2nd Workshop on Challenges in Next-Generation Analytics for the Future Power Grid

Workshop Report

1/13/2014

We report on the findings and recommendations of the 2nd Workshop on Challenges in Next-Generation Analytics for the Future Power Grid. This invitation only workshop was held on September 19th and 20th, 2013 in Seattle, WA. Power grid research experts from industry, academia, and government were invited. This report also summarizes key challenges facing the industry. Building on work from the first workshop in 2012, we had detailed discussion in the areas of decision support, numerical library, and software infrastructure architecture. We also discussed the role of different community models for creating and maintaining software. Specification recommendations for next steps are also made.

1.0 Executive Summary

The first workshop on Challenges in Next-Generation Analytics for the Future Power Grid, held on November 29 and 30, 2012 in Seattle, WA, recognized the challenges and validated the needs for concerted efforts by the community to advance the field of power grid analytics. The challenges are a result of the grid transition from its traditional paradigm to a future one that is more dynamic, stochastic, and more data-dependent for effective and efficient operation and planning. The confluence of the smart grid development and exploding data availability leads to three fusions in the power grid transition, namely, the *Fusion of Grid Networks and Data Networks*, the *Fusion of Transmission and Distribution*, and the *Fusion of Operation and Planning*. This transition results in an urgent need for power grid tools to scale up to handle bigger data, manage more complex systems and solve more complex models for ensuring reliability and improving efficiency. Major findings from the first workshop included:

- A power-grid-specific open software architecture allowing developers to focus on producing advanced and innovative solutions rather than being overwhelmed by the complexity of parallel computers and programming details. This architecture would also allow interoperability among various tools and methods supporting the three aforementioned fusions. Such grid-specific software architecture includes a software structure and a parallel numerical library to enable parallel simulation and a visualization framework to translate simulation results into actionable information.
- A community is needed to build such an open software architecture. This ambitious endeavor requires engagement by researchers, vendors, and practitioners. Research and development can be more productive and impactful if efforts are guided and coordinated by an accepted, grid-specific development architecture. The U.S. Department of Energy can play a major leadership role in fostering and facilitating such a community.
- It is important to define business models for the evolution and support of such an open architecture. Power companies depend on vendors to transition software and methods into commercial tools to ensure robustness and necessary training and technical support. Vendors would only do so if they saw commercial value from open source software and methods.
- Numerical libraries are critical to help overcome the barriers to parallel computing for the power grid. Such libraries need to be carefully designed so they offer a broad array of capabilities that are useful and easily exploited in a whole range of power grid models.
- Visualization for decision support is the important last mile for any methods and tools but has been underappreciated in the research community because of its applied nature. It is important for the community to advocate for its importance, share visualization experience, and attract researchers to work on this area. Examples include founding new journals for power grid visualization and decision support.
- Compared to many other domains, power grid applications and power grid software tools are highly diversified. There is a need to survey today's grid software tools in terms of the use of advanced computing and identify common and essential elements that numerical libraries need to include to maximize the reusability of software codes.

- Power grid datasets including measurement data, system models and parameters are sensitivesensitive in terms of infrastructure security and commercial competitiveness. But researchers and developers need realistic datasets (vs. textbook small examples) for developing and validating methods and software tools. This is especially true for advanced computing tools, which usually need large datasets. Building such datasets is a priority.
- Building a community is a journey. The workshop participants recommended having follow-on discussions and bringing a larger group together in about six months to focus on developing an action plan.

Following this first, by-invitation-only, workshop, a second workshop was held on September 19 and 20, 2013, in Seattle, WA at the University of Washington. The workshop built on the findings from the first workshop and focused on how to accelerate the use of advanced computing for the power grid. The intent of this workshop was to generate actionable outputs to rapidly move power grid analytics forward, with broad industry and academic participation. Establishing common goals for power grid analytics will benefit the community. These goals can inform national discussions on this topic and be translated into roles for researchers, vendors, and users. Specifically, the objectives of the workshop were:

1. Build a shared vision for a software architecture, numerical library and visualization framework to tackle power grid challenges;

2. Identify research requirements and needs to deliver the vision;

3. Share progress since prior workshop; and

4. Foster a community to move forward with research in a coordinated and accelerated manner and to establish a business model for technology transfer and sustaining engineering of research software.

The fundamental computing challenge is how to establish a computing environment capable of ingesting data for fast computation to gain timely knowledge and use that knowledge to enable more effective and timely actions by operators and autonomous controls.

Advanced computing, ranging from small multi-core computers to clusters to large-scale leadership class machines, holds the promise for addressing many aspects of this challenge. But simply putting "big computers" against the problem is not the solution. The key is how to adapt power grid problems to suit advanced computing and to most effectively leverage the power of advanced computing as it becomes more ubiquitous and affordable. However, there are high barriers preventing a fundamental transition from traditionally-sequential-computing-based tools to a parallel computing environment because of the complexity of parallel programming across multiple platforms. A grid-specific advanced computing architecture would facilitate the transition. Such architecture would include a data and software structure, a parallel numerical library and a visualization framework to translate data into actionable information. The workshop was structured to facilitate discussions on such an architecture. Major findings and tangible organizational and technological recommendations from the workshop are:

1. A general consensus among the participants is that a software infrastructure such as GridOPTICS™ should be developed as open-source software by an open community. This community needs input and acceptance from end-user consumers of the software (utility companies), intermediates who market production-ready software (“vendors”), and researchers creating proof-of-concept prototypes. To continue engaging such a community, the workshop recommended that future workshops be held at a frequency of once or twice per year, with smaller discussions in between to strengthen the collaboration among various participants. In particular, the next workshop should be scheduled about nine months after this workshop.
2. A business model continues to be a key aspect that needs to be developed to facilitate the forming of the community, the development of the architecture, and the appropriate sharing of technologies amongst stakeholders. The business model would include the governance structure of the community and the funding model to support the community. The community governing board would seek funding support for community activities. In the initial stage of the community, funding support from the Department of Energy to bring the effort to the level that the industry would see the value and in turn adopt the technology and support the community efforts. This funding will both enable raising the technology readiness of research software but also enable development of features and engineering processes to support cross-organization interoperation necessary for effective sharing and reuse of components. Eventually, funding for ongoing activities should come from a consortium of vendors and end-user utilities in addition to support from the Department of Energy.
3. The GridOPTICS software architecture needs to facilitate interoperability of software tools for data, computation, and visualization and the integrated modeling of transmission and distribution for seamless operation and planning functions. The workshop recommended that the community develop a basic design and a reference implementation. The reference implementation can then be extended by vendors to support various power grid applications. In order to strengthen the community, we recommend enhancements to be contributed back to the reference implementation. The community should create and maintain a common shared-source repository accessible to researchers, software vendors, and end-users. Establishing a repository from existing assets is a necessary step.
4. As part of the architecture, decision-support tools should be designed by using a human- or practice-centered approach. Cognitive systems engineering (CSE) takes a multidisciplinary, practice-centered approach with the goal of guiding the design of complex, computerized systems intended to support human performance. By applying a variety of methods to understand and support human cognitive performance (e.g., problem solving, judgment, decision making, attention, perception, and memory), tools can be designed and built to aid operators in accurately assessing the situation, gaining and maintaining situation awareness, and supporting them in successfully, efficiently, and safely conducting their work. Effort should be made to expand awareness of such techniques and successful patterns. We recommend creating forums such as workshops to this end. We further recommend creating venues to enable testing of new technologies with real-world data streams and models.
5. Predictive capabilities are recognized to be an important foundation for the future power grid. In addition to accelerating computation, new research on uncertainty quantification and

stochastic analysis is needed to provide confidence in predictions. This new research spans modeling, simulation, and decision making, and should be considered in the development of methods and tools from the design stage of the GridOPTICS architecture. Such predictive capabilities will benefit a wide range of applications from planning decisions, to operational decisions, to automation of control. They must be reliable enough for production use and robust enough to reduce cost or improve operational security. We recommend a research program be established by the community guided concretely by problems of high end-user importance and with an emphasis on understanding robustness of techniques in operational contexts.

6. In order to expose GridOPTICS™ and other proof-of-concept software to real world conditions, we recommend creation of a realistic control-center-like testing facility. This facility must receive timely power industry streaming data which can be used to create a test bed for grid-specific toolsets and applications built on top of it. Such a testing facility would satisfy the venue needed for new decision support tools by enabling operators to train with new tools and provide feedback to software developers, a feedback loop that would result in better, more usable tools. It would also provide an environment to test new strategies and techniques for visualization and alerting of operators to emerging operational conditions. As a resource shared with the research community, it will permit end-to-end demonstrations where new components can be studied in an operationally-complete environment. This facility would enable a co-design process that ensures proposed software tools are tested with operators to improve the usability of such tools.
7. Central to making measurable and credible progress is that the community defines a number of benchmark "challenge" problems that concretely and formally define success. These need to be vetted by the industry members of the community so that the benchmark problems can indeed be used to assess the impact of new technologies on the reliability of future power systems and on their economic feasibility to deploy and maintain. These problems will need datasets to support testing and careful consideration of operating requirements and the range of acceptable solutions. Delivering on a challenge problem is likely to require integrated innovations in all the three areas (i.e. data, computation, and visualization) discussed in this workshop.

2.0 Introduction

This report documents discussions, findings, and recommendations of the *2nd Workshop on Challenges in Next-Generation Analytics for the Future Power Grid*, held in Seattle, WA, USA on September 19th and 20th, 2013. This workshop built on the results of the first workshop held November 29th and 30th, 2012 at the Battelle Seattle Research Center and was motivated by the rapid pace of change in the power grid in two areas. First, an information revolution is taking place in the grid, resulting in large amounts of high frequency data from new measurement devices such as Phasor Measurement Units (PMUs) and smart meters. Second, physical power grid components are evolving to encompass new distributed generation sources that introduce dynamic and stochastic behaviors. The confluence of these two advances is leading to the grid undergoing a transition that encompasses three fusions, namely, the

Fusion of Grid Networks and Data Networks, the *Fusion of Transmission and Distribution*, and the *Fusion of Operation and Planning*. This transition is resulting in an urgent need for power grid tools to scale to handle bigger data, manage more complex systems and solve more complex models for ensuring reliability and improving efficiency.

The changing requirements and technology opportunities require significant research. The following areas were specifically called out as part of the first workshop:

- Software tools that address data intensity and can predict power grid behavior increasingly rely on new measurements and real-time data for monitoring and operation. Many of these measurements are high-speed streaming data gathering from geographically distributed sensors. Processing this data in a timely and reliable fashion requires scalable, adaptable software architecture to transfer, ingest, and manage very large amounts of data and coordinate computational processes.
- Power grid models need to have higher resolution of the end-to-end system over a much wider geographical area. This is due to changes in power generation and load, as well as an increased need to gain more efficiency out of the system. This requires new mathematics and computing methods that will be able to solve much larger models in real time.
- Information visualization and decision support is critical for converting large amounts of data into actionable information, enabling real-time operation, wide-area coordination and collaboration. Advances in these areas will drive new analytic methods, faster solutions, and new means of providing decision support.

The fundamental computing challenge is to establish a computing environment capable of bringing data to fast computation and quickly translating that into knowledge; and then, using that knowledge to enable more effective and timely actions by operators and autonomous controls.

Advanced computing, ranging from small multi-core computers to clusters to large-scale leadership class machines, holds the promise of addressing many aspects of this challenge. But simply putting “big computers” against the problem is not the solution. The key is how to adapt the power grid problems to suit advanced computing and to most effectively leverage the power of advanced computing as it becomes more ubiquitous and affordable. Today’s power grid tools are primarily designed for the sequential computing environment. Moving them to a parallel computing environment requires a software architecture that allows developers to focus on solving the problem rather than being overwhelmed by the complexity of parallel computers and programming details. Such a grid-specific advanced computing architecture includes a data and software structure, a parallel numerical library and a visualization framework to translate data into actionable information. The workshop was structured to facilitate discussions on such an architecture as well as key components of the software and use cases that drive requirements.

The workshop was structured to allow maximum participation and discussion of the attendees. A plenary session was followed by three parallel working sessions. The plenary session began with an address by Gil Bindewald from DOE’s Office of Electricity Delivery and Energy Reliability. Gil discussed power grid analytics challenges and the importance of advanced computing for the next-generation power grid software development. Then Esa Rantanen from the Rochester Institute of Technology spoke on lessons learned regarding human factors in operational decision-support from different domains. His focus was on aviation but he also discussed how those specifics are generalized to other areas. Finally, John McEntire, from PNNL, spoke about software intellectual property issues with an emphasis on the diverse business models that interact around shared and open source software.

Parallel sessions were focused on more narrow topics but included an open discussion on the current challenges. Each session had a lead domain expert to facilitate discussion and a scribe to take notes which are the primary basis for this report. The sessions were on the following topics which are elaborated in more detail below.

- Software infrastructure: data management, workflow support, and numerical methods. Data sets, what data sets are needed, how they can be collected or synthesized, curated, and shared.
- Numerical library: Library design, parallel computing techniques, and mathematics to enable more efficient computation.
- Operational decision support: visualization and other techniques to enhance situational awareness and decision making with an emphasis on cross-organizational scenarios.

GridOPTICS™ is a prototype implementation of a set of next-generation grid-specific data tools which span these three areas and was developed in part in response to the findings of the first workshop. It includes a prototype software infrastructure, work on numerical libraries, and experiments in decision support. These facets were part of the background information that was used to seed the discussions in the breakouts and to provide a model system for discussion. In many parts of this document we will use it as an example of the kinds of tools which will be required with the understanding that it will take a community-wide effort to make it or some similar software base complete and mature for production use.

The remainder of this report summarizes breakout session discussions and presents major findings from the workshop. The report includes the following sections: (3) an overview of the challenges to be addressed, (4) a report on the break out discussion on operational support, (5) a report on the breakout discussion on numerical libraries, (6) a report on the breakout session on software infrastructure, (7) a summary of the uses cases discussed in the three breakout sessions, (8) a summary of the discussion on community building, (9) recommendations, and (10) concluding remarks. The appendix includes a list of attendees, the agenda of the workshop and acknowledgement of the planning committee.

3.0 Challenges

The power grid industry is undergoing constant transformation. As shown in Fig. 1, the industry has moved from the early days of classic utilities that were vertically integrated with cost- and physical infrastructure-based operations; to the competitive era with open transmission access and wholesale electric markets; to today's emerging smart-grids with distributed intelligence, service valuations, and producer-consumer choices; and to the future smart cities that are sustainable, resilient, and wellconnected. Prior to competition, utilities behaved as loosely-connected entities. Now and in the future, connections are tighter. Blackouts will spread across borders. Utilities can no longer operate in isolation.

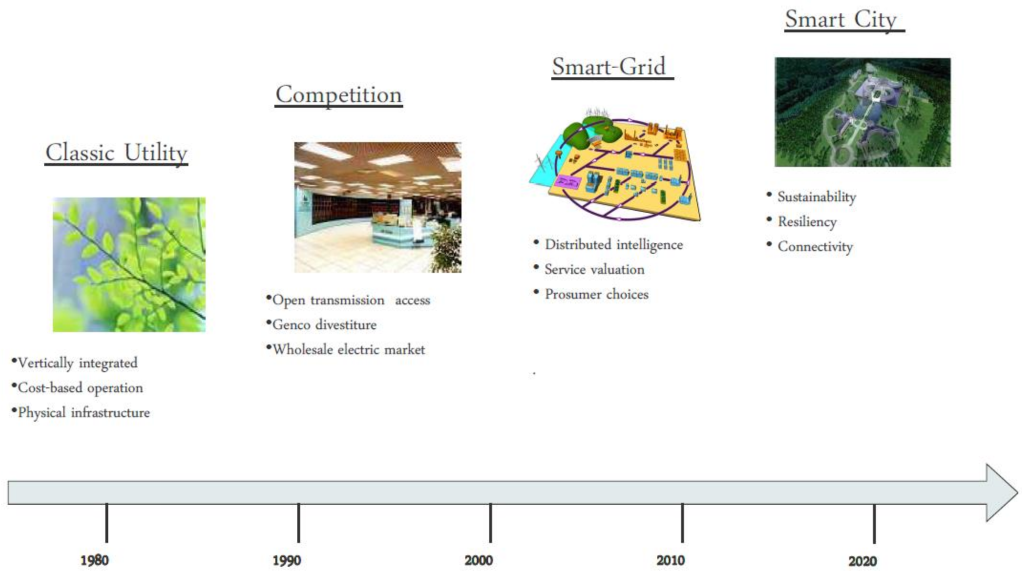

Figure 1. Power grid industry transformation.[1]

Industry transformations have resulted in greater complexity, interactions, and interconnectivity in power grid operations. Operators are relying less on intuition and more on analytics to make critical decisions and facilitate effective control. The industry trend is to bring more engineers into the control room. For the Western Electricity Coordinating Council (WECC), for example, engineering is now a 24/7 support function.

As shown in Fig. 2, the complexity of the power grid is expanding along certain dimensions. It is growing in degree of distribution as more devices are deployed and control decisions are distributed. It is expanding in timeliness and temporal requirements as accurate forecasting becomes more challenging and critical, and response times need to be shortened to deal with blackouts and emergency events. It is growing in uncertainty as more erroneous data and inaccurate complex models are produced, which impacts forecast accuracy. For decision support and visualization systems, one challenge is how to convey both time and geospatial information in intuitive displays. Another challenge is how to describe, convey, and understand uncertainty.

[1] Figure courtesy of David Sun as part of his break-out session presentation.

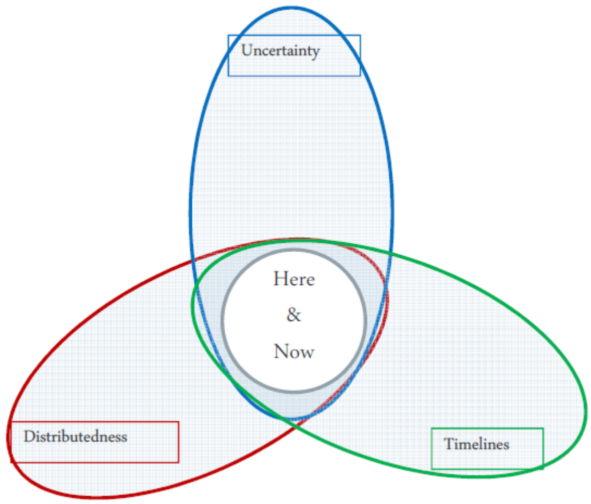

Figure 2. Power grid dimensions of complexity.[2]

Decision support may be viewed as an iterative process in which models and data are transformed into knowledge and actionable results as shown in Fig. 3. Decision support tools carry out the integration and transformation of models and data into results. Feedback and iteration is crucial in refining the results and updating them to reflect new information and circumstances. Decision analytics is required for both real-time grid operations as well as long-term planning with different requirements and constraints. As decision analytics is conducted, the operator or planner must continuously verify whether the right problem is being solved and the right data is being utilized.

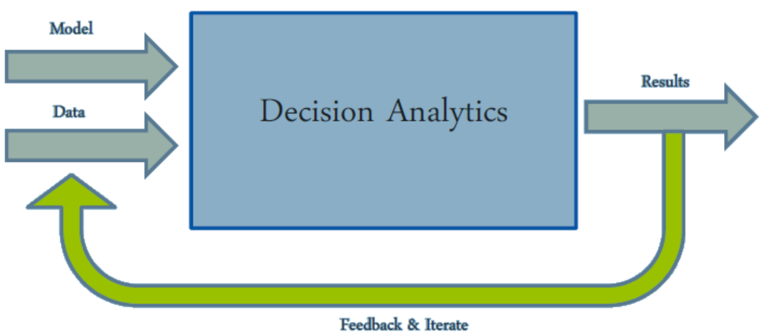

Figure 3. Decision analytics process[3]

Visualizations are ultimately applied as input into human decision making. The visualizations shown on a display are the end result of a potentially detailed process that begins with underlying models and large amounts of data. The dividing line between visualization, data analytics, and traditional power system

---

[2] Figure courtesy of David Sun as part of his break-out session presentation.
[3] Figure courtesy of David Sun as part of his break-out session presentation.

analysis applications are often blurred. A key challenge for both decision support and visualization systems is to ensure that the underlying analytics and models are behaving as expected and not producing glaring errors that will go unnoticed. Verification and validation are essential for data, models, and tools.

3.1 Decision Support and Visualization

Utilities can no longer operate in isolation. Blackouts and emergency events often extend beyond regional boundaries, providing impetus for neighboring utilities to share models and data. In practice, transmission information is commonly shared, but generation and outage data is often withheld due to competitive reasons. As more complex models and larger, more varied data are shared, it becomes harder to keep models and data in sync with reality. There are also open questionsquestions of how models should be validated, curated, and shared among organizations.

The sharing of near-miss black-out or emergency event information would also be valuable. Currently, utilities will document and review near-miss events, but do not share them with other utilities. This limits the ability of the open research community to learn from this data and validate new techniques.

While model and data sharing is more relevant to transmission networks today than distribution networks the links between transmission and distribution networks will grow more critical as transmission models penetrate further into distribution systems.

Challenges for decision support and visualization systems include:

- how to convey both time and geospatial information in intuitive displays
- how to describe, convey, and understand uncertainty, including uncertainty in data
- how to ensure underlying analytics and models are behaving as expected and not producing glaring errors that will go unnoticed
- how to verify and validate data, models, and tools
- how to design for extreme conditions or “black swan” events
- how to chart a course back to more normal conditions
- how to maintain consistency in interfaces and uses as users deal with different and changing problems, tasks, and collaborators
- how to accommodate differences in technology implementation between utilities
- how to allow tools to migrate across differently configured control rooms
- how to decide which staff have the ability to make changes to the analytic environment

3.2 Numerical Libraries

Key challenges for developing numerical libraries include multiple areas of the lifecycle of these libraries from basic research, software requirements, software development support, and maintenance of deployed software. Some high-points from the discussion at the workshop include:

- Performance (Robustness and consistency): For modeling and forecasting purposes, the requirements for scalability and performance may be different than for real-time control of power systems. These requirements need to be established and vetted with industrial consumers of the technologies.

- Verification using test cases and quantifiable metrics: Metrics for successful testing must be developed that will guarantee validated codes and will provide correct and valuable results for real world problems. Such test cases provide a formal statement of requirements.
- Verification using realistic synthetic test cases (Real environment): Many current test cases are based on real-life systems and have issues with access. Synthetic test cases that are widely available and do not have proprietary and/or security issues associated with them are needed
- Secure software: As complex software becomes an indispensable aspect of power delivery, appropriate software engineering practices should be used to ensure the development of secure software.
- Benchmarking, performance profiling for various tasks: Benchmarks that can be used to assess code performance across different platforms both in absolute terms (time to completion) and scalability are needed. Profiling tools and related diagnostic tools that can be used to assess different parts of an application and understand performance bottlenecks are also needed.
- Establishing a multidisciplinary community to develop libraries: Community development of software is important both to lower overall development costs and to standardize parts of the software and data stack to enable widespread use. Community development will also promote code robustness and efficiency of important components. Governance and support (both financial and in kind) are key issues for building a software development community.

3.3 Software Infrastructure

Key challenges for developing software infrastructure include:

- Access to high quality data sets. Significant barriers to data availability exist. These include security concerns, a lack of facilities for generating realistic simulated data, and lack of cooperation in constructing unified or interoperable data models
- Providing support for high speed streaming analytics. This implies that developers need to modify existing paradigms and consider how to translate existing simulation algorithms to the streaming analytic model
- Facilities to test software using streaming data
- Engaging end users (operators) in the software development process (co-design)
- Access to comprehensive use cases
- Building an open source software development community and addressing consequent needs for governance, and ongoing support of software.

4.0 Decision Support and Visualization

4.1 Areas where Existing Decision Support / Visualization Systems are Sufficient

In the context of control room and engineering support technologies, existing decision support and visualization systems work fairly well for over 99% of the time during routine operations or in cases where there are relatively small contingencies. Using existing technologies, operators are effective when the grid is performing normally or close to the projected scenario, or when dealing with normal disruptions such as lightning strikes that disable power lines or a generator going down. In such situations, a sufficient amount of time is available for the operator to consider and work problems. Day-ahead models are generally accurate and effective under near normal operating conditions. Incremental progress of decision support and visualization systems that support routine operations is a sufficient development approach.

4.2 Extreme Emergency Support

Many large-scale blackouts have timescales of several minutes to a few dozen minutes. These timescales allow for operator intervention, but this must occur quickly to be effective. During large-scale blackouts, operators also need rapid situational awareness to be effective. The control room environment during an extreme emergency situation is quite different from normal operating conditions. In an extreme emergency environment, advanced analysis tools may not function properly or be in states that are unfamiliar to the operator, a high level of stress is introduced, and many varied decision makers are required to collaborate.

Designing decision support and visualization systems for extreme conditions (often called "Black Swan" events) is challenging since operators seldom encounter the abnormal conditions. For example, in the August 14, 2013 East Coast USA / Canada incident, the contingency analysis tools produced a large number of alerts because of stressed system conditions (i.e., many arriving phone calls). Operators had difficulty in understanding and analyzing the contingency analysis results because they had never encountered so much information from that system. Furthermore, in extreme emergency situations, operators may have to rely more on intuition than analytics due to the need for timely action.

To support analysis in extreme emergency situations, decision support and visualization systems need to be able to chart a course back to a more normal condition. They also need to be timely – taking into account the time critical nature of making a decision in an extreme emergency context. To provide consistency, systems need to operate and be effective in both normal and extreme emergency conditions. Automated capabilities are needed in emergency situations to evaluate parameters and to determine and take extreme actions, but they must also continue to inform operators to provide situational awareness and avoid operator "out of the loop" scenarios. Furthermore, systems need to be able to provide features and mechanisms that allow operators to comprehend and act upon potentially thousands of entities or events such as violated contingencies. During the discussion of numerical methods, it was also observed that there is a need for new mathematical techniques for identifying critical but rare contingencies.

4.3 Dealing with Uncertainty

In extreme emergencies, operators have to deal with many uncertainties as they wade through an unfamiliar situation. Operators may lose trust in the normal decision support and visualization tools they use because the tools produce too much information or unintelligible or erroneous results. In the August 14, 2013 East Coast USA / Canada incident, operators lost trust in the state estimation tool because it flooded them with what the operators believed to be erroneous information. Operators may become suspicious of a tool if they believe it to be not smart enough or too smart.

Uncertainty exists in data as well. Bad data is abundant in the power grid. Decision support and visualization tools need to account for varying data rates and different data qualities and uncertainties. Questionable data needs to be cleaned up before it is presented to operators and before entering the decision support phase. Automated data cleaning methods would be valuable in replacing a usually manually-intensive process.

4.4 Models versus Model-Less

Physics naturally ties many of the elements of the power grid together. Power engineers are intimately familiar with physics-based electrical models as they are base representations in many of the analysis, decision support, and visualization tools they employ in their work. Such models are embedded in the consciousness of operators and planners and they think and act instinctively on these models.

Other kinds of models are also available to operators and planners such as retail and wholesale market models, different types of control models, weather or meteorological models, and geographical models. Other models of potential use for power grid analysis and decision support include models for telecommunications, computer systems, human in the loop, and social behavior. Such models have been developed for and applied in other domain areas, but have yet to substantially migrate to power grid analysis.

One open question is whether situations or problems exist where analysis and decision making should not involve complex models. An extreme emergency event is the most obvious situation where simpler model analysis and decision making would make the most sense, since time may not be available to understand and interact with complex models. Yet, in such instances, electrical models are still heavily used since their comprehension and understanding is so second-nature to power engineers.

4.5 Supporting Decision Making

To gain better trust and promote more effective use, decision support and visualization systems need to provide information and results that operators can more readily interpret. The limitations of the tools need to be documented and understood. Furthermore, tools need to support operator reasoning more directly. They need to be able to harness the expertise of the operator by providing a diversity of paths to a result, a variety of methods to look for anomalies or something out of place, and ways to triangulate on a problem. Interfaces and visualizations need to be configurable to display and highlight the right, critical information needed to make decisions.

Given information, an operator has to cross a mental barrier to declare an emergency. Decision support and visualization systems need to provide the operator with sufficient information to make that decision as well as to handle the emergency situation once it has arisen. Additionally, experienced operators will often follow a heuristic path of analysis regardless of what measurements and tools are reporting (e.g., the captain of the Titanic had 30 years of experience working against him). To counter cognitive inertia, decision support and visualization systems need to provide analysis results in a compelling way so as to break the operator from a singular line of thought. Furthermore, decision makers need to learn to recognize when they are following heuristics and to step back, apply analytics, and be more deliberate in generating decisions.

4.6 Transition Management

As previously mentioned, the power grid industry is undergoing constant transformation, but so are the models and tools, operators and planners, and operating environments. The design of decision support and visualization systems must account for many types of changes. With regards to information content, the volume and diversity of data will continue to grow. Change or event detection remains a valuable analysis approach where critical information is extracted based on changes that occur with respect to norms, time, scenarios, etc.

Users change over time as well. A particular user will evolve from a novice to an expert and apply decision support and visualization tools differently over time based on his/her growing domain knowledge and expertise and growing familiarity of analysis tools. Changes in the work environment are certain to continue as new roles such as 24/7 control room study engineers are defined and more and deeper collaborations across utilities and between transmission and distribution systems evolve. An important challenge for decision support and visualization systems is maintaining consistency in interface and use as users deal with different and changing problems, tasks, and collaborators.

Differences in technology implementation are also a challenge for decision support and visualization systems. Information and analysis tools are slow to change in any particular control room but may be significantly different across the control rooms of different utilities. The physical environment and layout may also be significantly different across control rooms. In industries that are strictly regulated, such as the airline and nuclear industries, consistency and uniformity in tools and environments are enforced. Whether such consistency and uniformity will emerge in the power grid industry is an open question. For decision support and visualization systems to be effectively deployed, one must consider how the tools might migrate across different control rooms and which control room staff members have the ability to make changes to the analytic environment. There is further a need to streamline the deployment of experimental tools to allow validation in real-world contexts and get design feedback from experienced operators.

4.7 Collaborative Decision Making

Blackouts and emergency events often extend beyond regional boundaries, providing impetus for neighboring utilities to share models and data. The sharing of various forms of data including text, videos, models, graphs, etc. is often desirable. In practice, transmission information is commonly shared, but generation and outage data is often withheld due to competitive reasons. The more models and data that are shared, however, the more work is required to keep the models and data in sync with reality. The question is open as to whether deregulation has improved or degraded model and data sharing among utilities.

Some progress in model and data sharing may be seen in the efforts of WECC. WECC makes study reports and data available to utilities and ISOs. They also intend to provide outage information in the future. Making available real-time applications and studies is also under consideration, where utilities may access real-time information specific to their regions. Currently, WECC runs different models such as state estimation, load factor calculations, and real-time sensitivity analysis for the full Western power grid. Through a centralized sharing model, WECC could allow utilities and ISOs to collect parts of external models and data to be stitched and integrated with local models and data. In contrast, a decentralized sharing model would involve utilities sharing models and data on their own. Whether a centralized or decentralized sharing model is more effective is another open question.

The sharing of near-miss black-out or emergency event information would also be valuable. Currently, utilities will document and review near-miss events, but do not share them with other utilities. Actual blackout reports, however, are generally shared across the industry.

Traditionally, model and data sharing is more relevant to transmission networks than distribution networks. On the distribution side, data is often bad, models may not be trusted, and thus, visualizations have limited value. In the future, however, the link between transmission and distribution will grow more critical as transmission models look more and more down into distribution systems.

4.8 Model and Tool Validation

A general approach to validate power grid models and tools is to tie them into simulations, dry run them with operators, and then collect and evaluate results to understand a model's or tool's effectiveness and usefulness. Methods for quantifying improvement to the analysis or decision making process need to be developed. A control center of the future facility or testbed would be useful in which to conduct usability evaluations of models, tools, and visualizations with real potential users. In such a facility, power engineers will need incentives to participate in studies such as receiving training credits. PNNL is currently the home to a number of power grid control room testbeds and would be a logical candidate for housing such a user facility.

A second approach to validation is simply to publish results in research literature and let the community at large evaluate and extend the results. More results need to be published in power journals and efforts need to be sustained towards building a stronger community.

4.9 Human Factors and Situational Awareness

Modern day large-scale electric power systems are characterized by large volumes of data, which must be transformed into useful and understandable information in an efficient manner to support operational and planning decisions. Also, during the operation of a large-scale power system, quick decisions are sometimes of the utmost necessity in preventing conditions leading to cascading failures and blackouts. Visually displaying real-time conditions can help power system operators maintain adequate situational awareness and respond in an expedited manner to conditions potentially threatening to system stability. As a result of the analysis of the August 14, 2003 blackout in the Northeastern United States and Canada, recommendations to develop adequate visualization and decision support tools for real-time system operators in a control room has stimulated widespread interest with universities, vendors, national laboratories and other organizations to get actively get involved in research in this area.

Representing the information required for operators to create and maintain situational awareness for making critical decisions and taking action given the specific situation is an essential task in maintaining the reliability of the electric grid. These tools are important in routine operations, but are a priority for providing notification in rapidly changing, non-routine situations.

The Human Factors discipline and the theory of situation awareness has existed since the mid-twentieth century, getting its start in the aviation industry. There are many theories and methodologies that have influenced the design of decision support and visualization tools for a multitude of disciplines. One workshop presentation highlighted some of the cognitive theories that may readily be used in the design of computer interfaces, including Rasmussen's Skills, Rules, and Knowledge (SRK) framework, Kahneman's Two-System View, and Hammond's Cognitive Continuum Theory.

Skills, Rules and Knowledge Framework (SRK). The skills, rules, and knowledge framework (Rasmussen, 1986) consists of three levels of behavior.

- Skill-based behavior (SBB) is fast, requiring little cognitive activity, and is accurate. Diagnostic troubleshooting at the skill-based level requires a direct match between the features of the problem observed and patterns that have been previously experienced and stored in long-term memory.
- Rule-based behaviors (RBB) consist of mental 'checklists,' and mental simulations of 'what if scenarios. Diagnosis at the rule-based level is done by applying sets of rules stored in long-term memory, e.g., a sequence of steps and procedures for doing so.
- Knowledge-based behavior (KBB) is explicit goal-controlled performance in unfamiliar situations where no rules or know-how are available. Iterative diagnostic testing and subsequent analyses are necessary for problem solving

Within the SRK taxonomy cognitive control may rely on a repertoire of automated behavioral patterns (i.e., SBB), a set of cue-action mappings (i.e., RBB) and/or problem-solving operations on symbolic representations (i.e., KBB). When applying this framework to the design of tools it is helpful to consider the experience level of the recipients of those tools and how domain-specific the interfaces need to be. For example, in the electric power domain, when designing for highly skilled and experienced operators who require system-specific computer displays, designers will target the skill-based and rule-based behaviors.

Two-System View. The two-system view (Kahneman, et al., 19xx) defines two types of cognitive processes: intuition and reasoning. In this theory intuition is evoked and generates a response particular to the situation at hand. Reasoning is then used to either evaluate the response generated by the intuition process or to perform effortful reasoning if there is no intuitive response. Operations of the intuition system are fast, automatic, effortless, associative, and often emotionally charged. The operations of the reasoning system are slower, serial, effortful, and deliberately controlled.

Cognitive Continuum Theory. Hammond's cognitive continuum theory includes two Meta theories: coherence and correspondence. Theories of coherence are similar to those that deal with the rationality of judgments and decisions and can be characterized by logical or mathematical consistence or the absence of contradictions. Theories of correspondence are concerned about the correspondence of judgment with empirical facts (i.e., empirical accuracy of judgment such as physicians' diagnoses). The psychological underpinnings of competence include both correspondence competence and coherence competence. Correspondence competence refers to such things as the remarkable perceptual abilities of humans, perceiving the 3D world that is projected on a 2D retina, as well as shape, size, color and constancy. Coherence competence is unique to humans and is an acquired skill. Different tasks require coherence competence than tasks demanding correspondence competence. Two aspects of cognitive competence are subject matter (or domain) competence, such as learning, memory, and deduction and judgment and decision-making competence, such as observation and inference. Cognitive continuum theory is based on correspondence and coherence, which can be achieved both by intuition and analysis.

The premises of cognitive continuum theory include the following:

1. Various modes, or forms of cognition, can be ordered in relation to one another on a continuum with intuitive cognition on one end and analytical cognition on the other.
2. The forms of cognition that lie on the continuum between intuition and analysis contain elements of both, and are termed quasi rationality (aka, common sense).
3. Cognitive tasks can be ordered on a continuum with regard to their capacity to induce intuition, quasi rationality, or analytical cognition.
4. Cognitive activities may move along the intuitive-analytical continuum over time; successful condition in stable environments requires stability along the continuum, but changing environments necessitate movement along it (dynamic cognition).
5. Human cognition is capable of both pattern recognition and the use of functional relations.

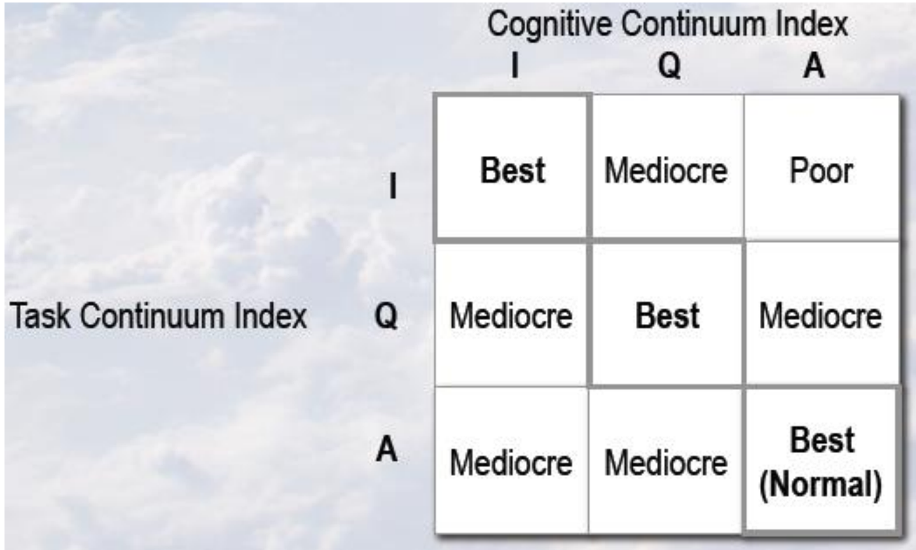

Figure 4. Cognition-Task Interaction[4]

The presenter also included different methodologies for conducting the analysis of cognitive work. One method is Cognitive Work Analysis (Vincente, 1999; Linterne, 2009). There are six stages of cognitive work analysis, each with resulting products:

Stage 1: Work Domain Analysis (WDA) – An Abstraction-Decomposition Space (ADS), an activity-independent representation of the functional structure of the work domain

Stage 2: Work Organization Analysis – A Contextual Activity Matrix

Stage 3: Cognitive Transformation Analysis – A suite of decision ladders

Stage 4: Cognitive Strategies Analysis – A detailed description of potential strategies that can be used to execute the cognitive processes identified in the cognitive transformation analysis

Stage 5: Cognitive Processing Analysis – A detailed description of the activity elements associated with the different modes of cognitive processing Stage 6: Social Transactions Analysis – A transaction network in which the transactions between agents are identified and characterized

Information Visualization. Following the discussion on Cognitive Work Analysis, there was a presentation on the Human Factors of information visualization. Some of the issues discussed included the following:

- The role of the information display in enhancing human capabilities and ways of mitigating limitations

[4] Figure courtesy of Esa Rantanen as part of his invited talk.

- Information Access Cost (IAC) – displays should reduce IAC during critical time-pressured events
- Information display should reduce human error, aid in error detection, mitigate error consequences, and not introduce new types of error

Ecological interface design is a theoretical framework for interface design, which attempts to extend the benefits of traditional direct-manipulation interfaces to apply to complex human-machine systems. Some of the foundational concepts used in ecological interface design (EID) are the previously discussed abstraction hierarchy and the skills, rules, and knowledge framework. In complex work domains there are three classes of events that operators must respond to, including

- Familiar events for which they have necessary skills through repeated exposure
- Unfamiliar but anticipated events, in which the lack of operator skill can be compensated by tools and aids anticipated by the designer
- Unfamiliar and unanticipated events (which the designer also did not anticipate) where there are no procedures, tools, or aids available so solutions must be improvised

Some ecological interface design principles address the different skills, rules, and knowledge behaviors:

1. Skills-based behavior (SBB)
    a. The operator should be able to act directly on the display
    b. Structure the display to support SBB
    c. Manipulation of on-screen elements
        i. Direct manipulation device preferred over command-language interface
        ii. Aggregation of elementary movements to into more complex routines corresponds to concurrent chunking of visual features into higher-level cues for these routines
        iii. Representation of multiple levels simultaneously
2. Rules-based behavior (RBB)
    a. Provide consistent one-to-one mapping between the work domain constraints and the cues or signs provided by the interface (i.e., provide the operators signs they can use to select appropriate actions)
        i. Will allow operators operate by relying in perceptual cues instead of having to resort to KBB
        ii. Will allow operators to take advantage of the economy of RBB while preserving the wide applicability of KBB
3. Knowledge-based behavior (KBB)
    a. Reveal the problem space in the form of an AH presentation
    b. Provide the operators a normative model of the work domain
    c. Support experimentation
    d. Relieve the operators of the burden of keeping track of causal nets within which they are reasoning

Ecological interface design provides the following guidelines:

1. Support experimentation by making boundaries of acceptable performance visible and their effects observable and reversible.
2. Provide feedback to support understanding and KBB monitoring during RBB performance, make visible latent constraints upon action.
3. At the RBB level, the display should represent cues for action as readily interpretable signs *and* indicate the preconditions to their validity (i.e., signs should have *symbolic* content)
4. To assist operators in coping with unanticipated situations at the KBB level, provide tools for experimentation and hypothesis-testing without a need to impact the plant, or make these actions reversible
5. Provide users with overview displays by which SB routines can be peripherally monitored
6. At RBB level, avoid procedural traps (strong-but-wrong rules) by providing integrated patterns as cues for action
7. At the KBB level, support memory by externalized schematics of mental models
8. Present information based on available data that is simultaneously suitable for SBB, RBB, and KBB processing
9. Provide an externalized mental model to aid in causal reasoning
10. Provide external memory aids for items, acts, and data that are not part of the present operational 'gestalt.'

5.0 Numerical Libraries

Numerical libraries are an essential element for enabling high-performance power applications with more dynamic and stochastic characteristics for the future power grid. Considerations include the design of the library, parallel computing techniques, mathematical algorithms required to enable more efficient computations, and verification/validation of the library. The purpose of such a library is to provide building blocks that are already parallel-computing compatible to developers, who can then focus on functional implementation instead of parallel computing details. The libraries also provide an opportunity to separate hardware capabilities from the applications. Suitable abstractions can make new functionality in hardware available to application developers without necessarily reworking large application codes.

The main discussion for the numerical library session focused on two topics: (1) what are the current issues we have in the area of power system applications, solvers, algorithms, numerical methods, and requirements; and (2) what are the use cases to select for moving the field of high performance computing (HPC) development forward. Details of the discussion can be found in the following sections.

5.1 Power System Applications

To develop effective and efficient numerical libraries, we need power system applications and well defined challenge problems to motivate library development and evaluate progress, especially considering the dynamic and stochastic future power grid. In the near future, the power grid will have a large amount of renewable energy and smart demand response, smart meters and hierarchically coordinated distributed control in the system. The power system applications we need to consider in the next 10-year term are more complex than what we have today. The power system applications have been discussed are listed below as the driver for the numerical library.

- Dynamic security assessment: Dynamic security assessment is critical for online evaluation of power system stability. As the size and complexity of interconnected grids increases, it becomes more and more challenging when a large number of contingencies are considered for near-real-time operation. The availability of PMU data needs to be considered in this application.
- Dynamic state estimation: Dynamic state estimation introduces dynamic models for real-time power grid operation. It provides a full dynamic view of a power grid, which further enables look-ahead dynamic simulation and dynamic contingency analysis. Observability analysis and how to extract information on dynamics are important for this application.
- Real-time path rating: Developing the ability to assess path capacity, including dynamics, in near time or real time is important for power grid operation. In today's practice, over-conservative path ratings are used because the execution time normally takes hours to weeks to complete the task. With a faster and more efficient tool, more accurate path rating can be calculated in near real-time, to reduce the cost of incurred by power companies.
- Extreme events with stochastic generations: Cascading failures and catastrophic events are key power system concerns. We need to be able to analyze these rare events (which may evade conventional approaches such as Monte Carlo) and assign probabilities to them. Traditional methods of enumerating scenarios are combinatorially intractable and so new approaches are needed.
- Renewable MW forecast: The penetration of intermittent renewable energy sources into the grid creates unique problems in developing accurate forecasts for renewable capacity and, more importantly, forecasts of grid reliability and efficiency. Wind generators and other non-deterministic power sources are not necessarily completely uncorrelated and need sophisticated models to predict overall contribution to the grid.
- Dynamic simulations: Dynamic simulation is used to simulate system dynamic behavior, including generator rotor angle and speed. When the dynamic simulation can be completed faster than real-time, we can predict the system future to reduce the risk of system failures.
- Other application areas include:
    - Multi-temporal-spatial scale model
    - Stochastic optimization/scheduling
    - Restoration planning – both transmission and distribution
    - Decision support
    - Aggregation of dynamic distribution load
    - Energy planning
    - Topology estimation
    - Better categorization based on utility business

The applications above pose tremendous challenges on developing different solvers for dealing with different problems. Detailed information about the solver issues can be found in the next section.

5.2 Solvers and Other Numerical Packages

The computational requirements for future power grid applications map into the following fundamental set of numerical solvers: linear algebra (linear system, non-linear and eigenvalue solvers), differential-algebraic equation solvers, and constrained optimization. In general, power system solvers involve sparse matrix manipulation. There are some existing solvers that can be used for solving power system applications, but they can be improved by utilizing the sparsity characteristics of power problems. In addition, an emphasis on renewable energy applications, which typically cannot be represented in a deterministic fashion, requires an increased focus on sampling approaches and, subsequently, integrated treatment of uncertainty. Based on the nature of different power system applications, the solver can be further categorized to following types.

- Linear algebraic equations: A linear algebraic equation solver is the basic solver for power system problems. Some example applications include power flow, contingency analysis, and state estimation. There is room for improvement of the linear algebraic solver by developing new suitable algorithms for power system and utilizing the power of high performance computing techniques.
- Nonlinear algebraic equations: Fast and efficient nonlinear algebraic solvers are also important for power system applications, such as power flow and state estimation.
- Eigenvalues: Mode analysis is one critical example for eigenvalue solvers, which can be used to determine the system's stability. Eigenvalues may be more useful than eigenvectors. This may provide some opportunities for performance optimization.
- Differential algebraic equations (DAE): DAE solver is a key part of solving dynamic simulation.
- Constrained and unconstrained optimization: Optimizations that include linear, nonlinear and discrete variables are needed. Example applications include optimal power flow and unit commitment with the involvement of renewable energy.
- Dynamic model reduction: Model reduction is also extremely important to reduce the complexity of power system to allow accurate approximation and then fast computation for multiple applications.
- Stochastic optimizer: The optimization under conditions of a stochastic environment requires advanced stochastic optimizers

All of these solvers would require advanced algorithms for making the solver efficient and robust.

5.3 Algorithms

To implement efficient and robust solvers, efficient algorithms are required. The algorithms that have been discussed are listed below. Areas where algorithms are needed include

- Non-linear optimization: many non-linear optimization problems are important to power systems. These problems can also contain discrete variables and a large number of constraints.
- Parallel linear and non-linear solvers: a broad range of solver capabilities are needed that can utilize specific features of different power grid problems

- Uncertainty quantification: Uncertainty quantification becomes more important when the stochastic behaviors of power system are considered. Developing probability distribution functions for computed quantities based on the statistics of the input with decision support identified as a use case
- Dimension reduction: Big data will be a problem in the near future, with the inclusion of PMU data and smart meter data. Dimension reduction algorithms are needed to extract useful information from the large amount of data and to present it to the user efficiently in a format that can be readily comprehended. This algorithm is selected as a use case, with the linkage to visualization and decision support
- Scenario generation: Stochastic generation of scenarios is useful for both modeling current system and for future planning. It can be used to generate stochastic realizations of details in existing networks that may not be available and to generate realizations for future networks that can be used for planning purposes. It can also be used to generate test sets for algorithm and library development.
- Least squares:
- Filtering (dynamic systems)
- Integration of discrete optimization with nonlinear power flow model
- Discrete events in continuous systems
- Optimal control of dynamic systems

5.4 Numerical Methods

Numerical methods are the study of algorithms that use numerical approximation for specified mathematical problems. Following numerical methods have been discussed in the workshop and are considered important to future power grid applications. No particular priority is implied nor is the list complete.

- (Bender's) Cut: for mixed-integer linear programming.
- Uncertainty propagation: Efficient methods are needed to propagate uncertainties and variability in inputs (e.g. uncertain parameters and unsteady or variable power sources) to generate characterizations of the distributions in output values
- Large scale, sparse, direct/iterative linear solver in parallel: linear algebraic equations are a major component of power grid analyses and reliable, robust and efficient solvers are a core need. Different problems may have different opportunities for optimization so a spectrum of solvers is needed.
    - Correlated random input: Methods are needed for generating correlated random inputs where some variables are random but correlated with each other. These are particularly needed for modeling renewables (*e.g.* wind farms) where behavior of a large group of variables is collectively random but individual variables within the group are highly correlated.
    - Graph theory for partitioning: Partitioning networks is an important part of distributing power grid problems to run on HPC platforms. Efficient partitioners not only distribute

work evenly across processors but can also minimize communication as well as creating desirable matrix layouts that can minimize the amount of work required to solve linear and non-linear algebraic equations based on the network.
  - Graph analysis is also necessary to detect island formation and other features which may affect the solvability of the power system equations and require special handling (*e.g.* creation of a new reference bus).
- Extreme statistics: Identifying extreme events and assigning a probability to them is important in assessing grid reliability and robustness. These events are unlikely and may evade conventional sampling approaches. Other non-sampling approaches may make it difficult to assign a probability to the event. This is a use case.
- Analysis and control of distributed dynamic systems: A power system consists of a large collection of distributed small dynamical systems that are spatially interconnected to achieve a global system behavior using local interactions. This item involves the interplay of control theory, system dynamics, distributed optimization, and graph theory.
- Semi-definite programming: This is a subfield of convex optimization concerned with the optimization of a linear objective function.

5.5 Requirements

During the workshop, we also discussed the requirements for developing numerical libraries. Some highlights are listed below:

- Test cases and quantifiable metrics for verifying and validating programs are required. These will need to be extensive enough to insure that successful solution of the test problems will lead to a high probability that the codes will solve a broad range of problems outside the test suite. Validation cases with associated metrics for success will also need to be developed that will guarantee that codes validated on the test cases will provide actionable results for other real world problems.
- Performance (Robustness and consistency): New libraries and algorithms will need to be robust so that a wide range of problems are solvable and high performing. For modeling and forecasting purposes the requirements for scalability and performance may be different than for real-time control of power systems.
- Realistic, synthetic test cases (Real environment): Many current test cases are based on real-life systems and have issues with access. Synthetic test cases that are widely available and do not have proprietary and/or security issues associated with them are desirable.
- Benchmarking, performance profiling for various tasks: Benchmarks that can be used to assess code performance across different platforms both in absolute terms (time to completion) and scalability are needed. Profiling tools that can be used to assess different parts of an application and diagnose non-scaling parts of the algorithm are also needed.
- Establish a community of libraries, build a multidisciplinary community: Community development of software is important. It will both to lower overall development costs and to standardize parts of the software and data stack to enable widespread use of data.

Community development will also promote code reuse and optimization of important components. Governance and support (both financial and in kind) are key issues for this effort.

- Diagnosis of software
- Goal-driven test cases
- Secure software

Table 1 summarizes the issues in numerical library.

Table 1. Summary of Numerical Libraries

Power system applications	Solvers	Algorithms	Math	Requirement
Real time dynamic security	Linear algebraic equations	Class of nonlinear optimization problems important to power system	(Bender) cut	Verification (test case, quantifiable metrics)
Dynamic state estimation	Nonlinear algebraic equations	Least square	Uncertainty propagation	Performance (robustness, consistency)
Renewable MW forecast	Eigenvalues	Filtering (dynamic systems)	Semi-definite programming	Realistic, synthetic test cases
Multi-temporal-spatial scale model	Dynamic simulation	Uncertainty quantification	Sparse direct linear solver in parallel	Real environment (added noise)
Stochastic optimization	DAE	Scenario generation	Correlated random inputs	Diagnosis tool
Restoration planning	Constrained optimization	Dimension reduction	Analysis and control of distributed dynamic system	Goal-driven test cases
Real time path rating	Non-constrained optimization	Integration of discrete optimization with nonlinear power flow model	Graph theory for partitioning	Secure software
Design support	Dynamic model reduction	Discrete event to continuous system	Extreme statistics (rare event)	
Extreme events w/ stochastic gen	Stochastic optimization	Optimal control dynamic system		
Aggregated distributed load modeling		Stochastic DAE		
Energy planning				

6.0 Software Infrastructure

6.1 Trends in Operational Data

Trends in operational power grid data point towards greatly increased data volumes (by approximately two orders of magnitude) and more sophisticated simulations and analytics. These advances will enable systems that provide real-time decision making and improved grid predictability, eventually leading to automated control of distribution and transmission systems. Several obstacles need to be overcome in order to reach these goals, including:

- Scaling issues related to data storage, communications network, and communications protocols
- Real-time applications will require extremely low latency networks
- Improving data quality (accuracy, missing data)
- Cultural changes that enable operational staff to take advantage of IT innovation
- Development and availability of sophisticated analytic tools
- Allowing heterogeneous data models to interoperate instead of enforcing a single model for all purposes
- Lack of comprehensive and modern training materials to educate the next generation of operators

As these obstacles are addressed, we will begin to see a convergence of IT and control technologies, which will require a much larger percentage of operations staff to possess IT skills. Cloud enabled technologies will be used more and more as evidenced by partnerships in the eastern US and other regions to form a collective private cloud for delivering advanced analytic tools.

To achieve these goals, the research community needs access to high quality data sets, but significant barriers to data availability exist. These include security concerns, a lack of facilities for generating realistic simulated data, and lack of cooperation in constructing unified or interoperable data models. Another issue is that researchers are focused on traditional high performance computing (HPC) computational architectures, whereas the power grid of the future will require architectures that support high speed streaming analytics. This implies that researchers need to modify their paradigm and consider how to translate existing simulation algorithms to the streaming analytic model. Additionally, researchers should develop and make available streaming architectures so that new algorithms can be tested against modern data sets.

Finally, this leads to one of the most significant issues facing researchers and operators, which is a tendency of researchers to look for technology solutions first over engaging with operators and learning about their needs. Going forward, computational research and software development must focus on operational needs first. To do this, comprehensive use cases should be developed in cooperation with operators that the research community and vendors can use to develop next generation tools.

6.2 Architecture and open source issues

The industry understands the need for a middleware to allow for the abstracting data from the applications being performed. That is an application request data but does not store the data internally. In the 3 presentations presented at the workshop, each outlined a product that is an information bus.

1. Grid Protection Alliance discussed how openHistorian was developed to be the integration bus for synchrophasor data and other high sampling rated data being received by a control center.
2. Pacific Northwest National Laboratory presented the GridOPTICS™ Software System as an example middleware to decouple data sources from the applications. The project also has a web service layer to allow for sharing select data with other utilities.
3. Incremental Systems presented PowerSimulator Architecture that uses Power Application Computing Environment that has PowerData real time database and Power Integrator – application integration framework.

These three concepts are unique in implementation but have common characteristics in that they believe multiple applications or services will want to use the same data. A difficulty in creating a single integration bus for the utility space is the name space problem. The integration layer must be able to handle 'n' configurations or it will not be used. An example of namespaces to be managed is the operations control room that uses high detailed node/breaker model for real time monitoring of the grid; compared to the planners that use the lower fidelity but larger footprint bus/branch model for long term studies. As planning and operation need to work on near term model, presently this is place of friction to get the models to align for studies. Other requirements for the middleware is to have fine grain data policies for who can use specific data points, built in user authentication, and ability to work with legacy programs.

The conversation about the middleware and architecture come back to having realistic data to test our software on. There is a push within IEEE to create the Smart Energy Repository. This repository has the goal of collecting realistic data creating data layer and data model. To obscure sensitive data such that NDA's and CIP compliance is not needed would require a large investment. The group believes researchers need access to this data. In the discussion of creating repository of the data for testing there was two possible formats pointed out. A suggested data type was the IEC TC57 CIM model of the power system which, was viewed by some as being too complex and updated versions being released frequently (i.e. the format is not stable). Another option presented was to augment the PSSE models with circuit breakers and substation locations.

The building of an open source community will be necessary to improve adoption of the tools presented at the workshop. During the workshop it was discussed that technology has an S-curve in which in the beginning there is a lot of software defects and at the end there nearly no defects. The control center software has been at the top of the S-curve for 3 generations. This has caused reluctance to change things in the operation technology systems in moving down the S-curve to having more software defects in the systems. The group realized that to get tools and architecture discussed at the workshop there need to be early adopters who will evangelize the technology to others. The group also recognized that the power industry is very slow to adopt new technologies and that long lead time for internal testing needs to be done prior to new technology being used in the production systems. To help with the quality of open source projects, there is a need for governance. This group of advisors is available to help guide the project and maintain high quality in the code being released from the project.

7.0 Use Cases

A thorough understanding of the emerging and anticipated use cases that are driving research and technology development will effectively enable investments into new technology. Each of the breakout sessions was asked to highlight uses cases that motivated the discussion. The following discussion is not intended to be exhaustive but represents some key problems discussed by the expert attendees. It is clear however that more focused and detailed discussion of relevant use cases with more input from power utility operators is needed. Ultimately, these use cases need to be concrete and benchmark problems with well-defined metrics to measure progress.

7.1 Dynamic Security Assessment

Dynamic security assessment (DSA) is critical to determine system security and region and operation limits, such as interface flow limits and generation limits. Transient analysis studies are performed with a certain number of predefined dynamic contingencies to ensure system security against these contingencies. These study results are then presented to operators through an Energy Management System to help in decision making. The objective of DSA is to improve system reliability, identify additional energy transaction potential and enhance use of engineering resources. In the real-time mode, it is required to run DSA every few minutes to provide operators needed information about system security status. Each DSA run includes contingency screening, transient stability analysis, and power transfer limit computation on selected contingencies. The mathematic model of transient analysis consists of a set of nonlinear differential equations and algebraic equations (DAE). The conventional solver approach is time domain based, i.e. uses numerical methods to discretize differential equations at each time step and iteratively solve machine equations and network equations. Today's DSA tool is relatively slow for a limited number of contingencies. The proposed DSA use case is faster, aiming for near-real-time computation with a large number of contingencies assuming a smart grid environment to provide more efficient and effective decision support to operators.

As the size and complexity of interconnected grids increases, the number of contingencies needing to be considered increases significantly and system dynamic models become more complex. These factors make performing near-real-time DSA in the control center challenging. To achieve real-time DSA, advanced High Performance Computing techniques are required to speed solving DAE problems for each individual contingency, and to concurrently process each contingency in a distributed way with an optimal dynamic load balancing scheme. The availability of more and more PMU data provides both a challenge and opportunity for near-real-time DSA: additional data impacts the computational burden, but also provides opportunities for more advanced algorithms for improving computational speed. Also, the penetration of stochastic renewable energy resources cannot be ignored. Thus, data and software architectures need to be designed to allow flexible data flows for the DSA simulation. An advanced DAE solver numerical library has to be developed to significantly reduce time-domain simulation time. While

DAE solvers have been studied in other domains, leveraging what has been done while taking advantage of the special properties of power system equations (such as sparsity, structural symmetry) is important. In the end, DSA is essentially a decision support tool to help operators make decisions maintaining system reliability and improving asset utilization. DSA results have to be presented to operators in a meaningful way. How to present DSA results is an important research topic, which involves not only visualization techniques, but cognitive science to understand what operators really want from real-time DSA.

7.2 Improved Asset Utilization

Transmission congestion poses significant challenges for power grid reliability in stressed conditions due to heavy loading and in uncertain situations due to variable renewable resources and responsive smart loads. However, building new transmission lines is not an easy option, involving both economic and environmental constraints. A non-wire solution that realizes unused transfer capabilities by performing realistic real-time path rating studies could be used to alleviate transmission congestion. A new approach to asset management could offer annual savings of billions of dollars if successfully developed and broadly adopted by the industry. The traditional method to determine a path rating is to perform several offline planning studies assuming worst case scenarios that determine the maximum power that can be transferred trough the path without violating thermal, voltage, and stability limits of the system. The path rating is hence a static number in operations that in the best case changes once a season (Spring, Summer, Fall, and Winter), but in a worst case scenario is set and left at the same value for multiple years. The proposed used case is to continuously calculate path ratings in near real time.

To perform path rating in real time, improvements need to be made to programs and algorithms used in determining the maximum power allowed to flow through a path. The thermal limit of the system is based on the maximum power that be transferred through a transformer or line segment. Dynamic inputs to this calculation include weather conditions (e.g. wind, temperature, and precipitation). The voltage stability limit is determined by finding the operating point of the system in which a voltage collapse happens. Finding such an operating point is presently done by iteratively performing multiple power flow studies with increased stress added to the system. The transient stability limit is based on a systems ability to find a new steady state after a system disturbance (e.g. a short on a line or generators removed from service). The calculations required to perform transient limit analysis is also discussed above in the use case for dynamic security assessment. Since each limit is computationally intensive to calculate it has previously not been possible to calculate all three limits (thermal, voltage stability, transient stability) in near real time. Advancements such as those discussed in the Numerical Libraries breakout session may make it possible to run algorithms fast enough. Each of these limits needs to be calculated using data from real-time observations, which are measurements from the field. Note that some of the inputs are not power grid measurement but wind forecasts, weather, and social activities. Thus the software architecture should be defined to allow for heterogeneous data sets to be used in the calculations. Finally, updated path ratings need to be used by operators as they make decisions about grid management. To have a path limit that is dynamically changing throughout the day will be drastically different than today's practice. Researchers will have to examine how the limits are used in the decisions made by operators prior to updating visualizations with the new information.

7.3 Predictive Control Systems

Dynamic complex systems require high reliability, demanding the need for highly skilled operators and advanced analytics and computing tools to deal with the increasing amount of high-frequency data from Phasor Measurement Units and smart meters. This evolving complexity suggests the need to create predictive capabilities in the tools that are being designed for the future power grid. The continuous task of monitoring the health of the system involves taking large amounts of data to be processed, applying fast computation to the data, which allows a rapid response by the automation or a human operator to the various events happening in the system at any given point.

In order to design a predictive control system that spans modeling, simulation, and decision making, several research gaps need to be filled. Initially, there is a need to understand how to design predictive capabilities that will allow the analysis of massive quantities of collected data and interpret it using fast computation. These capabilities need to further automation and quickly help human operators make sense of them, leading to timely decisions. Human operators require effective decision support and visualization tools designed in a way that represent the accuracy of the data (with the level of uncertainty quantified), and supports reasoning under uncertainty with the ability to predict the results of actions and decisions. The analytical tools also need to transform the great quantity of data into actionable information that can be represented visually to the operator in the right way at the right time.

7.4 Risk-based Decision Support

Disturbances to power system operations happen all the time, such as unplanned outages of generators and transmission devices, sudden drop of variable renewable generation, and unexpected congestion of transmission paths. Power system emergency situations caused by disturbances can be reduced, alleviated, or avoided by assessing operation risks during real time and take swift actions to steer the system away from security boundaries. Risk-based decision support can help operators extract the most essential information from large amount of raw data, perform predictions of system behavior under various possible load, variable generation, and market scenarios, and find optimal actions at the operator's disposal to resolve an emergency situation. Risks and benefits of different actions can also be evaluated by the decision support system to facilitate the operator's selection among generation redispatch, demand response deployment, load shedding, reconfiguration of lines, power flow and voltage control devices, and sometimes protection relay settings, etc.

A reliable decision support system requires sufficient and usually large amount of measurement data to feed into the underlying models and analytical tools, such as dynamic security analysis and stochastic optimal dispatch. It requires a spectrum of analytical algorithms and tools working in concert, including data mining, machine learning, algebraic and dynamic system solvers, optimization, and ecological human-machine interface design techniques. Software architecture design and numerical libraries are both critical to the development of such systems. A high performance computing platform is necessary for real-time screening of possible scenarios, prediction of system behaviors, and optimally selecting dispatch and control actions.

Demonstration of a risk-based decision support system requires the availability of underlying models and analytical tools, readiness of data feeds into these tools, and the integration of different types of information.

8.0 Community Building

Community building is an essential step in laying a solid foundation for any collaborative project. This will be an especially critical element in the development of the next generation grid analytics space. Often these types of communities coalesce around and develop communally a software infrastructure in source code form. It is crucial that an environment for such offers a sense of community that is based in trust and mutual sharing. In order to do so, it is proposed that the software (source and binaries) be made available under an open source license format. The collective group recognized the need for this work to be in an open workable format in order to have it become a mechanism to enhance community, collaboration and innovation. However, there are other goals that may appear to collide with this free and open market. Ultimately there is a goal to have the technology available to end users in a supported system(s). To facilitate vendor investment into the development of a commercially supported product, incentive that curtail the risks associated with such investment are needed. So, a balanced viewpoint will be critical.

8.1 Who is the community?

In order to know what structure is optimal and how to create this environment we need to investigate who will make up this community. It will most likely be made up of developers, integrators (vendors) and end users. Of those people, it is important to determine how the code will be used by each of the different groups. Who will require source code? Who needs the commercial off the shelf version of the application with user interfaces? What type of rights will each of these require in order to provide the biggest impact?

As for source code, these users will most likely be will be made up of developers and integrators. Developers will work on the research and development of the code continually improving the code so that it is useful. These are the players that make up the general body of the GridOPTICS™ project. Integrators and vendors are the companies that will ultimately develop the research grade code into something that looks more like a commercial grade product. Who are the integrators and vendors that can create such a thing? Will they also play a part in the initial development? These will need to be groups that would be interested in supporting and maintaining such software. The community will need to determine how best to keep necessary rights available so that these integrator types are interested in investing in the longevity of the product. There will need to be an organization supporting the need for such integration with these vendors.

In order to find more ‘birds of a feather’ players for the GridOPTICS™ project team, the community will need to discover other groups and/or conferences with common activities and interest. Where do we find such players? Where should the community be publishing papers? What are the most appropriate publications that will be of interest to those that could add value to GridOPTICS™?

8.2 How will the community be governed?

For any organization to be successful in the long term there needs to be a vision and mission for long term goals but also short term objectives. The group will need to investigate how this community is going to be governed so that such objectives materialize. There needs to be a structure in place so that the parties will work in accordance with common processes and procedures for dealing with research results. How is ownership of jointly developed results going to be handled? We will need to institute a process to review the code – assuring quality as well as security.

There will also need to be a determination made in regards to the best licensing mechanisms to allow both open collaborative research as well as ample opportunity to allow third party integrators and vendors to show interest in investing time and other resources in development of commercial grade products that can be serviced and supported for the ultimate end users. It is understood that utilization of an open source licensing mechanism will be needed for the development. What is the best open source licensing model for this activity and subsequent goals? A subgroup should be organized to study this issue in depth so that the research can be accomplished and the results further developed and integrators and vendors are incentivized to pursue commercialization.

8.3 What does the community need to use?

Another area to place emphasis on is the types of software tools that are necessary for this project to succeed. Some effort will be required to investigate what the community needs to use. Are there third party tools that are essential to the success of the project? What third party libraries will also be required? If so, what licenses are they offered under? Are those licenses compatible with the desired chosen operating license? How will these impact the desired licensing structure?

What other groups are complementary to GridOPTICS™? How can these groups intersect and interact with each other? How are these other groups structured? What types of licensing do they typically use? How do we make the GridOPTICS™ structure and governance compatible to these other players? Since the goal of the GridOPTICS™ framework is to become a universally usable code, it needs to be pluggable and very user friendly. Therefore, the licensing also needs to be protected from the viral components of the licenses other code may be licensed under. In order to be user friendly and to keep from infecting the total framework much thinking will need to be done in this regard.

8.4 Sustainability

Since there is the expectation that the software development will be done in an open source format, it becomes a great concern on how to maintain and sustain the code base. Initially there may be some support from government sources, but there may need to be a community that is contributing not only code, but also monetary support. Successful projects in the open source environment do not just happen overnight. They are organized with a strong foundation and structure and provided ample fiscal support to cover all of the needs of the base. Much thought and discussion will be required to determine how to keep the software infrastructure funded after the initial project ends. What does the endpoint look like? What does the end product look like? What will be required by the vendors and integrators for this to occur?

9.0 Recommendations

The findings and information above lead to specific organizational and technological recommendations on next steps:

1. A general consensus among the participants is that a software infrastructure such as GridOPTICS™ should be developed as open-source software by an open community. This community needs input and acceptance from end-user consumers of the software (utility companies), intermediates who market production-ready software ("vendors"), and researchers creating proof-of-concept prototypes. To continue engaging such a community, the workshop recommended that future workshops be held at a frequency of once or twice per year, with smaller discussions in between to strengthen the collaboration among various participants. In particular, the next workshop should be scheduled about nine months after this workshop.
2. A business model continues to be a key aspect that needs to be developed to facilitate the forming of the community, the development of the architecture, and the appropriate sharing of technologies amongst stakeholders. The business model would include the governance structure of the community and the funding model to support the community. The community governing board would seek funding support for community activities. In the initial stage of the community, funding support from the Department of Energy to bring the effort to the level that the industry would see the value and in turn adopt the technology and support the community efforts. This funding will both enable raising the technology readiness of research software but also enable development of features and engineering processes to support cross-organization interoperation necessary for effective sharing and reuse of components. Eventually, funding for ongoing activities should come from a consortium of vendors and end-user utilities in addition to support from the Department of Energy.
3. The GridOPTICS software architecture needs to facilitate interoperability of software tools for data, computation, and visualization and the integrated modeling of transmission and distribution for seamless operation and planning functions. The workshop recommended that the community develop a basic design and a reference implementation. The reference implementation can then be extended by vendors to support various power grid applications. In order to strengthen the community, we recommend enhancements to be contributed back to the reference implementation. The community should create and maintain a common shared-source repository accessible to researchers, software vendors, and end-users. Establishing a repository from existing assets is a necessary step.
4. As part of the architecture, decision-support tools should be designed by using a human- or practice-centered approach. Cognitive systems engineering (CSE) takes a multidisciplinary, practice-centered approach with the goal of guiding the design of complex, computerized systems intended to support human performance. By applying a variety of methods to understand and support human cognitive performance (e.g., problem solving, judgment, decision making, attention, perception, and memory), tools can be designed and built to aid operators in accurately assessing the situation, gaining and maintaining situation awareness, and supporting them in successfully, efficiently, and safely conducting their work. Effort should be made to expand awareness of such techniques and successful patterns. We recommend

creating forums such as workshops to this end. We further recommend creating venues to enable testing of new technologies with real-world data streams and models.

5. Predictive capabilities are recognized to be an important foundation for the future power grid. In addition to accelerating computation, new research on uncertainty quantification and stochastic analysis is needed to provide confidence in predictions. This new research spans modeling, simulation, and decision making, and should be considered in the development of methods and tools from the design stage of the GridOPTICS architecture. Such predictive capabilities will benefit a wide range of applications from planning decisions, to operational decisions, to automation of control. They must be reliable enough for production use and robust enough to reduce cost or improve operational security. We recommend a research program be established by the community guided concretely by problems of high end-user importance and with an emphasis on understanding robustness of techniques in operational contexts.
6. In order to expose GridOPTICS™ and other proof-of-concept software to real world conditions, we recommend creation of a realistic control-center-like testing facility. This facility must receive timely power industry streaming data which can be used to create a test bed for grid-specific toolsets and applications built on top of it. Such a testing facility would satisfy the venue needed for new decision support tools by enabling operators to train with new tools and provide feedback to software developers, a feedback loop that would result in better, more usable tools. It would also provide an environment to test new strategies and techniques for visualization and alerting of operators to emerging operational conditions. As a resource shared with the research community, it will permit end-to-end demonstrations where new components can be studied in an operationally-complete environment. This facility would enable a co-design process that ensures proposed software tools are tested with operators to improve the usability of such tools.
7. Central to making measurable and credible progress is that the community defines a number of benchmark “challenge” problems that concretely and formally define success. These need to be vetted by the industry members of the community so that the benchmark problems can indeed be used to assess the impact of new technologies on the reliability of future power systems and on their economic feasibility to deploy and maintain. These problems will need datasets to support testing and careful consideration of operating requirements and the range of acceptable solutions. Delivering on a challenge problem is likely to require integrated innovations in all the three areas (i.e. data, computation, and visualization) discussed in this workshop.

10 Appendix

The steering committee that organized the workshop was: David Callahan, Chair (PNNL), Gil Bindewald (DOE), Henry Huang (PNNL), Chen-Ching Liu (WSU), Tom Overbye (UIUC), David Sun (Alstom), and Paul Whitney (PNNL). The breakout sessions were facilitated by: Chen-Ching Liu (WSU) for Numerical Libraries, Tom Overbye (UIUC) for Decision Support, and Kevin Tomsovic (UTK) for Software Infrastructure.

The following table is the list of all attendees present.

Shrirang Abhyankar, ANL	**John McEntire, PNNL**
Naresh Acharya, GE	**Liang Min, LNNL**
Scott Backhaus, LANL	**Jodi Heintz Obradovich, PNNL**
Gil Bindewald, DOE	**Tom Overbye, UIUC**
David Callahan, PNNL	**Bruce Palmer, PNNL**
Yousu Chen, PNNL	**David Pinney, NRECA**
George Chin, PNNL	**Robin Podmore, IncSys**
Jeff Dagle, PNNL	**Esa Rantanen, RIT**
Kendall Demaree, Alstom	**Mark Rice, PNNL**
Jason Fuller, PNNL	**Russell Robertson, GPA**
Joe Gracia, ORNL	**Dick Russell, PNNL**
Santiago Grijalva, NREL	**Poorva Sharma, PNNL**
John Grosh, LLNL	**David Sun, Alstom**
Steven Hammond, NREL	**Brian Thomas, GE**
Jean-Baptiste Heyberger, RTE-France	**Kevin Tomsovic, UTK**
Henry Huang, PNNL	**Mani Vadari, MGS**
Carl Imhoff, PNNL	**Rob Vance, Alstom**
Wesley Jones, NREL	**Carol Woodward, LLNL**
Roger King, BCE	**Adam Wynne, PNNL**
Daniel Kirschen, UW	**Jinjun Xiong, IBM**
Hao Li, SCL	**Meng Yue, BNL**
Chen-Ching Liu, WSU	**Hongming Zhang, WECC**
Shuai Lu, PNNL	