

**Penn State, August 2013**

**Cloud-WIEN2k**

**A Scientific Cloud Computing  
Platform for Condensed Matter Physics**

**K. Jorissen**

**University of Washington, Seattle, U.S.A.**

Supported by NSF grant OCI-1048052

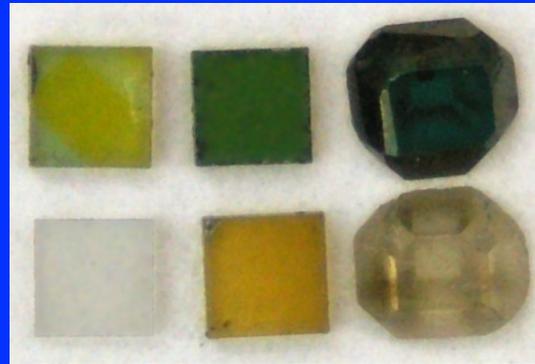
[www.feffproject.org](http://www.feffproject.org)

# Materials Science

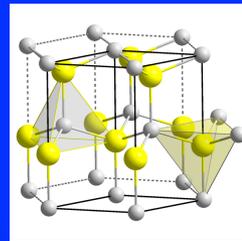
Materials Science research:

Theoretical models, evaluated on a computer, are usually needed for interpretation and quantification of measurements.

But HPC is often not readily available.



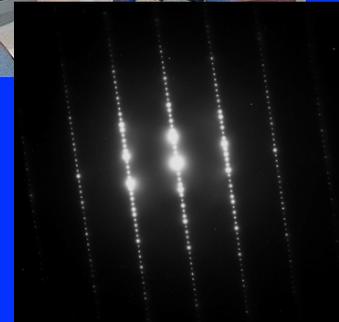
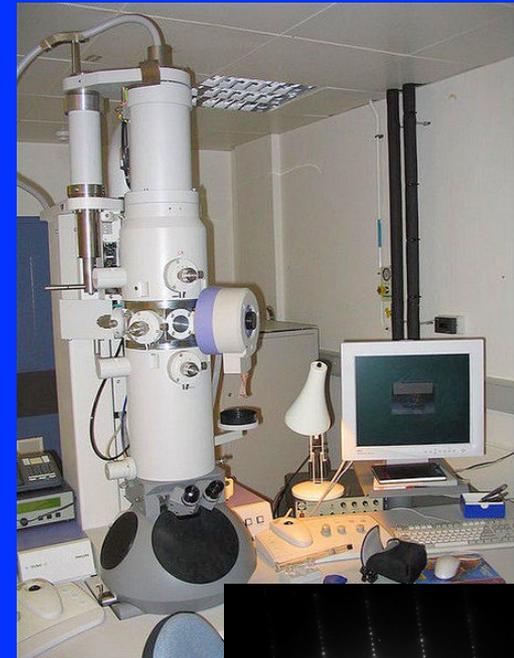
sample



theoretical model

measurement

interpretation



$$H\psi = E\psi$$
$$E = mc^2$$

## **Anecdote (High-Performance Computing is everywhere)**

### **Computational linguistics:**

“We automatically identify semantically related words in the 400 million word Dutch Twente corpus to

- Statistically find contextual associations and quantify association strength
- Identify syntactical relations between words
- Relevant to automatic translation software

Multivariate analysis with dozens of variables – large computational needs.”

--- an “English Lit major”

How do we bring the best theory and simulations to the scientists who need it?

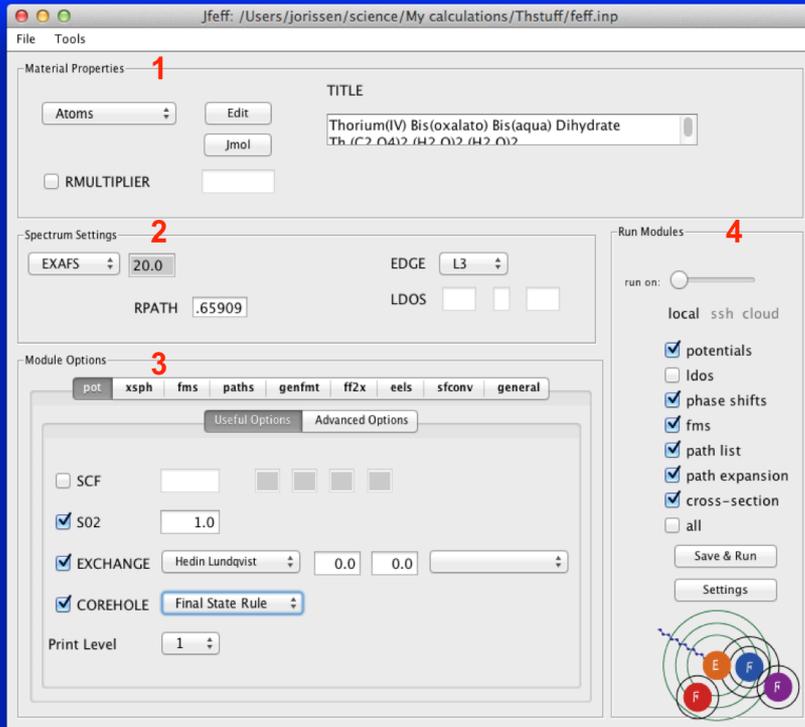
(often applied scientists - not computational specialists)

SOLUTION:

Scientific Cloud Computing

# Are state-of-the-art calculations “work for specialists”?

## FEFF-old (simple Einstein model for phonons)



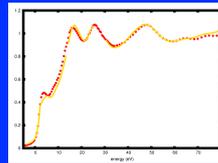
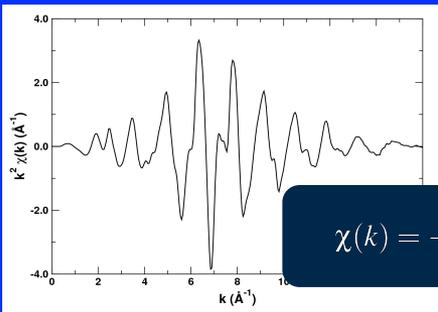
GUI

Easy install

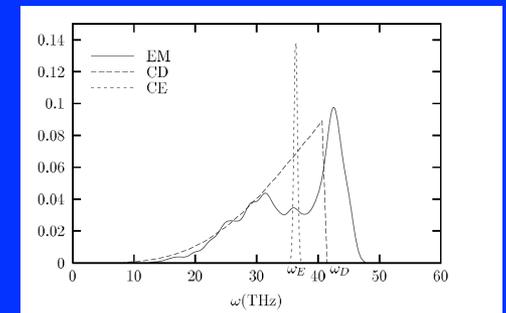
Runs on laptop

Load file & Click “Run”

~ 1 day to learn



$$\chi(k) = -k \sum_i S_0^2 N_i \frac{f_i^{eff}(k)}{R_i^2} e^{-R_i/\lambda} \sin(2kR_i + 2\phi(k)) e^{-\sigma_i^2 k^2 / 2}$$



# Are state-of-the-art calculations “work for specialists”?

FEFF-gold (accurate ab initio model for phonons)

Dynamical Matrix (DM) (DFT) -- ABINIT -- DFT requires cluster

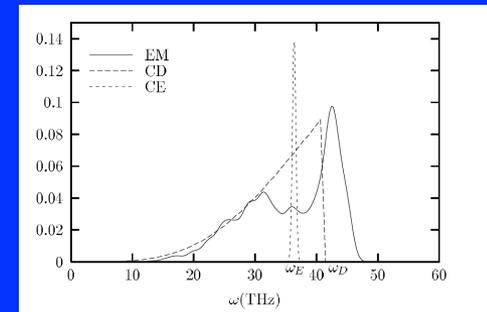
Debye

Invented / published 2006-2009  
Clearly an improvement  
Nobody uses it

X-ray Absorption

-- FEFF

~ 0.x grad students to learn



# *Are state-of-the-art calculations “work for specialists”?*

- Hardware barrier: advanced codes need clusters
  - Buy a cluster? IT support?
  - Supercomputing center?
  - Collaborate with specialists?
- Software barrier: running codes is difficult
  - Installation of || software tricky
  - lacking user-friendliness
  - multi-code workflows difficult

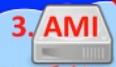
  $t \gg 1$  before improved theory reaches applied research

## Scientific Cloud Computing Platform



2. Java Interface Tools

launch



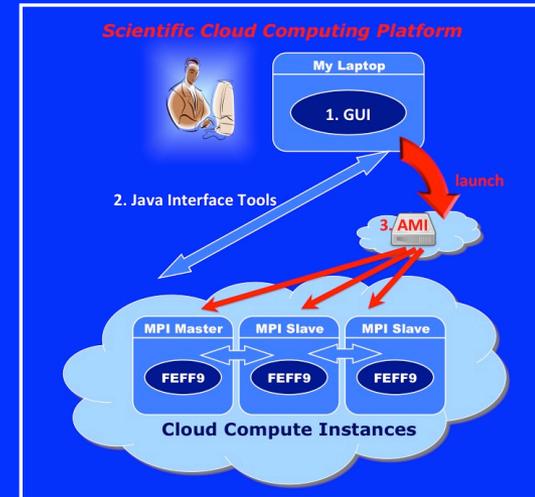
Cloud Compute Instances

Interface simplifies workflow (hides cloud -- app)

Developer makes virtual "XAS" compute node with preinstalled WIEN2k

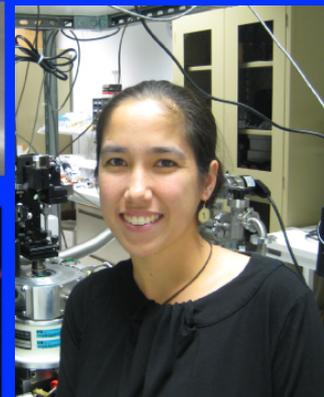
User requests 5 node Cloud Cluster for 3 hours when needed (\$20)

- Contains utilities for parallel scientific computing: MPI, compilers, libraries, NFS, ...
- Becomes compute node in SCC Cloud Cluster
- Developer-optimized Scientific codes for your research
  - **WIEN2k** for electronic structure calculations
  - latest version
  - optimized for performance
  - MPI parallelization for large calculations



“My new research group was looking for a way to implement *MEEP-mpi* (MIT Electromagnetic Equation Propagation) to simulate EM fields in nanoscale optical devices for cavity QED experiments. **We believe that Amazon EC2 is an economical and time saving solution for our finite difference time domain (FDTD) simulations.** My group’s research iterates between fabrication and simulation thus it is advantageous to buy computing power only when needed. **Moreover it is a relief not to have to maintain our own small cluster within our group.**”

Kai-Mei Fu, University of Washington (USA)





# WIEN2k-cloud

Cr3C2@giorgione.phys.washington.edu

giorgione.phys.washington.edu:7890/index.pl?SID=730047

Session: **[Cr3C2]** /Users/jorissen/science/calculations/Cr3C2

### SCF Cycle

**Options: (help)**

- spin polarized
- AFM calc.
- iterative diag
- iterative(0)
- in1orig
- parallel
- spinorbit
- dm
- orbital pot (LDA+U)
- eece (hybrid functionals)

**Expert options:**

- FSM: 0
- no HNS: 6
- in1new: 2
- q-limit: 0.05
- lt-number: 40

**Convergence criteria:**

- Energy: 0.0001 Ry
- Force: 1 mRy/au
- Charge: 0.001 e

Type of execution: cloud

E-mail notification  to

start SCF cycle Clear entries  only save parameters

**WIEN2k GUI (DFT)**

w2web © lutz.at

- Starts || cluster in EC2 cloud
- Uploads initialized calculation
- Runs || calculation in EC2 cloud
- Downloads results to laptop
- Deletes EC2 cluster

Other workflows / data flows can be added.

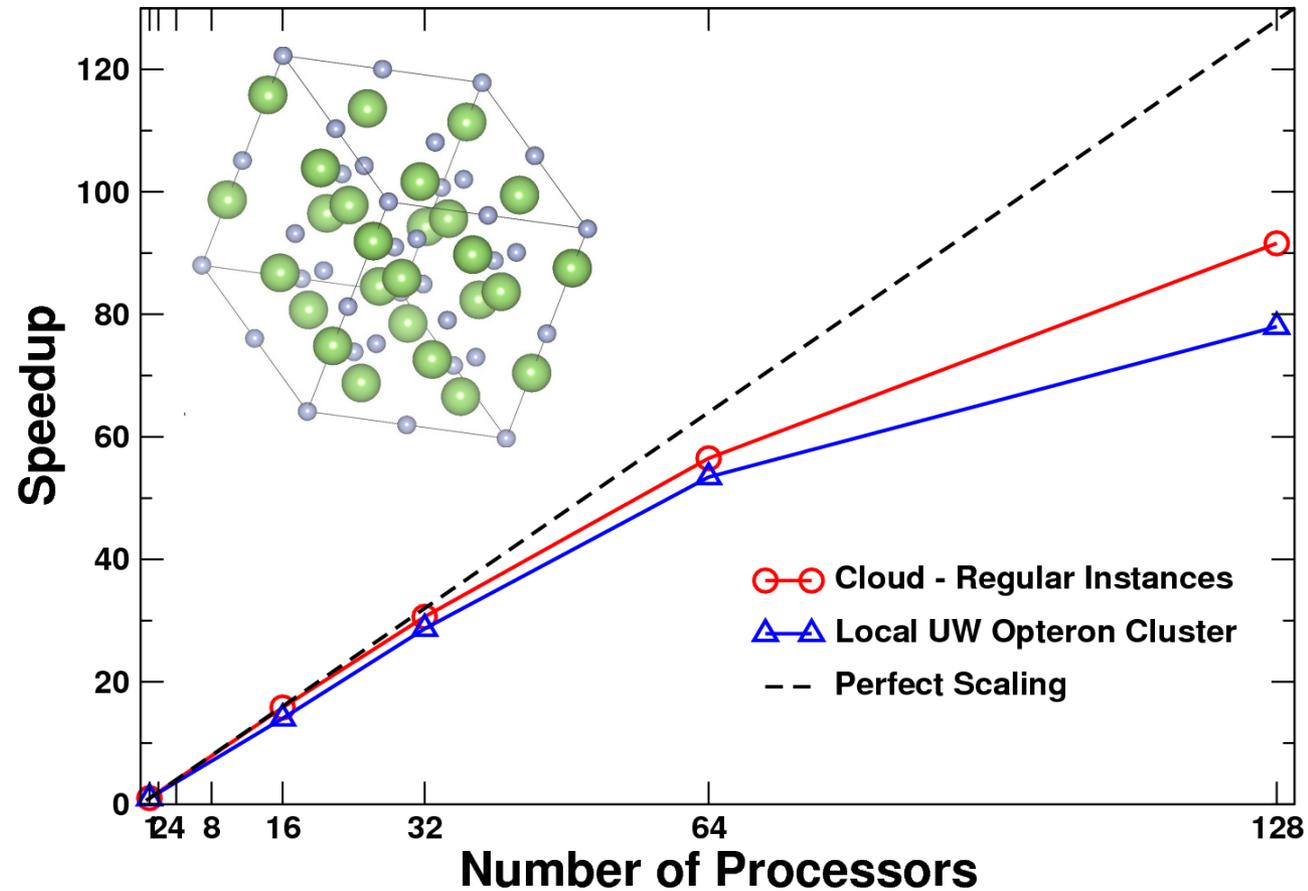
Requires:

- create EC2 account
- install SCC program

# Performance

## LOOSELY Coupled Processes

DFT KS equations on  
128 k-point grid



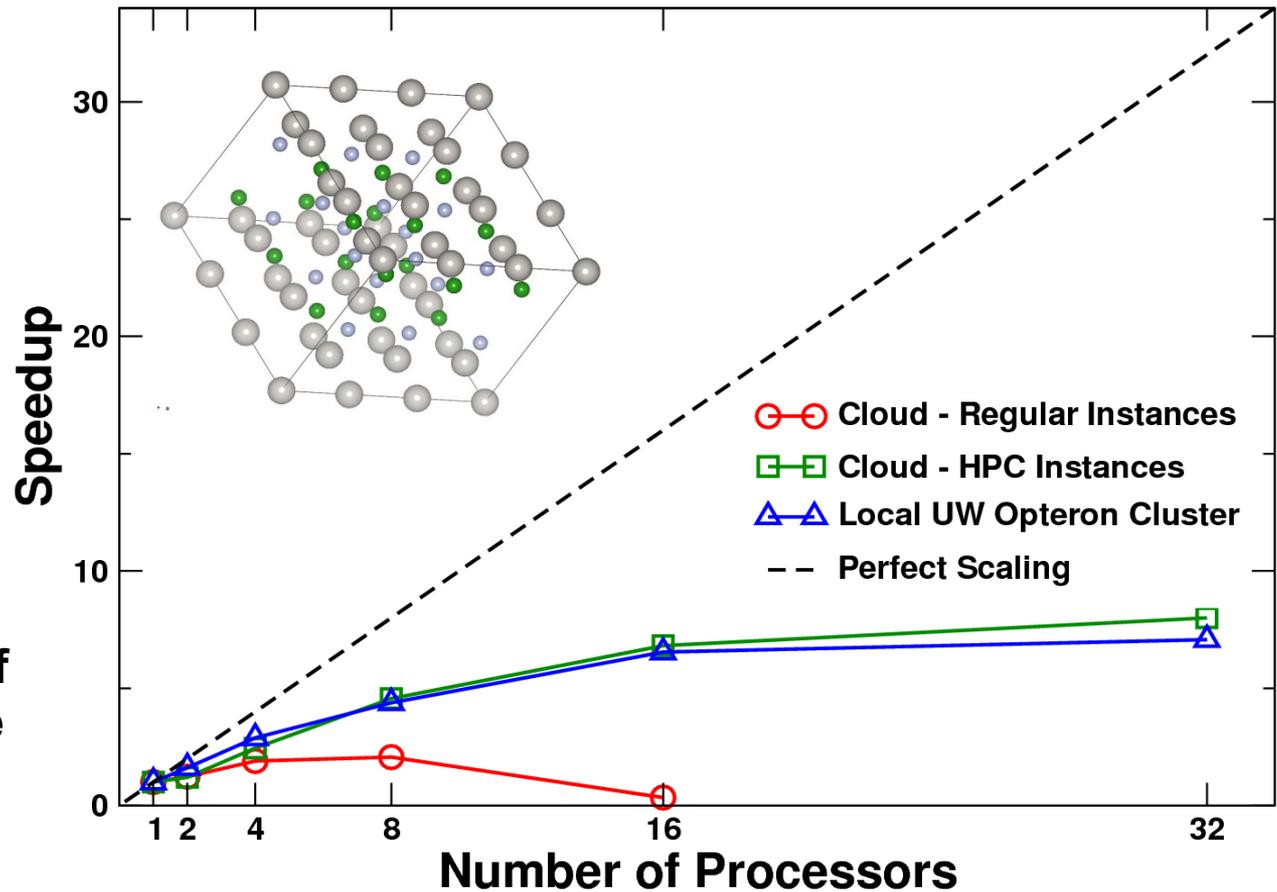
Good scaling

# Performance

## TIGHTLY Coupled Processes

KS for large system  
at 1 k-point

VERY DEMANDING of  
network performance



HPC cluster instances deliver good speedup

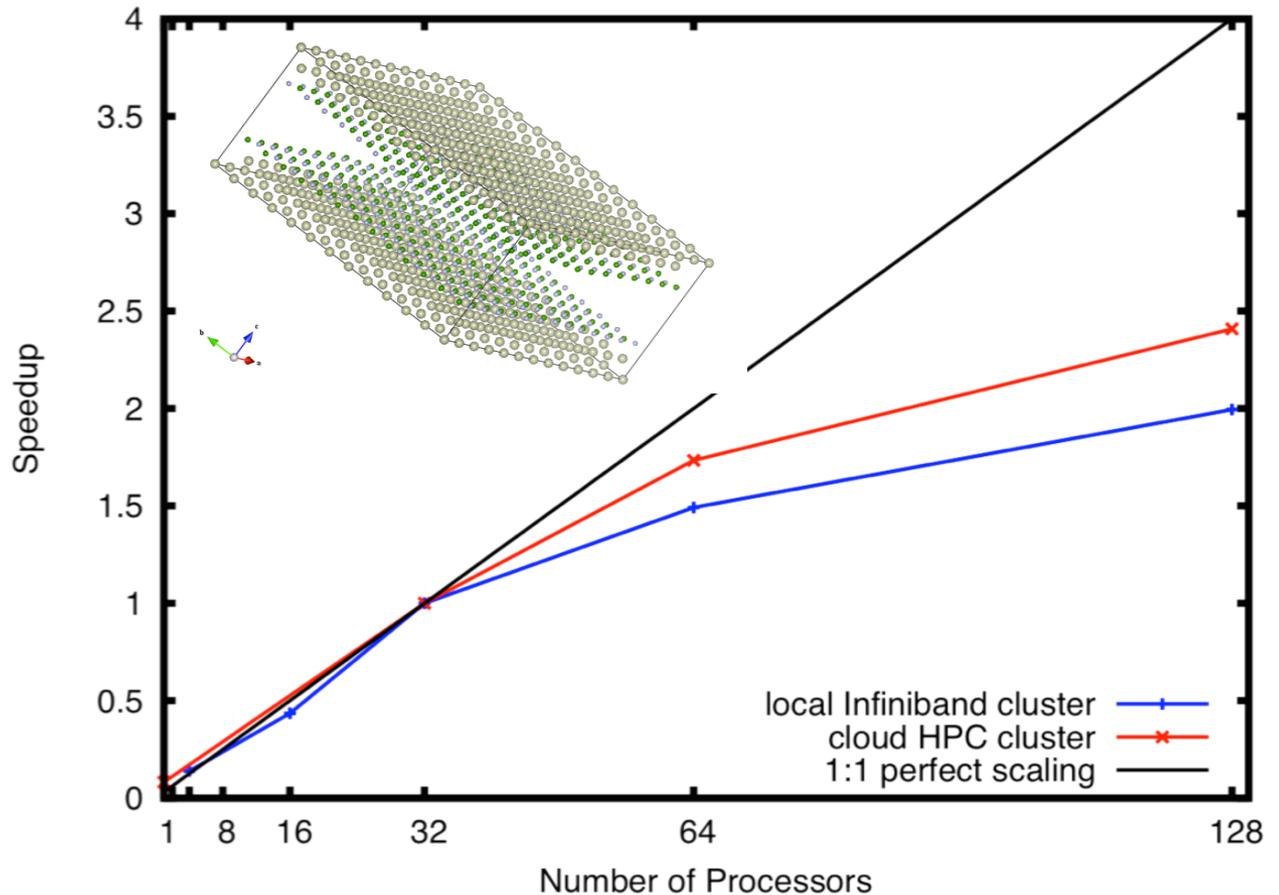
# 5. WIEN2k Performance Benchmarks

## TIGHTLY Coupled Processes

### KS for large system at 1 k-point

H size 56,000 (25GB)  
Runtime (16x8 processors) :  
Local (Infiniband) 3h:48  
Cloud (10Gbps) 1h:30 (\$40)

**VERY DEMANDING of  
network performance**



- 1200 atom unit cell; SCALAPACK+MPI diagonalization, matrix size 50k-100k
- HPC cluster instances deliver similar speedup as local Infiniband cluster

# “Scientific Cloud Computing can bring novel theory & HPC modeling to more researchers.”

Scientific Cloud Computing

A high-performance platform for taking materials simulations to the Cloud

Home The SCC Interface Tools JFEFF Publications Download The FEFF project

About Scientific Cloud Computing

Scientific Cloud Computing is a new way of doing scientific calculations. It means you don't have to apply for time at a supercomputing center, or buy a cluster, and spend your time installing and configuring software. Instead, you rent a virtual cluster by the hour. It comes preconfigured with scientific codes and tools and it's set up for parallel computing. You specify how much computing power you need today, and you let go of it once your work is finished. "Easy" is the key word here.

There has long been concern that Cloud Computing could not deliver the performance needed for parallel scientific computing. With the latest improvements in Cloud infrastructure, we have found that those concerns are no longer warranted. You can now use Cloud Computing for state-of-the-art parallelized scientific calculations.

The platform we offer contains:

- **SCZVP:** The Scientific Cloud Computing Virtual Platform, a scientific virtual machine prototype (AMI) for the Amazon EC2 cloud; it is preloaded with materials science software and scientific libraries and tools
- **SCZIT:** The Scientific Cloud Computing Interface Tools, a toolset that lets you create and manage virtual cloud clusters using simple commands (currently Linux and OS-X)
- **JFEFF:** A Java-based GUI for the FEFF9 spectroscopy code that can work in the cloud from your machine - no commandline work needed (currently Linux and OS-X only)

www.feffproject.org

Computer Physics Communications 183 (2012) 1911–1919

Contents lists available at SciVerse ScienceDirect

Computer Physics Communications

journal homepage: www.elsevier.com/locate/cpc

ELSEVIER

A high performance scientific cloud computing environment for materials simulations

K. Jorissen\*, F.D. Vila, J.J. Rehr

Department of Physics, University of Washington, Seattle, WA 98195, USA

ARTICLE INFO

ABSTRACT

Article history:  
Received 29 September 2011  
Received in revised form 11 April 2012  
Accepted 16 April 2012  
Available online 24 April 2012

Keywords:

We describe the development of a scientific cloud computing (SCC) platform that offers high performance computation capability. The platform consists of a scientific virtual machine prototype containing a UNIX operating system and several materials science codes, together with essential interface tools (an SCC toolset) that offers functionality comparable to local compute clusters. In particular, our SCC toolset provides automatic creation of virtual clusters for parallel computing, including tools for execution and monitoring performance, as well as efficient I/O utilities that enable seamless connections to and from the cloud. Our SCC platform is optimized for the Amazon Elastic Compute Cloud (EC2). We present

Comp. Phys. Comm. 183 (2012) 1911

1. Introduction

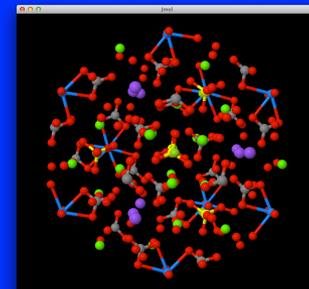
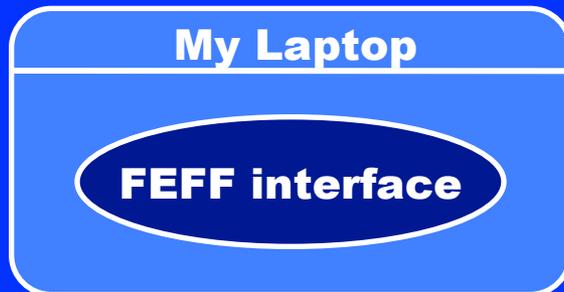
managing general purpose computing clusters on EC2. However, significant further developments were needed to create a platform

We acknowledge: ♦ **FEFF:** S. Story T. Ahmed B. Mattern M. Prange  
**J. Vinson** ♦ **UW:** R. Coffey E. Lazowska J. Loudermilk ♦ **Amazon:**  
**D. Singh** ♦ **NSF:** C. Bouldin ♦ supported by NSF OCI-1048052

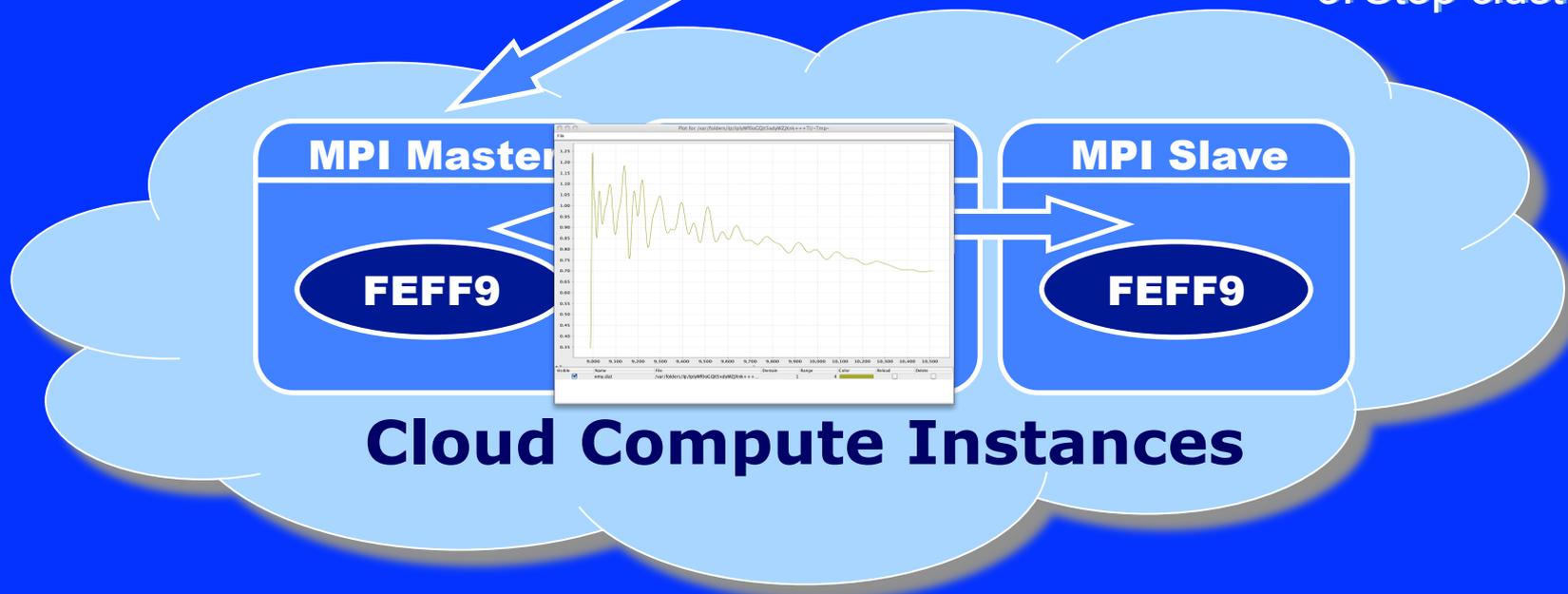




# 4. Cloud-Computing on the Amazon EC2 cloud



1. Create cluster
2. Calculations
3. Stop cluster



# Developer's view:

## ExecuteCloudContext.java:

```
import edu.washington.scc.*;

// Launch the new cluster with "cs" specifications:
ClusterResult rl = clust.launch(cs);

// Initialize the FEFF calculation on the cloud cluster:
// Copy feff.inp:
ClusterResult rp = clust.put(LocalWorkingDir+"/feff.inp", CloudWorkingDir+"/feff.inp");

// Run the FEFF9-MPI calculation:
ClusterResult rf9 = clust.executeCommand(Feff9CommandLine,CloudOut);

// Copy the output files back to the local computer:
ClusterResult rg = clust.get(CloudWorkingDir, LocalWorkingDir);

// Terminate the cloud cluster:
ClusterResult rt = clust.terminate();
```

# End User's view:

