Molecular Parameter Optimization Gateway (ParamChem):

Workflow Management through TeraGrid ASTA

Jayeeta Ghosh
National Center for Supercomputing Applications
1205 W. Clark St, Urbana, IL
jayeeta@ncsa.uiuc.edu

Nikhil Singh
National Center for Supercomputing Applications
1205 W. Clark St, Urbana, IL
nsingh@ncsa.uiuc.edu

Ye Fan
National Center for Supercomputing Applications
1205 W. Clark St, Urbana, IL
yefan@ncsa.uiuc.edu

Suresh Marru
Pervasive Technology Institute
2719 East 10th Street
Bloomington, IN 47408
smarru@cs.indiana.edu

Kenno Vanomesslaeghe
University of Maryland
20 Penn Street
Baltimore, MD 21201
kvanomme@umaryland.edu

Sudhakar Pamidighantam
National Center for Supercomputing Applications
1205 W. Clark St, Urbana, IL
spamidig@ncsa.uiuc.edu

Note: Additional authors MacKerell, Sheetz, Pierce, Polani and Gazula are listed in the 'Additional Authors' section.

ABSTRACT

Parameter optimization for chemical systems requires generation of initial guesses. These parameters should be generated using systematic sampling of parameter space, minimizing differences between output data and the corresponding reference data. In this paper we discuss the ParamChem project, which is creating reusable and extensible infrastructure for the computational chemistry community that will reduce unnecessary and eliminate redundancies in parametrized computations using modern software engineering tools.

The paper particularly focuses on constructing and executing coupled molecular chemistry models as complicated workflow graphs. These workflow management capabilities have been integrated with the GridChem Science Gateway infrastructure through the TeraGrid advanced user support program. Further, we describe how the project is enabling a sustainable growth for science gateway infrastructure by building upon tools provided by the Open Gateway Computing Environments. The paper also discusses plans for integrating TeraGrid information, monitoring and prediction services to provide automated job scheduling with resource maintenance and fault aware services.

Categories and Subject Descriptors

D.2 [**Software Engineering**]: Programming Environments—*Airavata*; D.2.11 [**Software Architectures**]: Domain-specific architectures

TeraGrid '11 July 18-21, 2011, Salt Lake City, Utah, USA.

General Terms

Design, Experimentation

Keywords

ParamChem, GridChem, OGCE, Scientific Workflows, TeraGrid Advanced User Support, Science Gateways

1. INTRODUCTION

Molecular mechanics (MM) and semi-empirical (SE) quantum chemical methods are routinely used for describing molecular properties of chemical systems [9]. These simulation methods are used to provide ways to compute energy/relative energy of a molecular system based on interactions at atomic or sub atomic (electron and nuclei – containing protons and neutrons) level of description of the molecules. However, the accuracy and range of applicability of these methods depend critically on the quality and coverage of the parameters that model inter atomic or inter sub atomic interactions included in the models.

The parameters of complex chemical processes, such as catalysis and biological function, need to be optimized within the context of the appropriate energy function or Hamiltonian. They must also be made consistent with each other. For example, to describe and exploit electronic characteristics of materials on bulk or bulk-like systems, approximate quantum chemical techniques are used. These characteristics depend on parameters to treat the description of the electron density or its correlation. If inaccurate parameters are used, significant computational resources are wasted while arriving at misleading results. However, large efforts are required to validate available parameters as well as extend parameter sets to novel systems. These challenges often represent a rate limiting step in theoretical investigations.

The fundamental quantities for the validation and optimization of parameters are the target reference data. This data is collected from reliable experimental models such as X-Ray crystal structures for structural data or computed data at a higher reliable model using abinitio quantum chemical(QM) simulations for structure, energetic and spectra. Such data must be integrated with energy of known parameters or optimization of novel molecular parameters to repro-

duce that data in a unified scheme. The reference data must be readily accessible as well as extensible allowing the automated identification. Examples of reference data include experimental solution and crystal data and ab initio QM data as well as cyber tools to rapidly extend the reference data set by preparing and undertaking QM calculations on novel molecules.

The reference data is systematically generated with conformational ensembles of targeted molecules, preconditioning of the conformational ensemble, performance of the QM calculations and systematic insertion of results into the extensible reference database. Proper organization of the database, including molecular species, preconditioned structures and experimental and QM data is essential to allow the parameter optimization. Parameter optimization requires generation of initial guesses for the parameter set in the context of the targeted energy function and systematic sampling of parameter space to minimize differences between output data from the model being optimized and the corresponding reference data.

To address such need for parameter optimization for chemical systems a virtual organization named **ParamChem** is being developed. ParamChem will be broadly applicable for a variety of chemical systems and should work for multiple models of energy functions or Hamiltonians. The project is creating a reusable and extensible infrastructure for the computational chemistry community that will eliminate unnecessary and inaccurate computations. Parametrized results will be made available through a publicly accessible database.

Further in the paper, we describe the computational aspects of the ParamChem project and the accomplishments through the TeraGrid advanced user support program. In addition, we detail the collaboration with the Open Gateway Computing Environments Project that addresses sustainability of a science gateway infrastructure.

2. PARAMCHEM REQUIREMENTS

The core of ParamChem gateway is to develop and provide automated Force Field parametrization capabilities. The parametrization can be broken down to several steps as illustrated in the Figure 1. The modeling steps relevant to optimization are enumerated as:

1. Atom typing
2. Generation of initial guess charges and Lennard-Jones parameter assignment
3. Generation of initial guess parameters
4. Generation of target data for charge optimization
5. Charge optimization
6. Generation of target data for optimization of bond, angle, dihedral and improper dihedral parameters
7. Optimization of bond angle, dihedral and improper dihedral parameters
8. Generation of target data for optimization of dihedral parameters about rotatable bonds
9. Optimization of dihedrals about rotatable bonds

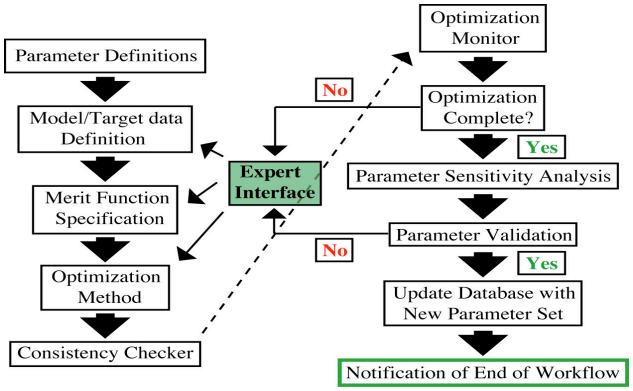

Figure 1: Optimization Flow

Each of these aspects has its own specific challenges and solutions; for instance, an algorithm that performs well for optimizing charges may be inappropriate for optimizing dihedral parameters about rotatable bonds. Significant progress has been made in the MacKerell lab towards automating each of the steps listed above. Specifically, an automatic atom typer for CGenFF as well as functionality for generating initial guess charges and parameters has been implemented, which is further discussed in Section 3.1.2.

The infrastructure for the specification and generation of target data, which will be used in tandem with target data from the database is underway and is further described in Section 3.3. Schemas for optimization of charges and dihedrals about rotatable bonds were previously published by the MacKerell lab. The optimization algorithms that play an important role in these methodologies are being generalized and improved. A number of approaches for the optimization of bond, angle, rigid dihedral and improper dihedral parameters are being evaluated.

The parametrization steps listed above are developed as software tools to initialize and optimize parameters required to provide reliable and accurate MM and SE models for a wide range of chemical systems. Finally, the ParamChem gateway derives requirements to ensure the infrastructures is readily extensible to allow for acquisition, generation and targeting of novel properties not already implemented as required to assure continued utility of the system for future, unforeseen applications.

MM and SE methods are extremely popular for their versatility in treating molecular structure and dynamic phenomena in a computationally efficient manner. These methods require accurate parameters for the chemical constituents of the system of interest in the context of the appropriate energy function or Hamiltonian. As new materials and molecular moieties are probed using simulations, the need for rapidly and accurately obtaining new optimized parameters to describe the chemical moieties becomes pronounced.

The OGCE workflow and middleware infrastructure are integrated to construct, execute and monitor coupled computational models, these tools and the integration efforts are further described in Section 4.1. The workflow tools interfacing with TeraGrid resource provide the needed computational capabilities to generate new and complementary data from calculations. The OGCE workflow tools wrap the above described command line applications in web services which are used as activities within the workflow. Further in the paper, we discuss how these requirements are imple-

mented driven by a proof of concept workflow of coupling two computational chemistry models.

3. GATEWAY ARCHITECTURE

ParamChem gateway is being implemented as layered architecture with functional components divided into major sub-systems as illustrated in Figure 2:

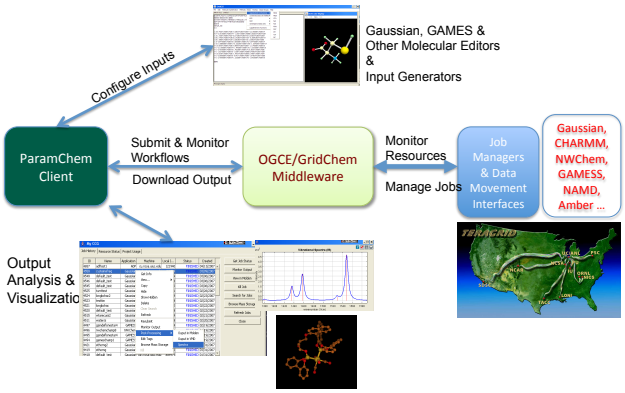

Figure 2: ParamChem Architecture

1. A user facing interface through to interact with the Cyberenvironment, named ParamCAD3D.
2. A back-end system to run computationally intensive calculations, consisting of
   (a) servers to provide initial atom typing and for generating initial guesses for (missing) parameters for a given molecular system
   (b) TeraGrid/XD resources for reference data generation.
3. A database from which experimental and quantum mechanical reference data as well as parameters of existing popular MM and SE Hamiltonians can be retrieved. The reference data and newly generated parameters will also be cataloged for further dissemination reuse.
4. Middleware that orchestrates communication between the three layers of the gateway infrastructure.

The system will steer the user through the following steps to execute parametric workflows:

1. As a one time step, construct and register applications and workflows with workflow registry
2. Browse workflows and select a workflow to be launched
3. Integrate with NanoCad to configure various parametric inputs and the molecule files will be uploaded to the staging area on ParaChem's TeraGrid hosted virtual machine gridchem.gateway.iu.teragrid.org.
4. Submit workflow to Workflow Engine Service
5. Monitor workflow in XBaya or polling workflow monitoring database.
6. Advance users will be able to interact with the workflow steering the execution.

The remaining sections of this paper review these steps.

3.1 Front-end Graphical User Interface

ParamChem has adopted the graphical user interface (GUI) of the GridChem project and enhanced it with additional features that are required for parametrization. We describe the ParamChem user interface in greater detail in this section.

3.1.1 Molecular editor

A new molecular editor, ParamCAD3D, has been implemented for the internal coordinate characterization and property assignment. Firstly, a Human Interface Guidelines (HIG)[15] document is created with inputs from a representative end-user (scientist) who will potentially use the parametrization specific molecular editor. Based on the HIG, a Java OpenGL based graphical application with the possibility to select internal coordinates is developed as illustrated in Figure 3. In addition to ParamCAD3D, several other molecular editors will be available through the ParamChem (and GridChem) client, including the "NanoCAD Classic" molecular editor, which was extended according to the HIG suggestions. These interfaces are currently available as downloadable stand alone components. In the future revisions we will integrate these tools with workflow execution and experiment management interfaces.

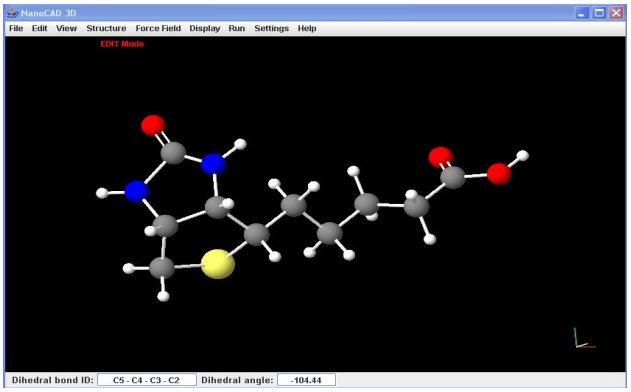

Figure 3: ParamCAD3D Molecular Editor

3.1.2 Atom typing, parameter and charge assignment infrastructure

Atom typing is a fundamental and critical first step in parameter optimization. In any molecular system the constituent atoms need to be assigned an identity of exchangeable or reusable chemical moieties. These moieties are defined to be used across a large set of molecular systems. These moieties may contain certain extensions to atomic descriptions. A programmable atom typer was developed and deployed to provide an improved, flexible and extensible infrastructure over more conventional approaches to atom typing by its programmable nature and by exploiting the hierarchical organization of the atom types in force fields such as the CHARMM General Force Field (CGenFF) [25], as demonstrated in Figure 4. A Sample Rule generated with this interfaces is illustrated in the Listing 1.

This atom typer was integrated with the ParamCAD3D molecular editor. In order to support a wide variety of force fields, different sets of atom typing rules can either be ported to the CGenFF atom typer, or implemented in the more conventional atom typer of the "NanoCad Classic" molecular

editor, which was extended to allow a wide variety of criteria for atom typing. After atom types are assigned, the next step in parameter optimization is the generation of initial guess parameters and charges. The resulting CGenFF program outputs per-parameter "penalty scores" that reflect the quality of the generated parameters and charges as shown in Figure 5 and the resulting toppar stream is illlustrated in the Listing 2. This functionality is currently integrated with the ParamCAD3D molecular editor. Additionally, the atom typing and assignment of parameters and charges by analogy is publicly accessible from ParamChem.org following a brief registration.

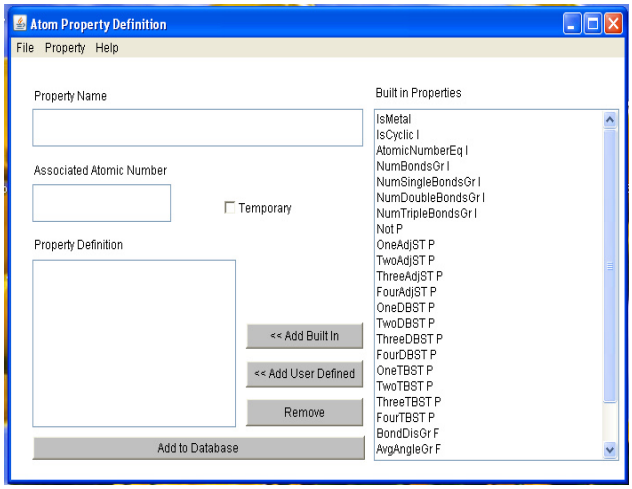

Figure 4: Example Nanocad Classic Atom Typer

Listing 1: Sample CGenFF atom typer Rule generated using the Atom Typer Interface

```
cat NG_
typ NG1T1 : ne ( bo 3 el C ) ( bo 1 ) err
        "nitrilium ion"
typ NG1T1 : ne ( bo 3 el C )
typ NG1T1 : ne ( bo 3 ) err
         "triple bonded non-nitrile nitrogen"
sub NG_2 : ne ( bo 2 )
sub NG_3P : nb 4
sub NG_N : nb 3
typ NG301 : err
        "negatively charged
                single bonded nitrogen"
end
cat NG_N
#Amidinium and guanidinium
typ NG2P1 : ne
        ( bo 1 el C ne ( bo 2 el N nb 4 ) )
#Single bonded neutral
        nitrogens that are sp2
sub NG_AM : ne
        ( bo 1 el C ne ( bo 2 elos ) )
typ NG2S3 : ne ( arom 6 )
typ NG2S3 : ne ( el P ne ( bo 2 el O ) )
typ NG2S3 : ne
        ( el P ne ( bo 1 el O nb 1 ) )
typ NG2RC0 : arom 5 arom 6
typ NG2R51 : arom 5
typ NG2R61 : arom 6
#Single bonded neutral
        nitrogens that are sp3
typ NG3N1 : ring 5 ne ( el N ) warn
        "5-membered ring hydrazine"
typ NG3N1 : ne ( el N )
typ NG3C51 : ring 5
typ NG331 : ne ( el H ) ( el H ) ( el H )
typ NG321 : ne ( el H ) ( el H )
typ NG311 : ne ( el H )
typ NG301 :
end
cat NG_AM
#The next one is pretty specific
        because of nasty exceptions
        like butyrolactam fused with pyrrole.
typ NG2R53 : ! arom 5 ! arom 6 !
        arom 7 ring23 5 ne
        ( bo 1 el C ring23 5 ne ( bo 2 elos ) )
typ NG2S2 : ne ( el H ) ( el H )
typ NG2S1 : ne ( el H )
typ NG2S0 :
end
```

Listing 2: Sample CGenFF generated Toppar stream

```
* Toppar stream file generated by
* CHARMM General Force Field
        (CGenFF) program version 0.8.5 alpha
* For use with CGenFF version 2b4-2
read rtf card append
* Topologies
RESI glycolal 0.000 ! penalty= 42.9
GROUP ! CHARGE Z EL NB
        NBE RNG1 TYP RNG2 TYP RNG3 TYP
ATOM C1 CG2O4 0.218 ! 6 C 3 4

ATOM O2 OG2D1 -0.401 ! 8 O 1 2
ATOM H3 HGR52 0.089 ! 1 H 1 1
ATOM C4 CG321 0.119 ! 6 C 4 4
ATOM H5 HGA2 0.090 ! 1 H 1 1
ATOM H6 HGA2 0.090 ! 1 H 1 1
ATOM O7 OG311 -0.624 ! 8 O 2 2
ATOM H8 HGP1 0.419 ! 1 H 1 1
! TYP INR
BOND C1 O2 ! 2 0
BOND C1 H3 ! 1 0
BOND C1 C4 ! 1 0
BOND C4 H5 ! 1 0
BOND C4 H6 ! 1 0
BOND C4 O7 ! 1 0
BOND O7 H8 ! 1 0
END
read param card flex append
* Parameters generated by analogy
*
ANGLES
CG2O4 CG321 OG311 112.00 122.50
        ! glycolal , from CG2O5 CG311
        OG311, penalty= 4.5
DIHEDRALS
OG2D1 CG2O4 CG321 OG311 0.0000 2 0.00
        ! glycolal , from OG2D3 CG2O5 CG311
```

```
        OG311, penalty= 14.5
HGR52 CG2O4 CG321 OG311 0.0000 3 180.00
        ! glycolal , from HGR52 CG2O4 CG321
        CG331, penalty= 42.9
CG2O4 CG321 OG311 HGP1 0.2200 1 0.00
        ! glycolal , from CG2O5 CG311 OG311
        HGP1, penalty= 4.5
CG2O4 CG321 OG311 HGP1 0.2300 2 180.00
        ! glycolal , from CG2O5 CG311 OG311
        HGP1, penalty= 4.5
CG2O4 CG321 OG311 HGP1 0.4200 3 0.00
        ! glycolal , from CG2O5 CG311 OG311
        HGP1, penalty= 4.5
END
```

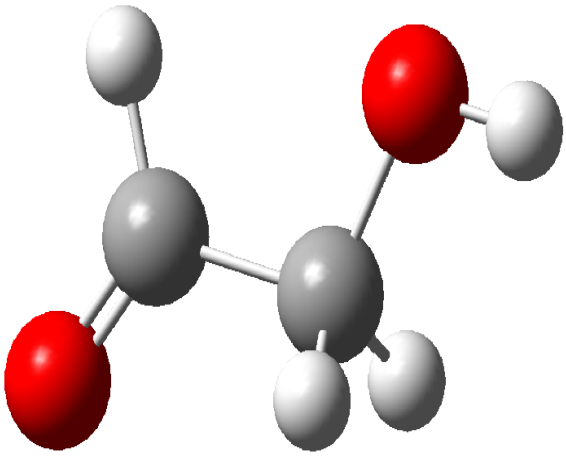

Figure 5: Sample CGenFF output with atom typing and assignment of parameters and charges by analogy

3.2 Computational Layer

3.2.1 Hardware Resources

ParamChem middleware is extended from GridChem [3], and has capabilities to interface with TeraGrid computational infrastructure. For workflow prototyping NCSA Ember HPC system [13] is primarily used. Eventually, other TeraGrid resources like San Diego Trestles, Pittsburgh Blacklight and Texas Ranger clusters will be certified in testing and made available as friendly user mode. The atom typing services are currently deployed on a ParamChem server at University of Kentucky and are integrated into the resource pool and seamlessly used for atom typing and initialization purposes. These components are currently deployed independent of the workflow system. As the next steps, the atom typing tools will be integrated with Workflow Infrastructure.

3.2.2 Computational Chemistry Software

To perform the computations needed for parametrization of molecular mechanics force fields in general and CGenFF [25] in particular (example: energy minimization, frequency calculations, determination of interaction energies, etc), CHARMM [10] MM software is utilized. CHARMM version 'c36a3' is installed on the Ember built as native MPI parallel as well serial executables. Also, QM software, Gaussian 09 rev B01 deployed on Ember are integrated. The Gaussian model is used for the quantum chemical reference data generation and is coupled to the CHARMM using workflow management system as described in 4.1. In addition to CHARMM and Gaussian, the Amber software [26] suite provides infrastructure for developing and optimizing parameters in the context of the GAFF (General Amber Force Field) force field using program modules such as antechamber, parmchk and tleap [26]. Version 11.0 of Amber suite has been obtained from Amber.org initially for this purpose but has since been licensed and is open to academic participants on NCSA HPC systems.

3.3 Metadata Formats and Services

ParamChem XML-based metadata schema [16] is developed to organize existing parameter data in a generic form. In particular, the scheme will assist parameter optimization workflow schema, the infrastructure for parameter specification, generalization and organization. This schema is already being used for the communications between ParamCAD3D and CGenFF program for atom typing and assignment of initial parameters and charges by analogy, and will ultimately be used for all transfer and storage of molecular data and parameters. Target data will be organized in a similar fashion in relational databases deployed in a mysql database.

3.4 Gateway Middleware & Security

The ParamChem middleware is co-opting the GridChem science gateway infrastructure [3] and building upon tools provided by the Open Gateway Computing Environments [19]. ParamChem's middleware consists of a workflow management server and a data management server. The GridChem data sub-system will be extended to incorporate more of workflow support and will reuse the functionality of simple job management and user management. OGCE-GridChem Bridge service [14] is developed to integrate the generic OGCE workflow system [17] with ParamChem/GridChem middleware.

ParamChem middleware uses TeraGrid MyProxy [1] based authentication for access to computational allocations shared with the GridChem project. The workflow system also authenticates to the MyProxy service and catalogs all workflow executions through the bridge service. The middleware to integrated with the TeraGrid science gateway attribute based authentication based on GridShib and SAML tokens[21]. Using this system end user information is automatically propagated to the TeraGrid audit databases and all computational jobs proxied through the community account are associated to the end users. OGCE tools use Globus [4] GridFTP for file management and GRAM for job management. Architectures for GridChem Middleware is further described in [27], and the workflow services are described in [17].

4. TERAGRID & OGCE COLLABORATIONS

4.1 Workflow Integration

Scientific workflow systems provide a programming model that allows the scientist to program experiments using service-oriented architecture that abstract the complexities of the underlying middleware. Scientific workflow systems act as

the point of interaction for the domain scientist with the gateway system. While implementing the workflow infrastructure for ParamChem project, Pamidighantam lab requested Advanced User Support through the TeraGrid ASTA program. Through peer review process the project was awarded the advanced support and the Indiana University Science Gateways Group was tasked to provide workflow management capabilities enhancing the existing GridChem single job submission system [18].

The OGCE workflow suite provides long running application support using asynchronous web services standards. The OGCE workflow system in advanced mode can be made to adapt to the dynamic changes during executing allowing the users to steer the application based on the intermediate results of the experiment. Most importantly, the workflow system is built upon a standardized workflow language WS-BPEL [22] and leverages community driven workflow enactment engines like the Apache ODE [5]. As shown in Figure 6 the workflow suite is flexible enough to leverage advanced workflow capabilities from various Scientific Workflow Engines. The OGCE suite will abstract the construction, deployment & monitoring differences between various workflow enactment engines providing the gateway users with a uniform interface. Further, the workflow suite provides seamless integration with gateway desktop and portal based interfaces for configuring, launching and monitoring workflow executions.

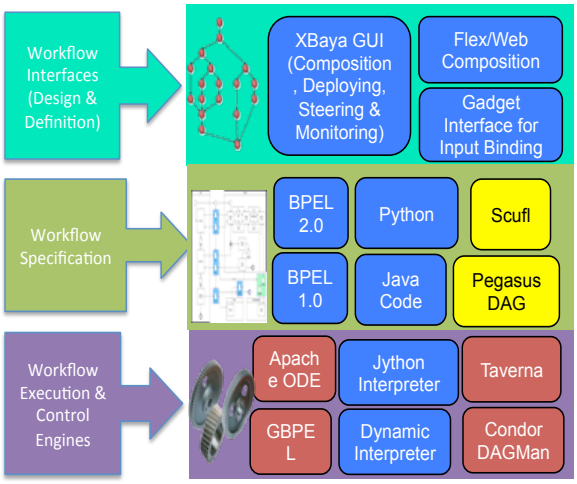

Figure 6: OGCE Workflow Suite Block Diagram

To provide the application and resource abstraction, OGCE tools are used to wrap applications as services [7], place services in a registry for discovery, and provide comprehensive messaging. As illustrated in the Figure 7, the GFac toolkit provides abstraction to various execution systems including ssh authenticated submission to campus resources, Globus based Grid Resources, Condor Submissions and EC2 API compatible Cloud resources like Amazon EC2 & EUCALYPTUS. Further details about the GFac Tools and the workflow suite are discussed in [17]. The system has been used extensively in other gateways and has undergone rigorous testing on the TeraGrid [11].

The team with ASTA and GridChem members has evaluated several workflow technologies such as Taverna [6], Pegasus [2] and [8]. The XBaya Workflow Suite is integrated with the gateway and the software is supported by Indiana

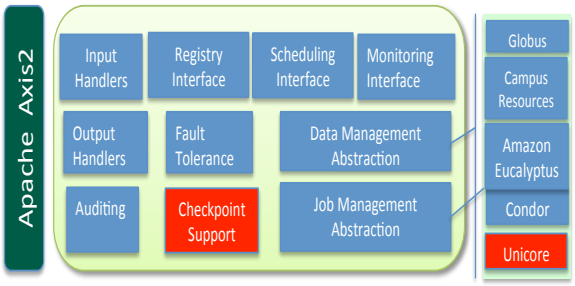

Figure 7: GFac Block Diagram

University through the OGCE project. The OGCE integration is being carried out in two paths: a loosely coupled path and a tightly coupled path. The loose coupling integration path has been completed through the ASTA support and provides the XBaya workflow construction, execution and monitoring capabilities in the GridChem/ParamChem user interface. However, the input configuration files are independently staged using GridChem client. The tighter coupling of ParamChem and OGCE us integrated through the web services based next generation GridChem middleware infrastructure. This effort is funded as part of the renewed OGCE Grant award.

A proof of concept parametrization workflow coupling CHARMM and Gaussian models is prototyped. As discussed in Section 2, the Pamidighantam lab implemented CHARMM-Gaussian workflows developed in the MacKerell lab using OGCE XBaya tools, as shown in Figure 8.

4.2 TeraGrid Advanced Support Accomplishments

The advanced support program accomplished the following tasks for the GridChem and ParamChem Projects:

- CHARMM-Gaussian workflow was implemented using the Open Gateway Computing Environment's workflow tools as described in Section 4.1. The task involved wrapping Gaussian and CHARMM models as web services, constructing the workflow graph and executing the jobs on TeraGrid resources.
- A new service called OGCE-GridChem Bridge [14] was developed to monitor and catalog workflow progress with ParamChem/GridChem data subsystem.
- Deployed a production-like instance of GridChem and a development environment of Indiana University Gateway Hosting environment
- Integrated Science Gateway Attribute based Authentication.
- Developed testing and monitoring components to roll out the workflow components into production environments.

5. CONCLUSIONS

In this paper the ParamChem project which is developing an integrated cyber environment to address the simulation needs of automated molecular parametrization is described. The project will specifically provide reference data organizers and generators as well as workflows for automatic parameterization of Molecular Mechanics Force Fields as well as Semi-Empirical methods.

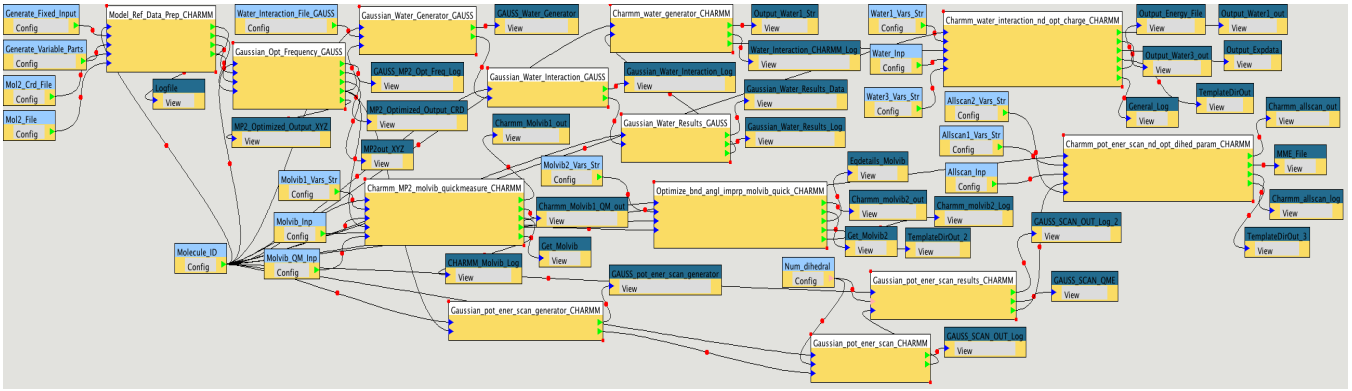

Figure 8: Example CHARMM-Gaussian parameter optimization workflow in XBaya Interface

A detailed description of the workflow integration efforts through the TeraGrid advanced user support program and collaboration between ParamChem and the OGCE Project are provided. The current version of the ParamChem interfaces with the CGenFF service hosted at paramchem.org to define and initialize parameters to be optimized further. A dedicated Java interface will give expert users tighter control over the optimization process, while the web interface will still be accessible to users to obtain reasonable initial guess based on analogy with a projected quality of the parameters to start a new simulation study and further optimization in house or though ParamChem infrastructure.

The paper also described how a proof of concept use case is derived from the ParamChem research goals. The model code is written to perform the needed parametrization. OGCE tools are used to wrap these models into workflow activities and the resulting workflows are executed managing jobs on remote TeraGrid resources.

6. FUTURE WORK

Through the renewed TeraGrid advance support and OGCE grant awards, the following ParamChem tasks are planned to be accomplished:

- Continued support of the ParamChem workflow development.
- Integrate with TeraGrid Information, Monitoring and Prediction Services to provide automated workflow scheduling binding jobs to resources with lowest start time and eliminating submissions to any clusters with non-working job-submission or data movement services. This service will mash information from exiting TeraGrid Information Services [12], INCA Monitoring Services [23] and Karnak queue predictions [24] and Speed Page bandwidth probes [20]. This integration will provide automated resource scheduling capabilities including near real-time middleware testing, information services and batch queue predictions.
- Develop a framework to assist with workflow cost estimation and reconfiguration. Enabling user interactivity during workflow configuration and submission.
- Refactoring GridChem client and middleware to monitor messages from the OGCE messaging system.
- Integrate GridChem account management with OGCE Security infrastructure
- Enhance existing single job GSISSH based submissions to use general purpose job management using Globus Gram.

7. ACKNOWLEDGMENTS

The authors would like to thank Nancy Wilkins-Diehr, Area Director for Science Gateways for cordinating the GridChem advanced support program. Further Rion Dooley (TACC) provided critical software development for the GridChem middleware and next generation architecture which is being enhanced for ParamChem development. Guarang Mehta and Karan Wahi for assisting in integrating Pegasus workflow system.

The National Science Foundation funded this work through various awards including: Chem-0823198 (CRI-CRIF:Collaborative Research: Extensible Cyberenvironments for Empirical and Semiempirical Hamiltonian Parameter Optimization and Dissemination), SCI-0503697 (ETF Grid Infrastructure Group: Providing System Management and Integration for the TeraGrid), OCI-0721656 (SDCI NMI Improvement: Open Grid Computing Environments Software for Science Gateways) and OCI-1032742 (SDCI NMI Improvement: Open Gateway Computing Environments - Tools for Cyberinfrastructure-Enabled Science and Education).

8. ADDITIONAL AUTHORS

Additional authors: Alex MacKerell (University of Maryland, `alex@umaryland.edu`), Michael Sheetz (University of Kentucky, `rmshee0@uky.edu`), Marlon Pierce (Pervasive Technology Institute, `mpierce@cs.indiana.edu`), Narendra Polani (University of Kentucky, `narendrakumar.polani@uky.edu`), Vikram Gazula (University of Kentucky, `gazula@uky.edu`)

9. REFERENCES

[1] J. Basney, M. Humphrey, and V. Welch. The MyProxy online credential repository. *Software: Practice and Experience*, 35(9):801–816, 2005.

[2] E. Deelman, J. Blythe, Y. Gil, C. Kesselman, G. Mehta, S. Patil, M. Su, K. Vahi, and M. Livny. Pegasus: Mapping scientific workflows onto the grid. In *Grid Computing*, pages 11–20. Springer, 2004.

[3] R. Dooley, K. Milfeld, C. Guiang, S. Pamidighantam, and G. Allen. From proposal to production: Lessons learned developing the computational chemistry grid cyberinfrastructure. *Journal of Grid Computing*, 4(2):195–208, 2006.

[4] I. Foster and C. Kesselman. Globus: A metacomputing infrastructure toolkit. *International Journal of High Performance Computing Applications*, 11(2):115, 1997.

[5] T. Gunarathne, C. Herath, E. Chinthaka, and S. Marru. Experience with adapting a ws-bpel runtime for escience workflows. In *Proceedings of the 5th Grid Computing Environments Workshop*, pages 7:1–7:10. ACM, 2009.

[6] D. Hull, K. Wolstencroft, R. Stevens, C. Goble, M. Pocock, P. Li, and T. Oinn. Taverna: a tool for building and running workflows of services. *Nucleic Acids Research*, 34(Web Server issue):W729, 2006.

[7] G. Kandaswamy, L. Fang, Y. Huang, S. Shirasuna, S. Marru, and D. Gannon. Building web services for scientific grid applications. *IBM Journal of Research and Development*, 50(2.3):249–260, 2006.

[8] B. Ludscher, I. Altintas, C. Berkley, D. Higgins, E. Jaeger, M. Jones, E. Lee, J. Tao, and Y. Zhao. Scientific workflow management and the Kepler system. *Concurrency and Computation: Practice and Experience*, 18(10):1039–1065, 2006.

[9] A. Mackerell Jr. Empirical force fields for biological macromolecules: overview and issues. *Journal of computational chemistry*, 25(13):1584–1604, 2004.

[10] A. MacKerell Jr, B. Brooks, C. Brooks III, L. Nilsson, B. Roux, Y. Won, and M. Karplus. CHARMM: the energy function and its parameterization. 1998.

[11] S. Marru, S. Perera, M. Feller, and S. Martin. Reliable and Scalable Job Submission: LEAD Science GatewayÕs Testing and Experiences with WS GRAM on TeraGrid Resources. In *TeraGrid Conference*, 2008.

[12] J.-P. Navarro. Teragrid information services. Proceedings of the Grid Computing Environments (GCE),Reno, Nevada: 11-12 November 2007, 2007.

[13] NCSA. Ember cluster. `http://www.ncsa.illinois.edu/UserInfo/Resources/Hardware/SGIAltixUV`.

[14] OGCE. Gridchem bridge. `http://www.collab-ogce.org/ogce/index.php/GridChem_ASTA`.

[15] ParamChem. Nanocad hig. `https://www.paramchem.org/wiki/index.php/Nanocad_HIG`.

[16] ParamChem. Xml schema. `https://www.paramchem.org/wiki/index.php/Main_Page#Database`.

[17] S. Perera, S. Marru, and C. Herath. Workflow Infrastructure for Multi-scale Science Gateways. In *TeraGrid Conference, June*, 2008.

[18] M. Pierce, S. Marru, R. Singh, A. Kulshrestha, and K. Muthuraman. Open grid computing environments: advanced gateway support activities. In *Proceedings of the 2010 TeraGrid Conference*, pages 1–9. ACM, 2010.

[19] M. Pierce, S. Marru, W. Wu, G. Kandaswami, G. von Laszewski, R. Dooley, M. Dahan, N. Wilkins-Diehr, M. Thomas, T. Center, et al. Open grid computing environments. In *Proceedings of the Fourth Annual TeraGrid Conference*, 2009.

[20] PSC. Speedpage. `http://speedpage.psc.edu`.

[21] T. Scavo and V. Welch. A grid authorization model for science gateways. In *International Workshop on Grid Computing Environments*, number 3, 2007.

[22] A. Slominski. Adapting BPEL to Scientific Workflows. *Workflows for e-Science*, pages 212–230.

[23] S. Smallen, C. Olschanowsky, K. Ericson, P. Beckman, and J. Schopf. The inca test harness and reporting framework. In *Proceedings of the 2004 ACM/IEEE conference on Supercomputing*, page 55. IEEE Computer Society, 2004.

[24] W. Smith. A service for queue prediction and job statistics. In *Gateway Computing Environments Workshop (GCE), 2010*, pages 1–8. IEEE.

[25] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of computational chemistry*, 31(4):671–690, 2010.

[26] J. Wang, R. Wolf, J. Caldwell, P. Kollman, and D. Case. Development and testing of a general amber force field. *Journal of computational chemistry*, 25(9):1157–1174, 2004.

[27] N. Wilkins-Diehr, D. Gannon, G. Klimeck, S. Oster, and S. Pamidighantam. TeraGrid science gateways and their impact on science. *Computer*, 41(11):32–41, 2008.