  - Docking, ligand design, binding affinities, screening libraries
- Year to Year licensing
  - Token based, so limited in number of concurrent uses
  - Currently owned by single group

Finally…..

- Let us know if there is some other package that does something that our current packages do not; we can look into the possibility of getting it.
  - Factors: cost, hardware/OS requirements, licensing issues, usage needs
- Any questions – contact me
  - anita.orendt@utah.edu
  - Phone: 801-231-2762
  - Office: 422 INSCC

Using Gaussian09

- Gaussian users must be in Gaussian users group
  - ***groups*** will show you the groups your are in
  - If **g05** is listed, then you are in gaussian users group
- Program website www.gaussian.com
- CHPC Documentation at https://www.chpc.utah.edu/documentation/software/gaussian09.php
- Sample script g09-module.slurm in /uufs/chpc.utah.edu/sys/pkg/gaussian09/etc/
- New – now have nbo6 installed to use with G09
  - **Pop=NPA6**, **Pop=NBO6**, **Pop=NBO6Read** and **Pop=NBO6Delete** keywords
- Use Gaussview to build molecular system and to create input files
  - ***module load gaussian09***
  - ***gv &***

G09 Script

- Two sections for changes:

  #SBATCH --partition=lonepeak
  #SBATCH --account=youraccount
  #SBATCH --time=72:00:00
  ##SBATCH --constraint=c20
  #SBATCH --nodes=1
  #SBATCH -o slurm-%j.out-%N

- And:

  setenv WORKDIR $HOME/g09/project
  setenv FILENAME input
  setenv SCRFLAG LOCAL
  setenv NODES 1

Scratch Choices

- **LOCAL** (/scratch/local)
  - Hard drive local to compute node
  - 400GB on ember; 337 -800 GB on KP
  - Fastest option - recommended **IF** this is enough space for your job
  - Do not have access to scratch files during run but log/chk files written to $WORKDIR/$SLURM_JOB_ID/$UUFSCELL
  - Automatically scrubbed at end of job
  - Will not have access to rwf, needed for restarts of some jobs
- **KPSERIAL** (/scratch/kingspeak/serial)
  - NFS mounted on all clusters (interactive and compute) EXCEPT lonepeak
  - 175 TB
- **LPSERIAL** (/scratch/lonepeak/serial)
  - Only on lonepeak
  - 33 TB
- **GENERAL** (/scratch/general/lustre)
  - NFS mounted on all clusters (interactive and compute)
  - 700 TB

Input File Structure

- *Filename*.com
- Free format, case insensitive
- Spaces, commas, tabs, forward slash as delimiters between keywords
- ! Comment line
- Divided into sections (in order)
  - Link 0 commands (%)
  - Route section – what you want calculation to do
  - Title
  - Molecular specification
  - Optional additional sections

Number of Processors

- %nprocs – number of processors on one node
  - ember – 12
  - kingspeak – 16, 20 or 24, 28
  - lonepeak 8, 12 or 20
  - ash – 12 or 20
  - tangent – 16

Memory Specification

- Memory usage: default is 6MW or 48MB – all nodes have much more than this!
- If you need more use %mem directive
  - Units : words (default), KB, KW, MB, MW, GB, GW
  - Number must be an integer
- Methods to estimate memory needs for select applications given in Chapter 4 of User's Guide
- %mem value must be less than memory of node
  - Ember general nodes have 24GB; a few owners with 498, 192 or 256GB
  - Kingspeak – most have 64GB; there are some owner nodes with 32 or 128GB, a few with 1TB; 4 general with 384GB
  - Lonepeak – 12 core have 96GB; 20 core have 256GB, owner mixed
  - Ash – 12 core have 24GB; 20 core have 64GB; 24 core have 256GB

Input - Route Specification

- Keyword line(s) – specify calculation type and other job options
- Start with # symbol
  - for control of the print level in the output file use #n, #t, #p for normal, terse, or more complete output
  - #p suggested as it monitors job progress; useful for troubleshooting problems
- Can be multiple lines
- Terminate with a blank line
- Format
  - keyword=option
  - keyword(option)
  - keyword(option1,option2,..)
  - keyword=(option1,option2,…)
- User's Guide provides list of keywords, options, and basis set notation http://www.gaussian.com/g_tech/g_ur/l_keywords09.htm

Input - Title Specification

- Brief description of calculation - for user's benefit
- Terminate with a blank line

Input – Molecular Specification

- 1st line charge, multiplicity
- Element labels and location
  - Cartesian
    - label x y z
  - Z-matrix
    - label atom1 bondlength atom2 angle atom3 dihedral
- If parameters used instead of numerical values then variables section follows
- Default units are angstroms and degrees
- Again end in blank line

Parallel Nature of Gaussian

- All runs make use of all core per node with nprocs
- Only some portions of Gaussian run parallel on multiple nodes (includes most of the compute intensive parts involved with single point energies/optimizations for HF/DFT)
- If time consuming links are not – job WILL NOT benefit from running on more than one node
- Not all job types are restartable, but more are restartable in G09 than were G03 (e.g., frequencies and NMR) – see ***restart*** keyword
  - Requires rwf from previous run
  - Still restart optimizations and single point energies the old way
- CHPC does allow for jobs over standard walltime limit if needed – but first explore using more nodes or restart options
- CHPC also has a few nodes outside of these clusters which have more memory and no wall time limits – talk to me if you think you need to use

Timings of G09 scratch system

Scratch setting	LOCAL	KPSERIAL	
time	1.75 hrs	1.75 hrs	

***depends strongly on amount of I/O and on other jobs usage on shared scratch systems

Scaling G09

B3PW91;650 bf; 8 opt steps; time in hours

# ember nodes	wall time	# KP 16 core nodes	wall time
1	3.25	1	1.5
2	1.5	2	<1

DFT Frequency of same case

# ember nodes	wall time	# KP 16 core nodes	walltime
1	3.25	1	1.25
2	1.75	2	0.75

Scaling G09 MP2 Opt (Freq)

8 opt steps 338bf; time in hours (w/ freq)

# ember nodes	wall time	# KP 16 core nodes	wall time
1	1 (7.5)	1	30min (3.75)
2	<1 (7)	2	24min (3.75)

RWF Sizes – Choice of Scratch

- For a DFT optimization with 462 basis functions
  - 150mb RWF
- For a DFT freq of above structure
  - 1.1gb RWF
- For a MP2 optimization with 462 bf
  - 55gb RWF AND 55gb SCR file
- For a MP2 frequency of above structure
  - 247gb RWF