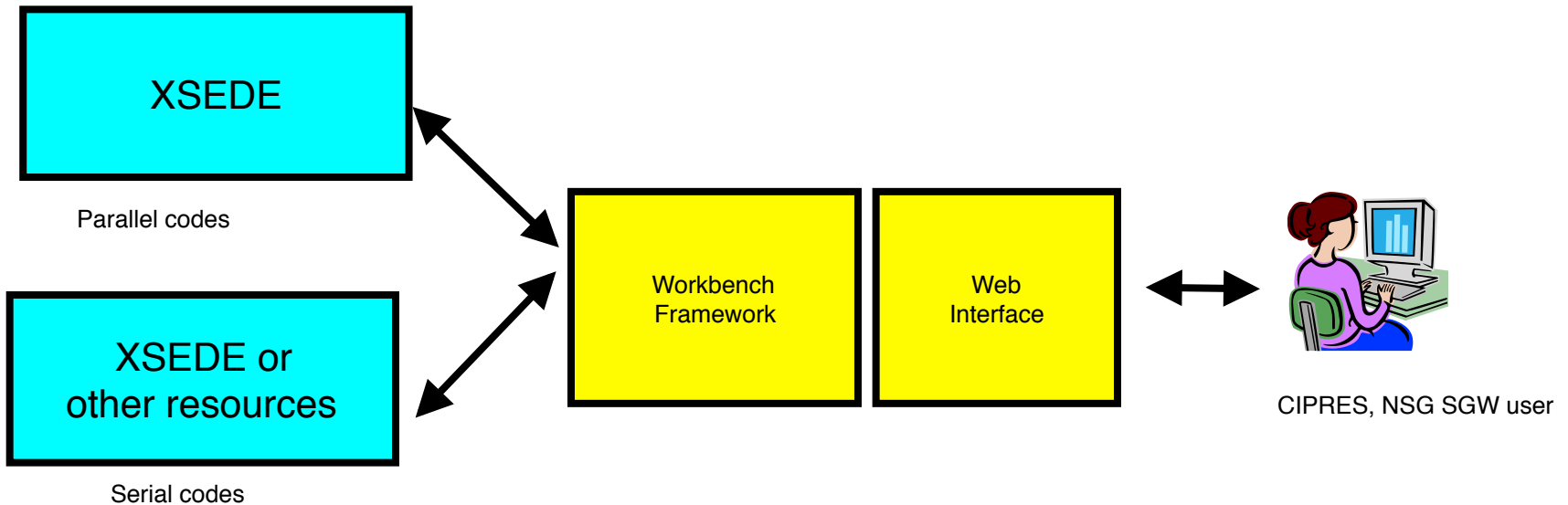# *Exploring Science Gateway Use Cases for Cloud Computing*

M. A. Miller, P. Papadopoulos, A. Majumdar, S. Smallen,
N. Wilkins-Diehr, R. P. Wagner, M. Tatineni, R. S. Sinkovits,
R. L. Moore, M. L. Norman

San Diego Supercomputer Center
University of California San Diego

SDSC

SAN DIEGO SUPERCOMPUTER CENTER

at the **UNIVERSITY OF CALIFORNIA; SAN DIEGO**

UCSD

Science Gateways – CIPRES, NSG

- Science Gateways provide easy and free access to HPC resources

- CIPRES: Science Gateway for phylogenetics research

- NSG: Science Gateway for computational neuroscience

- There are many other successful XSEDE Science Gateways (https://portal.xsede.org/science-gateways)

XSEDE

Parallel codes

XSEDE or other resources

Serial codes

Workbench Framework

Web Interface

CIPRES, NSG SGW user

NSF Funded CIPRES – Continually Growing

- CIPRES accounts for more than 30% of all XSEDE users
- More than 1300 publications enabled since 2009
- Staff size is less than 2 FTE
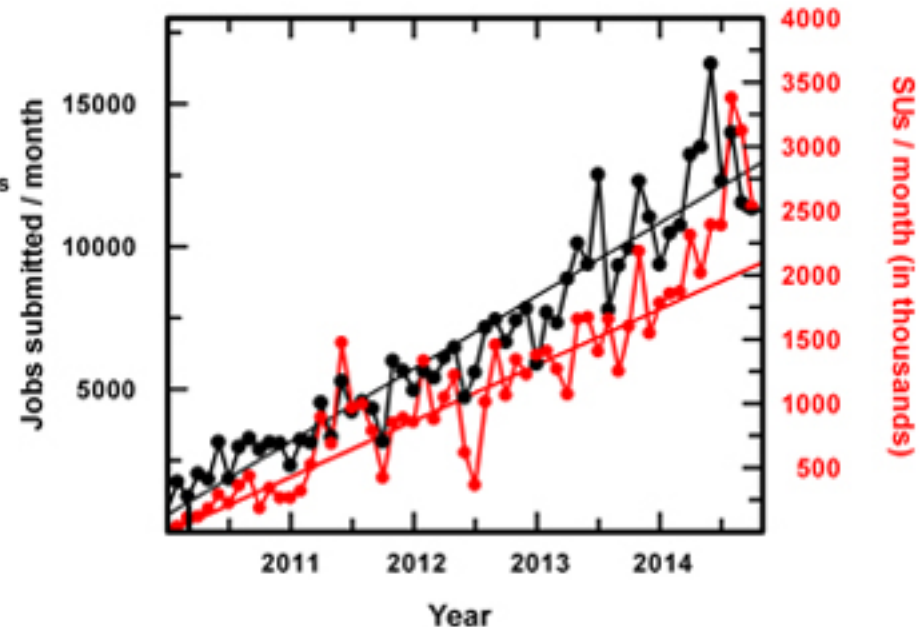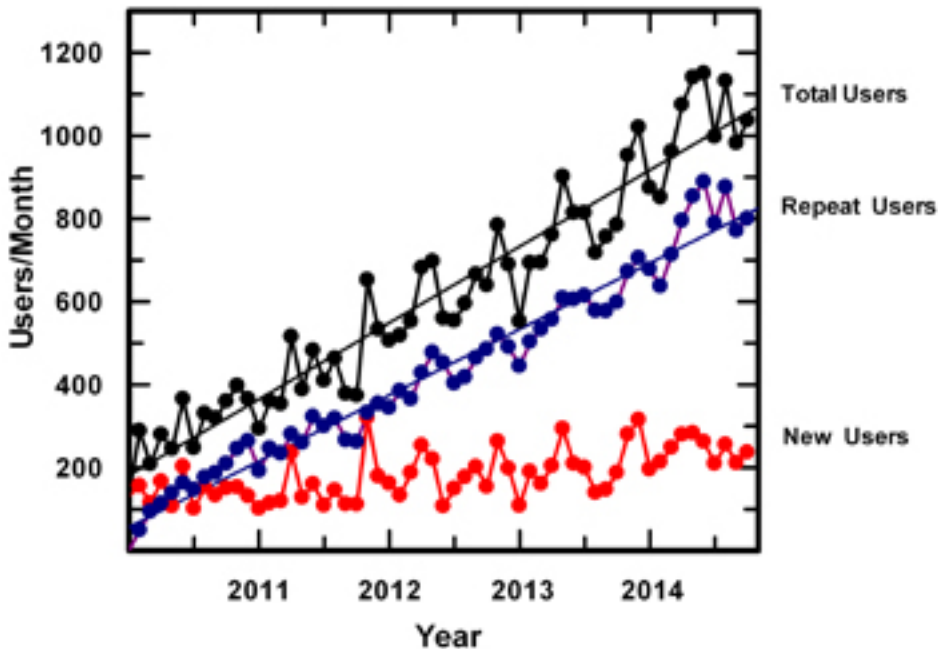
Usage 12/1/2009 – 10/1/2014

Neuroscience Gateway (NSG) – a NSF funded Science Gateway

- Went into friendly production phase in 2013

- In 2014 ~2M SUs used by neuroscience users; ~200 users

- Provides parallel spiking neuronal simulation codes – truly parallel simulation as well as parameter sweep type simulations
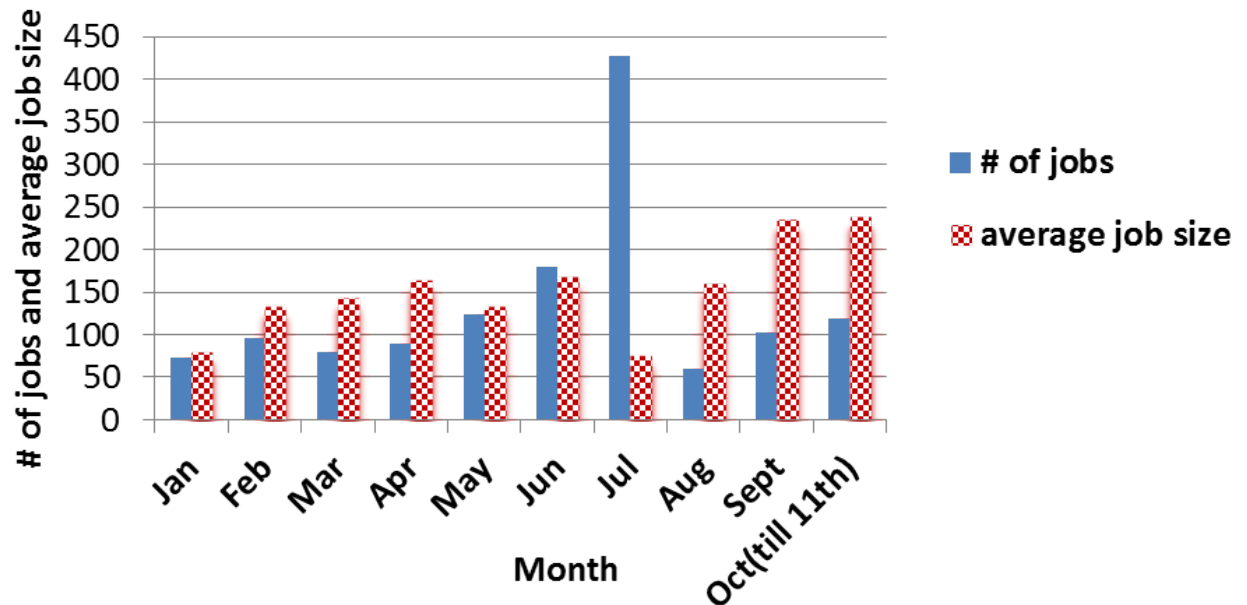
2014 (till mid October) - # of jobs and average job size per month

NSF HPC Resources

- NSF has made a significant investment in HPC resources with ~10s to 100s of thousands of cores, fast interconnects, parallel file systems etc.
  - Designed for codes that scales to many 1000s of cores
- This has accelerated the rate of scientific discovery in bio science (and other science domains)
- CIPRES codes run on these high end resources
  - Codes supported by CIPRES scale to no more than 64 cores
  - Significant fraction of CIPRES codes run on only 8 cores
- Neuronal simulations codes scale to ~100s to ~1000s of cores
  - But there are many parameter sweep type simulations using ~10s to ~100s of cores

Cloud Resources for Science Gateways ?

- What mix of high end HPC resources and Cloud resources are appropriate for SGW ?

- Provide the science impact needed by SGW while utilizing resources appropriately

- Oversubscription of XSEDE resources is an issue
  - Oversubscription threatens CIPRES operation
  - NSG ran out of allocated SUs in both 2013 and 2014 (and needed supplemental allocation)

SGW Usage Policies

- For fully allocated (and oversubscribed) resources (e.g. from XSEDE) limits are in place (within CIPRES, NSG) for max usage allowed by each SGW user

- Some populations of users do not use SGW as their computational needs are greater than what SGW provides
  - What is the size of this population ?
  - How are they meeting their needs ?
  - Can a fee-for service be provided for them ?

- Can a cloud provider offer lower cost solution on appropriate compute resources ?
  - Users can pay for service beyond the limit of free service
  - Thus avoid the limits in place for free resources

Cloud Experiments for SGW

- Establish minimum requirement of SGWs – CIPRES, NSG, others

- How do these requirements match with available provisioning options from cloud computing ?

- Benchmark codes on appropriate cloud resources
  - Network access
  - File systems
  - Data transfer
  - Code performance
  - Create distributable VM images
  - Any other GUI, workflow requirements etc.

Advantages of Clouds

- High end resources can be freed up for scalable codes
- Low scalability codes can move to less costly resources
- Portable work environment can be moved to commercial cloud
- Users can use home environment for low end runs and commercial resources for high end runs

Risks

- Poor scalability, performance of codes
- Poor file system, negative network impact, worse data management, portability issues
- The gain in throughput may not be worth the increased complexity
- Will savings in H/W equipment be offset by cost of software development