



LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

Using Supercomputers to Speed Execution of the CAISO/PLEXOS 33% RPS Study

*Carol Meyers, Fred Streit, Yiming Yao,
Steve Smith, and Alan Lamont*

September 2011

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

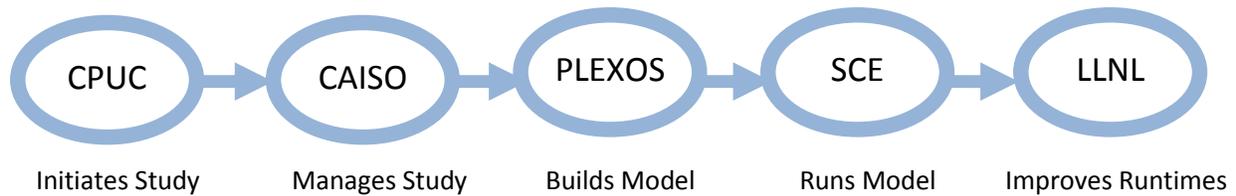
Using Supercomputers to Speed Execution of the CAISO/PLEXOS 33% RPS Study

Carol Meyers, Fred Streitz, Yiming Yao, Steve Smith, Alan Lamont

Study Sponsors and Key Players

The CAISO/PLEXOS 33% Renewable Portfolio Standard (RPS) study was initiated by the California Public Utilities commission (CPUC) in late 2008. The CPUC tasked the California Independent Systems Operator (CAISO) to manage and perform the study. In turn, CAISO partnered with a team at PLEXOS Solutions (a software vendor) to build the model, and tasked analysts at Southern California Edison (SCE) to run the model. The SCE analysts also worked directly with PLEXOS Solutions to refine and reconfigure the model as appropriate over the course of the study.

Lawrence Livermore National Laboratory (LLNL) was contacted by the CPUC after the analysts at SCE observed very slow runtimes in the model, necessitating a shift in the study completion date from April 2010 to February 2011. Specifically, a single month of the model was taking over a day to run, and a yearly calculation thus took several days. The CPUC asked whether LLNL high-performance computing (HPC) resources could be used to speed execution time of the model, and a team at LLNL investigated. An initial feasibility study was performed by LLNL in March-April 2010, followed by a comprehensive effort starting in August 2010 and continuing through mid-2011.



Objectives and Phases of the CAISO/PLEXOS 33% RPS Study

The study’s official title is “ISO Study of Operational Requirements and Market Impacts at 33% Renewable Portfolio Standard (RPS).” The stated objectives are twofold: 1) identifying operational requirements and resource options to reliably operate the ISO-controlled grid under a 33% RPS in 2020; and 2) inform market, planning, and policy/regulatory decisions by the ISO, state agencies, market participants, and other stakeholders.

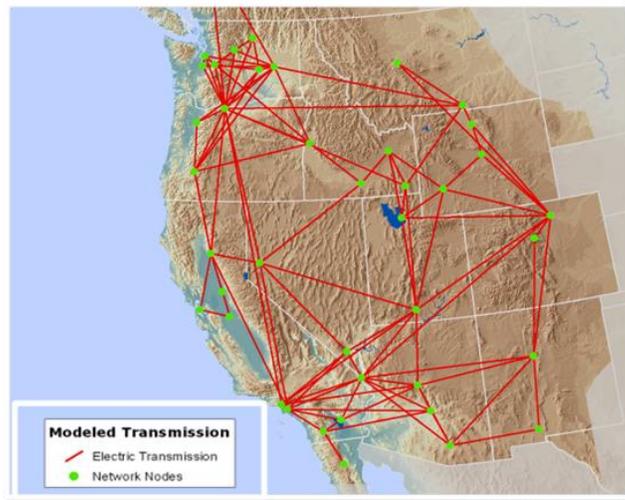
The first of these objectives requires the hourly estimates of integration requirements, measured in terms of operational ramp, load following and regulation capacity and ramp rates, as well as additional capacity to resolve operational violations. It also involves consideration of other variables that affect the results, such as the impact of different mixes of renewable technologies, and the impact of forecasting error and variability. The second objective entails supporting the CPUC to identify long-term procurement planning needs, costs, and options, as well as informing other decisions made by the CPUC and state agencies. For the ISO itself this includes informing state-wide transmission planning needs for renewables up to a 33% RPS, and informing design of wholesale markets for energy and ancillary services to facilitate provision of integration capacities.

The study is designed in two phases. The first (current) phase is focused on operational requirements and addressing these requirements with existing and new conventional fossil generation; for instance, gas turbines and/or combined cycle units. The second (planned) phase will address the same operational requirements with a combination of conventional fossil generation resources, new non-generation resources, and a renewable resource dispatch.

There are seven different scenarios considered in the current phase: a 20% RPS reference case; four 33% RPS cases (a reference case, a high out-of-state case, a high distributed generation case, and a low load case); an alternative 27.5% RPS case; and an all-gas case (no new renewables after 2008). In addition, the CPUC is planning a new set of cases that will alter the anticipated sets of runs.

Model Data and Framework

All data in the CAISO/PLEXOS 33% RPS study is maintained and provided by CAISO. This includes resource data and hourly load data for wind, import, hydro, solar, geothermal, and qualifying facility generation. The data has been vetted by CAISO and other stakeholders over the course of the last two years. The data ranges over 12 regions total, across the WECC grid (AB, AZ, BC, CA, CFE, CO, MT, NM, NV, NW, UT, WY), 42 utilities within these regions, and 2157 generators defined across the utilities. Transmission lines are defined between pairs of utilities in an aggregated manner (SCE analysts call them ‘transmission bubbles’), of which there are 104 total (see below). There are 7 different generator fuel types (coal, gas, oil, uranium, bio, wood, steam), and different designated sets of generators which are eligible to contribute to the CAISO and MUNI reserves. For storage, there are 7 different pumped hydro facilities, with head and tail storage at both.

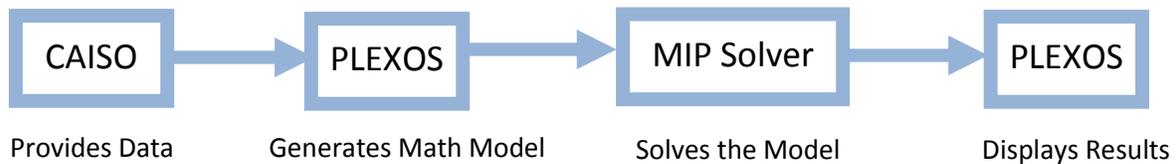


Transmission grid used in the CAISO/PLEXOS 33% RPS study

The full description of the energy model is implemented using the PLEXOS modeling software. The PLEXOS model contains a very high-level description of the system, in terms of the quantities that CAISO and SCE are familiar with (generators, transmission lines, fuels, emissions, storage, reserves, etc). The PLEXOS software takes the high-level description and converts it into a mathematical specification of the model (in this case, a mixed-integer linear program) describing the salient features of the grid

operations. In this specification, items to be optimized are represented by mathematical variables, and the behavior of the system is governed by a set of constraints over those variables. The overall quantity to maximize or minimize is represented by an objective function defined over these same variables.

The PLEXOS software takes the mathematical specification that it has generated and passes it to an external mixed-integer linear programming (MIP) solver. This solver is typically bundled with the PLEXOS software, but it is a separate code from a separate commercial entity. The MIP solver does *all* of the solving of the model, and passes the results back to PLEXOS, which postprocesses them and displays the results in a number of different formats.



Model Solution Procedure and Computational Bottleneck

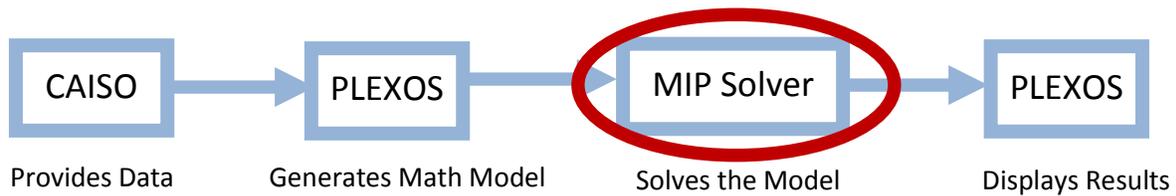
The basic problem that the CAISO/PLEXOS 33% RPS study is designed to answer is an operational one, specifically a unit commitment and *energy dispatch* problem: which generators need to be turned on, at which times, to satisfy the forecasted demand requirements across the grid, while obeying generation, storage and transmission capacity constraints? The simplest form of this problem is called a *unit commitment* problem; the CAISO/PLEXOS 33% RPS study is a bit more complicated in that there are a large number of additional side constraints. Some of these side constraints include the production of ancillary power, specifically the reserves. The study considers two different kinds of reserves: regulation reserves, which deal with minute-to minute fluctuations in demand and renewable generation; and load following reserves, which concern increases and decreases in generation directed by the ISO to meet changes on a sub-hourly basis. In both of these cases, only a subset of the generators is available to produce reserve power, and the model must choose which generators are used to produce this power.

Mathematically, the quantities that are optimized over are represented as *variables*, which are subject to *constraints* on the operation of the system over time. The CAISO/PLEXOS 33% RPS study model is a mixed-integer linear programming model, which means all of the constraints are linear. The objective of the model is to minimize the total operating cost, given predetermined levels of generation that are available.

A major part of the problem is the decision to start up or shut down generators during a day. Starting generators is expensive, so the model only starts them when necessary. The model must also take into account the fact that some generators may be relatively expensive to start, but less expensive to run. If they will be run for enough hours during the day, it may be less expensive to use those generators than a generator that is less expensive to start. This problem is difficult to solve because many of the variables are constrained to being *integers* that are only allowed to take on the values of 0 or 1. For instance, these are used to model the state of the generators as being either on or off.

Given these mathematical specifications, the question that the CAISO study seeks to answer is, “What are the cost impacts of different levels of renewable power generation on the system?” To solve this question, the model is run for the aforementioned 7 scenarios, and the results are compared. In each of the scenarios, the solution procedure itself is actually an *iterative process*: the generators start out with a certain base level of generation, and the model is solved to see whether any of the operating constraints are violated. Additional generation capacity is added to the system until none of the key constraints are violated. These violations are expressed as penalty terms on the objective function.

Computationally, the slowest performance *by far* is in the execution of the mixed integer programming solver, which occupies about 99% of the model’s solution time. The PLEXOS software itself generates the model very quickly, and the numerous exchanges between PLEXOS and the solver are likewise speedy. Within the solver, the largest part of the time is spent trying to find integer variables that satisfy the constraints and minimize the objective function. Unfortunately, the time required for this solution procedure can increase exponentially in the worse case, and there is no guaranteed way to solve it efficiently.



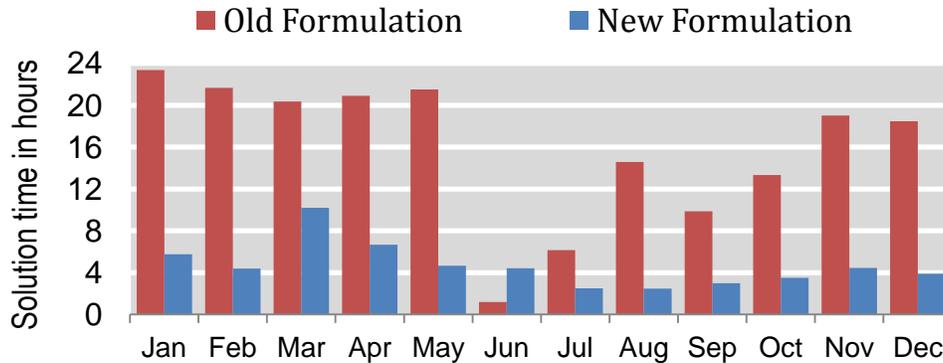
The MIP solver currently being used by SCE to solve the model is XPressMP, which is one of the better commercial solvers. At LLNL we have leveraged our ongoing relationship with IBM to acquire unlimited copies of another solver, CPLEX, which for many is considered the ‘gold standard’ in commercial solvers. While the solver used by SCE runs on a single node, using either 1 or 2 parallel threads, the CPLEX code can be leveraged to run on unlimited parallel threads on a single node.

LLNL Contribution: Using Supercomputers to Speed Execution of the CAISO/PLEXOS 33% RPS Study

The current LLNL effort seeks to improve the solution times of the CAISO/PLEXOS 33% RPS model and to use this as a demonstration of the potential benefits of high-performance computing applied to real-world energy problems. We obtained the most recent version of the model from CAISO in August of 2010, at the behest of the CPUC. Our first challenge was to modify the PLEXOS code to run on linux-based supercomputers, as it is natively a Windows .net-based executable. This was accomplished in September 2010, in partnership with the research and development staff at Energy Exemplar, the Australian parent company of the PLEXOS software. At the same time, we were able to leverage our existing connections with IBM to obtain unlimited copies of CPLEX for research purposes.

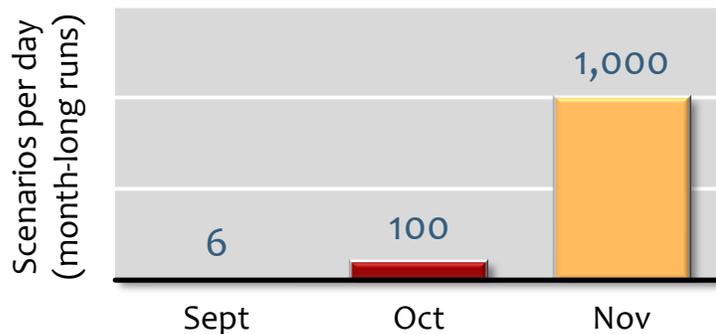


Our subsequent contributions were twofold: first, identifying improvements to the model formulation itself that led to an improvement in solution time; and second, demonstrating the capability to run thousands of scenarios at once using our massively parallel computing framework. For the first of these, we worked with the staff at Energy Exemplar to identify changes to the underlying mixed-integer programming formulation that enabled integer solutions to be found more quickly, while not sacrificing the accuracy of the solutions. The following graph shows the improvements in obtained solution times for each of the months in the simulated year; the overall average is a 4x speedup in execution time.



Improvements in solution times for each of the months in the CAISO/PLEXOS 33% RPS study

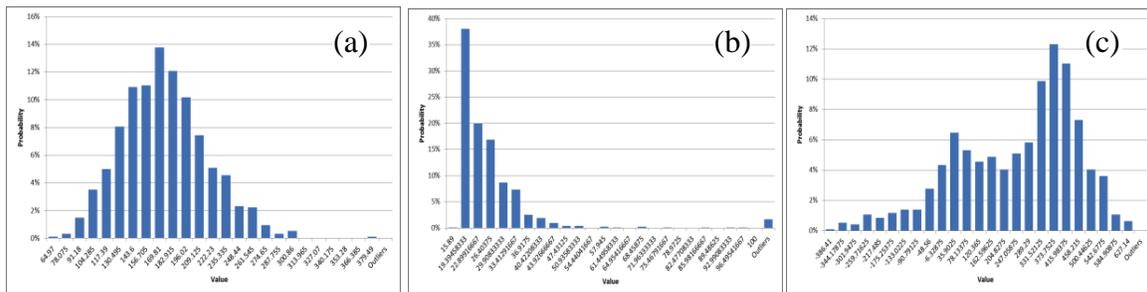
To greatly increase the number of scenarios that could be run at a time, we implemented customized scripts to execute the PLEXOS package using the underlying CPLEX MIP solver. By November 2010 we were able to execute a massively parallel suite of 1000 runs at a time, performed nearly simultaneously on the Hyperion supercomputer at LLNL. This demonstrates over a hundredfold increase in the throughput of scenario months that could be run in a day, as compared to the current setup at CPUC/SCE, which can execute one monthly scenario in a single day on at most 6 different processors.



Improvements in number of simultaneous scenarios able to be executed, September-November 2010

The ability to run thousands of scenarios in parallel also enables the ability to perform large-scale *sensitivity analyses* on the data and results, a capability the study previously lacked. To demonstrate the potential usefulness of such analyses, in November 2010 we performed a massively parallel suite of runs

varying the random seed used to control the pattern of generator outages. (In the current study, such outages are prespecified to a fixed pattern, whereas in the real world these outages are random and occur according to a probability distribution.) The results of this sensitivity study were enlightening: for some output parameters of interest, the impact of varying the generator outages caused a normally distributed pattern of ensuing results; for others, this distribution was exponential in shape; and for yet others, the distribution appeared bimodal (see below). This implies that constraining events such as generator outages to a fixed pattern can have a potentially large impact on the results. For instance, if the fixed pattern was chosen from one of those on the tail of an exponential distribution, this would produce markedly different results than a pattern from the mean. Altogether, this suggests that such sensitivity analyses may be very important in building confidence in the results obtained overall.



Patterns of ensuing results for different output parameters in the sensitivity analysis:

- (a) is an approximately normal distribution for the load following down price;
- (b) is an exponential distribution for the MUNI regulation down price;
- (c) is a bimodal distribution for Southern CA Edison's net export.

Starting in December 2010, our team has been collaborating closely with the CPLEX development team, to identify methods of better exploiting parallelism during the course of the MIP solution algorithm. Thus far, the benefit of executing the algorithm across multiple threads on the same node has been minimal, which was unexpected but is not uncommon in such formulations. Mathematically, the majority of the time spent by the solution algorithm is in a part of the code which does not parallelize well (known as the root node), rather than in the part for which effective parallel algorithms have been developed (the branch-and-bound tree). In January 2011 we identified some algorithmic settings which speed performance at the root node, leading to a (preliminary) 2x speed improvement. These collaborations are still ongoing.

Future Directions

In the future, our team at LLNL is interested in further the improving algorithmic performance of the model, specifically to exploit opportunities for massive parallelism during the solution of a single problem instance. This sort of distributed-memory parallelism has been applied only infrequently to mixed-integer problems in the past, and is on the cutting edge of research in the field. Our contacts at the CPLEX development team possess a prototype platform for solving MIP problems in a distributed memory fashion, and we are in the process of negotiating a research agreement to use this software.

Another area we would very much like to pursue is that of creating a user-friendly software front end, to allow native Windows PLEXOS users to interface with our high-performance computing environment in a seamless fashion. The current port of PLEXOS to linux is performed via command-line operations, whereas the Windows version of PLEXOS uses a customized GUI. We would like to interface the GUI with our machines to enable current Windows PLEXOS users to engage with a high-performance computing environment that is accessible from their desktops.

The success to date of the LLNL involvement in the CAISO/PLEXOS 33% RPS study has proved instrumental in demonstrating the value of high-performance computing to utilities and stakeholders in the state of California. We envision that the continuing success of this effort will lead to more direct support of this and future energy projects. This thus represents an exciting opportunity for both LLNL and energy stakeholders who can greatly benefit from these high-performance computing capabilities.