Design of a Bid-based Wholesale Energy Market, Ancillary Services and Capacity Market in Chile

Product 1: Tasks 1 & 2

FINAL VERSION

Prepared by ECCO International
San Francisco, California

For the

COORDINADOR ELECTRICO NACIONAL

ECCO International
268 Bush Street, Suite 3633
San Francisco, CA 94104
Phone: 415-731-9113
Fax: 415-731-9114

Table of Contents

List of Figures

List of Tables

1 Introduction to the International Benchmarking Study

Per ECCO's contract with the CEN in this Report we have chosen to present and analyze the following four (4) US ISOs for the international benchmark study.

- CAISO
- ERCOT
- PJM
- ISONE

This report will constitute Product 1. Product 1 is comprised of two Tasks (1&2).

In Task 1 we begin with a discussion of general market design principles that guided the market restructuring in the US and introduce a generic market design, often referred to as the Standard Market Design (SMD), that has been adopted by the US Federal Energy Regulatory Commission that captures these general principles. The specific implementations in different ISOs we present in details are essentially variants of the SMD. For accuracy some of these manuals and protocols are quoted verbatum with proper editing and clarifications. In particular, we focus on the challenges and solutions of integration of new low carbon assets. These challenges and solutions in existing market architecture, coupled with market reforms of these markets are the basis for lessons and recommendations we will offer for the Chilean energy market in subsequent reports.

The plan is to incorporate such lessons in the follow-up Tasks 3, 4 & 5 of this Project. The selection of these specific ISOs is intentional with the objective to cover a wide range of market conditions and Variable Resource Energy (VRE) penetration levels. On one hand we have the CAISO and ERCOT where we have a massive penetration of VRE and other low carbon assets and we also have PJM and ISONE where the VRE penetration is lower and modest at the present time. We plan to provide useful recommendations from the comparison of the benchmarking study of these four (4) ISOs in subsequent reports.

In Task 2 we will identify two (2) bid-based market design models that could be adopted in Chile.

Currently, the Chilean energy market is based on a cost-based, auditable short-term market. This approach has been proven to be, based on our experience, an easy way to produce a credible price for dispatching units in real time and clearing imbalances. Further this market is expected to limit the ability of suppliers to exercise unilateral market power while bid-based markets can be susceptible to the exercise of unilateral market power, depending on market conditions and unless other mitigation rules are implemented. In practice CEN used the cost-based market on using hydro resources to minimize the total audited cost of thermal generation units. However, a serious problem with the audited, cost-based market is the fact that Market Participants are denied the opportunity to determine the opportunity cost of water instead of exclusively relying on the CEN's SDDP solution.

Experience also has shown that it may be difficult to estimate the true opportunity costs of generators in cost-based markets, resulting in skewed prices and schedules and economic inefficiencies. ECCO strongly believes that a transition to a well-designed, bid-based energy market architecture will a) benefit the CEN market, b) assist hydro owners manage their opportunity costs and c) foster the transition to a decarbonized economy.

In a cost-based market architecture these problems are expected to magnify as the penetration of VRE generation in the systems increases since these assets exhibit physical constraints due to weather conditions. This is a clear case where the opportunity costs diverge from fuel costs due to energy or weather-based physical constraints.

In summary, in this Task we'll analyze in detail the benefits of a bid-based energy market and advantages and disadvantages compared to the current cost-based short-term market. We consider and analyze two (2) variations of bid-based design models. The first is based on the marginal pricing design and compared with the second model which is based on the pay-as bid pricing design. The focus is on their market efficiency implications along with the efforts for a practical implementation.

For each ISO market we present and analyze the following market elements:

1. Day-Ahead Energy Market
2. Reliability Unit Commitment
3. Intra-Day Market (Fifteen Minute Energy Market)
4. Real Time Market
5. Ancillary Services Market
6. Co-optimization of Energy and Reserves
7. Locational Marginal Pricing
8. Congestion Revenue Rights (CRR) Markets Or FTR Markets
9. Multi Settlements Market Design
10. Mass Low Carbon Generation and Market Impacts
11. Flexibility and Resource Adequacy
12. Market Power Mitigation, and
13. Financial Markets and Financial Market Players (Virtual Bidding)

We use the CAISO as the basis of our analysis and avoid to repeat some common elements which all markets have such as optimization aspects, pricing run, etc.

2 General Principles and The Standard Nodal Market Design

While differet regional transmission organizations (RTOs) in the US differ in their implementation and elements of their wholesale market design, they all follow a similar philosophy and general structure, often referred to as the Standard Market Design, which has been prescribed in the Federal Energy Regulatory Comission (FERC) Orders 888 and 2000 that guided the electricity industry restructuring in the US.

In this Section we describe the basic element of this standard wholesale market design which will then be refined with more granular details implemented in four selected ISOs. These refinments represents the end result of twenty-year market evolution and the best practices that have emerged. Some of the market components may be simplified if adopted in the Chilean electricity industry to reflect specific aspect such as the legacy industry structure and protocols, fuel sources, security considerations, scale of the system and network realities, and renewables integration. The basic design hinges on a set of products and contractual obligations that support the coordination and operation of the electricity system by the system operator. These products are procured and priced through a series of markets managed centrally by an independent system operator so as to meet load reliably, ensure contestability, efficiency incentives and nondiscriminatory access to the transmission network which is operated as a common carrier. The ISO owns no generation or transmission assets and has no financial interest in market outcomes. The ISO controls the transmission network by issuing dispatch istructions based on its procured resources and sets prices for the sellers and wholesale customers. Transmission owners and must run generators are compensated based on regulated tariffs approved by the regulator.

2.1 Wholesale Market Products

- Day-ahead forward energy
- Real-time balancing energy
- Forward capacity availability
- Financial transmission rights/obligations (FTR)
- Regulation – AGC (capacity and "mileage")
- Flexible Ramping
- Operating reserves:
  - Spinning
  - Non-spinning
  - Replacement

2.2 Bidding and Market Clearing

All the products listed above are procured through periodic auctions conducted by the system operator where generators submit price bids that can be changed with prescribed frequency (daily, hourly, half hourly) and specify operating constraints (min load, max load, ramp rate, min up time, min down time) which can be varied periodically. In principle the bids and operating constraints can be prescribed by the regulator based on technical characteristics of the generation plants and verified cost calculations. Such systems are referred to as cost based pools and has been implemented in Chile and other countries. However, such an approach may be inefficient since it does not allow generators to reflect their opportunity cost in their offers. For instance, the opportunity cost of a gas turbine with a “take or pay” fuel contract is zero and so should be its bid which will assure its dispatch as a price taker. Furthermore, a coal unit which should avoid shut down due to high startup cost and minimum down time should be allowed to bid a negative price during periods of over generation due to low demand or access renewables output. During such periods spot prices may be negative which is the correct price signal encouraging load shifting, and investment in storage. The California ISO which will be discussed below, allows spot energy prices to go down to minus $300/MWh and so do other ISOs. Such prices would not arise in a cost-based pool.

The markets for each of the products are cleared using a uniform price auction with the clearing price paid to all accepted product units and charged to users of the products, set to the bid submitted by the marginal accepted (unconstrained) product unit. According to auction theory, such a pricing rule provides the correct incentive for bidders to reveal in their offer their true opportunity cost and hence result in an efficient clearing price reflecting marginal opportunity cost. The economic rational for marginal cost pricing is described in detail in Appendix A. Understating the true opportunity cost may result in selling at a loss, while overstating it may result in missing out on a profitable sale.

The alternative approach of “pay as bid” which pays each accepted unit offered their offer price creates incentives for bidders to overstate their opportunity cost in an attempt to obtain the highest paid price. Such an auction, in a repeated setting results in a flat supply curve so that the “pay as bid” strategy does not achieve the primary objective of reducing purchase cost. Furthermore, because of forecasting errors, the procurement may be socially inefficient because some relatively cheap units that should have been procured may have overstated their price beyond the market clearing threshold. Pay as bid may work, however in a cost-based pool where the bids are closely regulated to reflect marginal cost but as mentioned above it is not practical to regulate bids so as to reflect opportunity cost. It should be noted however, that in spite of all economic arguments to the contrary, some systems, like in the UK, employed a “pay as bid” approach in procuring balancing energy and used a sale price based on an average buy price to charge the buyers. This approach was embedded within the performance-based regulation scheme and there is no empirical evidence to demonstrate the success or shortcomings of that scheme.

2.3 Multiple Markets

The centralized wholesale market operated by the ISO is organized as a sequence of separate markets for the different products as follows:

- Pre-day-ahead markets (pre-DAM)
- For transmission rights: FTRs
- For generation capacity availability
- Bilateral long term energy contracts (not involving the ISO)
- Day-ahead forward energy market (for next 24 hours)
- Based on centralized, bid based security constrained unit commitment
- Reliability unit commitment after DAM market close
- Bid cost recovery mechanisms for economic (not self-scheduled) bids
- Simultaneous market-clearing auctions for energy, ancillary services and congestion
  - Locational DAM uniform prices for energy
  - Multi-part bidding (startup, no load and multi-segment supply function
  - Allowed self-scheduling (price taker for congestion rents)
- Real-time balancing energy market (in US bidding every 15 minutes, rebalancing and RTM pricing every 5 minutes)
  - Energy markets are locational (nodal-based) marginal price (LMP)
  - All bid-based markets clear at uniform marginal prices
  - Convergence bidding for arbitrage between DA and RT prices (open to financial players)
- Market power mitigation
  - Screens are based on structural competitiveness or conduct and Impact
  - If screen fails then the bids are replaced by cost based "default energy bids"

2.4 Detailed Market Design Elements

2.4.1 Day Ahead Energy Market

The day ahead energy market is operated as a bid-based security constrained unit commitment. producers submit multi part bids for each hour or for the entire 24 hours of the next day specifying startup cost, no load cost and a multi-block energy supply function. Producers also specify technical constraints including min load, max load, ramp rate, forbidden operating regions, minimum up time, minimum down time. The extent of verification of technical parameters and frequency of allowed changes in the costs and technical parameters can be determined by the market designers in consultation with the market participants and system operator. The objective of the design is to allow sufficient flexibility for the producers to reflect their real costs and constraint while limiting the opportunities for market manipulation. Generators who wish to schedule bilateral physical transactions can do so and they become price takers for congestion charges if any. physical bilateral transactions are represented in the unit commitment as very large buy bids (above the bid caps) and very low (negative) sell bids so that these transactions are always scheduled before any economic bids by the unit commitment algorithm. The overall schedule for each of the 24 hours is determined by the following optimization algorithm which is run by the system operator:

Minimize Σ (fuel cost + no-load cost + startup cost)

Subject to:

- Load balance constraint at each node
- Unit output constrain for each generator
- Unit ramping limits for each gen
- Unit min up time and min down time for each gen
- Transmission constraints (DC approximation with thermal proxy limits)
- Reserves margin requirements
- Contingencies (n-1)

The above unit commitment optimization determines the hourly day ahead schedule for each generator, the reserves and the prices which may vary by location and are based on the local shadow prices corresponding to the nodal load balance constraint. The award quantities and corresponding prices are treated as financially binding forward contracts settled accordingly and any quantity deviation from such contractual obligations whether instructed or uninstructed is settled separately and may be subject to penalties. To the extent that the regulator wishes to socialize the cost differentials associated with some constraints (e.g. local congestion) or cost components (e.g. emergency use of nonstandard fuel), a separate pricing run may be used with the socialized constraints being relaxed and the socialized cost components replaced by the generic costs. Such a pricing run will produce a day ahead forward hourly settlement price at each node reflecting a Locational Marginal Cost (LMP). Socialization of constraints or various cost components may result in an uplift that is needed to be added to the LMP so as to cover the unaccounted costs.

2.4.2 Setting the LMPs

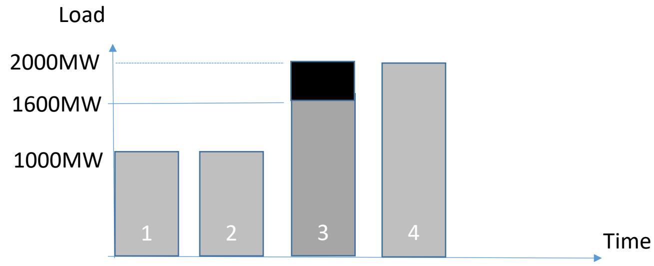
From a theoretical perspective the LMPs are determined by the “dual prices” corresponding to load variations. In other words, the change in the dispatch cost due to a marginal change in load at each location. These prices are automatically produced by typical commercial software but it is useful to understand intuitively how these prices are being affected by generation costs and constraints. First it is important to recognize that marginal cost should NOT average cost and should not include any amortized fixed or startup cost. To the extent that units are allowed to bid their marginal cost in a central auction (rather than submitting regulated cost based bids) it will be a mistake on their part to try to recover startup cost by markeing up their bids since that may result in not being dispatch when the market clearing price exceeds the fuel cost. In principle inframarginal profits when the clearing price is set by more expansive units should cover startup and fixed cost components, however when this is not the case a Bid Cost Recovery (BCR) provision described below will assure cost recovery over a 24 hour period. If the dispatch is not constrained by either generator limits on minimum output or intertemporal constraints, then the optimal dispatch follows merit order loading and hence the marginal cost, and therefore the LMPs are set to the cost of the most expansive MW hour produced at a location. This however, is not the case when some generators are loaded as a block and their output cannot be adjusted incrementally, when slow ramping generators are dispatched in certain hours in order to be available in subsequent hours or are producing energy while in the process of being ramped down, or when expansive generators are operating at minimum load for various considerations.

In principle one should specify all the constraints in the dispatch optimization and rely on the software to determine the marginal prices. Working around the dispatch optimization by making direct dispatch decisions, rather than specifying the constraints that such decisions are trying to address is ill advised and it will distort the correct economic price signals. For example, units that are dispatched at minimum load due to startup or minimum up and down time consideration should not participate in setting the LMP. Such constrained units will not be moved when load is increased or decreased incrementally and therefore the marginal cost of such units is irrelevant to the incremental cost of serving the marginal load increments.

If a unit is dispatched partially or at its minimum level during an off-peak hour (night) because it is needed during a peak hour (morning) its marginal cost does not necessarily set the LMP in the hours it produces even if it produces the most expansive MW in these hours. The LMP in each hour will be adjusted by a shadow price (Lagrange multiplier) on the ramp constraint that affects the intertemporal dispatch decisions. In the Example below, for instance, the off-peak energy is produced at a marginal cost of $30/MWh but the LMP in hour 2 is negative $10/MWh. This reflects the fact that an additional MW of load in hour 2 actually reduces the overall system cost because it allows displacement in hour 3 of a MWh produced by the fast ramping generator at $70 with a MWh produced by the slow generator at $30. Since serving that additional MWh of load in hour 2 cost $30, the net savings is $10 and hence the LMP in hour 2 is negative $10 which is owed to the incremental load. On the other hand, the slow ramping generator benefits in hour 3 from the fact that the LMP is set by the fast start unit at $70/MWh. So although it is penalized for its production in hour 2 at a negative price, its overall payment in hour 2 and 3 for the first 1000MW output averages $30/MWh which equals its production cost, while for the additional 600 MW it produces in hour 3 it is paid $70/MWh yielding an inframarginal profit of $40/MWh that can go toward availability payment covering fixed and startup costs. At the same time load and fast units that are not bound by ramp constraints and are sufficiently flexible to respond to the hourly prices, are getting the appropriate price signal that will induce efficient price response in each hour.

Illustrative Example:

- Two Types of Generators:
  - 800 MW of Fast start @ $70/MWh
  - 2100 MW of Slow ramp @ $30/MWh with max ramp rate of 600MW/hr.
- Four hours of dispatch with 2 hours off peak load 1000MW and 2 hours peak load 2000MW
- Optimal dispatch:



Fast	0	0	400	0	MW
Slow	1000	1000	1600	2000	MW
SMP	30	-10	70	30	$/MWh

Figure 2-1: Setting LMP

2.4.3 Bid Cost Recovery (BCR) Mechanism

The BCR mechanism is designed to incentivize producers to submit economic bids rather than physical schedules in the day ahead market. To achieve that it guarantees that any producer scheduled by the system operator will at least cover its, as bid, startup, no load and fuel costs. Because the unit commitment optimization is a non-convex mixed integer programming problem it is possible that a unit will be dispatched in a way that is optimal for the system as a whole although the unit's income from energy payments based on the LMP will not cover the unit's total cost of operation which includes startup and no-load costs. When this happens to a unit that was dispatched by the system operator based on economic bids it will receive a make whole payment that will cover any revenue shortfall over a 24-hour period. Such makewhole payments are covered through an uplift on the wholesale energy price. Units that self-schedule and submit physical schedules to the system operator are not entitled to such makewhole payments. Makewhole payments may also be awarded to units whose actual fuel cost is not reflected in the LMP due to a system emergency as discussed above, or to units that are dispatched out of merit order due to constraints but their cost is not reflected in the LMP which may be based on a pricing run where these constraints are not recognized[1].

BCR uplifts are common in all the US ISOs but they only amount to a small percentage of total wholesale energy cost. **According to the CAISO 2023 Market Monitors Annual Report, BCR payment by CAISO to California resources in 2022 was $255 Million which is 1.2% of total CAISO wholesale energy cost. Another $42 Million in BCR payments went to units in the Western Energy Imbalance Market which is also managed by the CAISO**. However, the rise in renewables dispatch led to an increase in BCR payments and new approaches are being adopted for dealing with the impact of nonconvexities on the market clearing prices.

2.4.4 Real Time Market

The real time market is a market for incremental and decremental deviations from the day ahead schedules which are traded for short term load balancing. In the US such deviations are dispatched every five minutes based on Optimal Power Flow (OPF) software that adjusts the output level of generators so as to meet the load balance constraint at minimum cost subject to technical constraints of the generators and transmission constraints that are represented by Kirchhoff laws. Specifically, the OPF solves the following optimization problem:

Minimize Σ (Generator Fuel Cost)

Such that:

- Energy balance (net supply = load at each node)
- Generator limits (including dynamic limits such as ramp rates)
- Transmission Constraints (AC model with voltage and thermal limits)
- Reserve requirements

The energy adjustments are co-optimized with procurement of AGC capacity, spinning reserves capacity and ramping reserves. Any load following within the five-minute dispatch intervals are met with AGC and replenished with spinning reserves and ramping reserves. Locational real time marginal prices are determined by the nodal shadow prices produced by the OPF software. While shorter time intervals for dispatch optimization are desirable some systems reoptimize the dispatch every half hour and use suboptimal adjustment procedures and reserve deployment for load following between the economic dispatches. In the US real time markets producers are allowed to submit bids for real time deviations every fifteen minutes. Such frequent rebidding has been introduced in FERC Order 764 which was intended to facilitate the integration of renewable resources

[1] Panagiotis Andrianesis, George Liberopoulos, George Kozanidis, Alex Papalexopoulos, "Recovery Mechanisms in Day-Ahead Electricity Markets with Non-Convexities – Part I: Design and Evaluation Methodology," IEEE Transactions on Power Systems, Vol. 28, no. 2, 2013, pp. 960-968, available online: doi: 10.1109/TPWRS.2012.2207920

such as wind which are intermittent and can be forecasted accurately over short time periods. Allowing intraday and intrahour rebidding of deviation prices improves efficiency as it allows producers to reflect short term variability in fuel cost (e.g. intraday gas price fluctuations) in their bids.

FERC Order No. 764, issued in June of 2012, was designed to address the increase in VERs being brought online and to remove barriers for VER integration. Specifically, in Order No. 764, FERC required transmission providers to give customers the option of adjusting their transmission schedules at 15-minute intervals, but also required VER generators to provide transmission owners with certain meteorological and forced outage data in order to improve power production forecasting.

After issuing Order No. 764, FERC received multiple requests for rehearing and clarification. Addressing these requests in Order No. 764-A, FERC affirmed all the basic determinations of Order No. 764 and provided clarification on multiple issues, including: that the intra-hour scheduling reform applies to all transmission customers that schedule transmission service under an Open Access Transmission Tariff; that when a transmission customer using 15-minute scheduling is taking service from a transmission provider using an hourly imbalance charge, the transmission provider must average the imbalances of each 15-minute scheduling period over the entire hour; and that a firm transmission schedule has curtailment priority over a non-firm transmission schedule using a 15-minute interval schedule. FERC also denied requests to allow transmission providers to unilaterally amend existing Large Generator Interconnection Agreements to include data reporting requirements for existing interconnection customers,

On September 19, 2013, FERC issued Order No. 764-B, further clarifying its Integration of Variable Energy Resources ("VER") rule originally established in Order No. 764. Powerex Corporation ("Powerex") and Iberdrola Renewables, LLC ("Iberdrola Renewables") had sought clarification, or in the alternative rehearing, of issues related to e-Tagging and the Bonneville Power Administration's ("Bonneville") practice of curtailments pursuant to its Dispatcher Standing Order.

However, in markets where intraday volatility in fuel prices is limited (**like in Chile**) and renewable resources are insignificant it is acceptable to use the day ahead bids for real time optimization and adjust locational real time (i.e. half hourly) prices based on changing load and supply conditions but the same day ahead price bids.

2.4.5 Two Settlement Systems

In some systems like Australia the day ahead dispatch and prices are treated as advisory while the settlements for the entire quantities produced or consumed in each hour or sub hourly period are determined ex-post based on the real time prices. Such an approach which is referred to as a single settlement system is flawed since it creates strong incentives for manipulation of the real time prices through withholding of capacity and exploitation of technical constraints. Settlements that are based entirely on ex-post real time prices also fail to provide a reliable Ex-ante price signal that will encourage demand response and will facilitate bilateral contracting by providing a reliable settlement price with reduced basis risk.

A superior approach that has been adopted in the US markets and in many other systems is the two settlement approach where the day ahead hourly dispatch and prices are financially binding and used for settlement of day ahead hurly awards and purchases, whereas the real time prices are used to settle deviation from the day ahead commitments regardless of whether the deviations are instructed, uninstructed, intentional or accidental. The real time prices provide the true value of real time energy and hence settling the deviations based on these prices automatically rewards deviations that improve system economic performance and penalizes deviations that increase system cost.

In the US markets the two settlement approach is enhanced by allowing financial players with no demand or production facilities to arbitrage the price differences between the day ahead and real time markets. Such arbitrage, referred to as convergence bidding or virtual bidding takes the form of supply or demand positions (bids) in the day ahead market that are automatically closed (as price taking bids) in real time. Such virtual positions allow producers and consumers to take positions in the day ahead market but settle them at real time prices without having to withhold their bids. This helps the system operator run the system reliably by reducing market participant's incentives for strategic real time deviations and surprises. Empirical studies have also shown that virtual bidding improves market efficiency by increasing liquidity and converging prices in the day ahead and real time market.

The arguments in favor of a two settlement system are similar to the arguments in favor of long term bilateral contracting. Day ahead prices are more stable and less volatile than real time prices that tend to be much more volatile. In a two-settlement system day ahead dispatch and day ahead prices are financially binding. The real time prices which are volatile, due to changes in weather conditions, contingencies, ramp constraints and other last minute variabilities and un-anticipated events, are only used to settle the differences between the day ahead schedules and real time production and consumption. This eliminates incentives for generators to take advantage of last-minute shortages in order to raise real time prices which would be applied to their entire production. In a two settlement system the real time prices are only applied to the differences between the real time production or consumption and the day ahead financially binding commitments. Increase in VER which raise uncertainty and real time volatility make the argument

in favor of a two settlement system with the real time dispatch within 15 minutes (or 5 minutes) of real time even stronger.

We understand that the market in Chile is based on a single settlement system and given the analysis above we plan to recommend a change to a two settlement market architecture.

2.4.6 Ancillary Services Procurement

Ancillary services consist of various reserves needed to provide load balancing, voltage stability frequency regulation, ramping capability and overall reliability. These are provided by synchronized and non-synchronized reserves differentiated by response time, governor settings set to respond to frequency signals for automatic generation control and control of reactive power needed to assure voltage stability. Some ancillary services such as black start capability can only be provided by specific units equipped for such tasks and therefore such capability is procured through long term contracts. However, most ancillary services can be provided by all generation units either as distinct products procured by the system operator through market mechanisms or as an obligation for self-provision of reserves which gives the system operator certain latitude in controlling units' output. From an economic perspective the main purpose of ancillary service markets is to procure the needed reserves at least cost from the generators that are in the best position to provide such reserves by providing compensation to these generators and avoiding "free ridership".

Compensation for reserves can be based on capacity and energy produced (or saved) when these reserves are deployed. In the case of AGC which is used for short term load following we distinguish between UP and DOWN regulation so the net energy does not properly reflect performance. Therefore, in the US, FERC order 764 prescribes performance payment based on "mileage" which is the term used to denote the sum of absolute values of up and down energy adjustments. Other types of reserves such as spinning, non-spinning and replacement reserves, which differ only in response time, are compensated for capacity and energy. However, some systems which co-optimize the dispatch of energy and reserves have opted to eliminate the capacity payments and compensate reserves held back out of merit based on lost opportunity cost The requirement for the different reserve categories is determined by the system operator based on reliability considerations which reflect the various contingencies. These reserve requirements may be allocated to generators in proportion to their output under self-provision or procured by the system operator through an auction and the procurement cost is prorated as an uplift to the wholesale energy price.

Procurement auctions for reserves can be uniform price or pay-as-bid and generators submit bids offering capacity of the different reserve types at prices that reflect their opportunity cost. For AGC the bids may specify a capacity price and mileage price while for other reserve types the energy price is based on the LMPs. An important aspect of the different reserve types is their downward substitutability. In other words, any unit used to provide regulation can also provide spinning reserves and any unit providing spinning reserves can provide non-spinning reserves, etc. Furthermore, units producing energy are interchangeable with units held for reserves. When congestion constraints the potential flow of reserves, as is the case in New York, the reserves hierarchy may also be locational reflecting feasible flow directions. The procurement of all reserve types is co-optimized subject to the substitutability constraints to meet overall reserve requirements at minimum (as bid) cost. The generators providing these reserves are compensated based on their bid (in a pay-as-bid setting) or the uniform clearing price for the reserve type they offered.

The massive integration of intermittent renewables (wind and solar) in California created a need for a new type of reserve referred to as Flexiramp, which was approved by FERC and is intended to assure sufficient real time ramping capability to follow rapid net load fluctuations. Such reserves are not bid as a separate product but are obtained by holding back energy offers from flexible generators that may be in merit order (i.e., their offer prices are below the LMP). In such cases the generators are paid opportunity cost for lost profit which is the LMP minus their bid price times the withheld energy output.

2.4.7 Resource Adequacy Mechanism and Capacity Remuneration

The economic “gold standard” for assuring resource adequacy is an energy only market where energy prices are allowed to rise high enough so as to reflect scarcity. Scarcity prices provide the necessary compensation to cover fixed cost of capacity and incentivize investment if necessary. A detailed description of the underlying theory is provided in Appendix A. Energy only markets for electricity have ben implemented in Texas, New Zealand and Australia where wholesale spot prices for electricity can rise to $9000/MWh (in Texas) or more (in Australia). However, in system where energy prices are capped or bids are constrained by regulated cost based pools, energy revenues are not sufficient to cover investment costs (“missing money”) and some form of capacity remuneration or an alternative mechanism is needed to ensure resource adequacy.

Capacity payment per MW of committed generation capacity is a common approach to remunerate generation capacity for availability. There are many variants to this approach with the payment determined administratively or determined through an auction as in PJM or ISO New England in the US. The capacity payment typically entail a must offer obligation with no restriction on the bid price for energy. However, the must offer obligation can be interpreted as a call option with a strike price at or below the energy bid cap. When the strike price is set below the bid cap, as was the case in the New England Forward Capacity Market (FCM) until 2015, the generator is liable for the difference between the LMP and the strike price known as Peak Energy Rents (PER). If the unit produces, then its excess revenues from selling energy above the strike price will offset the PER liability. However, if the generator is not available it will be liable for the energy cost at the LMP. The strike price in the original design of the New England FCM was set to a level attempting to estimate marginal cost of a generic Combustion Turbine but that turned out to be problematic because of fuel price fluctuations so in May 2015 the strike price was raise to $1000/MWh. An important feature of market-based capacity payments is allowing demand side resources to participate in the market. Allowing demand side participation amounts to permitting “opt out” for demand which is willing to curtail consumption in case of shortage and avoid the capacity charge which can be interpreted as an availability insurance premium.

One important aspect of capacity payment remuneration schemes around the world, whether determined administratively or via market mechanism, is that they are uniform and technology neutral. In other words, every MW of capacity regardless of the technology or whether it comes from generation, demand response or storage, get paid the same. Early attempts to have differential capacity payments based on technology (e.g. in Korea and other places) have proven to be ineffective and were changed to a uniform prices. The theoretical justification for having uniform capacity prices is that capacity payments should reflect the marginal cost of adding new capacity to the system which at the optimum must equal the loss of load probability tie value of lost load. Differences in the capacity cost of various technologies are covered by inframarginal profits of the different technologies when energy is settled at the LMP. This principle is described in detail in Appendix A.

Another resource adequacy mechanism which has been adopted in California, and could be appropriate for Chile, is an administrative “showing” approach. Under this scheme the regulator determines an annual and monthly capacity obligation for each load serving entity which is set to 115% of the historical peak load for that entity over the year and month. It is up to each load serving entity to enter into bilateral supply contracts with generation capacity and be able to demonstrate to the regulator that it has sufficient contract cover to serve its peak load. Specifically, in California a load serving entity must show at the beginning 85% of its contract obligation (i.e., 115% of peak load) for the year. And 100% of its monthly obligation at the beginning of each month. The California Public Utility Commission amended its resource adequacy requirement to include about 30% of flexible capacity (which can sustain a continuous 3-hour ramp) and prescribed amount of storage capacity (about 1500 MW total for California).

3 California Independent System Operator (CAISO)

3.1 Overview of the CAISO Energy Market

The California Independent System Operator runs one of the largest and most dynamic grids of the world. It manages the flow of buld electricity across 26,000 circuit miles of high-voltage power lines to utilities across California and in a small part of Nevada, ultimately reaching 30 million customers.

The ISO also operates California's wholesale electricity market, using state-of-the-art technology to match demand with the lowest cost energy available at any given time. The market runs an automated auction every day and every five minutes, across 9,700 price nodes, generating over 31,000 transactions a day.

As the only grid operator in the western US, the CAISO grands equal access to all market participants creating an economic based mechanism of diverse resources to compete in the power market. The CAISO also runs the Energy Imbalance Market (EIM) which gives western state utilities access to a real time trading market and sophisticated optimization technology, allowing them to provide the lowest cost power to their customers. The EIM market has proven to reduce costs, promote greater use of renewable energy and reduce greenhouse gas emissions for utilities in eight western states.

Another essemtial function of the CAISO is to provide transparent information about the state of the electrical system and market prices. The information helps its customers to effectively access the economics of the energy trading and manage the risks of participating in the wholesale market. The CAISO provides thousands of data points every day using a variety of mechanisms including a 100 of automated systems.

The CAISO currently manages 1,080 power plants, 239 million MWh per years and provides locational prices at 9,700 price locations.

3.2 Market Design basics

3.2.1 Market Participation

This section provides an overview of the basic elements of the CAISO energy market design and the participation in the market.

3.2.1.1 Scheduling Coordinators

Participants in the CAISO markets include generation companies, investor owned utilities, municipal utilities, federal agencies, state agencies, retail energy service providers, energy marketers, etc. All these participants have to be represented by Scheduling Coordinators (SCs) that are registered with the CAISO, except for Financial Transmission Rights (FTR) Owners which in the CAISO market are called Congestion Revenue Rights

(CRR) Owners that only need to be Business Associates. One participant can register multiple SCs with the CAISO. Many participants can do business with the CAISO using a single SC. The SC is responsible for bidding into the CAISO markets, reporting resource outages and derates to the CAISO, receiving market outcomes and dispatch instructions from the CAISO, submitting metering data to the CAISO, and settling market accounts with the CAISO.

3.2.1.2 Resource Identification and Participation

Resources include Generating Resources, Import Resources, Export Resources, Participating Load Resources, and Non-Participating Load Resources. Generating Resources may be single or aggregate resources. Non-Participating Load Resources are aggregate resources, with the exception of Transmission Ownership Right (TOR) and Existing Transmission Contract (ETC) loads that may be singe load resources. Participating Load Resources are hydro pump resources that are certified by the CAISO for real-time dispatch; they can participate in both Day-Ahead and Real-Time Markets. Each resource has a unique identifier regardless of the type of the resource. Each resource is mapped to a Location. The Location of an Import or Export Resource is a scheduling point at the external end of an inter-tie with the CAISO Control Area. The Location of a single Generating or Load Resource is a single network node, whereas the Location of an Aggregate Generating or Load Resource is an aggregate node mapped to multiple network nodes. In order to participate in the CAISO markets, each resource is represented by only one SC for any given Trading Day. An SC can represent multiple (also called a portfolio of) resources.

3.2.1.3 Generation Distribution Factors

An Aggregate Generating Resource is a logical unit that represents a collection of physical units that must be scheduled together. A typical example of an Aggregate Generating Resource consists a group of hydro units in the same water system. Aggregate Generating Resources must submit Generation Distribution Factors (GDFs) for each hour they submit bids in the CAISO markets. The GDFs describe how the energy output from the Aggregate Generating Resource is composed from the energy output of the associated physical units.

3.2.1.4 Load Distribution Factors

An Aggregate Load Resource is a logical load that represents a collection of individual loads that must be scheduled together. Aggregate Participating Load Resources must submit Load Distribution Factors (LDFs) for each hour they submit bids in the CAISO markets. LDFs describe how the energy consumption of the Aggregate Load Resource is composed from the energy consumption of the associated individual loads.

Non-Participating Load Resources may only bid in the Day-Ahead Market. Non-Participating Load Resources under default load aggregation are always Aggregate Load Resources on specific Load Aggregation Points (LAPs), also referred to as Load Zones, and they need not submit LDFs because the CAISO will use default LDFs from its LDF library for them. The LDF Library will contain historical LDFs for the Load Zones for various time periods.

3.2.2 Commodities

This section describes the commodities traded in the CAISO markets. These commodities may be bid into the CAISO spot markets (the Integrated Forward Market, the Residual Unit Commitment, the Hour-Ahead Scheduling Process, and the Real-Time Market) as applicable.

3.2.2.1 Energy

Generating and Import Resources can sell Energy to the CAISO markets. Energy is measured in MWh. If a unit's output for 15 minutes is 60 MW, the energy produced by this unit in that 15 minute interval is 60 MW $\times$ ¼ hr = 15 MWh.

Generating Resources registered as pumped-storage hydro resources can additionally buy Energy from the CAISO markets to serve their pumping load, and sell Energy in the Real-Time Market by reducing their pumping load.

Non-Participating Load and Export Resources can buy Energy from the CAISO markets.[2] Participating Loads Resources (hydro pumps) can buy Energy from the CAISO markets to serve their pumping load, and sell Energy in the Real-Time Market by reducing their pumping load.

3.2.2.2 Ancillary Services

Ancillary Services (AS) are services provided by certain resources using their capacity that meet specific technical requirements. Energy is produced when the capacity is dispatched as Energy in real time.
The CAISO procures the following four types of Ancillary Services:

- Regulation Up (Reg-Up)
- Regulation Down (Reg-Down)
- Spinning Reserve (Spin)
- Non-Spinning Reserve (Non-Spin)

---

[2] Non Participating Load Resources may bid only in the Day-Ahead Market.

Resources must be certified to provide Ancillary Services. Certified resources may self-provide and/or bid ancillary services in the CAISO markets.

3.2.2.2.1 Regulation

Regulation Up and Regulation Down are two separate Ancillary Services. Regulation (Up and Down) capacity is not dispatched by the Real-Time Market (RTM), but by the Automatic Generation Control (AGC) function of the Energy Management System (EMS) to maintain interchange schedules and system frequency.

Generating Resources that meet the necessary technical requirements can provide Regulation. Import and Export Resources registered as dynamic inter-changes can also provide Regulation.

3.2.2.2.2 Contingency Reserves

According to the Western Energy Coordinating Council (WECC) Minimum Operating Reliability Criteria (MORC) definition, Operating Reserve consists of Regulation and Contingency Reserves, which in turn consist of Spinning Reserve and Non-Spinning Reserve. Contingency Reserve is unloaded generation capacity reserved for contingencies.

The CAISO does not treat all Spinning and Non-Spinning Reserves as Contingency Reserves. The provider of Spinning and Non-Spinning Reserve can specify in the Day-Ahead Market (DAM) a “contingency flag” to indicate that the Spinning and Non-Spinning Reserves, if accepted, should be treated as Contingency Reserves. Otherwise, the Spinning and Non-Spinning Reserve capacity can be dispatched in real-time regardless of contingency conditions, provided that the CAISO continues to maintain adequate overall levels of Operating Reserve. Any spinning and Non-Spinning Reserves procured in the RTM would always be Contingency Reserves.

Spinning Reserves can be provided by online Generating Resources and by Import or Export Resources.

Non-Spinning Reserves can be provided by Generating Resources,[3] by Import or Export Resources,[4] and by Participating Load Resources.

---

[3] Both online and offline Generating Resources can provide Non-Spinning Reserve in the Day-Ahead Market, but only offline Generating Resources can provide Non-Spinning Reserve in the Real-Time Market. Offline Generating Resources must be capable of starting, synchronizing with the grid, and delivering the expected Energy within ten minutes.

[4] Import and Export Resources can provide Non-Spinning Reserve in the Day-Ahead Market, but only Spinning Reserve in the Real-Time Market.

3.2.2.3 Reliability Unit Commitment Capacity

Reliability Unit Commitment (RUC) capacity is available capacity (not scheduled for Energy or reserved for Ancillary Services in the Integrated Forward Market) secured by the CAISO in the Day-Ahead Market (DAM) to meet the difference between the load forecast and the scheduled load in the Integrated Forward Market (IFM).

Generating, Import, Export, and Participating Load Resources may bid RUC capacity in the DAM.

3.2.2.4 Congestion Revenue Rights

A Congestion Revenue Right (CRR) from a source Location to a sink Location, measured in MWh, is the financial right to receive a portion of the congestion revenue generated by congestion in the IFM from that source to that sink. Both source and sink Locations may be aggregate Locations, however, all relevant Price Nodes must reside in the CAISO transmission network including inter-tie scheduling points.

Most CRRs are “obligation” CRRs, although some are “option” CRRs. The entitlement of each MWh of an obligation CRR from a source to a sink is equal to the algebraic difference (may be negative) between the marginal congestion components of the IFM Locational Marginal Price (LMP) at the sink and the IFM LMP at the source. The entitlement of each MWh of an option CRR from a source to a sink is equal to the difference between the marginal congestion components of IFM LMP at the sink and the IFM LMP at the source, if this difference is positive, otherwise there is no entitlement or obligation. CRRs are allocated and auctioned yearly (or seasonally) and monthly.

3.2.3 Spot Markets and Timelines

This section describes the various CAISO spot markets.

3.2.3.1 Day-Ahead Market

The Day-Ahead Market (DAM) is a market for trading Energy, Ancillary Services, and RUC capacity for the next Trading Day that starts at the coming midnight, and ends at the following midnight. The bid submission for the DAM is allowed as early as one week ahead and up to 10:00 am one day ahead of the Trading Day. The results of the DAM are published by 1:00 pm[5] one day ahead of the Trading Day.

The Day-Ahead Market includes several functions that are performed in sequence:

1) Market Power Mitigation (MPM) and Reliability Requirement Determination (RRD);
2) Integrated Forward Market (IFM); and

---

[5] Except for unforeseen events.

3) Residual Unit Commitment (RUC).

3.2.3.1.1 Market Power Mitigation and Reliability Requirement Determination

The Market Power Mitigation (MPM) performs a test to determine which DAM bids are subject to mitigation for local market power based on specific criteria. If the test fails, the MPM mitigates the affected bids for the relevant Trading Hours. The Reliability Requirement Determination (RRD) determines the minimal and most efficient use of Reliability Must Run (RMR) resources to address local reliability in meeting the CAISO demand forecast over the Trading Day.

3.2.3.1.2 Integrated Forward Market

The Integrated Forward Market (IFM) is a market for trading Energy and Ancillary Services for each hour of the next Trading Day. The IFM uses the mitigated bids after MPM and RRD to clear supply and demand bids, and procure Ancillary Services to meet the ISO Ancillary Services requirements at least bid cost over the next Trading Day.

3.2.3.1.3 Residual Unit Commitment

The Residual Unit Commitment (RUC) is a reliability function for committing resources and procuring RUC capacity not scheduled in the IFM (as Energy or Ancillary Service capacity) to meet the difference between the CAISO load forecast and the scheduled load in the IFM, in each hour of the next Trading Day. Long-Start Resources with a Start-Up Time longer than a day may also be committed by RUC for the subsequent days as needed.

3.2.3.2 Real-Time Market

The Real-Time Market (RTM) is a market for trading Energy and Ancillary Services in real time. The bid submission for a given Trading Hour in the RTM is allowed after the DAM result publication for the corresponding Trading Day and up to 75 minutes before the start of that Trading Hour.

The Real-Time Market includes several functions that are performed in parallel, but with different periodicity:

1) Market Power Mitigation (MPM) and Reliability Requirement Determination (RRD);
2) Hour Ahead Scheduling Process (HASP);
3) Short-Term Unit Commitment (STUC);
4) Real-Time Pre-Dispatch (RTPD); and
5) Real-Time Economic Dispatch (RTED).

3.2.3.2.1 Market Power Mitigation and Reliability Requirement Determination

The MPM and RRD functions of the RTM are analogous to the same functions of the DAM. They are both performed hourly, 7½ min after the close of the RTM for a Trading Hour, i.e., 67½ minutes before the start of that Trading Hour. The MPM performs a test to determine which RTM bids are subject to mitigation for local market power based on specific criteria. If the test fails, the MPM mitigates the affected bids for that Trading Hour. The resultant mitigated bids are then used by all other RTM applications. The RRD determines the minimal and most efficient use of RMR resources to address local reliability in meeting the CAISO demand forecast over that Trading Hour. The MPM and RRD functions are performed simultaneously.

3.2.3.2.2 Hour Ahead Scheduling Process

The Hour Ahead Scheduling Process (HASP) is a process for trading hourly Energy and Ancillary Services based on bids submitted up to 75 minutes ahead of a Trading Hour. The HASP is performed hourly and immediately after MPM and RRD. Hourly Energy schedules and Hourly Ancillary Services awards for hourly pre-dispatched resources in that Trading Hour are published no later than 45 minutes before the start of that Trading Hour.

3.2.3.2.3 Short-Term Unit Commitment

The Short-Term Unit Commitment (STUC) is a reliability function for committing Quick-Start Resources (with start-up time plus minimum up time less than 255 min) to meet the CAISO demand forecast in each 15-min interval of the next four-to-five hours. The STUC is performed hourly after the HASP.

3.2.3.2.4 Real-Time Pre-Dispatch

The Real-Time Pre-Dispatch (RTPD) is a market for committing resources and for selling Ancillary Services at 15-min intervals. The RTPD runs automatically every 15 min, at the middle of each quarter of each hour, i.e., at 7½ min, 22½ min, 37½ min, and 52½ min into each hour. The RTPD Time Horizon is composed of a variable number of 15-min intervals that span the current and next Trading Hours. The first 15-min interval starts 22½ min after the time that RTPD runs, e.g., when RTPD runs at 7½ min into an hour, its Time Horizon starts at 30 min into that hour. The AS awards for the first 15 min interval of the Time Horizon are binding; the rest are advisory. The bids used in the next Trading Hour are the mitigated bids from the last execution of the MPM and RRD; these bids were submitted 75 minutes before the start of the next Trading Hour. The bids used in the current Trading Hour are the mitigated bids from the previous execution of the MPM and RRD; these bids were submitted 75 minutes before the start of the current Trading Hour. The RTPD execution at 52½ min into a given hour coincides with the HASP execution at 67½ minutes before the start of the latest Trading Hour for which bid submission time is closed; this RTPD is performed simultaneously with the HASP.

3.2.3.2.5 Real-Time Economic Dispatch

The Real-Time Economic Dispatch (RTED) is a market for trading Imbalance Energy and dispatching Ancillary Services at regular intervals. There are three modes for the RTED:

1) The Real-Time Interval Dispatch (RTID) is the normal mode of RTED. It runs automatically every 5 min, at the middle of each 5-min interval of each hour, i.e., at 2½ min, 7½ min, 12½ min, etc. into each hour. The RTID Time Horizon is composed of a variable number of 5-min intervals that span the current and next Trading Hours. The first 5-min interval starts 7½ min after the time that RTID runs, e.g., when RTID runs at 2½ min into an hour, its Time Horizon starts at 10 min into that hour. The dispatch for the first 5 min interval of the Time Horizon is binding; the rest are advisory. The bids used in the 5-min intervals of next Trading Hour included in the RTID Time Horizon are the mitigated bids from the last execution of the MPM and RRD; these bids were submitted 75 minutes before the start of the next Trading Hour. The bids used in the 5-min intervals of current Trading Hour included in the RTID Time Horizon are the mitigated bids from the previous execution of the MPM and RRD; these bids were submitted 75 minutes before the start of the current Trading Hour.
2) The Real-Time Manual Dispatch (RTMD) is executed manually and it has a single 5-min interval.
3) The Real-Time Contingency Dispatch (RTCD) is also executed manually, but it has a single 10-min interval.

3.2.4 Full Network Model

The CAISO markets employ a Full Network Model (FNM) with an accurate representation of the CAISO Control Area and all Embedded Control Areas. External Control Areas are not modeled, except for inter-connections that are modeled as radial inter-ties and for transmission facilities for which Participating Transmission Owners (PTOs) have converted their scheduling rights. The FNM is composed of network nodes interconnected with network branches. Physical generating units and loads are modeled at the relevant network nodes. The Location of a Generating Resource may be different from the node(s) of the corresponding physical generating unit(s) in the FNM. In general, the Location of a Generating Resource coincides with the node where the relevant revenue quality meter is connected. This Location is referred to as the “Point Of Delivery (POD).” Although the schedule, dispatch, and LMP of a Generating Resource refers to the POD, the energy injection is modeled in the FNM for network analysis purposes at the corresponding physical generating unit(s) (at the the interconnection point), taking into account of any losses in the transmission network leading to the POD. Import and Export Resources are modeled as generating units at the network nodes at the external end of inter-ties with external Control Areas. Aggregate Generating and Load Resources are not modeled explicitly in the FNM, but their energy supply and demand is mapped to their associated physical resources in the FNM using the relevant Generation and Load Distribution Factors, respectively.

The use of the FNM in the DAM and the RTM incorporates transmission losses and allows modeling and enforcing all network constraints. This results in Locational Marginal Prices (LMPs) for Energy that reflect the marginal cost of energy, losses, and congestion. Although the marginal energy component of the LMP is the same for all network nodes, the marginal loss and the marginal congestion components may vary across the network due to network characteristics and power flow patterns.

3.2.5 Pricing and Settlement

This section describes the pricing and settlement of the commodities in the CAISO spot markets.

3.2.5.1 Energy

Locational Marginal Prices (LMPs) are used to settle Energy supply or demand. LMPs are calculated for Price Locations and Aggregate Price Locations. A Price Location corresponds to exactly one network node. An Aggregate Price Location is defined as a collection of several network nodes. A Price Location is defined for each individual resource, and an Aggregate Price Location is defined for each aggregate resource. Furthermore, Aggregate Price Locations are defined for Trading Hubs for Inter-SC Energy Trade and CRR settlements.

The LMP is composed of three components: marginal energy, marginal loss, and marginal congestion cost. A resource is paid for Energy supply or charged for Energy demand the LMP at its Price Location. Similarly, an aggregate resource is paid for Energy supply or charged for Energy demand the LMP at its Aggregate Price Location.

3.2.5.2 Ancillary Services

Ancillary Service Marginal Prices (ASMPs) are used to pay the Ancillary Service providers for providing the AS through market bids. An ASMP is calculated for each resource for providing each type of AS in each market time interval for each market.

ASMPs are calculated based on the Regional Ancillary Service Marginal Prices (RASMPs). Regional Ancillary Service Marginal Prices are calculated for each AS region for each type of AS in each market time interval for each market.

Since a resource that provides an Ancillary Service can meet the requirement of multiple nesting AS regions, the ASMP represents the sum of the RASMPs of all the AS regions that the resource belongs to[6].

The cost of procuring the AS by the CAISO on behalf of the demand is allocated to metered demand on a system wide basis.

3.2.5.3 RUC Capacity

RUC LMPs are used to pay the RUC Capacities awarded through RUC bids in the RUC. RUC LMPs are similar to the Energy LMPs except that RUC LMPs are based on the RUC bids instead of the Energy bids. Resources under Resource Adequacy requirements are not eligible for RUC Award payments.

The cost of procuring the RUC by the CAISO on behalf of the load is allocated to the load using a 2-tier user rate approach. The first tier user rate applies to incremental load between DAM schedule and metered load; the second tier user rate applies to metered load to allocate the remaining cost that has not been fully allocated through the first tier allocation.

3.3 Market Optimization

This Section describes the optimization engines employed in the various market applications of the Day-Ahead and Real-Time Markets.

3.3.1 Security Constrained Unit Commitment Framework

The Security Constrained Unit Commitment (SCUC) is the optimization engine used in MPM/RRD, IFM, RUC, HASP, STUC, and RTPD. In general, it performs the following tasks:

- It determines resource commitment status;
- It determines Energy Schedules and prices;
- It determines Ancillary Services Awards and prices;
- It determines RUC Awards and prices.
- It provides data to Settlements to settle Energy, A/S, RUC, CRR payments/charges, and other uplift payments/charges.

---

[6] Tong Wu, Mark Rotheleder, Ziad Alaywan, Alex Papalexopoulos"Pricing Energy and Ancillary Services in Integrated Market Systems by an Optimal Power Flow," presented at the IEEE PES General Meeting, Toronto, Canada, July 13-18, 2003.

3.3.1.1 SCUC Algorithm

The Day-Ahead market clearing problem includes next-day generation offers, demand bids, depending on the market design virtual bids and offers, and bilateral transactions schedules. The objective of the problem is to minimize costs subject to all system resource and transmission constraints. Similar formulation is used to solve the Real-Time market problem as well as the Reliability Unit Commitment problem. In all cases, the SCUC accepts data that define bids (e.g., generator constraints, generator costs, and costs for other resources) and the physical system (e.g., load forecast, reserve requirements, and the security constraints). In Real-Time, the limited responsiveness of units and additional physical data (e.g., state estimator solution) further constrain the unit commitment problem. The SCUC selects resources based on minimum cost as reflected on bid prices and on physical deliverability by the transmission system.

The SCUC problem has long been viewed as a very complex and intractable optimization problem. One of the most popular methods for solving the SCUC problem has been the so-called Lagrangian Relaxation (LR) method. The accuracy of LR method was considered sufficient for industrial applications before the emergence of wholesale competitive energy markets. However, the development of competitive energy markets, where multiple generation owners compete to provide their products, has increased the pressure for more accurate Unit Commitment solutions. The disadvantage of the LR method is that heuristics are usually involved in obtaining feasible solutions and that the accuracy of the solution cannot be achieved in a controlled fashion within the convergence criteria.

An alternative approach to LR based method that effectively addresses the new modeling requirements imposed by the ISO markets is based on Mixed Integer Programming (MIP) implementations[7].

In recent years, significant progress has been made in developing efficient algorithms for solving general MIP problems. Many commercial MIP packages exist today in the market place. Some have been successfully applied to SCUC problems. The use of the MIP approach (which has been the standard SCUC algorithm in the US ISOs) with its new advanced features allows ISOs to deal effectively with a number of market design elements including co-optimized formulations for energy and ancillary services, a large number of transmission and other security constraints, dynamic ramp rates, forbidden regions, combined cycle units, hydro-scheduling, etc. For these reasons the MIP method had been adopted as the central algorithm for solving SCUC problems.

---

[7] Xiahong, Qiaozhu Zhai, Alex Papalexopoulos"Optimization Based Methods for Unit Commitment: Lagrangian Relaxation versus General Mixed Integer Programming," presented at the IEEE PES General Meeting, Toronto, Canada, July 13-18, 2003.

3.3.1.2 Network Applications

The CAISO SCUC interates between the Unit Commitment (UC) software and the Network Applications (NA), resulting in the need to solve the UC problem multiple times to obtain optimal results consistent with the limitation in the transmission system.

For this purpose the IFM SCUC engine includes a Full Network Model (FNM) that comprises of a detailed model of the physical power system network along with an accurate model of commercial network arrangements. These arrangements reflect the commercial scheduling and operational practices to ensure that the resulting Locational Marginal Prices (LMPs) reflect both the physical system and the actual scheduling practices. The commercial content of the Market FNM includes the following:

1) Load modeling considerations, such as Load Aggregations and System Load Distribution Factors, Custom Load Aggregations and Custom Load Distribution Factors, Trading Hubs and Hub Distribution Factors
2) Resource modeling considerations, such as Combined Cycle Units, Gen-ties, embedded generation, Participating Load, Aggregated generating resources, and Generation Distribution Factors
3) Grouping and zone definitions, such as UDCs, Price Locations, A/S regions, and RUC zones, and
4) Other scheduling elements, such as power system equipment schedules.

The SCUC clears the market by co-optimizing energy and ancillary services while managing congestion and losses. The Market FNM essentially represents the transmission network for California and is comprised of the following network components:

1) ISO Control Area encompassing the networks of the three (3) major investor owned utilities, referred to as Participating Transmission Owners (PTOs);
2) Non-ISO Control Areas that are embedded within the CAISO Control Area;
3) External Control Areas and External Transmission Systems;
4) Networks of New Participating Transmission Owners (NPTOs), and
5) Utilities (currently called UDCs).

The Market FNM includes an accurate reactive power model to ensure that reactive power related constraints are respected. The use of reactive power in power systems is an effective way for improving both power transfer capability and voltage stability. An AC power flow with local controls is implemented in the Network Applications. The operational status or schedules of the manually operated reactive power/voltage control equipment are accounted for in the Market FNM. Although the FNM is an AC model, SCUC is not pricing reactive power.

In addition to its physical and commercial components, several other model-related inputs are required in the optimization and processing of the Market FNM in the IFM markets. These inputs are a) the Ancillary Services regions and requirements, b) Constraint definitions and management, c) Branch Groups/Interfaces and Nomograms, and d) Contingency definitions and management. Each of these inputs has been generally regarded as separate applications, residing outside the vendor's software system. The Market System vendor will provide facilities for viewing and editing these market optimization inputs by the CAISO market operators.

The power system transmission constraints in both the base case and contingency cases will be included in SCUC optimization. The transmission power flows of the transmission system branches may be constrained in both directions. The set of transmission constraints selected to be included in the optimization are consistent according to specific constraint definition criteria. Any constraint loaded in base or contingency cases above a certain user adjustable percentage of the transmission equipment loading will be included in the optimization (the current default threshold is set to 90%). It should be noted that certain transmission constraints (user selectable) are required to be monitored only (i.e., not enforced). The monitoring is against the defined limits adjusted by certain user-defined percentage of the limit (user definable for groups of transmission equipment). The enforcement or monitoring status of a constraint is changeable by the authorized user.

The functions that use the Market FNM, called Network Applications (NA) are coordinated with the market clearing functions and are configurable by the user. NA can run on-line and in study mode. Both on-line and study sequences consist of a configurable set of applications using input initialized from appropriate sources or save cases and/or operator. The sequences are configured into a certain number of execution modes to run various markets. Each of these modes consists of a set of NA programs arranged in such a way that the mode performs its overall required function. An example sequence is listed below:

1) Schedule, limit, outage and forecast values retrieval
2) Aggregation/disaggregation and mapping of commercial and physical entities
3) Network Model Assembly and Topology Processor
4) Bus Scheduler
5) Power Flow
6) Contingency analysis; and
7) Constraint activation/deactivation.

For analytical functions, e.g., Power Flow, a number of slack bus options are provided. Classic slack bus options are also supported (such as distributed load, distributed generation and single user selectable slack). The proposed slack options enable a consistent approach for analysis and optimal allocation of system imbalances. Additionally, slack options used across various markets are compatible and consistent. Solution options are adjustable by authorized users in market runs and study sessions. Solution options and achieved quality for the analytical and optimization applications within each market run are archived and traceable for auditing purposes.

Lastly, there are two other very important NA functions that are used to produce network sensitivity information required to manage losses and congestion. These are:

1) Power Transfer Distribution Factor (PTDF) Calculations Function, and
2) Loss Sensitivity Calculations Function.

The PTDF Calculations Function produces the PTDFs. PTDFs are the sensitivities of injections at any location in the network with respect to flow on any transmission element (in a reference direction). PTDFs are used in the congestion management and the calculation of the LMPs. They are calculated following each AC power flow run.

The Loss Sensitivity Calculations Function calculates the marginal loss factors. These loss sensitivity factors are the sensitivities of losses with respect to injection at any network node. Loss factors are calculated following each AC power flow run using the distributed load slack option. Loss factors are accurately calculated for both physical and commercial portions of the model. This function also calculates losses after each AC power flow run.

3.3.1.3 Objective Function

The SCUC engine determines optimally the commitment status and the dispatch schedules of generation resources including curtailable demands and interchanges. The SCUC optimization model includes both system operation scheduling and the market clearing process. The objective is to minimize the total production costs (or bid cost) subject to network as well as resource related constraints over the entire time horizon, e.g., the Trading Day in the IFM. The time interval of the optimization is one hour in the DAM and 5, 10, or 15 minutes in the RTM depending on the application.

In IFM the overall production (or bid) cost is determined by the total of the start-up and minimum load cost of CAISO-committed resources, the energy bids of all scheduled resources, and the ancillary service bids of resources selected to provide ancillary services. This objective leads to a least-cost multi-product co-optimization methodology that maximizes economic efficiency, relieves network congestion and considers physical constraints. The maximal economic efficiency of the market operation can be achieved through a least-cost resource commitment and scheduling with co-optimization of energy, and Ancillary Services.

Both generation and load market participants can submit three-part energy bids (the three parts are start up cost in $/start, minimum load cost in $/hr, and energy bid price curve above minimum load in $/MWh). All online units provide energy service. Some of them can be selected to provide Regulation Up/Down and Spinning Reserve services. Generators can provide Non-Spinning Reserves regardless of their commitment status in the DAM. Costs of self-scheduled energy and self-provided Ancillary Services are represented by penalties of eventual uneconomic adjustments performed during the optimization.

Energy and Ancillary Service bid costs present integrated bid price curves. The bid price curves are stepwise functions of procured services, therefore bid costs are piecewise linear functions of service quantities. The maximum number of segments of energy bid price curves is ten, while Ancillary Service and Reliability Capacity bids present single segment price curves. The minimum segment size is configurable with a default value of 0.01 MW in all cases.

The objective function for the RUC optimization model is extended (compared to the IFM) to include the RUC bids and RUC self-schedules. Therefore, the RUC Capacity cost is considered as part of optimization objective only in RUC. The RUC capacity bids are single segment price bids. For partial RA units, a two segment RUC capacity bid will be accepted, where the first segment of $0 will represent the RA capacity and the second segment with a bid-in non-zero $ value will represent the remaining portion of the unit's capacity.

The constraints in the optimization include power balance constraints, ancillary service capacity requirement constraints, network constraints under both base case condition and contingencies and resource operating constraints. These constraints will be discussed next.

3.3.1.4 Power Balance Constraint

The power balance constraint states that the generation in the system should balance out with the load plus the losses. Only one market wide power balance constraint is considered. Both bid-in generation and bid-in load participate in market power balancing including network energy losses. The energy loss model is derived from the full AC network solution which is updated during the SCUC-NA iteration process. The network energy losses are linearized using incremental loss factors around the base operating point with respect to bid-in generators and loads.

3.3.1.5 Ancillary Services Constraints

The Ancillary Services Requirement a the CAISO are set up on a global, system-wide basis, or on a more granular regional level. The Ancillary Services Setter (ASRS) function allows the CAISO Operator to specify the AS procurement requirement for each AS Region. It calculates and expresses the gross estimated hourly objectives (minimum and/or maximum bounds) for AS procurement, both for the overall system and for pre-specified AS Regions. AS Regions are expected to align with current existing market system definitions of congestion zones or Load Zones. For each hour, the ASRS would publish the following AS requirement information[8]:

1) AS Region
2) Resources eligible to meet the Region's AS requirements
3) Region's minimum requirements for Spin, Non-Spin, Regulation-Up, Regulation Down, and
4) Region's maximum requirement for Regulation Down.

Ancillary Service bids and Ancillary Service self-provision can be submitted for each Ancillary Service. Additionally, Ancillary Service cascading is supported by the optimization, i.e., a lower quality of Ancillary Service can be substituted by the higher quality of Ancillary Service. Specifically:

1) Regulation Up can be used as substitution for both Spinning and Non-Spinning Reserves; and
2) Spinning Reserve can be used as substitution for Non-Spinning Reserve.

All AS are procured based on a ramp time of 10 minutes. The cascading sequence is configurable, but it is common for all Ancillary Service regions and all time intervals. Both Ancillary Service bids and Ancillary Service self-provisions are considered to satisfy regional requirements for that Ancillary Service. This means for example that a load serving entity with load in the San Diego LAP can self provide some or all of its AS obligation from resources in NP15 if the self provided AS clears the IFM.

3.3.1.5.1 Regulation Up and Down Requirements

For each AS Region and each Trading Hour a minimal requirement for Regulation Up capacity and a minimal and maximal requirement for Regulation Down can be specified. Both regulation bids and regulation self-provisions can participate in meeting these requirements. Only online generating units can be awarded regulation service to contribute to the Regulation Up and Regulation Down requirements.

3.3.1.5.2 Spinning Reserve Requirements

Separate Spinning Reserve minimal requirements can be specified for each AS Region and for each Trading Hour. The Spinning Reserve requirements can be met by Spinning Reserve bids and Spinning Reserve self-provisions as well as Regulation Up bids. Only online generating units provide Spinning Reserve service. According to Ancillary Service

[8] Tong Wu, George Angelidis, Alex Papalexopoulos, "Regional Ancillary Services Procurement in Simultaneous Energy/Reserve Markets," presented at the IEEE PES PSCE Meeting, New York, New York, October 10-13, 2004.

cascading, Regulation Up can be used as Spinning Reserve after the Regulation Up requirement is met. The substitution of Regulation Up self-provisions for Spinning Reserve is not allowed.

3.3.1.5.3 Non-Spinning Reserve Requirements

Separate Non-Spinning Reserve minimal requirements can be specified for each AS Region for each Trading Hour. Bids for Regulation Up and Spinning Reserve can be counted as Non-Spinning Reserve, too. The Non-Spinning Reserve requirements can be met by Non-Spinning Reserve bids and Non-Spinning Reserve self-provisions as well as Regulation Up and Spinning Reserve bids. The cascading of Regulation Up and Spinning reserve self-provisions is not allowed.

3.3.1.5.4 Maximum Upward Capacity Constraint

The total amount of upward Ancillary Service capacity is limited for each AS Region. Specifically, the sum of Regulation Up, Spinning Reserve and Non-Spinning Reserve procured in each AS Region using bids or self-provisions cannot exceed a limit maximum capacity at any time interval.

3.3.1.6 Network Constraints

Network constraints due to energy schedules are considered in the optimization for both the base case and contingency cases. The network power flow model is based on a full AC power flow solution. The branch flow MVA limits are translated into MW limits, assuming that MVAR branch flows and voltage magnitudes do not change significantly from one iteration to the next. The transmission line flows are expressed as linearized functions of the nodal power injections around the base operating state using calculated Shift Factors (PTDFs).

The set of critical transmission lines is selected according to the percentage of line MW loading. The lines loaded above the specified threshold (default 90%) are included in the optimization. To avoid oscillations in the SCUC-NA iteration process, lines are added into and never deleted from the critical set. The maximal number of enforced network constraints can be specified by the authorized user. The network constraints are ordered according to their percentage of loading. There are several types of network constraints as described next.

3.3.1.6.1 Network Branch Power Flow Limits

The network branch power flow limits are modeled as MVA ratings. They represent thermal limits of the transmission equipments. Normal and emergency ratings are specified for each branch for operation in normal and emergency conditions for each Trading Hour. Branch ratings are also derated for each Trading Hour. All branches are monitored but the operator may specify an exception list. The default branch rating is included with the EMS network model data imported into SCUC. Derated ratings are retrieved from the Outage Management application or entered manually by the user. The branch power flows of critical transmission lines are limited in both directions, but the limit is the same in each direction.

3.3.1.6.2 Transmission Interface Limits

A transmission interface is a branch group or a path that consists of one or more branches. All inter-ties and WECC paths are defined as transmission interfaces. A branch can be a member of multiple branch groups. The default branch group definition is included with the EMS network model data maintained in the Market FNM. The ratings for branch groups are referred to as Operating Transmission Capacity (OTC), usually determined by power flow analysis, transient stability analysis, voltage stability analysis and contingency analysis, performed by operation engineers and sometimes involving multiple neighboring control areas. These ratings are specified in MWs, are directional, and can change seasonally. Branch groups only have normal ratings. These ratings are used both in the base case and in contingency analysis. Usually these ratings already take into consideration the effects of significant contingencies.

3.3.1.6.3 Inter-Tie Energy-AS Constraints

Energy and Ancillary Service bids compete for the use of inter-ties when their demands for transmission capacity are in the same direction. Ancillary Service imports compete with energy schedules on designated inter-ties in the import direction. Moreover, energy does not provide counter-flow for Ancillary Service when the demands for transmission capacity are in opposite directions, and Ancillary Service does not provide counter-flow for energy when the demands for transmission capacity are in opposite directions. Finally, no netting is allowed among Ancillary Services. Obviously, only one of the inter-tie constraints may be binding in either direction at any given time.

3.3.1.6.4 Nomograms

A nomogram is a set of piece-wise linear inequality constraints relating resource output and transmission interface flows. Only constraints that relate AC branch MW flows, MW generation, and MW load having the standard format of single branch or interface constraints are considered in the current version of the SCUC. Resource statuses cannot be part of the nomogram model. The nomogram constraints must be piecewise linear constraints defining a convex set. Nomograms can consist of a family of piecewise linear constraints. The constraint curve is selected prior to the optimization. The following are examples of typical nomogram variables:

1) AC Interface MW Flow vs. AC Interface MW Flow;
2) AC Interface MW Flow vs. Area MW Generation; and
3) Area MW Generation vs. Area MW Load.

The nomogram constraint presents a single piecewise linear curve relating two or more nomogram variables.

3.3.1.6.5 Contingency Constraints

Contingencies are simulated forced outages of network elements. The SCUC performs contingency analysis using the Market FNM, to recognize network constraints in the commitment and dispatch of resources. The IFM system provides a facility for definition and maintenance of contingencies.

A defined contingency may involve any modeled element: line sections, transformers, switches, circuit breakers, shunts, synchronous condensers, etc. Generator and load contingencies can also be defined (for monitoring purposes). Equipment outages can be defined either by the element itself or its associated disconnect device. Contingency definitions can include actions beyond simply "opening" an element; thus, contingency definitions will allow for several different possible actions/commands. For example, a single contingency may involve opening a transmission line, closing an alternate switch or line section (automatic load transfer, as in a "flip-flop" arrangement), and/or bypassing a series capacitor/reactor. The sequence of these events will be pre-defined in contingencies. While most contingencies will likely involve only one or two elements, no single contingency will include more than 50 elements. The Contingency Application would also categorize each contingency into one of several groups (a configurable number of categories, initially set to 8). These categories would allow any individual contingency to be applied in one or many market environments (DAM, HASP, RTM).

In general, the security constraints corresponding to contingencies (except monitor only contingencies) are enforced in a preventive control mode, i.e., the optimal schedule is determined such that no security violations will arise if any defined contingency occurs.

3.3.1.7 Inter-Temporal Constraints

The constraints in the SCUC optimization include power balance constraints, ancillary service capacity requirement constraints, network constraints under both base case condition and contingencies, and resource inter-temporal constraints. In this Section we present the inter-temporal constraints in more details.

3.3.1.7.1 Minimum Up Time

Typically, a generating unit cannot change its commitment status at every time interval. It must stay online or offline for some minimal time period without changing its commitment status. The Minimum Up Time (MUT) constraint, specified in minutes, is the minimum amount of time that a unit must stay on-line between start-up and shut-down due to physical operating constraints.

3.3.1.7.2 Minimum Down Time

The Minimum Down Time (MDT) constraint, specified in minutes, is the minimum amount of time that a unit must stay off-line after the start of shut-down, including shut-down and start-up time. SCUC can commit and decommit units based on economics and consistent with the units' MUT and MDT constraints.

3.3.1.7.3 Start-Up Time

The generating unit Start-Up Time (SUT) is usually dependent on its cooling time, i.e., the time a unit needs to get into the online state depends on how much time the unit already was in the offline state. In this case, the total down time consists of the cooling time and the Start-Up Time, which is dependent on the cooling time. The total down time is enforced to be no shorter than the MDT. The CAISO SCUC models three cooling statuses: hot, intermediate and cold. These statuses are presented by separate segments of the Start-Up Time Function. These segments are the same as segments of the Start-Up Cost Function. The Start-Up Time Function is a monotonically increasing staircase curve of start-up cost versus cooling time.

3.3.1.7.4 Maximum Number of Daily Start-Ups

The SCUC enforces a resource constraint which is related to the maximum number of daily start-ups. Specifically, during the scheduling time horizon the total number of daily startups is limited by a number due to physical operating constraints.

3.3.1.7.5 Operational/Regulating/Reserve Ramp Rate

The operational ramp rate of resources limits the energy schedule changes from one time period to the next in the SCUC. The operational ramp rate constraints for energy schedule changes from one time period to the next are determined by the operational ramp rate function, or the regulating ramp rate if the resource provides regulation, multiplied by a configurable time interval (currently 60 minutes for the DAM).

The operational ramp rate function is described by a staircase function of up to four segments (in addition to ramp rate segments inserted by SCUC for modeling forbidden regions). The operational ramp rate function is submitted with the energy schedule and bid data. The ramp-rate function allows the SCs to declare the ramp-rate at different operating levels. However, the submitted ramp rate function is fixed throughout the time horizon. In order to mitigate possible capacity withholding through submitting low ramp rates, the SCUC uses the same ramp-rates up as ramp-rates down. The ramp rate changes as soon as the MW output ramps into a different operating level, (i.e., the ramp rate does not necessarily remain constant throughout a given range).

Similarly, regulating ramp rate constraints for procurement of Regulation Up and Regulation Down are determined by the submitted regulating ramp rate multiplied by a configurable time interval (currently 10 minutes). Moreover, resources taking longer than a prescribed number of (currently 20 in the DAM) minutes to ramp up or down to the next hour's energy schedule will not be able to provide Regulation (including Regulation Up and Regulation Down) for the next hour. The regulating ramp rate (for both Regulation Up and Regulation Down) is described by a single number. The regulating ramp rate is also used to evaluate both Regulation Up and Regulation Down bids and self-provisions.

Also, the operating reserve ramp rate constraints for procurement of the Spinning and Non-Spinning reserves are determined by the submitted operating reserve ramp rate multiplied by a configurable time interval (currently 10 minutes). Moreover, resources taking longer than a prescribed number of (currently 20 in the DAM) minutes to ramp up to the next hour's energy schedule are not able to provide operating reserve (i.e., Spinning and Non-Spinning Reserves) for the next hour. The operating reserve ramp rate (for both Spinning and Non-Spinning Reserves) is described by a single number. The operating reserve ramp rate is used to evaluate Spinning and Non-Spinning Reserve bids and self-provisions.

Note that the total amount of upward Ancillary Services is limited by the resource ramping capability over a specified time period (default is 10 minutes).

3.3.1.7.6 Daily Energy Limits

The CAISO SCUC enforces energy limit constraints which apply to a prescribed list of resources that can generate limited amount of energy for a given period of time because of hydro condition, emission allowance, or other regulatory or contractual reasons. Energy-limited resources indicate an energy limit in their DAM bids that applies to their schedule and dispatch throughout the Trading Day. The units are responsible for meeting their energy limit requirements for longer time periods, such as weekly, monthly or seasonal, subject to any applicable Resource Adequacy requirements. AS are not constrained by energy limits.

3.3.1.8 Forbidden Region Constraints

The CAISO SCUC enforces forbidden region constraints. The forbidden operating region is specified as a pair of low and high operating levels between which a resource may not operate in a stable manner. The forbidden regions lie between the resource minimum and maximum operating limits and they do not overlap. The SCUC optimization engine does not schedule a unit within a forbidden region unless the unit is operating at a full ramp rate in clearing the forbidden region. The unit ramping through a forbidden region cannot provide ancillary services, unless it takes less than 20 minutes (configurable) to cross the forbidden region. Also in this case it cannot set the LMP at its location. There is a separate ramp rate segment for each forbidden region, and a separate price curve segment lined up with each forbidden region. A generating unit can have up to four forbidden regions and up to five operating regions.

There are certain rules that the SCUC engine complies with in dealing with forbidden regions. These rules are:

1) If unit can cross the forbidden region in less than one time interval then it is never scheduled to operate inside the forbidden region.
2) If a slow unit cannot cross forbidden region without stepping inside it, the unit is scheduled to operate with full Ramp Rate inside a forbidden region.
3) The reversal of crossing direction is not allowed while the unit's schedule is going through the forbidden region. No hold time is required. I.e., once a unit crosses a forbidden region, the unit is allowed to cross back the forbidden region without requiring the unit to remain outside the forbidden region for a certain period of time.
4) The unit cannot provide Ancillary Services within a forbidden region, i.e. if unit is scheduled within forbidden region then both downward and upward Ancillary Services are equal to zero.
5) If a unit clears the forbidden region in less than 20 minutes, then it is allowed to provide Ancillary Services.
6) If a unit is scheduled within operating region then unit capacity available for scheduling is dependent on the size of surrounding forbidden regions. If the unit is slow and it cannot cross surrounding regions in configurable 20 minutes of ramping time, then the maximum point of lower forbidden region and minimum point of upper forbidden region define the unit capacity available for Ancillary Services.
7) If a unit is fast and it can cross surrounding regions in configurable 20 minutes of ramping time, then these regions are ignored and the maximum point of the next lower forbidden region and minimum point of the next upper forbidden region define the unit capacity available for Ancillary Services.

3.3.1.9 Unit Commitment

3.3.1.9.1 Commitment Status

The commitment status for each unit is the on/off state in each time period. A unit is off when it is offline or in the process of starting up or shutting down. A unit is on when it is online and synchronized with the grid. An off-on transition signifies a startup and an on-off transition signifies a shutdown. The SCUC software categorizes the reasons for which each resource is committed.

In the CAISO SCUC IFM there are some simple rules that determine the commitment status of a unit. These are:

1) If for any interval the unit is off line due to an outage, the unit's mode will be set to 'Unavailable' ('U') for that interval.

2) If for any interval the unit is forced on by the operator, the unit's mode will be set to 'MUST RUN' ('R') for that interval.

3) If for any interval the unit has a self-schedule, the unit's mode will be set to 'MUST RUN' ('R') for that interval.

4) If for any interval the unit is forced off by the operator, the unit's mode will be set to 'Unavailable' ('U').

5) If a unit is manually classified as an RMR unit by the operator, its status is set to "MUST RUN" in both the IFM and RUC runs.

6) In all other cases the unit is considered to have a "CYCLING" status in the IFM and its commitment status in each time interval depends on economics and the self-commitment status of the unit.

More complicated rules apply in real time in order to determine the commitment status of the unit.

The SCUC commitment period includes the hours when the resource is "ON" due to self-commitment, manual commitment by the CAISO through operator intervention (including certain RMR commitment), and optimal commitment by the SCUC application based on bid information. The SCUC commitment period extends from a startup to a shutdown and for some units it may span several days.

A Self-commitment period is a portion of the SCUC commitment period of a unit that has a non-empty Preferred Energy Schedule indicating its decision to self-commit the resource. The self-commitment period may include time periods where the unit does not have preferred schedules if it is determined that to meet the preferred schedule the unit must be ON due to MUT, MDT and the maximum number of start-up constraints.

3.3.1.9.2 Boundary Conditions

Each run in the IFM for a trading day needs to respect certain boundary conditions that result from the outcome of the IFM clearing of the previous day. For example, a unit that was committed in the last hour of the previous day and has a Minimum Up Time of 6 hours, will have a status "ON" for the first five (5) hours of the following day run. Similarly, a unit that was turned off in the last hour of the previous day and has a Minimum Down Time of 6 hours, will have a status "OFF" for the first five (5) hours of the following day run. Lastly, since the start-up costs and start-up time are a function of the cooling time, another boundary condition is to keep track of the time the unit has been offline.

3.3.1.10 Penalty Prices

A resource may decide to self-commit by submitting at least one (non-empty) preferred self-schedule. The total preferred self-schedule of a unit is composed of several specific type of preferred self-schedules associated with specific scheduling priorities such as

RMR, etc. These self-schedules need to be respected by the SCUC engine to the maximum extent possible. The CAISO SCUC achieves this objective by assigning to each specific type of preferred self-schedule a penalty (i.e., an uneconomic) bid price according to its scheduling priority.

All self-schedules are protected from curtailment in the congestion management process, if there are other economic bids that can be used to relieve congestion. If all economic bids are exhausted, the self-schedules between the minimum load and the first energy level of the first energy bid point will be subject to uneconomic adjustments based on the assigned penalty prices that reflect various scheduling priorities, such as RMR, price takers, etc. Uneconomic (penalty) adjustments are single-sided. I.e., imports and exports may be reduced to zero; load may be reduced to zero; generation may be reduced to lower operating (or regulating) limit. Any schedules below the minimum load level are treated as fixed schedules and are not subject to uneconomic adjustments for congestion management.

Furthermore, the CAISO SCUC software provides the functionality to classify and prioritize constraints among themselves and the control scheduling priorities discussed earlier. A common system of priority levels is supported for both control and constraint priorities. The priority level for any control or constraint class is configurable. The following constraint classes are supported and in the following priority order from high to low:

1) Power Balance Constraint;
2) Transmission constraints (separate levels for branch flows, interface flows, and nomograms) under base case;
3) Transmission constraints (separate levels for branch flows, interface flows, and nomograms) under contingency;
4) Ancillary Services minimum regional requirements; and
5) Ancillary Services maximum regional requirements.

3.3.1.11 Pricing Run

When the CAISO SCUC optimization engine converges, it produces schedules and LMPs for every time interval of the time horizon. However, in the event that resources are optimally scheduled or dispatched in the penalty region due to “uneconomic adjustments” required for feasibility, marginal prices would reflect the penalty prices of marginal resources scheduled or dispatched in the penalty region. Similarly, if binding constraints are violated for feasibility, marginal prices would reflect the penalty prices for these violations. The solution to this problem requires another run, called the “pricing run,” to “filter” these penalty prices out of the dual solution(which produces the prices). Specifically, resources scheduled or dispatched in the penalty region outside their energy bid (or their schedule if there is no energy bid) are scheduled or dispatched optimally based on specified configurable priorities, but the appropriate bid cap is used for pricing purposes. For supply increase or demand decrease in the penalty region the Energy Bid Ceiling (bid cap) is used. For supply decrease or demand increase in the penalty region the Energy Bid Floor is used.

Also, resources that are not allowed to set the marginal price, as identified by a flag in the Database, are filtered out in the pricing run. Specifically, in the pricing run these schedules and dispatches are fixed and not re-optimized. Moreover, energy and ancillary services bids that exceed specified and configurable “soft bid caps” are also not allowed to set the price, thus they are filtered out in the pricing run. The Energy Bid Ceiling is used instead of any bid price greater than the Energy Bid Ceiling and the Energy Bid Floor is used instead of any bid price lower than the Energy Bid Floor. The bid caps are configurable and different for energy and ancillary services.

To maintain consistency between resource scheduling and commodity pricing, the optimal solution of the scheduling run is preserved in the pricing run to the extent possible. The resource commitment statuses from the scheduling run are locked in the pricing run, i.e. committed units would be “must run” and uncommitted units would be “must not run”. All other constraints shall be considered in the pricing run. Also, energy self-schedules that have not been adjusted in the scheduling run are considered constant in the pricing run. In the pricing run these schedules are fixed and not re-optimized.

Specifically, the optimal resource schedules are bounded in the pricing run around the optimal solution of the scheduling run if they were scheduled in the penalty region or at a bid segment that violates the soft bid cap (if soft bid caps exist). These artificial bounds are as narrow as possible, but large enough to allow a feasible region without creating degeneracy of the optimization model. The default bounds are set equal to $\pm$0.001 MW range. All other resources, except ”Constrained Output Generators,(COG)”, are bounded by their original operating limits in the pricing run.

Finally, a software option allows “Constrained Output Generators,” as identified by a flag in the Database, to set the price in the pricing run if part of their minimum load energy is required to meet demand. The COGs are allowed to be scheduled continuously between zero MW and their minimum operating limit in the pricing run so that they could set the price at their location if their optimal schedule turns out to be above zero MW. The ramping of COGs in the operating region between zero MW and their minimum operating limit is not limited for all time periods of the time horizon. The energy bid of COGs in the pricing run is derived by dividing their Minimum Load Cost (MLC) by their minimum output (Pmin) extended down to 0 MW for all time periods of the time horizon.

Because of the COG treatment, the consistency between the schedule and the pricing run cannot be guaranteed at all times. Nevertheless, the proposed solution strives to minimize the discrepancy between dispatch and pricing.

3.3.2 Security Constrained Economic Dispatch

The Security Constrained Economic Dispatch (SCED) of the CAISO is the optimization engine used in RTED. It determines Energy dispatch and prices. SCED executes regularly at a dispatch time before each dispatch interval. There is a fixed time delay between each dispatch time and the following dispatch interval. The time delay accounts for the SCED execution time (120 sec.), the dispatch approval time (75 sec.), and the communication time for dispatch instructions (a configurable parameter set initially to 105 sec). The first dispatch interval of an hour starts at the start of that hour and the last dispatch interval ends at the end of that hour. The duration of the dispatch interval and the time delay are configurable parameters in SCED in multiples of 5 minutes, both initially set to 5 minutes. However, the duration of the dispatch interval is not less than the time delay, to prevent multiple SCED runs before any real dispatch. The Dispatch Operating Target (DOT) is the optimal dispatch calculated by SCUC/SCED based on telemetry. The Dispatch Operating Point (DOP) is the expected trajectory of the dispatched resource as it responds to dispatch instructions taking into account its ramp rate capability. The Preferred Operating Point (POP) is the same as the DOT, but it refers typically to units on AGC.

Specifically, SCED performs the following functions:

1) Correct future DOP of resources for which an outage has been reported in the Outage function;
2) Calculate the Imbalance MW requirement for the next dispatch target, which is the end of the next dispatch interval;
3) Calculate the Dispatch Operating Target (DOT) for each participating resource as the optimal dispatch for the next dispatch target to procure the required Imbalance Energy at least cost subject to resource and network constraints;
4) Perform a pricing run to determine the Locational Market Prices (LMPs) for the next dispatch interval; LMP prices are calculated for each node and in an aggregate level at LAPs;
5) Procure ancillary services as needed to meet the reserve requirement at least cost subject to resource and network constraints, and calculate associated opportunity costs;
6) Calculate the Dispatch Operating Point (DOP, a.k.a. POP) for each participating resource as a function of time as the expected trajectory of the resource operating point subject to resource capabilities;
7) At the start of the current dispatch interval, perform a compliance test for each participating resource based on its current telemetry and DOP calculated from the previous interval; and

8) Calculate the Ancillary Services capability of participating resources at the start of the next dispatch interval based on resource capabilities, and Ancillary Services schedules.

3.4 Market Power Mitigation

3.4.1 Overview

The market power mitigation process accomplishes the following objectives:

1) Reliability Requirement Determination (RRD), and
2) Local Market Power Mitigation (LMPM).

The market power mitigation process includes the following steps:

Step 1) Operator pre-specifies certain RMR requirements.

Step 2) SCUC Pass 1: Reliability and Market Power Mitigation with Competitive Constraint (RMPM-CC)

Step 3) SCUC Pass 2: Reliability and Market Power Mitigation with All Constraint (RMPM-AC)

Step 4) Reliability Requirement Determination and Bid Mitigation

The RRD and LMPM process as described here applies to both the DAM and the RTM. There are a few minor differences in its application to DAM and RTM as follows:

- The RRD and LMPM process will occur in the Day Ahead Market immediately after the close of the market at 10:00 A.M., after all bids and schedules are submitted by the SCs and validated by the ISO. The process occurs in Real-Time Market immediately after the close of the market at 75 minutes before the trading hour.
- The time horizon of the optimization for RRD and LMPM in the DAM is 24 hours (23 and 25 respectively on Daylight Saving transition days). The time horizon of the optimization for RRD and LMPM in the RTM is 105 minutes (i.e., from T-45 to T+60.)
- The load forecast time resolution in DAM is hourly, but it is 15 minutes for RTM.
- Each market interval for RRD and LMPM in the DAM is one hour. Each market interval for RRD and LMPM in the RTM is 15 minutes.
- The energy bid mitigation in DAM is done on hourly basis. The energy bid mitigation in the RTM is first done on 15-minute basis; and then the four 15-minute mitigated bids for each resource are synthesized to produce the hourly mitigated bid.

The operator pre-specifies certain RMR requirements because certain RMR resources are needed for reasons that cannot be determined automatically by SCUC such as voltage support, and certain RMR resources are saved for later use due to reasons that cannot be modeled accurately by the SCUC such as usage limit.

The RMPM runs optimally commit and dispatch resources as if they were procuring energy and ancillary services to meet 100 percent of the ISO's demand forecast and AS requirements. Only competitive network constraints[9] are enforced in the RMPM-CC runs, and all network constraints modeled in the FNM are enforced in the RMPM-AC runs. Comparing the dispatch levels between the RMPM-CC and RMPM-AC runs will determine RMR requirements and identify the resources subject to local market power mitigation.

3.4.2 Intermediate States of Bids

3.4.2.1 Bids Used in SCUC Pass 1

In SCUC 1, the pre-specified RMR requirements determine the commitment statuses and energy self-schedules with RMR priorities for the relevant RMR units.

The *Clean Bids*, i.e., the bids submitted by the SCs and validated by the CAISO, are used in SCUC Pass 1. The energy self-schedules that are part of the Clean Bids are represented by inserting uneconomic bid segments into the energy bids. These uneconomic bid prices (usually large negative numbers for supplies and large positive numbers for demands) determine in what order the energy self-schedules must be reduced in the event that economic bids are exhausted for relieving network constraints.

3.4.2.2 Bids Used in SCUC Pass 2

In SCUC 2, the pre-specified RMR requirements determine the commitment statuses and energy self-schedules with RMR priorities for the relevant RMR units.

The Clean Bids are further modified to protect the "competitive dispatch," i.e., the solution obtained from SCUC Pass 1, by inserting uneconomic bids to represent the energy schedules that are results of SCUC Pass 1. Such uneconomic bids also preserve the priorities of the energy self-schedules. These uneconomic bids prevent the "competitive dispatch" from being reduced and replaced economically by incremental dispatch from RMR resources that are considered in Pass 2 based on the estimated variable cost under the RMR agreements.

---

9 By default, competitive paths are today's inter-zonal paths. However, the ISO has adopted a "Competitive Path Assessment" methodology to determine the list of competitive paths. The study is repeated annually or upon major changes in the generation or transmission infrastructure.

3.4.3 Reliability Requirement Determination and Bid Mitigation

The Reliability Requirements and the Local Market Power Mitigation are determined by comparing the resource schedules derived from the RMPM-CC run and the RMPM-AC run as follows:

- For RMR units, the schedule obtained in SCUC Pass 2 is the RMR requirement in RUC if this schedule is greater than its RMPM-CC schedule. Moreover, in this case, the entire portion of the unit's bid above the RMPM-CC schedule and below the maximum contract quantity is mitigated to the lower of its Contract price or market bid,[¡Error! Marcador no definido.] but not lower than the highest bid price of the bid segments that have cleared SCUC Pass1. The unit's market bid (before modification for SCUC Pass 2) at and below the RMPM-CC schedule is retained. If the unit's RMPM-AC schedule is not greater than its RMPM-CC schedule, the unit's original market bid is retained in its entirety.
- For non-RMR resources, the resource is subject to LMPM if its RMPM-AC schedule (as obtained by SCUC Pass 2) is greater than its RMPM-CC schedule. In such cases, the portion of the unit's bid curve that is above its RMPM-CC schedule and below the maximum bid quantity is mitigated to the Default Energy Bid (DEB) if the DEB is lower. The resource-specific DEB can be Negotiated, LMP-Based or Cost-Based. The Cost-Based DEB may include an adder for a Frequently Mitigated Resource that meets certain criteria. The mitigated energy bid curve is made monotonically increasing by elevating the bid prices (above the bid segment prices for the segments that have cleared Pass 1) that would otherwise violate monotonicity.

The CAISO SCUC approach is shown in Figure 3-1, where for simplicity only two intervals are shown. In this case, the mitigated bid curve for the first interval dispatch is shown as the green dashed line and the mitigated bid curve for the second interval dispatch is shown as the black dashed line. The proposed methodology for developing a single mitigated bid curve for the entire hour is to take the minimum bid price of the two mitigated bid curves (i.e., the lower of the green or black dashed line. The mitigation extends to the end of the bid curve.

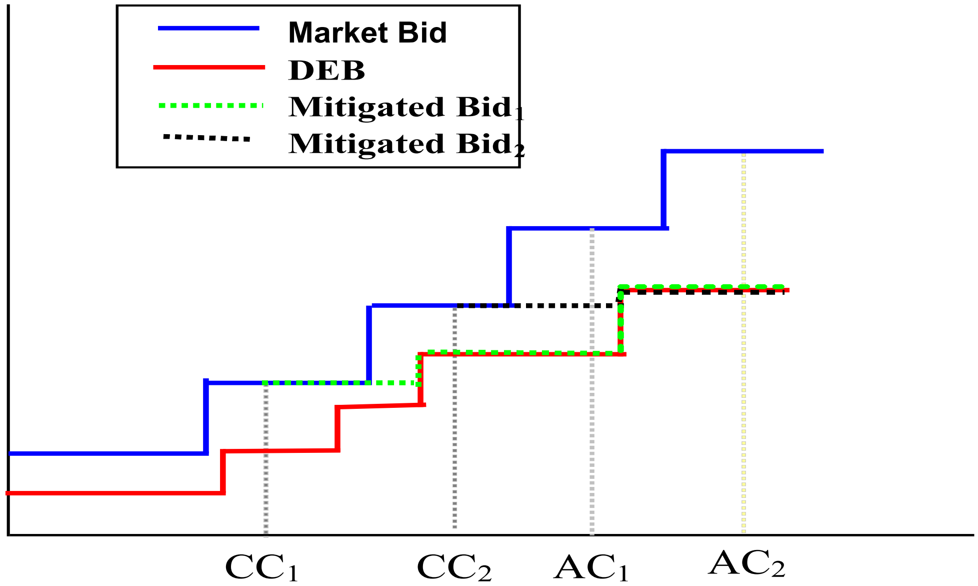


Figure 3-1: CAISO SCUC Mitigation Approach

Moreover, the pool of resources committed in the RRD and LMPM (Pre-IFM) constrain the resources that may be committed in DA IFM, but not RUC.

3.5 Integrated Forward Market

This section describes the Integrated Forward Market (IFM) that is conducted in the Day-Ahead Market (DAM) after the Market Power Mitigation (MPM) and Reliability Requirement Determination (RRD) process and before the Residual Unit Commitment (RUC) process.

3.5.1 Ancillary Services Markets

This section describes the Ancillary Services (AS) markets within the IFM.

3.5.1.1 Ancillary Services Regions

AS Regions are network partitions that are used to impose regional constraints in the procurement of AS. Without AS Regions, the procurement of AS would be system-wide without any locational considerations. AS regional constraints reflect transmission limitations between AS Regions that restrict the use of AS procured in one AS Region to cover for outages in another AS Region. AS regional constraints secure a minimum AS procurement in each AS Region so that AS awards are reasonably distributed across the system. AS regional constraints also limit the total AS procurement among Reg-Up, Spin, and Non-Spin in AS Regions with export transmission limitations.

There is always one AS Region defined for the entire CAISO Control Area: the “System Region.” AS are procured from certified Generating and Participating Load Resources within the System Region. AS are also procured from certified Inter-Tie Resources at Scheduling Points outside the CAISO Control Area. These Scheduling Points and the System Region define another AS Region: the “Super-Region,” which contains the System Region. The minimum AS regional constraints for the Super-Region reflect the AS requirements for each of the four AS: Reg-Down, Reg-Up, Spin, and Non-Spin. The minimum AS regional constraints for the System Region are only a percentage of the AS requirements, say 50%, to limit the AS procurement from Inter-Tie Resources for reliability purposes.

Besides the Super-Region and the System Region, other AS Regions (e.g., SP15) can be defined within the System Region to impose more granular regional limitations. Therefore, AS Regions are by definition hierarchical, i.e., larger regions include smaller ones. A Resource that is inside an AS Region is also inside all higher enclosing AS Regions.

3.5.1.2 Ancillary Services Requirements

The AS requirements are determined by the CAISO in accordance with the Minimum Operating Reliability Criteria (MORC):

- Operating Reserve requirements for 5% of forecasted load served by hydro-generation resources;
- Operating Reserve requirements for 7% of forecasted load served by non-hydro-generation resources;
- Operating Reserve requirements sufficient to cover the worst single contingency;
- At least half of the Operating Reserve is spinning;
- Locational considerations for a reasonable distribution of Operating Reserve across the system.

3.5.1.3 Ancillary Services Certification

Resources must be certified for AS provision since they must meet certain technical and operating requirements that depend on each AS. The AS certification and the associated maximum AS capacity are registered in the Master File after testing that demonstrates satisfactory delivery of each AS. Certified Resources are periodically subjected to unannounced testing by dispatching their AS awards in real time out of sequence to verify on-going AS provision capability.

3.5.1.4 Ancillary Services Self-Provision

Generating and Participating Load Resources certified for AS may self-provide these AS in the IFM. An AS self-provision is an AS capacity offer in a given Trading Hour without a price. Self-provided AS are evaluated for feasibility with respect to the relevant resource operating characteristics and regional constraints, and are then qualified (accepted) prior to AS bid evaluation. If a regional constraint imposes a limit on the total amount of Reg-Up, Spin, and Non-Spin, and the total self-provision of these AS in that region exceeds that limit, self-provided AS are qualified pro rata in three tiers: Reg-Up first, followed by Spin, and then by Non-Spin.

Qualified AS self-provision effectively reduces the AS requirements that need to be met by AS bids within the same AS Region. Qualified AS self-provision also reduces the AS obligation for each SC in the AS cost allocation.

3.5.1.5 Ancillary Services Bids

Resources certified for AS may bid these AS in the IFM. An AS bid is an AS capacity offer in a given Trading Hour at a single price. Resources may both self-provide and bid AS in a given Trading Hour as long as the total offered AS capacity from both self-provision and bid does not exceed the applicable certified maximum AS capacity. Resources must specify through a flag whether their Spin and Non-Spin awards are to be treated as contingency reserve in real time, i.e., be available for dispatch under contingency conditions only, or whether they can be dispatched optimally under all conditions. This specification is for all Trading Hours of the Trading Day and does not affect AS procurement, only AS dispatch in real time.

3.5.1.6 Ancillary Services Procurement

AS bids are evaluated simultaneously with Energy bids in the IFM to clear bid-in supply and demand, and to meet the AS requirements net of qualified AS self-provision. Thus, the IFM co-optimizes Energy and AS; the capacity of a resource with Energy and AS bids is optimally used for an Energy schedule, or it is reserved for AS in the form of AS awards. Furthermore, AS bids from Inter-Tie Resources compete with Energy bids for inter-tie transmission capacity.

AS are procured in the IFM to meet the AS requirements, net of qualified AS self-provision, subject to resource operating characteristics and regional constraints. The IFM employs a *cascaded* AS procurement where Reg-Up may be used to meet Spin and Non-Spin requirements, and Spin may be used to meet Non-Spin requirements, if this *AS substitution* results in a lower overall AS procurement cost. Note that this AS substitution is performed optimally among the AS bids, whereas no such substitution takes place in qualifying self-provided AS.

3.5.1.7 Ancillary Services Pricing

The IFM calculates a Regional Ancillary Service Marginal Price(RASMP) for each AS Region as the cost sensitivity of the relevant binding regional constraint at the optimal solution, i.e., the marginal reduction of Energy-AS procurement cost for a marginal relaxation of that constraint. If no regional constraint is binding for an AS Region, the corresponding RASMP is zero.

Since AS Regions are nested, the Ancillary Service Marginal Price (ASMP) for a given resource and AS is the sum of all RASMPs for that AS overall AS Regions that include this resource. The ASMP would adequately compensate resources for their AS awards (from selected bids), i.e., the ASMP for a given resource and AS would not be lower than the selected AS bid from that resource for that AS. Furthermore, the ASMP would also reflect any opportunity costs in reserving capacity for AS instead of scheduling that capacity as Energy.

3.5.1.8 Ancillary Services Settlement

There are three elements in AS settlements: AS Payments, No Pay, and AS cost allocation. There are briefly described as follows. AS awards from selected AS bids are paid the applicable ASMP. The ASMP for a given resource and AS is calculated as the sum of all RASMPs for that AS over all AS Regions that include this resource. No Pay calculates ex post AS capacity from AS awards that was not available because of any of the following reasons:

1) Undispatchable AS Capacity due to outages or derates affecting the resource regulating capability, its maximum capacity, or its operational ramp rate.
2) Unavailable AS Capacity due to uninstructed deviation.
3) Undelivered AS Capacity due to insufficient Energy in response to an AS dispatch instruction.
4) Declined AS Dispatch Instruction; this is the only No Pay AS Capacity applicable to Inter-Tie Resources, and only to Inter-Tie Resources.

No Pay AS Capacity is charged back to the relevant SC.

The net cost of AS procurement, i.e., all the AS payments for AS awards from selected AS bids in the DAM and RTM is recovered through a system-wide AS rate charged to SCs in proportion to their AS obligation, adjusted by Inter-SC AS Capacity Trades.

3.5.2 Energy Market

This section describes the Energy market within the IFM.

3.5.2.1 Generating Resource Bids

Generating Resources may participate in the IFM by submitting multi-part Energy bids for one or more Trading Hours of the Trading Day. The various parts of a multi-part Energy bid are:

Start-Up Cost

Up to three progressively increasing start-up costs are supported depending on the duration of the time that the resource remains offline before starting up: a warm start-up cost, an intermediate start-up cost, and a cold start-up cost. In each Trading Day, start-up costs are obtained from long-term (e.g., 6-mo) registered values, calculated based on proxy cost data, or bid in, according to Market Participant election. In either case, the start-up costs remain the same for the entire Trading Day in both the IFM and the RTM.

Minimum Load Cost

The minimum load cost is the hourly cost of operating at minimum load and it occurs while the resource remains online. In each Trading Day, minimum load costs are obtained from long-term (e.g., 6-mo) registered values, calculated based on proxy cost data, or bid in, according to Market Participant election and applicable CAISO policy. In either case, the minimum load costs remain the same for the entire Trading Day in both the IFM and the RTM.

Energy Bid

The Energy bid is the incremental cost of producing Energy above the minimum load and may be bid for each Trading Hour of the Trading Day. Up to ten incremental costs are supported for different operating ranges, thus the Energy bid is a staircase curve with up to ten steps (segments). The Energy bid for Generating Resources must be monotonically increasing, i.e., the incremental Energy cost of the next segment (if any) must be greater than the incremental cost of the previous segment.

Generating Resources with no operating range (equal minimum load and maximum capacity), commonly referred to as Constrained Output Generators (COG), may not have Energy bids; their optimal scheduling (commitment) is based solely on their start-up and minimum load costs.

Energy bids are subject to Local Market Power Mitigation (LMPM) prior to the IFM.

Pumping Cost

The pumping cost is the hourly cost of operating a hydro pump and it occurs while the pump remains online. In each Trading Day, pumping costs are bid in and remain the same for the entire Trading Day in both the IFM and the RTM. Pumping costs apply only to pumped-storage hydro resources and hydro pumps. Pumped-storage hydro resources are modeled as resources with three modes of operation: offline, generating mode, and pumping mode.

Pumping Load

The pumping load is the energy consumption of a pumped-storage hydro resource or a hydro pump while in a pumping mode of operation.

3.5.2.2 Participating Load Bids

Hydro pumps certified as Participating Load Resources are modeled as pumped-storage hydro resources without a generating mode of operation. Hence they may only bid pumping cost and pumping load.

3.5.2.3 Import/Export Resource Bids

Import and Export Resources may bid an Energy bid for each Trading Hour of the Trading Day. The Energy bid is a staircase curve with up to ten steps (segments). The Energy bid for Import Resources must be monotonically increasing, whereas the Energy bid for Export Resources must be monotonically decreasing. Import and Export Resources may also bid a minimum number of consecutive hours where their Energy bids may be scheduled flat throughout the Trading Day (this is referred to as a “block Energy bid”). In this way, whenever an Energy bid is accepted and an import or export schedule is produced, a flat schedule is guaranteed for at least the minimum number of consecutive hours. There may be more than one schedule block within the Trading Day. Although the schedule is flat across a given schedule block, the schedules of different schedule blocks may differ.

3.5.2.4 Non-Participating Load Bids

Non-Participating Load Resources may bid an Energy bid for each Trading Hour of the Trading Day. The Energy bid is a staircase curve with up to ten steps (segments) and it must be monotonically decreasing.

3.5.2.5 Aggregate Resource Energy Bids

The Energy bids of Aggregate Generating and Load Resources are not distributed in the Security Constrained Unit Commitment (SCUC). However, the resultant Energy schedules are distributed to the corresponding physical generating units or loads, using the relevant Generation Distribution Factors (GDFs) or Load Distribution Factors (LDFs). This is necessary in the Network Applications (NA) to perform power flow calculations and contingency analysis, and to calculate marginal loss factors and power transfer distribution factors. Consequently, the distribution of aggregate generation and load is locked at the relevant distribution factors and these resources are scheduled on an aggregate basis.

3.5.2.6 Self-Schedules

Resources may self-schedule for each Trading Hour of the Trading Day in addition to or without providing Energy bids. However, no self-schedules are allowed for Import or Export Resources that submit block Energy bids in the DAM. A self-schedule is modeled as an Energy bid with an extreme price (penalty price) that effectively provides scheduling priority over Energy bids. The penalty price is only for modeling purposes and it does not affect the Energy LMP, which is calculated by the Pricing Run. For settlement purposes, self-schedules are price-takers; i.e., self-scheduled supply is paid the relevant Energy LMP, whatever it is, and self-scheduled demand is charged the relevant Energy LMP, whatever it is.

There are several different types of self-schedules at different scheduling priorities.

A self-schedule indicates self-commitment, i.e., the IFM shall not decommit self-scheduled resources. The self-schedule, although at a higher scheduling priority than Energy bids, it may be reduced by the IFM if this is necessary to resolve network constraints. Self-schedules may also be adjusted by the IFM as necessary to resolve any resource operational or inter-temporal constraint violations.

Self-committed resources are not eligible for recovery of their start-up costs. Self-committed resources are also not eligible for recovery of their minimum load costs during the Trading Hours where they self-schedule.

3.5.2.7 Resource Capacity Constraints

The CAISO IFM optimally commits and schedules resources to balance supply and demand subject to resource and network constraints. Network constraints. Resource Capacity constraints limit the Energy schedule and the AS awards to the available bid in resource capacity. For example for online units not scheduled for Regulation the Energy schedule must be greater than or equal to the Lower Operating Limit and the sum of the Energy schedule and the Non-Spin and Spin awards must be less than or equal to the lower of the Upper Economic Limit or the Upper Operating Limit.

When resource capacity constraints are binding at the optimal solution, any opportunity costs between Energy schedules and AS awards are included in the Energy LMPs and in the ASMPs.

3.5.2.8 Energy Scheduling

Resources are committed and scheduled in the IFM for each Trading Hour of the Trading Day. Self-committed resources with self-schedules and/or AS self-provision are modeled as “must run” in the relevant Trading Hours. RMR resources pre-dispatched manually before the DAM are also modeled as “must run” in the relevant Trading Hours with a RMR self-schedule at the applicable RMR level. Resources with outages are modeled as “unavailable” in the relevant Trading Hours. Resources with multi-part Energy bids and/or AS bids, but without self-schedules or AS self-provision are modeled as “cyclable” in the relevant Trading Hours, which means that these resources are available for optimal commitment in these hours, subject to applicable inter-temporal constraints and initial conditions.

The difference between the Energy schedules across two consecutive Trading Hours is limited to the 60-min ramping capability of the resource. If the resource is scheduled for Regulation in either of these hours, the regulating ramp rate is used to determine the 60-min ramping capability, otherwise, the operational ramp rate is used. Energy scheduling across Trading Hours may impose limitations in AS awards across these hours, and vice versa. More specifically, Reg-Down awards in a Trading Hour would limit from below the Energy schedule for the next hour to the 20-min ramping capability of the resource using the regulating ramp rate. Similarly, Reg-Up awards in a Trading Hour would limit from above the Energy schedule for the next hour to the 20-min ramping capability of the resource using the regulating ramp rate.

Furthermore, Spin or Non-Spin awards in a Trading Hour would limit from above the Energy schedule for the next hour to the 20-min ramping capability of the resource using the operational ramp rate. AS awards in these limitations include both selected AS bids and qualified AS self-provision. AS bid selection is optimal, hence an economic evaluation would determine whether it is better to select AS bids and restrict Energy schedules to a 20-min ramp capability, or use the full 60-min ramp capability and not select AS bids.

3.5.2.9 Energy Pricing

The IFM employs a Full Network Model (FNM) and thus calculates Locational Marginal Prices (LMPs) for Energy at each network node. The LMP at a given node is the marginal cost of serving load at that node. This is a theoretical definition; there is no requirement for load to be connected to that network node. A LMP is calculated for all nodes, including the ones without load. The LMP at a given node is composed of the following three components:

1) Marginal Energy cost;
2) Marginal Loss cost; and
3) Marginal Congestion cost.

The Marginal Energy cost is the same for all nodes in the network; it is the sensitivity of the power balance constraint at the optimal solution. The Marginal Loss cost reflects the marginal cost of transmission losses in the network; it is the Marginal Energy cost multiplied by the marginal loss factor at that node. The Marginal Loss cost may be positive or negative depending on whether a power ejection at that node marginally increases or decreases losses, using a distributed load slack to balance it. The Marginal Congestion cost reflects the marginal cost of congestion in the network; it is a linear combination of the shadow prices of all binding constraints in the network, each multiplied by the corresponding power transfer distribution factor. The Marginal Congestion cost may be positive or negative depending on whether a power ejection at that node marginally increases or decreases congestion.

The IFM calculates LMPs and their components for all nodes, including aggregate nodes, and all resources, including aggregate resources. The LMP of a resource is the LMP of the corresponding Location, aggregate or not. LMPs for Aggregate Generating and Participating Load Resources are calculated as weighted averages of the LMPs at the relevant nodes, weighted by the resultant Energy schedules. Therefore, since the distribution of the Energy schedules for these resources is fixed, the normalized weights are equal to the relevant GDFs or LDFs. The same holds for Non-Participating Load Resources under Default Load Aggregation; the weights in their aggregate LMP are the default LDFs. By contrast, the distribution of the Energy schedules of Non-Participating Load Resources under Custom Load Aggregation is not fixed, hence the resultant Energy schedules are used as weights in their aggregate LMP instead of the relevant custom LDFs.

The IFM also calculates aggregate LMPs for Trading Hubs for the settlement of CRRs and Inter-SC Energy Trades, other than Physical Inter-SC Energy Trades. The calculation formula for Trading Hub LMPs is still subject to debate.

The IFM also calculates the shadow prices of all binding network constraints at the optimal solution. Of these shadow prices, only the shadow prices on inter-tie constraints are significant for settlements.

3.5.2.10 Energy Settlement

Generating and Import Resources are paid for their Energy schedule the LMP at their Location. Load and Export Resources are charged for their Energy schedule the LMP at their Location. The LMP at an aggregate Location for an aggregate resource is an aggregate LMP. The net revenue from these payments and charges is attributed to marginal loss surplus and congestion revenue and is credited to the CRR (i.e., FTR) Balancing Account.

Inter-SC Energy trades are paid (for trade in) or charged (for trade out) the relevant Trading Hub, LAP, or Generating Resource LMP.

Obligation CRRs from a source to a sink are paid the algebraic difference between the marginal congestion LMP components at the sink and the source. Option CRRs from a source to a sink are paid the positive difference between the marginal congestion LMP components at the sink and the source. These payments are debited to the CRR Balancing Account.

Finally, unrecovered start-up and minimum load costs for non-self-committed resources are conditionally recovered through the Bid Cost Recovery (BCR) mechanism.

3.6 Residual Unit Commitment

This Section describes the Residual Unit Commitment (RUC) that is conducted in the Day-Ahead Market (DAM) after the Market Power Mitigation (MPM) and Reliability Requirement Determination (RRD) process and after the Integrated Forward Market.

3.6.1 RUC Requirement

The CAISO RUC process occurs after the DA Integrated Forward Market (DA IFM). Since the DA IFM only commits sufficient resources to meet scheduled loads and exports, and it is likely that the scheduled loads are less than the load forecast, there is a need for a process to commit additional resources to meet load forecast. RUC is the process designed to ensure sufficient on-line resources to meet real-time load. The load forecast is provided by the CAISO load forecasting tools and adjustable by operators on RUC zone basis. A RUC zone is a collection of nodes. RUC zones may not necessarily coincide with Load zones.

3.6.2 RUC Bids

The CAISO RUC process is a process where resources submit RUC bids and are paid for their services. Resources that are eligible for participating in RUC are registered with the CAISO. The eligible resources can submit RUC bids to the RUC market. A RUC bid is a ($/MW, MW) pair. The meaning of a RUC bid differs depending on whether the resource that submits the RUC bid is under Resource Adequacy (RA) obligation.

If a resource is not under RA obligation, the RUC bid that the resource submits is interpreted as an incremental amount of capacity that the resource is willing to provide in the RUC market in addition to its DA energy schedule.

If a resource is under RA obligation, a certain amount of capacity of this resource is registered with the CAISO as RA capacity. The expectation is that the RA capacity must participate in both the DA IFM and the RUC process. Moreover, the RA capacity must participate in the RUC process with a $0/MW RUC bid price for the entire RA capacity. Therefore, the software automatically inserts a $0/MW RUC bid for the entire RA capacity; a RA resource only needs to submit a RUC bid for the non-RA capacity. In other words, a RUC bid submitted by a RA resource is interpreted as a RUC bid for the non-RA capacity in addition to the $0/MW RUC bid for the RA capacity.

In the CAISO energy market design, a resource cannot skip the IFM to participate in the RUC market directly; the total amount of RUC award (which considers both the RA capacity plus the submitted RUC bid quantity for an RA resource) is limited by the quantity of the energy bid minus the sum of DA energy schedule and the upward ancillary service awards. In other words, the sum of the DA energy schedule, the upward ancillary service awards including ancillary self-provisions, and the RUC award is limited by the quantity of the energy bid.

While IFM will honor multi-hour blocks inter-tie bid when procuring energy, post DA-IFM process (RUC, HASP and RTM) are not designed to enforce multi-hour block constraints. Therefore, RUC will evaluate all inter-tie RUC availability and HASP will evaluate inter-tie energy bids on an hourly basis instead of a multi-hour block basis.

3.6.3 RUC Constraints

In addition to the resource constraints and network constraints that are common in all SCUC runs as described in §3.1, RUC has three constraints that worth special explanation:

- Capacity constraints;
- Energy constraint; and the
- Quick-Start Resource constraint.

The capacity constraints ensure that sufficient RUC capacity is procured to meet load forecast. This is done by enforcing the power balance between the total supply (which includes DA IFM energy schedules and RUC capacity) and the total demand (which includes DA IFM export schedules and load forecast.) The load forecast can be adjusted to increase the RUC target if there is A/S bid insufficiency in DA IFM. The energy constraint ensures that RUC will not commit excessive amount of minimum load energy from internal resources. Specifically, the constraint ensures that the sum of the following terms will not be greater than a configurable percentage (e.g., 95%) of the system load forecast:

- DA IFM energy schedules of generators
- DA IFM energy schedules of net imports
- DA IFM energy schedules of load reductions of participating loads, and
- Minimum load energy committed by RUC.

The Short-Start Resource constraint ensures that RUC will not commit excessive amount of capacity from Short-Start Resources. Specifically, the constraint ensures that the total IFM energy schedules and RUC capacity from Short-Start Resources is not greater than a configurable percentage of the total available capacity of all Short-Start Resources. The list of Short-Start Resource will be managed by the CAISO taking into consideration of any exception required by the RA policy.

3.6.4 RUC Commitment and RUC Capacity Awards

The time horizon of the RUC optimization can be configured in several ways as follows:

- RUC optimizes over a 24-hour time horizon for the next day (default option);
- DA IFM optimizes for two consecutive 24-hour time horizons, and RUC optimizes over a 48-hour time horizon for the next day and the day after the next; and
- RUC optimizes over a maximum of seven consecutive 24-hour time horizons.

When the optimization time horizon is extended to more than 24 hours, the RUC bids for the first 24 hours are duplicated for the remaining hours of the time horizon.

RUC capacity awards are the incremental amount of capacity above the DA IFM energy schedules, which are needed to meet the load forecast in RUC optimization. There are a few exceptions:

- The portion of the capacity that covers the minimum load is not considered RUC capacity award, in other words, it is not eligible for RUC availability compensation since the minimum load energy is compensated through the commitment cost compensation (including minimum load cost).
- The capacity of a RMR unit used in the RUC optimization to meet load forecast is not considered RUC award since the capacity is already compensated through RMR contract. The RMR needs will be re-evaluated in the real-time.

The CAISO will only issue RUC instructions to resources that must be started in DA in order to be available to meet real-time load. In other words, the CAISO will re-evaluate the commitment decisions in HASP for resources that can be started in the HASP process. However, the RUC advisory availability as determined by the RUC application will be made available to the market participants even if a RUC start-up instruction is not issued in DA process.

3.6.5 RUC Pricing

RUC Locational Marginal Prices (RUC LMPs) are calculated by the RUC optimization based on the RUC bids. A resource that receives a RUC instruction will be compensated by the product of the RUC capacity award and the RUC LMP of its Location. The determination of the RUC LMP is similar to the determination of the energy LMP, except that RUC bids are used for the RUC LMP.

3.6.6 RUC Settlement

Although RUC providers are paid by the RUC LMP, the cost incurred to the CAISO for procuring the RUC capacity is allocated to load on a system wide basis. The RUC cost allocation uses a 2-tier settlement approach. The first-tier settlement allocates a portion of the RUC cost to the load deviations. The second-settlement allocates the remaining RUC cost to all loads.

3.7 Real Time Market

This Section describes the Real Time Market (RTM) applications that are executed in real time. Most RTM applications are periodic and run concurrently with rolling Time Horizons, but with different periodicity. The timeline characteristics and the objectives of the RTM applications differ; each one is designed for specific tasks as shown in Table 3-1.

Table 3-1: RTM Applications

Application	Periodicity	Interval	Time Horizon	Task
MPM	Hourly	15-min	105 min	Market power mitigation and reliability requirement determination for RTM bids submitted at *T*–75' for the Trading Hour from *T* to *T*+60'.
HASP	Hourly	15-min	105 min	Hourly pre-dispatch for hourly pre-dispatched resources for the Trading Hour from *T* to *T*+60'. Advisory 15-min dispatch for 5-min dispatchable resources for the Trading Hour from *T* to *T*+60'. Hourly AS awards for hourly pre-dispatched resources for the Trading Hour from *T* to *T*+60'. Advisory 15-min AS awards for 5-min dispatchable resources for the Trading Hour from *T* to *T*+60'.
STUC	Hourly	15-min	255 min	Unit commitment for the Time Horizon from *T* –15' to *T*+240'.

Application	Periodicity	Interval	Time Horizon	Task
RTPD	15-min	15-min	105-60 min	Unit commitment for the 105-60 min Time Horizon. 15-min AS awards for 5-min dispatchable resources for the first 15-min interval of the Time Horizon. Advisory 15-min AS awards for 5-min dispatchable resources for the remaining 15-min intervals of the Time Horizon. Advisory 15-min dispatch for 5-min dispatchable resources for each 15-min interval of the Time Horizon.
RTID	5-min	5-min	35 min	5-min dispatch for 5-min dispatchable resources for the first 5-min interval of the Time Horizon. Advisory 5-min dispatch for 5-min dispatchable resources for the remaining 5-min intervals of the Time Horizon.
RTMD	5-min	5-min	5-min	5-min dispatch for 5-min dispatchable resources.
RTCD	On demand	10-min	10-min	10-min dispatch for 5-min dispatchable resources.

The HASP functionality is embedded in the RTPD application; one of the four RTPD runs in an hour is special, performing the HASP function. Table 3-2 lists the output information from the RTM applications and whether it is sent to the ADS or the SC GUI.

The Automated Dispatching System (ADS) is the application developed by CAISO to communicate real-time dispatch instructions to Market Participants.

Table 3-2: RTM Output

Application	Output	ADS	SC GUI
MPM	Mitigated Energy Bids	No	Yes
HASP	Hourly pre-dispatch for hourly pre-dispatched resources	Yes	Yes
	Hourly LMP for hourly pre-dispatched resources	No	Yes
	Advisory 15-min dispatch for 5-min dispatchable resources	No	Yes
	Advisory 15-min LMP for 5-min dispatchable resources	No	Yes
	Hourly AS awards for hourly pre-dispatched resources	Yes	Yes
	Hourly ASMP for hourly pre-dispatched resources	No	Yes
	Advisory 15-min AS awards for 5-min dispatchable resources	No	Yes
	Advisory 15-min ASMP for 5-min dispatchable resources	No	Yes
STUC	Binding start-up and shut-down instructions	Yes	No
	Advisory start-up and shut-down instructions	No	No
RTPD	Binding start-up and shut-down instructions	Yes	No
	Advisory start-up and shut-down instructions	No	No
	Binding 15-min AS awards for 5-min dispatchable resources	Yes	Yes†
	Binding 15-min ASMP for 5-min dispatchable resources	No	Yes†
	Advisory 15-min AS awards for 5-min dispatchable resources	No	No
	Advisory 15-min ASMP for 5-min dispatchable resources	No	No
	Advisory 15-min dispatch for 5-min dispatchable resources	No	No
	Advisory 15-min LMP for 5-min dispatchable resources	No	No
RTID	Binding 5-min dispatch for 5-min dispatchable resources	Yes	Yes†

Application	Output	ADS	SC GUI
	Binding 5-min LMP for 5-min dispatchable resources	No	Yes†
	Advisory 5-min dispatch for 5-min dispatchable resources	No	No
	Advisory 5-min LMP for 5-min dispatchable resources	No	No
RTMD	Binding 5-min dispatch for 5-min dispatchable resources	Yes	Yes†
	Binding 5-min LMP for 5-min dispatchable resources	No	Yes†
RTCD	Binding 10-min dispatch for 5-min dispatchable resources	Yes	Yes†
	Binding 5-min LMP for 5-min dispatchable resources	No	Yes†

† ex post publication after the Trading Hour.

3.7.1 General Description

This Section highlights the differences between the IFM and the RTM, and it describes the concepts, functions, and general principles for real time dispatch that are common to all RTM applications.

3.7.1.1 Differences from the Integrated Forward Market

The CAISO RTM applications are multi-interval optimization functions minimizing the cost of dispatching imbalance energy and procuring additional Ancillary Services, when applicable, subject to resource and network constraints. The main differences between the IFM and the RTM are the following:

1) The IFM application uses hourly time intervals, whereas the RTM applications use sub-hourly time intervals within their Time Horizon.
2) The Time Horizon of the IFM application spans the next Trading Day, whereas the Time Horizon of the RTM applications is variable and spans the current and next few Trading Hours at most. The Time Horizon of the RTM applications is a rolling Time Horizon that rolls one time interval into the future as the application executes its periodic cycle. This means that results for time intervals past the first one are advisory since they will be recalculated the next time the application runs.
3) The IFM application uses Non-Participating Load bids to clear them against supply bids, whereas the RTM applications use actual load prediction and demand forecast. Non-Participating Load bids are not accepted in the RTM.

4) The RTM applications use the latest available information about resource availability and network status; in fact, the optimal dispatch is initialized at the State Estimator (SE) solution that is provided by the Energy Management System (EMS).
5) The IFM application commits resources optimally for the next Trading Day using three-part energy bids. Almost all resources can be considered for optimal commitment, except for resources with exceptionally long start-up and/or minimum up times because the full cost impact of commitment decisions for these resources cannot be evaluated within the IFM Time Horizon. Similarly, the RTM applications that have unit commitment capabilities can commit resources optimally within their Time Horizon, however, because that Time Horizon is short (a few hours at most), only fast-start and medium-start resources can be committed. Consequently, any long-start resources that are not scheduled in the IFM or RUC, are effectively not participating in the RTM.
6) Unlike the IFM application, the RTM applications need to interface with the Automated Dispatch System (ADS) to communicate financially binding commitment and dispatch instructions, and with the Inter-Tie Transaction System (ITS) for confirmation of Inter-Tie Resource schedules and dispatch.
7) The RTM applications provide more control to the operator with the capability to bias the imbalance energy requirements, block commitment or dispatch instructions, or issue out-of-sequence instructions. This operator input is necessary to address any unexpected system conditions that may occur in real time.
8) The RTM applications also provide the functionality to the operator to switch the system or individual resources into a contingency state under which contingent Operating Reserves are dispatched optimally to address system contingencies. Contingent Operating Reserves are otherwise reserved and not dispatched by the RTM applications.

3.7.1.2 Real Time Dispatch Principles

Generally, the objective of a real-time market is system balancing and load following above and beyond the normal function of the Automatic Generation Control (AGC). AGC is mainly a control rather than an energy service. As AGC units depart from their Preferred Operating Point (POP) responding to frequency and net inter-change deviations, they temporarily supply or consume balancing energy. This energy is then purchased from or sold to resources that participate in the RTM at regular intervals. As the selected resources supply or absorb the procured energy, the AGC units return to their POP and their control margin is restored.

Irrespective of the particular dispatch methodology that is employed in a real-time market, the schedule deviations can be classified into *instructed* and *uninstructed*. Instructed deviations are the result of participating resources responding to dispatch instructions. Uninstructed deviations are the result of load forecast errors, forced outages and contingencies, strategic behavior, modeling limitations, failure to follow dispatch instructions, etc. Instructed deviations are usually price setters, whereas uninstructed deviations are price takers and may be subjected to penalties. Uninstructed deviations prompt the response of AGC to balance the system creating imbalance energy requirements that are met through instructed deviations calculated optimally by the RTM. Figure 3-2 illustrates the feedback loop between uninstructed and instructed deviations in the operation of the CAISO RTM and its interction with the AGC.

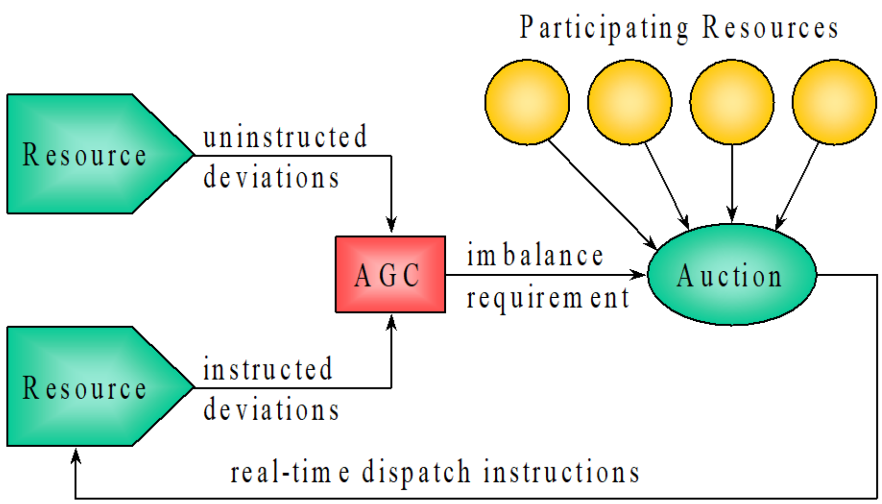


Figure 3-2: Real-Time Energy Market & AGC Interactions

The redispatch of the selected resources should result in a feasible outcome, i.e., no resource, network, or security constraints should be violated. Furthermore, if there are any such violations due to system condition changes, resources should be redispatched to remove these violations even if there is no system imbalance.

3.7.1.3 Real Time Dispatch Initialization

The CAISO RTD executes regularly before the start of each dispatch interval to allow a time delay for execution time, operator review, and communication of dispatch instructions through ADS to SCs and eventually to participating resources.

The RTM applications employ a Full Network Model (FNM) and enforce all network constraints calculating a Locational Marginal Price (LMP) at each network node. The CAISO dispatch approach is to initialize the optimal dispatch of each participating resource at the POT, obtained from the current SE solution, or optionally the telemetry (when available and reliable), considering the applicable ramp rate and capacity limits, as shown in Figure 3-4. When the SE solution or the telemetry is bad or unavailable, the last DOT can be used for initialization[10].

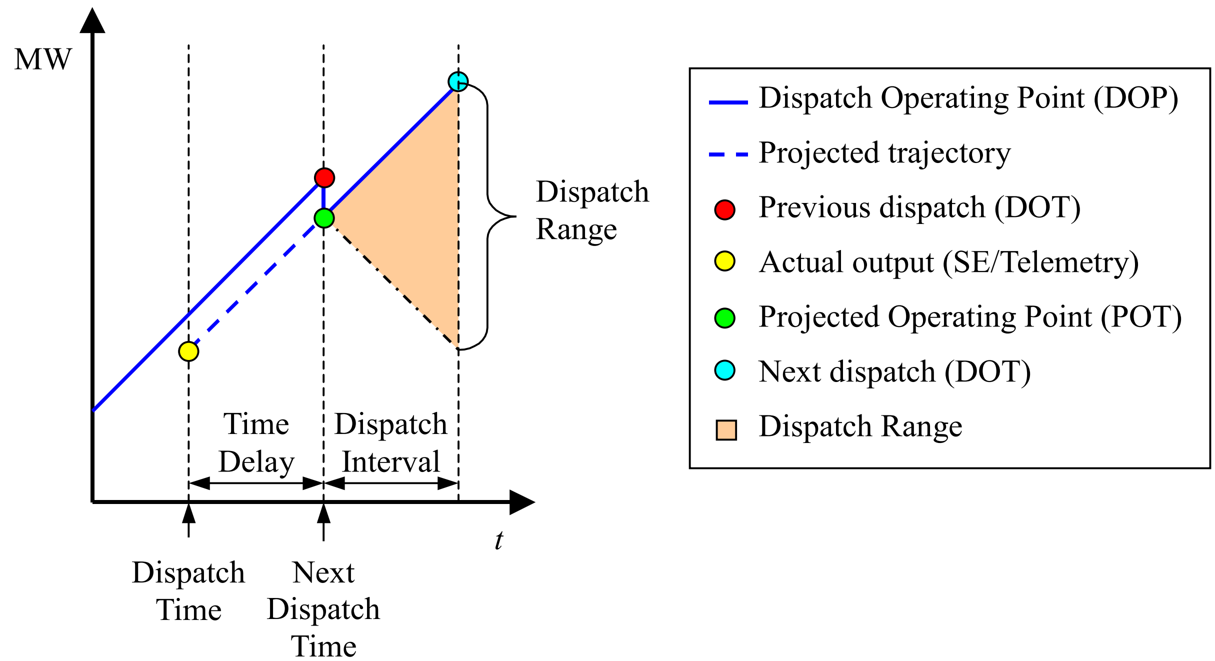


Figure 3-3: CAISO RTM Dispatch Initialization

The CAISO dispatch approach ensures that uninstructed deviations are taken into account in the calculation of the optimal dispatch so that dispatch instructions are always feasible at the resource level. For the same reason, no modification is needed in the imbalance energy requirements to account for any uninstructed deviations, leading to a simpler and more compact design.

It is important to note that the CAISO design deploys symmetrical cross-interval ramping, which means that the DOT is calculated for the middle of each dispatch interval.

3.7.1.4 Real Time Dispatch Symmetrical Ramping

Cross-Interval Ramping

For some ISOs the Dispatch Operating Target (DOT) is at the end of the dispatch interval and the Real-Time Dispatch (RTD) application calculates the optimal dispatch to meet the forecasted demand at that point in time. RTD dispatches resources at full ramp (according to their bid) starting at the beginning of the dispatch interval until they reach their DOT,

[10] George Angelidis, Alex Papalexopoulos "On the Operation and Pricing of Real-Time Competitive Electricity Markets," presented at the 2002 IEEE/PES Winter Power Meeting, New York, New York, January 29, 2002

which could happen at the end of the interval. This ramping rule leads to asymmetrical ramping for resources that are either marginal or capacity-limited. Assuming a smooth load variation and all other factors aside, this asymmetry results in a temporary Area Control Error (ACE) due to imperfect load following within the dispatch interval. The ACE **i**s biased resulting in non-zero regulating energy over the hour.

The CAISO RTM applications calculate the optimal dispatch to meet the average forecasted demand in each dispatch interval, effectively placing the DOT at the center of the dispatch interval. With this calculation, dispatched resources are expected to follow dispatch instructions ramping symmetrically across each dispatch interval boundary, as shown in Figure 3-4. A standard ramp (shown in blue in Figure 3-4) between the centers of consecutive dispatch intervals provides the best possible load following and simplicity since it applies uniformly to all dispatched resources. Faster symmetrical ramps (shown in red in Figure 3-4) contribute to ACE oscillations due to imperfect load following across dispatch interval boundaries. Nevertheless, the ACE is symmetric across interval boundaries resulting to lower ACE magnitudes and zero net regulating energy across intervals. The standard ramp was adopted because of its simplicity, transparency, and uniformity, the lower reliance on regulation, and the beneficial impact in meeting Control Performance Standards (CPS).

Variable Ramp Rate Impact

Variable ramp rates introduce some complexity in these otherwise simple cross-interval and cross-hourly ramping rules. If the ramp rate changes during the ramp, the DOP can no longer be symmetric across the dispatch interval boundary. In that case, the ramp should be the

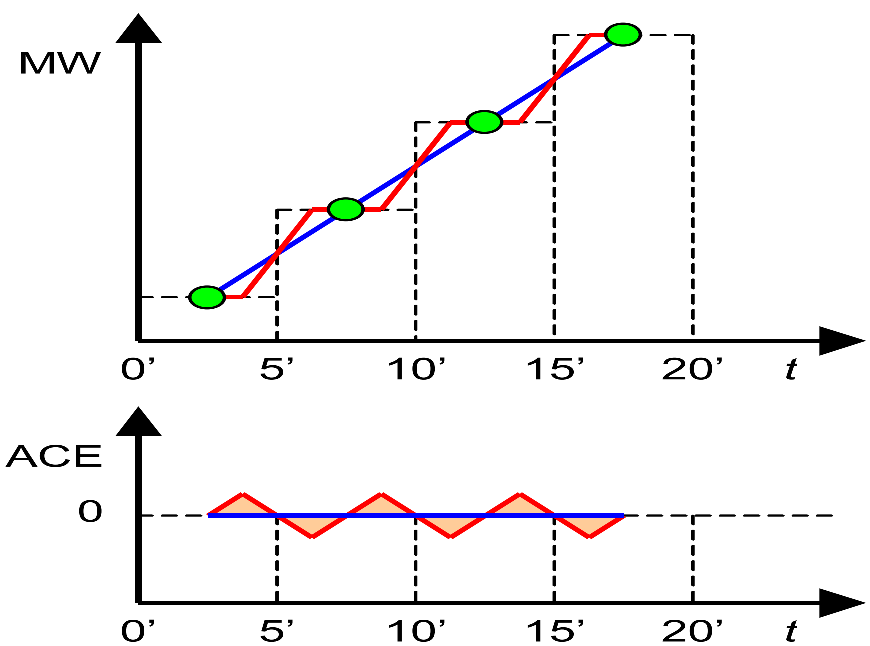


Figure 3-4: CAISO Dispatch Symmetric Ramping

smoothest possible ramp across the 5 min ramping window that results in symmetric (canceling) ramping energy, if possible. Similarly, if the ramp rate changes during the ramp, the DOP can no longer be symmetric across the hourly boundary. In that case, the ramp should again be the smoothest possible ramp within the 20-60 min ramping window that results in symmetric (canceling) ramping energy, if possible.

3.7.1.5 Imbalance Energy Requirement and Bias

The CAISO RTM applications provide an interface to the operator to bias the imbalance energy requirements within the relevant Time Horizon to reflect information that is known to the operator, but not to the RTM. For example, a pending outage that is not yet reflected in the Outage Scheduler, a pending curtailment of a large load, or an expected large uninstructed deviation in general. The imbalance energy bias is used by all RTM applications while in effect.

The imbalance energy requirement is the right hand side of the power balance constraint in the optimal dispatch formulation. The imbalance energy requirement for each time interval in the RTM is calculated as the load prediction obtained from the Very Short Term Load Predictor (VSTLP) for that interval, reduced by fixed supply and the marginal loss contribution of participating resources at the base operating point using loss sensitivity factors. The marginal loss contribution of participating (dispatchable) resources must be subtracted from the right hand side because the load projection includes the system losses at the base operating point. The base operating point is a power flow solution obtained at 15-min intervals from the SCUC solution using a distributed load slack and is used to linearize the non-linear power flow equations. The loss sensitivity factors are calculated from sensitivity analysis at that power flow solution, and thus represent a linear loss approximation due to both participating resource and non-participating load deviations.

There is no need to reflect uninstructed deviations in the imbalance energy requirement because the dispatch is initialized at the POT, which is extrapolated from the State Estimator solution, and thus any uninstructed deviations are already accounted for.

3.7.2 Hour Ahead Scheduling Process

This Section describes the Hour Ahead Scheduling Process (HASP) of the RTM.

3.7.2.1 HASP Timeline

The CAISO HASP timeline is illustrated in Figure 3-5. The HASP uses the same real-time energy and ancillary service bids that are submitted by *T*-75', where *T* is the beginning of the Trading Hour. The HASP is performed immediately after the MPM and RRD process that is executed at *T*-67.5' once every hour. The HASP execution is completed by *T*-45'.

The Hourly Energy schedules and Hourly Ancillary Services awards for hourly pre-dispatched resources in that Trading Hour are published no later than *T*-45'. Since control area checkout is usually completed by *T*-30', the pre-schedulers will have 15 minutes to complete the control area checkout after the inter-tie schedules are known from the HASP. The HASP Time Horizon is one hour and 45 minutes starting from *T*-45' and ending at *T*+60'. The bids that are submitted at *T*-75' for the Trading Hour between *T* and *T*+60' are mitigated by the MPM and RRD process that is executed at *T*-67.5'. The bids that are used for the portion of the time horizon between *T*-45' and *T* were submitted at *T*-135' and mitigated by the MPM and RRD process that was executed at *T*-127.5'.

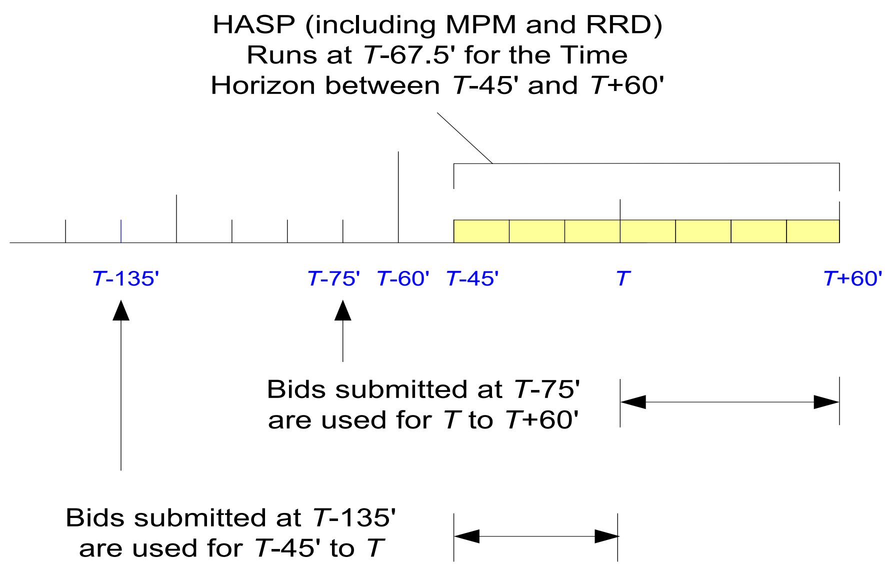


Figure 3-5: Illustration of the CAISO HASP Timeline

3.7.2.2 HASP Ancillary Services

The CAISO HASP determines the ancillary service awards, through both self-provision qualification process and market optimization, for each of the 15-minute intervals in the time horizon. Inter-tie resources cannot self-provide ancillary services because ancillary services must compete for the transmission capacity on the inter-tie. Inter-tie resources can submit ancillary service bids in order to provide ancillary services. However, the types of ancillary services that an inter-tie resource can bid depend on the energy type of the inter-tie resource.

3.7.2.3 HASP Pre-Dispatch

In the CAISO RTM, non-participating loads are not optimization variables; therefore, non-participating loads cannot submit energy bids. The only kind of participating load that the software supports is pumps that are modeled as pump-storage generating units.

Generating Resources and Import Resources can provide energy in the RTM by submitting self-schedules and energy bids. Export Resources may also submit energy bids in the RTM. The CAISO HASP determines energy schedules for hourly pre-dispatched resources for the Trading Hour (i.e., between *T* and *T*+60') on an hourly basis instead of on a 15-minute basis. This is accomplished in the software by enforcing constraints that ensure that the energy schedules for the four 15-minute intervals are equal. The LMP used to settle the hourly energy schedule is computed as the simple average of the four LMPs of the four 15-minute intervals of the Trading Hour.

3.7.3 Short-Term Unit Commitment

This Section describes the Short-Term Unit Commitment (STUC) application of the RTM.

3.7.3.1 STUC Timeline

The STUC timeline is illustrated in Figure 3-6. The STUC uses the same real-time energy and ancillary service bids that are used by HASP after MPM and RRD. The STUC is performed at *T*-52.5' once every hour. The STUC Time Horizon is four hours and 15 minutes starting from *T*-15' and ending at *T*+240'. The STUC determines whether some resources need to be started early enough to meet the demand within the Time Horizon. The bids that are used for the portion of the Time Horizon between *T*-15' and *T* were submitted at *T*-135' and mitigated by the MPM and RRD process that was executed at *T*-127.5'. The bids that are used for the portion of the Time Horizon between *T* and *T*+240' were submitted at *T*-75' and mitigated by the MPM and RRD process that was executed at *T*-67.5'.

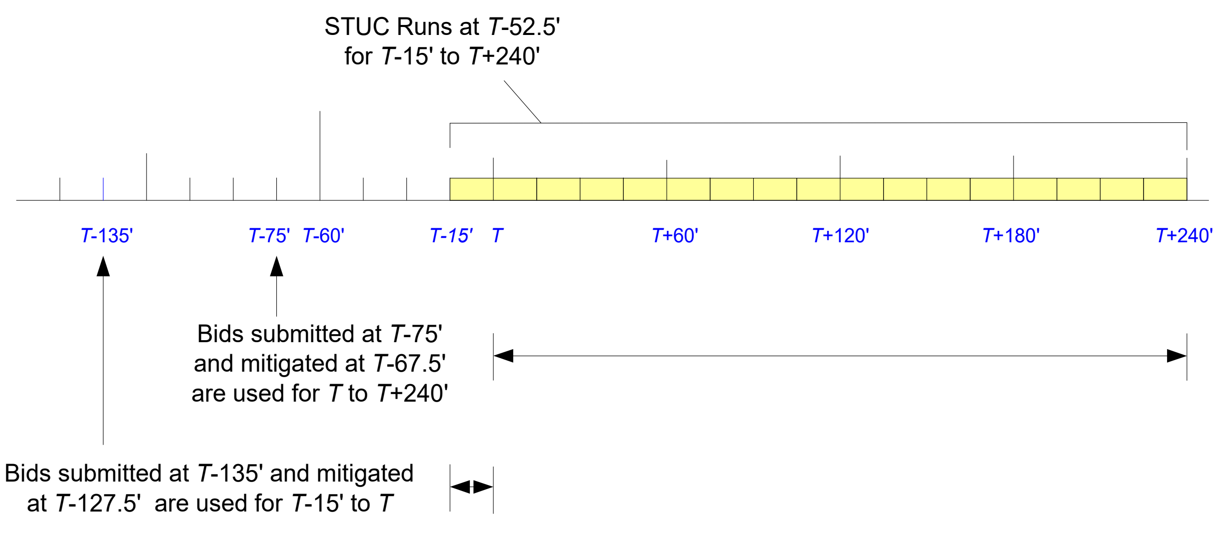


Figure 3-6: Illustration of the STUC Timeline

3.7.3.2 STUC Unit Commitment

The STUC produces a unit commitment solution for every 15-minute interval within the Time Horizon. The STUC recognizes initial conditions and respects binding commitment instructions (start-ups or shut-downs) that have been instructed by RTPD and STUC previously. Therefore, some of these binary decision variables are pre-determined or constrained by previously issued binding commitment instructions and initial conditions. For the remaining commitment decisions that are reflected in the commitment solution, STUC issues binding commitment instructions only if the commitment decision cannot be postponed for reevaluation by the next RTPD or STUC execution, i.e., when a resource must be notified immediately, considering its start-up time, so that it can be online in the needed interval. If the commitment of a resource can be reevaluated by the RTPD or STUC in subsequent runs and there is sufficient time to start-up the resource then, the commitment decision is advisory and not sent to ADS.

3.7.4 Real-Time Pre-Dispatch

This Section describes the Real-Time Pre-Dispatch (RTPD) application of the RTM. The main purpose of the RTPD is to commit fast- and medium-start resources and adjust day-ahead schedules on short-term basis due to expected operating conditions in the near future. The other purpose of the RTPD is to procure Ancillary Services to meet any additional AS requirements in real time due to changes in the demand forecast, outages, and non-contingent operating reserve dispatch during the hour. The RTPD uses a Security Constrained Unit Commitment (SCUC) co-optimization engine to minimize costs as reflected in the market bids subject to capacity and inter-temporal resource constraints and using a Full Network Model (FNM).

The RTPD application performs optimal commitment of resources participating in the real-time market to meet load and ancillary services requirements over a variable time period from one to two hours with 15-minute granularity. The resulting RTPD-based commitment instructions are the final decisions regarding resource commitments to adjust day-ahead and hour-ahead market schedules. The HASP functionality is embedded in the RTPD application; one of the four RTPD runs in an hour is special, performing the HASP function.

3.7.4.1 RTPD Timeline

The RTPD runs every 15 minutes at the midpoint of each 15-min interval in each hour. The RTPD execution at *T*–67.5' is special in that it also performs the HASP function. In this run, the RTPD application performs a full optimization of the time period beginning at *T*–45' and ending at *T*+60. Thus, the maximum number of 15-minute intervals in this run is 7. The bids that are used for the first hour (the operating hour, *T*–60' to *T*) in the RTPD

Time Horizon had been submitted at *T*–135' and mitigated in the MPM that was executed at *T*–127.5'.

In subsequent RTPD runs, the RTPD Time Horizon shrinks one 15-minute interval each time until it only covers the four 15-minute intervals of the Trading Hour from *T* to *T*+60', when RTPD runs at *T*–37.5'. Therefore, the RTPD Time Horizon varies between four and 7 15-minute intervals.

Also, binding Commitment Instructions (CIs) are issued for the two hours (the current hour and the pre-dispatch hour). These are resource start-up and shut-down instructions and are optimally determined by RTPD. All binding start-up and shut-down instructions from the RTPD runs are respected in all subsequent RTM applications.

3.7.4.2 RTPD Ancillary Services

Ancillary Services can be procured in the Real-Time market. The RTPD function processes non-zero real-time AS bids and real-time AS self-provisions. Furthermore, it co-optimizes real-time AS procurement with energy and will calculate AS Marginal Prices (ASMPs). The AS requirements in real time default to the same requirements used in the DAM, but they may also reflect additional requirements due to changes in the demand forecast. The Day-Ahead AS awards from the DAM, the hourly AS awards from the HASP, and the qualified Real-Time AS self-provisions are treated as fixed in the RTM, but incorporating any changes due to outages as they may occur.

The RTPD determines ancillary service awards to 5-minute dispatchable resources for every 15-minute interval in the RTPD Time Horizon on a 15-minute basis. However, only the ancillary service awards for the first 15-minute interval of the Time Horizon are financially binding. The ancillary service awards for 5-minute dispatched resources during the subsequent 15-minute intervals are determined by the subsequent RTPD runs. The 15-minute AS award is compensated by the ASMP of the service for the interval.

3.7.4.3 RTPD Unit Commitment

The RTPD produces a unit commitment solution for every 15-minute interval within the Time Horizon. Both RTPD and STUC can issue bidding instructions. The RTPD function can issue commitment instructions every 15 minutes, whereas the STUC can issue commitment instructions hourly. These commitment instructions are for any 15-minute interval in the RTPD or STUC Time Horizon.

The RTPD function runs at the top of the hour and optimizes the time period from *T*–45' to *T*+60'. The units committed in the first time interval (from *T*–45' to *T*–30') receive binding commitment instructions to start-up or shut-down. Units committed in some subsequent time interval also receive a commitment instruction if they need adequate time to comply with the binding commitment instruction.

Both RTPD and STUC recognize initial conditions and respect binding commitment instructions (start-ups or shut-downs) that have been instructed by RTPD and STUC previously. Therefore, the unit commitment of some resources is pre-determined or constrained by previously issued binding commitment instructions and initial conditions.

The RTPD also produces a dispatch for 5-minute dispatchable resources for every 15-minute interval in the RTPD Time Horizon on a 15-minute basis. However, these 15-minute dispatches are not communicated to resource schedulers; they are advisory information to the CAISO only. The dispatch for these resources will be finalized by the Real-Time Economic Dispatch (RTED) application on a 5-minute basis.

3.7.5 Real-Time Economic Dispatch

This Section describes the Real-Time Economic Dispatch (RTED) function of the RTM. The RTED uses the SCED optimization engine and does not perform Unit Commitment or enforce any inter-temporal constraints that rely on the use of integer decision variables. The RTED also does not procure any additional Ancillary Services (AS). The resource commitment status and their AS awards are obtained from the RTPD and they remain fixed in the RTED, except for any Out Of Sequence (OOS) commitment instructions and outages that may occur. The RTED dispatches resources to meet Imbalance Energy requirements over the relevant Time Horizon at minimum cost based on the available energy bids, and subject to resource and network constraints as applicable. These energy bids may be mitigated by the MPM) and RRD application that is executed prior to HASP; they may include penalty prices to enforce self-schedule and contingent Operating Reserve priorities.

The RTED produces dispatch instructions that after operator review are sent via ADS to the relevant SCs. The RTED also calculates Locational Marginal Prices (LMPs) for each dispatch interval and the Dispatch Operating Point (DOP) for each dispatchable resource as a function of time. The DOP is the *expected* trajectory of the power output of a resource as it responds to dispatch instructions. The DOP traverses through the Dispatch Operating Targets (DOTs) based on the cross-interval and cross-hour ramping rules as described earlier.

The RTED has three mutually exclusive operating modes that are described in the next three sections.

3.7.5.1 Real-Time Interval Dispatch

The Real-Time Interval Dispatch (RTID) is the normal operating mode of the RTED. RTID runs every 5 minutes at the middle of each 5-minute interval. The RTID Time Horizon is currently composed of 7 5-minute intervals and it starts 7½ min (the time delay) after each dispatch time. The dispatch for the first dispatch interval of the Time Horizon is financially binding and is communicated to the SCs via the ADS. The dispatch for the remaining dispatch intervals in the Time Horizon is advisory and is not communicated to the SCs.

Figure 3-7 illustrates the periodic execution of the HASP, STUC, RTPD, and RTID, and their Time Horizons and time interval compositions. The time delay between the dispatch time and the start of the Time Horizon allows for execution time, operator review, communication of dispatch instructions through ADS, and resource ramping toward the DOT. For RTID, the time delay is 7½ min and is composed of the following: a) execution time of up to 2 min; b) operator review period of at least 75 sec; c) ADS time window of 105 sec before the next dispatch time when the ramp toward the DOT begins; and d) 2½ min, i.e., half a dispatch interval, of cross-interval ramping before the start of the first dispatch interval of the Time Horizon. The dispatch instructions are automatically sent through ADS at the end of the review period, i.e., 105 sec before the next dispatch run, unless they are blocked by the operator.

All RTM applications exchange data continuously. Binding hour-ahead schedules and AS Awards for hourly pre-dispatched resources calculated by the HASP are kept fixed in STUC, RTPD, and RTED, notwithstanding any operating adjustments. Binding AS Awards calculated by the RTPD are also kept fixed in RTED, notwithstanding forced outages. Binding commitment decisions calculated by the HASP, RTPD, or STUC, are respected by all RTM applications in subsequent executions.

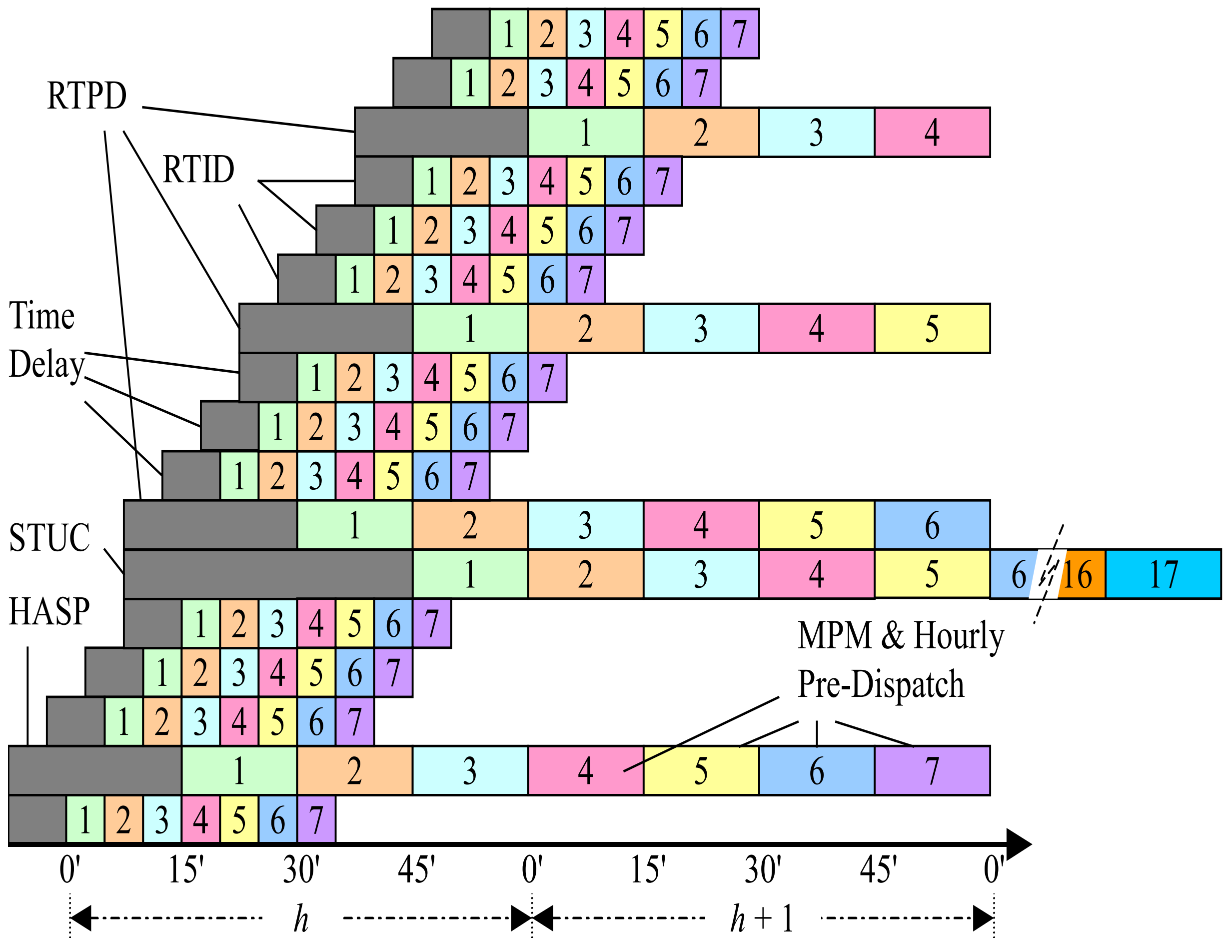


Figure 3-7: RTM Time Horizon Layout

RTID does not enforce any inter-temporal constraints, except for ramp rate limits and energy limits. RTID employs the Full Network Model, however, it does not iterate with the Network Analysis function to perform power flows or contingency analysis. RTID uses the loss penalty factors and the power transfer distribution factors calculated by the last network linearization by RTPD. Therefore, it is assumed that the network remains linear as the RTPD dispatch over a 15-min interval is successively refined by RTID on a 5-min basis.

3.7.5.2 Real-Time Manual Dispatch

The Real-Time Manual Dispatch (RTMD) application is a backup dispatch process that can be activated at any time when the RTID fails to converge because of bad data, network model issues, or optimization engine failure[11]. When activated, the RTMD executes every 5 minutes in place of the RTID at the same dispatch times, until the operator switches back to the RTID. However, the other RTM applications, namely RTPD, and HASP and STUC, continue to execute normally every 15' and every hour, respectively.

The RTMD is different from the RTID in the following aspects:

- RTMD has a single 5-minute dispatch interval.
- Resources are dispatched in merit order based on Energy bids and ramp rate limits, but ignoring transmission losses and network constraints. The operator is presented a merit order list of feasible bids and a system Imbalance Energy requirement that needs to be met. Bids are selected in the merit order list to meet the Imbalance Energy requirement.
- The dispatch must be sent to ADS manually by the operator; no automatic transfer takes place.
- A system-wide Market Clearing Price (MCP) (as opposed to LMPs) is produced for the dispatch interval.

3.7.5.3 Real-Time Contingency Dispatch

The Real-Time Contingency Dispatch (RTCD) application can be activated at any time to address contingencies by dispatching contingent Operating Reserve. The RTCD is activated when a State Estimator (SE) solution is available after the occurrence of a contingency that qualifies as a significant event. The SE solution is available normally at the middle of 5-minute clock intervals to provide a base operating point for RTID.

The RTCD is different from the RTID in the following aspects:

- The dispatch time can be any time.
- When the RTCD is activated, the RTID or RTMD, whichever mode is active, is immediately suspended and may not resume until after the next 5' clock Dispatch Time.
- Any pending dispatch instructions before the RTCD is activated are discarded and not sent to ADS.
- The RTCD has a single 10-minute dispatch interval. A 10-minute dispatch interval is used so that Non-Spinning Reserve from resources with zero minimum load and no inter-temporal constraints can be optimally dispatched.

---

[11] Harry Singh, Alex Papalexopoulos "Alternative Design Options for a Real-Time Balancing Market," presented at the 2001 IEEE PICA Conference, Sydney, Australia, May 19-24, 2001

- The time delay between the dispatch time and the start of the ramp for the DOT is configurable and initially set to two minutes.
- Contingent Operating Reserve bids are automatically released for optimal dispatch.
- The dispatch must be sent to ADS manually by the operator; no automatic transfer takes place.

The RTCD calculates a contingent dispatch for the 10-minute dispatch interval, presents the dispatch to the operator, and awaits approval. The operator may block or edit individual dispatches as usual, and then manually sends the dispatch to ADS, or blocks and discards the entire dispatch. The RTCD stays idle afterwards awaiting operator instructions, however, the RTPD is still running every 15 minutes. The operator may manually initiate another RTCD run, or switch to RTID or RTMD, which resumes at the next applicable dispatch time (the midpoint of a 5-minute clock interval).

3.8 Congestion Revenue Rights (CRR or FTR) Market

3.8.1 Overview

CRRs (or FTRs) are financial instruments that enable holders of such instruments to manage variability in Congestion costs that occur under Congestion Management protocol that is based on locational marginal pricing. CRRs are mainly acquired for the purpose of offsetting costs associated with IFM Congestion costs that occur in the Day-Ahead Market. They can also be used for other legitimate activities, many of which will increase the liquidity of the CRR market. Only CRR Obligations can be acquired through the CRR Allocation and CRR Auction processes. CRR Options are not available through the CRR Allocation and CRR Auction processes and are only available for Merchant Transmission Facilities.

There are two types of CRRs: CRR Obligations and CRR Options:

CRR Obligation – A CRR Obligation entitles its holder to receive a CRR Payment if the Congestion in a given Trading Hour is in the same direction as the CRR Obligation, and requires the CRR Holder to pay a CRR Charge if the Congestion in a given Trading Hour is in the opposite direction of the CRR. CRR Payments to CRR Holders of CRR Obligations are based on the per-MWh cost of Congestion, which equals the positive amounts of Marginal Cost of Congestion (MCC) at the CRR Sink minus the MCC at the CRR Source multiplied by the MW quantity of the CRR. CRR Charges for CRR Obligations associated with Congestion in the opposite direction are based on the negative amounts of the difference in MCC between the CRR Sink and CRR Source.

CRR Option – A CRR Option entitles its Holder to a CRR Payment if the Congestion is in the same direction as the CRR Option, but requires no CRR Charge if the Congestion is in the opposite direction of the CRR. CRR Payments to CRR Holders of CRR Options are based on the per-MWh cost of Congestion, which equals the positive amounts of Marginal Cost of Congestion (MCC) at the CRR Sink minus the MCC at the CRR Source multiplied by the MW quantity of the CRR. There are no CRR Charges associated with Congestion in the opposite direction of CRR Options.

All CRRs held by CRR Holders are settled with revenue collected in the IFM Congestion Fund. CRR Obligations can be acquired as Point-to-Point (PTP) CRRs. A PTP CRR is a CRR Obligation defined from a single CRR Source to a single CRR Sink.

There are four terms for CRRs:

Monthly CRR – A CRR acquired for one calendar month. Monthly CRRs are made available on a TOU basis.

Seasonal CRR – A CRR acquired through the annual CRR Allocation or CRR Auction process that has a term of one season and either on or off peak. For the purpose of the CRR processes, a season is defined as follows: season 1 is January through March, season 2 is April through June, season 3 is July through September and season 4 is October through December.

Long Term CRR – One of the tiers in the annual allocation process is the Tier LT. Long Term CRRs have a term of 10 years and are allocated on a seasonal/TOU basis.

Merchant Transmission CRR – The Merchant Transmission CRR has a term of 30 years or the pre-specified intended life of the facility, whichever is less. The acquisition of the Merchant Transmission CRR is performed through a separate process.

The following processes exist for the creation and acquisition of CRRs:

- CRRs are created by the CAISO through the CRR Allocation and CRR Auction processes and through the allocation of Merchant Transmission CRRs.
- Only internal Load Serving Entities (LSEs) can participate in the CRR Allocation.
- After the annual (including the Long Term CRR Allocation process) and monthly CRR Allocation processes there is an annual and monthly CRR Auction for any entity interested in acquiring CRRs. The annual auction will not include the auction of Long Term CRRs.
- Parties may also acquire CRRs from CRR Holders through the Secondary Registration System (SRS) through which CRRs are traded bilaterally.
- Transferees must also qualify as Candidate CRR Holders prior to acquiring CRRs.

3.8.2 Yearly Calendar of Allocations & Auctions

The CAISO conducts an annual CRR Allocation and CRR Auction once a year. The annual CRR Allocation and CRR Auction release Seasonal CRRs for four seasonal periods and two time-of-use periods, on peak and off peak. These seasonal/TOU periods coincide with the calendar quarters (season 1 – January through March, season 2 – April through June, season 3 – July through September, and season 4 – October through December). Part of the annual CRR Allocation process includes the release of Long Term CRRs (i.e. Tier LT), which if an entity chooses to participate in, provides the ability to obtain allocated CRRs for a period of ten years. The Long Term CRRs are also allocated based on the four seasonal and two time-of-use periods mentioned above. The CAISO also conducts monthly CRR Allocations and CRR Auctions twelve times a year in advance of each month. Within each annual and monthly CRR Allocation and CRR Auction process, ISO performs distinct processes for each on-peak and off-peak period.

Each CRR Allocation process is based on nominations submitted to the CAISO by LSEs eligible to receive CRRs. The CAISO posts the specific timelines each year no later than June 30 of each year.

3.8.3 Key Steps Performed in the CRR Allocation & Auction Processes

This Section provides an overview of the key steps that are completed as part of the CRR Allocation and Auction processes.

1. The CAISO prepares network model, constraints, Aggregated Pricing Node (APNode) mapping, and contingencies.
2. Candidate CRR Holders register for the CRR Allocation and/or CRR Auction.
3. The CAISO performs verification data collection process for entities participating in the CRR Allocation.
4. The CAISO announces dates of CRR Allocation and Auction markets.
5. Nominations that reflect the rights under Transmission Ownership Rights (TOR).
6. The CAISO inputs Transmission Ownership Rights (TOR) nominations and runs the Simultaneous Feasibility Test (SFT).
7. The Annual CRR Allocation process begins with the simultaneous running of Seasons 1 through 4.
8. The annual allocation process consists of four tiers for each of the four seasons.
9. The annual CRR Auction begins once the annual CRR Allocation process is completed. Interested Candidate CRR Holders or CRR Holders provide the necessary creditworthiness requirements to the CAISO accThe CAording to the CRR Auction timeline.

10. The CAISO opens the bid submission window for the annual CRR Auction markets for all four seasons and runs the SFT for all four seasons and two TOUs, reviews and posts results.
11. The monthly CRR Allocation and CRR Auction follows a similar process as described above for the annual CRR Allocation and CRR Auction.

3.8.4 Annual CRR Allocation

The CAISO uses the following sources for submitting CRR nominations in the CRR Allocation process:

- Generating Unit PNodes
- Trading Hubs
- Scheduling Points
- Points of Delivery associated with Existing Transmission Contracts

The CAISO makes available 65% of Seasonal Available CRR Capacity for the annual CRR Allocation and CRR Auction processes, 60% of wjhat is left of the Seasonal Available CRR Capacity in Tier LT, and 100% of Monthly Available CRR Capacity for the monthly CRR Allocation and CRR Auction processes. The percentages noted above are also applied to the Operating Transfer Capability values to be used for each of the Scheduling Points.

In the annual or monthly CRR Allocation and CRR Auction processes, the CAISO accounts for any Merchant Transmission CRRs as Fixed CRRs on the DC CRR FNM that is used in the SFT for the CRR Allocation and CRR Auction.

For the purpose of the annual CRR Allocation and CRR Auction, the CAISO assumes all transmission facilities, within the ISO Controlled Grid, are in service unless it is aware of a major outage that is scheduled for a long portion of one or more of the seasons in the annual process. If the outage is deemed to be significant then the CAISO may choose to reduce the operating limit of the facility or to take the facility completely out of service. For the purpose of the annual CRR Allocation and CRR Auction, the CAISO assumes that all lines are in-service unless a scheduled outage of a significant facility is known in time to reflect that outage in the FNM for the annual process.

3.8.5 Annual CRR Auction

The annual CAISO CRR Auction process takes place after the four-tiered annual CRR Allocation process. Any Candidate CRR Holder or CRR Holder may participate in the CRR Auction subject to the creditworthiness requirements under the ISO Tariff. Candidate CRR Holders or CRR Holders interested in the annual CRR Auction submit bids, once the market is opened, via the CRR MUI to the extent that they do not exceed their Aggregate Credit Limit.

Any Market Participant (MP) wishing to participate in the CRR Auction must complete the registration process in advance of any CRR market in which they want to participate. Once an entity has successfully registered and fulfilled all the requirements, they become a Candidate CRR Holder. An entity that is already a CRR Holder is not required to go through the registration process again to participate in the CRR Auction, but is still subject to the CAISO's creditworthiness requirements.

Once the bid submittal period is over, the CAISO runs the SFT and optimization and returns results to the CRR Holders via the CRR MUI. ISO publishes a detailed timeline for the annual CRR Allocation and CRR Auction processes 30 days prior to the beginning of the CRR Allocation and Auction process.

3.8.6 CRR Bid Submission

Each bid to buy a Point-to-Point CRR must specify the following information:

- The associated season and time-of-use period
- The associated CRR Source and CRR Sink
- A monotonically non-increasing piecewise linear bid curve in quantities (denominated in thousandths of MW) and prices ($/MW)

In general, any bid point (quantity, price) is allowed, as long as the first MW quantity is zero. However, when the last bid point of any bid curve leads to having a vertical bid segment, such a point is not further considered downstream in the CRR application as it has no meaning in the optimization engine. Figure 3-8 shows a two-segment bid. The last bid point leads to a vertical segment and, therefore, is ignored. In this case, both the pre-auction credit requirements and the auction clearing processes use the equivalent one-segment bid shown in Figure 3-9.

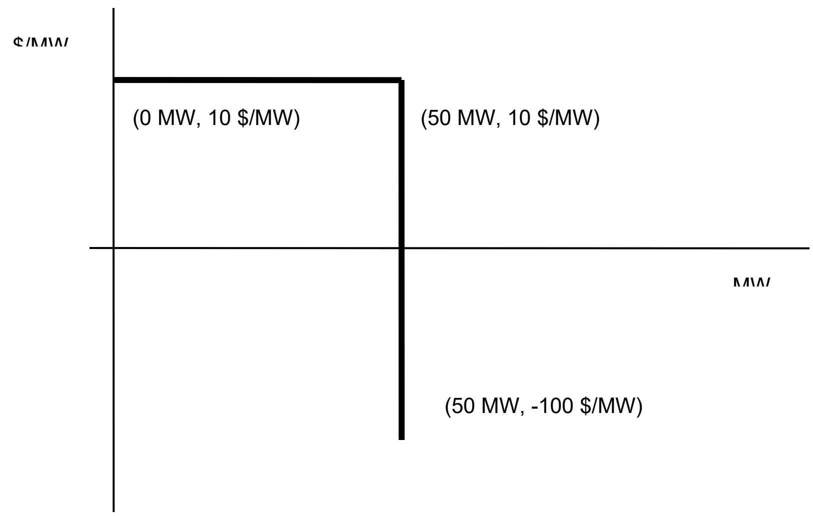


Figure 3-8: CRR Buy Bid

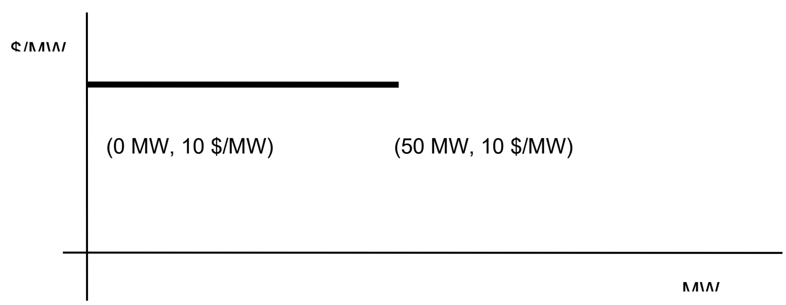


Figure 3-9: Equivalent One-Segment Bid

If an entity is interested in removing a CRR from its portfolio there are two mechanisms by which this can be performed; the Secondary Registration System, or by offering to sell the CRR in the auction process. To sell a CRR in the auction the holder of the CRR must offer it for sale during an auction period for which the CRR is active. For example, if a CRR was acquired in the 2022 annual process for season 2, then that CRR could only be offered for sale in the 2019 monthly auctions for April, May and June of 2022. The MW of the first bid point must be zero and the maximum MW offered for sale must be less than or equal to the available MW of the CRR. If a CRR has been offered for sale in the SRS a CRR sell bid cannot be submitted in the auction until the SRS offer is expired.

3.8.7 Monthly CRR Allocation & Auction

The monthly CRR Allocation process is similar to the annual CRR Allocation process. Each month the CAISO uses a monthly Demand Forecast submitted through the CRR

MUI to calculate two load duration curves (one on-peak and one off-peak load duration curve for the applicable month) to form the basis for monthly allocations.

The monthly CRR Auction process takes place after the two-tiered monthly CRR Allocation process. Any Candidate CRR Holder or CRR Holder may participate in the monthly CRR Auction subject to the creditworthiness requirements under the ISO Tariff. Candidate CRR Holders interested in the monthly CRR Auction submit bids, once the market is opened, via the CRR MUI to the extent that they do not exceed their Aggregate Credit Limit. Once the bid submittal period is over, the CAISO closes the market and runs the SFT and optimization and returns results to the CRR Holders via the CRR MUI. ISO publishes a detailed timeline for the monthly CRR Allocation and CRR Auction processes 30 days prior to opening a market.

The allowable CRR Sources and CRR Sinks in the CRR Auction process are Generator PNodes, Scheduling Points, Trading Hubs, LAPs, and Sub-LAPs. The submitted bids have similar format like in the Annual CRR process.

3.8.8 Simultaneous Feasibility Test

The annual and monthly CRR Allocation and CRR Auction processes release CRRs to fulfill CRR nominations and bids as fully as possible, subject to a Simultaneous Feasibility Test (SFT). For the CRR Allocation, to the extent that CRR nominations are not simultaneously feasible, the nominations are reduced in accordance with the CRR Allocation optimization formulation until simultaneous feasibility is achieved. For the CRR Auction, to the extent that bids are not simultaneously feasible, the bids are reduced in accordance with the CRR auction optimization formulation.

The main purpose for applying the SFT is to help ensure that the CRRs created through an allocation or auction process are revenue adequate. Revenue adequacy is the situation in which, over a given period, at least as much congestion revenue is collected by the CAISO than is paid out in CRR entitlements to the CRR Holders.

In the CRR Allocation process, the SFT is applied by modeling CRR Source nominations and CRR Sink nominations as injections and withdrawals, respectively, onto a Full Network Model (FNM). The location and amount of injection is based on the definition of the CRR Source in terms of its member Pricing Nodes (PNodes) and allocation factors[12]. The location and amount of withdrawal is based on the definition of the CRR Sink in terms of its member PNodes and allocation factors. The same process applies to the CRR Auction process. The application of the sources and sinks onto the FNM creates flows on operating constraints based on the nominated MW amount in the case of an allocation and the MW amount from the bid curve in the case of the auction. These flows, along

[12] Allocation Factors are the weights that define the fractional MW amount that is mapped from a Source or Sink to a Pnode.

with any flows due to Fixed CRRs, are compared to the constraint limits. All nominated CRR Sources and CRR Sinks are applied simultaneously in the CRR Allocation process and likewise all CRR bids are applied simultaneously in the CRR Auction process. The comparison of the resulting flows against the constraint limits is performed simultaneously. This simultaneous comparison is the SFT. The SFT is employing a DC network. Note that Point-to-Point CRR nominations and bids are pre-balanced in terms of injection and withdrawal amounts for the SFT.

If the SFT fails, for a given set of nominations or bids, feasibility must be achieved. This is accomplished by reducing the MW quantity amounts associated with the nomination or bids. This reduction is performed through an optimization process. In fact, the SFT is embedded within an optimization formulation. The optimization formulation has an objective function and a set of constraints. During the optimization process, the objective function is either maximized (or minimized which ever is the case) while all constraints are simultaneously satisfied (i.e., not violated).

There are two basic objective function formulations that can be utilized for allocating CRRs:

- Maximizing CRR MW (Max CRR)
- Weighted Least Squares (WLS)

When the CRR process first started in 2009 the optimization formula was to maximize CRRs. Under this formulation the nomination that is most effective in relieving the constraint was curtailed completely before going to the next most effective nomination. In contrast, the current WLS CAISO SFT optimization algorithm distributes the curtailment across all CRR nominations that are effective in relieving the congestion, and thus spreads the curtailment among multiple allocation participants. The WLS is considered a more equitable formulation for the CRR allocation process; in that CRR nominations share the available capacity.[13]

[13] This problem is relevant to the allocation process only, not to the CRR auction. In a CRR auction the auction participants use their bid prices to convey their value on each CRR, and the auction objective is to maximize the CRR MW amounts resulting from clearing the auction. As a result, when there is a congested constraint the SFT curtails the CRR bids based on the participants' bid prices so as to minimize the reduction in the CRR MW amounts. In an allocation process there are no economic bids, so all nominated CRRs are identical from a financial perspective.

3.9 Resource Adequacy (RA) Mechanism

3.9.1 RA Requirements, Deficiencies & Timelines

SCs submit to the CAISO annual and monthly Resource Adequacy Plans and identify the specific resources that the LSEs use to rely on that satisfy their forecasted monthly Peak Demand and Reserve Margin as well as their local requirements for the relevant reporting period. The annual and monthly RA Plans must be submitted pursuant to the schedule set forth on the Reliability Requirements website. The monthly RA process is illustrated in the Figure 3-10 below.

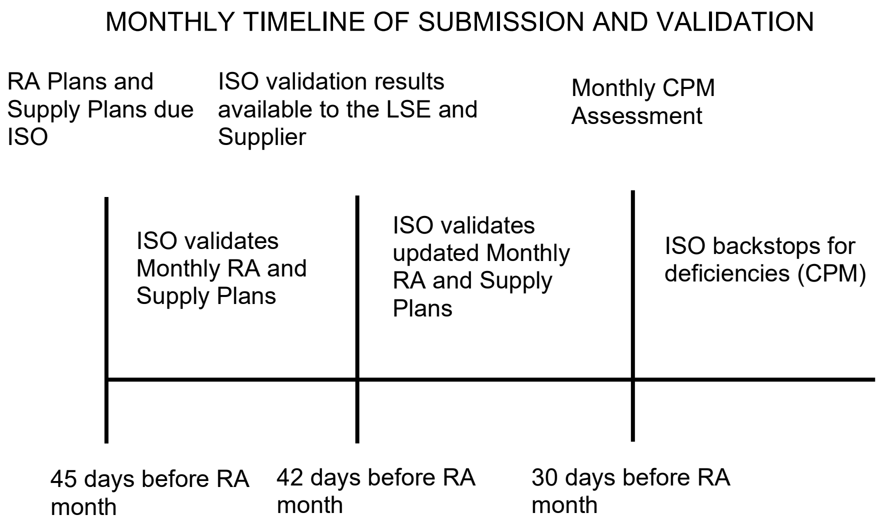


Figure 3-10: Monthly RA Timeline

A Cross Validation between RA and Supply Plans, RA Resources, is performed by the CAISO following the completion of individual validations of RA and Supply Plans. The Cross Validation is performed to ensure that the information contained in the RA Plan correctly matches its corresponding Supply Plan. The Cross Validation includes, but is not limited to, the following information:

- All Resources in the RA Plan are present in the Supply Plan
- The RA Capacity present in the RA Plan is equal to or less than the RA capacity in the Supply Plan
- All flexible capacity resources in the RA Plan are present in the Supply Plan

- The Flexible RA Capacity identified for each category in the RA Plan is equal to or less than the Flexible RA Capacity present for each category in the Supply Plan
- The submitted Local Capacity Area Resources are also validated.

Once plans are submitted the CAISO analyzes for flexible RA capacity deficiencies on a cumulative system-wide basis, at the Local Regulatory Authority level, and at the individual LSE level. The CAISO evaluates whether the flexible capacity shown by all Load Serving Entities in aggregate meets the cumulative CAISO system flexible need. In performing this analysis, the CAISO caps the flexible capacity in the Peak Ramping and Super-Peak Ramping categories at their respective total Peak Ramping and Super-Peak Ramping maximums calculated by the CAISO.

If the CAISO cumulative system flexible capacity needs are not met, the CAISO analyzes for each Local Reliability Area (LRA), whether the flexible capacity in the plans of the LSEs under the LRA's jurisdiction cumulatively meet the LRA's total flexible need and the LRA's Base Ramping need. In performing this analysis, the CAISO caps the flexible capacity in the Peak Ramping and Super-Peak Ramping categories at their respective total LRA Peak Ramping and Super-Peak Ramping maximums.

If the cumulative CAISO system is short of its flexible needs, and an LRA is short of its flexible needs, the CAISO calculates the individual LSE deficiency for LSEs under the jurisdiction of the deficient LRA. The CAISO provides the results of this calculation to the deficient LRA and the SC for each LSE of that LRA to show the risk of Capacity Procument Mechanism (CRM) designation and cost allocation, calculated using the CAISO method.

LSEs under the jurisdiction of Local Regulatory Authorities that have established a flexible need allocation methodology may re-submit RA Plans to cure deficiencies identified by their Local Regulatory Authority. These LSEs may also choose to provide updated RA Plans in order to mitigate possible CPM cost allocation risk if they are individually deficient.

Any other LSE that receives an ISO notice of deficiency may provide, no less than 30 days prior to the first day of the month covered by the plan, an updated RA Plan demonstrating that it meets the flexible capacity needs in order to cure the deficiency and avoid possible CPM cost allocation.

3.9.2 Competitive Solicitation Process (CSP)

The CAISO procures and prices backstop capacity designated under the capacity procurement mechanism (CPM) through a competitive solicitation process (CSP). This process is set up to run annually, monthly, and intra-monthly to cover all potential CPM designations. Participation in a CSP is voluntary.

The foundation of the CSP is based on the following three principles

- Supplier to submit offers to the solicitation process
- Mitigation rules

- Determining which resource the CAISO will offer a CPM designation

Suppliers submit offers into the CSPs according to a defined set of rules that outline the offer format, product definition, and other terms and conditions of offering into the CSP. The CSP has mitigation measures. First, the CSP includes a soft offer cap on all bids.Second, the timeline forces suppliers to bid resource capacity into the CSP prior to the knowledge of whether and what type of CPM event may occur. Finally, any offers above the soft offer cap price must be cost-justified at FERC to recover up to a resource-specific cost of service rate. If FERC approves the price, the SC submits relevant information to the CAISO so that its settlements system is updated accordingly.

There is a set procedure for determining which resource to offer a CPM designation. The procedure is based on the relative offers of capacity a CSP and transparent evaluation criteria.

If an annual or monthly CPM event is triggered due to deficiency in RA showings or a collective deficiency in the RA showings then the CSP offers are optimized and a selection of resources is based on minimizing overall cost of meeting designation criteria performed by a CAISO optimization engine. If there is insufficient capacity offered in to the CSP then ISO inserts offers for non RA capacity at the soft offer cap plus a penalty price defined in CAISO Database.

3.9.3 RA Capacity DAM & RTM Market Participation Requirement

SCs representing Resource Adequacy Capacity procured by LSEs must make the Resource Adequacy Capacity listed in the Scheduling Coordinator's monthly Supply Plan available to the DAM and RTM CAISO markets. Specifically,they must Self-Schedule or submit Economic Bids for all Resource Adequacy Capacity into the IFM and RUC for all hours that the resource is physically available, unless an Outage affecting Resource Adequacy Capacity has been reported to the CAISO. SCs for Resource Adequacy Resources that do not submit Self-Schedules and instead submit Economic Bids reflecting all their Resource Adequacy Capacity are subject to CAISO optimization for that capacity in the DAM. Resource Adequacy Capacity selected in RUC is not eligible to receive a RUC Availability Payment. Resource Adequacy Capacity subject to RUC is optimized at a zero dollar RUC Availability Bid.

Resource Adequacy Resources that are committed by the CAISO in the IFM or RUC for Resource Adequacy Capacity or have Self-Schedules for part of their Resource Adequacy Capacity must remain available to the CAISO for their full amount of RA Capacity through the RTM. Resource Adequacy Capacity from Short Start Units that were not scheduled in the IFM or committed in RUC, are required to be bid or self-scheduled in the HASP or RTM, subject to any limitations for use-limited resources. Resource Adequacy capacity from System Resources is <u>not</u> required to be offered in the RTM if not scheduled in the DAM.

To the extent Resource Adequacy Resource Capacity is not scheduled for Energy or as RUC capacity in the DAM, such capacity may also be offered or bid in the Real-Time

Market to support a Self-Scheduled export in HASP that would have an equal priority as the ISO Forecast of CAISO Demand.

The CAISO determines if all dispatchable Resource Adequacy Capacity not otherwise selected in DAM or RUC, is reflected in a Bid into the RTM and automatically inserts a Generated Bid in the RTM for any remaining dispatchable Resource Adequacy Capacity for which the CAISO has not received notification of an Outage. As the ISO does not automatically submit bids for Use-Limited Resources, SCs must actively submit all required Energy Bids or Self-Schedules into the RTM for these resources.

The CAISO also optimizes flexible capacity participating in RUC using $0/MW-hour for all Flexible RA Capacity that is not reflected in an IFM Schedule in only the required hours for the resource's committed Flexible RA Capacity categories.

In determining the amount of capacity that is optimized in RUC at $0/MW-hour, the CAISO assumes maximum overlap of RA Capacity and Flexible RA Capacity on the given hour. For example, a resource with 100 MW of RA Capacity and 150 MW of Flexible RA Capacity that is awarded an IFM Schedule of 100 MW in HE13 is automatically optimized in RUC using $0/MW-hour for the additional 50 MW of Flexible RA Capacity.

3.9.4 Resource Adequacy Substitution

RA resources are expected to be available during the entire month. The CAISO substitution rule provides opportunities for RA resources to take maintenance outages under specific conditions when there is advance notice of the outage. Resources also experience forced outages, when advance notice is not possible. The availability incentive mechanism is designed to provide resources with incentives to undertake actions to reduce the occurrences of forced outages in a month. In order to allow resources to manage their availability incentive risk, the CAISO has developed substitution rules that allow capacity from resources to "substitute" for RA capacity which has experienced a planned/forced outage. A resource on an outage has the option to provide substitute RA capacity to mitigate any potential impact to the original RA resource's availability incentive calculation. Substitution mechanism allows the supplier of Resource Adequacy Capacity that is tied to a specific generating resource the ability to substitute that capacity in the event the Resource Adequacy resource is on an Outage.

There are three types of Resource Adequacy Capacity – Local, System and Flexible. Local and System RA are commonly known as generic RA capacity type while Flexible is known as Flexible RA Capacity type. The substitution capacity is further categorized as Day Ahead or Real Time Substitution depending on the timeline of submission. An RA resource on outage has the option to provide substitute RA capacity to mitigate any potential impact to the original RA resource's availability incentive calculation.

It is the responsibility of the SC to provide sufficient substitution capacity to satisfy the Resource Adequacy Substitute Capacity (RASC) assignment to avoid getting the outage denied.

Example: Outage on resource A is from 10/1 to 10/31 and the Outage was submitted to ISO on 9/20 at 10QAM. On 9/21 at 8AM, ISO outage assessment considers this outage and assigns eligibility to resource A. Supplier then has to provide substitute capacity for trade days 10/1 to 10/31 before 9/22.

If the RA resource on an outage provides substitute capacity, the obligation on the resource on outage transfers to the substitute resource up to the MW amount provided. The must offer obligation and assessment is transferred to the substitute capacity and the original resource's capacity is not assessed under the availability incentive mechanism for each day substitute capacity is provided.

A non-Resource Adequacy Resource that the CAISO approves to substitute for Resource Adequacy Capacity becomes a Resource Adequacy Resource for the duration of the substitution.

3.9.5 Procurement Mechanism

The CAISO has the authority to designate Eligible Capacity to provide CPM Capacity services under the CPM to address the following circumstances:

- Insufficient Local Capacity Area Resources in an annual or monthly Resource Adequacy Plan
- Collective deficiency in Local Capacity Area Resources
- Insufficient Resource Adequacy Resources in an LSE's annual or monthly Resource Adequacy Plan
- A CPM Significant Event
- A reliability or operational need for an Exceptional Dispatch CPM

Eligible Capacity is the capacity of Generating Units, System Resources or Participating Load that is not already under a contract to be a Resource Adequacy Resource, is not under an RMR Contract, and is not currently designated as CPM Capacity. Eligible Capacity must be capable of effectively resolving a procurement shortfall or reliability concern.

An "CPM Significant Event" is defined as a substantial event, or a combination of events, that is determined by the CAISO to either result in a material difference from what was assumed in the resource adequacy program for purposes of determining the Resource Adequacy Capacity requirements, or produce a material change in system conditions or in CAISO Controlled Grid operations, that causes, or threatens to cause, a failure to meet Reliability Criteria absent the recurring use of a non-Resource Adequacy Resource(s) on a prospective basis.

The process incldues three (3) steps:

In the first step, the CAISO may procure CPM Capacity which can have an initial term of thirty (30) days.

In the second step, if the CAISO determines that the CPM Significant Event is likely to extend beyond the thirty (30) day designation period, the CAISO may extend the CPM Capacity designation for another sixty (60) days.

In the third step, the CAISO conducts an assessment of any proposed solutions to determine whether they totally or partially can mitigate the need for ongoing CPM Capacity. The CAISO can consider and implement such alternative solutions provided by Market Participants in a timely manner, but no sooner than the day after the end of the 90-day designation period. If Market Participants do not submit any alternatives to the designation of CPM capacity that are fully effective in addressing the deficiencies in Reliability Criteria resulting from CPM Significant Event, the CAISO will extend the term of the designation.

Exceptional Dispatch CPM designations occur either in the post-DAM time frame or in real-time during the trading day. In the post-DAM time frame, which is a 12 hour period occurring up to 6 hours in advance of the next trading day, the CAISO's tools are not always capable of identifying and committing a "final" quantity as there are many variables the CAISO relies upon that may change during that time frame.[14] Real-time Exceptional Dispatch CPM designations commit or dispatch the resource, depending upon start-up requirements and previous schedules and awards, to an actual megawatt quantity.

3.10 Virtual Bidding Market

3.10.1 Overview

"Convergence" or Virtual Bids are financial bids submitted only in the Day-Ahead Market. The Integrated Forward Market (IFM) clears Virtual and physical bids in a non-discriminatory manner. If cleared in the IFM, the resulting Virtual Supply and Virtual Demand Awards settle first at the locational Day-Ahead LMP and then be automatically liquidated with the opposite sell/buy position at the RTM LMPs (specifically the 15 minute market (FMM).

Convergence bidding provides Market Participants with several financial functions. First, there is the opportunity to earn revenues (and to risk losses) resulting from any differences in the Day-Ahead and RTM LMPs. Market Participants, using their insights into system and market conditions, may be able to identify Virtual Bidding opportunities that result in more efficient market outcomes. The potential for financial reward encourages Virtual Bidding activity that would tend to minimize any systematic differences between Day-Ahead and FMM LMPs, thus minimizing incentives for under or over-scheduling physical Demand in the Day-Ahead Market. A generator owner can also use

[14] The post-day ahead window begins at 6:00 p.m. on the day before the Trading Day and ends at 6:00 a.m. on the Trading Day,

a Virtual Bid to mitigate the risk impact of an outage that happens after the close of the Day-Ahead Market. By increasing market liquidity through Virtual Bidding, the potential for the exercise of market power also decreases.

3.10.2 Virtual Bidding & DAM Market Participation

SCs submit Bids for Virtual Supply and Virtual Demand for each resource to be used in DAM. SCs may submit Virtual Bids for DAM as early as seven days ahead of the targeted Trading Day and up to Market Close of DAM for the target Trading Day.The CAISO validates all Bids submitted to DAM. In the case of Virtual Bids (Supply and Demand), credit checks are performed against the Parent SC's (which provides financial collateral for itself and subordinate SCs) available credit limit prior to passing the Virtual Bids to the Day-Ahead Market.

The Virtual Bids are explicitly flagged as Virtual Bids when submitted to the Day-Ahead Market. Their submission and processing includes an indicator that identifies them as Virtual Bids rather than physical Bids. This indication;

- Allows for their exclusion from the automated Local Market Power Mitigation process;
- Allows the Virtual Bids to be tracked and associated with the Convergence Bidding Entity;
- Allows the CAISO to be able to suspend Virtual Bids by location or by Convergence Bidding Entity when necessary; and
- Allows the CAISO to exclude virtual bids from the RUC market

Virtual Supply Bids at the Aggregated PNode locations will have the GDF applied just as physical Supply Bids thereby treating physical and Virtual Bids consistently in the Day-Ahead Market.

Virtual Energy Bids are Economic Bids and do not include Self-Schedules. These Bids can be at any Eligible PNode, or Eligible Aggregated PNode location and be a Virtual Supply Bid and/or Virtual Demand Bid at that location.

Day-Ahead Economic Virtual Bids for Supply are limited to the Energy Curve defined in the Bid. For Virtual Bids this is required and the Resource Type selected must be Virtual Supply. Virtual Supply Bids must start at zero (0) MW.

The Energy Bid Curve must be monotonically increasing. Virtual Supply Bids are validated by the CAISO upon submission to ensure that the Energy Bid Curve complies with bid validation rules. Virtual Supply Bids are subject to the energy bid caps.

Day-Ahead Economic Virtual Bids for Demand are limited to the Energy Curve defined in the bid. For Virtual Demand Bids this is required and the Resource Type selected must be "Virtual Demand". Virtual Bids must start at 0 MW. Virtual Demand Bids are validated by the CAISO upon submission to ensure that the Energy Bid Curve complies with bid validation rules. Virtual Demand Bids are subject to the Hard Energy Bid Cap. Virtual Demand Bids are subject to the energy bid caps.

At any time, the energy cap for resource-specific resources is $1,000/MWh. SCs may submit bids above $1,000/MWh and up to $2,000/MWh. However, bids above $1,000/MWh must be cost-justified through the submission of a Reference Level Change Request to the resource's Default Energy Bid (DEB). Bids above $1,000/MWh will be reduced to the higher of $1,000/MWh and the resource's Revised DEB as modified by an approved Reference Level Change Request. The Revised DEB cannot exceed $2000/MWh, meaning that the CAISO will reject bids that are submitted above $2,000/MWh.

3.10.3 Virtual Bidding DAM Market Clearing

At the Day-Ahead Market close (currently 10:00 a.m.) the application aggregates the Virtual Bids at each Eligible PNode/APNode to create one aggregate Virtual Supply Bid and one aggregate Virtual Demand Bid at each location (the aggregate bid can contain many more than 10 segments). For aggregation of Bids, the application follows the standard of stacking up Bid segments when Energy Prices are different while adding MWs if Energy Prices are the same.

The CAISO IFM performs Unit Commitment and Congestion Management, clears Virtual Bids submitted by SCs and clears the Energy Bids as modified in the MPM, taking into account transmission limits, inter-temporal and other operating constraints, and ensures that adequate Ancillary Services are procured in the CAISO Balancing Authority Area based on 100% of the CAISO Forecast of CAISO Demand.

After the CAISO DAM completes, the cleared Virtual Bid results will be de-aggregated at the eligible SC ID level before the Day-Ahead Market results, which include Virtual Awards, are published to Market Participants. For the de-aggregation of a non-marginal segment, it is straight forward to assign the individual cleared MW to the eligible SCID. For the marginal segment, the relevant MW cleared amount is associated with multiple bid segments and hence a prorating is needed to obtain the individual cleared MW amount at the SCID level. The CAISO prorates the awarded MWs proportional to the submitted MWs of the marginal segment of each Virtual Bid contributing to the marginal aggregate segment.

Demand Bids and Virtual Bids are not accepted in the RUC or the RTM. Virtual Bids are not considered in RUC, but they may influence the RUC outcome based on the amount of unit commitment, Virtual Awards, and physical schedules awarded in the IFM.

3.10.4 Virtual Bidding Settlements

Virtual Supply Awards are paid the Day-Ahead LMPs at their location and charged in Real-Time at the applicable FMM LMPs at the applicable PNodes or APNodes. Virtual Demand Awards are charged the Day-Ahead LMPs at their locations and paid in Real-Time at the applicable FMM LMPs at the applicable PNodes or APNodes.

It is possible that an excessive amount of Virtual Supply versus Virtual Demand is cleared in IFM, such that there is “virtual” overgeneration. Since RUC only runs with physical Bids and the CAISO Forecast of CAISO Demand, and to the extent that Virtual Supply has displaced physical Supply, RUC may need to commit more physical resources and/or more RUC capacity maybe awarded in order to make sure that there is enough physical capacity covering the CAISO Forecast of CAISO Demand.

3.11 Low Carbon Assets and Market Design

CAISO is the leader among the US ISO markets in modifying its market design to foster market participation of low carbon assets, such as Demand Response (DR), Renewable Energy Sources (RES), Baterry Energy Storage Sources (BESS), etc. In this Section we present various energy market design elements for these low carbon assets, specifically DR and BESS assets.

3.11.1 Demand Response

3.11.1.1 Demand Response Products

The CAISO offers three demand response products:

1. Proxy Demand Resource (PDR)
2. Reliability Demand Response Resource (RDRR)
3. Proxy Demand Resource – Load Shift Resource (PDR-LSR)

The CAISO developed the Proxy Demand Resource (PDR) product to increase demand response participation in the CAISO’s wholesale Energy and Ancillary Services markets. Additionally, PDR helps in facilitating the participation of existing retail demand response into these markets.

The CAISO developed the Reliability Demand Response Resource (RDRR) product to further increase demand response participation in the CAISO markets by facilitating the integration of existing emergency-triggered retail demand response programs and newly configured demand response resources that have reliability triggers and desire to be dispatched only under certain system conditions.

The CAISO developed the Proxy Demand Resource - Load Shift Resource (PDR-LSR) to recognize the ability for demand to consume during oversupply conditions, facilitating its ability to shift and shape load with market signals reflecting grid conditions.

In addition, RDRR enables the integration of California Regulator (CPUC)-jurisdictional emergency responsive demand response resource programs. RDRRs can economically bid and be dispatched in the DAM but only be dispatched for reliability in the RTM. RDRRs cannot offer or self-provide Ancillary Services or submit RUC availability bids.

PDR, RDRR, and PDR-LSR products each provide the capability for an aggregator of retail customers, working with a certified CAISO Demand Response Provider (and SC), to bid demand response on their behalf directly into the CAISO's organized markets to the extent permitted by applicable laws and regulations regarding retail customers. PDRs and RDRRs may elect to bid and be scheduled in five-, fifteen-, or sixty-minute intervals (i.e., in RTD, FMM, or HASP). If these resources do not make an election, the default will be 60 minutes. PDRs and RDRRs using the sixty-minute (hourly block) bidding option are ineligible for energy bid cost recovery. PDRs and RDRRs electing to use the five- or fifteen-minute bidding options must ensure their selection complies with operational and technical constraints in the CAISO Database (the Master File). The ISO may request documentation from resources selecting the five- or fifteen-minute to confirm that their resource has the ability to comply with five- or fifteen-minute dispatches, respectively.

Proxy Demand Resources using the load-shift methodology (PDR-LSR) must elect to bid and be dispatched in the RTM using either the five or fifteen minute interval bidding option. PDR-LSR does not have the sixty minute (hourly block) bidding option.

In general, the three (3) products is a combination of Load scheduled by a Load Serving Entity at the Default LAP and a Bid to curtail submitted by the Demand Response Provider (DRP) using a separate proxy generator with a distinct Resource ID.

The PDR-LSR product is modeled as a PDR consisting of registered location(s) that include at least one storage device (BESS). Under this model, the resource has the ability to bid and be dispatched for both load curtailment (discharging, generation), and load consumption (charging, negative generation) from the BTM storage.

A PDR, RDRR, and PDR-LSR are treated in the markets as a proxy generator bid as an aggregate generator, which may be defined at a single node or across multiple nodes within a CAISO defined Sub-LAP. The scheduling, dispatch, and settlement of the PDR, RDRR or PDR-LSR are as a proxy generator resource as a distinct Resource ID. The LSE base Load will be scheduled and settled at the Default LAP (DLAP). Settlements for energy provided by PDRs and RDRRs are based on the Demand Response Energy Measurement which is calculated using an approved Performance Evaluation Methodology. The Demand Response Energy Measurement applicable to use of the Performance Evaluation Methodology is the resulting Energy quantity calculated by comparing the Customer Baseline of a PDR, RDRR or PDR-LSR (Curtailment) against its actual underlying Load for a Demand Response Event.

A PDR, RDRR or PDR-LSR are separately measured behind the meter generation, utilize Meter Data consisting of its total gross consumption when using the Customer Load Baseline Methodology. The Demand Response Energy Measurement for a PDR or

RDRR consisting of registered behind-the-meter generation is the quantity of Energy equal to the difference between (i) the Energy output, and (ii) the Generator Output Baseline for the behind-the-meter generation registered in the PDR or RDRR, which derives from the Energy output of the behind-the-meter generation only, independent of offsetting facility Demand. For a PDR or RDRR using the combination of both methodologies, the Demand Response Energy Measurement is their independently derived Demand Response Energy Measurements' resulting sum.

3.11.1.2 Demand Response Market Design

The following summarizes the Proxy Demand Resource, Reliability Demand Response Resource or Proxy Demand Resource – Load Shift Resource product design attributes:

- A DRP may participate in the CAISO Markets separately from the LSE;
- The LSE and Utility Distribution Company (UDC) have the opportunity to review location information for a registration requested by a DRP;
- A PDR is eligible to participate in the Day-Ahead Energy market, Real-Time Energy market and Ancillary Services market to provide Spinning and Non-Spinning Reserves;
- A PDR and RDRR may elect the bid dispatchable options of 60, 15, or 5 minutes for the Real-time market;
- A RDRR is eligible to participate in the Day Ahead Energy market and Real-Time Energy market;
- A PDR-LSR participates under the PDR model providing functionalities for the BTM storage to bid and be dispatched for both load curtailment and load consumption. PDR-LSR is eligible to participate in the Day-Ahead and Real-Time Energy markets through the use of two separate Resource IDs for load curtailment (CUR), and load consumption (CON). The PDR-LSR (CUR) Resource ID is designed in the same fashion as the standard PDR product. The PDR-LSR (CON) Resource ID is designed in similar fashion as the existing NGR – Dispatchable Demand Resource (DDR) model. Both Resource IDs must be registered or updated in the CAISO MasterFile at the same time, and must bid in the same Market with a bid dispatchable option of either 15 or 5 minutes.
- Load curtailment (CUR) Resource ID is eligible to provide Resource Adequacy (RA), and may participate in the Ancillary Services (AS) market to provide Spinning and Non-Spinning Reserves.
- Load consumption (CON) Resource ID is ineligible for RA and provision of AS, and must bid from the bid floor up to a value less than $0.
- PDR and RDRR are load curtailment products. Performance for the resource will be measured in aggregate based on individual location load curtailment only and must not include measured export of energy from any of these individual locations;

- The CAISO does not prohibit net-energy metered (NEM) locations from participating in PDRs or RDRRs, however, meter data from NEM locations must only represent load or the resulting load offset when using the MGO methodology;
- The DRP's SC submits a PDR or RDRR bid to curtail Load, or PDR-LSR bids either to curtail or consume loads, and receive Automated Dispatch System (ADS) instructions as if it were a generator. The PDR or RDRR is bid and settled at a PNode (which could be a specific location or an aggregation of PNodes, and Settlement occurs directly between the CAISO and the DRP's Scheduling Coordinator;
- The LSE continues to forecast and schedule its total Load at the Default LAP;
- PDRs, RDRRs and PDR-LSRs consist of Residential End Users and Non-residential End Users.

3.11.2 Battery Energy Storage System (BESS) Market Participation

3.11.2.1 Overview

CAISO is a leader in the BESS penetration in the energy market. As such it has the most advanced BESS energy market design elements. The CAISO favors the Non-Generation Resource (NGR) model) which allow these resources to participate in the markets and provide services while accounting for their unique capabilities and characteristics. In this Section, we present the NGR model as the optimal way to encourage efficient and reliable market participation by BESS resources.

We also discuss various challenges regarding modeling and State of Charge (SOC) issues.

3.11.2.2 CAISO BESS Design Model

The increased penetration of BESS resources in the market requires the implementation of sophisticated models that can accommodate their participation in the market. NGR is a generic resource model where a resource can produce or consume energy within a continuous operating range that can span both generation and demand. This model can be used for demand response, when the operating range spans demand only, or for a hybrid plant than can operate continuously in charging and discharging modes. Figure 5-1 shows a typical energy bid from a BESS plant.

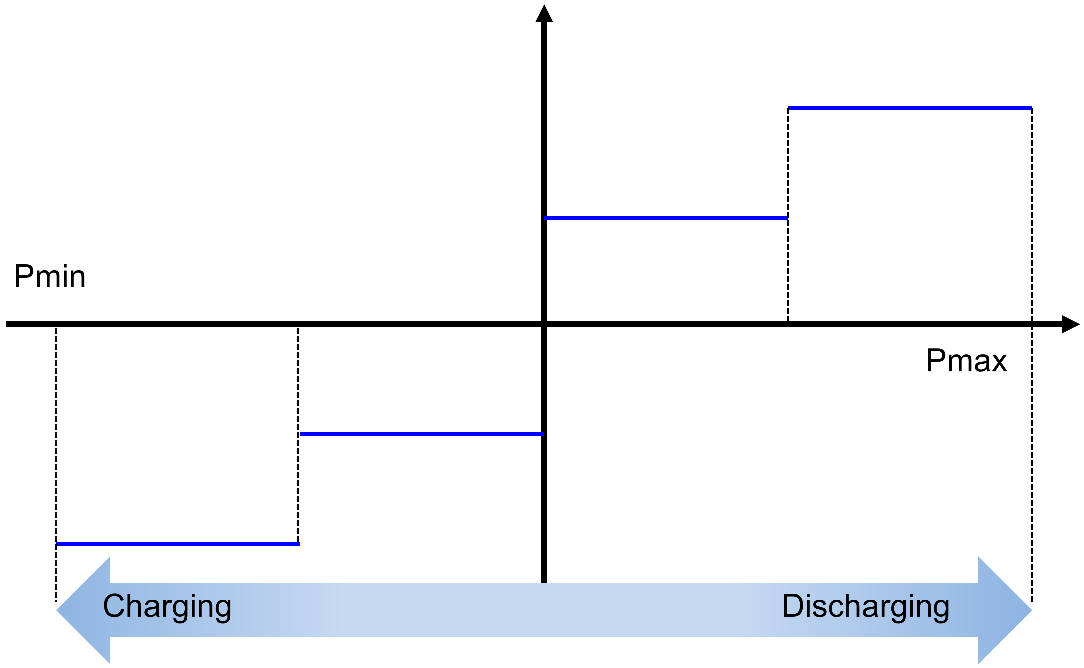


Figure 3-11: BESS Energy Bid

With the NGR model, a BESS plant can participate in the following Ancillary Services markets:

Regulation Down;

Regulation Up;

Spinning Reserve;

Non-Spinning Reserve

Flexible Ramping Up; and

Flexible Ramping Down

BESS plants are uniquely qualified to provide all these ancillary services, and especially regulation, because they possess superior ramp rates.

Since the availability of hybrid plants for ancillary services depends on the SOC, special constraints must be enforced in the market solution to only award ancillary services that are deliverable based on the SOC. These constraints impose additional telemetry requirements for the BESS plants. Specifically, the SOC must be telemetered to the CAISO along with the power output.

In the general NGR model, the Market Participant is responsible for managing the SOC for a BESS plant by strategic participation in the market. For example, submitting a self-schedule and/or an energy bid below 0MW for charging, or above 0MW for discharging.

It is important also to note that for BESS plants, the key question for the market participant is whether to interface with the market as a single resource (hybrid configuration or as two or more separate resources (co-located plant). In the co-located model, where each individual technology is treated separately, with the only connection being an injection limit at the point of interconnection. In the hybrid model, the information being exchanged with the CAISO is about the hybrid and not any of its individual components.

3.11.2.3 DAM BESS SOC Management

The CAISO NGR is subject to capacity and operational range limits as a generator, with adjustments to account for their unique operational characteristics. Unlike most traditional generators, NGRs are able to withdraw energy from the grid to charge, and have a limited energy storage capacity. The operating range of an NGR can be negative to account for the ability to withdraw energy from the grid to charge. The NGR model will take into account the resource's charging efficiency when it is withdrawing energy from the grid. The charging efficiency is the percentage of charging energy that, after losses, is ultimately available for generation.

The ability of an NGR to provide energy and ancillary services depend on the NGR's stored energy level, or state of charge (SOC). An NGR is subject to the same requirements as generators for providing energy and Ancillary Services, with additional constraints to manage SOC. For example, the operating reserve capability of an NGR is limited to what the resource can provide in 10 minutes, similar to a generator, while it is also limited by the SOC.

3.11.2.4 RTM BESS SOC Management

For NGRs designated as Limited Energy Storage Resources (LESRs), state of charge (SOC) constraints are applied to both the binding and non-binding intervals in FMM and RTD based on their Master File parameters, Lower and Upper Charge Limit bids, End-of-Hour (EOH) SOC bids limits, and, if applicable, the reliability-induced Minimum SOC.

The CAISO to properly manage the SOC for each interval, RTED receives the latest SOC for each NGR via telemetry and uses this information to calculate an initial condition SOC for each NGR, similar to the way generator initial operating levels are calculated, by projecting the actual status toward the last dispatch.

In RTD, the SOC remaining at the end of the RTD time horizon is constrained to ensure the LESR is able to meet its self-schedules in intervals beyond the scope of the RTD time horizon. The CAISO requires that resources scheduled to provide Spinning Reserve and Non-Spinning Reserve must be capable of maintaining that output for at least 30 minutes from the point at which the resource reaches its award capacity. Resources offering Regulation in the real-time market must also have the capability to meet a continuous energy requirement for at least 30 minutes.

Consistent with these requirements, when an NGR receives a Spinning Reserve or Non-Spinning reserve award, the CAISO reserves its SOC to ensure the NGR can continuously deliver that capacity for 30 minutes. When an NGR receives a Regulation award or qualified self-provision, the CAISO reserves its SOC to ensure it can continuously deliver that capacity for 30 minutes in the applicable fifteen-minute market and RTD interval.

Energy schedules and ancillary service awards are co-optimized in FMM subject to these SOC constraints. The ancillary service awards are fixed in RTD; however, the SOC constraints are still enforced in RTD to constrain the energy dispatch so that the sustained energy requirement is satisfied, and to ensure future awards and self-schedules can be met.

Further, SCs for LESRs may submit optional end-of-hour (EOH) state-of-charge (SOC) limits in the real-time market, as part of their energy bids, to manage the optimal use of their resources throughout the day. These EOH SOC limits are different from the minimum and maximum daily SOC limits, which ensure that resources receive dispatches that respect such daily SOC limits at the end of every market interval. Instead, the market may dispatch LESRs to meet imbalance energy needs within the hour, while ensuring the resource's SOC satisfies the EOH limits only at the end of the corresponding hour, if submitted as part of the energy bids.

The SC submits a bid with minimum and maximum EOH SOC limits in MWh to reflect a target range. If the SC desires a single target EOH SOC, then the minimum and maximum EOH SOC limits may be set the same. The RTM will respect an LESR's minimum and maximum EOH SOC limits. The market ignores EOH SOC limits if they conflict with the resource's daily minimum and maximum SOC limits. However, there are bid validation rules that limit these occurrences in the market.

3.11.3 Flexible Ramp Product

The CAISO implemented a new market feature known as the flexible ramping product on November 1, 2016 to enhance reliability and market performance by procuring flexible ramping capacity in the real-time market. This flexible ramping capacity can be utilized to help manage volatility and uncertainty of real-time imbalance demand due to massive penetration of RES energy amd load uncertainty.

This uncertainty stems from the amount of net load that exists in the real-time market. Net load refers to the difference between system loads minus output from wind and solar generation. This represents the portion of load that needs to be met by other sources of energy, including dispatchable gas-fired generation that is used to balance changes in intermittent sources of renewable energy and other factors affecting demand for real-time energy. When the real-time market software runs the optimization for the current binding interval, the net load values in future advisory intervals are not known. The market software uses a forecast for the net load in these future advisory intervals. However, the net load in those future intervals may differ from this advisory forecast. Therefore, flexible ramping capacity is needed not only to make the advisory interval's forecasted net load feasible (or maintain power balance) but also to make a larger range of potential net loads in the advisory interval feasible.

In other words, the flexible ramping product is designed to ensure a margin of sufficient ramping capacity beyond the forecasted ramping needs to protect against power balance violations. As more wind and solar capacity is added to the system, this is increasing the need for such flexible capacity to manage any increase in net load volatility and reduce the frequency of power balance constraint relaxations.

A key component of the flexible ramping product market design is that the amount of flexible capacity that the product procures is derived from a demand curve which reflects a calculation of the optimal willingness-to-pay for that flexible capacity. The demand curves allow the market optimization to consider the trade-off between the cost of procuring additional flexible ramping capacity and the expected reduction in power balance violation costs. T his approach increases market efficiency by signaling an explicit trade-off between the costs and benefits of procuring more or less flexible ramping capacity for a given interval.

The demand curves for the flexible ramping product are intended to be calculated based on the expected ramp needed to meet scheduled loads, as well as the uncertainty surrounding ramping needs. This uncertainty was intended to be calculated using historical data on the error surrounding forecasted ramping needs. The demand curves are calculated independently for each hour of the day, and differ by market (15-minute and 5-minute) and direction (upward ramping and downward ramping). There are separate demand curves calculated for each energy imbalance market area in addition to a system-level demand curve. These demand curves are intended to be calculated from historical net load forecast error data.

The flexible ramping product is incorporated into the ISO’s market optimization as a “soft” constraint which can be relaxed depending on the cost of meeting the constraint. With this approach, the demand curves are first entered into the market software as segments of relaxation capacity that reflect the expected cost of a power balance constraint violation for the level of foregone capacity procurement. The maximum amount of capacity on the demand curve (or uncertainty) is then treated as a requirement and is met in every interval through a combination of flexible ramping capacity procurement or relaxation capacity.

The flexible ramping product procures both upward and downward flexible capacity, in both the 15-minute and 5-minute markets. Procurement in the 15-minute market is intended to ensure that enough ramping capacity is available to meet the needs of both the upcoming 15-minute market runs and the three 5-minute market runs with that 15-minute interval. Procurement in the 5-minute market is aimed at ensuring that enough ramping capacity is available to manage differences between consecutive 5-minute market intervals.

4 ERCOT

4.1 General Overview

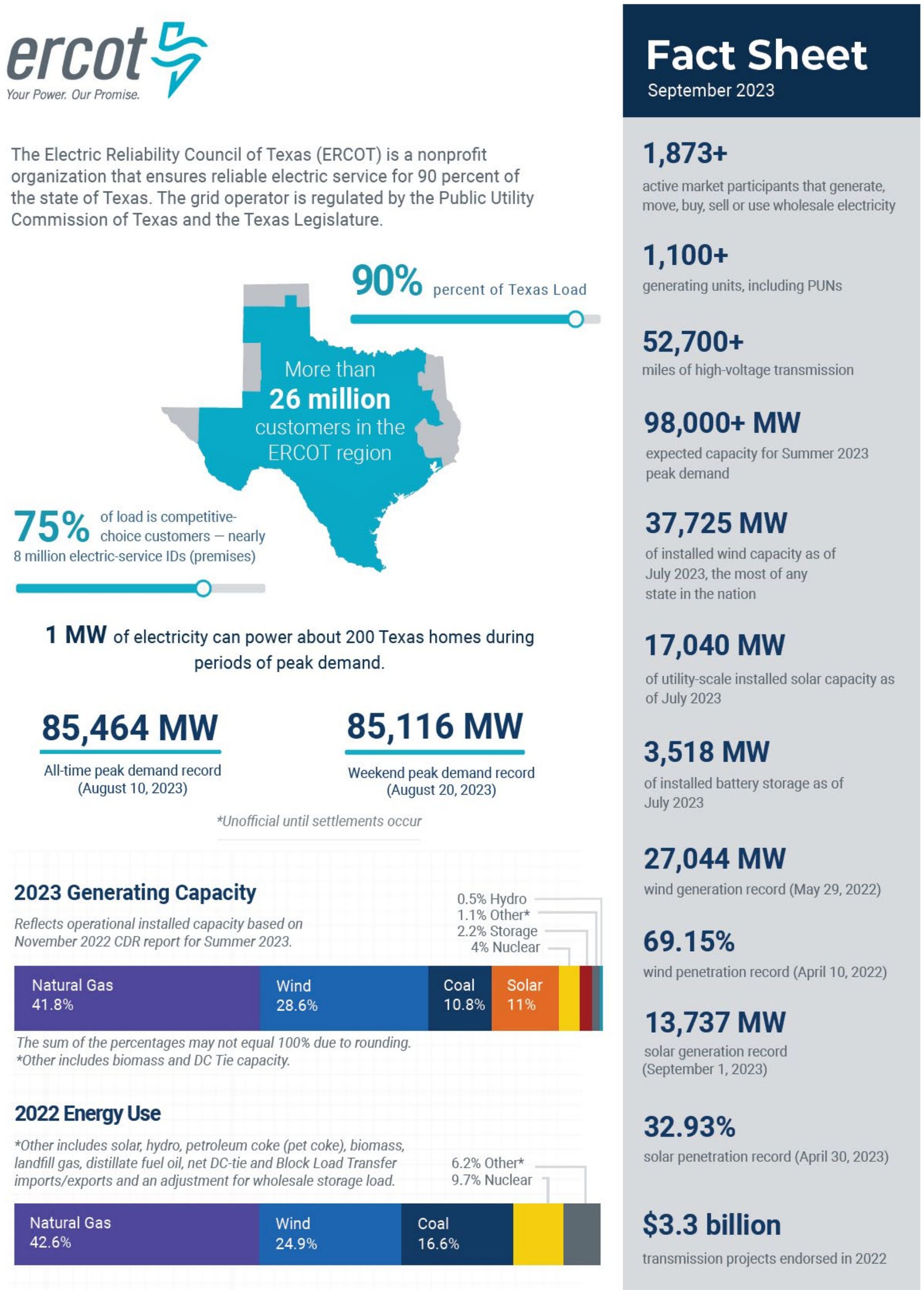


Figure 4-1: ERCOT Fact Sheet

(Source: www.ercot.com)

Electric Reliability Council of Texas, Inc. (ERCOT), is a non-profit corporation that manages the flow of electric power to more than 26 million Texas customers who are located within the Texas Interconnection, which along with the Eastern and Western interconnections, comprise the three synchronously connected electric grids in the continental United States These three interconnections are not connected synchronously, but do have the ability to transfer relatively small amounts of power over DC ties. The Texas Interconnection is located entirely within the State of Texas and covers over 75% of the State's geography and represents about 90 percent of the state's total electric consumption. The region does not include the El Paso area in far West Texas, which is within the Western Interconnection, nor the northern Panhandle area in the Eastern Interconnection (except for some transmission facilities built to move renewable generation from that region into the population centers of Texas located in the ERCOT region), nor does it include some relatively small areas around North East and South East Texas, which are also located in the Eastern Interconnection. Figure 4-2 shows an exploded map of the Western Interconnection, ERCOT, and Eastern Interconnection. While Figure 4-1 above shows a map of Texas with the footprint of ERCOT and basic facts regarding the ERCOT system.

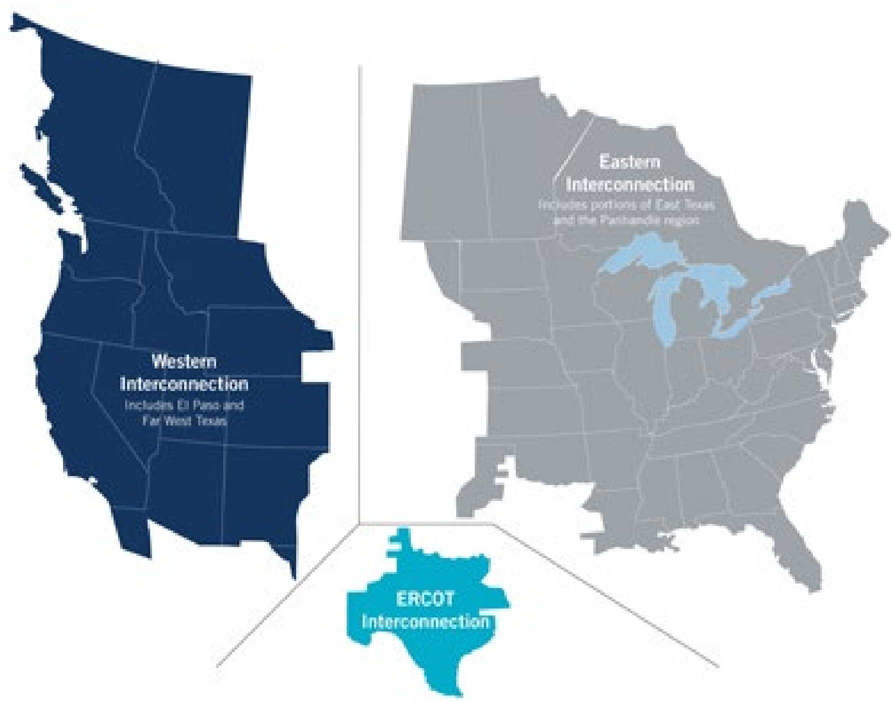


Figure 4-2: Map of the Western Interconnection, ERCOT, and Eastern Interconnection

(Source: www.ercot.com)

The origins of the Texas Interconnection lay in the Second World War when, in 1941, a group of Texas electric utilities joined together as the Texas Interconnected System (TIS) to support the war effort. TIS sent excess power supplies to industrial manufacturing companies on the Texas Gulf Coast to provide power for energy-intensive production processes. Recognizing the reliability advantages of remaining interconnected, the TIS members continued to use and develop the interconnected Texas grid, including agreeing not to interconnect or sell power outside of the interconnection in order to avoid becoming subject to the Federal Power Act, or the jurisdiction of the Federal Power Commission, the predecessor of the Federal Energy Regulatory Commission (FERC). For this reason ERCOT and the Texas Interconnection have avoided becoming subject to FERC jurisdiction under the Federal Power Act except for certain reliability matters relating arising from the Energy Policy Act of 2005 and for wholesale sales to and from ERCOT over its asynchronous ties with the Eastern Interconnection. TIS members adopted official operating guides for their interconnected power system and established two monitoring centers within the control centers of two utilities, one in North Texas and one in South Texas. TIS formed ERCOT in 1970, to participate with the North American Reliability Council as utilities nationally increased coordination following the Northeastern blackout of 1965. ERCOT continued to evolve over the ensuing decades to take on more responsibilities and roles.

In 1995, the Texas Legislature acted to deregulate the wholesale generation market within the Texas Interconnection and the PUCT began the process of expanding ERCOT's responsibilities and capabilities to enable wholesale competition and facilitate efficient use of the power grid by all market participants. In August of the following year the PUCT adopted an electric utility joint task force recommendation that ERCOT become an independent system operator (ISO) to ensure that an impartial, third-party organization was overseeing equitable access to the power grid by competitive market participants. About a month later, this change was implemented officially, when the ERCOT Board of Directors restructured its organization and initiated operations as a not-for-profit ISO, making it one of the first electric ISOs in the United States. Three years later, Texas Legislature completed its restructuring of the Texas Interconnection by enacting Senate Bill 7 which required investor-owned utilities (IOU) to unbundle their functions (generation, delivery and retail sales of electricity) and required by January 1, 2002, the creation of a competitive retail electricity market to give customers the ability to choose their retail electric providers. Senate Bill 7 also gave the PUCT authority to oversee ERCOT and to develop rules to protect competition in the wholesale market and consumers in the retail market. Over time the Texas legislature has continued to give the PUCT increased authority over ERCOT and its operations, including the ability to confirm ERCOT’s independent directors and approve its budget.

From 1999 to 2000, ERCOT sponsored a stakeholder process to address how ERCOT would administer its responsibilities to support the competitive retail and wholesale electricity markets while maintaining the reliability of electric services. In thousands of hours of meetings and mark-up sessions, the market participants worked together to develop new ERCOT protocols, which, when approved by the PUCT, are enforceable rules and standards for implementing market functions regarding: energy scheduling and dispatch, ancillary services, congestion management, outage coordination, settlement and billing, metering, data acquisition and aggregation, market information systems, transmission and distribution losses, renewable energy credit trading, registration and qualification, market data collection, load profiling, and alternative dispute resolution.

Historically, there had been 10 interconnected control areas in ERCOT. Each corresponding control area operator had been responsible for balancing supply and demand within its area in cooperation with the other areas. At the end of July 2001, the existing 10 control areas in the ERCOT region were consolidated into a single control area (now called a "balancing area" under North American Reliability Corporation (NERC) terminology). Wholesale power sales among parties began to operate under the new electric industry restructuring guidelines, including centralization of power scheduling and procurement of ancillary services to ensure reliability. Commercial functions were centralized to facilitate efficient market operations, including meter data acquisition and aggregation, load profiling and statewide registration of retail premises to facilitate switching by customers between competitive retail electricity providers (REP). On January 1, 2002, ERCOT and its market participants launched the competitive retail electric market, under the auspices of the PUCT, allowing individuals and businesses in areas previously served by incumbent IOUs to choose power suppliers. The 1999 restructuring law applied to IOUs, but allowed public power entities, municipally owned utilities and electric cooperatives, then approximately 24 percent of the ERCOT load, to decide if they wanted to participate in retail competition.

A day-ahead "scheduling" process was established in 2001 where market participants provided matched generation and consumption information for each 15-minute interval in the following day and made offers to provide ancillary services. After initial operation of the ERCOT wholesale market without any representation of transmission limits, ERCOT was divided in 2002 into four zones for the purposes of dispatching and pricing power purchases from generators and sales to retail customers. Supply was specified based on portfolios of generation in each zone and demand was also specified zonally. A "balancing market" then sought bids and offers to deviate from the day-ahead schedules.

It soon became evident that a system of zonal portfolio dispatch was inefficient in maintaining grid reliability and expensive for wholesale market participants. As a result, in September 2003, the PUCT ordered ERCOT to develop a nodal wholesale market design, with the goal of improving market and operating efficiencies through more granular pricing and scheduling of energy services. The ERCOT stakeholder process worked from 2003 through the middle of 2005 on developing nodal protocols and in September of that year ERCOT submitted draft nodal protocols to the PUCT. In April 2006 the PUCT issued an order approving the stakeholder-developed protocols for the nodal market that was eventually launched on December 1, 2010. The implemented market design included locational marginal pricing for generation, a day-ahead energy and ancillary services co-optimized market, day-ahead and hourly reliability-unit commitment, and congestion revenue rights. The real-time market prices and dispatches generation in 5-minute periods and settles in 15-minute increments. Instead of zonal portfolio offers and bids as in the previous market, the nodal market required offers specifically from each generating unit.

As ISO for the region since 1996, ERCOT has managed the open-access regime and today schedules power on an electric grid that connects more than 52.700 miles of transmission lines owned by transmission-only service providers, transmission and distribution-only IOUs, municipally owned utilities, and electric cooperatives with over 1100 generation units in order to deliver electricity to multiple public entities and private companies for distribution to their retail customers. In addition, ERCOT performs financial settlement for the competitive wholesale bulk-power market and administers retail switching for nearly eight million premises in competitive choice areas of Texas, which represents approximately 75% of ERCOT's electric load. As of July 2023, ERCOT had over 98,000 MWs of resources available to serve electric consumers, including **37,725 MW of installed wind capacity, 17,040 MW of installed utility scale solar capacity and 3518 MW Batttery Storage.**

4.1.1 Retail Choice

A unique feature of the ERCOT electricity reform has been the implementation of retail choice that allowed independent Retail Energy Providers (REPs) to compete with the incumbent utilities in selling electricity over the distribution system to retail customers. About 60 REPS have been operating in the ERCOT territory. The implementation of retail choice has been a primary driver in the ERCOT market design. For instance the implementation of an energy only market and rejection of a capacity market mechanism was largely influence by the REPs who view capacity markets that facilitate capped energy price, as public price insurance that undercuts the price hedging function provided to consumers by the REPs in an uncapped (or highly capped) energy only market.

In the law that restructured the power industry in ERCOT in 1999 , the Texas Legislature explicitly wanted to facilitate full competition across all retail customer classes: large commercial and industrial (C&I), small commercial, and residential. The start date for retail choice was set for January 1, 2002. However, Texas Legislature decided that the market for residential and small commercial customers (also known as mass market customers) should have a five-year transition before full retail choice was implemented. Nevertheless, it was intended that the benefits of restructuring would accrue eventually to all market segments and not just to large customers.

The Texas retail market design had some uncommon or unique features compared to restructuring in other parts of the United States and internationally. For instance, competitive retailers were given responsibility for customer billing rather than giving responsibility to the wires companies that remained rate and service regulated, so customers would more readily see who their retail agents were. Also, residential and small commercial customers were not given the easy fallback of "basic or standard offer" or "default service" from regulated wires companies, which also encouraged mass market customers to actively shop for retail electric service rather rely on a regulated default service provider.

Instead of a default service, the Texas Legislature chose to use a "Price-to-Beat" mechanism, which put a price floor on the offerings to mass market customers from the retail affiliate of the incumbent unbundled utility, allowing competing REPs to undercut the incumbent. At the start of the retail market in 2002, the previously regulated prices had been high enough to permit all customers to immediately benefit from the Price-to-Beat, while the Price-to-Beat was nevertheless at a high enough level to enable profitable entry by new REPs. The Price-to-Beat was also indexed to the cost of natural gas to maintain headroom for competitive retailers if wholesale power prices rose as a result in increases in the cost of natural gas (a fuel for a significant portion of the ERCOT generation fleet).

Another ongoing goal of the Texas Legislature was to have a set of policies to maintain the ease of entry into both wholesale and retail markets, which provided consumers with an ample choice of REPs and REPs with an ample choice of wholesale suppliers. For instance, (i) the use of ERCOT as centralized switching registration agent, and (ii) having simple, straightforward registration requirements for power generation companies (PGCs) at the PUCT, both facilitated entry of independent REPs. Furthermore, to help spur supply competition and ensure that prices were reasonably fair and not the product of market power abuse, the restructuring law limited any one PGC or their affiliates from owning and controlling more than 20% of the installed generation capacity within ERCOT.

After restructuring, the Texas Legislature encouraged the deployment and use of smart meters, and the PUCT adopted rules to facilitate their installation by the regulated wires companies. Currently, ERCOT has achieved nearly system-wide installation of smart meters (over 98 percent of all energy settled in areas with retail choice). As a result, customers can switch retailers faster (often within a day) and more easily than in many other jurisdictions. As part of the system-wide deployment of smart meters, PUCT required ERCOT to settle smart meters in 15-minute increments, which facilitated new products and services for residential and small commercial customers, among which are voluntary aggregated load reduction programs administered by REPs. The smart meter deployment has also enabled REPs to offer these customer classes more varied and flexible products to manage their energy use based upon time of consumption, such as "free nights or weekends" (with a related increase in the price during other hours of the day) and even products that provide for the real-time wholesale settlement price to be passed through to the retail customer with no mark-up other than a small monthly charge together with the TDU charge. However, offering unmitigated exposure by retail customers to wholesale rates (by a REP called Griddy) has been banned after the "Big Freeze" of February 2020.

Strong political mandates by state legislators, who are responsive to voters who are also consumers, ensured that prices were as fair (not the product of market power abuse) and uniform across the relevant geographic footprints as much as reasonably possible. The PUCT has responded by making it a consistent policy to remove import constraints into large load centers and to make the transmission network more robust in the South and West Load Zones.

4.2 The ERCOT Nodal Protocols

This section describes the structure of the ERCOT market and its elements. ERCOT is the Control Area Operator (CAO) for the ERCOT interconnection and performs all Control Area functions as defined in the Operating Guides and the North American Electric Reliability Corporation (NERC) policies. ERCOT procures Ancillary Services to ensure the reliability of the ERCOT System.

(ERCOT is the central counterparty for all transactions settled by ERCOT pursuant to these Protocols and is deemed to be the sole buyer to each seller, and the sole seller to each buyer, of all energy, Ancillary Services, Reliability Unit Commitments (RUCs), Emergency Response Service (ERS), and other products or services for which ERCOT may pay or charge a Market Participant, except for those products or services procured through bilateral transactions between Market Participants and those products or services that are self-arranged by Market Participants. ERCOT may not profit financially from its activities as the Independent Organization in the ERCOT Region. ERCOT may not use its discretion in the procurement of Ancillary Service capacity or deployment of energy to influence, set or control prices

4.2.1 Day-Ahead Operations

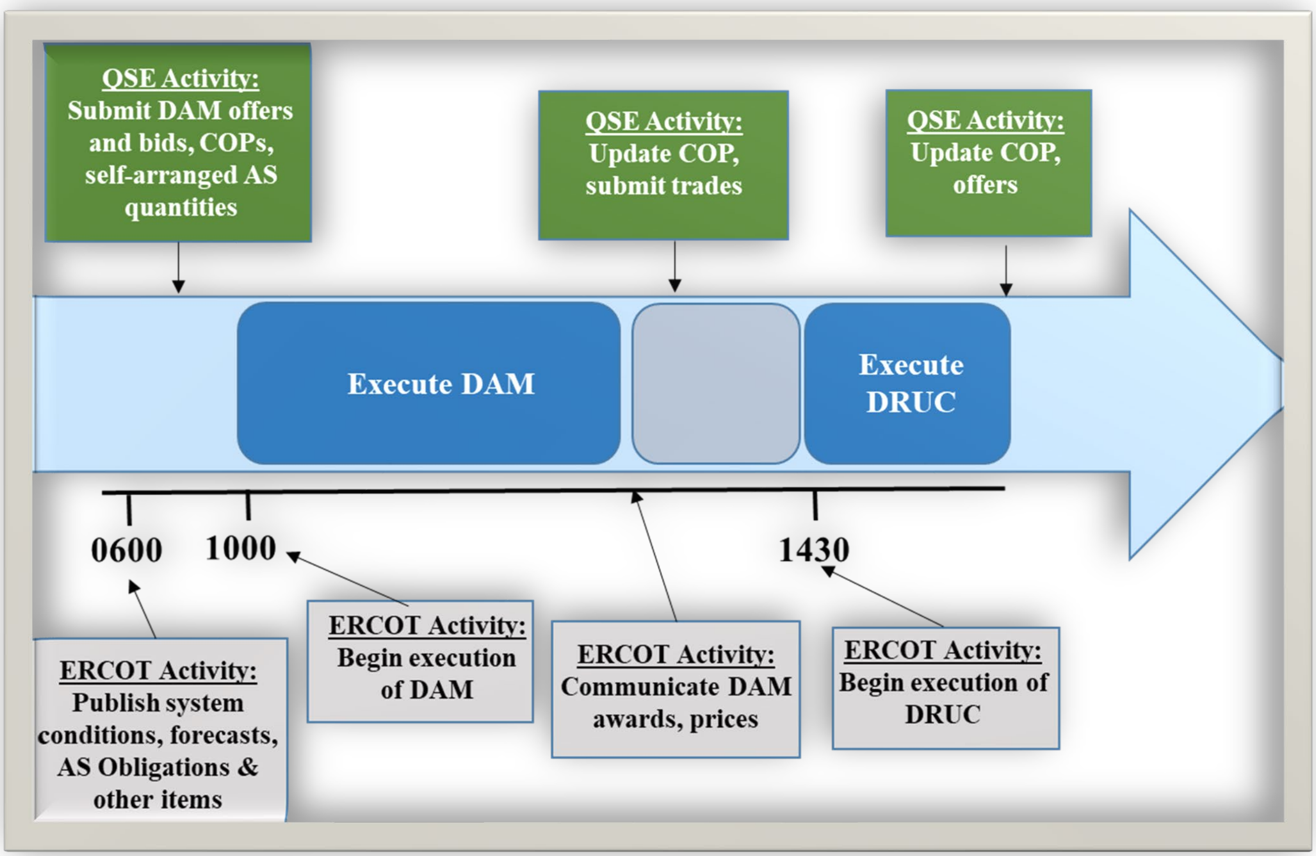


Figure 4-3 Day Ahead Operations

Source: https://www.ercot.com/mktrules/nprotocols/current

The Day-Ahead Market (DAM) is a daily, co-optimized market in the Day-Ahead for Ancillary Service capacity and forward financial energy and congestion transactions. DAM energy settlements use DAM Settlement Point Prices that are calculated for Resource Nodes, Load Zones, and Hubs for a one-hour Settlement Interval using the Locational Marginal Prices (LMPs) from DAM. In contrast, the Real-Time energy settlements use Real-Time Settlement Point Prices that are calculated for Resource Nodes, Load Zones, and Hubs for a 15-minute Settlement Interval. Figure 4-3 illustrates the time line for various activities performed in the Day Ahead Market by Qualified Service Entities (QSE) and by ERCOT.

ERCOT shall analyze the expected Load conditions for the Operating Day and develop an Ancillary Service Plan that identifies the Ancillary Service MW necessary for each hour of the Operating Day. The MW of each Ancillary Service required may vary from hour to hour depending on ERCOT System conditions. ERCOT must post the Ancillary Service Plan to the ERCOT website by 0600 of the Day-Ahead. If ERCOT determines that an Emergency Condition may exist that would adversely affect ERCOT System reliability, it may change the percentage of Load Resources that are allowed to provide ERCOT Contingency Reserve Service (ECRS) and Responsive Reserve (RRS) from the monthly amounts determined previously

ERCOT shall assign part of the Ancillary Service Plan quantity, or total Ancillary Service procurement quantity, if different, by service, by hour, to each Qualified Scheduling Entity (QSE) based on its Load Serving Entity (LSE) Load Ratio Shares (LRSs) (including the shares for Direct Current Tie (DC Tie) exports) aggregated by hour to the QSE level. QSE shares will be adjusted on a pro rata basis such that the sum of all shares is equal to one. The resulting Ancillary Service quantity for each QSE, by service, by hour, is called its Ancillary Service Obligation. ERCOT shall base the QSE Ancillary Service allocation on the QSE to LSE relationships for the operating date and on the hourly LSE LRSs from the Real-Time Market (RTM) data used for Initial Settlement for the same hour and day of the week, for the most recent day for which Initial Settlement data is available, multiplied by the quantity of that service required in the Day-Ahead Ancillary Service Plan. By 0600 of the Day-Ahead, ERCOT shall notify each QSE of its advisory Ancillary Service Obligation for each service and for each hour of the Operating Day, based on the Ancillary Service Plan, as well as that QSE's proportional limit for any Self-Arranged Ancillary Services

4.2.1.1 Wind-Powered Generation Resource Production Potential

ERCOT shall produce and update hourly a Short-Term Wind Power Forecast (STWPF) that provides a rolling 168-hour hourly forecast of wind production potential for each Wind-powered Generation Resource (WGR) and for each wind generation component of a DC-Coupled Resource. ERCOT shall produce and post to the ERCOT website every five minutes an Intra-Hour Wind Power Forecast (IHWPF) by wind region that provides a forecast of ERCOT-wide wind production potential for each five-minute interval over the next two hours from each forecast model. The posting shall indicate which forecast model was being used by ERCOT for Generation To Be Dispatched (GTBD) calculation purposes. ERCOT shall produce and update an hourly Total ERCOT Wind Power Forecast (TEWPF) providing a probability distribution of the hourly production potential from all wind-power in ERCOT for each of the next 168 hours. A Resource Entity with a WGR or DC-Coupled Resource that has a wind generation component shall install equipment to enable telemetry of site-specific meteorological information that ERCOT determines is necessary to produce the STWPF and TEWPF forecasts, and the Resource Entity's QSE shall telemeter such information and Resource status information to ERCOT. ERCOT shall establish procedures specifying the accuracy requirements of meteorological information telemetry for WGRs and DC-Coupled Resources with a wind generation component.

4.2.1.2 PhotoVoltaic Generation Resource Production Potential

ERCOT shall produce and update hourly a Short-Term PhotoVoltaic Power Forecast (STPPF) that provides a rolling 168-hour hourly forecast of PhotoVoltaic (PV) production potential for each PhotoVoltaic Generation Resource (PVGR) and for the PV component of each DC-Coupled Resource. ERCOT shall produce and post to the ERCOT website every five minutes an Intra-Hour PhotoVoltaic Power Forecast (IHPPF) by PhotoVoltaic region that provides a forecast of ERCOT-wide PhotoVoltaic production potential for each five-minute interval over the next two hours from each forecast model. The posting shall indicate which forecast model was being used by ERCOT for GTBD calculation purposes. ERCOT shall produce and update an hourly Total ERCOT PhotoVoltaic Power Forecast (TEPPF) providing a probability distribution of the hourly production potential from all PhotoVoltaic Generation Resources and the PV components of all DC-Coupled Resources in ERCOT for each of the next 168 hours. A Resource Entity with a PVGR or DC-Coupled Resource that has a PV component shall install equipment to enable telemetry of site-specific meteorological information that ERCOT determines is necessary to produce the STPPF and TEPPF forecasts, and the Resource Entity's QSE shall telemeter such information and Resource status information to ERCOT. ERCOT shall establish procedures specifying the accuracy requirements of meteorological information telemetry for PVGRs and DC-Coupled Resources with a PV component.

4.2.2 QSE Activities and Responsibilities in the Day-Ahead

During the Day-Ahead, a Qualified Scheduling Entity (QSE) Must submit its Current Operating Plan (COP). It may submit Three-Part Supply Offers, Day-Ahead Market (DAM) Energy-Only Offers, DAM Energy Bids, Energy Bid/Offer Curves, Energy Trades, Self-Schedules, Capacity Trades, Direct Current Tie (DC Tie) Schedules, Resource-Specific Ancillary Service Offers, DAM Ancillary Service Only Offers, Ancillary Service Trades, Self-Arranged Ancillary Service Quantities, and Point-to-Point (PTP) Obligation bids as specified in this Section.

By 0600 in the Day-Ahead, each QSE representing Reliability Must-Run (RMR) Units, Firm Fuel Supply Service (FFSS) Resources (FFSSRs), or Black Start Resources shall submit its Availability Plan to ERCOT indicating availability of RMR Units, FFSSRs, and Black Start Resources for the Operating Day and any other information that ERCOT may need to evaluate use of the units.

A Capacity Trade is the information for a Qualified Scheduling Entity (QSE)-to-QSE transaction that transfers financial responsibility for capacity between a buyer and a seller. Such trades for hours in the Operating Day that is reported to ERCOT before 1430 in the Day-Ahead are accounted for by ERCOT in the Day-Ahead Reliability Unit Commitment (DRUC). A Capacity Trade submitted at or after 1430 in the Day-Ahead for the Operating Day is accounted for in any Hourly Reliability Unit Commitment (HRUC) processes executed after the Capacity Trade is reported to ERCOT.

An Energy Trade is the information for a QSE-to-QSE transaction that transfers financial responsibility for energy at a Settlement Point between a buyer and a seller. Energy Trade for hours in the Operating Day that is reported to ERCOT before 1430 in the Day-Ahead creates a capacity supply or obligation in the DRUC process. Energy Trades submitted after 1430 in the Day-Ahead for the Operating Day create a capacity supply or obligation in any HRUC processes executed after the Energy Trade is reported to ERCOT. Energy Trade may be submitted for any Settlement Interval within an Operating Day before 1430 of the following day.

A Self-Schedule is the information that a QSE submits for Real-Time Settlement that specifies the amount of the QSE's energy supply at a specified source Settlement Point to be used to meet the QSE's energy obligation at a specified sink Settlement Point. A Self-Schedule may be submitted for any Settlement Interval before the end of the Adjustment Period for that Settlement Interval. As soon as practicable, ERCOT shall notify the QSE through the Messaging System of any of its Self-Schedules that are invalid Self-Schedules. The QSE may correct and resubmit any invalid Self-Schedule within the appropriate market timeline.

4.2.2.1 Self-Arranged Ancillary Service Quantities and Ancilary service offers

For each Ancillary Service, a QSE may self-arrange all or a portion of the advisory Ancillary Service Obligation allocated to it by ERCOT, subject to the QSE's share of system-wide limits as established by the standards for Determining Ancillary Service Quantities. If a QSE elects to self-arrange Ancillary Service capacity, then ERCOT shall not pay the QSE for the Self-Arranged Ancillary Service Quantities for the portion that meets its final Ancillary Service Obligation; ERCOT shall pay the QSE the respective Day-Ahead Ancillary Service price for any Self-Arranged Ancillary Service Quantities that exceed a QSE's final Ancillary Service Obligation. A QSE may submit a negative Self-Arranged Ancillary Service Quantity in the DAM. ERCOT shall procure all negative Self-Arranged Ancillary Service Quantities submitted by a QSE. Such negative Self-Arranged Ancillary Service Quantities will be considered by DAM to be equivalent to a bid to buy Ancillary Services at the highest price.

By 1000 in the Day-Ahead, a QSE may submit Resource-Specific Ancillary Service Offers from Generation Resources and ESRs to ERCOT for the DAM and may offer the same Generation Resource or ESR capacity for any or all of the Ancillary Service products simultaneously with any Energy Offer Curves from that Generation Resource or Energy Bid/Offer Curves from that ESR in the DAM. Offers of more than one Ancillary Service product from one Generation Resource or an ESR may be inclusive or exclusive of each other and of any Energy Offer Curves, as specified according to a procedure developed by ERCOT.

By 1000 in the Day-Ahead, a QSE may submit Load Resource-Specific Ancillary Service Offers for Regulation Service, Non-Spin, RRS, and ECRS to ERCOT and may offer the same Load Resource capacity for any or all of those Ancillary Service products simultaneously. Offers of more than one Ancillary Service product from one Load Resource may be inclusive or exclusive of each other, as specified according to a procedure developed by ERCOT

Ancilary services may also be traded among QSEs and reported to ERCOT subject to specif rules and restrictions.

4.2.3 DRUC HRUC and RMR Offers

ERCOT shall decide, in its sole discretion, to commit a Reliability Must-Run (RMR) Unit using the DRUC or HRUC process only when it has determined that the RMR Unit is likely to be needed in Real-Time for reliability reasons, taking into consideration whether SCED will solve transmission constraints without the RMR Resource, contractual constraints on the Resource, and any other adverse effects on the RMR Unit that may occur as the result of the dispatch of the RMR Resource.

If ERCOT has determined that an RMR Unit will be needed in Real-Time to resolve a transmission constraint, then ERCOT shall manually commit the Resource for the capacity required to resolve the transmission constraint using the DRUC or HRUC process. ERCOT may submit Energy Offer Curves at the effective Value of Lost Load (VOLL) in $/MWh on behalf of RMR Units committed in the DRUC or HRUC, and subsequently available for Dispatch by SCED, unless ERCOT declares a Market Suspension, in which case no Energy Offer Curves will be submitted, and ERCOT may, at its discretion, Dispatch RMR Units to restore the ERCOT Transmission Grid.

4.2.4 Energy Offers and Bids

4.2.4.1 Three-Part Supply Offers

A Three-Part Supply Offer consists of a Startup Offer, a Minimum-Energy Offer, and an Energy Offer Curve. **ERCOT must validate** each Startup Offer and Minimum-Energy Offer and the Energy Offer Curve before it can be used in any ERCOT process. In the absence of such validation ERCOT uses **generic offer caps** resource category. The DAM uses all three parts of the Three-Part Supply Offer and also uses Energy Offer

Curves submitted without a Startup Offer and without a Minimum-Energy Offer. The RUC only uses the Startup Offer and the Minimum-Energy Offer components for determining RUC commitments, but the Energy Offer Curve may be used in Settlement to claw back some or all of a RUC-committed Resource's energy payments. The Energy Offer Curve may also be used by SCED in Real-Time Operations. The Startup Offer component represents all costs incurred by a Generation Resource in starting up and reaching its Low Sustained Limit (LSL). The Minimum-Energy Offer component represents a proxy for the costs incurred by a Resource in producing energy at the Resource's LSL. Startup Offers and Minimum-Energy Offers are not applicable to Energy Storage Resources (ESRs)

A QSE may submit an Energy Offer Curve without also submitting a Startup Offer and Minimum-Energy Offer for the DAM and during the Adjustment Period, but only Three-Part Supply Offers are used in the RUC process. A QSE that submits an Energy Offer Curve without also submitting a Startup Offer and a Minimum-Energy Offer is considered not to be offering the Resource into the RUC, but that does not prevent the Resource from being committed in the RUC process like any other Resource that does not submit an offer in the RUC.

A QSE that submits an Energy Offer Curve without a Startup Offer and a Minimum-Energy Offer for the DAM for any given hour will be considered by the DAM to be **self-committed** for that hour, as long as an Ancillary Service Offer for Off-Line Non-Spin Service was not also submitted for that hour. A Combined Cycle Generation Resource will be considered by the DAM to be self-committed if: i) Its QSE submits an Energy Offer Curve without a Startup Offer and a Minimum-Energy Offer for the DAM for that Combined Cycle Generation Resource and no other Combined Cycle Generation Resource within the Combined Cycle Train; and ii) Its QSE submits no Ancillary Service Offer for Off-Line Non-Spin for any Combined Cycle Generation Resource within the Combined Cycle Train.

For any Operating Hour, the QSE for a Resource may submit or change Energy Offer Curve information at any time prior to SCED execution, except for the percentage of Fuel Index Price (FIP) and percentage of Fuel Oil Price (FOP), and SCED will use the latest updated Energy Offer Curve available in the system. The QSE must provide a brief freeform reason at the time of the submission of the Energy Offer Curve if submitted after the end of the Adjustment Period. For the percentage FIP and percentage of FOP within the Energy Offer Curve, submissions and updates must be received by ERCOT's systems in the Adjustment Period. If a new Energy Offer Curve is not deemed to be valid, then the most recent valid Energy Offer Curve available in the system at the time of SCED execution will be used and ERCOT will notify the QSE that the invalid Energy Offer Curve was rejected. Once an Operating Hour ends, an Energy Offer Curve for that hour cannot be submitted, updated, or canceled

4.2.4.2 Make whole Payments and Mitigation

Resources dispatched in the DAM based on their three part offers are entitled to make whole payments if their energy settlements do not cover their total cost over the day (including startup and minimum energy cost). The FIP and FOP used to calculate the Energy Offer Curve Cap for Make-Whole Payment calculation purposes shall be the FIP or FOP for the Operating Day. In the event the Energy Offer Curve Cap for Make-Whole Payment calculation purposes must be calculated before the FIP or FOP is available for the particular Operating Day, the FIP and FOP for the most recent preceding Operating Day shall be used. Once the FIP and FOP are available for a particular Operating Day, those values shall be used in the calculations.

Energy Offer Curves and Energy Bid/Offer Curves may be subject to mitigation in Real-Time operations under Security Constrained Economic Dispatch, using a Mitigated Offer Cap (MOC) calculated by ERCOT for each resource category.

For each Resource contracted by ERCOT under as Reliability Must Run, the Resource's MOC curve for use in the SCED process is determined by ERCOT when the Resource's offer is subject to mitigation in accordance with Constraint Competitiveness Tests. The single price ($/MWh) value will be used as the MOC curve for the full operating range of the Resource. The calculations occur within the SCED process as well as during the process for determining Real-Time On-Line Reliability Deployment Price Adder. This analysis will only be applied to active constraints for which the contracted Resource has a more than 2% unloading Shift Factor on the Transmission Facility(s), more than 5% unloading impact on the Transmission Facility(s) based on telemetered HSL, and if at least one other Resource not contracted by ERCOT has an unloading Shift Factor of 5% or more relative to the constraint(s).

4.2.4.3 Energy Only Offers by Self-scheduled Resources

The DAM Energy-Only Offer Curve represents the QSE's willingness to sell energy at or above a certain price and at a certain quantity at a specific Settlement Point in the DAM. DAM Energy-Only Offers are not resource specific. The offered energy can be a fixed quantity block, variable quantity block, or curve indicator for the offer. A fixed quantity block consists of single price (in $/MWh) and single quantity (in MW) for all hours offered in that block. A variable quantity block consists of a single price (in $/MWh) and single "up to" quantity (in MW) contingent on the purchase of all hours offered in that block. And an offer curve that is monotonically increasing for both price (in $/MWh) and quantity (in MW) with no more than ten price/quantity pairs within the range of -$250.00 per MWh and the Day Ahead System Wide Offer Cap (DASWCAP), which can be as high as 9000 dollars per MWh

4.2.4.4 DAM Energy Bids

A DAM Energy Bid represents the QSE's willingness to buy energy at or below a certain price and at a certain quantity at a specific Settlement Point in the DAM. DAM Energy Bids can be fixed quantity blocks, variable quantity blocks, or curve indicator for the bid. A Ffixed quantity block has a single price (in $/MWh) and single quantity (in MW) for all hours bid in that block. A variable quantity block has a single price (in $/MWh) and single "up to" quantity (in MW) contingent on the purchase of all hours bid in that block. And a curve is a monotonically decreasing energy bid curve for price (in $/MWh) and monotonically increasing for quantity (in MW) with no more than 10 price/quantity pairs.

4.2.4.5 Energy Bid/Offer Curve

The Energy Bid/Offer Curve represents the willingness of a QSE representing an Energy Storage Resource (ESR) to buy energy at or below a certain price and sell energy at or above a certain price and at a certain quantity in the DAM or its willingness to be dispatched by SCED in Real-Time Operations. ERCOT must validate each Energy Bid/Offer Curve before it can be used in any ERCOT process. A QSE may submit Resource-Specific Energy Bid/Offer Curves to ERCOT. Such Energy Bid/Offer Curves will be bounded in the DAM for each Operating Hour by the Low Sustainable Limit (LSL) and High Sustainable Limit (HSL) of the ESR specified in the Current Operating Plan (COP), and bounded in SCED by the LSL and HSL of the ESR as shown by telemetry.

In the DAM, ERCOT will not consider COP Resource Status when evaluating Energy Bid/Offer Curves. In the Real-Time Market (RTM), SCED will consider an ESR unavailable for SCED Dispatch when the ESR's Resource Status is OUT.

In the RTM, a QSE may submit or change an Energy Bid/Offer Curve at any time prior to SCED execution, and SCED will use the latest updated Energy Bid/Offer Curve available in the system. If a new Energy Bid/Offer Curve is not deemed to be valid, then the most recent valid Energy Bid/Offer Curve available in the system at the time of SCED execution will be used and ERCOT will notify the QSE that the invalid Energy Bid/Offer Curve was rejected. Once an Operating Hour ends, an Energy Bid/Offer Curve for that hour cannot be submitted, updated, or canceled.

4.2.4.6 Day Ahead and Real Time System Wide Offer Caps

The DASWCAP and RTSWCAP shall be determined in accordance with the Public Utility Commission of Texas (PUCT) Substantive Rules. The methodology for determining the DASWCAP and RTSWCAP is as follows: The Low System-Wide Offer Cap (LCAP) is set at $2,000 per MWh for energy and $2,000 per MW per hour for Ancillary Services. At the beginning of each year, the DASWCAP and RTSWCAP shall be set equal to the respective High System-Wide Offer Cap (HCAP) and maintained at this level as long as the Peaker Net Margin (PNM) during a year is less than or equal to the PNM threshold per MW-year. Additionally, the Value of Lost Load (VOLL) used to determine the Ancilary Service Demand Curves (ASDCs) for DAM and RTM shall be set to the HCAP for DAM.

If the PNM exceeds the PNM threshold per MW-year the DASWCAP and the VOLL used to determine the ASDCs for DAM and RTM shall be reset per the Scarcity Pricing Mechanism.

ERCOT shall set the PNM threshold at **three times the cost of new entry** of generation plants. The specific parameters are as follows: HCAP – DAM (DASWCAP)= $5000/MWh, HCAP – RTM (RTSWCAP)=$2000/MWh, PNM threshold $315,000/MW-year. When the calculated PNM exceeds PNM threshold per MW-year, the DASWCAP and the VOLL used to determine the ASDCs for DAM and RTM shall both be changed to the LCAP for the remainder of the calendar year.

4.2.5 DAM Clearing Process

The DAM uses a multi-hour mixed integer programming algorithm to maximize bid-based revenues, including revenues based on Ancillary Service Demand Curves (ASDCs), minus the offer-based costs over the Operating Day, subject to security and other constraints. The bid-based revenues include revenues from ASDCs, DAM Energy Bids, bid portions of Energy Bid/Offer Curves, and Point-to-Point (PTP) Obligation bids. The offer-based costs include costs from the Startup Offer, Minimum Energy Offer, and Energy Offer Curve of any Resource that submitted a Three-Part Supply Offer, DAM Energy-Only Offers, offer portions of Energy Bid/Offer Curves, Ancillary Service Only Offers, and Ancillary Service Offers. Security constraints specified to prevent DAM solutions that would overload the elements of the ERCOT Transmission Grid include the following: Transmission constraints – transfer limits on energy flows through the ERCOT Transmission Grid, e.g., thermal or stability limits. These limits must be satisfied by the intact network and for certain specified contingencies. These constraints may represent: Thermal constraints – protect Transmission Facilities against thermal overload. Generic constraints – protect the ERCOT Transmission Grid against transient instability, dynamic stability or voltage collapse. Power flow constraints – the energy balance at required Electrical Buses in the ERCOT Transmission Grid must be maintained. Resource constraints – the physical and security limits on Resources that submit Three-Part Supply Offers or Energy Bid/Offer Curves: Resource output constraints – the Low Sustained Limit (LSL) and High Sustained Limit (HSL) of each Resource; and Resource operational constraints – includes minimum run time, minimum down time, and configuration constraints. The DAM may not select any one part of a resource capacity to provide more than one Ancillary Service or to provide both energy and an Ancillary Service in the same Operating Hour. The DAM may, however, select part of that Resource capacity to provide one Ancillary Service and another part of that capacity to provide a different Ancillary Service or energy in the same Operating Hour, provided that linked Energy and Off-Line Non-Spinning Reserve (Non-Spin) Resource-Specific Ancillary Service Offers are not awarded in the same Operating Hour.

The sum of the awarded Resource-Specific Ancillary Service Offer capacities for each Resource must be within the Resource limits specified in the Current Operating Plan (COP) and Resource Limits in Providing Ancillary Service, and the Resource Parameters.

Block Resource-Specific Ancillary Service Offers for a Load Resource – blocks will not be cleared unless the entire quantity block can be awarded. Because block Resource-Specific Ancillary Service Offers cannot set the Market Clearing Price for Capacity (MCPC), a block Ancillary Service Offer may clear below the Ancillary Service Offer price for that block.

Block DAM Energy Bids, DAM Energy-Only Offers, and PTP Obligation bids – blocks will not be cleared unless the entire time and/or quantity block can be awarded. Because quantity block bids and offers cannot set the Settlement Point Price, a quantity block bid or offer may clear in a manner inconsistent with the bid or offer price for that block. The DAM may commit a Combined Cycle Generation Resource in a time period that includes the last hour of the Operating Day only if that Combined Cycle Generation Resource can transition to a shutdown condition in the DAM Operating Day. The energy cleared for an ESR may be negative, indicating purchase of energy, or positive, indicating sale of energy.

Ancillary Service needs will be reflected in ASDCs for each Ancillary Service. Self-Arranged Ancillary Service Quantities will first be used to meet the ASDCs, and the remaining Ancillary Service needs are met from Ancillary Service Offers, as long as the costs do not exceed the ASDC value. ERCOT may not buy more of one Ancillary Service in place of the quantity of a different service.

As soon as practicable, but no later than 1330 in the Day-Ahead, ERCOT shall notify the parties to each cleared DAM transaction (e.g., the buyer and the seller) of the results of the DAM as follows: Awarded Resource-Specific Ancillary Service Offers, specifying Resource, MW, Ancillary Service type, and price, for each hour of the awarded offer; Awarded Ancillary Service Only Offers, specifying MW, Ancillary Service type, and price, for each hour of the awarded offer; Awarded energy offers from Three-Part Supply Offers and from DAM Energy-Only Offers, specifying Resource (except for DAM Energy-Only Offers), MWh, Settlement Point, and Settlement Point Price, for each hour of the awarded offer; Awarded DAM Energy Bids, specifying MWh, Settlement Point, and Settlement Point Price for each hour of the awarded bid; Awarded Energy Bid/Offer Curves, specifying Resource, MWh, Settlement Point, and Settlement Point Price, for each hour of the awarded bid/offer; and Awarded PTP Obligation Bids, number of PTP Obligations in MW, source and sink Settlement Points, and price for each Settlement Interval of the awarded bid.

As soon as practicable, but no later than 1330, ERCOT shall post on the ERCOT website the hourly: Day-Ahead MCPC for each type of Ancillary Service for each hour of the Operating Day; DASPPs for each Settlement Point for each hour of the Operating Day; Day-Ahead hourly LMPs for each Electrical Bus for each hour of the Operating Day; Shadow Prices for every binding constraint for each hour of the Operating Day. Energy bought in the DAM consisting of the following: The total quantity of awarded DAM Energy Bids (in MWh) bought in the DAM at each Settlement Point for each hour of the Operating Day; The total quantity of awarded PTP Obligation Bids (in MWh) cleared in the DAM that sink at each Settlement Point for each hour of the Operating Day; and The total absolute value quantity of awards to bid portions of Energy Bid/Offer Curves (in MWh) cleared in the DAM at each Settlement Point for each hour of the Operating Day. Energy sold in the DAM consisting of the following: The total quantity of awarded DAM Energy Offers (in MWh), from Three-Part Supply Offers and DAM Energy Only Offers, bought in the DAM at each Settlement Point for each hour of the Operating Day; The total quantity of awarded PTP Obligation Bids (in MWh) cleared in the DAM that source at each Settlement Point for each hour of the Operating Day; and total quantity of awards to offer portions of Energy Bid/Offer Curves (in MWh) cleared in the DAM at each Settlement Point for each hour of the Operating Day.

Aggregated Ancillary Service Offer Curve of all Ancillary Service Offers (including both Resource-Specific Ancillary Service Offers and Ancillary Service Only Offers) for each type of Ancillary Service for each hour of the Operating Day; Electrically Similar Settlement Points used during the DAM clearing process; Settlement Points that were de-energized in the base case; System Lambda; and Ancillary Services sold in the DAM consisting of the total quantity of awarded Resource-Specific Ancillary Service Offers and Ancillary Service Only Offers, for each Ancillary Service for each hour of the Operating Day.

4.2.6 Transmission Security Analysis and Reliability Unit Commitment

4.2.6.1 Overview

Transmission security analysis and Reliability Unit Commitment (RUC) are used to ensure ERCOT System reliability and to ensure that enough Resource capacity and qualified Ancillary Service capacity are committed in the right locations to reliably serve the forecasted Load and Ancillary Service needs on the ERCOT System including Direct Current Tie (DC Tie) Load that has not been curtailed. ERCOT conducts at least one Day-Ahead Reliability Unit Commitment (DRUC) and at least one Hourly Reliability Unit Commitment (HRUC) before each hour of the Operating Day. The DRUC, which is conducted after the close of the DAM, uses Three-Part Supply Offers, capped at the maximum of generic or verifiable minimum energy and Startup Costs, submitted before the DAM by Qualified Scheduling Entities (QSEs) that were considered in the DAM but not awarded in the DAM. A QSE may not submit a Three-Part Supply Offer to be considered in the DRUC unless the offer was also submitted for consideration in the DAM.

The HRUC process is conducted at least one hour before the Operating Hour to fine-tune the Resource commitments using updated Load forecasts and updated Outage information. However, HRUC may decommit Resources only to maintain the reliability of the ERCOT System. The RUC considers capacity requirements for each hour of the RUC Study Period with the objective of minimizing costs. The RUC process may not be used to buy Ancillary Service unless the Ancillary Service Offers submitted in the DAM are insufficient to meet the requirements of the Ancillary Service Plan.

After the use of market processes to the fullest extent practicable without jeopardizing the reliability of the ERCOT System, any ERCOT Dispatch Instructions for additional capacity that order a QSE to commit a specific Generation Resource to be On-Line shall be considered a RUC Dispatch for the purpose of the Settlement of payments and charges related to the committed Generation Resource.

ERCOT may also issue a Weekly Reliability Unit Commitment (WRUC) Verbal Dispatch Instruction (VDI) to inform a QSE that a Resource is required to be On-Line for all or part of a future Operating Day. Following such instruction the QSE may self-commit the Resource for the WRUC-instructed hours by updating the Resource's Current Operating Plan (COP) prior to the DRUC process execution. :

4.2.6.2 RUC Timeline and Process

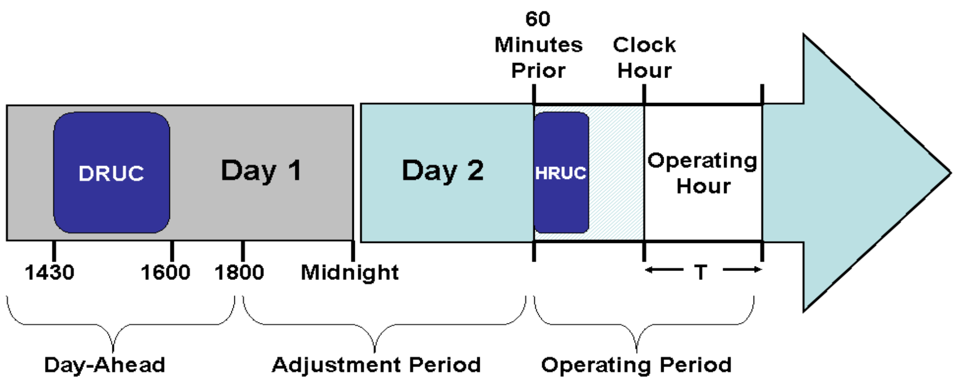


Figure 4-4 Ruc Timeline

Source: https://www.ercot.com/mktrules/nprotocols/current

Figure 4-4 describes the RUC normal timeline. Detailed rules provided in Section 5 of the ERCOT Protocols (https://www.ercot.com/mktrules/nprotocols/current) address cases where the DAM is delayed, the market is aborted or suspended for any reason.

The RUC process recommends commitment of Generation Resources, to match ERCOT's forecasted Load including Direct Current Tie (DC Tie) Schedules, subject to all transmission constraints and Resource performance characteristics. The RUC process takes into account Resources already committed in the Current Operating Plans (COPs), Resources already committed in previous RUCs, Off-Line Available Resources having a start-up time of one hour or less, and Resource capacity already committed to provide Ancillary Service. The formulation of the RUC objective function penalizes violations of security constraints. The objective of the RUC process is to minimize costs based on the Resource costs. For all hours of the RUC Study Period within the RUC process, Quick Start Generation Resources (QSGRs) with a COP Resource Status of OFFQS are considered as On-Line with Low Sustained Limit (LSL) at zero MW. QSGRs with a Resource Status of OFFQS shall only be committed by ERCOT through a RUC instruction in instances when a reliability issue would not otherwise be managed through Dispatch Instructions from Security-Constrained Economic Dispatch (SCED).

The RUC process can recommend Resource decommitment. ERCOT may only decommit a Resource to resolve transmission constraints that are otherwise unresolvable. Qualifying Facilities (QFs) may be decommitted only after all other types of Resources have been assessed for decommitment. In addition, the HRUC process provides decision support to ERCOT regarding a Resource decommitment requested by a Qualified Scheduling Entity (QSE). During the RUC process, ERCOT may also review and commit, through a RUC instruction, Combined Cycle Generation Resources that are currently planned to be On-Line but are capable of transitioning to a configuration with additional capacity. ERCOT may deselect Resources recommended in DRUC and in all HRUC processes

If a Resource receives a RUC Dispatch Instruction that it cannot meet due to a physical limitation the QSE representing the Resource shall notify the ERCOT Operator of the inability to fully comply with the instruction and shall comply with the instruction to the best of the Resource's ability. A QSE shall be excused from complying with any portion of a RUC Dispatch Instruction that it could not meet due to a physical limitation that was reflected, at the time of the RUC Dispatch Instruction, in the Resource's COP, startup time, minimum On-Line time, or minimum Off-Line time.

The HRUC process uses current Resource Status for the initial condition for the first hour of the RUC Study Period. All HRUC processes use the projected status of transmission breakers and switches starting with current status and updated for each remaining hour in the study as indicated in the COP for Resources and in the Outage Scheduler for transmission elements. While the DRUC process uses the Day-Ahead weather forecast for each hour of the Operating Day, the HRUC process uses the weather forecast information for each hour of the balance of the RUC Study Period.

The settlement details for RUC dispatched units are omitted and can be found in Section 5 of the ERCOT Nodal Protocols at https://www.ercot.com/mktrules/nprotocols/current.

4.2.7 Adjustment Period and Real-Time Operations

4.2.7.1 Overview

The Adjustment Period provides each Qualified Scheduling Entity (QSE) the opportunity to adjust its trades, Self-Schedules, and Resource commitments as more accurate information becomes available. During the Adjustment Period, ERCOT continues to evaluate system sufficiency and security by use of Hour-Ahead Reliability Unit Commitment (RUC) processes, Transmission Security Analysis and Reliability Unit Commitmen. During Real-Time operations, ERCOT dispatches Resources under normal system conditions and behavior based on economics and reliability to match system Load with On-Line generation while observing Resource and transmission constraints. The Security-Constrained Economic Dispatch (SCED) process produces Base Points and Ancillary Service awards for Resources. ERCOT uses the Base Points from the SCED process and uses the deployment of Regulation Up Service (Reg-Up), Regulation Down

Service (Reg-Down), ERCOT Contingency Reserve Service (ECRS), Responsive Reserve (RRS), and Non-Spinning Reserve (Non-Spin) to control frequency and solve potential reliability issues.

Real-Time energy settlements use Real-Time Settlement Point Prices that are calculated for Resource Nodes, Load Zones, and Hubs for a 15-minute Settlement Interval, using the Locational Marginal Prices (LMPs) from all of the executions of SCED in the Settlement Interval. Similarly, Real-Time Ancillary Service Settlements use Real-Time Market Clearing Prices for Capacity (MCPCs) for a 15-minute Settlement Interval, using the MCPCs from all of the executions of SCED in the Settlement Interval. In contrast, the Day-Ahead Market (DAM) energy settlements will use DAM Settlement Point Prices that are calculated for Resource Nodes, Load Zones, and Hubs for a one-hour Settlement Interval, and DAM Ancillary Service Settlements will use DAM MCPCs for a one-hour Settlement Interval

4.2.7.2 Real Time Market Timeline

Figure 4-5 below highlights the major activities that occur in the Adjustment Period and Real-Time operations. Activities for the Adjustment Period begin at 1800 in the Day-Ahead and end one full hour before the start of the Operating Hour. ERCOT produces using SCED and monitors the resulting Real-Time Locational Marginal Prices (LMPs), Real-Time Market Clearing Prices for Capacity (MCPCs), and Real-Time Settlement Point Prices, including Real-Time prices for energy metered, Real-Time Reliability Deployment Price Adders for Energy, and Real-Time Reliability Deployment Price Adders for Ancillary Service

The SCED process is designed to simultaneously manage energy, the system power balance and network congestion through Resource Base Points and calculation of LMPs every five minutes. The SCED process uses a two-step methodology that applies **mitigation prospectively** to resolve Non-Competitive Constraints for the current Operating Hour. The SCED process evaluates Energy Offer Curves, Output Schedules and Real-Time Market (RTM) Energy Bids to determine Resource Dispatch Instructions by maximizing **bid-based revenues** minus **offer-based costs**, subject to power balance and network constraints. The SCED process uses the Resource Status provided by SCADA telemetry, Operational Data Requirements, and validated by the Real-Time Sequence, instead of the Resource Status provided by the COP. The SCED solution must monitor cumulative deployment of Regulation Services and ensure that Regulation Services deployment is minimized over time. In the Generation To Be Dispatched (GTBD) determined by Load Frequency Control (LFC), ERCOT shall subtract the sum of the telemetered net real power consumption from all Controllable Load Resources available to SCED. For use as SCED inputs, ERCOT uses the available capacity of all committed Generation Resources by creating proxy Energy Offer Curves for certain Resources for which its QSE has submitted an Output Schedule instead of an Energy Offer Curve, for Dynamically Scheduled Resources (DSRs) that submitted only incremental and decremental Energy Offer Curves and for RUC committed resources with no offer curves. Specific details on how the proxy Energy curves are constructed can be found at . : https://www.ercot.com/mktrules/nprotocols/current

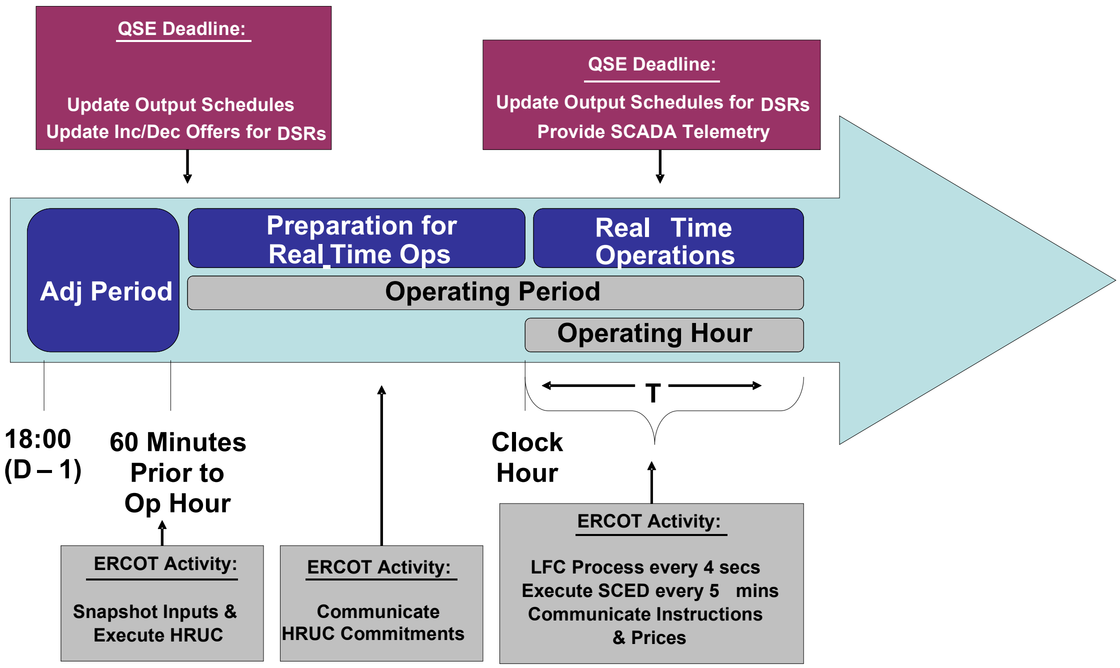


Figure 4-5 Real Time Operations

Source: https://www.ercot.com/mktrules/nprotocols/current

.For each SCED process, ERCOT shall calculate a Real-Time On-Line Reserve Price Adder and a Real-Time Off-Line Reserve Price Adder based on the On-Line and Off-Line available reserves in the ERCOT System and the **Operating Reserve Demand Curve** (ORDC). The Real-Time Off-Line available reserves shall be administratively set to zero when the SCED snapshot of the Physical Responsive Capability (PRC) is equal to or below the PRC MW at which Energy Emergency Alert (EEA) Level 1 is initiated. In addition, for each SCED process, ERCOT shall calculate a Real-Time On-Line Reliability Deployment Price Adder. The sum of the Real-Time Reliability Deployment Price Adder and the Real-Time On-Line Reserve Price Adder shall be averaged over the 15-minute Settlement Interval and added to the Real-Time LMPs to determine the Real-Time Settlement Point Prices. The price after the addition of the sum of the Real-Time On-Line Reliability Deployment Price Adder and the Real-Time On-Line Reserve Price Adder to LMPs approximates the pricing outcome of the impact to energy prices from reliability deployments and the Real-Time energy and Ancillary Service co-optimization since the Real-Time On-Line Reserve Price Adder captures the value of the **opportunity cost** of reserves based on the defined ORDC. An Ancillary Service imbalance Settlement with

Real-Time Ancillary Service Imbalance Payment or Charge, is used to make Resources indifferent to the utilization of their capacity for energy or Ancillary Service reserves.

4.2.8 The Operating Reserves Demand Curve (ORDC)

An ORDC develops a proxy to the demand for reserves based on value of lost load and probabilistic assessment of the likelihood of demand curtailment versus the level of reserves. The simplified implementation currently used in ERCOT augments the prices resulting from the day ahead and real time market clearing auctions with an adder reflecting operating reserves use and grid stress.

For purposes of the grid operation, the value of generation availability is a combination of the locational and operational characteristics of the unit. Congestion and scarcity pricing are needed to send generation owners the right reliability value signals. Scarcity pricing in ERCOT initially relied on limited exercise of market power because the PUCT objected to administrative imposition of scarcity rents based on the value of lost load (VOLL), which was necessary in the absence of meaningful demand response, and instead insisted that scarcity prices should result from high offers by suppliers.

The ORDC arose out of an extensive, prolonged, and exhausting debate that occurred at the PUCT over whether to replace the, then, non-binding target reserve margin with a mandatory minimum reserve margin. This debate began in the fall of 2011 following a year of extreme weather and drought and continued until early 2014. On February 2, 2011 Texas experienced below freezing weather across the entire state, the likes of which had not been experienced in 22 years, which resulted in 146 generator forced outages that required ERCOT to implement involuntary rotating outages for more than four hours across ERCOT in order to maintain grid stability despite ERCOT's having an actual installed capacity reserve margin of over 16% that year. At the peak of the rolling outages up to 4,000 MW were required to be curtailed. This experience was quickly followed by a record hot summer that began in early spring and continued through most of September. By August of that year Texas had experienced a record number of days in excess of 100 degrees and record temperatures across the state, coupled with a very severe drought. As a result, ERCOT came very close on several days to having to implement system-wide load curtailments because of generator forced outages and de-rates coupled with record peak loads, again despite having what was then an otherwise healthy installed capacity reserve margin on paper.

Later that year, the then Chairman of the PUCT called for the establishment of a mandatory minimum reserve margin because of ERCOT's forecasts of declining target reserve margins. Taking this action would have resulted in the commission having to abandon ERCOT's energy-only market design by adopting some mandatory mechanism to maintain that required reserve margin. The PUCT's Chairman was unable to move the proposal forward because the PUCT only had two members and was deadlocked over the issue. The commission took a number of steps to improve the performance of the existing energy market including establishing a timetable to raise ERCOT's offer cap from $3,000/MWh to $9,000/MWh. These actions helped relieve the some of the pressure to "do something."

While these and other enhancements to the energy market were being discussed and considered, the use of an ORDC in ERCOT was first brought to the attention of the PUCT commissioners individually through an original paper on ORDC authored by Dr. William W. Hogan. The commissioners decided to pursue the idea as a possible enhancement or even solution to the resource adequacy debate. Ultimately, the proposed ORDC went through several iterations to reflect the ERCOT settlement systems and processes and was finally adopted on September 12, 2013. The commission decided that the ORDC would produce a system wide adjustment to the price of operating reserves utilizing a price adder to the real-time clearing price in each five minute dispatch and pricing interval equal to the product of the value of lost load (VOLL) multiplied by the LOLP. The VOLL was set at $9,000 because that was where the offer cap was going to be set by 2015 under the PUCT's prior order. The minimum contingency was set at 2000 MW under the rationale that when operating reserves fell below that figure ERCOT's operators ceased relying upon the market and began taking out of market actions to maintain the stability of the grid, (although it should be recognized that this level of 2000 MW neither matches the largest single or largest double contingency in ERCOT). Consequently, the ORDC's LOLP reaches 100% at a reserve level of 2000 MW and the ORDC adder should be $9,000/MWh at this reserve level. Lastly, it was decided that ERCOT would use a continuous function advocated by the IMM, rather than the piecewise linear approximation used by ERCOT and that when the ORDC was implemented, the energy offer price floors for deployment of ancillary services would be removed. These offer floors were intended to mitigate the price reversals that would otherwise occur when such services were deployed but became superfluous once the ORDC was implemented. The LOLP was based on an ERCOT calculation that does not literally calculate the LOLP in the relevant dispatch interval. The PUCT then instructed ERCOT to draft the protocols necessary to implement the commission's decisions and to push them through the stakeholder process. ERCOT subsequently filed a draft copy of the necessary protocols for commission review and ERCOT and the stakeholders moved forward to implement the commission's unanimous decisions.

4.2.9 Real Time Settlement Prices for energy and Ancilary Services

Real-Time energy Settlements use Real-Time Settlement Point Prices that are calculated for Resource Nodes, Load Zones, and Hubs. For each Security-Constrained Economic Dispatch (SCED) Locational Marginal Price (LMP) calculated at each Settlement Point in the SCED process, an administrative price floor of -$251/MWh will be applied to Real-Time Settlement Point Prices after adding the Real-Time Reliability Deployment Price Adders for Energy. ERCOT shall assign an LMP to de-energized Electrical Buses for use in the calculation of the Real-Time Settlement Point Prices by using heuristic rules applied in the following order:

The Real-Time Settlement Point Price for a Load Zone Settlement Point is based on the state-estimated Load in MW and the time-weighted average Real-Time LMPs at Electrical Buses that are included in the Load Zone. The Real-Time Settlement Point Price for a Load Zone Settlement Point for a 15-minute Settlement Interval is calculated as follows:

The Real-Time Market Clearing Price for Capacity (MCPC) for Reg-Up is the time-weighted average of the sum of the Real-Time MCPCs for Reg-Up and Real-Time Reliability Deployment Price Adder for Ancillary Service for Reg-Up of each SCED interval in the 15-minute Settlement Interval. Likewise, The Real-Time MCPC for Reg-Down is the time-weighted average of the sum of the Real-Time MCPCs for Reg-Down and Real-Time Reliability Deployment Price Adder for Ancillary Service for Reg-Down of each SCED interval in the 15-minute Settlement Interval

The Real-Time MCPC for Responsive Reserves (RRS) is the time-weighted average of the sum of the Real-Time MCPCs for RRS and Real-Time Reliability Deployment Price Adder for Ancillary Service for RRS of each SCED interval in the 15-minute Settlement Interval. Likewise The Real-Time MCPC for Non-Spin is the time-weighted average of the sum of the Real-Time MCPC for Non-Spin and Real-Time Reliability Deployment Price Adders for Ancillary Service for Non-Spin of each SCED interval in the 15-minute Settlement Interval/

In June 2023 ERCOT launched a new daily procured ancillary service called ERCOTContingency Reserves (ECRS) which has been created and implemented to support grid reliability and mitigate real-time operational issues to keep supply and demand balanced. ECRS complements and provides support to the four procured Ancillary Services ERCOT has been using: Regulation Up, Regulation Down, Responsive Reserve Service and Non-Spin Reserve Service. ECRS is introduced to restore RRS availability once RRS resources are depleted or to mitigate a reliability concern if there is a deficiency in the ramping capacity. By design, ECRS can be dispatched by SCED and should respond within 10 minutes to the deployment instructions. The Real-Time MCPC for ECRS is the time-weighted average of the sum of the Real-Time MCPC for ECRS and Real-Time Reliability Deployment Price Adder for Ancillary Service for ECRS of each SCED interval in the 15-minute Settlement interval/

4.2.9.1 Real-Time Congestion Payment or Charge for Self-Schedules

The congestion payment or charge to each QSE submitting a Self-Schedule calculated based on the difference in Real-Time Settlement Point Prices at the specified sink and the source of the Self-Schedule multiplied by the amount of the Self-Schedule

4.2.9.2 Real-Time Ancillary Service Imbalance Payment or Charge

Based on the Real-Time On-Line Reliability Deployment Price Adders, Real-Time On-Line Reserve Price Adders and a Real-Time Off-Line Reserve Price Adders, ERCOT shall calculate Ancillary Service imbalance Settlement, which will make Resources indifferent to the utilization of their capacity for energy or Ancillary Service reserves. The payment or charge to each QSE for Ancillary Service imbalance is calculated based on the price calculation in the Security Constrained Economic Dispatch, and applied for each QSE: to the amount of Real-Time Metered Generation from all Generation Resources, represented by the QSE for the 15-minute Settlement Interval;

4.2.10 Congestion Revenue Right

4.2.10.1 Definition and Aquisition

A Congestion Revenue Right (CRR) is a financial instrument that entitles the CRR Owner to be charged or to receive compensation for congestion rents that arise when the ERCOT Transmission Grid is congested in the Day-Ahead Market (DAM) or in Real-Time. CRRs do not represent a right to receive, or obligation to deliver, physical energy. Most CRRs are tradable in the CRR Auction, in the DAM, or bilaterally, as described in more detail in this Section.

CRRs may be acquired as follows:

(a) CRR Auction conducted periodically where CRR account holders can procure ands sell CRRs.
(b) PCRR Allocations where ERCOT allocates Pre-Assigned Congestion Revenue Rights (PCRRs)) to eligible Municipally Owned Utilities (MOUs) and Electric Cooperatives (ECs)
(c) Bilateral Market where CRR Account Holders may trade privately or through ERCOT, Point-to-Point (PTP) Options and PTP Obligations
(d) DAM – Qualified Scheduling Entities (QSEs) may bid for PTP Obligations in the DAM.

CRR may be of the following types:
(a) Point to Point (PTP) Option, some of which may be PCRRs;
(b) PTP Obligation, some of which may be PCRRs;
(c) PTP Option with Refund, all of which are PCRRs;
(d) PTP Obligation with Refund, all of which are PCRRs; and
(e) Flowgate Right (FGR).

Each CRR is characterized by Quantities in MWs, duration of one hour, ability to be be fully tradable, designated source and sink settlement points in case of PTP and directional network element or a bundle of directional elements in case of FGRs

ERCOT CRR Auctions include PTP Options, PTP Obligations and FGRs. PTP Options are evaluated hourly as the positive power flows on all directional network elements created by the injection and withdrawal at the specified source and sink points in the quantity represented by the CRR bid or offer (MW), excluding all negative flows on all directional network elements. PTP Obligations are evaluated hourly as the positive and negative power flows on all directional network elements created by the injection and withdrawal at the specified source and sink points of the quantity represented by the CRR bid or offer (MW). Consequently, PTP Options can only result in payments from ERCOT to the CRR Owner of record while a PTP Obligation may result in either a payment or a charge to the CRR Owner of record. FGRs are offered only on a select number of links or bundle of links (Called Flowgates) and they are evaluated hourly based on the quantity represented by the CRR bid or offer (MW).on the selected directional network elements. Currently ERCOT has no designated flowgates.

Pre-Assigned Congestion Revenue Rights (PCRRs) are allocated to Non-Opt-In Entities (NOIEs) that either have established ownership prior to September 1, 1999 in a specific Generation Resource or have a long-term (greater than five years) contractual commitment for annual capacity and energy that was entered into prior to September 1, 1999 from specific Generation Resources. The NOIE is entitled to capacity in the form of PCRRs from specific Generation Resource(s) pursuant to the long-term contract(of more than five years); long-term contracts that are not backed by specific Generation Resources are not eligible for PCRRs unless the contract is a portfolio supply contract with multiple generation resources. Also eligible to PCRRs is a contract with A federally-owned hydroelectric Generation Resource that is the subject of a series of sequential long-term contracts between the NOIE nominating the PCRR(s) and the federal government based upon a long-term (greater than five years) allocation from the federal government for annual capacity and energy produced at such federally-owned hydroelectric Generation Resource, and that allocation was in place prior to September 1, 1999.

4.2.10.2 CRR Auctions

The Congestion Revenue Right (CRR) Auction auctions the available network capacity of the ERCOT transmission system not allocated, or sold in a previous auction. The CRR Auction also allows CRR Owners an opportunity to offer for sale CRRs that they hold. Each CRR Auction allows for the purchase of CRR products. Types of Congestion Revenue Rights to Be Auctioned, in strips of one or more consecutive months and allows for the reconfiguration of all CRR blocks that were previously awarded for the months covered by that CRR Auction.

The CRR Network Model must be based on, but is not the same as, the Network Operations Model. For the purposes of CRR Network Model construction for a CRR Long-Term Auction Sequence, The CRR Network Model must, to the extent practicable, include the same topology, contingencies, and operating procedures as used in the Network Operations Model as reasonably expected to be in place for each month. The expected network topology used in the CRR Network Model for any month or set of months must include all Outages from the Outage Scheduler and identified by ERCOT as expected to have a significant impact upon transfer capability during that time. These Outages included in the CRR Network Model shall be posted on the Market Information System (MIS) Secure Area consistent with model posting requirements by ERCOT with accompanying cause and duration information, as indicated in the Outage Scheduler. Transmission system upgrades and changes must be accounted for in the CRR Network Model for CRR Auctions held after the month in which the element is placed into service.

ERCOT shall model bids and offers into the CRR Auction as flows based on the MW offer and defined source and sink. When the Simultaneous Feasibility Test (SFT) is run, the model must weight the power flow buses and Hub Buses included in a Hub or Load Zone appropriately to determine the system impacts of the CRRs. To distribute injections and withdrawals to buses within a Hub, ERCOT shall use distribution factors that are consistent with planning and operations.

ERCOT shall conduct CRR Auctions as follows: The CRR Monthly Auction, held once per calendar month, shall include the sale of one-month terms of Point-to-Point (PTP) Options and PTP Obligations for the month immediately following the month during which the CRR bid submission window closes. Twice per year, a CRR Long-Term Auction Sequence shall be held, selling PTP Options and PTP Obligations, subject to the following constraints: Each CRR Long-Term Auction Sequence shall consist of six successive CRR Auctions, each of which offers for sale CRRs spanning a term of six consecutive calendar months (either January through June, or July through December). In each such CRR Auction, CRRs shall be offered in one-month strips or in strips of up to six consecutive months within the term covered by the auction. The CRR Long-Term Auction Sequence shall operate in chronological order, first providing a CRR Auction covering the next six-month (January through June, or July through December) period that has not yet commenced, and then five successive CRR Auctions for the five six-month periods thereafter. For each CRR Auction, the CRR Auction Capacity shall be defined as follows: For the CRR Monthly Auction, 90%. For any CRR Auction that is part of a CRR Long-Term Auction Sequence, 70%, 55%, 40%, 30%, 20%, or 10% for the first, second, third, fourth, fifth, and sixth six-month windows sold in the sequence, respectively.

A CRR Auction Offer indicates a willingness to sell CRRs at the auction clearing price, if it equals or exceeds the Minimum Reservation Price. Likewise, A CRR Auction Bid indicates a willingness to buy CRRs at the auction clearing price, if it is equal to or less than the Not-to-Exceed Price. Both Offers and Bids must be submitted by a Participating CRR Account Holder and must include the following: The unique identifier for each CRR, type of CRR, The source Settlement Point and the sink Settlement Point; The month, or strip of consecutive months, time-of-use designated, The quantity of CRRs in MW, and a dollars per CRR (i.e. dollars per MW per hour) for the Reservation Price. The Participating CRR Account Holder may submit a self-imposed credit limit for the CRR Monthly Auction or for each time-of-use in a CRR Auction that is part of a CRR Long-Term Auction Sequence, if desired.

The CRR Auction system prevents a CRR Account Holder from being awarded bids and offers that exceed the lesser of the CRR Account Holder's self-imposed credit limit or the credit limit as prescribed by ERCOT Credit Requirements for CRR Auction Participation.

4.2.10.2.1 CRR Auctions Process

The CRR Auction must be a single-round, simultaneous auction for selling the CRRs available for all auction products. ERCOT shall enter into the CRR Auction system a credit limit for each Counter-Party that has at least one CRR Account Holder. A Counter-Party's CRR Auction credit limit is equal to the lesser of the credit limit imposed by ERCOT, or, if provided, the Counter-Party's self-imposed CRR Auction credit limit for the CRR Monthly Auction or for a time-of-use within a CRR Auction held as part of a CRR Long-Term Auction Sequence. Prior to the CRR Auction, ERCOT will conduct a two-part pre-auction screening process. First, if the Counter-Party's CRR Auction credit limit is greater than that Counter-Party's credit exposure as defined below using the CRR bid volumes rather than awarded volumes, then the Counter-Party's CRR Auction credit limit will be ignored as the CRR Auction is solved. Second, for each CRR Account Holder of a Counter-Party, if the CRR Account Holder's self-imposed credit limit is greater than that CRR Account Holder's credit exposure as defined below, then the CRR Account Holder's self-imposed credit limit will be ignored as the CRR Auction is solved.

The calculated exposure for the pre-auction screening for each CRR Account Holder is the sum of the credit exposure for PTP Obligation bids, PTP Obligation offers, and PTP Option bids for that CRR Account Holder. The calculated exposure for the pre-auction screening for each Counter-Party is the sum of the credit exposure for PTP Obligation bids, PTP Obligation offers, and PTP Option bids for that Counter-Party. PTP Option offers have zero credit exposure. Separately, for PTP Obligation bids, PTP Obligation offers, and PTP Option bids, for each source/sink Settlement Point combination, the credit exposure will use the bid price and MW quantity that produces the maximum credit exposure that could result from the CRR Auction for that source/sink Settlement Point combination.

4.2.10.2.2 Simultaneous Feasibility Test

The Simultaneous Feasibility Test (SFT) is a market feasibility test that confirms that the transmission system can support the awarded set of CRRs during normal system conditions, assuming that the Network Operations Model updated with Real-Time network topology is the same as that modeled (for the CRR Auction), while observing all security constraints. The SFT uses a Direct Current (DC) power-flow model **with no losses**, to model the effect of CRR Auction bids and offers on the expected system network topology during the auction term. SFT is not a system reliability test and is not intended to model actual system operating conditions. SFTs are run during the determination of the winning bids and offers for the CRR Auction. Inputs to the SFT model include: CRR bids and offers for the auction; All previously awarded or allocated CRRs for each month; Transmission line Outage schedules; Expected configuration of Transmission Facilities, adjusted for oversold CRRs, as specified in paragraph (e) below; Increased capacity of each element that has been oversold in prior CRR Auctions and CRR allocations to exactly match the amount of CRRs that have been sold or allocated on that element (this ensures the feasibility of the CRR Auction); Thermal operating limits (including estimates for Dynamic Ratings) for transmission lines;

For a CRR Long-Term Auction Sequence, ERCOT shall use Dynamic Ratings based on a historical analysis of the maximum peak-hour temperatures for the previous ten years; and For the CRR Monthly Auction, ERCOT shall use Dynamic Ratings for the maximum peak-hour temperature forecast for the month; Voltage and stability limits that are valid for the study period converted to thermal limits; ERCOT Transmission Grid pre- and post-contingency ratings; All Transmission Element contingencies expected to be used by ERCOT in Real-Time operations.

4.2.10.3 CRR Balancing Account

In the Day-Ahead Market (DAM), if the congestion rent (resulting from buying and selling energy at LMPs and from congestion charges paid by self scheduled transaction) is equal to or greater than the net amounts due to all Congestion Revenue Right (CRR) Owners for any Settlement Interval, then ERCOT shall pay the net amounts due to the CRR Owners and put any excess amount into the CRR Balancing Account (CRRBA).

However, if the congestion rent is less than the net amounts due to all CRR Owners for any Settlement Interval, then ERCOT shall short-pay each CRR Owner on a prorated basis and shall keep track of how much each CRR Owner has been short-paid. The proration must be calculated using only the amounts due to the CRR Owner for CRRs settled in both the DAM and Real-Time and not using amounts due to ERCOT for Point-to-Point (PTP) Obligations owned by the CRR Owner.

ERCOT shall pay any positive balance in the CRRBA to each short-paid CRR Owner, with the amount paid to each CRR Owner being the lesser of (a) a prorated amount based on the short-paid amount for that CRR Owner compared to the total short-paid amount, and (b) the short-paid amount for that CRR Owner. Any remaining positive balance in the CRRBA will first be used to fund the CRRBA fund up to the fund cap and any surplus must be allocated to all Qualified Scheduling Entities (QSEs) based on the QSE’s ratio shares for the month.

4.2.10.4 Day-Ahead CRR Payments and Charges

ERCOT shall pay or charge the owner of each **Point-to-Point (PTP) Obligation** based on the difference in the Day-Ahead Settlement Point Price between the sink Settlement Point and the source Settlement Point. For PTP Obligations that have a positive value and sink at a Resource Node, the PTP Obligation payment may be reduced due to directional network elements that are oversold in previous Congestion Revenue Right (CRR) Auctions. PJM

ERCOT shall pay the owner of a **PTP Option** the difference in the Day-Ahead Settlement Point Price between the sink Settlement Point and the source Settlement Point, if positive.

For PTP Options that sink at a Resource Node, the PTP Option payment may be reduced due to Transmission Elements that are oversold in previous CRR Auctions.

In theory owners of FGRs in the direction of the Day ahead flow on the specific link or bundle of links should be paid by ERCOT the shadow price resulting from the Day Ahead SCUC. However, currently ERCOT has no defined Flowgates so the protocol does not specify FGR settlements.

It should be noted that ERCOT also auctions off and allocates Real Time CRRs which are distributed and settled in the same way as the Day Ahead CRRs based on the Real Time LMPs. A Point-to-Point (PTP) Obligation bid is, as before, a bid that specifies the source and sink, a range of hours, and a maximum price that the bidder is willing to pay ("Not-to-Exceed Price"). PTP Obligations that are bought in the Day-Ahead Market (DAM) are settled based on the applicable Real-Time Settlement Point Prices.

4.2.11 State Of Texas Renewable Energy Program

On May 9, 2000, the Public Utility Commission of Texas (PUCT) appointed ERCOT as Program Administrator of the Renewable Energy Credits (REC) Trading Program The purposes of the REC Trading Program have been: To ensure that the cumulative installed generating capacity from renewable energy technologies in this state with a target of 10,000 MW of installed renewable capacity by January 1, 2025 and ensure that all Customers have access to providers of energy generated by renewable energy Resources ERCOT determines and enforces the annual Renewable Portfolio Standard (RPS) requirement for each Retail Entity in Texas and on a quarterly basis, awards RECs or Compliance Premiums earned by REC generators based on verified MWh production data and retires RECs or Compliance Premiums as they expire.

ERCOT conducts a REC Trading Program Settlement process annually;Monitors the operational status of participating renewable energy generation facilities in Texas and records retirements; ERCOT may conduct site visits to renewable energy generation facilities on a random basis to ensure integrity of the REC Trading Program, as deemed necessary. A REC or Compliance Premium is a tradable instrument that represents all of the renewable attributes associated with one MWh of production from a certified renewable generator. A REC or Compliance Premium may trade separately from energy. RECs are distributed to REC generators on a quarterly basis by ERCOT. The number of RECs distributed to a certified generator is based on physically metered MWh production. RECs may be traded, transferred, and retired. Compliance Premiums are awarded by the Program Administrator in conjunction with a REC that is generated by a renewable energy Resource that is not powered by wind and meets PUCT criteria.

Renewable Energy Credit (REC) generators or REC aggregators must apply to the Public Utility Commission of Texas (PUCT) for certification to produce or aggregate RECs. On receipt of a copy of a notification from the PUCT certifying that a renewable energy generation facility is eligible to generate or an Entity is eligible to aggregate RECs, ERCOT shall establish a REC trading account for the facility or Entity. All Renewable Energy Credit (REC) generators and REC offset generators must report quarterly MWh production data to ERCOT no later than the 38th day after the last Operating Day of the quarter, in an electronic format prescribed by ERCOT. The reported MWh quantity shall be solely produced from, and attributable to, a renewable generator as so designated by the Public Utility Commission of Texas (PUCT). Information relevant to quarterly reporting shall be handled in one of the following processes:

REC aggregators shall report production from microgenerator renewable energy Resources that are not interval metered for energy Settlement, in accordance with the methodology approved by the PUCT for the purposes of measuring the REC production of such Resources, in the format prescribed by ERCOT, including applicable supporting documentation. To enable Retail Entities the ability to calculate their Renewable Portfolio Standard (RPS) requirements, all Retail Entities serving Load in the state of Texas shall provide Load data to ERCOT on a monthly basis, and no later than the 38th day after the last Operating Day of the month, in an electronic format prescribed by ERCOT. The reported MWh quantity shall be solely the energy consumed by Customers in Texas. Load data.

On the receipt of a request from the owner of a Renewable Energy Credit (REC) or Compliance Premium and purchaser of the REC or Compliance Premium, ERCOT will transfer the REC or Compliance Premium from the owner's REC trading account to the REC trading account specified in the transfer request. Transfer requests received by ERCOT shall be effective upon confirmation by the receiving Entity.

Finally, a REC offset represents one MWh of renewable energy from a renewable energy generator placed in service before September 1, 1999 that may be used in place of a REC to meet a renewable energy requirement but may not be traded.

5 PJM

5.1 PJM at a Glance

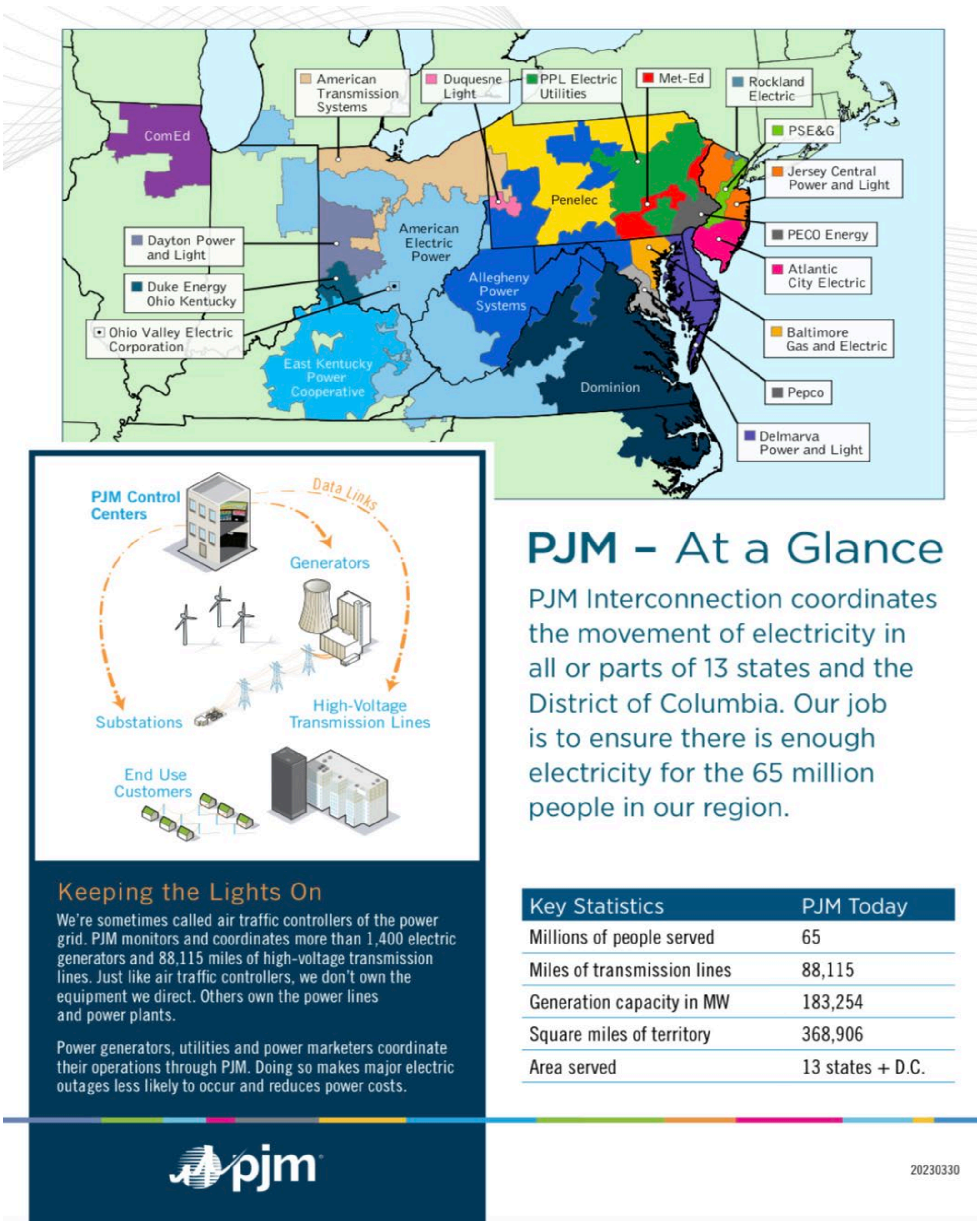


Figure 5-1: PJM At a Glance

(Source: https://www.pjm.com/-/media/about-pjm/newsroom/fact-sheets/pjm-at-a-glance.ashx)

PJM Interconnection serves as the regional transmission organization in the United

States for a 369,089-square-mile area that covers 13 states including all or parts of Delaware, Illinois, Indiana, Kentucky, Maryland, Michigan, New Jersey, North Carolina, Ohio, Pennsylvania, Tennessee, Virginia, West Virginia and the District of Columbia. As an RTO, PJM plays a vital role in the U.S. electric system. PJM: It ensures the reliability of the high-voltage electric power system by coordinating and directing the movement of electricity in its region, It operates a competitive wholesale electricity market. It plans generation and transmission expansion to ensure reliability, It operates independently and neutrally to provide real-time information to its members and customers and to support their decision-making.

The scope of PJM's operations is extensive: It serves a population of approximately 65 million. It controls 1,379 generation resources with diverse fuel types. Total generation capacity in PJM amounts to 180,086 megawatts, serving peak demand of 165,492 megawatts. In 2018, 806,546 GWh of annual were transacted in PJM. The PJM grid has over 84,000 miles of transmission lines. PJM serves 1,018 members with an annual billing of $49.80 billion.

5.2 Overview of the PJM Energy Market

PJM Energy Markets consist of two main markets: the Day-ahead Market and the Real-time Balancing Market. These markets go through a two-step process involving dispatch and pricing to ensure the efficient operation of the system. During the dispatch run, a security-constrained economic dispatch is performed, which aims to meet load and reserve requirements at the lowest cost while considering transmission constraints. This step is focused on determining how much power each resource should generate in order to meet the demand. After the dispatch run, the pricing run takes place. In this step, Locational Marginal Prices (LMPs) are calculated. LMPs indicate the cost of producing an additional unit of electricity at specific locations within the transmission system. During the pricing run, an optimization problem is solved to determine LMPs, taking into account the committed resources from the dispatch run.

One notable aspect during the pricing run is the concept of Integer Relaxation. It allows Eligible Fast-Start Resources, which are capable of operating with short notification and startup times, to set prices and include their associated commitment costs. This means that Fast-Start Resources, such as Fuel Cells, CTs, Diesels, Hydro, Battery, Solar, Landfill, Wind, Hybrid Resources, and Economic Load Response, can be fully dispatched between zero and their Economic Maximum in the determination of LMPs, even if they have limited dispatchable ranges.

It's worth noting that resources cannot be committed in the pricing run if they were not committed in the dispatch run. This constraint ensures that the optimization problem considers only the resources that were already committed during the dispatch run, including their associated costs. Overall, the PJM Energy Markets' process aims to efficiently manage the energy system by determining the dispatch of resources in the most cost-effective manner and calculating LMPs to reflect the costs of producing additional electricity at different locations within the transmission system.

5.2.1 Day-ahead Energy Market

The Day-ahead Energy Market is a forward market in which hourly clearing prices are calculated for each hour of the next operating day based on generation offers, demand bids, increment offers, decrement bids and bilateral transaction schedules submitted into the Day-Ahead Energy Market. The Day-Ahead Energy Market enables participants to purchase and sell energy at binding Day-Ahead LMPs. It also allows transmission customers to schedule bilateral transactions at binding Day-Ahead congestion charges based on the differences in the Congestion Prices between the transaction source and sink. Load Serving Entities (LSEs) may submit hourly demand schedules, including any price sensitive demand, for the amount of demand that they wish to lock-in at Day-ahead prices. Any generator that is a PJM generation capacity resource that has a Reliability Pricing Model (RPM) or Fixed Resource Requirement (FRR) Resource Commitment must submit a bid schedule into the Day-ahead Energy Market even if it is self-scheduled or unavailable due to outage. Other generators have the option to bid into the Day-ahead Energy Market.

Transmission customers, that is customers that are self-scheduled and only require transmission services, may submit fixed, dispatchable or 'up to' congestion bid bilateral transaction schedules into the Day-ahead Energy Market and may specify whether they are willing to pay congestion charges or wish to be curtailed if congestion occurs in the Real-time Energy Market. Curtailment Service Providers (CSPs) may submit demand reduction bids. After the daily quote period closes, PJM calculates the Day-ahead schedule based on the bids, offers and schedules submitted, using the scheduling programs based on least-cost, security constrained resource commitment and dispatch for each hour of the next operating day. The Day-ahead scheduling process incorporates PJM reliability requirements and reserve obligations into the analysis. The resulting Day-ahead hourly schedules, generated by the dispatch run, and Day-ahead LMPs, generated by the pricing run, represent binding financial commitments to the market participants. The Day-ahead Market settlement is calculated for each Day-ahead Settlement Interval (hourly interval) based on scheduled hourly quantities resulting from the dispatch run and on Day-ahead hourly prices resulting from the pricing run.

Market Buyers and Market Sellers may submit virtual increment offers or decrement bids at any hub, node at which physical generation or load is settled. Virtual offers and bids do not have to be backed by physical capacity and such day ahead positions are typically closed in real time at prevailing prices. Virtual offers and bids enable market participants to arbitrage the difference between day ahead and real time prices or to be able to schedule energy transactions in the day ahead but settle it at real time prices (instead of withholding the offers or bids till real time). Energy market transactions, except generation resource offers and price sensitive demand bids, may be submitted with an energy bid/offer price of no greater than $2,000/ MWh. PJM may require, however, that a Market Participant shall not submit in excess of 3000 virtual bid/offer segments in the

Day-ahead Energy Market, when PJM determines that such limit is required to avoid or mitigate significant system performance problems related to the volume of virtual bids.

PJM also allows transmission service customers to to self-schedule bilateral transactions between specified pricing source and sink nodes with 'Up to' congestion bids that will limit their exposure to congestion charges. However, Up-to congestion bids shall be no greater than $50/MWh, and no less than -$50/ MWh. Any 'Up-to' congestion transaction that bids higher than $50/MWh or less than -$50/MWh will be rejected. PJM maintains an up-to date list of source/sink combinations that will be available for 'Up-to 'congestion bidding. The eligible 'Up-to' bidding locations are posted at https://www.pjm.com/markets-and-operations/energy.aspx.

'Up-to" congestion bids are cleared based on the total LMP price difference between the source and the sink. PJM may require that a Market Participant shall not submit in excess of 3000 'Up-to' congestion transactions in the Day-ahead Energy Market, when PJM determines that such limit is required to avoid or mitigate significant system performance problems related to the volume of transactions.

5.2.2 Real-time Energy Market

The Real-time Energy Market uses the Real-time Security Constrained Economic Dispatch (RTSCED) program, known as the dispatch run, to determine the least cost solution to balance supply and demand. The dispatch run considers resource offers, forecasted system conditions, and other inputs in its calculations. For more information regarding the RT SCED program, Generators and Demand Resources may alter their bids for use in the Real-time Energy Market during the following periods:

- During the Generation Rebidding Period which is defined from the time PJM posts the results of the Day-ahead Energy Market until 14:15.
- Starting at 18:30 (typically after the Reliability Assessment and Commitment (RAC) Run is completed) and up to sixty-five (65) minutes prior to the start of the operating hour.

Real-time LMPs and Regulation and Reserve Clearing Prices are calculated every five (5) minutes by the Locational Price Calculator (LPC) program, in a process referred to as the pricing run, and are based on forecasted system conditions and the latest approved RT SCED program solution. The balancing settlement is calculated for each Real-time Settlement Interval (five (5) minute interval) based on actual five (5) minute Revenue Data for Settlement MW quantity deviations from Day-ahead scheduled quantities resulting from the dispatch run and on the applicable Real-time prices resulting from the pricing run.

5.2.3 Locational Marginal Prices

Locational Marginal Price (LMP) is defined as the marginal price for energy at the location where the energy is delivered or received and is based on forecasted system conditions and the latest approved Real-time security constrained economic dispatch program solution. LMP is expressed in dollars per megawatt-hour ($/MWh). LMP is a pricing approach that addresses Transmission System congestion and loss costs, as well as energy costs. Therefore, each spot market energy customer pays an energy price that includes the full marginal cost of delivering an increment of energy to the purchaser's location.

- When there is transmission congestion in PJM, one or more of the generating units are dispatched out of economic merit order to keep transmission flows within limits. There may be many resources that are dispatched to relieve the congestion. The LMP reflects the cost of re-dispatch for out-of-merit resources and cost of delivering energy to that location.
- LMPs are calculated at all injections, withdrawals, EHVs (nominal voltage of 500 KV and above), Interfaces, and various aggregations of these points.
- LMPs are calculated in both the Real-time Energy Market and Day-ahead Energy Market.
    - Day-ahead LMP is calculated based on the security-constrained economic dispatch for the Day-ahead Market.
    - The Real-time LMP is calculated based on the approved security constrained economic dispatch solution for the target dispatch interval
- The LMP calculation determines the full marginal cost of serving an increment of load at each bus from each resource associated with an eligible energy offer as the sum of three (3) separate components of LMP. In performing this LMP calculation, the cost of serving an increment of load at each bus from each resource associated with an eligible energy offer is calculated as the sum of the following three components of Locational Marginal Price:
    - System Energy Price – This is the price at which the Market Seller has offered to supply an additional increment of energy from a generation resource or decrease an increment of energy being consumed by a Demand Resource. The System Energy Price may include a portion of the defined reserve penalty factors should a reserve shortage exist.
    - Congestion Price – This is the effect on transmission congestion costs (whether positive or negative) associated with increasing the output of a generation resource or decreasing the consumption by a Demand Resource, based on the effect of increased generation from or consumption by the resource on transmission line loadings. The Congestion Price is set to the specified transmission constraint penalty factor in the event a transmission constraint cannot be controlled below the penalty factor value. The Congestion Price may include a portion of the defined reserve penalty factors should a reserve shortage exist.
    - Loss Price – This is the effect on transmission loss costs associated with increasing the output of a generation resource or decreasing the consumption

by a Demand Resource, based on the effect of increased generation from or consumption by the resource on transmission losses.

- The energy offer or offers that can serve an additional increment of load at a bus at the lowest cost, calculated in this manner, shall determine the Locational Marginal Price at that bus.

5.2.4 Reliability Unit Commitment

After the posting of Day-Ahead Market results, PJM performs a second resource commitment, known as the Reliability Assessment and Commitment (RAC) run, which includes updated resource offers and availability, as well as updated load forecast information, to commit any additional resources needed for reliability. The timing of the RAC relative to the day ahead and real time markets is illustrated in the figure below

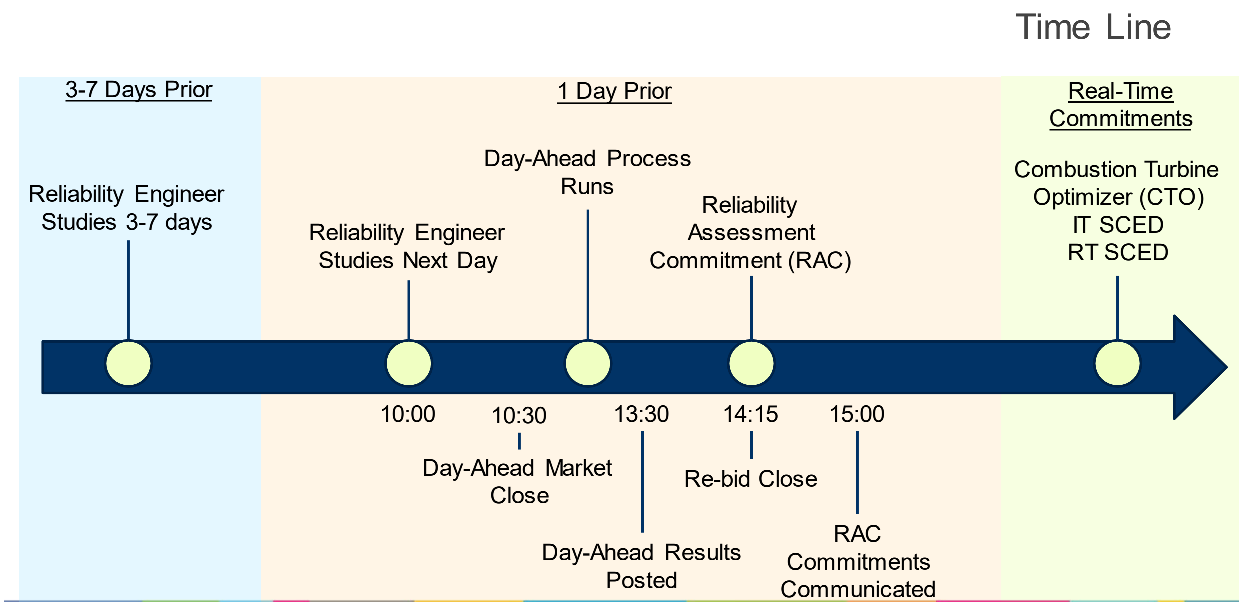


Figure 5-2: Timeline of the energy market and reliability assessments

(Source: https://www.pjm.com/-/media)

Reliability Studies conducted for the next day and for next 3-7 days include

- Power flow studies (full contingency list) are preformed for scheduled transmission outages
  - Thermal constraints
  - Reactive constraints (real time or post contingency voltage)
  - Units are identified based on distribution factors (dfax)
  - For voltage - proximity to the problem, thermal surrogate
  - Extreme weather Hot, Cold, Hurricane, etc.
- If required Long Lead units are called or run through
- Unit with >32 hours total time to start cannot be committed in the DA

- Before the Day-Ahead Market is run, next day commitments are given to the Day-Ahead Market operators

RAC Commitment that follows the Day Ahead market close, is based on an optimization run that utilizes PJM forecast to schedule additional long lead units for reliability concerns at minimum startup and no load cost

- Load is based on PJM forecast
- Interchange is based on PJM forecast
- Energy and Reserve co-optimization
- Focuses on Steam & Combined Cycle Commitments

The Real Time Commitment consists of three parts: Combustion Turbine Optimization (CTO), Intermediate Term Security Constrained Economic Dispatch (IT SCED) and Real Time Security Constrained Economic Dispatch (RT SCED).

The ***CTO*** utilizes updated forecasts and system conditions to schedule additional long lead Combustion Turbines (CT) at minimum total as bid production cost

- Typically run 03:00-07:00 and throughout day if needed
- Same inputs as RAC but updated closer to the peak

Updated load forecast

Updated unit information

- Focus on CT commitment

Used to commit > 2 hour time to start and long minimum run time units

IT SCED and ***RT SCED*** utilize updated forecasts and system conditions to minimize (as bid) production cost while enforcing security constraints and co-optimizing energy and reserve. The IT SCED is run with a 2 hour look ahead and focuses on CT commitment whereas the RT SCED has a 15 minute look ahead and only dispatches online units. The input for these commitment runs is the current system conditions which include:

- Available Short Termed Load Forecast
- Topology
- Generation
- Load
- Interchange
- Operator selected EMS constraints

5.2.5 Uplift and Make whole Payments

Generation units that submit economic bids (as opposed to self-schedule) and are centrally dispatched through the day ahead and real time markets or reliability unit commitment are guaranteed to recover their costs including startup cost, no load cost and fuel cost in every 24 hour period through uplift payments.

5.2.6 Energy Storage Resources (ESR) Participation Model

An Energy Storage Resource (ESR) is a resource capable of receiving electric energy from the grid and storing it for later injection to the grid that participates in the PJM Energy, Capacity and/or Ancillary Services Markets as a Market Participant.

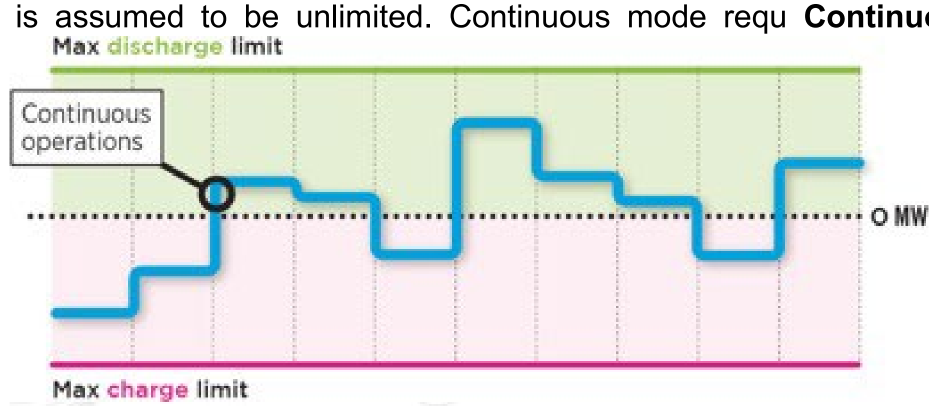
The Energy Storage Resource participation model is an optional model for ESRs to schedule their operation into PJM Markets. Energy Storage Resources participating in the model make their own commitment decisions in Energy and can be dispatchable within their specified operating limits. Energy Storage Resources that elect to be in the ESR participation model cannot also elect to be optimized by PJM in the pumped storage hydro optimizer.

ESR Participation Model Election (i.e. Opt In/Opt Out)

- Resources must opt into the Energy Storage Resource Participation Model by sending a request to Member Relations at custsvc@pjm.com.
- Once a resource opts in for ESR participation, the opt-in status remains until an opt- out request is received.
- Existing resources must send opt-in requests no later than September 30 for the upcoming January 1 to December 31 participation months.
- Resources within the new resource queue process must send an opt in request no later than three (3) months in advance of their initial start in the Energy Markets.
- An opt out request for an existing resource must be sent to Member Relations at custsvc@pjm.com no later than September 30 to remove the resource for the upcoming January 1 to December 31 participation months.

ESR model participants are not optimized for commitment decisions in Day-ahead and Real- time because they are managed directly by participants through specification of the four modes of operation:

Mode – shall mean the mode of operation of an Energy Storage Resource model participant that includes both negative and positive MW quantities (i.e., the Energy Storage Resource model participant is capable of continually and immediately transitioning from withdrawing MW quantities from the grid to injecting MW quantities onto the grid). ESR model participants operating in Continuous Mode cannot specify a ramp rate as it is assumed to be unlimited. Continuous mode requ **Continuous** ires the



Maximum Discharge Limit to be greater than or equal to zero and the Maximum Charge Limit to be less than or equal to zero.

Figure 5-3: Continuous ESR Mode

(Source: https://www.pjm.com/~/media/documents/manuals/m11.ashx)

Charge mode - shall mean the mode of operation of an Energy Storage Resource model participant that only includes negative MW quantities (i.e., the Energy Storage Resource model participant is only withdrawing MWs from the grid). Charge Mode requires that the Energy Storage Resource model participant's Minimum Charge Limit and the Maximum Charge Limit be less than or equal to zero, and the Energy Storage Resource model participant is required to define a ramp rate.

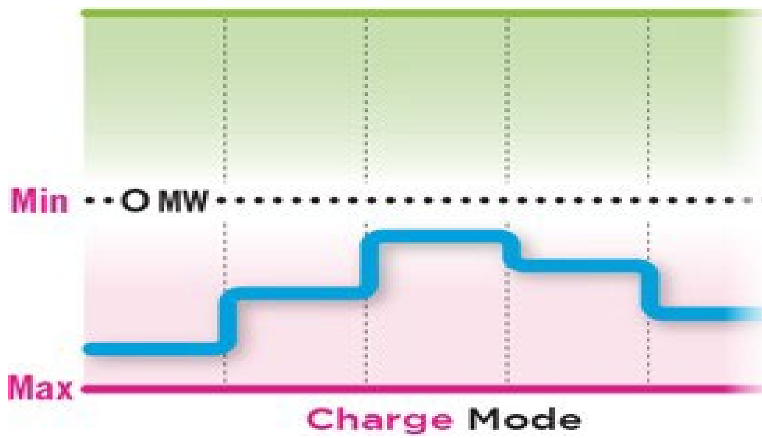


Figure 5-4: Charge Mode

(Source: https://www.pjm.com/~/media/documents/manuals/m11.ashx)

Discharge mode - shall mean the mode of operation of an Energy Storage Resource

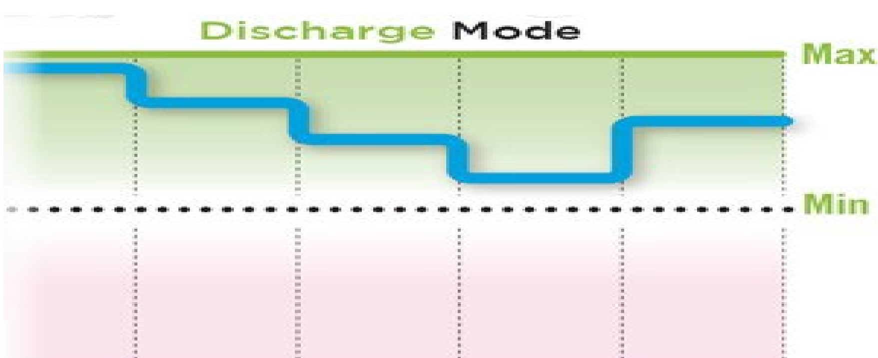


Figure 5-5: Discharge Mode

model participant that only includes positive MW quantities (i.e., the Energy Storage Resource model participant is only injecting MWs onto the grid). Discharge Mode requires the Minimum Discharge Limit and the Maximum Discharge Limit to be greater than or equal to zero. A ramp rate is required in this operating mode.

(Source: https://www.pjm.com/~/media/documents/manuals/m11.ashx**)**

These modes are to be used by both the Day-ahead and Real-time Markets. These modes are to be submitted by the Market Participant on an hourly basis through Markets Gateway by 11:00 the day before the Operating Day for Day-ahead and sixty-five (65) minutes before the operating hour for Real-time.

5.2.7 PJM Energy Market Activities per Manual 11

The following provides the activites of the PJM operator.

- Post on the Markets Gateway System, the PJM load forecast, total bid demand, and Day-ahead Reserve objective for each hour of the next Operating Day by 13:30 at the completion of the Day-ahead scheduling process.
- Post forecasts of total hourly demand for the next four (4) days and peak demand for the subsequent three (3) days.
- Post hourly LMP, Congestion Price, and Loss Price values for the next Operating Day at the completion of the Day-ahead scheduling process by 13:30.
- Post the schedule of demand, supply, and bilateral transactions for private viewing by Market Participants.
- PJM may perform supplemental resource commitments after the Day-ahead schedule is posted in order to maintain reliable operation. Such supplemental commitments are based on minimizing startup and no-load costs.
- PJM may limit its dependence on Combustion Turbines (CTs) to provide reserves in order to maintain reliable operational standards. Such limits shall be based on past performance of these resources.
- Market Operators will commit in the Day-ahead Market any generation resources that were scheduled by PJM dispatch in advance of the Day-ahead Market and are still required for the Operating Day and therefore not cancelled. The scheduled hours for the pre-committed generation resource in the Day-ahead Market will at least include the hours where PJM dispatch has scheduled the resource as well as any additional hours where the resource was deemed to be economic as a result of the Day-ahead Energy Market solution.

5.2.8 Local Market Power Mitigation

PJM mitigates the potential abuse of market power in the energy markets using a structural screen that tests for the concentration of local market power under transmission constrained conditions and applies measures to mitigate such power when detected. If congestion is identified during the Day-ahead scheduling process or during Real-time operations, then generators impacting a congested transmission line that fails the Three-Pivotal Supplier Test ("TPS Test") have their offers capped. Resources remain eligible to set LMP when their offer-are capped. Both pool-scheduled and self-scheduled resources are eligible for offer capping. The Three-Pivotal Supplier Test is performed in IT SCED, as part of the dispatch run. The test checks whether removing the offers of the three generators having the larges impact in producing counterflow on a congested line (accounting for generator capacity and Power transfer distribution Factor (PTDF)), will result in infeasible power flow. When this is the case the congested line fails the test and all generators that can potentially reduce congestion on that line have their offers capped.

Offer-capping is applied as follows:

- Resources are offer-capped at lesser of their cost-based or price-based schedules, including startup and no-load components.
- For resources scheduled in the Day-ahead Market, the offer caps are applied at the time of commitment and apply for the length of time the unit is scheduled in the Day-ahead Market at the schedule that results in the lowest overall system production cost.
    - If the incremental energy offer, no load cost or startup cost for any portion of the offer capped hours is updated subsequent to the Day-ahead commitment, the offer caps are recalculated for each hour that was updated and apply at the schedule that results in the lowest dispatch cost for each updated hour. However, once the resource is dispatched on a cost-based offer, it remains on a cost-based offer regardless of the determination of the cheapest schedule.
- For resources scheduled in the Real-time Market, the offer caps are applied at the time of commitment and apply at the schedule which results in the lowest dispatch cost..
    - If the incremental energy offer, no load cost or startup cost for any portion of the offer capped hours are updated subsequent to the Real-time commitment, the offer caps will be recalculated for each hour that is updated and will apply at the schedule that results in the lowest dispatch cost for each updated hour. However, once the resource is dispatched on a cost-based offer, it will remain on a cost-based offer regardless of the determination of the cheapest schedule.
- Non-CT resources, as well as CTs that are committed in the Day-ahead Market and are expected to run in the Real-time Market without additional notification from PJM Dispatch, that are offer-capped in the Day-ahead Market are offer-capped for those same hours in the Real-time Market and at the same schedule.
- Pool-scheduled CTs that are committed in the Day-ahead Market and are not expected to run in Real-time unless notified by PJM Dispatch and are offer-capped in the Day- ahead Market and are re-evaluated for market power at the time of commitment in the Real-time Market. Such resources are offer-capped in

accordance with the results of the TPS Test that is conducted at the time of the Real-time commitment.

- Pool-scheduled resources brought on-line for economics prior to constrained conditions are not offer-capped at the time of commitment.
- Resources that passed the TPS Test at the time of commitment remain uncapped and are not be subject to additional market power testing until the end of the initial capping determination period, which is defined as follows:
  - For pool-scheduled or self-scheduled resources committed in the Day-ahead Market, the end of their Day-ahead commitment.
  - For pool-scheduled resources committed in the Real-time Market (and not in the Day-ahead Market): the end of their minimum run time.
  - For self-scheduled units committed in the Real-time Market (and not in the Day- ahead Market): the end of the first hour of their commitment.
- Resources running in Real-time beyond the initial capping determination period are subject to evaluation for market power on an hourly basis and are offer-capped as follows:
  - Resources operating on a price-based schedule whose owner pass the TPS Test will not be offer-capped and will remain on the price-based offer.
  - Resources operating on a price-based schedule whose owner does not pass the TPS Test will be offer-capped.
  - Resources operating on a cost-based schedule will remain on a cost schedule regardless of the results of the TPS Test.
- Once a unit is offer-capped in the Real-time Market it shall remain offer-capped until the earlier of:
  - The resource's release from its commitment by PJM Dispatch.
  - The end of the Operating Day.

5.2.9 Market Clearing Engine and Time Line

A multi-module software platform is utilized by PJM to dispatch Energy and ensure adequate Reserves in Real-time and Regulation in near time as shown in Figure 5-6 below. The Real-time Market Clearing Engines and various other applications communicate jointly and the most recent information from each application is stored, and upon request, provided to each application. The Real-time Market, data is processed from the markets database and other PJM systems. The applications jointly optimize the products for a defined target time to ensure that all system requirements are met using the least cost resource, thus minimizing production cost.

The Ancillary Services Optimizer (ASO) performs the joint optimization function of Energy, Reserves and Regulation in the dispatch run. The main functions of ASO are the clearing and commitment of all Regulation resources and inflexible Reserve resources for a one hour time period. The ASO is executed one (1) hour prior to the beginning of an operating hour and is normally solved and approved up to thirty (30) minutes prior to the operating hour. Upon approval, the assignments are posted in the Markets Gateway system. In the event the ASO case is not approved, previous assignments are effective into the next hour. The ASO engine uses the hourly offers for Energy, Reserves and Regulation that are effective at the target time for each case solution and also performs the Regulation Three Pivotal Supplier Test. The ASO does not calculate market clearing prices.

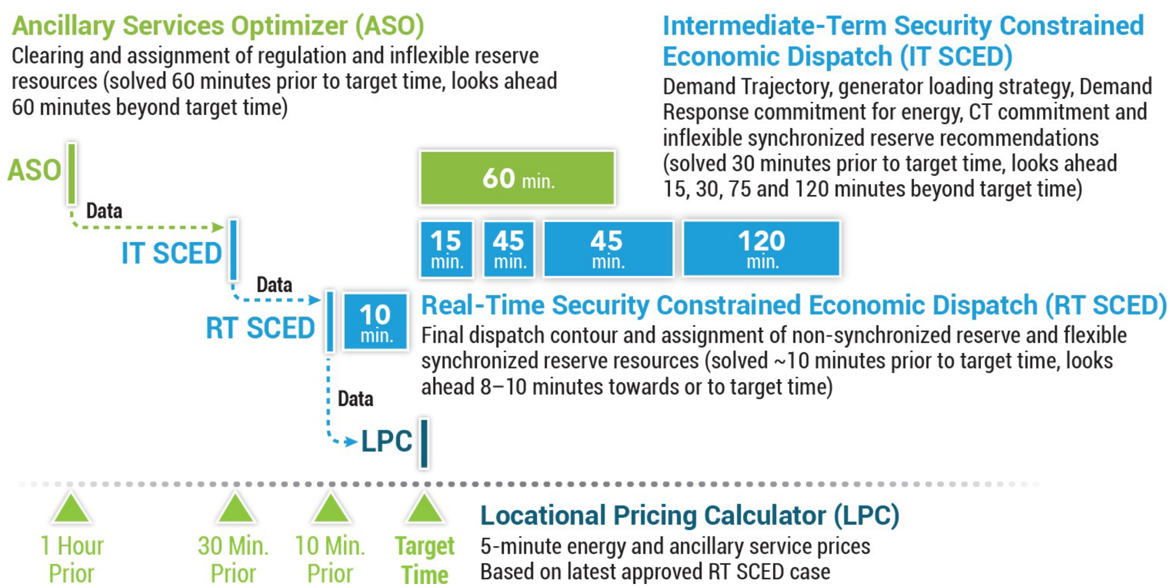


Figure 5-6: Market Operation Software Platform Time Line

(Source: https://www.pjm.com/~/media/documents/manuals/m11.ashx)

5.3 Overview of the PJM Regulation Market

The PJM Regulation Market is an essential component of PJM's operations, ensuring that Regulation ancillary services are procured efficiently and at the lowest cost while maintaining grid reliability. The ASO and RT SCED programs play crucial roles in optimizing the utilization of Regulation resources and other ancillary services to meet the requirements of the electric grid.

The Regulation Market within PJM serves as a market-based system for the buying and selling of Regulation ancillary services. Ancillary services are essential for maintaining the balance between electricity supply and demand, ensuring grid stability, and responding to real-time fluctuations in electricity consumption and generation. Various entities, including resource owners and other market participants, are involved in the Regulation Market. Resource owners submit specific offers for Regulation Capability and

Regulation Performance. PJM utilizes the Ancillary Service Optimizer (ASO), which is an hour-ahead Market Clearing Engine, to optimize the dispatch of Regulation resources. It takes into account Regulation offers, energy offers, and resource schedules as input data. RT-SCED optimizes both energy and reserves in real-time, considering transmission constraints, reserve requirements, and prior committed Regulation resources. It's part of the process to ensure grid security and reliability. PJM calculates the five-minute Regulation Market Clearing Price (RMCP) and Regulation Market Performance Clearing Price (RMPCP) using the Locational Price Calculator. These prices are used to derive the five-minute Regulation Market Capability Clearing Price (RMCCP). These clearing prices play a crucial role in determining payments to providers and charges to purchasers of Regulation services during market settlements.

The total PJM Regulation Requirement for the PJM RTO is determined in whole MW for the ramp and non-ramp periods. Demand Resources are limited to providing 25% of the regulation requirement. PJM aims to achieve the lowest cost alternative for procuring Regulation services for each hour of the operating day. This is done through a simultaneous co-optimization process that considers Synchronized Reserves, Non-Synchronized Reserves, Secondary Reserves, and Energy. The result is an optimized RTO dispatch profile and Locational Marginal Prices (LMPs) for the market hour.

For each eligible resource capable of providing Regulation, an opportunity cost is estimated using the dispatch profile and forecasted LMPs. This cost is adjusted by factors like Performance Score and Benefits Factor. The estimated opportunity cost is added to the adjusted Regulation Capability cost and adjusted Regulation Performance cost, resulting in the adjusted total regulation offer cost. The adjusted total regulation offer cost is used to determine the merit order price. Self-scheduled Regulation resources have a merit order price of zero.

All available regulating resources are ranked in ascending order based on their merit order prices. The lowest-cost set of resources is identified to simultaneously meet the PJM Regulation Requirement, Synchronized Reserve Requirement, Primary Reserve Requirement, 30-minute Reserve Requirement, and provide Energy for a given hour. If there are more self-scheduled and zero-cost offers than needed to meet the Regulation Requirement, resources with the highest historic performance scores are selected as tie-breakers to determine which set of resources to commit.Prices for Regulation are calculated every five minutes, along with Energy and Reserves, using the Locational Pricing Calculator (LPC).

The highest merit order price associated with the lowest cost set of resources awarded Regulation becomes the Five-Minute Regulation Market Clearing Price (RMCP). The RMPCP is determined as the highest adjusted performance offer from the set of cleared resources. The RMCCP is calculated as the difference between RMCP and RMPCP. Resources that self-scheduled to provide Regulation are compensated ex-ante based on a processes outlined in the PJM Manual on Operating Agreement Accounting.

These processes ensure the efficient procurement of Regulation services while considering factors like cost, performance, and historic performance scores. The aim is to maintain grid stability by selecting the most cost-effective resources to meet the various requirements in real-time.

5.3.1 Regulation Market Eligibility and Offers

Regulation offers may be submitted only for those resources electrically within the PJM RTO. To regulate, a resource must meet the following criteria:

- Generation resources as well as Demand Resources must be able to provide 0.1 MW of Regulation Capability in order to participate in the Regulation Market. Generation resources must have a governor capable of automatic generation control (AGC).
- Generation and Demand Resources must be able to receive and respond to an AGC signal. A resource's MW output must be telemetered to the PJM control center in a manner determined to be acceptable by PJM.
- Regulation providers are pre-certified for responding to type A or type D regulation signals (or both) and submit offers accordingly for the regulation type they are certified to provide . The Regulation D signal is a fast, dynamic signal that requires resources to respond almost instantaneously. Regulation A is a slower signal that is meant to recover larger, longer fluctuations in system conditions. These two signals communicate with each other and work together to match the system need for regulation. Regulation resources that are dual certified as RegA and RegD may submit a set of offers for each signal type. In such case, the Market Clearing Engine evaluates both offers but will clear the resource for either one or neither of the two signal types based on economics and system needs
- Resources should give priority to the regulation signal by not allowing the sum of the regulating ramp rate and energy ramp rate to exceed the economic ramp rate. A generator providing regulation may follow the energy dispatch signal only after accounting for the regulation capability.
- When a Demand Resource that is eligible for the Regulation Market is called for a mandatory Emergency or Pre-Emergency Load Management Event, it is de-assigned from Regulation for any intervals that overlap with the Load Management Event and PJM will not assign the resource to Regulation for the remainder of the mandatory portion of the Load Management Event.
- Regulation offers may be either **Cost Based** or **Price Based**
  - Cost-Based Regulation Offer ($/MWh): This value must be validated using the unit-specific operating parameters submitted with the regulation offer and the applicable $12/MWh regulation margin adder.
  - Price-Based Regulation Offer ($/MWh, optional): This value is capped at $100/MWh, and its submission is optional on the part of the market participant.

In both cases offers are split into components

- Regulation Capability portion capturing the Fuel Cost Increase and Unit Specific Heat Rate Degradation due to Operating at Lower Loads and VOM Rate [$/MWh of Regulation]. In the case of cost based offers, the margin adder may only be added to this component
- Regulation Performance portion representing Cost Increase due to Heat Rate Increase during non-steady state operation or the resource owner's price to provide regulation movement in $/ΔMW. The ΔMW value represents the sum of absolute value of MW movements (up and down)

In the case of cost-based regulation offer price, each market participant may also submit additional information to support the cost-based offer price including Heat Rate at the economic maximum [BTU/kWh], Heat Rate at the default regulation minimum for a resource [BTU/kWh], Variable Operation and maintenance (VOM) rate increase [$/MWh of Regulation], Fuel Cost [$/MBTU .PJM validates the cost-based regulation offer price to ensure that it does not exceed actual regulating cost. An example of this calculation is available on the PJM website in the 'Regulation Two Part Cost-Based Offer' document, located at https://www.pjm.com/-/media/markets-ops/ancillary/regulation-two-part-cost-based-offer-effective-20181204.ashx?la=en. If a market participant does not submit a cost-based regulation offer price they are not permitted to participate in the PJM Regulation Market until such offer has been validated. Any participants that do not submit any of the supporting parameters will have their cost-based regulation offer price capped at the margin adder of $12/MWh

5.3.2 Operation of the Regulation Market

The PJM dispatcher regularly assesses the resources providing regulation services to the power grid. Regulation involves maintaining a balance between electricity supply and demand to stabilize the grid frequency. The PJM dispatcher's goal is to minimize the overall cost of regulation. To achieve this, they make adjustments to the regulation assignments of resources as needed. When there is an excess of regulation capacity, the PJM dispatcher starts deselecting resources. They begin with the highest cost resources providing regulation and work their way down. If there is a deficiency in regulation capacity the PJM dispatcher uses IT SCED to select resources. They start with the lowest cost resources that are not currently providing regulation and increase regulation capacity.

The costs associated with regulation resources, such as Regulation Market Clearing Price (RMCP), may change based on the adjustments made in real-time. The PJM Energy Management System (EMS) communicates with Local Control Centers (LCCs) and individual resources or plants as needed. Changes in resource regulation assignments are communicated to LCCs.

The total regulating capabilities of the company's resources are monitored and sent back to the PJM EMS through telemetering methods like the PJM data link. During transitions between on-peak and off-peak periods, resource regulation assignment changes begin

30 minutes before the new period and are completed no later than 30 minutes after the period begins. This ensures a smooth transition. Resources that are qualified to provide both A and D type regulation services cannot have their signal assignment change within the same operating hour so they continue with the regulation signal type they were initially committed to.

5.4 Overview of the PJM Reserve Markets

The PJM Reserve Markets provide PJM participants with a market-based system for the purchase and sale of the Synchronized Reserve, Primary Reserve, and 30-minute Reserve Services (“Reserves”). The PJM Reserve Markets are conducted in both the Day-ahead Market and Real-time Market processes. In the Day-ahead Market, PJM schedules Reserves on a simultaneously, co-optimized basis with Energy for each hour of the next Operating Day. In Real-time, PJM procures Reserves on a simultaneously, co-optimized basis with Energy for each hour and each interval.

Both the Day-ahead and Real-time Reserve Markets are offer-based and procure resources to meet the required Reserve Services:

- Synchronized Reserve Service: can only be satisfied by online resources that are able to respond in ten (10) minutes or less.
- Primary Reserve Service: can be satisfied by online or offline resources that are able to respond in ten (10) minutes or less.
- 30-Minute Reserve Service: can be satisfied by online or offline resources that are able to respond in thirty (30) minutes or less.

There are three reserve products that can meet these required Reserve Services:

- Synchronized Reserve Product: online resources that are able to respond in ten (10) minutes or less.
- Non-Synchronized Reserve Product: offline resources that are able to respond in ten (10) minutes or less.
- Secondary Reserve Product: online or offline resources that are able to respond between ten (10) and thirty (30) minutes.

The Synchronized Reserve, Non-Synchronized Reserve and Secondary Reserve products have a priority sequence based on the level of reliability which each provides. Synchronized Reserve, being the most reliable as it is online and can respond in ten (10) minutes or less, can also meet the Primary and 30-Minute Reserve requirements. Likewise, Non-Synchronized Reserve can also meet the 30-Minute Reserve requirement. The co-optimization of these reserves which reflects this hierarchy guarantees that the clearing prices of less reliable reserve product will not exceed that of a more reliable one so the price of each reserve product are consistent with their priority ranking.

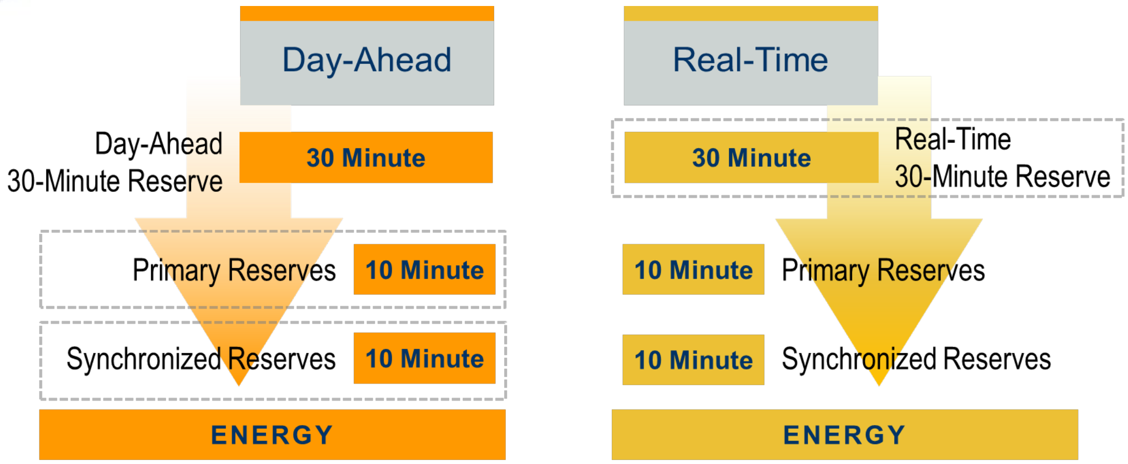


Figure 5-7: Day-ahead and Real-time Reserve Services

(Source: https://www.pjm.com/~/media/documents/manuals/m11.ashx)

Each Load Serving Entity (LSE) on the PJM system has a Reserve Obligation in kWh based on their Real-time load ratio share and the procured supply to meet each Reserve product. The total resrves requirement for each reserve type is characterized by an operating reserves demand curve (ORDC) which determines the scarcity adder to the energy price when reserves falls below a minimum required level

The following subsections apply to both the Day-ahead and Real-time Market, unless specifically stated.

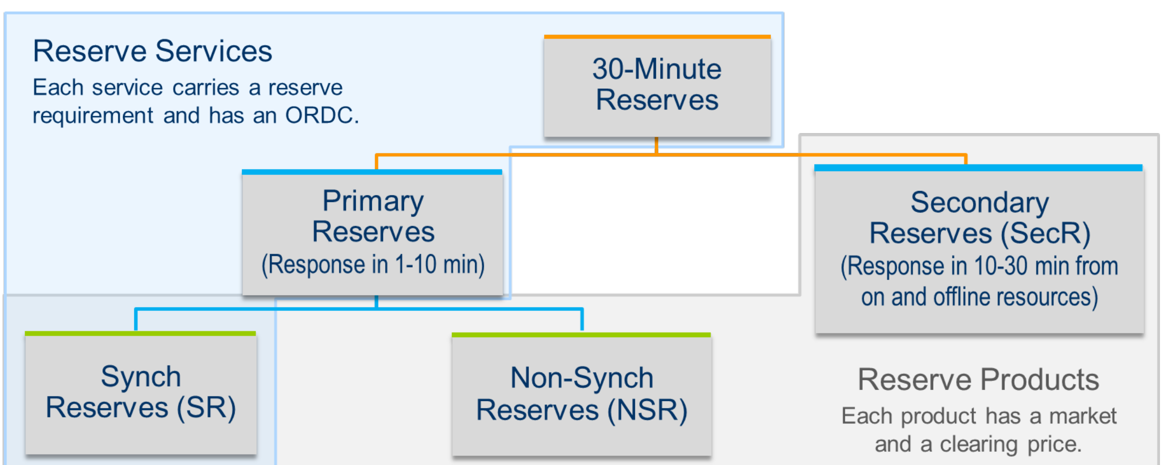


Figure 5-8:_Reserve Services and Corresponding Products

(Source: https://www.pjm.com/~/media/documents/manuals/m11.ashx)

5.4.1 Reserve Market Eligibility

In general, with few exceptions, Generation, Hybrid Resources, Energy Storage Resources, and Economic Load Response resources are eligible to provide Synchronized Reserves, Non-Synchronized Reserves, and Secondary Reserves. However, Economic Load Response, Hybrid Resources, Energy Storage Resources enrolled in the ESR participation model and Pumped hydro resources that are not participating in the PJM optimized pumped storage model in the Day-ahead market are not eligible to provide Non-Synchronized Reserves.

Nuclear, Wind, or Solar resource must obtain approval to be considered eligible to provide reserves by submitting to PJM and the Market Monitoring Unit (MMU) a written request for such approval and provide documentation to support the resource's ability to follow dispatch at the direction of PJM, such as historical operating data showing voluntary response to reserve events and/or technical information about the physical operation of the resource.

Generation resources must be able to provide 0.1 MW of Reserve Capability in order to participate in the Reserve Markets. Furthermore in the event PJM forecasts a credible natural gas pipeline contingency(s), PJM Dispatch will determine the eligibility of resources to provide Reserves depending on the severity of the contingency and other system conditions in after order to ensure system reliability is maintained.

5.4.2 Economic Load Response Eligibility

Economic Load Response must successfully complete Ancillary Services certification in PJM's DR Hub system and must be able to provide 0.1 MW of Reserve Capability in order to participate in the Reserve Markets. Furthermore, Economic Load Response providing reserves are required to provide -the-fact meter data at a one (1) minute interval for each location called on to respond in a reserve event. Residential locations without meters recording at a one (1) minute interval or shorter may participate using a statistical sampling method subject to PJM approval.

Economic Load Response Curtailment Service Providers not providing complete, accurate and timely meter data for locations called on to respond in a reserve event may be suspended from participating in the Reserve Markets until corrective measures are implemented and may be referred to the PJM Market Monitor and/or the FERC Office of Enforcement for further investigation as necessary.

Whenever Economic Load Response assigned in the Reserve Markets is called on to respond to a mandatory Emergency or Pre-Emergency Load Management Event, it will be de-assigned from Reserves for any intervals that overlap with the Load Management Event, starting from the notice time of the Load Management Event, unless otherwise approved by PJM. PJM will not assign the resource to Reserves for the remainder of the mandatory portion of the Load Management Event. Economic Load Response that demonstrate the ability to receive reserve commitments via approved telemetry (e.g. Jetstream) qualify as a Flexible resource. Otherwise, Economic Load Response resources are regarded as Inflexible resources.

5.4.3 Reserve Must Offer Requirements.

Any generator that is a PJM generation capacity resource that has a Reliability Pricing Model (RPM) or Fixed Resource Requirement (FRR) Resource commitment that is eligible to provide Reserves **must offer** their 10-minute and 30-min reserve capability, unless the unit is unavailable due to an approved planned outage, maintenance outage or forced outage.

If a resource that has a reserve **must offer** requirement chooses to not make its reserve capability available, for example through self-scheduling or offering a fixed output, when the resource is otherwise able to operate with a dispatchable range, the resource is defined to be violating the reserve must offer requirement.

All other generation resources that are eligible to provide reserves that have submitted Energy offers are considered to have offered the unit's applicable capability into the reserve markets. However, Hydroelectric, Economic Load Response, Hybrid Resources and Energy Storage Resources (ESR) are not considered available by default, and must submit specific reserve offers to be considered. Hydroelectric, Economic Load Response, Hybrid Resources and Energy Storage Resources reserve offers must be based on their realistic achievable ramp rate and current operating conditions taking into account all constraints (e.g. regulatory, environmental, operational).

For multiple physical units that are modeled as an aggregated resource in the Market system, if at least one of the physical units are online, then the aggregate is considered online and may be considered for online synchronized and/or online secondary reserves, but not eligible for non-synchronized reserve. Any offline units belonging to an online aggregated resource that have the capability of providing reserves are subject to the must offer requirement and PJM will treat the offline reserves as synchronized reserves. Hydroelectric, Hybrid Resources and Energy Storage Resources must include the available offline reserves in their synchronized reserve offer.

5.4.4 Reserve Market Resource Offer Structure

All generation resources that have submitted energy offers and are eligible to provide reserves, will be considered as offered into the Reserve markets. This excludes hydropower resources, Hybrid Resources and Energy Storage Resources who must submit specific Reserve offers to be considered. Reserve offers consist of three elements: availability, offer MW, and offer price and vary depending on the type of resource and the market in which the resource is participating.

PJM will calculate the Reserve MW quantity available from each generation resource, not including ESR, Hybrid Resources and Hydroelectric resources, based on the bid in energy parameters, reserve parameters, Regulation status and current energy output data. Other resources, may specify offer MW values as described below, where:

- Synchronized Reserve offer MW is reserve capability of a resource that can be converted fully into energy in ten (10) minutes, or the load reduction achievable in ten (10) minutes and is provided by equipment electrically synchronized to the system.
- Non-Synchronized Reserve offer MW is reserve capability of a resource that can be fully converted into energy within ten (10) minutes and is provided by equipment not electrically synchronized to the system.
- Secondary Reserve offer MW is reserve capability of a resource that can be converted fully into energy after ten (10) minutes and before thirty (30) minutes, or the load reduction achievable after ten (10) minutes and before thirty (30) minutes by equipment which may not necessarily at the time of the request be electrically synchronized to the system.

The ability to specify an Offer MW varies by resource type.

As to offer price, all resources may specify a Synchronized Offer Price ($/MWh) that must be cost-based and capped at the Expected Value of Synchronized Reserve Penalty. Resources listed as available for Synchronized Reserves with no Offer Price amd all Non-Synchronized Reserve and Secondary Reserve offer prices are set to $0.00/MWh.

5.4.5 Reserve Requirement DeterminationStructure

PJM models a reserve requirement at the RTO and sub-zonal level in whole MW for each hour of the operating day based on the greatest MW loss of all potential Largest Single Contingencies on the system. The table below describes the reliability and reserve requirements for each Reserve Service

In order to meet Regional Reliability Criteria, PJM may schedule additional Contingency Reserves on a temporary basis in order to meet the Largest Single Contingency, as necessary to account for resource performance. PJM shall post details regarding additional scheduling of reserves in Markets Gateway. The Largest Single Contingency in Day-ahead is normally the largest Economic Maximum value for all available schedules or the summation of the largest Economic Maximum value for all available schedules of an active reserve group for the hour. The Largest Single Contingency in Real-time is normally the higher of [max of (the largest online generator's output or Economic Maximum) or the sum of the higher of (Economic Maximum values or outputs of an active reserve group)].

Table 5-1: Reserve Service Determination

(Source: https://www.pjm.com/~/media/documents/manuals/m11.ashx)

	Reserve Service		
	Synchronized Reserve (SR)	Primary Reserve (PR)	30-Minute Reserve (30-Min)
Reliability Requirement	Largest Single Contingency	150% of Synchronized Reserve Reliability Requirement	Greater of (Primary Reserve Reliability Requirement, 3000 MW, or largest active gas contingency)
Reserve Requirement	SR Reliability Requirement + Extended Reserve Requirement	PR Reliability Requirement + Extended Reserve Requirement	30-Min Reliability Requirement + Extended Reserve Requirement

An active reserve group is a model of a station with multiple generation resources with a total capacity in excess of 800 MW, where there is a single outlet or where a single fault would trip multiple generation resources at the station. For purposes of the 30-Minute Reserve Requirement, the largest gas contingency is calculated as the summation of the Economic Maximum values of the identified resources. Only those potential Largest Single Contingencies communicated by PJM Operations and modeled in the market clearing software will be eligible to set the applicable reserve requirements used in the market clearing process. At times, anticipated heavy load conditions may result in PJM operators carrying additional reserves to cover increased levels of operational uncertainty. PJM may extend the 30-Minute Reserve, Primary Reserve and Synchronized Reserve Requirements in the Market Clearing Engine during the on-peak period in order to incorporate these actions in Energy and Reserve Pricing when a Hot Weather Alert, Cold Weather Alert or an escalating emergency procedure has been issued for the Operating Day.

The extended Synchronized Reserve Requirement, extended Primary Reserve Requirement and extended 30-Minute Reserve Requirement will be equal to the existing extended applicable Reserve Requirement plus the sum of any additional MW brought online for that hour by PJM dispatch to account for operational uncertainty. Each Reserve Requirement will have an associated reserve demand curve These demand curves are used to articulate the value of maintaining reserves at specified levels and ensure product substitution between energy and reserves up to a specified penalty factors.

Due to transmission security considerations on the PJM system, it is necessary to carry a minimum amount of Synchronized Reserve, Primary Reserve, and 30-Minute Reserve in each specific sub-zone in PJM. The main goal of procuring locational reserves is to not overload critical transmission constraints when reserves are deployed. Sub-zone list that are updated periodically can be found at https:// www.pjm.com/markets-and-operations/ancillary-services, As system conditions dictate, PJM may need to model new sub-zones into the Reserve Markets to better support reliable operations and produce market results that are more consistent with system operating conditions. New reserve sub-zones may be defined for different constraints categories

5.4.6 Reserve Market Clearing

PJM schedules resources as needed to meet the Reserve Requirements of each Reserve Zone and active sub-zone via co- optimization with energy in both the Day-ahead and Real-time markets. Resources are scheduled based on the resource-specific offer data submitted and the product substitution cost of providing energy or any other product the resource is capable of providing. The joint optimization seeks to procure and minimize the total production cost of energy while meeting the various reserve requirements. Reserves and energy will be co-optimized the same way in the Day-ahead and Real-time Market. The same reserve zone configuration is modeled in the Day-ahead and Real-time markets unless there is an operational emergency requiring it to be changed in the Real-time market

As described above, Synchronized Reserve, Non-Synchronized Reserve and Secondary Reserve products have a priority sequence based on the level of reliability which each provides. Synchronized Reserve, being the most reliable as it is online and can respond in ten (10) minutes or less, can also meet the Primary and 30-Minute Reserve requirements. Likewise, Non-Synchronized Reserve can also meet the 30-Minute Reserve requirement. In addition, the location of the reserves procured also have a hierarchy. Reserves procured in a sub-zone can also meet the requirement of the RTO. As a result, a megawatt cleared in a subzone for Synchronized Reserves can also be used to meet the RTO Synchronized Reserve Requirement, the subzone Primary Reserve Requirement, the RTO Primary Reserve Requirement, the subzone 30-minute Reserve Requirement (if modeled), and the RTO 30-minute Reserve Requirement. PJM will commit the most economic combination of resources to simultaneously meet all energy and reserve requirements while accounting for product and locational substitutability illustrated in Figure 5-9 below. This results in cascading clearing prices with higher products and location in the hierarchy entail a higher price.

The Day-ahead Reserve Market clearing results in an hourly price for Synchronized Reserves, Non-Synchronized Reserves, and Secondary Reserves for the next day, and is posted along with the resource-specific reserve assignments from the dispatch run. The hourly Reserve Clearing Prices are fixed once calculated and posted the day before the Operating Day. The hourly Reserve clearing prices are based upon the offer prices submitted by the selected resources, together with the summation of the applicable product shadow prices in the Day-ahead joint-optimization process, from the pricing run. Sixty (60) minutes prior to the operating hour PJM executes the Ancillary Services Optimizer (ASO). The ASO jointly optimizes Energy, Synchronized Reserves, Non-Synchronized Reserves, Secondary Reserves, and Regulation based on forecast system conditions to determine an economic set of inflexible reserve resources to commit for the operating hour. Inflexible resources are defined as those resources that physically require an hourly commitment due to minimum run time constraints or staffing constraints. Inflexible resources include:

- Synchronous condensers that are operating in condensing mode solely for the purpose of providing Synchronized Reserves
- Energy Storage resources enrolled in the ESR participation model (unless the resource has elected to be flexible)
- Economic Load Response that are prepared to curtail in response to a PJM Reserve Event (unless the resource has elected to be flexible)

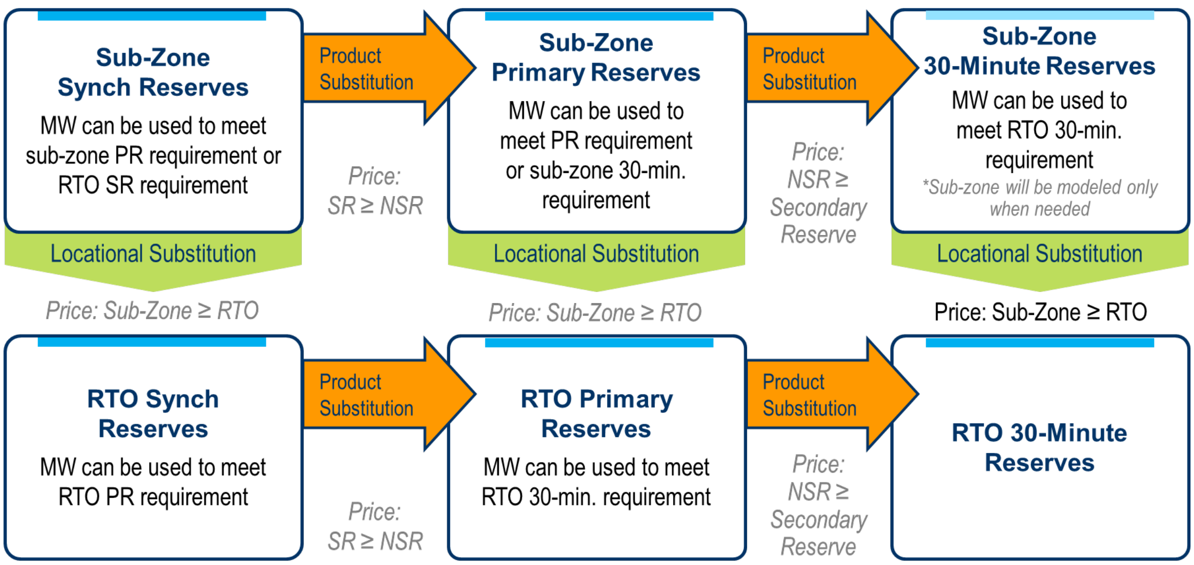


Figure 5-9: Product and location substitutability for reserves

(Source: https://www.pjm.com/~/media/documents/manuals/m11.ashx)

Additional Real-time Reserve commitments may be made on flexible reserve resources by the RT SCED application and additional inflexible reserves resources recommendations by the IT SCED application. Commitments on flexible reserves resources may change with each execution of the RT SCED application while commitments on inflexible reserve resources will respect the minimum run time of those resources. PJM Operator, if necessary, may manually request an inflexible resource previously committed for reserves to provide energy. The Real-time Security Constrained Economic Dispatch (RT SCED) program jointly optimizes the remaining RTO Reserve needs simultaneously with Energy while honoring effective regulation assignments.

Clearing Prices (SRMCP), Non-Synchronized Reserve Market Clearing Prices (NSRMCP), and Secondary Reserve Market Clearing Prices (SecRMCP) are used for market settlement. During each execution of RT SCED, additional Reserve MWs may be committed to meet the Reserve Requirements from flexible resources for all services based on forecasted system conditions by re-dispatching online generating resources. In addition, RT SCED will commit offline resources to meet the balance of the Primary Reserve Requirement and 30-minute Reserve Requirement.

The Real-time Market Reserve clearing prices are set based upon the offer prices submitted by the selected resources, together with the summation of the applicable product shadow prices in the joint-optimization process from the pricing run. Resources cannot clear the Real-time Reserve and Regulation Markets for the same interval. The requirement for regulation is first met before reserves because regulation is a higher priority service.

5.5 Overview of Demand Resource participation

The integration of Demand Response (DR) into the PJM Markets recognizes the importance of Load Response to a fully functioning market, the effect of Load Response on the reliability of the grid and on **the ability of the system to decarbonize the electricity supply through the use of renewable resources**.

5.5.1 DR Participation through Curtailment Service Providers (CSPs)

The purpose of these rules is to enable Demand Resources under the direction and control of Curtailment Service Providers (CSPs) to participate in the various PJM markets. CSPs are Members or Special Members of PJM that participate in the PJM Markets by mobilizing Demand Resources to reduce demand.

PJM Emergency or Pre-Emergency Load Response enables Demand Resources that reduce load during emergency or pre-emergency conditions to receive payment for those reductions. Demand Resources can participate in Emergency Load Response as Energy Only, Capacity Only or both Energy and Capacity Resources, and ar being compensated accordingly.

PJM Economic Load Response enables Demand Resources to respond to PJM Energy, Synchronized Reserve, and/or Secondary Reserve prices by reducing consumption and receiving a payment for the reduction or following PJM signal to reduce or increase load if providing regulation services.

- The Day-ahead Option provides a mechanism by which any qualified Market Participant may offer Demand Resources the opportunity to reduce the load they draw from the PJM system in advance of Real-time operations and receive payments based on Day-ahead LMP for the reductions.
- The Real-time Option provides a mechanism by which any qualified Market Participant may offer Demand Resources the opportunity to commit to a reduction and receive payments based on Real-time LMP for the reductions.

Demand Resources may be registered simultaneously as Economic Load Response Resources and Emergency or Pre-Emergency Load Response Resources. However, Energy Settlements shall be limited to demand reductions that are executed in response to the Real-time and/or Day-ahead LMP or as dispatched by PJM and that are not implemented as part of normal operations.

Emergency and Pre-Emergency resource offer price may not exceed the following:

- 30 minute lead time: $1,000/MWh, plus the applicable Primary Reserve Penalty Factor from the first step of the demand curve, minus $1.00.
- Approved 60 minute lead time: $1,000/MWh, plus the applicable Primary Reserve Penalty Factor from the first step of the demand curve divided by 2.
- Approved 120 minute lead time: $1,100/MWh.

Qualified CSPs may also offer the load reductions of demand resources into the Day-ahead and/or Real-time Energy Market pursuant to the PJM Manuals, Markets Gateway User Guide, and additional rules and requirements.

CSPs that would like to participate in the Energy Market shall submit a bid for each Demand Resource which includes: Transmission zone and pricing point based on where the Demand Resource is located and the associated pricing point used to settle the load in the retail market and/or as defined by PJM.

Hourly Incremental Offer curve (minimum increments of 0.1MW) may represent up to ten (10) combinations of MW load reduction and offer price. This determines the price offered into the Day-ahead Market for respective MW amount in each hour and the price offered for dispatch in the Real-time Market. CSPs may only submit an energy offer greater than $1,000/MWh for an Economic DR resource if the CSP has verified that the end use customer's incremental cost for each offer is greater than $1,000/MWh pursuant to the DR Validation Process. The offer must be equal to or less than the end use customer(s)' incremental cost. CSP verified energy offers greater than $1,000/MWh and less than or equal to $2,000/MWh are eligible to be used in the calculation of the applicable LMP as defined in the Tariff. Energy offers with CSP verified incremental costs greater than $2,000/MWh are compensated through Operating Reserves. Demand Resources are eligible to set Day-ahead and Real-time Energy market prices if selected as the marginal resource

CSPs who manage the load curtailments must be registered with PJM and provide information about each Demand Resource that they manage, including its location the load reduction method and the associated load reduction kilowatt capability. Load reduction methods indicate the type of electrical equipment that is controlled to provide the demand response activity and include: Heating, Ventilation and Air Conditioning (HVAC), Lighting, Refrigeration, Manufacturing, Water Heaters, Batteries, Plug Load and Generation, Locations that use generators, in whole or in part as a load reduction method, shall provide PJM with the primary fuel type used for each generator which includes: Coal, Diesel, Natural Gas, Oil, Gasoline, Kerosene, Propane, Wood, Landfill Gases and Waste products. In cases where the On-Site Generator has a mixed fuel type, CSPs should report on the primary fuel source as the On-Site Generator fuel type.

Compensation of DR based on the LMP is subject to a Net Benefit Test as mandated in

FERC Order 745. The Net Benefits Threshold (NBT) is the point on the aggregate supply curve at which the participation of DR resources results in a greater overall savings to the load on the system compared to the DR resources remaining on the system as load. PJM shall compute the NBT following an approved procedure and post the NBT and associated supporting information for each month by the 15th of the prior month on pjm.com. CSPs only receive compensation for Demand Resources cleared in Day-ahead Market or dispatched by PJM in the Real-time Market if the applicable LMP is greater than or equal to the monthly NBT level.

Demand Resources must be equipped with interval meters recording electrical usage at the EDC account level. The interval of data collection must be sufficient to provide PJM with hourly, one minute or real time load data as applicable for the wholesale market. Residential Direct Load Control (RDLC) aggregates may have interval meters installed on a statistical sample of EDC accounts per PJM Manual 19: Load Forecasting and Analysis, Attachment C, and subject to PJM approval. Any CPS submitting a settlement for load reduction made in the Energy Market that is not metered directly by PJM is responsible for uploading the appropriate meter data into PJM's DR Hub system within sixty (60) days of the reduction.

Load data must be provided for all hours of the day and for all days necessary for PJM to calculate the Customer Base Line (CBL) for settlements or to measure compliance as necessary. When On-Site Generation is used solely to enable the Participant to provide demand reductions then the CSP may provide qualified meter generation output data, upon approval by PJM, from the On-Site Generator for each hour of the event day instead of actual load metered data. Provision of hourly meter data from the On-Site Generator is deemed a certification by the CSP that the On-Site Generator was not used for any purpose other than to support the load reduction during the event day.

5.5.2 Participation by Price Responsive Demand (PRD)

The implementation of dynamic and time-differentiated retail rates, along with investments in Advanced Metering Infrastructure (AMI), has allowed consumers to modify their electricity consumption patterns in response to changing wholesale electricity prices. AMI enables more accurate tracking of electricity usage and facilitates the implementation of dynamic retail rate structures. Consumers can now adjust their electricity consumption based on real-time price signals without being centrally dispatched by PJM or participating in PJM markets. This price responsiveness by consumers is known as Price Responsive Demand (PRD).

PRD is predictable, as it involves consumers voluntarily reducing or shifting their electricity usage in response to price fluctuations in the wholesale market. The relationship between dynamic retail rate structures and wholesale prices is crucial, as it influences consumer behavior. Given the predictable nature of PRD, wholesale market designs and operations need to account for it. The wholesale market must consider the impact of PRD on demand and pricing.

PRD can help reduce the need for excess installed capacity to meet reliability standards, such as Loss of Load Expectation (LOLE). During capacity emergencies or price spikes, PRD can lead to a drop in demand, helping to stabilize the grid.. in sum, PRD is a mechanism that leverages consumer response to dynamic retail rates and wholesale electricity prices to manage electricity demand more efficiently, reduce the need for excess capacity, and enhance grid reliability. A PRD Provider is a PJM Member representing retail customers capable of reducing their electricity consumption in response to price signals. PRD Providers must meet eligibility requirements to offer PRD services on behalf of retail customers. Coordination between wholesale markets, retail rate design, and PRD providers is essential for maximizing its benefits to consumers and the overall electricity system.

5.6 Market Clearing and Scheduling Process

PJM schedules resources based on economics to control potential transmission limitations that are binding in the Transmission Reliability analysis that is performed in parallel with and subsequent to the Day-ahead Market analysis. The scheduling process evaluates the price of each available resource compared with every other available generating resource. The process for scheduling the PJM RTO requires: The scheduling objective varies depending on the scheduling time horizon (Day-ahead, Reliability Assessment and Commitment (RAC) Run, Real-time Operations). The PJM scheduling process in the Day-ahead Energy Market is to schedule generation to meet the aggregate Demand bids that result in the least-priced generation mix, while maintaining the reliability of the PJM RTO. During the RAC Run, PJM schedules additional resources as needed to satisfy the PJM Load Forecast and the Operating Reserve objective based on minimizing the scheduling cost.

In the Day-ahead Energy Market, the goal is to schedule sufficient generation to cover aggregate demand bids and Day-ahead Reserve requirements calculated as a function of such demand bids. Subsequently In the Reliability Assessment and Commitment (RAC) Run (subsequent to the Day- ahead Energy Market), PJM schedules addition generation to cover the PJM Load Forecast and Operating Reserve requirements.

In Real-time Operations, sufficient generation is scheduled to control potential transmission limitations that are binding in the Transmission Reliability analysis, Sufficient generation is scheduled to satisfy the PJM Regulation Requirement, Primary Reserve Requirement, and other ancillary service requirements of the PJM RTO.

Figure 5-10 below illustrates the components of the requirement met by the Market scheduling. Market scheduling Scheduling of resources is performed economically on the basis of the prices and operating characteristics offered by the Market Sellers, using security constrained economic dispatch and continuing until sufficient generation is dispatched in each five minute interval to serve all energy purchase requirements, as well as the PJM RTO requirements

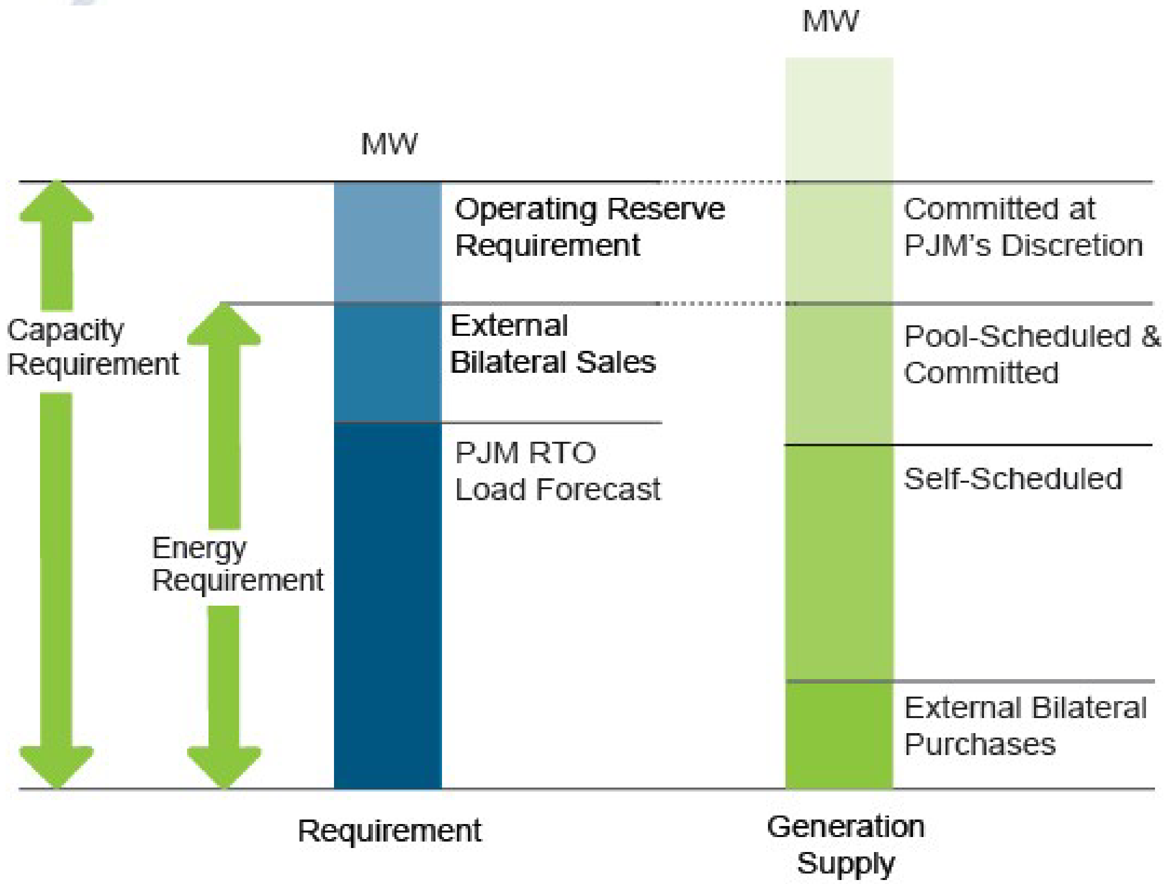


Figure 5-10: Requirements met by Market Scheduling

(Source: https://www.pjm.com/~/media/documents/manuals/m11.ashx)

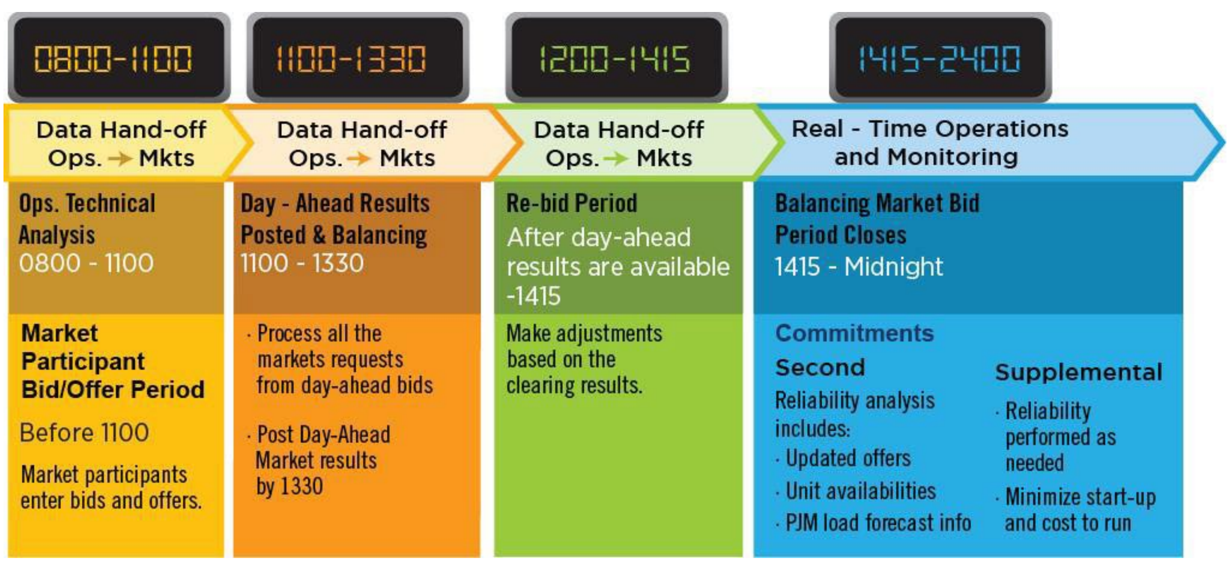


Figure 5-11: Energy Market Daily

(Source: https://www.pjm.com/~/media/documents/manuals/m11.ashx)

Figure 5-11 above illustrates the time line for posting of data and executing the scheduling process by employing an array of analytical scheduling tools. These tools permit PJM scheduling staff to analyze numerous scheduling scenarios and schedule the resources for short-term, hourly and daily activities.

The Markets Database is a very large database that contains information on each generating resource that operates as part of the PJM Energy Market, Economic Load Response information, Demand Information, Virtual Bids, Regulation Offers, Reserve Offers, Day-ahead Energy and Reserve Market Clearing Prices, Regulation Market Clearing Prices and Real-time Reserve Market Clearing Prices. Market Participants may access the Markets Database by using the PJM Markets Gateway Web site via the Internet using manual entry or bulk upload/download via XML format.

The Energy Market technical software develops the Day-ahead Market results so as to minimize production cost of energy and reserve to meet the demand bids and decrement bids that are submitted into the Day-ahead Market while respecting the PJM RTO security constraints and reliability requirements that are necessary for the reliable operation of the PJM RTO.

Subsequent to the close of the generation Re-bidding Period at 14:15, steam resource commitment status may be changed in the RAC run so as to minimize the additional startup costs and costs to operate steam resources at economic minimum in order to provide sufficient operating reserves to satisfy the PJM Load Forecast and adjusted Day-ahead Reserve requirements. Combustion Turbine (CT) resources are included in the scheduling process and are scheduled in the Day-ahead Market. However, the decisions concerning actual operation of pool-scheduled CT resources during the Operating Day are not made until the current operating hour in Real-time dispatch.

5.6.1 Calculation of Day-ahead Prices

PJM's Day-ahead Energy Market uses security constrained economic dispatch optimization software to determine the least-cost means of procuring supply to meet demand and meet Day- ahead Reserve Requirements in the PJM Region. The Day-ahead Market includes offers for energy and reserves from generation, bids from fixed and price-sensitive load, increment offers, decrement bids, Up-to Congestion Transactions, offers for demand reductions, and interchange transactions. Day-ahead Locational Marginal Prices (LMPs) and reserve Market Clearing Prices (MCPs) are calculated for each hour, using optimization software

In performing this calculation, the Day-ahead Market calculates the cost of serving an increment of load at each bus from each eligible offer as the sum of the following components of Locational Marginal Price: (1) System Energy Price, (2) Congestion Price, which is the effect on transmission congestion costs (whether positive or negative) and

(3) Loss Price. The offers that can serve an increment of load at a bus at the lowest cost, calculated in this manner, determine the Day-ahead Price at that bus for each hour. These LMPs and MCPs shall be the basis for purchases and sales of energy and transmission congestion charges resulting from the Day-ahead Energy Market.

Day-ahead economic dispatch is performed in the Day-ahead security constrained economic dispatch software program, known as the dispatch run. Day-ahead prices are calculated in a subsequent execution of the Day-ahead security constrained economic dispatch optimization software program, known as the pricing run. The pricing run executes the same optimization as the dispatch run but additionally applies Integer Relaxation to Eligible Fast-Start Resources. Integer Relaxation is the process by which the commitment status variable for an Eligible Fast-Start Resource is allowed to vary between zero and one. Day-ahead prices are determined for every hour, using the applicable marginal energy offer of the resources being dispatched using the offer schedule on which the resource is committed in the dispatch run. The applicable marginal energy offer is determined by comparing the megawatt output of the resource from the pricing run with the Market Seller's Incremental Energy Offer curve or, for Eligible Fast-Start Resources, the Composite Energy Offer which adds amortized Start-Up Costs and amortized No-Load Costs, expressed in dollars per megawatt-hour ($/MWh), to the resource's Incremental Energy Offer. The integer relaxation allows units operating at their minimum load to participate in price setting, in spite the fact that the shadow prices on the load balance constraint in the dispatch run excludes units operating at their minimum load constraint.

Allowing such participation of Fast Start units operating at minimum load in price setting has become important with the increased deployment of renewable resources that often results in many Fast Start units, dispatched to back up these renewables, operating at minimum load.

5.7 Overview of the PJM Transmission Rights Market

PJM uses two types of financial instruments to allocate entitlements to the transmission network under its control to its firm transmission service customers who jointly own the property rights to the network, and to convert such rights into tradable hedging instruments against uncertain hourly congestion charges.

ARRs - Auction Revenue Rights: are entitlements allocated annually to Firm Transmission Service Customers that entitle the holder to receive an allocation of the revenues (or charges) from the Annual FTR Auction

FTRs - Financial Transmission Rights: are financial instruments awarded to bidders in the FTR Auctions that entitle the holder to a stream of revenues (or charges) based on the hourly Day Ahead congestion price differences across the path

Both ARRs and FTRs are defined as point-to-point instruments specifying points of injection and withdrawal with MW denominations, and are subject to a simultaneous feasibility test (SFT). The SFT applied to the ARR allocation ensures that all the allocated ARRs can be scheduled simultaneously as energy injections and withdrawals transactions and meet the DC power flow transmission constraints. Likewise, the SFT applied to the FTR auction ensures that all outstanding FTRs can be scheduled simultaneously as energy injections and withdrawal transactions and meet the DC power flow transmission constraints.

PJM transmission customers are transacting energy over the grid are exposed to congestion hourly congestion charges that are determined by the day ahead dispatch. The congestion charge per MWh transaction from an injection point A to a withdrawal point B is the congestion component of the Locational Marginal Price (LMP) at point B minus the corresponding price at node A. The marginal loss components of LMP are excluded from this calculation. FTRs are financial instruments that enable market participants to hedge against such uncertain congestion costs and achieve cost certainty in their forward energy transactions.

FTRs at PJM are offered both as “obligations” or “options”. A one MW FTR obligation from point A to point B entitles/obligates its holder to receive/pay an hourly cash flow that equals to the difference in the congestion components of the nodal prices (per MWh) between point B (withdrawal) and point A (injection), as determined by the day ahead dispatch, over the term of the FTR. By contrast, an option entitles the holder to a positive cash flow with no obligation when the payoff is negative. In either case, the FTR payoffs per MWh when there is congestion in the direction of a transaction, equal to the congestion charges and can offset such charges.

The FTRs are allocated trough an annual auction in which bidders can submit price and quantity bids or offers for eligible point to point FTR obligations or options. The auction is cleared by a DC Optimal Power Flow (DCOPF) algorithm that maximizes auction net proceeds subject to simultaneous feasibility of all the cleared FTR and the outstanding FTRs issued through other allocation mechanisms (e.g. long erm FTRs). The clearing prices for the FTRs are determined by the nodal price differences resulting from the auction DCOPF, i.e., the auction price for a one MW FTR from point A to point B is the difference between the auction nodal price at B minus the auction nodal price at point A.

The resulting auction prices for FTRs are also used for distributing the auction proceeds to ARR holder who can schedule their ARRs as price takers in the FTR auction or just keep them and collect their corresponding proceeds. The SFT applied to the allocated ARRs ensures that the auction revenues will cover the payment to the ARRs, assuming that the same topology, is used in in the ARR SFT and in the FTR auction. This revenue adequacy is because the mix of allocated ARRs represents a feasible point in the FTR auction revenue maximization problem, so the objective function value at that point (which is the total payment to ARRs) is less than the maximum revenue achieved by the auction.

Likewise, assuming that the topology used by the FTR auction SFT is the same as that used in the day ahead dispatch or the feasible set of the FTR auction SFT is a subset of the day ahead dispatch feasible set, then the congestion charges will exceed the FTR payoffs. Again, such revenue adequacy, is assured by the fact that the FTR mix represents a feasible but suboptimal solution for the day ahead dispatch, which maximizes the value of the transmission network and the congestion charges.

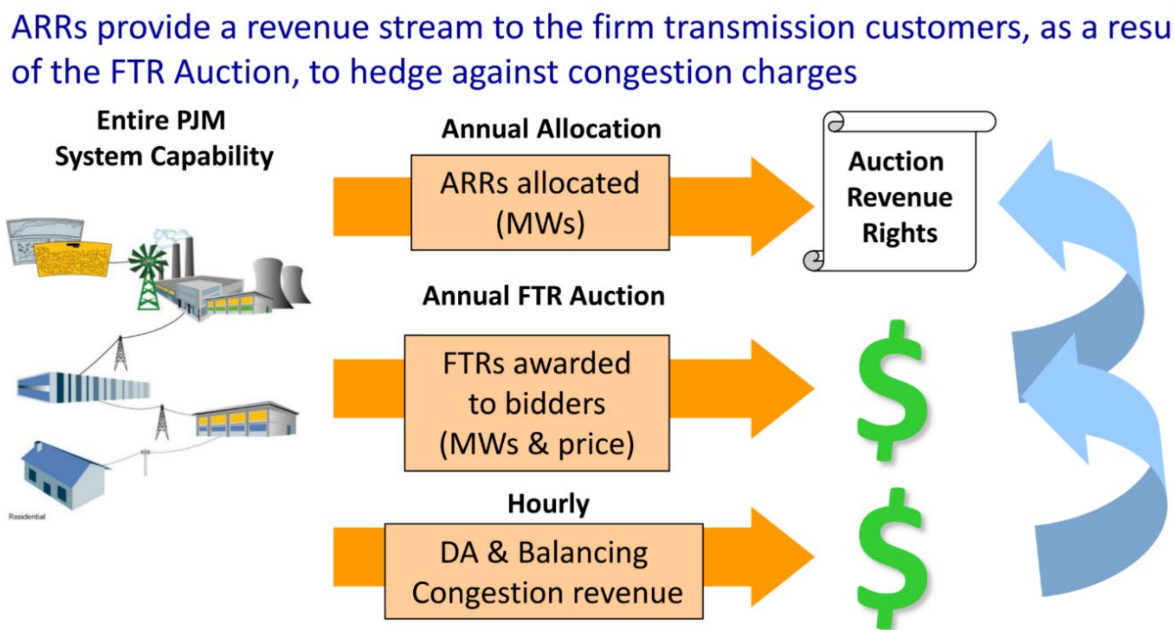


Figure 5-12: PJM ARR/FTR Relationship

(Source:https://www.pjm.com/-/media/training/nerc-certifications/markets-exam-materials/mkt-optimization wkshp/financial-transmission-rights.ashx?la=en)

In the following sections we describe the key elements of the ARR allocation process and FTR auctions as well at the mechanisms for exercise and conversion of ARRs to FTRs, and the use of FTRs as hedges against congestion charges. In order to increase transparency of the presentation we omit some details such as the handling of transitional FTRs that are issued between annual FTR auctions for new load in zones associated with market growth. We also omit Residual Auction Revenue rights (ARRs) defined as Network Service or Firm Point-to-Point Auction Revenue rights that were made available as a result of transmission upgrades that occur after the Annual ARR Allocation. These specialized topics are addressed in the detailed PJM Manual 06.

5.7.1 ARR Allocation

Auction Revenue Rights (ARRs) are the mechanism by which the proceeds from the Annual FTR Auction are allocated. ARRs are allocated to Network Service Customers and to Firm Point-to-Point Transmission Customers for the duration of the Annual Planning Period. Auction Revenue Rights will be distributed to Network Service Customers and Firm Point-to-Point Transmission Customers. Market Participants submit ARR requests for the planning period during the Annual ARR Allocation process. The Annual ARR Allocation is a two-stage allocation process designed to provide long-term certainty along with increased flexibility. The first stage of the allocation consists of two parts, Stage 1A and Stage 1B. In this first stage, Network Service Customers make ARR requests based on active generation resources that historically served load in each transmission zone or Qualified Replacement Resources.

Also in Stage 1, Firm Transmission Customers that are deemed as Qualifying Transmission Customers can make ARR requests based on the megawatts of firm service provided between the receipt and delivery points to which the Customer had Firm Point-to-Point Transmission Service during the historical reference year.

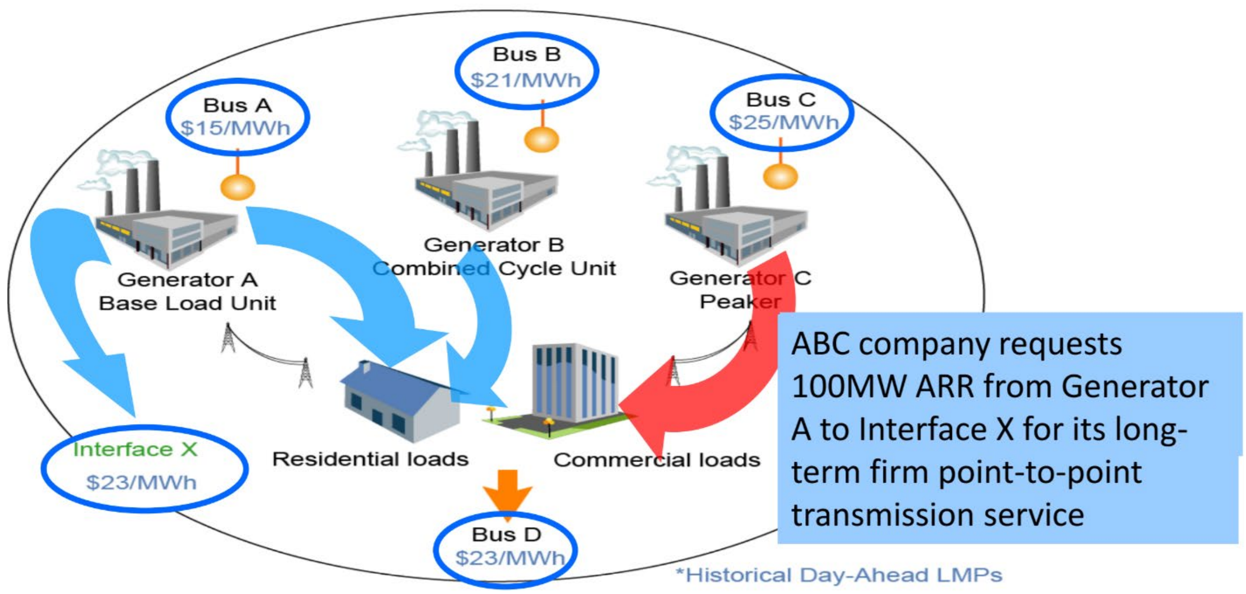


Figure 5-13: ARR Candidates

(Source: https://www.pjm.com/-/media/training/nerc-certifications/markets-exam-materials/mkt-optimization-wkshp/financial-transmission-rights.ashx?la=en)

The second stage, Stage 2, is a three-round allocation procedure that allows market participants to adjust their hedging paths on an annual basis. PJM will allocate ARRs that pass a Simultaneous Feasibility Test to Firm Transmission Customers based on priority (feasibility).

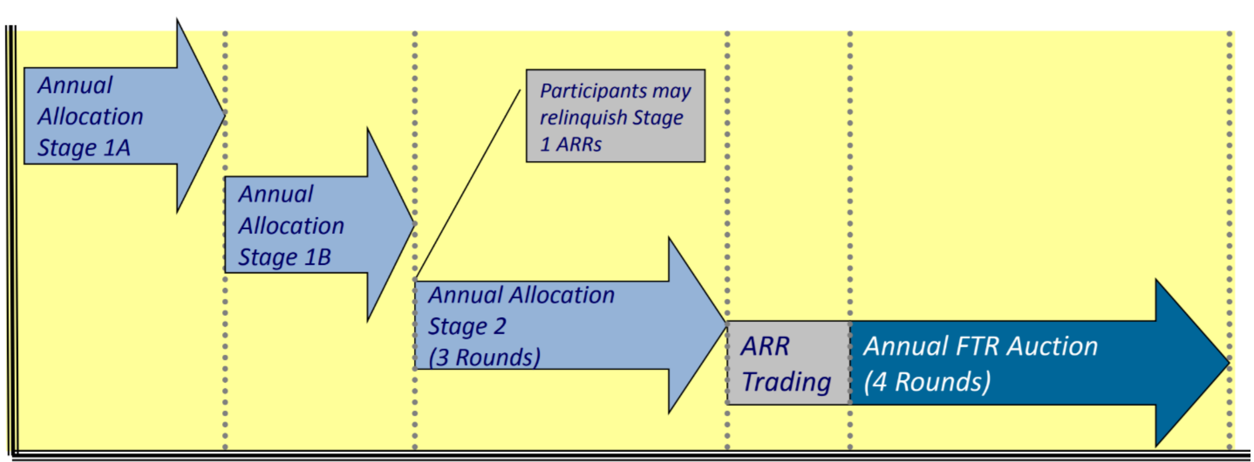


Figure 5-14: Annual ARR/FTR Market Timeline

(Source:https://www.pjm.com/~/media/etools/ftr-center/ftr-module-oflmp-and-ftr-course.ashx)

5.7.2 Annual Allocation of Auction Revenue Rights (ARRs) Stage 1A

In Stage 1A, Network Services Customers must specify specific active historical generation resources or Qualified Replacement Resources (source) to aggregate Network Customer Load in the Transmission Zone or other designated Load Aggregation Zone (sink) up to value of base load. Qualifying Transmission Customers are any Firm Transmission Customers with an agreement for Long-Term Point-to-Point Transmission Service used to deliver energy from a designated network resource to load located either outside or within the PJM Region, and that was confirmed and in effect during the historical reference year for the zone in which the resource is located. “Zonal Base Load” is the lowest daily zonal peak load from the twelve month period ending October 21 of the calendar year immediately preceding the calendar year in which an annual Auction Revenue Right allocation is conducted, increased by the projected load growth rate for the relevant Zone, when non-extraordinary conditions exist for the applicable twelve month period, as determined by PJM. If the lowest daily zonal peak load from the applicable twelve month period is abnormally low due to extraordinary circumstances, as determined by PJM, Zonal Base Load shall mean the next lowest daily zonal peak load that was not affected by extraordinary circumstances during the applicable twelve month period, increased by the projected load growth rate for the relevant Zone. In Stage 1A,

Firm Transmission Customers that are deemed as Qualifying Transmission Customers may request ARRs up to 50% of the megawatts of firm service provided between receipt and delivery points as to which the Transmission Customer had Point to-Point Transmission Service during the historical reference year. All Network Integration Service ARRs allocated in Stage 1A are designated from an active historical generation resource or Qualified Replacement Resource. All requests received during each stage of the Annual ARR Allocation are deemed to have arrived simultaneously. A Network Service Customer’s total ARR amount allocated to a transmission zone or load aggregation zone cannot exceed the participant’s total network base load in that zone or load aggregation zone.

PJM performs a Simultaneous Feasibility test to determine the set of ARRs that can be awarded to each Network customer and notifies each Load Serving Entity (LSE) of the ARR awards resulting from the Stage 1 allocation process. A participant may surrender any portion of the ARR awards resulting from Stage 1 of the Annual ARR Allocation process prior to the commencement of Stage 2 of the Annual ARR Allocation process provided that all remaining outstanding ARRs are simultaneously feasible following the return of such ARRs. ARRs may be traded but trades must be made no later than the opening of the first round of the Annual FTR Auction and all trades are effective for the entire planning period. An LSE wishing to trade its ARRs must trade all of its ARRs associated with a particular zone. The LSE’s zonal network service peak load is also automatically transferred to the new ARR owner for purposes of ARR allocation and reassignment. The new ARR owner is then subject to ARR reassignment associated with shifts in the original owners zonal network service peak load. Within the planning period, as load changes from one LSE to another within a transmission zone, a proportionate share of the ARRs defined to sink into the zone are reassigned from the old LSE to the new LSE. The reassignment of ARRs must be initiated by a request made by the LSE gaining load.

5.7.3 Annual Allocation of Auction Revenue Rights (ARRs) Stage 1B

In Stage 1B of the Annual ARR Allocation, Network Services Customers must specify specific active historical generation resources or Qualified Replacement Resources (source) to aggregate Energy Settlement Area in the Transmission Zone or other designated Load Aggregation Zone (sink) up to value of network service peak load minus awarded ARRs from Stage 1A. Qualifying Transmission Customers are any Firm Transmission Customers with an agreement for Long-Term Point-to-Point Transmission Service used to deliver energy from a designated network resource to load located either outside or within the PJM Region, and that was confirmed and in effect during the historical reference year for the zone in which the resource is located.

Firm Transmission Customers that are deemed as Qualifying Transmission Customers may request ARRs up to the remainder of the megawatts of firm service not awarded in Stage 1A provided between receipt and delivery points as to which the Transmission Customer had Point-to-Point Transmission Service during the historical reference year. All Network Integration Service ARRs allocated in Stage 1B are designated from an active historical generation resource or Qualified Replacement Resource. All requests received during each stage of the Annual ARR Allocation are deemed to have arrived simultaneously. A Network Service Customer’s total ARR amount allocated to a transmission zone or load aggregation zone cannot exceed the participant’s total network peak load in that zone or load aggregation zone minus the awarded ARRs from Stage 1A. PJM determines the set of eligible ARR sources for each transmission zone or for each historic load aggregation zone within a transmission zone based on the historical reference year that corresponds to the LMP-based market implementation for the transmission zone. Only long-term supply contracts or historical capacity contracts that were in place during the reference year and have a contract term of ten (10) years or greater (or were contracts with renewable options that have been exercised, and such exercised option term(s) plus the original contract term were or will be, equal to a term of ten (10) years or more prior to the reference year are eligible to be considered historical generation resources for the purposes of Stage 1 allocation. This would include generation that was owned by an LSE and later sold but retained under a supply contract such that the generation was designated to the serve the load continuously for ten (10) years or greater.

PJM performs a Simultaneous Feasibility test to determine the set of ARRs that can be awarded to each Network customer. PJM notifies each LSE of the ARR awards resulting from the Stage 1B allocation process. A participant may surrender any portion of the ARR awards resulting from Stage 1B of the Annual ARR Allocation process prior to the commencement of Stage 2 of the Annual ARR Allocation process provided that all remaining outstanding ARRs are simultaneously feasible following the return of such ARRs. ARRs may be traded but trades must be made no later than the opening of the first round of the Annual FTR Auction and all trades are effective for the entire planning period. An LSE wishing to trade its ARRs must trade all of its ARRs associated with a particular zone. The LSE's zonal network service peak load is also automatically transferred to the new ARR owner for purposes of ARR allocation and reassignment. The new ARR owner is then subject to ARR reassignment associated with shifts in the original owners zonal network service peak load. Within the planning period, as load changes from one LSE to another within a transmission zone, a proportionate share of the ARRs defined to sink into the zone are reassigned from the old LSE to the new LSE. The reassignment of ARRs must be initiated by a request made by the LSE gaining load.

5.7.4 Annual Allocation of Auction Revenue Rights (ARRs) Stage 2

The second stage, Stage 2, of the allocation is a three-round allocation procedure. In Stage 2 of the Annual ARR Allocation, participants submit ARR requests for the planning period. All Network Service ARR requests must pass a Simultaneous Feasibility Test before being given PJM approval and PJM can approve all, part, or none of the ARR request based on the results of the Simultaneous Feasibility Test. A participant's total ARR amount to a transmission zone or load aggregation zone cannot exceed the participant's total network load in that zone or load aggregation zone. As before, the source point of each ARR request may be any available generator bus, Hub or external interface or load zone for which PJM calculates and posts a Day-ahead Congestion Price value. The sink point of each ARR request must be the Network LSEs aggregate load in the Energy Settlement Area in the transmission zone or load aggregation.

PJM performs an iterative allocation process that consists of three sequential rounds, with one third of the remaining system ARR capability allocated in each round. Network Service customers and Firm Transmission customers can view the results of each allocation round before submitting their ARR requests for a subsequent round. In every round of the three-round allocation process, the Network Customer's ARR requests are limited to one third of the Network customer's peak load remaining unallocated after the Stage 1 allocation process. For example, if the Network customer's peak load is 100 MW and they had received 70 MW of ARRs in Stage 1, then the ARR requests in each round of Stage 2 are limited to (100-70)/3 = 10 MW.

At the end of each round PJM notifies each Network Customer or Firm Transmission Customer of the ARRs that they were awarded as a result of Round 1. After viewing the results, Network Customers or Firm Transmission Customer submit ARR requests for next round. The allocation process continues in an iterative manner for three rounds. In each round, PJM staff performs the Simultaneous Feasibility test to determine the feasible set of ARRs that can be awarded. If all ARR requests are not simultaneously feasible, then proration is required. If ARR proration is required due to infeasibility then ARRs are allocated in proportion to the MW value requested and in inverse proportion to the effect on the binding constraint. As illustrated in the example below.

Line A to B Capacity is 50 MW

Two ARR requests are submitted exceeding 50 MW capability of Line A-B:

Table 5-2: ARR Prorating Example

ARR #	Requested ARRs	Path	Effect per MW on Line AB	Resulting Line AB Flow Impact
#1	200 MW	A to B	.50	100 MW
#2	200 MW	C to D	.25	50 MW

Because the total flow impact on Line A-B resulting from the two ARR requests is 150 MW, proration is required because the capability of Line A-B is only 50 MW.

Prorated ARR Amount Awarded =
(Line Capability MW) x (Requested MW / Total Requested ARR MW) x (1/ ARR Effect per MW on Line A-B)

ARR #1 Prorated MW Awarded = *(50 MW) x (200 MW / 400 MW) x (1/.50) = 50 MW*

ARR #2 Prorated MW Awarded = *(50 MW) x (200 MW / 400 MW) x (1/.25) = 100 MW*

At the end of each round, PJM notifies each Network Customer or Firm Transmission Customer of the ARRs that they were awarded. After viewing the results, Network Customers or Firm Transmission Customer submit ARR requests for the next round. The allocation process continues in an iterative manner for three rounds. As before the allocated ARRs may be traded prior to the opening of the first round of the Annual FTR Auction subject to the stipulation detailed in stage 1.

5.7.5 Reassignment of Auction Revenue Rights (ARRs)

ARRs allocated for the planning period will be reassigned on a proportional basis within a zone as load switches within the planning period. As load shifts from one LSE to another within a transmission zone, a proportionate share of the ARRs defined to sink into the zone are reassigned from the old LSE to the new LSE. The reassignment of ARRs is an automatic process that is conducted on a daily basis. ARRs are only reassigned from those LSEs that have lost load in a zone and have a net positive economic ARR position to that zone. An LSE that loses load will lose ARR MWs in proportion to the amount of load lost and this same proportion will reduce each individual ARR assigned to the LSE. ARRs are initially allocated to the nearest 0.1 MW but reassigned to the nearest .001 MW.

The total set of ARRs to be forfeited by LSEs losing load in a zone will be reallocated to LSEs gaining load in the zone in proportion to each LSE's MW load gain relative to the total load shifted in the zone.

On a daily basis, Auction Revenue Rights are reassigned by comparing each LSE's Network Service Peak Load in a zone to the Network Service Peak Load of the previous day. Each ARR owned by an LSE losing load is reduced in proportion to its lost load. The forfeited ARRS are then reallocated to the LSEs gaining load in proportion to their relative gain.

5.7.6 New Stage 1 Resources

A Network Service User may request the addition of new stage 1 resources to the stage 1 resource list if the capacity of the historical resources for a zone is less than the zonal base load. Such requests are subject to the simultaneous feasibility tests and are limited to generation resources either owned by the requesting party or subject to ten-year or longer firm energy and capacity contracts that have been executed to serve load eligible for stage 1A ARRs and remains in effect for a minimum term of ten years.

Simultaneous feasibility tests for new stage 1 resource requests shall ensure that the requests for a new base resource does not increase the MW flow on facilities binding in the current ARR allocation or in future stage 1a allocations and does not cause MW flow to exceed applicable ratings in either set of conditions. A simultaneous feasibility test of new stage 1 resource requests will assess the feasibility of the requests with all existing stage 1 and stage 2 ARRs modeled as fixed injection-withdrawal pairs. Only requests for new stage 1 resource that are received by by November 1 will be processed for the next Annual ARR Allocation and will be deemed to have been submitted simultaneously and will be evaluated at the same time.

5.7.7 Alternate Stage 1 Resources

A Network Service User may replace one or more of its existing stage 1 resources and its associated MW amount of ARRs with an alternate resource. Such requests are subject to the simultaneous feasibility tests described below and are limited to generation resources either owned by the requesting party or subject to ten-year or longer firm energy and capacity contracts that have been executed to serve load eligible for stage 1A ARRs and remains in effect for a minimum term of ten years. Simultaneous feasibility tests for alternate stage 1 resource requests shall ensure that relative to the existing resource, the alternate base resource does not consume a greater amount of transmission capability on facilities binding in the current ARR allocation or future stage 1A allocations, and does not allow MW flow(s) to exceed applicable ratings on any other facilities. A simultaneous feasibility test of alternate stage 1 resource requests will assess the feasibility of the requests with all existing stage 1 and stage 2 ARRs modeled as fixed injection-withdrawal pairs. Based on the 10 year allocation model with all eligible stage 1 ARRs for each year including base load growth for each year. The amount of MWs that can be nominated from the alternate resource cannot exceed the original awarded stage

1 MW amount of ARRs associated with the original stage 1 resource. Only requests for new stage 1 resource that are received by November 1 will be processed for the next Annual ARR Allocation and will be deemed to have been submitted simultaneously and will be evaluated at the same time. If the Network Service User accepts the MW amount of ARRs associated with the alternate resource as established by the simultaneous feasibility test, the alternate resource will replace the relevant existing stage 1 resource beginning with the next Annual ARR Allocation. Otherwise, the original stage 1 resource will remain in effect.

5.7.8 Allocation of Incremental Auction Revenue Rights (IARRs)

5.7.8.1 Merchant and Generation Interconnection IARRs

Transmission expansion projects associated with new generation interconnection and Merchant Transmission Expansion projects will be allocated incremental ARRs in a three-round allocation process in which the customer requests incremental ARRs for three pairs of point-to-point combinations (one point-to-point combination is requested per round). In each round, one-third of the Incremental ARRs made available by the expansion project will be assigned to the requester. After each of rounds one and two, the requester may accept the assigned Incremental ARRs or refuse them. Acceptance of the assignment will remove the assigned Incremental ARRs from availability in the next rounds. Refusal of the assignment will result in the Incremental ARR being available for the next round. The Incremental ARR assignment made in round three will be final and binding. The final and binding Incremental Auction Revenue Right assignment for a requested point-to-point combination in each round shall in no event be less than one third of 80% and no greater than one-third of 100% of the non-binding estimate of Incremental Auction Revenue Rights for that point-to-point combination that was provided to the New Service Customer.

Incremental ARRs will be effective for thirty years or the life of the facility or upgrade, whichever is less. At any time during this thirty-year period, in lieu of continuing this thirty-year ARR, the Interconnection Customer shall have a one-time choice to switch to an optional mechanism, whereby, on an annual basis, the customer has the choice to request an ARR during the Annual ARR Allocation process between the same source and sink, subject to simultaneous feasibility. Once this option is chosen, the Interconnection Customer must request the Incremental ARR during each annual ARR enrollment window for the upcoming planning period. If no request is made, the Incremental ARR is forfeited for that planning period. At any time during this thirty-year period, an Interconnection Customer may return Incremental ARRs that it no longer desires at any time, provided that all remaining outstanding ARRs can be simultaneously accommodated following the return of such ARRs. In the event an Interconnection Customer returns Incremental ARRs, the Interconnection Customer shall have no further rights regarding such Incremental ARRs.

5.7.8.2 IARRs for Regional Transmission Expansion Plans (RTEP) and Elective Upgrades

IARRs will be allocated to any RTEP upgrade whose costs is born on a regional basis. The IARRs created by a such projects or elective upgrades will be allocated to the Responsible Customers in proportion to each Responsible Customers' share of the project or upgrade cost. The detailed IARR Calculation Method for RTEP and elective upgrades is described in the PJM Manual 06 available at:

https://www.pjm.com/-/media/documents/manuals/m06.ashx.

IARRs become effective on the first day of the first month that the upgrade is included in the transmission system model for the monthly FTR auction. As before, Incremental ARRs will be effective for thirty years or the life of the facility or upgrade, whichever

is less. For IARRs that become effective at the beginning of a planning year, their value will be determined identically to that of annually allocated ARRs, based on the nodal prices of the annual FTR auction. If IARRs become effective during a planning year, then their value for each month remaining in that planning year will be based on the results of the monthly FTR auctions. For each planning year thereafter, the value of IARRs will be determined identically to that of annually allocated ARRs, based on the nodal prices of the annual FTR auction.

5.7.9 FTR Auctions

5.7.9.1 Overview

Throughout the year, PJM oversees the process of selling and buying FTRs through FTR Auctions. Market Participants purchase FTRs by participating in Long-term, Annual and Monthly FTR Auctions.

Long Term FTRs - PJM conducts a Long-term FTR process of selling and buying FTRs through a multi-round process for FTRs with duration of three consecutive Planning periods immediately following the Planning Period during which the Long-term FTR Auction is conducted. The capacity offered for sale in Long-term FTR Auctions shall be the residual system capability after the assumption that all ARRs allocated in the immediately prior ARR allocation process, are self-scheduled into FTRs and modeled as fixed injections and withdrawals in the Long-term FTR Auction. The Long-term FTR Auction is a multi-round auction consisting of three rounds. In each round 1/3 of the feasible FTR available capability is awarded. FTRs that are purchased in one round may be offered for sale in later rounds. The long-term FTR auction model includes all upgrades

planned to be placed into service on or before June 30th of the first Planning Period within the three year period covered by the auction. The transmission upgrades to be modeled for this purpose shall only include those upgrades that, individually, or together, have 10% or more impact on the transmission congestion on an individual constraint or constraints with congestion of $5 million or more affecting a common congestion path. Transmission upgrades modeled for this purpose are modeled in the subsequent long-term FTR auctions.

Annual FTR Auction - The Annual FTR Auction offers for sale the entire transmission

entitlement that is available on the PJM system on an annual basis. The Annual FTR Auction is a multi-round auction consisting of four rounds. In each of the four rounds, 25% of the feasible FTR capability of the entire PJM system is awarded. FTRs that are purchased in one round may be offered for sale in subsequent rounds.

Monthly FTR Auctions - In each calendar month, Monthly FTR Auctions provide a method of auctioning the residual FTR capability that remains on the PJM Transmission System after the Long-term and Annual FTR Auction is conducted. The Monthly FTR Auctions are single-round auctions, where the residual FTR capability is awarded. The Monthly FTR Auctions also allow Market Participants an opportunity to offer for sale any FTRs that they currently hold. An auction participant must own any FTR that is offered for sale (no short sales). In the Monthly FTR Auctions, Market Participants may bid to buy or offer to sell FTRs that cover either one month for any of the next three months remaining in the planning period or three months for any of the quarters remaining in the planning period with some further restrictions on overlaps that are specified in the PJM Manual 06.

The clearing mechanism of the FTR Auctions will maximize the quote-based value of FTRs awarded in each auction. The proceeds from the Annual FTR Auction

are distributed to ARR holders. All Long-term and monthly auction revenues are first allocated among ARR holders in proportion to the holder‘s deficiencies from the Annual FTR Auction. Any monthly auction revenues remaining after this allocation are treated as excess congestion charges and are distributed starting with Stage Two as described in the “Market Settlements” section. FTRs are awarded in the FTR Auctions as options or obligations for the Annual and monthly FTR Auctions and obligations only for the Long term FTR Auctions. FTR products are available for on-peak hours, off-peak hours and 24 hours. FTRs acquired in the FTR auctions entitle the holder to credits (or debit obligations) for transmission congestion charges for the term and validity period of the FTR. They hedge the FTR holder against congestion payments when energy delivery is consistent with the FTR definition. However FTRs do not hedge their holder against payment for losses.

Valid FTR sources and sinks in the auctions are limited to those designated by PJM and for which an LMP is being calculated and posted. The list of available sources and sinks for each auction is posted before the start of the bidding window. Only a subset of paths are eligible for FTR options in the Annual and Monthly FTR Auctions in order to prevent potential auction clearing performance issues. FTR Options are not available in the Long-term FTR Auctions. FTRs in all FTR Auctions may be designated from injection buses outside PJM and withdrawal locations outside PJM OR buses with injections and withdrawals within PJM. In the Annual FTR Auction, an ARR holder may self-schedule an FTR Obligation (up to the ARR MW reservation amount) into the Annual FTR auction as a “price-taker” auction buy bid. The self-scheduled FTR must have exactly the same source and sink points as the ARR. This feature can only be used in Round 1 and must be for a 24-hour FTR Obligation product. 25% of the MW amount self-scheduled in Round 1 will clear in each round. In all FTR Auctions, FTRs can be reconfigured, meaning that the FTR auction not only allows Market Participants to purchase the FTRs offered into the auction by sellers, but also enables buyers to purchase FTRs that are different from any of the FTRs offered into the auction by sellers. Quotes in the Auctions with a $0 bid price are allowed in the auctions. However, $0 bids will not be awarded on paths with a clearing price of $0 to avoid potential performance issue of the market clearing algorithm (due to degeneracy).

5.7.9.2 Auction Clearing

The winning quotes are determined by the set of simultaneously feasible FTRs with the highest total auction value, as determine by the bids of the buyers and taking into account the reservation prices of the sellers. To ensure feasibility, each constraint is monitored for limit violation by the worst case realization of awarded FTR Options. Hence, counterflow created by an FTR Option is ignored.

The valuation of the awarded FTRs in the auctions is based on the bids submitted into each FTR Auction. Therefore, the set of bids that maximizes the bid-based value of the awarded FTRs to the Market Participants that would receive them is the winning set. This ensures that PJM awards the set of FTRs and allocates them among auction participants in such a way that the value-based transmission utilization is maximized. The optimization algorithm determines market clearing nodal prices at each bus and corresponding shadow prices on all binding constraints. Market clearing prices for FTR obligations (whether bought or sold) can be computed directly from the nodal prices. Specifically, the clearing price of an A-to-B FTR Obligation is equal to the nodal price at bus B minus the nodal price at bus A and it is the negative of the clearing price of a B-to-A FTR Obligation. This is not true for the FTR Options since the clearing prices of FTR Options are never negative. The clearing price of an FTR Option is a function of the shadow price on each binding constraint and cannot be computed directly from the nodal prices. Specifically, the clearing price for an FTR option from point A to point B is calculated as the sum of the corresponding shift factors on all binding transmission lines in the direction of the flow multiplied by the corresponding shadow prices on these lines.

In addition to the FTR auctions, PJM operates an FTR secondary trading markets that enables and facilitates bilateral trading of existing FTRs between market participants. A bulletin board system called FTR Center allows trading of existing FTRs only, options and obligations (but does not permit conversions or reconfiguration). Independent trading is also allowed but then PJM is not able to adjust members billing to reflect such changes of ownership.

5.7.10 Market Settlements of FTR Auction Revenues to ARRs

Long-term, Annual and Monthly FTR Auction revenues are distributed to ARR holders in proportion to (but not to exceed) the economic value of the ARRs when compared to the

Annual FTR Auction clearing prices for FTR Obligations from each round proportionately. Long-term FTR auction revenues associated with FTRs that cover multiple Planning Years shall be distributed equally across each planning period in the effective term of the FTR. Excess revenues after distribution to ARR holders will be used to fund any shortfall in FTR Target Allocations over the Planning Period. These funds are accounted for on a monthly basis as Excess Congestion Charges and they are distributed with other Excess Congestion Charges as described in the Section entitled “Distributing Excess Transmission Congestion Charges" in the PJM Manual 06.

The settlements for ARRs is based on the clearing prices from each round of the Annual FTR Auction. The amount of the credit that the ARR holder should receive for each round is equal to the MW amount of the ARR (divided by the number of rounds) times the price difference from the ARR delivery point to the ARR source point:

$$\text{ARR Target Allocation} = \textit{ARR MWs / \# of Rounds} * (LMP_{Delivery} - LMP_{Source})$$

The LMP values in the above calculation are results for FTR Obligations from the appropriate round of the Annual FTR Auction. The ARR Target Allocation can be positive or negative. An ARR can be either a benefit or liability to the holder depending on the direction of transmission congestion in the annual auction analysis. If sufficient funds are collected in the Annual and Monthly FTR Auctions to satisfy all ARR Target Allocations then the ARR Credits = ARR Target Allocations for all ARR holders. The ARR Credits may be prorated proportionately if there are insufficient Annual and Monthly FTR auction revenues collected to cover all of the ARR credits. If the ARR Credits are prorated, the difference between ARR Target Allocations and ARR Credits are called ARR Deficiencies. The ARR Deficiencies may be funded by Annual Excess Congestion Charges. The settlements for the Annual FTR Auction and the corresponding ARR settlements are performed on a daily basis.

5.7.10.1 FTR Settlement

The Transmission Congestion Credit Target Allocation is the amount of credit the FTR holder should receive in each constrained hour due to the value of an FTR. PJM determines a target allocation of Transmission Congestion Credits for each hour for each FTR which is determined by the number of FTR megawatts times the difference in the hourly day ahead LMP between the source and sink buses. If the source or sink is an aggregate zone than the calculation uses a load weighted average.

Any revenue deficient transmission rights (ARRs or FTRs) remaining at the end of the Planning Period are satisfied through a transmission rights uplift credit, the costs of which are allocated as charges to FTR holders on a pro-rata basis according to their net FTR target allocation position, relative to the total net FTR target allocation positions of all FTR holders in the PJM Interchange Energy Market. An entity with a net negative FTR target allocation position is not subject to transmission rights uplift allocation charges and are excluded from the uplift charge calculations.

5.7.11 Simultaneous Feasibility Test (SFT)

The Simultaneous Feasibility Test (SFT) is a market feasibility test run by PJM that provides revenue adequacy by ensuring that the Transmission System can support the subscribed set of FTRs or ARRs during normal system conditions. If the FTRs or ARRs can be supported under normal system conditions and congestion occurs, PJM will be collecting enough congestion charges to cover the FTRs or ARR credits, thus becoming revenue adequate. The purpose of the SFT is to preserve the economic value of FTRs or ARRs to the holders by ensuring that all FTRs or ARRs awarded can be honored. An SFT is run for each ARR or FTR requested. The SFT uses a DC power flow model that models the requested firm transmission reservations and expected network topology during the period being analyzed. It is not a system reliability test and is not intended to model actual system operating conditions. FTRs and ARRs for Firm Point-to-Point Service are modeled as generation at the injection point(s) and load at the delivery (sink) point(s). FTRs and ARRs for Network Integration Service are modeled as a set of generators at the source point and a network load at the delivery (sink) point. SFTs are run for yearly, monthly, and weekly analysis periods, when network resource changes are submitted and during the determination of the winning bids for the Annual FTR Auction and the Monthly FTR auction.

Inputs to the SFT model include all newly-requested FTRs and ARRs for the study period, all existing FTRs and ARRs for the study period, transmission line outage schedules, Thermal operating limits for transmission lines, that are expected to last for 2 months or more are included in the determination of simultaneous feasibility for the Annual PJM FTR Auction, and outages of five days or more are included in the determination of simultaneous feasibility for monthly PJM FTR auctions. Outages of shorter duration that are determined through PJM analysis to be likely to cause FTR revenue inadequacy if not modeled are also included. Simultaneous Feasibility determinations take into account outages based on reasonable assumptions about configuration and availability. the SFT evaluates the ability of all system facilities to remain within normal ratings during normal, extended-period operation, while maintaining an acceptable bulk system voltage profile. The system must also be able to sustain any single contingency event with all system facilities remaining within applicable short-term, emergency ratings while maintaining an acceptable bulk system voltage profile and a maximum bulk system voltage drop of five percent. To ensure feasibility, each constraint is monitored for limit violation by the worse-case scenario combination of awarded FTR options and obligations. Hence, counterflow created by an FTR option is ignored.

5.8 Overview of Resource Adequacy and the Capacity Market

The PJM Capacity Market is designed to ensure the adequate availability of necessary resources that can be called upon to ensure the reliability of the grid. In PJM, the Capacity Market structure, through either the Reliability Pricing Model (RPM) consisting of a Base Residual Auction and three Incremental Auctions for a Delivery Year or the Fixed Resource Requirement (FRR) Alternative, provides transparent information to enable forward capacity market signals to support infrastructure investment. The capacity market design provides a forward mechanism to evaluate the ongoing reliability requirements in a transparent way to provide opportunity for generation, demand response, energy efficiency, price responsive demand, and transmission solutions.

The Reliability Pricing Model (RPM) is designed to align capacity pricing with system reliability requirements and to provide transparent information to all market participants far enough in advance for actionable response to the information. The basic elements of RPM are:

- Locational capacity pricing recognizing locational differences in requirements and resource availability
- Variable Resource Requirement mechanism to adjust price based on the level of resources procured
- Forward commitment of supply by generation, demand resources, energy efficiency resources, and qualified transmission upgrades cleared in a multi-auction structure
- A Reliability Backstop mechanism to ensure that sufficient generation, transmission and demand response solutions will be available to preserve system reliability

The PJM Capacity Market also contains an alternative method of participation, known as the Fixed Resource Requirement (FRR) Alternative. The Fixed Resource Requirement Alternative provides a Load Serving Entity (LSE) with the option to submit a FRR Capacity Plan and meet a fixed capacity resource requirement as an alternative to the requirement to participate in the RPM, which includes a variable capacity resource requirement.

The current Capacity Market design recognizes two product types for Capacity Resources, Capacity Performance Resources and Base Capacity Resources. The Capacity Market design also includes the ability to offer Seasonal Capacity Performance Resources. The market clearing optimization algorithm is designed to clear equal quantities of off-setting seasonal capacity sell offers thereby creating an annual capacity commitment by matching a Summer- Period Capacity Performance Resource with a Winter-Period Capacity Performance Resource.

Participation by Load Serving Entities (LSEs) in the RPM for load served in the PJM region is mandatory, except for those LSEs that have elected the Fixed Resource Requirement (FRR) Alternative and submitted an approved FRR Capacity Plan for their load served in an FRR Service Area. Under RPM, each LSE that serves load in a PJM Zone during the Delivery Year shall be responsible for paying a Locational Reliability Charge equal to their Daily Unforced Capacity Obligation in the Zone multiplied by the Final Zonal Capacity Price applicable to that Zone. LSEs may choose to hedge their Locational Reliability Charge obligations by directly offering and clearing resources in the Base Residual Auction and Incremental Auctions or to self-supply resources to cover their obligation in the Base Residual Auction. Such action may wholly or partially offset an LSE’s Locational Reliability Charges during the Delivery Year.

Resource providers with Existing Generation Capacity Resources, Planned Generation

Capacity Resources, bilateral contracts for unit-specific Capacity Resources, existing Demand Resources, Planned Demand Resources, Energy Efficiency Resources, PRD providers, and Qualifying Transmission Upgrades may participate in PJM's Capacity Market, either in the RPM or the Fixed Resource Requirement (FRR). Participation is mandatory for resource providers with: Available unforced capacity from Existing Generation Capacity Resources located within the PJM market footprint; or Bilateral contracts for available unit-specific Capacity Resources that are Existing Generation Capacity Resources located within the PJM market footprint. Generation is treated as existing for the purpose of must-offer requirement and mitigation provisions when the generation is (a) in service at the commencement of an RPM Auction or (b) not yet in service but has cleared an RPM Auction for any prior Delivery Year.

5.8.1 The RPM Model

The RPM Model is PJM's resource adequacy construct that ensures that adequate Capacity Resources, including planned and existing Generation Capacity Resources, Energy Efficiency Resources and planned and existing Demand Resources will be made available to provide reliable service to loads within the PJM Region. The RPM provides a long term pricing signal for Capacity Resources and LSE obligations that is consistent with the PJM Regional Transmission Expansion Planning Process. The RPM is a multi-auction structure for procuring generation resources and PRD commitments to satisfy the region's unforced capacity obligation. It includes a Base Residual Auction, Incremental Auctions and a Bilateral Market.

- The Base Residual Auction is held three (3) years prior to the start of the Delivery Year.
- At least three Incremental Auctions are conducted after the Base Residual Auction to adjust resource commitments up or down as needed to satisfy potential changes in market dynamics that are known prior to the beginning of the Delivery Year.
- The bilateral market provides resource providers an opportunity to cover any auction commitment shortages. The bilateral market also provides LSEs the opportunity to hedge against the Locational Reliability Charge. The bilateral market is facilitated through a Capacity Exchange system.

5.8.2 Resource Adequacy

The purpose of PJM RTO resource adequacy is to determine the amount of Capacity Resources required to serve the forecast load and satisfy the PJM reliability criterion. PJM performs an assessment of resource adequacy each year for an eleven-year future period. The analysis considers load forecast uncertainty, forced outages of Generation Capacity Resources, as well as planned and maintenance outages. The studies are performed using the installed capacity values of resources. The reliability value of a resource depends on two variables: the installed capacity of the resource and a measure of the probability that a resource will not be available due to forced outages or forced de-ratings. The reliability criterion is based on Loss of Load Expectation (LOLE) not exceeding one occurrence in ten years. The resource requirement to meet the reliability criterion is expressed as the Installed Reserve Margin (IRM) as a percentage of forecast peak load.

The process of determining the Installed Reserve Margin (IRM) that meets the PJM reliability criterion assumes that the internal RTO transmission is adequate and any generation can be delivered to any load without transmission constraints. This process helps in determining the minimum possible IRM for the RTO. This is followed by a Load Deliverability analysis where the RTO is divided into 27 Locational Deliverability Areas (LDAs). The analysis determines transmission import capability needed for each LDA to meet the area reliability criterion of Loss of Load Expectation of one occurrence in 25 years. This import capability requirement is called Capacity Emergency Transfer Objective (CETO), expressed in megawatts and valued as unforced capacity.

When a capacity market does not have the ability to price capacity on a locational basis, all the resources in the market are valued equally. When this occurs, it is possible to have excess reserves in the RTO and relatively low capacity prices. This market signal will result in generation capacity retirements. In some areas of the RTO these retirements will create violations. These conditions will indicate that a higher value for resources is required to be recognized in constrained locations to incent existing generating capacity to remain in service, and new capability to be built in the form of generation resources, demand resources, energy efficiency resources, or merchant transmission upgrades. One of the key features of RPM is the recognition of locational value of capacity. Locational Constraints are localized intra-PJM capacity import capability limitations that are caused by transmission facility limitations or voltage limitations that are identified for a Delivery Year in the PJM Regional Transmission Expansion Planning

5.8.3 Variable Resource Requirement (VRR)

The Variable Resource Requirement (VRR) Curve is a demand curve used in the clearing of the Base Residual Auction that defines the price for a given level of Capacity Resource commitment relative to the applicable reliability requirement. Variable Resource Requirement Curves are defined for the PJM Region and each of the constrained LDAs within the PJM region. The purpose of the Variable Resource Requirement concept is to recognize the value of excess resources above the reliability requirement and provide revenue to resources. The price on the Variable Resource Requirement is higher when the resources are less than the reliability requirement and lower when the resources are in excess. The Variable Resource Requirement Curve for the PJM Region is based on a target level (i.e., the PJM Region Reliability Requirement), Cost of New Entry, and Net Energy & Ancillary Services (E&AS) Revenue. Figure 5-15 below illustrates a typical VRR Curve.

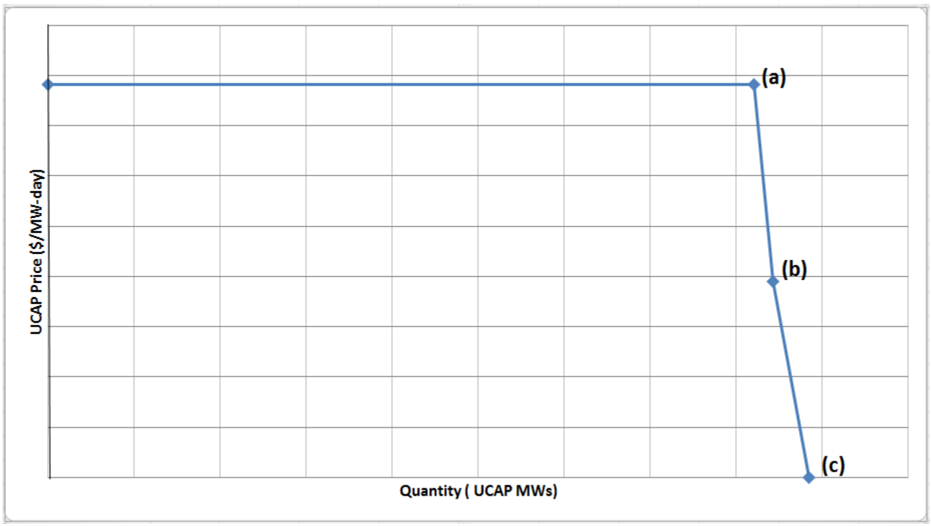


Figure 5-15: Illustrative Example of a Variable Resource Requirement Curve

The First, Second and Third Incremental Auctions provide both a forum for capacity suppliers to purchase replacement capacity, and a means for PJM to adjust previously committed capacity levels due to reliability requirement increases or decreases. The demand curve in these auctions will be built based on a combination of buy bids submitted by market participants and buy bids, if any, submitted by PJM.

5.8.4 Demand Resources in the PJM Capacity Market

5.8.4.1 Load Management Resources

Both Existing and Planned Demand Resources may participate in RPM Auctions or commit to FRR Capacity Plan. Existing Demand Resources are those MWs on a demand resource identified in a pre- registration process in the Capacity Exchange system prior to the RPM Auction. Nominated DR Values (in MWs) associated with end-use customer sites that the Curtailment Service Provider (CSP) has under contract for the current Delivery Year and that the CSP intends to have under contract for the auction Delivery Year are considered Existing MWs.

Planned Demand Resources are defined as resources that do not currently have the capability to provide reduction in demand or to otherwise control load in PJM, but that is scheduled to be capable of providing such reduction or control on or before the start of the Delivery Year. Planned Demand Resources are those MWs on a demand resource that the CSP intends to offer in the RPM Auction or commit to an FRR Capacity Plan in excess of the CSP’s Existing MWs on such demand resource. A resource provider may offer Demand Resources (Planned or Existing) associated with Behind the Meter Generation for an entire Delivery Year into the Base Residual or Incremental Auctions. If the DR offer clears in an RPM auction for a given Delivery Year, the operating Behind the

Meter Generation cannot be netted from load for the purposes of calculating he Peak Load Contributions for the subsequent Delivery Year Demand Resources offered and cleared in a Base Residual or Incremental Auction will receive the corresponding LDA product-specific Resource Clearing Price determined by the optimization algorithm.

5.8.4.2 Price Responsive Demand

The development and implementation of dynamic and time-differentiated retail rates, together with utility investment in advanced metering infrastructure (AMI) has lead an increasing quantity of load in PJM to be responsive to changing wholesale prices. Through enabling technology and behavioral changes, consumers modify their demand as prices change without being centrally dispatched by PJM or bidding demand reductions into the PJM markets. Given the linkage between dynamic retail rate structures and wholesale prices, this price responsiveness is predictable and needs to be accounted for in the wholesale market design and operations.

This predictable reduction in consumption in response to changing wholesale prices is known as Price Responsive Demand (PRD). In the PJM Capacity Market, a PRD Provider may voluntarily make a firm commitment of the quantity of Price Responsive Demand that will reduce its consumption in response to real time energy price during a Delivery Year.

In order to commit PRD for a Delivery Year, a PRD Provider must submit a PRD Plan in advance of the Base Residual Auction or Third Incremental Auction for such Delivery Year that demonstrates to PJM's satisfaction that the maximum nominated amount of price responsive demand will be available by the start of the Delivery Year. Additional PRD may participate in the Third Incremental Auction only in the event, and to the extent that the LDA final peak load forecast for the Delivery Year increases relative to the LDA preliminary peak load forecast used for the Base Residual Auction.

A PRD Provider that is committing PRD in Base Residual Auction or Third Incremental Auction must also submit a PRD election in the Capacity Exchange system which indicates the Nominal PRD Value in MWs that the PRD Provider is willing to commit at different reservation prices ($/MW-day). The PRD election by PRD Providers will result in a change in the shape of the RTO/LDA VRR Curves used in the RPM Auctions. Based on the PRD elections and Resource Clearing Price in the RPM Auction, PJM will determine the Nominal PRD Value committed by each PRD Provider. Those PRD Providers that elected to provide PRD at reservation prices equal to or less than the Resource Clearing Price will have the corresponding value of PRD committed in the RPM Auction. Once committed in a Base Residual Auction, Third Incremental Auction or committed for load served under the FRR Alternative, Price Responsive Demand may not be uncommitted or replaced by available capacity resources or Excess Commitment Credits. However, a PRD Provider may transfer the PRD obligation to another PRD Provider bilaterally.

6 ISO New England

6.1 Market Design Basics

The ISO-New England market first opened its first version of markets in 1999. This market was limited to an intra-day, real-time market with a single region-wide clearing price for energy and prices for ancillary services. After a period of exploring a custom market design, ISO-New England opted in 2001 to purchase rights to the market design of the PJM Interconnection and the associated software.[15] This decision shortened the development and implementation timeline, and the new "Standard Market Design" was implemented in 2003.

The adopted market design incorporated the key features seen in US market operators:

1. Locational marginal prices (LMP) for energy based on security constrained economic dispatch
2. Multi-settlement with both a day-ahead market (hourly) and real-time market (5 to 15 minute pricing intervals)
3. Co-optimiztion of energy and reserve
4. Financial Transmission Rights based on congestion in the Day-Ahead Market

Capacity markets (Resource Adequacy) is not considered a part of the standard market design among US market operators.

At the time in 2003 PJM did not have bid-based reserve markets and hence ISO-New England implemented a Forward Reserve Market (FRM) that did not require changes to the dispatch software.

6.1.1 Commodities

The full set of market commodities (products with clearing prices) include:

1. <u>Energy</u>. The energy product is defined with a 5-minute pricing and settlement. The energy product is nodal with LMPs. The day-ahead energy market uses hourly increments.
2. <u>Regulation Reserves</u>. This is a secondary frequency response responding to 4 second instructions from the system operator under automatic generation control.
3. <u>Contingency Reserves (10-minute operating reserve)</u>. These are secondary frequency response services responding to instructions from the economic dispatch optimization. These contingency reserves include two products. The off-line non-synchronized (non-spinning) reserve is considered of lesser quality due to the probability of start failures.

---

[15] "ISO-NE, PJM Team Up to Standardize Power Markets," *Natural Gas Intelligence* (blog), April 2, 2001, https://www.naturalgasintel.com/iso-ne-pjm-team-up-to-standardize-power-markets/.

a. Ten-minute spinning reserve (TMSR): supplied by online resources capable of converting capacity to energy within 10 minutes.
b. Ten-minute non-spinning reserve (TMNSR): supplied by offline resources that can start and synchronize to the grid within 10 minutes.

4. Tertiary Reserve (30-minute operating reserve). System operators in the Northeast of the US that belong to the Northeast Power Coordinating Council are required to also maintain thirty-minute reserve available to it that is at least equal to one-half its second contingency loss. ISO-New England includes this reserve product as Thirty-Minute Reserve (TMOR). This reserve product has local requirements.
5. Forward Reserve. Forward reserve is market that obligates the seller to provide either TMNSR or TMOR in the real-time market and places extra real-time performance requirements on the seller.
6. Financial Transmission Rights (FTRs). This the right to receive (or pay) the congestion rents between two points. ISO-New England defines this product in monthly and annual terms and durations between one and 12 months.
7. Capacity. ISO-New England has a Resource Adequacy product of capacity. The Forward Capacity Market (FCM) provides an additional revenue stream for the purpose of allowing revenue adequacy across the markets. FCM payments are paid to suppliers that agree and are selected to supply capacity three years in advance of delivery, and the ISO procures sufficient capacity to satisfy the 1-in-10 year loss of load planning objective on average.

Non-market products, not subject to an auction format include the following:

8. Voltage support service. This service helps the ISO maintain an acceptable range of voltage on the transmission system. The ISO regulates voltage through reactive power dispatch, and the generators that provide this service receive voltage support payments. This service is not purchased through a market.
9. Blackstart service. This service is provided by generators that are able to start quickly without outside electrical supply. The ISO selects and compensates strategically located generators for providing blackstart service in the event of a partial or complete system shutdown. This service is not purchased through a market.

Imports and exports from neighboring systems are referred to as “external transactions”. These imports and exports operate on an hourly schedule if they are private transactions and every 15 minutes if arranged based on a joint economic dispatch between the New York ISO and ISO-NE.

The products and their timelines are summarized below.

Time Horizon

Products	Forward	Day-Ahead	Spot
Energy		Day-Ahead Market	Real-Time Market
Capacity	Forward Capacity Market		Performance Payments
Ancillary Services	Forward Reserve		Performance Penalties
			Operating Reserve Mkt. Regulation
			Voltage Support
	Blackstart Procurement		
Congestion	Financial Transmission Rights	Congestion Payments	

Figure 6-1: Products by Time Horizon[16]

6.1.2 Market Participation

6.1.2.1 Market Participants

Any entity that meets the minimum requirements of financial assurance can become a Market Participant and participate in the markets. A Market Participant is a participant in the New England Markets (including a FTR-Only Customer) that has executed a Market Participant Service Agreement. This includes generation companies, investor-owned utilities, municipal utilities, federal agencies, state agencies, retail energy service providers, energy marketers, etc. The Market Participant is responsible for making supply and demand offers and is the financially responsible entity.

Those with physical assets must identify a “Designated Entity” to receive dispatch signals, report outages and maintain constant communication with ISO-New England operators. The Market Participant and Designated Entity can be distinct companies with the Designated Entity providing service to the Market Participant for a fee.

6.1.2.2 Participation by Type

ISO-New England markets allow participation of two types of physical entities: dispatchable resources such as generators and batteries and controllable load.

[16] “An Overview of New England’s Wholesale Electricity Markets: A Market Primer” (ISO-New England, June 5, 2023), https://www.iso-ne.com/static-assets/documents/2023/06/imm-markets-primer.pdf.

Resources make offers to inject energy or control load levels (demand response). Load refers to a physical point of withdrawal that is generally assigned a cost in different markets.

ISO-New England also allows participation by financial traders without a physical asset. These financial traders take positions in forward markets, such as the day-ahead market, that are settled at a spot market. Financial traders also can purchase FTRs and collect the associated congestion revenues. This is illustrated in the following tables:

Table 6-1: Participation in ISO-New England Markets

Participants	Day-Ahead Market	External Transactions (Imports and Exports)	Real-Time Market
Resources	✓	✓	✓
Load	✓	✓	✓
Financial Traders	✓	✓	

Table 6-2: Participation in ISO-New England Markets

Participants	FCM Auction	Forward Reserve Market Auction	FTR Annual Auction	FTR Monthly Auction
Resources	✓	✓	✓	✓
Load			✓	✓
Financial Traders			✓	✓

The Day-Ahead Market and the Real-Time Market differ slightly in the types of Resources (supply offers) and Load (demand bids) that participate, as seen in Table 6-3 and Table 6-4. Note that storage devices are modeled two types of resources, as a generator and as a dispatchable demand.

Table 6-3: Demand Bid Types in the Day-Ahead Market

Type	Description
Fixed Demand	A bid to purchase a specified MW amount at any price. This bid has no associated bid price and is willing to clear regardless of the market-clearing price. These must be bid in at the load zone level and are typically associated with physical load.
Price-Sensitive Demand	A bid that includes both a specified MW quantity and price. The participant is willing to clear this bid as long as the clearing price is no greater than their specified bid price. These must be bid in at the load zone level and are typically associated with physical load.
Asset-Related Demand (ARD)	ARDs are physical demands that are discretely modeled by the market software. They settle at a node, have average consumption exceeding 1 MW and may be either an end-use customer or one or more storage facilities. *Dispatchable* ARDs (DARDs) submit bids to consume energy with segments specifying a MW amount and price they are willing to pay. ARDs submit additional bid parameters, including min/max consumption levels, ramp rates, maximum daily starts, and others. An example of a dispatchable ARD is the pumping side of a pumped-storage facility, or the charging side of a battery.
Virtual Demand ("Decrement")	A virtual bid that is not associated with physical demand. It is a bid to purchase a specified MW amount at a chosen node for no more than the stated bid price.
Exports	A bid submitted to move energy out of the New England system.

Table 6-4: Demand Bid Types in the Real-Time Market

Type	Description	Dispatchable?
Load	The ISO produces short-term load forecasts based on expected wholesale demand from load-serving entities (LSEs). Load is non-dispatchable electricity consumption and does not respond to price. Load is ultimately settled based on metered consumption at the nodal level.	No
Dispatchable Asset-Related Demand (DARD)	Demand that is capable of having its energy consumption modified in Real-Time in response to Dispatch Instructions.	In ramp-feasible range between min and max consumption levels.
Export	Energy moving out of the New England syst6m.	Priced transactions are dispatchable. Fixed transactions are not dispatchable, but can be cut to ensure constraints are not violated.

The corresponding supply offer types in the Day-Ahead Market and the Real-Time Market are illustrated in Table 6-5 and Table 6-6.

Table 6-5: Supply Offer Types for Day-Ahead Market

Type	Description
Generator (including battery injection)	An offer submitted by a physical generator in New England to sell energy. A generator submits price and quantity pairs representing the MW amounts it is willing to supply at certain prices. Generators submit additional parameters, including start-up costs, ramp rates, minimum and maximum output levels, and others.
Demand Response Resource (DRR)	An offer submitted by a DRR indicating a willingness to reduce load at a specified offer price. DRRs submit price and quantity pairs representing the MW amounts they are willing to reduce their load at different price levels. DRRs submit additional parameter, including initiation cost, ramp rate, min/max reduction levels, and others.
Virtual Supply ("Increment")	A type of virtual offer that does not represent supply backed by a physical asset. It is an offer to sell a specified MW amount at a particular node for no less than its stated offer price.
Import	An offer submitted to deliver energy into the New England.

Table 6-6: Supply Offer Types for Real-Time Market

Type	Description	Dispatchable?
Generator (including battery injection)	Physical generation. Generators submit offers to supply energy with segments specifying a MW amount and price reflecting their willingness to produce energy. Generators submit several other parameters, including start-up costs, ramp rates, and minimum and maximum output levels.	In ramp-feasible range between minimum and maximum output levels.
Demand Response Resource (DRR)	Dispatchable load reduction. Demand	In ramp-feasible range between Min Reduction

	response resources submit offers to reduce their load with segments specifying a MW amount and price reflecting their willingness to consume less energy. DRRs submit other parameters, including initiation cost, ramp rate, and min/max reduction levels.	and Max Reduction
Import	Energy moving into the New England system.	Priced transactions are dispatchable. Fixed transactions are not dispatchable but can be cut to ensure constraints are not violated.

6.1.3 Aggregation and Distributed Resources

ISO-NE currently allows aggregated demand response resources (dispatchable demand) to aggregate within pre-identified dispatch zones where congestion is not expected. These zones must not cross reserve zones with local reserve requirements.

On February 2, 2022, ISO-NE, joined by the Market Participant organization and transmission owners filed a compliance proposal for FERC Order 2222 that would reduce the size of participating resources to 0.1 MW. The proposal creates two new participation models for the energy and ancillary services market (called Demand Response DERA and Settlement Only DERA) and modifies existing models to accommodate the physical and operational characteristics of DERAs. FERC has requested changes to proposal and no changes have been implemented.

6.1.3.1 Financial Assurance Requirements

Market participation requires meeting financial security requirements. This is particularly important for FTR participation. In ISO-New England this is described as financial assurance. Some key points for financial assurance are discussed below.[17]

Market Participation Minimum. The ISO reviews market participant financial reports, risk

[17] ISO-New England, "ISO New England Financial Assurance Policy, ISO-NE Open Access Tariff, Section I.i, Exhibit IA," 2022, https://www.iso-ne.com/static-assets/documents/2017/09/sect_i_ex_ia.pdf.

management procedures, credit rating agencies reports, and credit guarantees from financial institutions. The market participant is required to disclose significant events such as bankruptcy proceedings, litigation risks, sanctions, mergers, etc. There is a financial viability review that includes review of audited financials, credit rating reports, current. If there are credit ratings, a minimum of BBB-/Baa3 is required. Proof of a net worth greater than $1 Million is accepted for unrated participants, or total assets greater than $10 Million.

<u>Credit limits for market activity</u>. A higher credit rating allow allows a higher credit limit as a percent of tangible net worth. For example, an AA rating results in a credit limit equal to 4.5% of tangible net worth, whereas a higher AAA rating allows a higher credit limit fo 5.5% of tangible net worth. Purely financial traders who only participate in the FTR market are not granted a credit limit for market activity.

<u>New participants</u>. New participants and participants returning from defaults require a bank letter of credit or share of mutual funds.

<u>FTR financial assurance</u>. This requirement is calculated per FTR path. There is no fixed collateral requirement. Instead, the risk is measured based on the standard deviation of congestion on that path. In addition, the value of the possible default is measured with a mark-to-market process.

6.1.4 Spot Markets and Timelines

The spot markets identified by ISO-New England as illustrated in Figure 1 are:

1. Energy in the real-time market
2. Ancillary services
    a. Regulation
    b. Contingency reserve: TMSR
    c. Contingency reserve: TMNSR
    d. Contingency reserve: TMOR
    e. Voltage support (a cost-based service)

Energy and ancillary services (except voltage support) are co-optimized.

Regulation is not divided into up and down services.

It should be noted that the contingency reserve prices are determined by a demand curve and there are no financial offers. Instead, competition is based on availability and successful performance. Regulation has both a capacity offer, and a regulation service offer of $/MW for instructed movement. More details are provided on the ancillary service markets below.

Capacity scheduled for reliability purposes through the Resource Adequacy Assessment (RAA) is settled at Real-Time Market prices.

The timeline for the energy market is illustrated below in Figure 2. Key points are:

- The security constrained economic dispatch is every 10 minutes or less. This requires electronic dispatch signals to generators and not voice instructions.
- There is a continuous rolling window of short-term commitment that optimizes the use of fast-start units such as gas turbines. This short-term unit commitment looks out 2 hours.
- The pricing is every 5 minutes. Settlement is also at this 5-minute level.
- Energy offers can be constantly updated, even during the operating day. This was introduced to enable adjustments to natural gas prices during periods of fuel price volatility.
- Reliability unit commitment (Resource Adequacy Assessment) is a manual process that is continuous.
- The Re-Offer period that occurs after the Day-Ahead Market publication of 13:30 and before 14:00 allows Market Participants to adjust their offers that will be included in the first reliability commitment, i.e., RAA, for the next operating day. However these offers are settled at Real-Time Market prices.

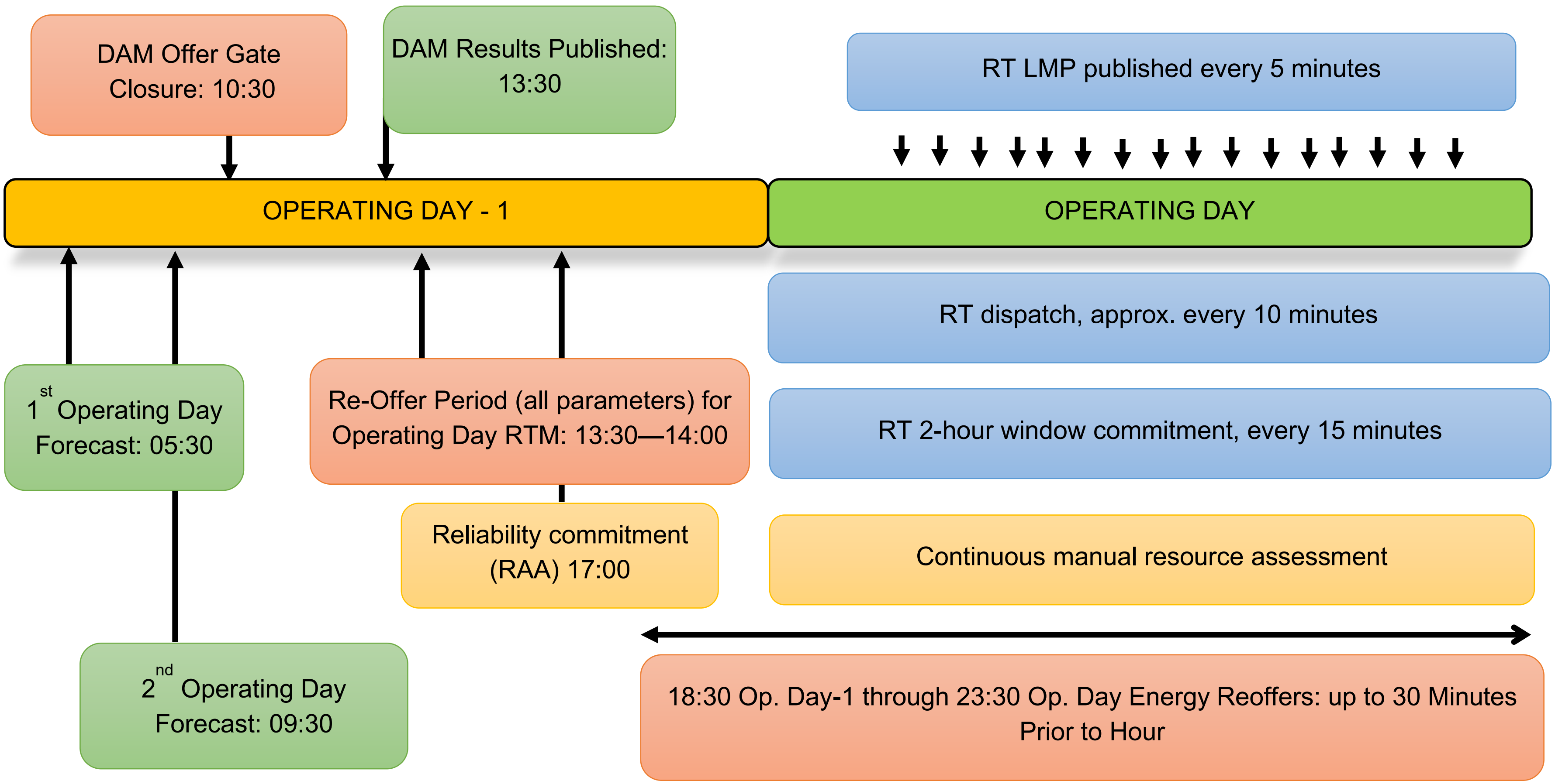


Figure 6-2: Energy Market Timeline, Unit Commitment and Dispatch

Imports and exports are referred to as “external transactions.”

6.1.5 Network Model

In calculating Day-Ahead Market prices, ISO-NE bases the system conditions on the expected transmission system configuration and the set of offers and bids submitted by Market Participants. In calculating Real-Time Locational Marginal Prices and Real-Time Reserve Clearing Prices, the ISO obtains a complete and consistent description of conditions on the electric network in the New England Control Area by using the power flow solution produced by the state estimator for the pricing interval, which is also used by ISO-New England for other functions within power system operations. The State produces a power flow model based on available real-time metering information, information regarding the current status of lines, generators, transformers, and other equipment, bus load distribution factors, and a representation of the electric network, to provide a complete description of system conditions, including conditions at Nodes and External Nodes for which Real-Time information is unavailable. In calculating Real-Time Locational Marginal Prices and Real Time Reserve Clearing Prices, ISO-NE obtains a State Estimator solution every five minutes, which provides the megawatt output of generators and the loads at Locations in the New England Control Area, transmission line losses, penalty factors, and actual flows or loadings on constrained transmission facilities. External Transactions (imports/exports) between the New England Control Area and other Control Areas are included in the Real-Time Locational Marginal Price calculation on the basis of the Real-Time transaction schedules implemented by the ISO’s dispatcher.

6.1.6 Pricing and Settlement

6.1.6.1 Energy

Locational Marginal Prices (LMPs) are used to settle Energy supply or demand. LMPs are calculated for over 1000 pricing nodes, load zones (an aggregation of pricing nodes), and the central trading hub (an aggregation of pricing nodes with historically low congestion). Each dispatchable generator has a separate pricing node. Load is charged the LMP for its load zone.

6.1.6.2 Ancillary Services

Operating reserves (TMSR, TMNSR, and TMOR) are procured in real time through a dispatch and pricing process that co- optimizes energy and reserves. The market dispatch and pricing software determines real-time reserve quantities and the prices for each reserve product. There are no distinct offer prices for reserve products. Clearing prices are opportunity-cost based, with the software determining prices that ensure a generator is no worse off providing energy or reserves. In other words, when a generator is instructed to a lower dispatch level to provide reserves rather than energy, it is compensated for any economic losses, due to not producing energy at a given LMP, through reserve clearing prices.

A reserve price above zero occurs when the pricing software must re-dispatch resources that would otherwise provide energy to satisfy the reserve requirement. This effectively holds a unit's dispatch point down when they would otherwise provide energy, and creates a price so they will be indifferent between providing reserves or energy. In nearly all cases, the reserve constraint is reflected in the energy price.

The regulation market consists of several key elements:

- the amount of regulation capability needed for a particular time interval (i.e., the regulation requirement),
- the supply offers of regulation-capable resources to provide regulation service and capacity,
- market clearing based on the least-cost combination of resources to satisfy the regulation requirements; the highest-priced resource(s) chosen to provide regulation sets the regulation clearing prices for service and capacity, and
- the regulation clearing prices which are used to determine the compensation (i.e., payments or "settlement") provided to resources providing regulation.

To administer the market, ISO-NE develops hourly regulation requirements for the real-time energy market (i.e., quantities of regulation "capacity" and "service" needed to ensure the reliable operation of the grid). Regulation resources provide supply offers indicating their availability and the cost (offer prices) for providing regulation. Regulation resources provide offers for "service" (the up and down movement of the resource while providing regulation) and "capacity" (a measurement of the MW range within which the resource is being moved up and down while providing regulation).

ISO-NE utilizes an optimization model to determine the least-cost combination of regulation resources (based on their offers) to meet the regulation requirement. When the market optimization has chosen the least-cost cohort of regulation resources, the regulation prices for "service" and for "capacity" are determined by the highest offer prices of the resources selected to provide regulation. Regulation prices are determined for each five-minute pricing interval in the real-time energy market.

The costs for ancillary services are allocated to load on a pro-rata MWh share basis.

6.2 Market Power Mitigation

6.2.1 Overview

In the context of the ISO-NE market rules, market power refers to "any actions or transactions that are without a legitimate business purpose and that are intended to or foreseeably could manipulate market prices, market conditions, or market rules for electric energy or electricity products." Market power mitigation refers to actions undertaken by the Internal Market Monitor (IMM) to "minimize interference with open and competitive markets, and thus to permit to the maximum extent practicable, price levels to be determined by competitive forces under the prevailing market conditions." The IMM administers defined mitigation processes for addressing the potential exercise of market power in both the ISO's energy and capacity markets.

6.2.2 Energy Market Mitigation

In the day-ahead and real-time energy markets, the IMM has implemented an ex-ante mitigation process to prevent the exercise of certain types of supplier-side market power. Energy market mitigation focuses on the economic withholding of generating capacity (supply) in the day-ahead and real-time energy markets. Absent mitigation, the economic withholding of generating capacity – depending on the need for the capacity – may result in (1) energy market clearing prices (LMPs) that exceed expected competitive price levels or (2) elevated uplift (NCPC) payments to generators.

To limit economic withholding and its market impacts, the ISO-NE IMM reviews energy market supply offers for generators in both the day-ahead and real-time energy markets. Under certain conditions, the IMM will mitigate generator offers; that is, the IMM will replace the financial parameters of a generator's supply offer (i.e., start-up, no load, and segment energy offer prices) with "reference" values. The reference values are intended to replicate a competitive offer for the generator.94 Use of the reference values allows energy market prices and payments to come closer to approximating a competitive outcome, by limiting the ability of a market participant with market power to economically withhold energy from the market.

6.2.3 Energy Market Mitigation Types

The IMM administers seven types of ex-ante energy market supply offer mitigation. Mitigation of offers is applied as the result of three types of tests: structure, conduct, and impact.

- **Structural test**: Represents a determination that market circumstances may confer an advantage to a supplier. This may result from (1) a supplier being "pivotal" (i.e., load cannot be satisfied without that supplier) or (2) a supplier operating within an import- constrained area (with reduced competition).

- **Conduct test**: Represents a determination that the financial parameters of a supply offer appear to be excessively high, relative to a benchmark offer value (a “reference” value). The conduct test applies to all mitigation types.
- **Impact test:** Represents a determination that the original supply offer would have a significant impact on energy market prices (LMPs). This test only applies to general threshold energy and constrained area energy mitigation types.

Supply offers are only mitigated when a violation of each applicable test occurs. For example, general threshold mitigation only applies when a supplier is pivotal, the offer prices for one (or more) of its generators exceed the conduct test thresholds, and the market impact of the economic withholding exceeds the impact test thresholds. The variation in tests across mitigation types reflects either market conditions associated with potential market power (transmission-constrained area vs. unconstrained area (general threshold)) or the likelihood that a participant’s offers could directly impact uplift payments (e.g., reliability commitment and manual dispatch).

The overall energy mitigation framework is summarized in Table 6-7.

Table 6-7: Energy Market Mitigation Types

Mitigation type	Structure test	Conduct test threshold	Impact test
General Threshold Energy (real-time only)	Pivotal Supplier	Minimum of $100/MWh and 300%	Minimum of $100/MWh and 200%
General Threshold Commitment (real-time only)		200%	n/a
Constrained Area Energy (import- constrained)	Constrained Area	Minimum of $25/MWh and 50%	Minimum of $25/MWh and 50%
Constrained Area Commitment (real-time only, import-constrained)		25%	n/a
Reliability Commitment	n/a	10%	n/a
Start-Up and No-Load Fee	n/a	200%	n/a
Manual Dispatch Energy (real-time only)		10%	n/a

6.2.4 Capacity Market Mitigation

The IMM administers two forms of mitigation for Forward Capacity Auction (FCA) bids and offers: supplier-side mitigation for existing resources and buyer-side mitigation for new resources (i.e., the Minimum Offer Price Rules (MOPR) for new resources).

6.2.4.1 Supplier-Side Mitigation

A market participant attempting to exercise supplier-side market power will try to economically withhold capacity during the FCA – for a single year or permanently – in an effort to increase the clearing price above a competitive level. An inflated clearing price can benefit the remaining resources in the market participant's portfolio, as well as the portfolios of other suppliers.

Delist bids are the mechanism that allow capacity resources to remove some or all of their capacity from the market for one or more commitment periods.[100] Delist bids specify the lowest price that a resource would be willing to accept in order to take on a capacity supply obligation (CSO). To restrict resources from leaving the market at a price greater than their competitive offers, the IMM reviews delist bids above a proxy competitive offer threshold called the dynamic delist bid threshold (DDBT) price.

A competitive delist bid is consistent with the market participant's net going forward costs, expected capacity performance payments, risk premium, and opportunity costs. All existing capacity resources, as well as certain types of new import capacity resources, are subject to the pivotal supplier test. If the IMM determines that a delist bid is uncompetitive and the supplier fails the pivotal supplier test, the IMM mitigates the delist bid to a competitive price (i.e., IMM's estimate of a competitive offer).

6.2.4.2 Buyer-Side Mitigation

A market participant attempting to exercise buyer-side market power will try to offer capacity below cost in an effort to decrease the clearing price to benefit the capacity buyer. In practice, the risk of price suppression in the ISO-NE market is largely due to out-of-market revenue streams inherently designed to incent new build of renewable generation to meet the states' environmental goals, as opposed to the exercise of market power.

To guard against price suppression, the IMM evaluates requests to offer capacity below pre- determined competitive threshold prices, or Offer Review Trigger Prices (ORTPs). The associated rules are referred to as the Minimum Offer Price Rule (MOPR). Market participants that want to offer below the relevant ORTP must submit detailed financial information to the IMM about their proposed project. The IMM reviews the financial information for out-of-market revenues or other payments that would allow the market participant to offer capacity below cost. The out-of-market revenues are either replaced with market-based revenues or removed entirely and the offer is recalculated to a higher, competitive price (i.e., the offer is mitigated).

6.2.4.3 Other IMM Monitoring and Potential Market Power Mitigation

The IMM also monitors the ISO's markets for other forms of potentially uncompetitive behavior. Several types of behavior or activities are specifically mentioned in the ISO's Tariff. These include:

- Physical withholding: Physical withholding involves participants attempting to influence market prices or other outcomes by physically withdrawing available capacity from the energy markets (e.g., false outage declarations or declining to make supply offers when it would have been in the participant's economic interest to do so).
- Physical supply offer parameters: Physical supply parameters that are not subjected to limitation within the ISO's supply offer software (eMarket) can be reviewed by the IMM for potentially uncompetitive behavior and failure to comply with Tariff- prescribed limits.

For example, an economic minimum offer is limited to being no more than double (100% greater than) the IMM's reference value; an economic maximum offer may be no less than 50% of its reference value.

- Increment Offers and Decrement Bids: Deviations between day-ahead and real-time energy market LMPs are monitored to determine whether they are consistent with competitive outcomes. The IMM will review participant activities that might have contributed to these deviations and the role of increment offers and decrement bids in that activity.
- FTR Revenues: The IMM monitors – and mitigates - the use of increment offers and decrement bids by the holders of financial transmission rights (FTRs). Increment offers and decrement bids can be used to create or magnify congestion that benefits the holders of FTRs.
- Cost of Service Agreements: The IMM reviews the supply offers for generators that have cost-of-service agreements with the ISO.

6.3 Day-Ahead Market

6.3.1 Overview

The Day-Ahead Market is a financial market that establishes 24 hours of financial obligations to either inject energy or withdraw energy. However, it also allows optimal unit commitment for resources with inter-temporal constraints and optimizes the output of limited energy resources such as hydro generation and pumped storage.

The Day-Ahead Market respects all expected transmission constraints, start-up and shut-down profiles of generators and other operational constraints.

The Day-Ahead Market permits bids and offers without associated physical assets that settle at the imbalance price established by the Real-Time Market. These are referred to as virtual bids and offers ("convergence bidding" in the California market).

6.3.2 Ancillary Services

ISO-New England does not have ancillary service bidding in either the Day-Ahead or Real-Time Market. In the Day-Ahead Market the ISO commits sufficient capacity to assure that that the operating (contingency) reserve can be met in each hour. In addition, the ISO commits additional capacity to cover uncertainty in generation by intermittent resources.

ISO-New England is proposing to create four new reserve products that would be co-optimized with energy and priced in the day-ahead market. These day-ahead reserve products are not forward sales of reserves that settle against real-time reserve prices; instead, they have a unique call-option settlement structure, where the settlement is based on a predetermined strike price and the actual real-time Hub LMP.

The proposed structuring of reserve as call options is distinct from other markets in the US. Under the new design, resources will submit a single energy option offer. That offer may be cleared, in the day-ahead market, for energy imbalance reserve, generation contingency reserve, or replacement energy reserve – the three new day-ahead ancillary services, discussed in detail below. That is, a market participant does not submit separate option offers for each type of ancillary service; it submits a single offer, and the market clearing process determines how that offer (and the physical capabilities of the associated resource) can most cost effectively serve the system's day-ahead ancillary service requirements. The award of all day-ahead ancillary services will be co-optimized (i.e., simultaneously cleared) with participants' day-ahead energy supply and demand awards.

The proposed new ancillary services are:

- Energy Imbalance Reserve to cover shortfalls in forecasted, primarily renewable, generation.
- Day-Ahead operating reserve including Ten-Minute Spinning Reserve, Ten-Minute Non-Spinning Reserve, and Thirty-Minute Operating Reserve.

The call option is proposed to have a strike price that represents the expected LMP. The strike price would be available to suppliers prior to when they make their offer.

This proposal has not been filed with FERC yet.

6.3.3 Energy Market

6.3.3.1 Offers

As previously described above in Table 1-5 the types of supply offers in the Day-Ahead Market are:

- Generator (including battery injection)
- Demand Response Resource (DRR, includes dispatchable demand "DARD")
- Virtual Supply ("Increment")
- Import

Note that storage devices are modeled as generators when injecting and dispatchable demands when withdrawing. ISO-New England distinguishes between technologies that can continuously switch between injection and withdrawal and those, such as pumped storage, that must pause between injection and withdrawal. Batteries (continuous) are not allowed start-up and no-load offers.

More details on each type off supply offer is provided below.

Generator Supply Offers

1. Energy offers are increasing blocks of ($/MWh, MW) pairs for each hour. These offers may change hour by hour.

2. Generator supply offers may include a start-up fee ($) and no-load fee ($/hour). These fees may change hourly.
3. Dual-fuel generators may identify the fuel used. For example, a generator that normally offers based on natural gas costs, may have to offer based on higher-cost diesel fuel due to scarcity of natural gas pipeline capacity. This would normally trigger offer mitigation, but if the offer identifies diesel as the input, the estimatedM cost is higher.
4. Minimum run time less than 24 hours.
5. Minimum down time.
6. Other physical operating parameters, including outages. Energy limitations (hydro or battery) can be represented as limitations on the hourly maximum.
7. Self-scheduled generation is treated as an energy offer at the energy offer floor price ($150/MWh)and no start-up and no-load fee.

<u>Demand Response Resource (DRR) Demand Reduction Offers</u>

Demand reduction offers are demand acting as supply.

1. DRR demand reduction offers, if cleared are paid the LMP.
2. DRR demand reduction offers are ($/MWh, MW) pairs for each hour. The $/MWh offer must be below a prescribed value calculated by ISO-New England to represent the price at which cost savings to consumers would be zero.
3. Interruption cost similar to a start-up fee may be offered each hour.
4. Minimum reduction time, maximum reduction time.

<u>Virtual offers</u>

Virtual offers can be made at any pricing node and must specify a single ($/MWh, MW) pair.

<u>Imports (External Transaction)</u>

Imports in the Day-Ahead Market are treated as purely financial, similar to a virtual offer, and may or may not flow in real-time.

Imports may be a) self-scheduled (energy offered at the offer floor price), b) priced (energy offered at a price and cleared if the external node LMP equals or exceeds the offer price, c) offered with a ($/MWh, MW) pair, and “up-to-congestion”. “Up to congestion” are imports that only clear if the LMP at importing node minus the LMP at the sink node is less than a $/MWh threshold.

6.3.3.2 Bids

As previously described above in Table 1-3 the types of demand bids in the Day-Ahead Market. Demand bids can vary by the hour. Demand is bid at a nodes and can take several forms.

- Fixed Demand

- Price-Sensitive Demand
- Dispatchable Asset-Related Demand (DARD)
- Virtual Demand (Decrement)
- Exports (External Transaction)

Fixed Demand

Fixed demand at a node is modeled as a bid for energy at the demand bid cap price.

Price Sensitive Demand

Price sensitive demand is a bid for energy with descending price ($/MWh, MW) pairs down to the bid floor.

Dispatchable Asset-Related Demand (DARD)

A DARD has both economic and physical characteristics, for example, minimum and maximum consumption levels. A DARD such as a pumped load may also require commitment and have associated physical characteristics.

Virtual Demand (Decrement)

Virtual demand bids can be made at any pricing node and must specify a single ($/MWh, MW) pair.

Exports (External Transaction, a virtual load)

Exports may be a) self-scheduled (energy offered at the bid cap price), b) priced (energy offered at a price and cleared if the external node LMP equals or is less than the offer price, c) offered with a ($/MWh, MW) pair, and “up-to-congestion”. “Up to congestion” are exports that only clear if the LMP at importing node minus the LMP at the injection node is less than a $/MWh threshold.

6.4 Resource Adequacy Assessment (RAA) Commitment

6.4.1 RAA Requirement

The RAA process uses the ISO’s load forecasts to make supplemental generator commitment decisions. During the RAA process, the ISO may determine that, based in part on their load forecast—or renewable generation forecast errors—that the day-ahead market scheduled insufficient capacity. The ISO can also commit additional non-fast start generators over what cleared in the day-ahead market to satisfy real-time load and reserve requirements.

6.4.2 RAA Commitment Process

The RAA is a continuous evaluation process that begins after the Day-Ahead Market results are posted at 13:30 and continues through the operating day. This is illustrated below. The RAA has a 36-hour horizon.

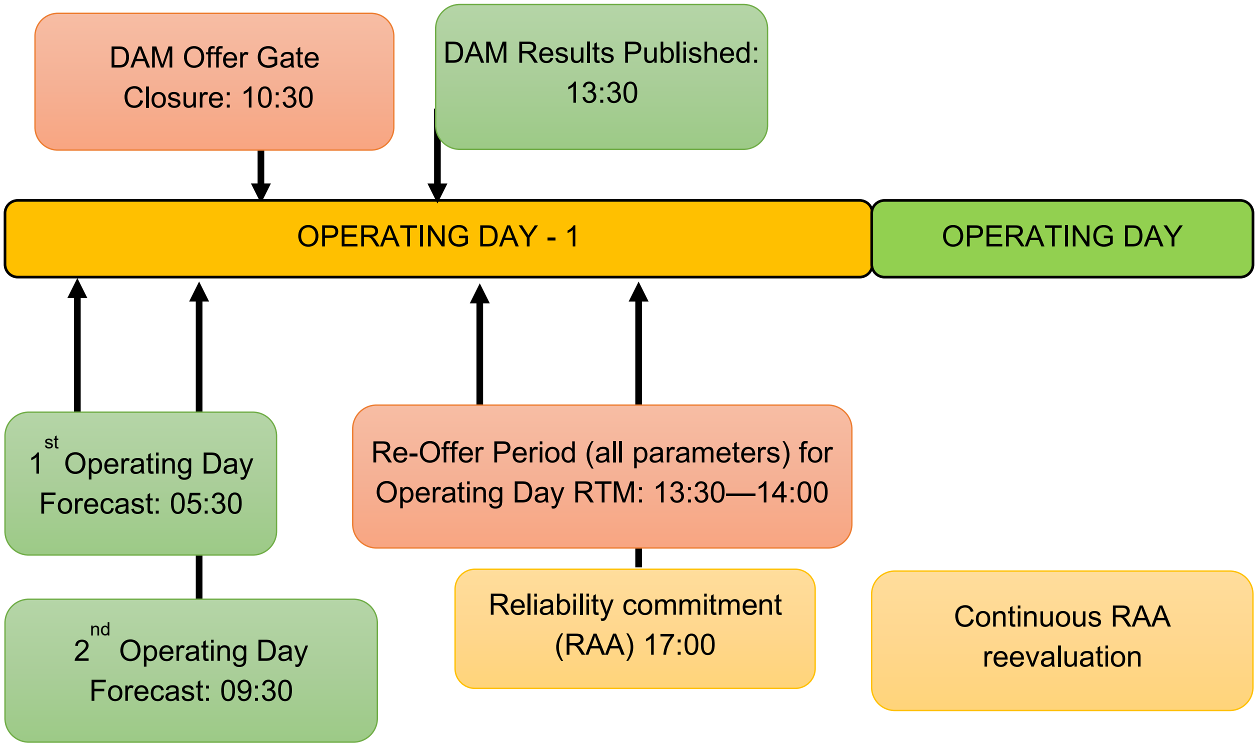


Figure 6-3: RAA Commitment Process

6.4.3 RAA Bidding and Selection

In the ISO-New England market the RAA is considered part of the Real-Time Market. The bids submitted at 13:30 are bids into the Real-Time Market. The selection is based on an estimation least cost but is not a full optimization.

6.4.4 RAA Pricing and Settlement Though Net Commitment Period Compensation

The RAA committed generators are paid the Real-Time Market LMP. The RAA does not result in a clearing price. Since the RAA generators were not economic in the Day-Ahead Market they usually have costs beyond the market prices in the Real-Time Market as well. When this is the case, they are paid through Net Commitment Period Compensation (NCPC).

6.4.5 Net Commitment Period Compensation

Net Commitment Period Compensation (NCPC) is referred to informally as “uplift”. Uplift is a make-whole payment made to electricity generators (mostly) when their as-offered production costs, as well as certain types of opportunity costs, are not fully recovered through the LMP. This revenue adequacy mechanism is crucial to ensuring generators face the appropriate incentives to follow ISO dispatch instructions.

Units that are committed or dispatched out of economic merit order, such as to satisfy a particular reliability need like local reserve or voltage support in the RAA, are more likely to rely on uplift payments. This is because their production costs were otherwise deemed to be too expensive.

The payment of NCPC seeks to ensure that generators are incentivized to follow ISO instructions as they are compensated to assure no losses due to ISO-NE instructions.

Table 6-8: NCPC Reasons

Reason	Reason Description	Market
Economic (First Contingency)	*Economic merit order* commitment and dispatch to meeting the system load and reserve requirements.	Day-ahead and Real-time
Local Second Contingency	*Out of economic merit order* commitments providing local second contingency protection in import-constrained areas.	Day-ahead and Real-time
Voltage	*Out of economic merit order* commitments providing voltage control in specific locations.	Day-ahead and Real-time
Distribution Reliability	*Out of economic merit order* commitments providing support to local distribution networks, also known as special constraint resource (SCR) payments.	Real-time only
Generator Performance Audit	*Out of economic merit order* commitments and dispatch paid to satisfy the ISO's performance auditing requirements, specifically for dual fuel generators testing on oil.	Real-time only

6.5 Real Time Market

6.5.1 Overview and Timeline

The Real-Time Market is a physical market that is based on security constrained economic dispatch (SCED). The SCED software runs approximately every 10 minutes to produce MW target dispatch points. A separate program runs every 5 minutes to produce LMPs. The Real-Time Market optimizes the joint and simultaneous procurement of energy and reserve ("co-optimization").

6.5.2 Short-Term Unit Commitment

A market software process called Multi-Interval Look Ahead Commitment (also known as real- time unit commitment or "RTUC") makes fast-start resource commitment recommendations.20 RTUC is automatically executed every 15 minutes, and operators approve RTUC process solutions manually during the operating day. Operators can also execute RTUC manually if needed. Another market software process called Contingency Dispatch (CDSPD) can also commit fast-start resources in real time. CDSPD does not run regularly; operators only use this process in emergencies when the supply/demand balance is tight and quick software optimization times are important.

Additional non-fast start real-time commitments occur in the Security Constrained Reserve Adequacy (SCRA) process, which runs multiple times during the operating day, at pre- determined schedules and as required by system conditions. If this process determines that the existing schedule is not sufficient to meet operating reserve requirements, then the operators can commit additional resources.

6.5.3 Real-Time Security-Constrained Economic Dispatch

Real-time dispatch instructions and LMPs can vary at the five-minute level but, in practice, vary with the frequency in which the ISO runs the Unit Dispatch System (UDS). The ISO generally runs UDS every five to 40 minutes to optimize for a period of 10 to 15 minutes ahead of market clearing. LMPs are effective in the UDS target interval (i.e., the time period 10-15 minutes after UDS is run). UDS produces unit dispatch instructions with the same frequency. The ISO communicates dispatch instructions immediately when the operators approve the market clearing result.

6.5.4 Ancillary Services

The ancillary services procured in the Real-Time Market are:

1. Regulation
2. Contingency reserve: TMSR
3. Contingency reserve: TMNSR
4. Contingency reserve: TMOR

Voltage support is procured through non-market means in the same time horizon.

6.5.4.1 Regulation

The regulation market consists of several key elements:

- the amount of regulation capability needed for a particular time interval (i.e., the regulation requirement),
- the supply offers of regulation-capable resources to provide regulation service and capacity,
- market clearing based on the least-cost combination of resources to satisfy the regulation requirements; the highest-priced resource(s) chosen to provide regulation sets the regulation clearing prices for service and capacity, and
- the regulation clearing prices which are used to determine the compensation (i.e., payments or "settlement") provided to resources providing regulation.

The ISO develops hourly regulation requirements for the real-time energy market (i.e., quantities of regulation "capacity" and "service" needed to ensure the reliable operation of the grid). Regulation resources provide supply offers indicating their availability and the cost (offer prices) for providing regulation. Regulation resources provide offers for "service" (the up and down movement of the resource while providing regulation) and "capacity" (a measurement of the MW range within which the resource is being moved up and down while providing regulation).

The ISO-NE regulation market is not split into up and down regulation products.

The ISO adjusts the generator's regulation compensation to reflect performance while providing regulation. For a generator to receive full compensation for providing regulation, it must control the generator's output, such that the output is within a performance envelope. The generator's performance score can be negatively affected by failing to maintain output within the performance envelope.

The regulation service compensation represents the absolute value of the movement in the target instantaneous output multiplied by the regulation service prices, and is adjusted by a generator's performance score (i.e., its ability to provide regulation within the performance envelope). The capacity compensation similarly is the amount of capacity provided by the generator multiplied by the capacity prices, and is adjusted by a generator's performance score. Regulation capacity is defined as the minimum of (1) five times the automatic response rate (i.e., the rate (MW/Minute) at which the generator can change its output), and (2) one-half of the difference between the regulation high and low limits.

6.5.4.2 Operating Reserve

The ISO maintains three contingency reserve products with varying levels of quality to respond to a contingency loss. Assets are assessed on their capability to respond to dispatch instructions within:

- Ten-Minute Spinning Reserve (TMSR), provides energy within 10 minutes from an online state (highest quality)

- Ten-Minute Non-Spinning Reserve (TMNSR), provides energy within 10 minutes from an offline state (middle quality),
- Thirty-Minute Operating Reserve (TMOR), provides energy within 30 minutes from an offline or online state (lowest quality).

In addition to system reserve requirements, there are local reserve requirements in three import-constrained zones: Connecticut (CT), Southwest Connecticut (SWCT), and NEMA/Boston. Similar to system reserve requirements, local reserve capacity requirements are based on the largest loss (generator or transmission) in the import-constrained zone.

The market dispatch and pricing software determines real-time reserve quantities and the prices for each reserve product. There are no distinct offer prices for reserve products. Clearing prices are opportunity-cost based, with the software determining prices that ensure a generator is no worse off providing energy or reserves. In other words, when a generator is instructed to a lower dispatch level to provide reserves rather than energy, it is compensated for any economic losses, due to not producing energy at a given LMP, through reserve clearing prices.

A reserve price above zero occurs when the pricing software must re-dispatch resources that would otherwise provide energy to satisfy the reserve requirement. This holds a unit's dispatch point down when they would otherwise provide energy, and creates a price so they will be indifferent between providing reserves or energy.

Penalty Factors (also known as demand cuves for reserve)

Reserve Constraint Penalty Factors (RCPFs) are the maximum cost (a price cap) the market is willing to incur in order to meet a reserve constraint. The software will not re-dispatch resources to meet reserves at any price; when the re-dispatch costs exceed the RCPF for a product, or the available reserve capacity is less than the requirement, the price cap takes effect. At this point, the market software stops re-dispatching resources to meet reserves, limiting the re-dispatch costs incurred to satisfy reserve requirements.

The RCPF is added to the energy price (LMP) due to their interdependence in procurement and signals reserve scarcity in real time. Further, certain RCPF prices trigger capacity scarcity conditions under the Pay for Performance rules (see Section 9). Each reserve product has a corresponding RCPF.

The RCPF is added to the energy price (LMP) due to their interdependence in procurement and signals reserve scarcity in real time. Each reserve product has a corresponding RCPF, as shown in Table 6-9.

Table 6-9: Reserve Constraint Penalty Factors

Requirement	RCPF ($/MWh)
Ten-Minute Spinning Reserve Requirement (10-min spinning)	50
System Ten-Minute Reserve Requirement (10-min non-spinning)	1,500
System Minimum Total Reserve Requirement (30-min)	1,000
System Total Reserve Requirement (30-min)	250
Local Zonal Reserve Requirement	250

The optimization formulation of the reserve constraints assures that the lowest quality reserves are sacrificed before the higher quality reserves are sacrificed. The resulting shadow prices are cumulative so that if TMOR is sacrificed, the prices of TMNSR and TMSR are equal to or exceed the price of TMOR. If both TMOR and TMNSR are sacrificed, then the price of TMSR equals or exceeds their prices.

6.6 Forward Reserve

ISO-NE procures operating reserves on a forward basis through an auction. It conducts two auctions each year, one for each summer and winter reserve period (June through September and October through May, respectively). The auctions procure operating reserve capacity for both the control area and local reserve zones. The auctions create a contractual obligation between the buyer of operating reserves (the ISO) and sellers of operating reserves (Market Participants) to provide operating reserves in the real-time energy market. The sellers of forward-contracted operating reserves receive forward operating reserve payments in lieu of “spot” payments for real-time reserves.

Two types of reserve products are procured in the auctions: ten-minute non-spinning reserves (TMNSR) and thirty-minute operating reserves (TMOR). Prior to the auctions, the ISO determines the quantity of TMNSR and TMOR needed to ensure system and local reliability. To do this, the ISO identifies significant operational contingencies at the system and local level. The contingencies are converted to reserve requirements that are used in the auction to determine the amount of operating reserve to be procured. The TMNSR requirement for the system is based on the forecasted first contingency, while the TMOR requirement for the system is based on the forecasted second contingency.

In the auction, resource owners make offers to supply operating reserves. The supply offers specify the q*uantity* of reserve supply available, the *location* of the reserve supply, the *price* needed to provide the reserve capacity, and the *type* of reserve supply (i.e., TMNSR or TMOR). The supply offers are used to construct supply curves for the auction; the auction's clearing prices and quantities are determined by the intersection of the reserve requirements for each product and the respective supply curves.

Participants that assume a forward reserve obligation must follow several steps during the delivery period to satisfy the requirements associated with that obligation:

1. The first step is the "assignment" of the forward reserve obligation. Assignment refers to a participant with a forward reserve obligation assigning that obligation (Forward Reserve MWs) to specific energy market assets (e.g., generators, demand reduction resources) capable of delivering the obligation in the real-time energy market. Typically, these are offline, fast-start capable assets that can respond to the ISO's dispatch instructions within either 10 or 30 minutes; however, online assets also may have obligations assigned to them.
2. The second step is the "delivery" of the obligation. Delivery requires that a portion of dispatchable capacity (equal to or greater than the FRM obligation MWs) must be offered into the real-time energy market at a price equal to or greater than the FRM "threshold price."[56] If a participant is unable to deliver its forward reserve obligation MWs into the real-time energy market, it may enter into a bilateral contract to temporarily transfer the obligation to another Market Participant.

Participants with forward reserve obligations have agreed to accept the forward reserve payments in lieu of real-time energy market (spot) reserve revenue. This applies during the hours when participants have a delivery obligation and have satisfactorily delivered upon that obligation in the real-time energy market. Outside of the FRM delivery period, all resources providing reserve capacity (designations) in the real-time energy market receive real-time energy market (spot) reserve revenue.

Participants with forward reserve obligations are subject to two types of penalties: failure to deliver and failure to activate. Failure to deliver occurs when a participant's delivered (or assigned) MWs are less than the FRM obligation MWs. Failure to activate occurs when an asset with an FRM assignment is unable to physically deliver the reserve capacity in the real-time energy market (such as failing to start-up at the ISO's dispatch instruction). Both a failure to reserve and failure to activate result in the forfeiture of FRM payments for the applicable delivery period and additional financial penalties.

6.7 Financial Transmission Rights

6.7.1 Overview

Financial Transmission Rights (FTRs) in ISO-New England, as elsewhere in the US, are forward products that are settle based on Day-Ahead Market congestion differences between two locations. Because they are a forward market products and not options, the settlement may require the owner to make payments in the case that the direction of congestion is counter to the direction of flow contracted for.

6.7.2 FTR Products and Auctions

The FTR market of the ISO-NE is very similar with the FTR markert of PJM so details of the markets are included in the PJM Chapter.

Market participants can obtain FTRs by participating in ISO-administered auctions for annual and monthly products. There are separate auctions for on-peak and off-peak hours.81 The FTRs awarded in the two annual auctions have a term of one calendar year (i.e., January 1 to December 31), while the FTRs awarded in one of the monthly auctions have a term of one month. FTRs can be purchased in all auctions but can only be sold in the second annual auction or the monthly auctions as only FTRs that are owned (i.e., have been purchased) can be sold by participants (i.e., there is no short selling).

The offer parameters are contained in the Table 6-10 below.

Table 6-10: FTR Offer Paarameters

Element	Description
Path	FTRs are defined between two points (i.e., pricing nodes): 1) the point of injection (or the “s ource”) and 2) the point of withdrawal (or the “sink”)
Price	The $/MW value the participant i s willing to pay to acquire the FTR
MW-amount	The size of the FTR (in MWs) the participant is willing to buy
Term	The monthly or annual period to which the FTR applies (e.g., November 2021)
Period	The hours in which the FTR applies (i.e., on-peak or off-peak)

FTRs can be purchased for any month remaining in the calendar year. This allows FTR holders to, for example, purchase FTRs for the peak summer months three months in advance of the summer. FTRs can also be sold and resold among market participants for the future months.

The FTRs auctions are based on a simultaneous feasibility test. Only a fraction of the transmission capacity is sold in any one auction so as to allow auctions in following months and to account for unforeseen transmission capacity reductions.

The revenue from the FTR auctions are allocated to holders of Auction Revenue Rights (ARRs). These holders include both: (1) market participants that paid for transmission upgrades that made the additional sale of FTRs possible and (2) congestion-paying LSEs. The former group is referred to as Incremental ARR (IARR) holders. The MW-value of ARRs they receive is based on the additional amount of FTRs in the FTR auction that their transmission upgrade made possible. The remaining ARRs are allocated to congestion-paying LSEs. The majority of auction revenue is allocated to congestion-paying LSEs.

6.8 Resource Adequacy (RA) Mechanism: Forward Capacity Market

In the ISO-New England market the resource adequacy mechanism is the Forward Capacity Market (FCM). This resource adequacy mechanism uses an auction that acquires rights to capacity three years in advance. This is the Forward Capacity Auction (FCA).

6.8.1 Forward Capacity Auctions

The FCA is designed to procure sufficient capacity to satisfy the 1-in-10 year loss of load planning objective on average. It includes local requirements for foreseen transmission constraints and accounts for retirements. Resources that do not receive a Capacity Supply Obligation (CSO) because their bids were too high and they are not required for a local constraint, are “delisted” and can retire. Delisted resources are not required to offer into the energy market.

After the FCA, capacity resources can modify their positions in one of the three annual reconfiguration auctions (ARA1, ARA2 and ARA3), The last opportunity that a capacity resource has to alter its CSO before the delivery period is during a monthly reconfiguration auction.

The FCA is based on descending clock auction design. The auctioneer starts at a high price at which all, or almost all, resources wish to stay in the market. As the auctioneer decreases the price more and more resources exit the auction. The FCA includes an administratively determined demand curve representing the value of capacity to the region. The auction continues until the bid prices exceed the price on the demand curve.

The FCA includes both import-constrained zones and export-constrained zones.

After the completion of the FCA, capacity resources have numerous opportunities to adjust their CSO position. For example, a resource owner may wish to adjust its position to account for expected availability of capacity for the delivery period or expected performance during capacity scarcity conditions.

The ISO holds three annual reconfiguration auctions (ARAs) in the three years between the FCA and delivery period. The ARAs are sealed-bid auctions where capacity resources submit bids and offers before the auction. Capacity resources that participate in the ARAs can submit supply offers (increase CSO position) or demand bids (decrease CSO position) depending on their desired adjustment. There are also shorter-term monthly reconfiguration auctions.

6.8.2 Delivery

6.8.2.1 Energy Must-Offer Requirement

The must-offer rule creates the physical obligation that capacity resources must be available for energy dispatch by the ISO. It requires a capacity resource to submit energy offers into the day- ahead and real-time energy markets at a quantity greater than or equal to their CSO to the extent that it is physically available.

6.8.2.2 Failure-to-Cover Charges

In the event that a resource faces a delay to the start of their commercial operation, it is subject to a failure-to-cover charge. The failure-to-cover charge is applied to capacity resources whose maximum demonstrated output (MDO) is less than their CSO. A resource's MDO is the highest MW output measured in the prior six years. For new resources without any historical MW output, their MDO is replaced with a demonstration of sufficient installed capacity just prior to energy delivery. Any resource with unproven capacity (CSO > MDO) will be charged for failing to cover their obligation.

6.8.2.3 Pay-for-Performance (PfP) Incentives

Resource performance is measured during a capacity scarcity event, which occurs when the system is short of offline reserves and the associated reserve-constraint penalty factors are triggered in the Real-Time Market. Performance is measured as availability of either energy or reserve in an amount equal to the CSO.

The rules establish a performance penalty (under-performance) or performance premium (over-performance) that is applied to the resource. This performance rate is set to escalate over time in the following schedule.

Table 6-11: Forward Capacity Market Performance Rates

Performance Period	Performance Rate (negative for under-performance, positive for over-performance)
2021-2024	\$3,500/MWh
2024-2025	\$5,455/MWh
2025 and beyond	\$9,337/MWh

Resources that have not sold capacity and do not have an obligation but do so anyway are eligible to *receive* the performance rate. These payments come from those resources that have sold capacity, have an obligation, and did not perform. The total paid out equals the total paid in.

6.9 Low Carbon Assets and Market Design

6.9.1 Demand Response

ISO-New England provides ways for demand response to participate in energy, ancillary services, and the forward capacity market.

Demand response resources (DRR) are dispatched based on price, not an emergency declaration. A DRR must satisfy the following requirements:

1. Have a capacity ≥ 0.1 MW
2. Dispatchable
3. Offer into energy market, dispatched on basis of price
4. Metered at the retail delivery point
5. May be aggregated if all physical assets are within the same reserve zone and dispatch zone

<u>Energy</u>: Demand Response Resources (DRR) submit demand reduction offers into the Day-Ahead and Real-Time Energy Markets, and will be committed and dispatched when economic compared to other resources.

<u>Ancillary Services</u>: DRR can be co-optimized to provide energy and/or reserves to meet both energy and reserve requirements. DRR can participate in regulation.

<u>Forward Capacity Market:</u> DRR may be aggregated in dispatch zones as one asset in the FCM. The DRR must provide dispatchable energy or reserve in real-time to meet performance requirements.

Demand response assets can be aggregated into dispatch zones as long as these do not cross reserve zone boundaries. ISO-NE has identified 20 DRR zones, although these can change.

The following bidding parameters are allowed. Those in *italics* can be reoffered after the Day-Ahead market has cleared.

Available/Unavailable

- Notification Time
- Start-up Time
- Minimum Reduction
- Maximum Reduction
- Price/Quantity Pairs
- Interruption Cost
- Minimum Reduction Time
- Min. Time Between Reductions
- Ramp Rate
- Offered Claim 10 (10-minute reserve)
- Offered Claim 30 (30-minute reserve)

In order to provide reserve service, DRR must have telemetry at a minimum of one-minute granularity.

Settlement uses a baseline methodology.

6.9.2 Solar and Wind Generation

ISO-New England requires each solar and wind generator to submit supply offers. These offers are used to reduce generation output when local or system reliability constraints become binding. When there is a binding constraint that must be respected each solar or wind generator receives a "Do Not Exceed Dispatch Point." This dispatch MWh amount is the lesser of: (1) the maximum output level at which the resource would operate based on its offer curve and Real-Time Prices, and; (2) a reliability limit representing the maximum acceptable output that is consistent with reliability constraints. In effect the solar or wind plant must have a low offer in order to continue generating.

The reliability limit is determined, in part, using a "high confidence" forecast of the potential unconstrained output of each solar or wind plant for the next dispatch interval. During a dispatch interval, solar or wind plants are free to operate at any level between zero and the resource's Do Not Exceed Dispatch Point.

ISO-NE maintains a centralized forecast of wind and solar generation for real-time operation and the calculation of the Do Not Exceed dispatch point. This centralized forecast is based in part on current meteorological data.

Telemetry requirements frequency as thermal generators but with additional meteorological data including: wind speed and wind direction, ambient air temperature, ambient air pressure, and ambient air relative humidity.

6.9.3 Storage Market Participation

ISO-NE does not directly manage the state-of-charge of batteries. Rather, it requires certain duration of market services, such as the ability to provide one hour of contingency response in order to qualify as reserve. Thus, while not directly managing the state-of-charge of a storage device, the charge is recognized as a limitation on the ability of the resource to provide a service.

The ISO-NE distinguishes between two types of storage:

a) Continuous Storage Facilities can transition seamlessly between charging and discharging (and vice versa) and that can charge or discharge at any MW level within their range.
b) Binary Storage Facilities, e.g., pumped-storage hydroelectric units, cannot switch nearly instantaneously from charging to discharging nor operate continuously across the boundary between their negative and positive MW ranges. These facilities have separate models for their generation and pumping states.

The available energy for each (state-of-charge for a battery) and available storage are both telemetered to ISO-NE in real-time. ISO-NE uses this information to automatically adjust the maximum output, maximum consumption and other operating limits. The change in operating limits affects eligibility for market compensation for capacity, energy, and ancillary services. It is also used to determine whether a resource can be considered as dispatchable.

ISO-NE requires that dispatchable energy resources be able to maintain a dispatch point for 15 minutes. Reserve markets require that the resource be able to sustain its output for one hour. For example, to ensure that an Electric Storage Facility (battery) would be able to follow a dispatch instruction to consume at its maximum capability for 15 minutes, the Maximum Consumption Limit of its dispatchable operation must be revised down if its Available Storage drops below 15 minutes at its bid-in Maximum Consumption Limit.

To ensure that ISO-NE is able to rely on storage resources to address reliability requirements and to make storage resources have an adequate state of charge to provide the service in question, ISO-NE provides for the automatic de-rating of Continuous Storage Facilities to meet the minimum run times. The telemetered values provide ISO-NE with the maximum amount of time the facility is able to receive or inject energy at the facility's operating limits and at the facility's current rate of charge or discharge.

Other parameters (as defined by FERC) that the ISO-NE uses are:

a) Maximum Charge Limit and Maximum Discharge Limit,
b) Discharge Ramp Rate and Charge Ramp Rate, and
c) Minimum Charge Limit and Minimum Discharge Limit
d) Minimum Charge Time and Minimum Run Time

Binary storage facilities (pumped storage) must offer a combined notification and start time of less than 30 minutes.

Continuous storage facilities must have a zero value for the minimum down time, and combined notification and start time or zero. The start-up and no-load fee are also set to zero.

Electric storage facilities can make use of two other existing bid parameters in the Day-Ahead Market to voluntarily allow ISO-NE to optimize limited energy generation and consumption in the Day-Ahead Market. The Maximum Daily Energy Limit parameter is the maximum amount of MWhs that a limited energy resource expects to be able to supply in the next operating day, and the Maximum Daily Consumption Limit parameter is the maximum number of MWhs that a dispatchable demand (DARD) expects to consume in the next operating day. If used, these parameters set a limit on the number of MWhs the Electric Storage Facility will clear in the day-ahead market, for discharging and charging, respectively.

Like all generators and dispatchable demands energy storage facilities can electronically revise the price-quantity pairs included in their offers and bids prior to each operating day, and in real-time up to 30 minutes prior to the start of each hour. This helps then manage their charge and discharge activity.

Finally, SO-NE the right to manage any resource (including an Electric Storage Facility) for reliability purposes. This has historically been the case for pumped storage facilities that have had their output restricted during periods of forecasted reserve shortage. Thes reliability actions make the facility eligible for make-whole payments, they eligible to receive NCPC credits for the out-of-merit dispatch of a DARD, e.g., mandated pumping at high energy prices, and the lost opportunity costs of a generator, e.g., reduced dispatch of a pumped storage generation during high energy price periods.

6.10 Virtual Bidding Market

As noted in the section on the Day-Ahead market, virtual offers (“increment offer”) and virtual bids (“decrement offer) may be made at any pricing node (or zone). However, two protections are in place to prevent undesirable impacts on the Day-Ahead Market.

a) <u>Bidding restrictions</u>: If the Internal Market Monitor (IMM) determines that (i) the average hourly deviation between the Day-Ahead Market and the Real-Time Market computed over a rolling four week period is greater than ten percent (10%) or less than negative ten percent (-10%), and (ii) the bid and offer practices of one or more Market Participants has contributed to a divergence between LMPs in the Day-Ahead Energy Market and Real-Time Energy Market, then the IMM investigates and may prevent the Market Participant’s from making any virtual bids or offers for six months.

b) Cap on FTR revenues (recuperation of FTR revenue): If a Market Participant makes virtual offers or bids near an FTR path owned by the Market Participant that result in Day-Ahead Market congestion exceeding that in the Real-Time Market, then the FTR revenues for that path are set to zero. “Near” is defined by a power flow that results in 75% or more of the virtual power flowing over the path of the FTR.

7 Identification and Analysis of Various Design Options for the Bid-Based Design Models Suitable for Chile

7.1 Overview

In this Chapter we present a detailed analysis of various energy design parameters and offer recommendations which will be used in subsequent Tasks to determine the optimal energy market design for CEN.

The design parameters are the following:

1. Price Formation (Pay-as-Clear vs Pay-as-Bid)
2. DAM Market Design Architecture (Unified vs. Split DAM by Technology)
3. Long Term Contracts PPAs (Physical and Financial) for RES deployment with simple CfDs vs CfD variants with increased price exposure and CfDs with Revenue cap and floor
4. Financial Market vs Non-Financial Market provisions
5. Energy Only Markets with Scarcity Provisions vs Capacity Adequacy Mechanisms (or Both)
6. Capacity Market with flexibility enhancements vs Capacity Adequacy Schemes (like Optimized CM or Capacity payment or Centralized reliability options)
7. Competition in Concentrated Markets (Ex-Ante Market Power Mitigation vs. Ex-Post Market Power Mitigation) (or both)
8. Virtual Bidding Market vs. Physical Market only
9. Unit Commitment Costs (Startup/No load costs) & Ancillary Services Bidding Format
10. Hydro Power Treatment in a Bid-Based Energy market Architecture

7.2 Price Formation (Pay-as-Clear or Marginal vs. Pay-as-Bid)

7.2.1 Overview

In this Section we analyze the merits and drawbacks of these two pricing mechanisms. The auction format can have an impact on the market outcomes. Under the pay-as-clear format all the winning projects receive the highest accepted price offer; and in a pay-as-bid format all the winning projects receive their own bid. Under both formats, the projects that offer the lowest prices are selected first until all the demand in the auction has been covered. These formats have been widely studied and we have actual practical experience from both. While some of the results are mixed, several robust conclusions emerge. First, if the auctions are competitive, the outcome of the two pricing rules is the close to each other. However, this is rarely the case. Most energy markets deploy the marginal pricign rule.

Based on our experience the debate to transition from a marginal pricing-based architecture to a pay-as-bid scheme is typically predicated on the belief that the energy price spikes reflect the exercise of monopoly power. This is clearly the framework of the debate for the 2022 EU energy crisis.

It is obvious that each option will trigger a very different response from the market. Based on our experience, under the uniform-pricing model, suppliers have every reason to bid approximately their marginal opportunity costs for energy in each of the blocks of power that they offer. On the contrary under the pay-as-bid market architecture participants tend to second guess the marginal price and bid accordingly.

It has been established that the consequence of the transition to a pay-as-bid scheme would be a radical change in bidding behavior that would:

- Never result in anticipated savings
- introduce unmeasurable inefficiencies in the dispatch of power and impose new costs on generating companies, which would inevitably tend to increase rather than decrease average prices
- tend to weaken the competition in generation that is the best safeguard against exertions of monopoly power such as may contributed to elevated prices; and
- impede—again to an unmeasurable extent—the expansion of capacity

The analysis below provides extensive arguments that any attempt to regulate the generators' bids (possibly for each technology) is futule and immensely counterproductive.

7.2.2 Pay-As-Bid vs. Uniform Price in a Price-Taking (Competitive) Environment

We now examine the merits and drawbacks of these two bidding systems. The theoretical foundation of the marginal pricing is provided in Appendix A.

Under uniform-pricing rules, generators have every reason to bid approximately[18] their marginal opportunity costs for energy[19] in each of the blocks of power that they offer. They know that if any of those bids is rejected because there are sufficient lower bids to satisfy the demand, they will be better off, because they will not have committed themselves to sales at prices that fail to cover their avoidable costs. More important, they

---

[18] Bids will tend to deviate from marginal cost to the extent units of capacity are large relative to total load.

[19] In most cases the marginal opportunity cost is just the incremental cost of generating additional energy. For hydro power, however, it has little to do with physical operating costs, consisting rather of the revenue or value sacrificed by using or selling it today rather than later or in one place in Greece rather than elsewhere, both of which depend in turn on how full the reservoirs are and expectations about future prices. Even for fossil plants, the marginal opportunity cost may differ from incremental operating costs to the extent there are opportunities to sell the energy in other markets in and out of their market.

know also that on their accepted bids they will receive the full benefit of whatever price above that level is necessary to equate demand and supply in the market, regardless of the level of their own bids, permitting them to pocket the difference between their avoidable costs and the market-clearing price as a necessary contribution toward recovery of their fixed charges and profits. Specifically, for markets where a dominant player exercises monopsony power and bids below Mimimum Variable Costs other Market Participants should bid based on their marginal costs rather than their MVC. This market behavior is consistent with "healthy" economic behavior and fully in compliance with the regulator's monitoring mechanisms.

The naïve expectation of advocates of pay-as-bid schemes is that since all the infra-marginal bids—the ones below the highest marginal cost output necessary for the sum total of accepted bids to satisfy market demand — will under **uniform pricing receive more than their bid prices** (by margins successively larger as the accepted bids range downward from the marginal, highest to the lowest cost), the pay-as-bid architecture would simply wipe out those markups. This in turn would mean that the average price purchasers will have to pay under pay-as-bid will incorporate no markup above their bids. For example, if the successful bids for a particular hour were of equal blocks of output with incremental costs, successively, of €30, €40, €50, €60 and €70 per MWh, the market clearing price of €70 will under the uniform price system bestow on the successful bidders markups above marginal costs of €40, €30, €20, €10 and zero, respectively. On the other hand, the pay-as-bid scheme will reduce those markups all to zero: the block bid at $30, reflecting avoidable costs of $30, will receive a price of only $30; and so on.

The critical assumption is, of course, that under the pay-as-bid market rules, generators will bid just as they had before. **The one absolute certainty, however, is that they will not**. Knowing that this scheme does not yield any contribution to their fixed costs, let alone profits—they obviously will universally change their practice immediately, bidding instead at what they *expect* will turn out to be the market-clearing price—$70 in the foregoing simple example. This observation has been established for many years now in the international markets.

All markets in the US and most EU markets settle with the marginal pricing rule. Rarely though we observe the pay-as-bid rule implemented in practice. One example is the RTBM market in Italy (TERNA). TERNA settles the RTBM with the pay-as-bid pricing rule. It was set from the beginning of the market and now it is hard to change even though many people at TERNA and many people in the Regulator's office prefer the marginal pricing rule. Another example is the UK NETA protocols which also deploy the pay-as-bid pricing rule.

The rationale of using the pay-as-bid pricing rule in the general case and the market conditions which may justify this scheme are given in Appendix B.

To the extent that the several bidders were able perfectly to predict the market-clearing price, in short, the savings from the change in the rules for consumers would prove to be zero. The *only* difference between the average prices actually realized under the two systems would, therefore, be the extent—and only the extent—to which their predictions proved to be mistaken.

Setting aside for later consideration the possibility that pay-as-bid pricing might be more effective in curbing exertions of monopoly power — if there is no reason to expect that prices will be consistently higher or lower under pay-as-bid, this scheme has the fopllowing effects:

1. **<u>Pay-as-bid introduces some inevitable reduction in efficiency</u>** as generators find themselves forced to depart from bidding their marginal costs if they are to receive any compensation for their fixed costs or contribution to profits. With all bids exceeding the marginal costs of all blocks of power, by amounts that depend upon the varying estimates of the several bidders of what will prove to be the marginal, market-clearing bid, the perfect, total cost-minimizing merit order dispatch will, inevitably, no longer be assured: some lower-marginal cost bids will be rejected — because their bidders have overestimated the market-clearing price — in favor of other, higher-marginal - cost power offered with more conservative markups. Since so very much is at stake in terms of the bidders recovering their total costs or any profit, and since the constantly changing demand and supply conditions that will determine the market clearing price are mainly unpredictable, their several bids will vary correspondingly in the markups above marginal cost that they incorporate. This observation holds even more true due to the fact that the ability of the several sellers to predict the market clearing price and their markups likely to differ substantially. **<u>With bids selected, then, on the basis not just of the marginal costs they reflect but also of these varying markups, the consequent inefficiencies stemming from departures from merit order dispatch of their plants are likely to be large</u>**.

   Inefficiencies will not be a consequence only of forecasting errors if bidders differ substantially and consistently in their relative marginal costs. Practice has shown that occasional inefficient outcomes are a consequence of rational strategic bidding.[20] For example, if there are two bidders with uncertain costs—uncertain in the sense that the individual sellers do not know the costs of the other—and one is known to have lower costs than the other on average, the bidder likely to have higher costs will rationally bid less aggressively, with a smaller markup over its operating costs than the bidder with lower costs; the latter will feel free to incorporate a larger markup in its bid, because it knows its rival is relatively unlikely to underbid it. The consequence will be that the disadvantaged bidder will be called on to supply too

[20] See, for example, Eric Maskin and John Riley. "Asymmetric Auctions." *Review of Economic Studies*, July 2000.

often, because it will have submitted a lower bid in some instances in which it has higher costs than its more efficient rival. **In any industry, competitive or otherwise, it is consumers who end up bearing the costs of such inefficiencies**.

2. **Another inefficiency inescapably introduced by the pay-as-bid pricing rule is the cost of forecasting market prices that it would impose on all participants**. Under the uniform, market-clearing price system, as we have experienced in international market over many years, sellers have every motivation to bid their marginal costs, which are of course readily available to them. The pay-as-bid scheme introduces uncertainties into their calculations and costs of attempting to minimize or dispel them by forecasting what the market-clearing price or prices would turn out to be. These costs, too, would ultimately be borne by consumers.

3. Finally, and in a sense worst of all, it is likely to discourage competition—to which consequence we now turn. Specifically, the pay-as-bid pricing scheme does not produce the right price signals for consumption. If we assume that bidders submit offers close to their costs, the pay-as-bid approach generates an average cost price signal for loads as opposed to the more desirable marginal cost price signal. **The net result is inefficient levels of consumption. As load becomes more elastic, the effects of this inefficiency are likely to increase**. Also, if bidders try and guess the marginal price, it is possible that many or all bids will be clustered around a single price based on the load forecast. This will increase the incidence of tied bids and may also weaken temporal price signals.

7.2.3 The Effect of the Pay-as-Bid Pricing Rule on the Exercise and Dissipation of Market Power

Most markets are not effectively competitive and at best they are imperfectly competitive. Then it becomes necessary to try to decide, first, whether uniform price or pay-as-bid pricing rule is likely to be more conducive to the exercise of such market power. Next we analyze the market power issue for both small and large players under both pricing rules.

7.2.3.1 Small Bidders are Disadvantaged Under the Pay-As-Bid Pricing Rule

Under the uniform price rule, competitors prosper or fail on the basis of their relative generating efficiencies alone; that is not only a consequence but also a prerequisite of an effectively competitive market. **Under pay-as-bid, their profitability depends heavily also on their successful forecasting.** From the standpoint of making markets more effectively competitive, even more troublesome than the effect of pay-as-bid in creating uncertainties and imposing the costs of forecasting would be the differential relative burdens of these uncertainties on small and large firms. Experience has shown that there are large economies of scale in the efforts to gather the requisite information and make such forecasts on a continuing hour-by-hour and day-by-day basis. The small firm would have to mount the same kind of effort as a large one; and if those efforts prove to be necessary (and the behavior of participants in such markets suggests a general belief that they are), the cost per unit of output would be much greater for the small than large competitors. Not only will the uncertainties introduced by pay-as-bid tend to discourage future investments, it will have an especially discouraging effect on investment by small firms. **In this respect, we can certainly claim that the pay-as-bid pricing rules will discourage the increased competition that is a critical part of the long-run of the economic efficiency of the energy market.**

One powerful impetus behind the pay-as bid scheme is the conception—which we will proceed to assess next—that the uniform pricing system is susceptible to gaming by large bidders. But under uniform price, smaller competitors likewise benefit from any such exertions of monopoly power: they too automatically receive any monopolistically elevated prices.

7.2.3.2 The Relative Susceptibility of Uniform Price and Pay-As-Bid Pricing Rules to Monopolistic Gaming

A substantial number of international studies have concluded that the price spikes in many markets were magnified by some large generators "gaming" the system by raising their offers or withholding physical capacity. For a generator to benefit from such a strategy several conditions must hold. First, demand must, in the aggregate, be inelastic. Second, the generator must control a mix of capacity such that withholding a portion of its capacity from the market will lever up the market clearing price received by its remaining capacity or other successfully bidding units sufficiently to more than compensate for the sacrificed net revenue on its withheld capacity. Note, this strategy does not require a high degree of industry-wide concentration. In these circumstances, it takes only a modest amount of withholding relative to the size of both the entire market and the total capacity of the game-playing generator.[21] The withholding can take two forms: (1) physical withholding—not bidding some fraction of one's operable capacity, or (2) economic withholding—bidding some fraction of one's operable capacity at a price markedly above its incremental cost. Both can have the same outcome: a higher clearing price.

This kind of behavior, in the surface may give credence to the argument to substitute the pay-as-bid pricing rule for the uniform pricing architecture. Under pay-as-bid, such generators would have to bid the estimated monopolistically elevated price on all their offered sales in order to reap those gains, at the immensely increased risk[22] that some or

[21] The fact that prices at such times may exceed the marginal operating costs of the least efficient generator in use—the usual indicator of monopolistic withholding of output—does not in itself prove that monopolistic withholding has occurred: when demand reaches the absolute physical limit of capacity—i.e., demand becomes totally unresponsive to price—the competitive price will rise to whatever level necessary to equate supply and demand: the more inelastic the demand, in these circumstances of an absolute limit to supply, the greater the margin by which the market price can exceed marginal generating cost.

[22] This risk is not great at peak times, with little aggregate excess capacity. It is much greater off-peak.

all of those higher bids will prove to have been excessive and therefore be rejected, with a consequent loss of the entire difference between their actual marginal costs and the ultimate market price. This change in the pricing method would, by this reasoning, therefore dramatically alter the balance of risks and potential gains of such exertions of market power.

Just as the naïve expectation that a shift to pay-as-bid will produce a substantial reduction in the average prices consumers pay ignores the certainty that generators will radically alter their bidding practices to frustrate achievement of that result. Thus, the argument that the pay-as-bid pricing rule would discourage monopolistic withholding by changing the balance of risks and potential benefits, fails to take into account the ways in which bidders will respond by changing their bidding behavior correspondingly. **This means that bidders woud, under the pay-as-bid pricing rule, attempt to predict the consequent behavior of the market prices in their several bids and, to the extent they succeed, the anticipated gains for consumers will prove to have been illusory.**

Also the claims that the pay-as-bid pricing rule will diminish the ability of the parties to collude tacitly to increase prices is illusory. A bidding process that is repeated daily is precisely the kind of game that lends itself to such collusion; changing the pricing rule would not alter that.

Another possibly important difference between the two pricing schemes is the greater transparency of bidding behavior under uniform pricing than pay-as-bid in detecting collusive or quasi-collusive pricing. Monopolistic behavior would be far more readily detectable under uniform pricing than pay-as-bid pricing. Since, as we have already emphasized, if the market were competitive all bidders would have every incentive to bid approximately their true marginal costs under the former system, and since costs—at least marginal operating costs—are easily measured to a first-order approximation in the electricity industry, the bid data would clearly provide evidence of imperfect competition and market power abuse.

Under the pay-as-bid pricing, in contrast, every seller would be forced to bid above its marginal cost, even if the market were perfectly competitive. Hence, there would be no direct way for the Regulator and market observers to identify from the bid data parties that appear to be exercising market power. Although it is difficult to assess the value of this additional transparency, it is another advantage of the uniform pricing rule.

7.2.4 Recommendation

The shift from uniform (Marginal) to pay-as-bid pricing — that it would provide purchasers of electric power substantial relief from the high prices of electric power — is simply mistaken. The immediate consequence of the introduction of the pay-as-bid pricing rule would be a radical change in bidding behavior that would:

- forestall anticipated consumer savings
- introduce unmeasurable inefficiencies in the dispatch of power and impose new costs on generating companies, which would inevitably tend to increase rather than decrease average prices
- tend to weaken the competition in generation that is the best safeguard against exertions of monopoly power; and
- impede — again to an unmeasurable extent — the expansion of capacity and future investments.

In summary we strongly believe that the uniform pricing rule is much superior compared to the pay-asbid pricing scheme. It offers the following advantages:

- Provides incentives for cost reflective bidding
- It is much more efficient
- It is consistent with optimal dispatch
- It is less prone to manipulation
- It provides marginal cost signals for demand side response
- It is better suited for use in double auctions that incorporate demand side bidding
- It reduces the advantages of large players over smaller ones, thus fostering competition
- It is less likely to result in tied bids.

7.3 DAM Market Architecture (Unified vs. Split DAM by Technology)

7.3.1 DAM Bifurcation Overview

One important design papameter which has emerged mainly in the EU energy markets resulting from the last year's EU energy crisis is whether energy auctions should be technology neutral (i.e., multiple technologies compete within the same auction) or whether they should be technology-specific (i.e., there is some degree of discrimination across technologies, either by type, location, and scale).

The massive penetration of RES plants in the grid has resulted in this debate among various policy makers if it is prudent to split the DAM by technology given the different operating and capital costs of these assets compared to the costs fossil fuel-based conventional power plants. Specifically, in a split DAM architecture there will a fossil fuel-based power auction based on the marginal pricing rule while a separate market will be created for RES energy where prices would be defined at the time of the auction. Note, that in the EU markets, the dynamics of fossil-fuel-based generation are wholly different from those of renewables, where the price for renewables is usually fixed at auction and connected to a Contract for Difference (CfD) design.

The justification is that the price differences between new renewable and fossil fuels can be very large with substantial price oscillations. The objective of the DAM bifurcation is to reduce the procurement costs. Specifically, the capital expenditure of RES dominates their cost structure and there are no changes in unit cost when increasing or decreasing their operation. Resources with such features include in addtion to RES, nuclear, high-efficiency co-generation, mandatory hydro and electricity storage bundled with intermittent renewables (hybrids).

7.3.2 DAM Bifurcation Design Proposal

Proponents of the DAM bifurcation design propose the following principles for the design modifications:

- The RES plants should submit volume-based offers in the day-ahead market (DAM), not economic bids. The volume-based offers reflect the best possible forecasts of their operation on the next day. With this offer they assume responsibility for the real- time operation, are subject to deviation costs and can participate in the intra-day and balancing markets.
- For their volume-based offers in the DAM, these resources get remuneration depending on contracts for differences concluded with private third parties or the public sector, regardless of the DAM.
- In case these resources declare no coverage by bilateral or public contracts for differences, they may participate in a non-mandatory pool (green power pool) operated by a public body (or a private body adequately empowered) acting as a single buyer and seller to load-serving entities and consumers.
- The volume-based offers of these resources may correspond to bundled resources that may include storage and possibly an aggregation of RES plants.
- The ISO scrutinizes the volume-based offers from the perspective of forecasting accuracy and system operation possibilities and may accept or curtail the volumes declared. The eventual curtailment follows pro-rata rules.
- In the next step, the DAM considers that the accepted volumes of the above resources that operate have a MUST RUN status. Thus, the ISO subtracts the accepted volumes from the load declarations. The remaining load (net load) corresponds to the demand that the on-demand resources dispatchable resources must meet. Then, the resources submit combined economic and volume offers according to the same rules currently applied and the market is cleared with the same way it is cleared today.
- The load-serving entities and consumers pay at market-clearing prices for the energy purchased in the net-load DAM. They may also buy from the green power pool, if this operates. They also have payment obligations in the context of CfDs which are independently concluded.

- The above points describe a two-stage bifurcated DAM.
- The intra-day and balancing markets remain unchanged.
- Consumers pay a weighted average of the prices of the two DAM markets. Thus, if the first stage of the DAM corresponds roughly to two-thirds of electricity consumption and for example has an average cost of 80 US$/MWh and the second stage of the DAM clears at 250 US$/MWh reflecting gas generation costs, the consumer would pay (2/3 x 80)+(1/3x250)=137 US$/MWh.

7.3.3 DAM Bifurcation Design Analysis

Under the technology-neutral approach, the final technology mix is decided through the power auctions based on the technologies' current costs as reflected by the offers/bids. Under a technology-specific approach, the regulator must decide how much to procure from each technology. Thus, the former might be subject to market failures, while the latter might be subject to regulatory failures. Proponents of the DAM bifurcation approach claim that technology neutrality which effectively minimizes current costs, it may result in over-compensation for low-cost technologies, such as RES, unnecessarily increasing procurement costs. Further they argue, that technology-neutrality favors technologies whose costs are currently low at the expense of less-mature technologies whose costs are higher. Obviously, this argument is not true or soon it will not be true since RES energy is currently achieving grid parity in many jurisidictions.

The asymmetric costs of various technologies may result in high rents of low-cost assets in technology-neutral auctions if the high-cost assets set the auction price. This could be avoided under technology-specific auctions that pay each technology at its market-clearing price that mitigate the existing asymmetries among the projects and, thus, the resulting rents.

This argument however, is contrary to the well established principle of designing and operating wholesale energy markets over the last twenty years where the infra-marginal rents in technology agnostic power auctions is critical for providing a clear and robust market price signal for investments in future capacity in the long-term and optimal dispatch operations in the short-term. In addition, the DAM bifurcation option separates the DAM in two segments, decreasing the liquidity of DAM and producing a marginal price of the DAM that does not express the short-term marginal cost of the energy system.

Further, the new "distorted DAM price" shall also create inconsistencies in the Forward Markets, since the generating units shall sell forward contracts considering their expected revenues (irrelevant with the DAM price) and the Load Representatives shall buy forward contracts considering the weighted average cost of the generating units and the DAM price. These prices shall be different, distorting the Forward Market contracts. Thus, a fundamental issue arise with regard to the underlying asset of Forward Contracts.

Finally, the DAM bifurcation approach, relies on regulatory decisions, instead of market outcomes. Specifically, the regulator is required to have precise information about the profitability of the various technologies, in order to run technology-specific auctions to reduce the rents of low-cost technologies without distorting the allocation across technologies. **In other words, the DAM bifurcation approach attempts to substitute the relative advantage of technology-neutral auctions, which is to select low-cost investments, with the regulator which should have enough information to replicate the same outcome.**

As RES technologies mature and costs are reduced and are similar across them (both in the value they provide as well as in their costs) the market solution is much preferable and therefore the DAM bifurcation apporach is deemed unnecessary and counterproduction. In this case, having the regulator decide how much to procure from each technology can be very challenging and ultimately costly.

7.3.4 Recommendation

Based on the above analysis we clearly recommend to reject the DAM bifurcation approach for the following reasons:

- It will create substantial regulatory uncertainty
- It wil lead to inefficient market outcomes
- It reduces the liquidity of the DAM
- The "distorted DAM price" creates problems as an underlying asset for forward contracts
- It will not reduce the over-compensation problem
- It will put an undue burden on the Regulator which in essence should substitute the energy market
- It will destroy the infra-marginal rents property which is so critical for providing the proper market signals for investments and the well functioning of the energy markets

7.4 Long Term Contract PPAs (Physical and Financial) with Simple CfDs vs CfD Variants

7.4.1 Overview of the Bilateral Contracting Market

There is overwhelming theoretical and practical evidence regarding the competiveness of wholesale power markets and the impact of forward contracting and vertical arrangements in mitigating market power and enhancing competitive outcomes. **The**

theoretical work shows that Bilateral Contracts (BCs), in general, reduce the incentives for market power abuse and the adverse impact of such abuse, by reducing the profitability of spot price manipulation and the impact of spot price spikes on transfers between customers and suppliers.

Empirical evidence supports the theoretical predictions that bilateral contracting reduces market power. Further the presence of financial hedge contracts mitigates market power as well. There is substantial evidence that firms that do not divest generation and remained vertically integrated in generating power and serving retail load behave more competitively in wholesale power markets relative to generation owners that serve little or no retail load.

Further, multiple studies that deployed data from the PJM, ISO New England, and the original California ISO markets and simulated the outcomes with and without vertical bilateral arrangements or retail contracting of wholesale power output have shown that the PJM and ISO New England markets were much more competitive than the California market, and this difference could be explained by retail contracting of wholesale power output in PJM and ISO New England which was not permitted in the original California ISO restructured market model. In fact, the studies showed that if PJM and ISO New England had not permitted these vertical bilateral contracting arrangements as had been the case in California, those markets would have seen substantial price increases. The overall production costs were estimated to increase 45% in the absence of such bilateral contracting arrangements.

Our analysis below gives credence to the claim the presence of a Forward BC market is very beneficial to the market. It is our experience that such a market is desirable by Market Participants, and if available, they'll make use of it to optimize their portfolio. If Market Participants, for some reason, are not willing to transact in such a market, it is our recommendation that the Regulator should force such transactions between producers and suppliers in imature markets. The introduction of such a market should be done gradually and methodically.

At the same time it is very important to ensure that the Forward BC market does not dry up the "Pool" market which creates important price signals. In mature market these market find a natural equilibrium where the volumes that clear in each market are stabilized.

7.4.2 Financial Bilateral Contracting Benefits Analysis

The lack of significant forward contracting in a market increases the incentives for generation unit owners to exercise market power in the spot energy markets. To substantiate this claim, consider the following example of a firm with some ability to affect the market-clearing price in the spot electricity market.

Let QS denote the amount of energy it produces, PS the spot price of energy, and MC is marginal cost of producing electricity. Suppose this firm has previously sold a two-sided contract-for-differences (CFDs) at a price PC. Let QC denote the quantity of CFDs sold. The payoff to the seller of a two-sided CFD is (PS – PC)*QC. If PS is greater than PC, the seller pays to the buyer the difference between PS and PC times QC. If PC is greater than PS, the buyer pays to the seller the difference between PC and PS times QC. For simplicity, assume that MC is the same value for all output levels. The variable profit earned by the firm is then:

$$\text{Variable\_Profit}(PS) = (PS - MC)*(QS - QC) + (PS - PC)*QC \qquad (7\text{-}1)$$

The first point to note from this variable profit function is that until the generator covers its forward financial contract position, QC, with physical sales, QS, it will use its ability to influence the market price, PS, to set it lower than its marginal cost, MC. This incentive operates because when QS is less than QC, the only way for the generator to make a positive contribution to variable profits in Equation (1) is if PS less than MC. A second point to note is that if QS is greater than QC, and QC is non-zero, then the generator does have an incentive to use its market-power to raise prices. However, the presence of forward financial contract dulls this incentive to raise the spot price, PS, because the generator only earns this price for its spot market sales beyond QC. **Consequently, to the extent that QS is greater than QC, the firm has less of an incentive to raise prices through its bidding behavior. Consider the case that QC is equal to zero. Here the marginal incentive of the firm to raise PS by exercising its market power is greatest, because it earns this higher price for all of its spot sales, QS.**

This example illustrates a very important point associated with assessing the benefits to load-serving entities of forward market purchases. Specifically, the forward market commitments made by or imposed upon a generation unit owner significantly alters its incentives to raise prices or withhold capacity from the spot market. **However, generation unit owners understand this mechanism, and are reluctant to commit to forward financial contracts at prices that do not yield the same expected profit stream as they could obtain from their forecast spot market sales.** Consequently, to make a forward market sale attractive to a generation unit owner, the load-serving entity may have to offer a higher forward market price. However, once this contract has been signed, this generation unit owner will now bid more aggressively in the spot market, with the result being lower spot prices. Deeming these forward contracts imprudent after fact because of the lower spot price would be inappropriate. The reason is that these lower spot prices would not have occurred if the forward contracts were not in place. **This aspect of forward contracting and impact on generation unit behavior in spot markets considerably complicates any assessment of the prudence after the fact, of any forward market purchases.**

An additional benefit of forward financial contracts is the protection from spot price

fluctuations they provide to load-serving entities. A load-serving entity that holds the other side of the two-sided CFD in the above example is completely hedged against spot price risk if its consumption is equal to QC. To the extent its consumption differs from QC, it bears spot price risk only on the deviations of its actual consumption from QC. A load-serving entity holding a significant fraction of its expected sales in CFDs, has effective price certainty on its wholesale energy obligations and can therefore set a fixed retail price and be reasonably assured of covering its costs regardless of what happens to spot electricity prices. If all load-serving entities held forward contracts for all of their energy obligations to small business and residential consumers, the regulators would know that these entities have wholesale price certainty for these customers. This wholesale energy cost certainty would allow the regulators to set fixed default retail rate for these two customer classes. The regulators could also allow other retail pricing plans where these customers voluntarily take on wholesale price risk in exchange for the opportunity to receive lower average electricity prices (because they alter their demand in response to wholesale price changes) than under the fixed retail price.

A final benefit of forward financial contracting is that it effectively renders less important any discussion of the relative advantages of pay-as-bid versus uniform price auction mechanisms for electricity spot market designs. In a competitive electricity market, regardless of whether the spot market is cleared using a pay-as-bid or uniform price auction, electricity that is produced and delivered within a given hour is being paid according to wide variety of forward market contract prices.

For example, one would expect that the owner of a low-variable-cost, high-fixed-cost unit would prefer to operate it as a base-load facility. Consequently, the owner of this unit would be willing to sign a multi-year two-sided CFD at a price close to the expected average annual price of electricity because it expects to be operating this unit in virtually all hours of the year at a constant rate of output. The owner of a peaking unit that expected to only operate during 100 to 200 peak hours of the year, would probably engage in a different set of forward sales. This unit owner might instead sell a one-sided CFD for a significant fraction of the unit's expected output. Under a one-sided CFD, in exchange for an up-front payment from the buyer, the unit owner pays out the maximum of zero and the difference between PS and PC times the number of units of the contract sold, QC, to the buyer of the CFD, where PS is the spot price and PC the contract strike price. This CFD provides the purchaser with insurance against price spikes in the spot market for QC units of output, but does not require the purchaser to make any payments to generation unit owner if PS is less than PC, only an up-front payment at the time the contract is signed. This up-front payment should help the unit owner cover the annual fixed costs associated with running its unit. For the hours covered by this one-sided CFD, the unit owner can earn a maximum price of PC by selling QC units in the spot market.

The unit owner could cover the remainder of its annual revenue needs through sales in the spot market. Finally, a unit owner that primarily serves intermediate load levels, may choose a combination of two-sided and one-sided CFDs, as well as some sales into the spot market. Each of these contracts would be negotiated with individual buyers, so that the unit owner would have a portfolio of forward market positions at a variety of prices.

In competitive electricity markets with active forward markets, in any given delivery hour all market participants—loads and generation unit owners—have a portfolio energy purchases and sales at a variety of prices. Consequently, regardless of whether the spot market is cleared using a uniform-price auction or pay-as-bid auction, electricity is delivered during a given hour according to a large number of forward prices negotiated at a number of different times in the past and under a variety of contract forms.

It is important to emphasize the reason that energy is delivered under a variety of prices under either spot market price-setting process. Forward contracts are negotiated under different terms and conditions at different times before delivery takes place. Presumably, these prices reflect the best information at the time the contract is negotiated of its value to the buyer and seller. New information about the market-clearing price of electricity at a given time in the future continually arrives and it is processed by buyers and sellers of forward electricity contracts. **The continual arrival, over time, of new information about spot market conditions at the delivery or clearing date of a forward contract is the major reason for the large number of prices for electricity delivered in the same hour**. We would expect that a forward financial contract for delivery of 1 MWh energy in a given hour in the future could not consistently sell for a higher price through a bilateral negotiation pay-as-bid market, versus a uniform price auction market. Otherwise, the buyer of this contract would instead purchase from the uniform price auction. Conversely, if the price was lower under a uniform price auction, we would expect the seller to move to the pay-as-bid bilateral negotiation market.

For these reasons, as well as many others, virtually all market designers strongly agree that robust forward financial markets play an important role in the market. However, there are currently various impediments, regulatory or otherwise, to the development of this market in various places. For example in some markets the major impediment is the fact, that energy and ancillary services prices in the ISO markets reflect the exercise of significant market power. Consequently, any forward contract price that a generation unit owner would voluntarily offer to a load-serving entity would reflect this market power. This calls for the implementation of specific mitigation market rules to ensure that the potential for exercise of market power is reduced or eliminated.

7.4.3 Long-Term Bilateral Contracting and Low Carbon Generation

7.4.3.1 Overview

We now analyze how corporate BCs between RES generation owners and major loads, such as intense industries or even smaller consumers, can facilitate market evolution, reduce costs and improve market efficiency. We discuss the challenges for executing such contracts and propose ways to improve the liquidity and reduce market uncertainty. Finally we'll propose how the sliding premium method (curently the dominant remuneration mechansims for RES generation) could be structured as a contract for differences (CFDs) to facilitate corporate PPAs. We'll show that CFDs provide incentive for a wide range of actors to invest in renewable energy, enabling them to participate in the energy transition and ensure the lowest total costs to consumers.

7.4.3.2 Analysis of a Private RES-PPA Wholesale Market

RES plants have become significantly more affordable in recent years. As we discussed in Section 7.3 RES plants have low variable costs but relatively high capital costs. This is why the costs of financing the investment are a key part of their overall cost. In turn, financing costs depend on the level of certainty involved in the future revenues from electricity generation. Uncertain revenues increase the cost of financing significantly compared to a market designs with renewable energy remuneration mechanisms that reduce this uncertainty by avoiding policy risk and serving as hedge for market price uncertainties.

If the revenue from electricity production is ensured at the time of the investment decision, investors can use this certain revenue stream to raise low cost capital from bonds or debt to finance their investment. On the contrary, to the extent that revenues are uncertain at the time of the investment decision, investors require higher shares of risk-taking equity capital. This increases the overall cost of financing their investment.

A wholesale market for trading private Bilateral Contracts (physical or financial) of various characteristics of duration, balancing conditions, the strike price and various profiles between a retailer or a trader, and a RES company can create the conditions for containing the cost of finacing the RES investment In this case, the RES developers will use the BCs as collateral for raising funding from banks or investors.

The bilateral contract can be financial (like a CFD), or a contract with physical delivery between an offtakers (a major load for example) and a RES generator. These contracts should not enjoy any public support subsidies and should specify various parameters, such as duration, clauses, balancing conditions, and the exact definition of the contract's strike price or strike price collars or conditions.

Based on our experience there are several barriers for developing such a market.

These include: a) funding uncertainty of RES projects, b) hesitation on the part of the off-takers to sign very long-term contracts (mismatch of preferences of generators and loads), c) lack of RES-PPA market liquidity. Market developments, such as storage investments, reduction of RES procurement costs, etc., can relieve some of these problems but a key issue remains: the availability of balancing and complementary energy from the wholesale and balancing markets with enough liquidity and hedging instruments to facilitate retail market competition. A key objective of this organized market is to provide the tools to manage the mismatch between RES volatility and the load profiles and mitigate uncertainties through economies to scale.

For such organized market to develop, a critical mass of bilateral contracts need to be concluded to establish certain liquidity levels and enable favorable financing conditions for RES investments. Typical bilateral contracts with CFDs with physical delivery optionality can be traded in this voluntary market. **The Government has a key role to play intially in establishing such a market by a) becoming in a sense the "Market Maker", b) ensuring efficient funding of RES projects, and c) subsidizing a portion of the costs incurred by the small off-takers.**

The most prevalent RES remuneration mechanism in many ISOs is based on **the sliding premium**. The aim is to drive RES market integration and at the same time, safeguard low financing costs for investors. Prior to the investment, RES developers submit their strike price in an auction. The projects that submit the lowest strike prices win the auction and thus a guarantee for this price for 20 years. During operation, RES operators then sell their electricity production in the market. On a regular basis (say monthly basis), the average revenue of all plants of a technology, e.g. PV, is calculated – the so called "market value of the electricity". If the market value is below the strike price specified by the RES developer, the project receives the difference between the strike price and the market value – as so called "sliding premium".

With continued declines of the cost of RES technologies as well as more likelihood attributed to scenarios of increasing electricity prices, triggered by increasing gas prices and carbon prices, it is increasingly likely that the market value will exceed the "strike price" from the auctions. As a result, the price of electricity above and beyond this strike price will contribute an increasing share to possible revenues. At the same time, the share of revenue that is secured with the strike price will fall. Since electricity market revenue above the minimum price is uncertain, this implies that an increasing share of equity is required for the investments. **This means that financing costs will rise and electricity consumers will no longer be able to benefit from falling RES technology costs to the fullest extent.**

Under the evolving market circumstances, we propose two options for the design of the RES remuneration mechanism to consider for the market of RES-PPAs:

1. Sliding premium
2. Developing them further as PPAs with CfDs

We briefly analyze both options below.

Option 1: Sliding premium-rising financing costs counter cost reductions

The current premium scheme will reinforce the process by which uncertain electricity market revenue is increasingly and is included in bid calculations, ultimately resulting in rising financing costs.

In the case of lower realized market values, the strike price will effectively be the minimum price, because electricity market values are topped up with the sliding premium. Additional revenue from electricity market sales materialize if the realized market value is above the strike price. But unlike the strike price, this additional revenue is uncertain and thus cannot be used for financing with low-cost debt. For this reason, more equity is required for financing. Hence, higher financing costs off-set part of the lower technology costs of renewable energies, and consumers do not fully benefit from the cost savings.

Option 2: Contracts for difference - hedging for investors and protection for consumers

With only a few adjustments, the sliding premium could be structured as a contract for difference (a CFD). Instead of only providing investors with a hedge against low market values, investors and the public could sign CFDs. The United Kingdom is already using this remuneration mechanism. CFDs are long-term PPAs at strike prices determined in auctions. As with the sliding premium, RES operators receive additional revenues when the market value of the electricity they generate is below the agreed strike price. However, if the electricity price that RES plant operators obtain is above the strike price, the operators must pay the difference back to the ISO. This lowers the renewable energy levy and can even render it negative in the long run. This can increase acceptance of the energy transition, as electricity consumers, who have safeguarded renewable energy operators for many years against low electricity prices, are protected against high electricity prices in the future to the extent of existing CFDs.

CFDs ensure that RES project developers do not expect electricity market revenues in addition to the strike price. As a result, uncertain revenue no longer needs to be included in the financing calculation, which would result financing costs to rise.

The CfD option implies that when technology costs fall, the strike price under the sliding premium falls more strongly than the strike price of CFDs. But in the sliding premium case, electricity consumers have additional costs for electricity consumption in case of high prices or when the market value of renewable energy is high. Falling technology costs are partially offset by rising financing costs in that case, and electricity customers will therefore benefit from only part of the technology cost reduction.

7.4.4 Recommendation

Comparison of policy instruments shows that electricity consumers via private RES-PPAs only fully benefit from falling technology costs in the case of CFDs. Only CFDs provide incentive for a wide range of actors to invest in renewable energy, enabling them to participate in the energy transition. **Uncertain electricity market revenue must be taken into consideration when refinancing investment when the premium is sliding or fixed or in the absence of renewables remuneration mechanisms. This requires higher shares of equity for financing.** This can be a detriment to actor variety, as small actors tend not to have large balance sheets of large energy suppliers that are in a better position to dedicate larger shares of equity to a RES project. Reduced participation may result in a decline of support from local groups in the energy transition.

Reduced participation also reduces the competitiveness of the RES auctions. Hence, a remuneration system that allows all actors (large and small) to finance investments with low equity requirements is of particular relevance.

In the absence of CFDs, the investors with the most optimistic market expectations are increasingly winning the RES energy auctions. **The "winner's curse" could cause RES project realization rates to drop. Significant declines in expected longer-term power prices between the time of the RES and final investment choice (closure) may trigger investors to revise their previous assumptions and abandon a RES project and pay the relevant penalties instead.** This scenario would endanger Chile's expansion targets for renewable energy.

Changing electricity price or market value trends in the period between the auction and plant construction have no influence on the RES operator revenue in a CFD system. This leads to higher realization rates. CFDs also lead to symmetrical hedging of investors and electricity consumers. Investors obtain a hedge against low market values for renewable energy, and electricity consumers obtain a hedge against increasing wholesale power prices.

In summary we strongly encourage long term PPAs with CfDs to foster the energy transition. CFDs ensure that additional revenues are returned to consumers. Further, investors have a revenue hedge under the CFD system, which enables favorable financing costs and in turn, a lower total cost of renewable electricity.

7.5 Financial Market vs Physical Markets

7.5.1 Overview

International experience in restructured electricity markets in the US and world wide has shown that reliance on spot market to set the price for a high portion of electricity consumption is problematic. Long term forward contracting between suppliers and demand is necessary for three major reasons:

1. Long term contracts incentivize and support investment in generation capacity to achieve resource adequacy and reliability, by reducing uncertainty for suppliers and facilitating capital for such investments
2. Long-term contracts provides hedging instruments for fuel supplies, adverse weather and systems contingencies and ensures electricity supply at reasonable prices
3. Long-term forward contracting reduces the incentives for price manipulation and exercise of market power **exploiting market concentration** and short term uncertainties on the demand and supply side. **While spot prices will continue to reflect scarcity conditions and provide incentives to producers and consumers to respond to these conditions, long term contracts minimize the transfers from consumer to producers affected by spot price excursions and hence reduce the potential gains by suppliers from price manipulation**.

In California, for instance, during the energy crisis of 2000-2001, divestiture of generation resources without proper vesting contracts, combined with market design flaws, adverse weather conditions and lack of incentives for load serving entities to insure their supply through long term contracts, resulted in catastrophic price hikes due to market power abuse, market collapse, and financial failure of major load serving entities. In Colombia during the 2015-2016 El-Ninio, flawed design and inadequate enforcement mechanisms of the Firm Energy Contracts that were supposed to insure resource adequacy, resulted in near market collapse, default of a major generation company (900MW Thermocandelaria) and “Bail Out” at customer’s expense.

Given the lessons from international markets it is clear that **long term forward contracts should be an integral part of the market design**, however, it is important to differentiate between the distinct objectives of resource adequacy vs. price hedging as we do below. A resource adequacy mechanism is intended to insure sufficient investment in generation resources and incentivize the desired mix of resources that will account for extreme weather driven events, provide adequate reserves guided by technical considerations and support the environmental objectives of decarbonization.

Resource adequacy mechanisms should not be designed to mitigate market power or mitigate high energy price, which should be addresses separately with financial energy price hedging contracts discussed below. Consequently, the price caps serving as strike price should be sufficiently high (typically $1000-$2000 per MWH). Capacity or Firm Energy obligations with low strike prices, such as the Reliability Options implemented ISO-NE and the Colombia Firm Energy Options, failed because the **strike price was not high enough to absorb spikes in fuel cost**. At times of scarcity, generators called to produce energy under the resource adequacy mechanisms at the strike price opted for default because the fuel cost exceeded the strike price (as was the case with Thermocandelaria in Colombia). ISO-NE attempted to fix the problem by indexing the strike price to fuel cost and account for fuel switching options but that fix failed and the strike price was eventually reset to the system price cap.

7.5.2 Long Term Forward Energy Contracts: Physical vs. Financial

Long term forward contracts are agreements between two entities to sell and buy energy over a specific time duration at a fixed agreed upon price. Such contracts typically specify fixed quantities at a fixed price, however the specified quantities may be in the form of a monthly, weekly or daily totals or allocated over hours. Contracts may also specify fixed delivery rate or allow rate variations within limits. Typically Load Serving Entities serving retail customers bear the volumetric risk due to load fluctuations and adjust the quantities they procured through fixed quantity contracts through spot market transactions. However, some “Load Following Contracts” may commit to supplying a percentage of the LSEs load, in which case the quantity risk is borne by the supplier.

Long term forward contracts which are typically referred as Power Purchase Agreements (PPA) can be either physical or financial. Physical PPAs are exercised in the day ahead market as self-scheduled transactions submitted to the ISO and settled directly among the parties to the contract. On the other hand, financial PPA, are settled among the parties as **Contracts for Differences or CfDs**, independently of the physical transactions scheduled by the parties to the contract and settled with the ISO. In other words, parties to a financial PPA buy and sell energy through the Day Ahead Market (DAM) operated by the ISO and settle their transaction at the ISO Day Ahed prices. Subsequently, they settle directly among themselves the differences between the DAM prices and the agreed upon Contract Price. If the DAM price in a particular hour exceeded the contract price, then the seller pays to the buyer the price difference and vice versa so the net effect is equivalent to transacting the contract quantity at the contract price. **The financial settlement among the parties to the contract is independent of the physical schedules so if the seller or the buyer are over or under covered by the contract compared to their physical schedule, the quantity differences are subject to the DAM prices. Consequently, a financial entity that does not represent demand or generation capacity could enter into a Financial PPA with an energy resource or LSE participating in the ISO market and provide financial hedging without being involved in the physical market.**

Financial PPAs which enable participation by financial entities should be closely regulated to prevent market manipulation by speculators. **However, decoupling the physical market from the financial settlement is beneficial for several reasons**. Such decoupling provides flexibility to both supplier and demand to buy and sell energy efficiently without being bound by physical obligations. For instance on the demand side, a manufacturer can reduce consumption when spot prices are high due to scarcity, and earn the difference between the spot price and contract price for the unused portion of the contracted energy. On the supply side, with the increase in the supply of renewable energy, a gas generator may choose to reduce output when renewable energy is abundant and spot prices are low, and rely on the renewable resources to provide unmet contracted energy supply.

Allowing financial participants to participate in the contract market, which we strongly recommend, increases market liquidity and improves hedging efficiency by squeezing out arbitrage opportunities. Such financial participants also facilitate capital availability for investment in energy resources. They also enable **cross commodity hedging** by participating in other markets that are correlated with or impact electricity prices. For example **Spark Spread forwards and options** are instruments that track the spread between the price of electricity and natural gas which captures the profitability of gas generators and hence provides a hedge for investment in such a resource. **Weather Derivatives** is another example of financial instruments traded outside the arena of electricity contracts, which are used to hedge against the risk of weather related losses, enabling cross hedging between agriculture, tourism, and energy. New investments in electrolizers for Green Hydrogen production and facilities for Green Ammonia can be greatly facilitated by cross commodity hedging between markets for Hydrogen related products, Fertilizer industries and electricity. Such hedging practices are uncommon in contract markets that are restricted to electricity producers and consumers.

7.5.3 Standardized Forward Energy Contracts

Financial Long Term Forward Energy contracts are typically traded over the counter among producers and consumers with possible participation of financial entities. **We recommend the adoption of such contracts to supplement a resource adequacy mechanism** to provide hedging and stability of energy prices. Furthermore, given the important role of such contracts in assuring supply to consumers of energy at reasonable prices and hedging against the risk of price spikes, **we recommend the establishment of Standardized Forward Energy Contracts (SFEC)** traded through a central clearing house organized by the regulator, system operator or a private exchange that is in a position to manage counterparty risk efficiently. Such contracts will specify energy prices, source and sink, as well as compliance period but exclude coverage of congestion risk which can be handled through FTRs, addressed separatly. **We also recommend that trading of such contracts be open to financial player participation**, subject to regulatory oversight. Under this proposed approach, hedging by load serving entities that are subject to retail price regulation should be incentivized and possibly mandated through hedging requirements on a percentage (e.g., 90%) of forecasted load to be covered by contracts. **Regulatory practices of allowing regulated retailers to pass**

through wholesale prices to their customer while holding them responsible for adverse outcomes of hedging contracts should be avoided since they create disincentives for long-term contracting and encourage retailers to under forecast their load or submit low bids that may not clear the SFEC market. On the contrary, regulated retailers should be encouraged and incentivized to manage spot market risk on behalf of their customers and rewarded for successful risk management practices. Unregulated retail customers are free to manage their own risk and decide how much contract cover they procure and should be allowed to diversify their risk through multiple counterparties.

On the supply side, new plants typically require fixed-price or indexed long-term forward contracts for a significant share of the expected output of the plant in order to obtain the necessary up-front financing to construct the plant. Existing plants sign fixed-price or indexed forward contracts to reduce their exposure to short-term prices and reduce the volatility of their financial results. These contracts supplement the revenues from resource adequacy mechanisms such as capacity payments or reliability charges for firm energy contracts. Generation unit owners and retailers may also participate in bilateral over-the-counter (OTC) markets to hedge their medium-term price risk. **We recommend, however, that suppliers be required to offer a mandated fraction (e.g., 90%) of their installed capacity or firm energy obligated by the resource adequacy mechanism as SFECs** with possible waiver for verified long-term bilateral contracts. Such contracting requirements imposed on regulated buyers and on sellers of energy will mitigate the exercise of market power by generators and control self-dealing in vertically integrated companies. On the other hand, it is important to allow sufficient headroom for market participant exposure to the spot market to facilitate market efficiency through incentives for supply and demand response to spot prices. We note that the quantities obligated through the resource adequacy mechanisms are guided and backed by physical generation capacity, firm energy and forecasted demand. However, if we allow financial participation, generators can, like any financial entity to sell SFECs in excess of their physical capacity and demand representatives can buy SFECs in excess of their forecasted demand in accordance to their hedging strategies.

An advantage of a centralized contract market trading SFECs as compared to OTC trading of bilateral contracts is data transparency that allows estimation of a forward curve that facilitates risk management. A centrally cleared market also enables anonymity of sellers and buyers, which **decouples energy price risk from credit risk of the participants**, thus providing small participant indiscriminate access to cheap resources. The credit risk can be handled through collateral requirements and strong enforcement rules which may involve **monitoring and verification of fuel supply risk management practices** by the supplier and verification of load forecasts. One of the problems with the default of Thermocandelaria during the Colombian 2015-2016 crisis was the lack of contract enforcement which resulted in consumers bearing the excess fuel cost.

7.5.4 Mechanics of Centralized SFEC Procurement and Settlement

The procurement of SFEC can be done through a staggered annual uniform clearing price auction for monthly contracts with a 3-5 years forward compliance. Quarterly two-sided uniform price auctions will enable buyers and sellers to adjust their positions, accounting for updated load forecasts, resource availability, transmission investments, and fuel markets volatility. Supply offers and demand bids for SFEC should be segregated by location into supply nodes and load aggregation points (LAP) and the auctions conducted simultaneously for all locations. Offers by nodal resources will be adjusted up or down by a locational premium reflecting long run average locational marginal prices, and so will the contract price paid to the resource. Likewise load bids will be adjusted by the long run average LAP premium and so will the contract price paid by the load. Contract settlements between the central underwriter and the contract holder are based on hourly LMP or LAP price realization, where the contract holder and the central underwriter settle the difference between the contract price and the locational hourly price as a contract for differences (CfDs). Hence the difference between the long term LMP or LAP price average used in the contract auction and the corresponding hourly prices are borne by the contract holder as basis risk.

7.5.5 Recommendations

1. Implement a physical resource adequacy mechanism with a capacity or firm energy must offer obligation with a high price cap. A mixed approach with a capacity obligation for thermal resources and firm energy obligation for hydro is also possible. This mechanism is restricted to physical participants or investors in future resources and load and is subject to verification.
2. Introduce a standardized financial forward energy market with a central underwriter which is open to financial participants with regulatory oversight.
3. Impose a forward energy contract procurement requirement on load serving entities serving regulated load and on resources with capacity or energy obligated under the resource adequacy mechanism.
4. Eliminate spot price pass through provisions for entities serving regulated load, exposing them to spot price risk, and institute incentives for managing price risk of regulated load.
5. Institute credit worthiness and contract enforcement for the standardized forward energy contracts with possible verification of fuel supply hedging.

7.6 Energy Only Markets with Scarcity Provisions vs Resource Adequacy Mechanisms (or Both)

7.6.1 Missing Money Analysis Overview

The competitiveness of energy markets depends on the diversity of generating system's cost structure and essentially on the diversity of generators' marginal costs. Only at scarcity supply times fixed costs matter and may drive market prices way above marginal costs. Under normal market conditions, marginal costs prevail in the large majority of market instances. Then, revenues above marginal costs depend exclusively on the diversity of marginal costs. In other words, the revenues of a given generator depend on whether competitors with higher marginal costs are included in the energy market clearing.

Markets dominated with renewable energy with zero margincal costs and gas power plants exhibit a market price duration curve which is flat most of the time except rare cconditions of scarsity. These markets cannot guarantee capital and fixed cost recovery for the dispatchable gas plants. Revenues above marginal costs are highly uncertain as occurring only during scarcity times due to the low diversity of marginal costs among the remaining operating units. Scarcity pricing can help, but based on our experience, investors are reluctant to rely on uncertain revenue streams which may occur sporadically and at uncertain time intervals only when the system reaches scarcity conditions. Under these conditions energy markets exhibit, in various degrees, "**Missing Money**" problems.

The term "Missing Money" refer to a situation where infra-marginal energy revenues in excess of operation and maintenance cost are systematically lower than the amortized cost of new entry for a marginal generator. Extensive analysis of the performance of organized energy-only wholesale markets indicate that they do not appear to produce enough net revenues to support investment in new generating capacity in the right places and consistent with the administrative reliability criteria that are applicable in each region. In some cases, even adding capacity revenues, the total net revenues that would have been earned by a new plant would have been less than the fixed costs that investors would need to recover in order to make investment in new generating capacity profitable. **Every organized market we have studied exhibits a similar gap between net revenues produced by energy markets and the fixed costs of investing in new capacity measured over several years time.**

What are the causes of these shortfalls? We focus on those that depress energy market revenues. The following provides an illustrative, but not exhaustive, list:

- Price caps on energy markets
- Market power mitigation mechanisms that do not allow prices to rise high enough during conditions when generating capacity is fully utilized
- Actions taken by ISOs that have the effect of keeping prices from rising fast enough and high enough to reflect the value of lost load
- Reliability actions taken by ISOs that rely on Out of Market (OOM) calls on generators that pay some generators premium prices to solve certain problems but depress the market prices paid to other generators
- Payments made by ISOs to keep inefficient generators in service due to

transmission and related constraints rather than allowing them to be retired or be mothballed

- Regulated generators operating within a competitive market (e.g., “MUST RUN” generators that have poor incentives to make efficient retirement decisions, depressing market prices for energy.

All the actions taken together depress energy market revenues and create the “missing money” problem. Missing money is one of many factor that could hinder investment in new capacity when needed but not the only one. Uncertainties due to future environmental regulation, transmission expansion, renewables penetration, demand response policies, and interventions by the ISO, Regulator and market monitors can create a risk premium that will indirectly raise the revenue requirement for new entry and hence the level of missing money, if any.

7.6.2 Energy Only Markets vs Resource Adequacy Approaches

Under an Energy Only mechanism generators are only remunerated for energy sales (with additional payments for ancillary service provision). In the US markets, if generators are centrally dispatched in a DAM or as part of RUC there is a “make whole” provision that guarantees recovery of startup and no-load cost over a 24 hour cycle. **From an economics perspective, the energy only approach represents the theoretical “gold standard” and has been strongly advocated in some jurisdictions, especially in some countries in Europe, Australia, New Zealand and in the US in ERCOT**. However, there are certain realities pertaining to electricity markets that present obstacles to the proper functioning of such an idealized framework, at least in some jurisdictions. One characteristic of electricity markets that distinguishes them from other commodity markets is the presence of a “System Operator, i.e., the ISO" who is responsible for ensuring reliability and hence mitigate scarcity conditions. This is done typically through out of market actions such as “Reliability Unit Commitment” (RUC), deployment of reserves and demand side resources. Such actions tend to mask scarcity conditions and reduce infra-marginal rents which may result in “missing money” for generators and reduce investment incentives. Limited demand response mechanisms that would allow scarcity prices to be set by the load, as the theory ascribes, **as well as inability to establish an individual reliability market due to lack of technology and engineering infrastructure**, also impedes the market's ability to set true scarcity prices.

To avoid such market interference and obstruction of scarcity price signals, the ISO could determine scarcity conditions based on use of reserves and other mitigation actions or emergency states, and then set the clearing prices to the price cap administratively during scarcity conditions. In ERCOT for instance, the ISO has adopted **an administrative operating reserve demand curve (ORDC) to calculate an energy price adder that reflects the cost of operating reserves depletion.** Some jurisdictions, however, have rejected such rules and insist that scarcity prices should be based solely on offer prices. This policy effectively requires that generators exercise market power during scarcity. Specifically, if the price cap is very high, it may be sufficient to attract forward capacity at acceptable levels. Sustaining the price cap at very high levels helps the market achieve a volatile profile of a typical commodity market and creates the conditions for future investments to be weighted more heavily towards equity and less towards debt. **The debt to equity ratio is definitely an important factor affecting investment in an energy only market.**

In some jurisdictions, high volatility is not encouraged. One could argue that the adverse effects of price volatility could be mitigated in a well functioning market by forward contracting and other risk management practices. This is indeed true, however, the realities of electricity markets suggest that vertical disintegration in many restructured electricity markets and the re-regulation of some segments (e.g., the retail market) has resulted in improper distribution of risk along the electricity supply chain. **This misallocation of risk results in improper risk management, as was the case in the original design of the California market in the 90s.** Consequently, some regulatory intervention in some markets, at least on a temporary basis, might be needed in order to achieve socially efficient risk management. Such regulatory intervention, requires caution since measures taken to ensure generation adequacy may have the effect of suppressing energy prices due to excess capacity or perverse incentives so that the necessity of such measures becomes self-perpetuating. This has clearly been the case in several restructured markets, such as in Argentina for instance, where a large capacity payment paid on the basis of generated energy induces generators to bid below marginal cost so as to increase production and, consequently, capacity payment revenues. As can be seen, the Energy Only market option requires robust market design and effective risk management practices.

Policy makers, in some jurisdictions, raise several concerns when reliance on energy prices is deployed to cover capacity costs through scarcity rents. Non-storability of electricity in massive scale, demand and supply uncertainty, inelastic demand and the steepness of the supply curve at its high end all contribute to high price volatility when reserve margins are low. While some temporary high prices reflect legitimate economic signals that are needed in order to attract investment, they are politically unacceptable in many jurisdictions especially since it is impossible to differentiate between legitimate scarcity rents and high prices resulting from market power abuse, or from strategies such as the “Hockey Stick” bidding strategy that exploit the inelastic demand and flawed or incomplete market rules. Furthermore, they claim that, even if high prices do reflect legitimate scarcity rents, which induce investment, sustained levels of scarcity rents **while new capacity is being built** will result in unacceptable transfer of wealth from consumers to producers.

Such concerns have prompted the impositions of price caps and market mitigation

procedures that provide a disincentive for investment by suppressing legitimate scarcity price signals.

Furthermore, the reliance on scarcity rents to recover investment cost also has an undesirable side effect with regard to the implementation of innovative demand response mechanisms. The primary purpose of such mechanisms is to mitigate the impact of supply uncertainty and reduce the escalating cost of ancillary services due to massive penetration of intermittent renewables resources. However, one of the byproducts of demand response is abatement of scarcity prices and hence such demand response initiatives, particularly those that entail a capacity payment for interruptible loads, face objection from generators who regard such payments as preferential treatment of demand side resources.

Currently, some energy markets in Europe such as the Nordpool and the Australian Victoria pool are Energy Only markets but over the last few years we see an emergence of capacity market in places such as the UK, Alberta Canada as well as discussion favoring such markets in Germany, even though in Germany capacity markets are disguised under the "strategic Reserves" market. All US markets have capacity markets, except the ERCOT market in Texas.

In summary, if the energy only market framework is adopted it is critical to incorporate a robust scarcity pricing through highly caped energy prices and operating reserves demand curves (ORDC) as in ERCOT. But we need to emphasize that such market architecture leads to a volatile profile and creates the conditions for future investments to be weighted more heavily towards equity and less towards debt as we discused earlier.

Even in energy only markets with very generous scarcity pricing it is doubtful that that supply can be secured without a resource adequacy mechanism. This is an important practical observation due to the fact the scarcity pricing alone is not expected to provide the required degree of certainty to revenue streams to secure sufficient investments.

The alternative approach is a resource adequacy approach that compensates generators for installed capacity with provisions that limit energy prices and hence reduce scarcity rents. These approaches are motivated by the fact that that energy markets encounter difficulties in discovering efficient prices during scarcity periods as we discussed in the Missing Money Section 7.6.1. Such resource adequacy approaches are determined in a variety of ways ranging from administrative capacity payments to payments embedded in bilateral contracts in response to a mandated capacity obligation or payments determined through a centralized capacity market. Some proponents of capacity payments have characterized generation capacity or the "supply reliability" provided by such capacity as a distinct product from the energy provided by the capacity. In the next Section we'll analyze various resource adequacy approaches that can be applicable to the Chilean market.

7.6.3 Recommendation

Based on the analysis above we strongly recommend the following:

1. Ensure a robust scarcity pricing methodology implemented through highly caped energy prices and operating reserves demand curves (ORDC) as in ERCOT
2. Implement a Capacity Remuneration Mechanism (CRM) in addition to scarcity pricing; the CRM apporach can take different forms as we'll discuss in the next Section as in PJM or CAISO or CREG in Colombia.

7.7 Capacity Market with Flexibility Enhancements vs Capacity Adequacy Approaches (like Optimized Capacity Market or Capacity Payment Schemes or Centralized Reliability Options)

7.7.1 Capacity Market with Flexibility Enhancements

Capacity Markets for a single product may provide enough capacity to meet reliability requirements, but it may not be sufficient to ensure that enough "flexible" capacity is procured to deal with the unpredictability and intermittency of RES resources. In the new environment flexible generation should be valued and procured by deploying Capacity Markets with specific provisions for that purpose.

One way to reward flexibility is to incorporate ramping services as an element in the Capacity Market mechanism. This is important because, based on our experience, traditional Capacity Markets are likely to guarantee a certain *quantity* of capacity, they often fail to incentivize the flexible type of capacity that the system needs. This consideration point to the substantial value of flexible power plants which can provide flexible capacity ideally situated to deal with the intermittency of the renewable-based generation.

Therefore, we strongly recommend to take the above analysis into account and design the Capacity Market in such a way to result in higher remuneration for those generators that are best able to change their production quickly, reliably and by large increments. Creating multiple "capacity segments", that have different value and are subject to separate capacity auctions, is one way of doing so. In such a system, there are multiple categories of capacity that each contain resources with a specified level of ramping capabilities. The segment with the most flexible generation capacity is auctioned first and clears at the highest price. Then the second segment, with less flexible firm capacity, is auctioned and clears at a lower price. This process of clearing increasingly less flexible firm capacity continues until the needed amount of firm capacity is procured. All generators bid into the highest capacity segment that they qualify for and receive the clearing price of that segment.

In contrast to a single clearing price system where all firm capacity receives the same price, a multiple clearing system rewards conventional generators according to the type of ramping services they have to offer. **Under both types of auctions, the same amount of firm capacity is procured and the overall cost can remain equal as long as the rise in capacity payments for more flexible capacity is offset by a decline in revenues for less flexible capacity.**

A second way to acquire the type of resources a power system needs is to tender for these resources specifically. Identifying the exact resource needs of the national or regional electricity system is no straightforward exercise. Nonetheless, the ISO can make a forecast of total electricity demand and production by variable renewables. By taking the difference between the two – and taking imports and exports into consideration – the ISO would then find the "net demand". Simply put, the net demand indicates how much firm capacity needs to be available in the power system and what ramping qualities this capacity should possess.

On the basis of this net demand forecast, the ISO can then define specific products that it needs to ensure the efficient management of the power grids for the years to come. For instance, in a system with a high installed capacity of solar PV, one can imagine a tender for a ramping product that is available between 6 and 8 PM – the period where output by PV is likely to decline and where electricity consumption is reaching its daily peak. By tendering for these products on a forward basis, the ISO would offer multi-year contracts to the winning bidders and ensure reliability by penalizing non-delivery of the contracted flexibility services. Such tendering systems should allow for equal participation of parties offering demand side management or storage opportunities.

Such a system carries two key advantages: first, it aligns the needs of the power system with the electricity products that market players are bidding to offer, and secondly, it gives those generators who can offer the flexibility services that the market needs access to additional long-term revenues.

Both solutions provide clear frameworks that allow for an optimal delivery of the flexibility services that ISOs require to efficiently manage the power grid. Nonetheless, they require tailor making to meet national and regional needs. The urgency of incentivizing flexibility, as well as the most efficient way to attain the desired level, will differ between countries and regions as a function of their electricity mix, the availability of interconnection capacity, the age and lifetime of generation assets and prevailing legal prerogatives.

7.7.2 Capacity Adequacy Approaches

We present three capacity adequacy options:

1. The Bilateral Market with mandatory capacity contracting imposed on load serving entities, with possibly prescribed resource mix and flexibility, as in CAISO
2. Capacity Markets with auction for forward generation capacity commitments and possible demand side participation for demand side reduction, as at PJM
3. Firm Energy Options with a regulated strike price, procured centrally by the

regulator and financed through a “reliability charge” imposed on the demand, as implemented by CREG in Colombia

7.7.2.1 Bilateral Market with Mandatory Capacity Contracting

An example of this capacity market mechanism is California. The California generation adequacy mechanism relies on bilateral procurement of generation capacity by California Public Utility Commission (CPUC) jurisdictional Load Serving Entities (LSEs), (i.e., Utilities) in response to a regulatory annual “showing” requirement. The showing requirement is based on input from the California Independent System Operator (CAL ISO) and California Energy Commission (CEC). The procured capacity is reviewed and supplemented by CAISO using a backstop procurement mechanism. An advantage of this mechanism is that it gives jurisdictional control to the CPUC and enables the CPUC to enforce environmental objectives such as Renewable Portfolio Standard (RPS), as part of the showing requirement. However, the main problem with this mechanism is the short planning horizon of one year which does not allow for sufficient look-ahead to enable new timely entry in case of a capacity shortfall. Increasing the showing horizon to three of four years, as suggested by the CAISO in a recent proceeding, would **disadvantage the unregulated LSEs** who do not have sufficiently long term contracts with load to cover a long term capacity contracts and are exposed to load migration risk. Consequently, the CPUC decided to maintain the current mechanism on a temporary basis until further review, exploiting the fact that there is no capacity shortage in California for the time being.

In general, one of the key arguments against the Centralized Capacity Market (CCM) is that there is no consumer benefit to paying the clearing price to all capacity that clears, even if it enables existing resources to earn a return above their costs of staying in business. **Although, a Bilateral Market might appear to be cost-effective in the short run, it can easily result in an excessive amount of retirements by facilities that are unable to earn enough to invest in environmental upgrades or repowering**. A good illustration of this fact is certain design features of existing power plants that have significant adverse environmental impacts and need to be upgraded. Under a Centralized Capacity Market where such capacity can earn the clearing price, owners of these resources will be able to make economically efficient decisions whether to cease operating or invest in environmental upgrades in response to any policy initiatives to develop environmentally sustainable capacity.

Under the Bilateral Market approach with the avoidance of paying a market clearing capacity price, these existing resources may have little or no choice but to exit the market, removing a potentially large amount of generation capacity which tends to be concentrated in load pockets and which could, if their revenues justified the investment, remain in operation with less ultimate environmental impact than developing alternative generation capacity for these areas.

Also, the absence of a transparent market provides opportunities for third-party intermediaries to capture a significant share of the consumer and producer surpluses that the Bilateral Market proponents assert will be realized as savings to consumers. Extensive experience has shown that when centrally-clearing transparent markets, such as a CCM, are not available such intermediaries provide valuable ‘market-maker’ services by reducing transaction costs for buyers and sellers, but they do so less efficiently than a CCM because each such intermediary controls only a portion of the market.

Thus, the Bilateral Market versus CCM distinction can be viewed as a distinction between non-transparent, less efficient markets in which consumer and producer surpluses are captured by private market makers, versus transparent efficient markets where the surpluses are realized by the buyers and sellers. **The result is that the purported cost savings from adopting a Bilateral Market approach rather than a CCM has little chance of being realized by the consumers to any great extent.**

The advantage of the Bilateral Market approach is its flexibility to allow regulatory intervention to establish specific goals for environmental policies such as GHG reduction and promotion of renewable generation. Specifically, the concern is that a policy of having a large pool of generators bidding against each other for a standard capacity product may provide insufficient regard to the environmental and operational aspects of the resource. Intervention by policy makers to enforce capacity targets by resource type may substantially complicate the auction. Indeed, a centralized auction via a CCM may not be successful in that respect. While a centralized auction approach may be well-suited to achieving system reliability, it is less clear that this is true for satisfying specific environment goals across multiple local capacity areas. So, it seems that a Bilateral Market regime is more conducive to the development of specialized resources that can suport a specific jurisdiction’s environmental objectives. An example of implementing such specific objectives is the mandating of 1.4GW of storage in California as part of the resource adequacy (RA) contracting obligation, imposed on load serving entities. Another example is a requirement that one third of the RA capacity meet specific ramping capability and be offered as flexible resources through economic (rather than physical) offers.

7.7.2.2 Capacity Markets with Auction for Forward Generation Capacity Commitments

The main goal of Centralized Capacity Markets (CCM) is to ensure availability of adequate generation capacity (determined based on technical criteria) at a future date and set a uniform price for all existing capacity based on the marginal capacity cost of new entry (CONE). An important advantage of CCM as compared to bilateral approaches is the assumption of credit risk by the ISO which enables longer term capacity contracts over time horizons that enable participation by potential new investors or existing generators offering new capacity to be ready three or four years in the future. The ISO is in the best position to assume such risk since regardless of load migration the ISO is in a position to collect payment for the capacity from the load at the time of consumption and pay to the seller of the option at the time the energy is delivered.

However, it is not necessary that all the capacity in the market be cleared through a CCM. Various forms of self-provision or opt-out procedures are possible. For instance, Load Serving Entities can contract bilaterally with generators for their capacity (as part of a long-term contract) and deduct that capacity from their procurement obligation or clear that capacity through the centralized market as a price taker by offering it at a zero price. A CCM can fulfill its purpose even if it functions as a residual market for a portion of the installed capacity and new entry. It is important, however, to protect a CCM against the exercise of market power on the buyer side (monopsony power). **If we rely on a CCM to set prices for all installed capacity, Load Serving Entities might have an incentive to keep that price low and accomplish that goal by procuring new capacity bilaterally and then offering it at low cost in the CCM so as to crash the market by lowering the clearing price for all capacity. Such exercise of monopsony power was attempted by the state of Connecticut in the ISO New England FCM, triggering a FERC ruling that prohibited such practice and the institution of rules to prevent such market abuse.**

As a general matter, we have observed over the years that a CCM would better promote investment, compared to the bilateral market mechanism, and do so more cost effectively, because it leads to greater price transparency and symmetry of information available to market participants. These properties help reduce transaction costs as well as mitigate market power.

Regardless of how the capacity payment is being set, the recipients of capacity payments undertake a “must offer obligation." Capacity contracts procured centrally also entail a “strike price” on energy, which can be set at a relatively low price cap ($1000 per MWh for most ISOs in the USA. It is important to note that the objective of the strike price is not to address market power concerns and set at very low levels.

In the initial design of the New Englad ISO Forward Capacity Market the strike price was set at the cost of a generic CT with heat rate of 22,000 Btu/kWh. Revenues accrued to such a generic unit when prices exceed the strike price are labeled “**Peak Energy Rents**” and deducted from the capacity payments. (Subsequent versions of the FCM have raised the strike price and eventually eliminated the Peak Energy Rents in FCM 10.). This is a form of a clawback mechanism.

A capacity market can be framed as a procurement of physical call options on energy (also referred to as Reliability Options, or ROs on future energy deliveries) with an energy price cap being the option strike price and the capacity payment or capacity component of the mandatory contract being the option premium (see Figure 7-1).

In essence an RO is a Contract **for Differences 'CfD'** between the ISO and the capacity providers; the ROs are bought by the ISO as a hedging instrument to hedge suppliers from high prices. In other words, ROs cap the exposure of Suppliers at the strike price. The Reliability Option Payback obligation (i.e., the clawback obligation) may be based on a **monthly floating strike price reflecting the cost of a hypothetical low efficiency peaking unit**. The same strike price will apply across all markets (Day Ahead, Intraday and Balancing). The rationale of a floating strike price is to ensure that fuel prices are reflected in the calculation and hence the strike price does not fall below the marginal costs of plants therefore promoting security of supply and efficient dispatch.

In addition to the energy payment generators are paid a "reliability premium" as a supplemental payment on top of the wholesale spot price. It is determined through an ascending clock auction. This option premium is charged to consumers as a "reliability charge", per committed MWh. The clearing price of the auction which sets the reliability charge based on offers by new entrants, whereas incumbent units are price takers, exactly as envisioned in the capacity market in Colombia.

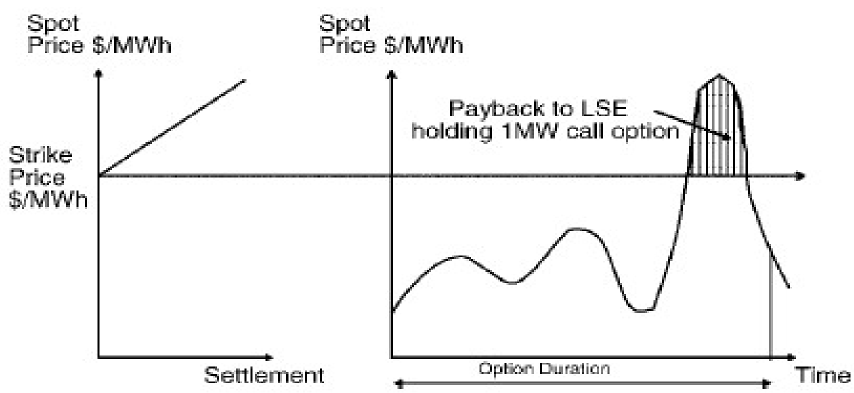


Figure 7-1: Payoff of Reliability Options from Capacity Providers to Load Serving Entities or ISOs

The choice of the strike price of the call option is a critical design parameter in this setup. At one extreme, a strike price of zero would correspond to a mandatory forward contract between the ISO or the LSEs and capacity providers, which would arguably interfere excessively with the risk management practices of stakeholders, who may not be willing to hedge their entire supply / demand in a mandatory forward arrangement. On the other extreme, an excessively high strike price invalidates the call option, since the option is rarely if ever exercised, and essentially becomes void.

Based on our experience from the ISO-NE and the Colombia markets we strongly recommend the strike pricr to be set at high levels. Any attempt to set it close to the marginal cost of a hypothetical thermal generator to deal with market power concerns is misguided. It should exceed the spot price by a large margin except perhaps during extreme rare events. **The general principle here is that reliability is assured even if the energy available is expensive while assuring low prices should be done through energy purchase agreement or energy price hedging.**

Such a high strike price can be set to a fixed level that should not be adjusted based on fuel costs and is independent of particular type of fuel used. So the strike price should be technology agnostic. Finally, it should be set at a level to encourage demand response participation.

7.7.2.3 Firm Energy Options with a Regulated Strike Price

As we discussed earlier alternative approaches to resource adequacy include:

1. Scarcity pricing in energy only markets, implemented through highly caped energy prices and operating reserves demand curves (ORDC) as in ERCOT (presented in Section 7.6.5)
2. Mandatory capacity contracting imposed on load serving entities, with possibly prescribed resource mix and flexibility, as in CAISO (presented in Section 7.6.2.1)
3. Capacity Markets with auction for forward generation capacity commitments and possible demand side participation for demand side reduction, as at PJM (presented in Section 7.6.2.2)
4. Firm Energy Options with a regulated strike price, procured centrally by the regulator and financed through a "reliability charge" imposed on the demand, as implemented by CREG in Colombia (which is discussed in this Section).

There is also an option of the Energy only market with mandatory long term Standardized Fixed Price Energy Forward Contracts procured through a central clearing house as proposed by Frank Wolak

(https://docs.cpuc.ca.gov/PublishedDocs/Efile/G000/M372/K082/372082582.PDF)

As we discussed in Section 7.7.2.2 the capacity-based alternatives whether implemented through a capacity market or mandated procurement, can be interpreted as procurement of physical call options on energy (also referred to as **reliability options, or ROs**) with an energy price cap being the option strike price and the capacity payment or capacity component of the mandatory contract being the option premium. The primary objective of these mechanisms is to ensure reliability in the sense of ensuring the availability of sufficient resource capacity to prevent outages due to resource shortage and provide income to incumbent generators and investors to support "iron in the ground" and cover for the "missing money" due to price caps. Further hedging to protect customers against high prices can be handled through forward financial energy contracts which we discuss in Section 7.5.

In hydro dominated system that are subject to possible weather driven energy shortages rather than capacity shortages, it is appropriate to introduce a resource adequacy mechanism based on **Firm Energy Options as in Colombia**. Such Firm energy Option may be used as an alternative or in parallel with capacity-based reliability options.

Firm Energy Options will be procured by the regulator or the ISO through a periodic single buyer auction for compliance in a future year (e.g., five to fifteen years in the future) to cover a high percentage of forecasted load, with additional periodic adjustment auction to account for forecast updates. Sellers will include incumbent resources and investors in new generation. New generation offers are subject to oversight of construction progress while existing energy constrained resources such as hydro must maintain minimum levels of energy in their reservoirs subject to high penalties. Premium payments for the Firm Energy Options based on the auction clearing prices are covered by a reliability charge imposed on load on a prorate basis. The Firm Energy options are designed to assure resource availability (regardless of price) so the reliability charge needs to cover investment cost or water holding costs but they are not intended to hedge fuel supplies or water opportunity cost so the strike price for the FIRM Energy Options will be of the magnitude as the price cap. The intent is to provide sufficient head room to resources so they can cover compliance cost when their Firm Energy Options are called. Price excursions up to the option strike price can be hedged separately through financial forward contracts described in Section 7.5.

In sum, like capacity mechanisms, Firm Energy Options should be viewed as a form of catastrophic insurance guaranteeing the reliability of the system and availability of energy regardless of price, while price hedging and market power mitigation should be handled separately. As such, it is appropriate to target resource adequacy mechanisms to physical market participants on the supply and demand side, and have the regulator or system operator administer and underwrite such insurance. Further price insurance for customers and load serving entities can be provided through financial bilateral contract or a centralized contract market as we discuss in Section 7.8.

We recommend against combining the resources adequacy insurance and the energy price hedging.

7.7.3 Recommendation

Based on the analysis above we recommend the following:

1. Implement a centralized single-buyer capacity auction with forward generation capacity commitments like in most US markets
2. Define the capacity auctions as Reliability Options, as the best way of risk management
3. Ensure that the strike prices are technology agnostic and very high. They should not be used to mitigate market power; other mechanisms such as bilateral

contracts or other hedging mechanisms should be used for this purpose
4. The definition of procured quantities remains an important challenge and requires special attention (we'll address this important issue in Task 3)
5. Scarcity pricing should complement CRM mechanisms
6. Firm energy options with a regulated strike price can also be adopted for energy constrained resources such as hydro resources as an alternative or in parallel with capacity-based reliability options.

7.8 Competition in Concentrated Markets (Ex-Ante Market Power Mitigation vs. Ex-Post Market Power Mitigation) (or Both)

7.8.1 Overview

This design market element contains an analysis of the efficacy of **ex-ante** vs **ex-post** mitigation measure options, the conditions under which each option may be preferable, especially in concentrated markets, the costs of each option to the market and elaborates on a potential hybrid approach where a combination of both options may be effective.

7.8.2 Ex-Ante vs. Ex-Post Market Power Mitigation Analysis

The **ex-ante** energy market supply offer mitigation is applied before the market clearing by deploying three types of tests: **structure, conduct, and impact.** (A detailed design of these mitigations screens will be proposed in Task 3).

- **Structural test**: Represents a determination that market circumstances may confer an advantage to a supplier. This may result from (1) a supplier being "pivotal" (i.e., load cannot be satisfied without that supplier) or (2) a supplier operating within an import- constrained area (with reduced competition).
- **Conduct test**: Represents a determination that the financial parameters of a supply offer appear to be excessively high, relative to a benchmark offer value (a "reference" value). The conduct test applies to all mitigation types.
- **Impact test:** Represents a determination that the original supply offer would have a significant impact on energy market prices (LMPs). This test only applies to general threshold energy and constrained area energy mitigation types.

In the **ex-post** approach bids are not mitigated and intervention is applied by the regulator if harmful behavior is observed. Specifically, the enforcement against abuses of market power may arise in the form of pre-specified **ex-ante** restrictions on firms and their behavior—such as price caps, bidding restrictions, or mandated prices that reflect anticipated costs—or the **ex-post** deterrence of harmful conduct through the prospect of investigations, after-the-fact mitigation, and costly punishment (*e.g.*, fines, damages payments, etc.) In either case, the choice of specific mitigation regimes (as well as the screening methods that trigger mitigation) has much to do with the Regulator's beliefs about two key issues:

- What is the likelihood and cost of mistakenly applying pre-emptive controls that may prevent sellers from charging prices that actually enhance economic efficiency? In diagnostic terms, this problem (i.e., the false identification of market power abuse when it does not exist) is known as a "**false positive" or "type I" error**.

- Conversely, what is the possibility that the enforcement regime will fail to detect market power abuse, or fail to do so in a timely manner, and what will be the economic and political cost of this failure? In diagnostic terms, this problem (i.e., the failure to identify market power abuse when it exists) is known as a "**false negative," which is sometimes referred to as a "type II error."**

In order to choose the optimal mitigation regime, policymakers and regulators need to develop an understanding of the "**loss functions**" that accurately represent their assessment of the likelihood of false positives or false negatives and the associated costs of such errors. This is especially relevant in concentrated markets where the potential for market power abuse is magnified. One also needs to consider the costs of enforcement regimes relative to alternatives, which include the costs associated with monitoring, evaluating, and mitigating market power, as well as the costs of evaluating and modifying the monitoring and mitigation processes as experience is gained and market conditions change over time.

Once such information is developed, or a decision is made based on policy and political considerations, one can choose among enforcement regimes so as to reduce mitigation errors and the expected costs. For example, a policymaker/regulator who believes that false negatives are associated with much higher societal cost than false positives might choose a mitigation approach that errs on the side of avoiding false negatives (*i.e.,* a relatively more "stringent" approach from the perspective of the market participants being examined.) Conversely, a policymaker/regulator who is concerned that the mitigation of false positives creates costly inefficiencies (e.g., distorted resource usage and investment disincentives) might choose an approach that errs on the side of avoiding false positives (i.e., a less stringent approach from the perspective of the market participants being examined.). In practice in all ISO markets in the US **ex-ante** market power mitigation is enforced followed by **ex-post** mitigation if necessary.

While it is well recognized that consumer harm resulting from false negatives and inadvertently unmitigated market power abuse can be extensive, the long-term cost of false positives and associated over-mitigation must not be underestimated, especially in concentrated markets. Mitigation actions, if they are erroneous or unnecessary, can promote both short-term and long-term inefficiency. This can lead to costly changes in the operations of generating plants and distorted prices that adversely affect investment incentives, contracting behavior, demand response, innovation, and dynamic (*i.e.*, long-run) efficiency. **Even if over-mitigation does not have significant price impacts, it may create a perception of having such price impacts, which may in turn create a perception of regulatory risk and undermine supplier and investor confidence—which can also result in under-investments and higher long-term costs to consumers.**

The implementation of automatic **ex-ante** mitigation in U.S. organized electricity markets (or, more precisely, the addition of **ex-ante** mitigation to **ex-post** monitoring and enforcement capability) differs significantly from the almost sole reliance on **ex-post** mitigation regimes (*e.g.*, enforcement of the antitrust laws) used in most other markets, including all organized electricity markets in the EU region. The combination of **ex-ante** and **ex-post** mitigation is also generally viewed to be a more stringent enforcement regime than those arising in most other markets, since electricity markets present unique challenges. As a theoretical matter, the stringency of a combined **ex-ante** mitigation and **ex-post** enforcement regime depends on the specified screening methods and the nature of mitigation (or sanctions) chosen by the policymaker/regulator.

In general, advocates for **ex-ante** mitigation argue that market participants prefer this approach because of its greater transparency and the reduced regulatory risk compared to sole reliance on **ex-post** enforcement. In addition, **ex-ante** mitigation avoids the often slow, potentially costly, uncertain, and burdensome investigations associated with **ex-post** enforcement regimes. It is also feared that, due to their costs and delays, **ex-post** enforcement processes do not reliably deter or mitigate market power abuses (particularly in markets where such abuses are likely to arise frequently as in concentrated markets or markets where the regulator does not have enough enforcement powers or even it lacks sufficient expert resources) or are unable to undo fully the harm caused by such abuses. Thus, concerns naturally arise that exclusive reliance on **ex-post** enforcement may lead to excessive regulatory risk and under-mitigation of market power abuse in markets where the conditions are ripe for frequent abuse of this nature. ECCO is clearly supporting these concerns and favors **ex-ante** market power mitigation in combination with **ex-post** market power mitigation.

Advocates for **ex-post** mitigation only claim that, while **ex-ante** mitigation can be comparatively transparent, quick and very analytical based on pre-determined mathematical methodologies, it risks being too prescriptive, overly broad, and having unintended consequences—notably a larger fraction of **false positives** (*i.e.*, the implementation of mitigation actions when market power abuse does not exist). Thus, they claim, the **ex-ante** market architecture may impose costs that exceed their corresponding benefits if they force market participants to alter their behavior under conditions where the market is performing efficiently. By contrast, **ex-post** mitigation is less transparent, can be less analytical and more specifically tailored to those instances in which a market participant is demonstrated to have engaged in anticompetitive or otherwise inefficient behavior. By conducting a full investigation that considers the specific facts and circumstances of a claimed abuse of market power, **ex -post** mitigation regimes can more reliably avoid **false positives** and the costs of over-mitigation.

Despite these claims, based on our practical experience, we advocate the implementation of **ex-ante** market power mitigation methodologies even at the risk of resulting in some **false positives.** This problem can be effectively addressed by applying more lenient **ex-ante** mitigation methodologies, like the NY ISO as opposed to more stringent one like the CAISO and PJM. Further, our experience from the energy crisis of the EU energy markets in 2022 and 2023 gives credence to the claim that the ex-ante market power mitigation is needed and if it were available in the EU markets based simulation analysis performed by ECCO it would result in 30% lower market prices on the average.

7.8.3 Additional Commentary

All ISOs in the US deploy structural and conduct-and-impact screening approaches as core elements of their market power mitigation processes. In simple terms, these approaches ultimately alter the bids of certain generators to a pre-defined reference level in order to prevent abuses of market power. Structural tests are used to: (i) identify geographic regions to be subject to more stringent conduct testing and (ii) identify the particular transmission constraints that will be subject to default mitigation (not applicable in EU markets). Whenever a structural test is failed, bids are frequently capped at a reference level that is meant to prevent an exercise of significant market power. Alternatively, structural screens are used in certain cases to identify markets in which a more stringent conduct- and-impact mitigation processes are applied.

PJM, CAISO, and ERCOT deploy a structural approach throughout the sequence of market monitoring and deploy price mitigation when bids are deemed to be submitted in what is considered a non-competitive market. **There is, however, important variation in the timing of structural techniques applied to energy markets.** In PJM, there is automatic mitigation of bids from generating units dispatched for congestion relief unless a structural screen (**the Three Jointly Pivotal Supplier or 3JPS test**) is passed on a day-ahead and real-time basis. Thus, the structural screen in PJM is unique in that it is performed after bids are submitted in the day- ahead and real-time markets. In contrast, under their new market designs, ERCOT and CAISO will use their structural screens only on a periodic basis (well before bids are submitted by market participants) to identify when the market is non-competitive and be subject to default mitigation.

Conduct-and-impact screens can be used for **ex-ante** mitigation only after bids are submitted in the day-ahead and real-time markets. Under such conduct-and-impact approaches, each generator's bids are compared to a predefined reference level that approximates competitive bidding. If such bids exceed predefined thresholds over those reference levels, the generator is said to have failed the conduct test. Then, the market price impact of the observed bidding behavior is measured, and if the unmitigated bids result in price increases above some predefined market impact threshold, generators that have failed the conduct test have also failed the impact test. Suppliers who fail the conduct-and-impact tests have their bids replaced with a reference level, called Default Energy Bids (DEBs) which is meant to approximate bidding under competitive conditions.

The ISO-NE, NYISO, and MISO rely on such conduct-and-impact tests for **ex-ante** mitigation. However, conduct-and-impact tests based on more stringent thresholds are often applied in transmission-constrained sub-regions that are more prone to market power abuses. These sub-regions are selected using structural techniques. For example, the energy and capacity bids of generators located in the transmission-constrained load pocket of New York City are subject to tighter thresholds. Similarly, the bids of generators located in transmission- constrained regions of New England, such as the Boston metropolitan area, also are subject to tighter thresholds. MISO employs a structural test to define Narrow Constrained Areas, to which stricter conduct-and-impact tests are applied. **One common theme unites all of the ISOs: structural tests in some form identify transmission-constrained regions requiring more mitigation.**

7.8.4 Recommendation

1. We strongly recommend the implementation of **ex-ante** market mitigation approach especially given the fact that the Chilean market is higly concentrated
2. We also recommend the application of **ex-post** measures to deter harmful market behavior, in addition to the ex-ante market power mitigation architecture.

7.9 Virtual Bidding Markets vs. Physical Markets Only

7.9.1 Overview

Two-settlement electricity markets consist of two interrelated markets: the Day-Ahead Market (DAM) and Real-Time Market (RTM). The DAM market is a forward market, where energy can be purchased at forward day-ahead prices, also called day-ahead LMPs

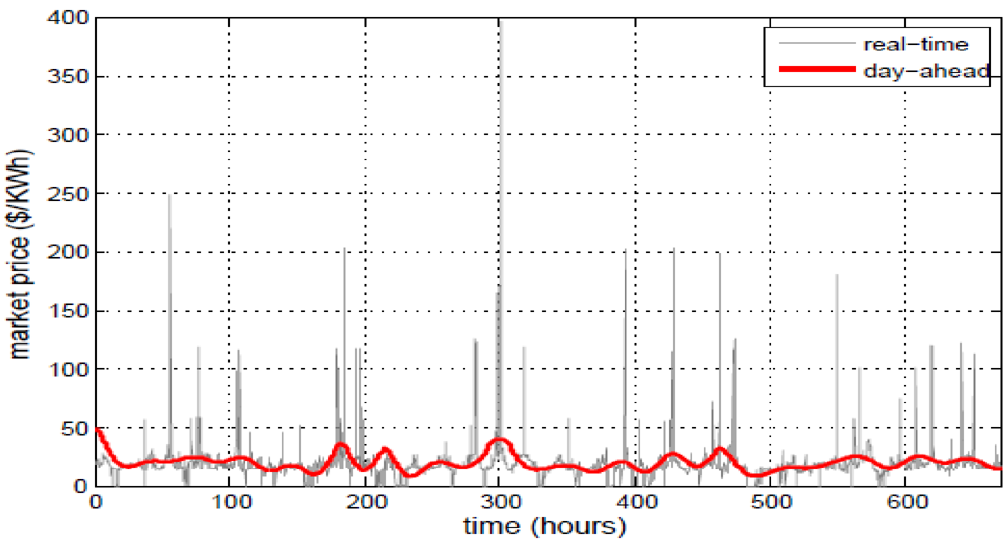


Figure 7-2: Real Time vs. Day Ahead Market Price Spread

(DAM LMPs). The RTM market is a physical spot market, where energy can be purchased at spot prices, also called real-time LMPs (RTM LMPs). DAM LMPs are usually more stable than RTM LMPs as illustrated in Figure 7-2.

In the RTM market, price spikes are often triggered by unplanned outages of generation plants and transmission facilities, and unpredictable weather, while the DAM market is less affected due to a longer planning horizon. The DAM market includes three sequential processes: market power mitigation and reliability requirement determination (MPM-RRD), integrated forward market (IFM), and residual unit commitment (RUC). The MPM-RRD starts the day before delivery.

In purely physical markets, participants are allowed to submit supply and demand offers and bids in the day ahead which result in forward market award that are binding contracts for specific quantities and price. Subsequently, they can submit real time offers and bids which modify the forward day ahead awards. The incremental real time awards are settled at the real time clearing prices. The combination of the day ahead commitment and the real time awards become physical dispatch instructions. Deviations from these instructions are treated as “uninstructed deviations” and settled according to rules specified in the tariff.

In markets like the CAISO, both physical and virtual (financial) participation is allowed until the start of the MPM-RRD. In the MPM-RRD, the ISO mitigates bids from physical resources that exercise locational market power, and ensures the availability of physical resources whose outputs are required to maintain local reliability. The results of the MPM-RRD are a pool of bids that is ready for the IFM. In the IFM, the ISO economically clears the supply bids against the demand bids with the transmission constraints enforced, determines DAM schedules and DAM LMPs, and procures ancillary services. When the demand forecast exceeds the total physical supply cleared in the IFM, the additional capacity is procured by the ISO in the RUC to satisfy reliability requirements. Note that the additional resources procured in the RUC are not directly used for production, and hence do not receive DAM LMPs. However, there are still costs to keep these resources staying online, namely start-up costs and minimum load costs, as discussed later.

In the RTM market, the ISO runs the economic dispatch process every 5 minutes to rebalance the residual demand, which is the deviation between the instantaneous demand and the scheduled demand in the DAM market. For one hour, RTM LMPs are the arithmetic averages of 12 5-min prices over the hour to settle the residual demand and the supply used to balance the residual demand. While DAM and RTM LMPs reflect the cost of energy production, generation plants also incur start-up and minimum load costs which they submit as part of their bids. Start-up costs are the costs that are incurred when generation plants are turned on, and minimum load costs are the costs that maintain generation plants to operate at the minimum load level. The CAISO guarantees that all dispatched resources who submit economic bids will cover their costs in the DAM and RTM markets. Hence, if a resource does not cover its total cost including start-up and minimum load cost through its energy revenue at DAM and RTM LMPs, its shortfall is covered by an uplift payment which is allocated to market participants based on a two-tier cost allocation scheme that considers both causation and socialization. The tier 1 uplift costs account for cost causation, and the tier 2 uplift costs account for cost socialization. Some uplift costs are allocated to virtual bids as discussed later.

7.9.2 Convergence Bidding (CB) in Two-Settlement Electricity Markets

CB allows market participants to arbitrage between the DAM and RTM markets through a financial mechanism, exempting them from physically consuming or producing energy. A virtual demand bid is to make financial purchases of energy in the DAM market, with the explicit requirement to sell back that energy in the RT market at the same location. Conversely, a virtual supply bid is to make financial sales of energy in the DAM market, with the explicit requirement to buy back that energy in the RTM market at the same location. On the physical side, the positions taken in the DAM market are offset by the opposite positions in the RTM market, which leaves market participants with no physical obligation. In anticipation of DAM LMPs being less than RTM LMPs, market participants can make profits by using virtual demand bids to effectively buy energy in the DAM market and sell it back in the RTM market. These virtual demand bids result in the additional demand in the DAM market that increases DAM LMPs, and the additional supply in the RTM market that decreases RTM LMPs. This yields the desired outcome of CB—price convergence. Price convergence is regarded as a benefit to the DAM and RTM markets. It reduces the incentives for market participants to defer their physical resources to the RTM market in expectation of favorable RTM LMPs. The improved stability of the DA market is beneficial from reliability perspectives. To ensure reliability of the power grid, the system operator is required to procure sufficient capacity in the RUC, when the total physical supply cleared in the IFM is not enough to meet the system operator’s demand forecast. With physical resources withheld by market participants, the system operator tends to over-procure capacity in the RUC. This raises the RUC uplift costs, and increases the risk of decommiting scheduled resources in the RTM market when deferred physical resources show up.

The benefit of CB also comes from the fact that it relieves market participants from using physical resources to arbitrage price differences between the DAM and RTM markets, also called **implicit virtual bidding** in some literature. Implicit virtual bidding is the bidding strategy where market participants intentionally defer their physical resources to the RTM market to take advantage of favorable RTM LMPs, by bidding at prices that are unlikely to be cleared in the DAM market rather than their economic costs and benefits.

Although implicit virtual bidding can achieve price convergence in the absence of CB, it can also lead to reliability problems that jeopardize the efficiency of the DAM and RTM markets. Without the revelation of the true economic costs and benefits of physical resources, it is difficult for the system operator to allocate resources efficiently and optimally. In addition, the prices at which market participants bid their physical resources largely depend on their own anticipation of DAM and RTM LMPs, and this introduces uncertainty into the DAM market. In some cases, the system operator can either over-schedule physical supply in the IFM that has to be sold back in the RT market, or under-schedule physical supply in the IFM that relies on the procurement in the RUC to balance. These variations decrease the stability of the DAM market, and could undermine reliability of the power grid.

CB can be conducted at both nodes and trading hubs. In comparison to nodes, trading hubs provide more liquidity to trade large volumes of virtual bids. In the CAISO, for instance, there are three trading hubs that correspond to three congestion management zones. DAM and RTM LMPs at the trading hub represent the weighted average of prices at generation nodes within the corresponding congestion management zone. The weights are determined annually based on the seasonal generation in the previous year, and are differentiated by peak and off-peak hours. The virtual bids submitted at the trading hub are distributed to generation nodes in proportion to their weights, and are bound together so that they are cleared as a whole in the DAM market.

The credit policy for CB requires that the current exposure of virtual bids submitted by a market participant may not exceed the collateral established with the ISO. The current exposure of virtual bids is calculated by the sum of the product of the quantity and the corresponding reference price of each virtual bid. For one node, the reference price can be set, for instance, the 95th percentile value of the historical price differences between DAM and RTM LMPs. After the settlement of virtual bids, the collateral is adjusted based on the realized profits and losses of virtual bids.

We do not recommend, a transaction fee imposed on submitted virtual bids, but cleared virtual bids should be required to pay uplift costs. The costs allocated to cleared virtual bids include may include the IFM tier 1 uplift costs, and the RUC tier 1 uplift costs. In particular, cleared virtual demand bids should be obligated to pay a proportion of the IFM tier 1 uplift costs, as virtual demand bids tend to increase physical supply procured in the IFM. Cleared virtual supply bids should be subject to a proportion of the RUC tier 1 uplift costs, as SOs tend to under-schedule physical supply in the IFM due to virtual supply bids and increase additional capacity procured in the RUC. In CAISO, the costs allocated to 1 MWh of cleared virtual position are estimated to be between $0.065 and $0.085.

The CAISO initial market reform (MRTU) excluded virtual bidding out of concern that the market can be manipulated by speculators. After the first year convergence bidding was introduced and Figure 7-2 illustrates the transaction volume and weighted price differences (spread) between the DAM and RTM for one quarter. Overall the volume of virtual bidding has been about 10% of the total traded DAM cleared volume. We also see that arbitrage hains and losses of the virtual transactions were quite evenly distributed

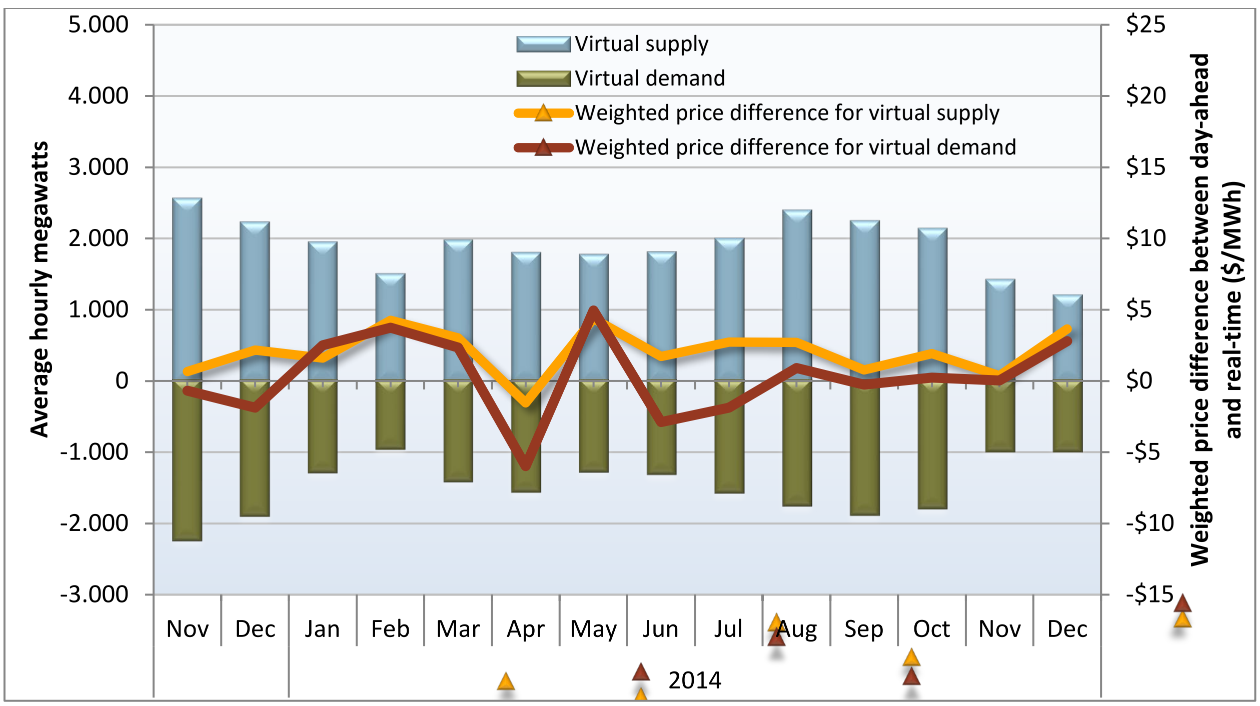


Figure 7-3: Virtual Trading Volume and Weighted Price Spread

Analysis of the year before and after the adoption of convergence bidding in several academic papers show clear reduction of the spread and arbitrage opportunities between the day ahead and real time prices. Fugure 7-3 adopted from a paper by Li, Oren and Svoboda (2016)[23] show the sharp decline in potential profit from arbitrage between DAM and RTM prices after the adoption of Convergence Biding in CAISO, which indicates price convergence and financial efficiency.

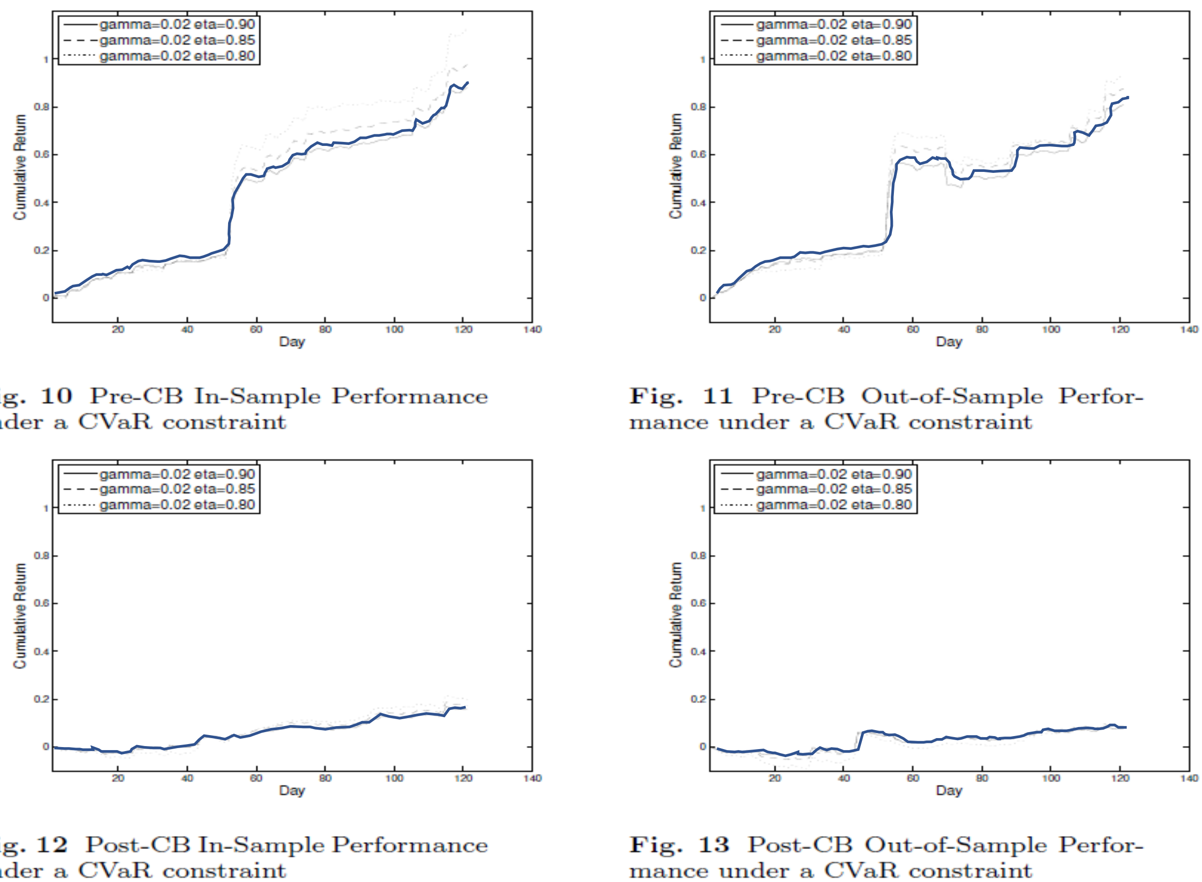


Figure 7-4: Cumulative Profitability of Arbitrage between DAM and RTM prices in

[23] Li Ruoyang, Alva J. Svoboda, Shmuel S. Oren," Efficiency Impact of Convergence Bidding on the California Electricity Market", Journal of Regulatory Economics Vol. 48, No. 3 (2015) pp245-284.

The above discussion clearly demonstrates the benefit of allowing convergence bidding in a market with a two settlement system as implemented in all US ISO market. However, an alternative that may achieve similar goals may be a physical market with intraday trading. Such intraday trading allowing market participants to readjust their hourly day ahead positions before real time may reduce the price spread between the adjusted forward clearing prices and real time prices which may make convergence bidding unnecessary. Nevertheless. given the experience in the US **we recommend a two settlement system with convergence bidding open to financial participants**.

7.9.3 Recommendation

1. Institute a two settlement system
2. Allow virtual transactions
3. Open virtual bidding to financial entities

7.10 Unit Commitment Costs (Startup/No load costs) & Ancillary Services Bidding Format

7.10.1 Overview

The Day ahead market procures resources for 24 hours of energy and ancillary services. The ancillary services should include regulation up and down, synchronized reserves that can be activated with 10 minute notice, non-synchronized spinning reserves that can be activated with a 30 minute notice and replacement fast start reserves that can be committed on short term notice. Additional products procured in the day ahead market may include flexible ramping reserves as in the CAISO or Contingency reserves as in ERCOT. These products are procured through concurrent procurement auctions to meet bid-in or forecasted load and reserve requirements that are determined by the system operator based on reliability criteria. **We recommend that these procurement auctions be Uniform Clearing Price Auctions.**

Pay As Bid auctions are possible, as was implemented in the UK, but such a market clearing approach is problematic from an incentives perspective if we adopt a bid rather that cost based approach, **which we recommend** and it would require administrative determination of the whole sale price charged to load. Such an approach is also incompatible with LMP **which we recommend**. So for the rest of this section we will assume Uniform Price Clearing.

7.10.2 Demand Side Bids for Energy

Bid-in load needs to be specified by location which can be a Node or load aggregation point (LAP) representing a zone or collection of nodes, for each of the 24 hours covered by the day ahead market. Such load bids can be submitted directly by load serving entities and whole sale customers or through **Scheduling Coordinators** as in CAISO or **Qualified Service Entities** as in ERCOT. Load bids can specify a maximum price for load increments. Bilateral self-scheduled transactions can be treated as paired load bids and supply offers where the load maximum price is very high (above the offer cap) and the supply offer is at a very low price. Such bilateral schedules can control their exposure to congestion charges by submitting “up to bids” (as at PJM) that puts a ceiling on the LMP difference between their source and sink which is the congestion charge they will face for the bilateral scheduled transaction.

7.10.3 Day Ahead Energy Offers

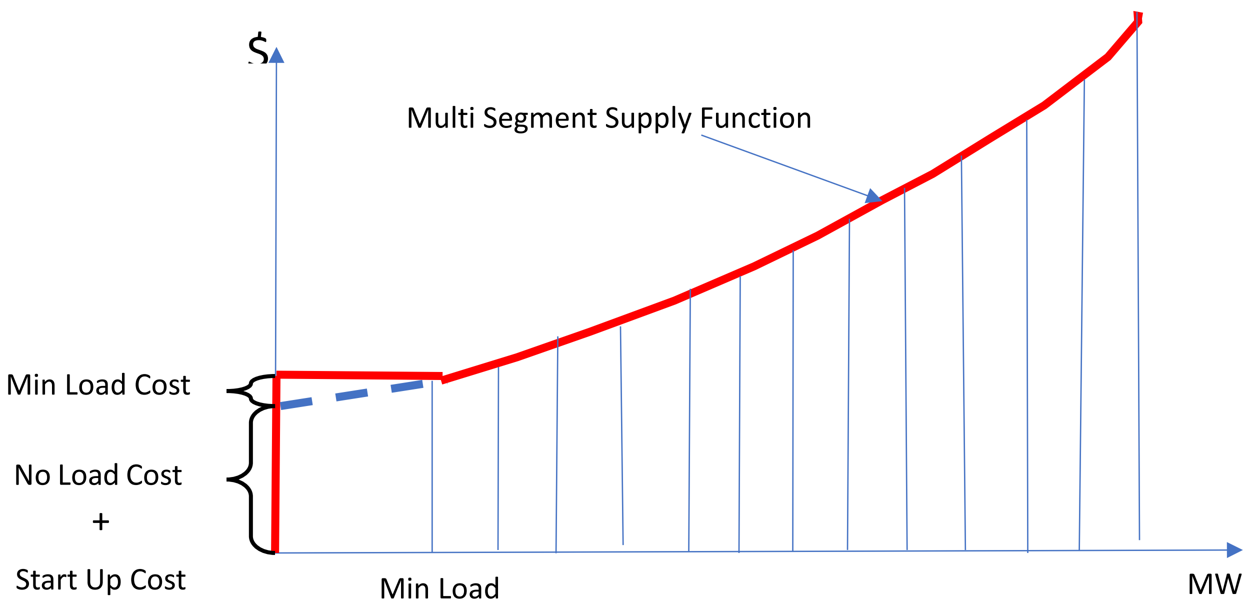
In a **centrally cleared market which we recommend,** resources are scheduled are committed in the day ahead in hourly increments for 24 hours using security constraint unit commitmeny (SCUC). The inputs to SCUC typically include cost parameters such as Startup Cost, No Load Cost, Minimum Load, and incremental energy cost, often represented by a convex piecewise linear function. The SCUC also accounts for technical parameter for each generation unit including, Max output, Ramp Rate, Minimum up time, Minimum down time and other technology specific constraints, as well as transmission constraints. Technical resource parameter are typically submitted by each resources and updated periodically, subject to verification. Some of these parameters are affected by weather conditions such as ambient temperature, maintenance protocols and emission constraints, but updating them should be infrequent to avoid manipulation. For instance, in the early days of the UK market Resource used flexibility declaration which could be varied hourly, to manipulate prices and exercised market power.

Cost components such as Startup Cost, No-load Cost, and Energy Cost can be Cost Based or Offer-Based. In a cost-based approach which currently prevails in Chile, the cost components are subject to verification and typically do not reflect opportunity cost. On the other hand under the offer based approach, the cost components are determined by offers submitted by the resources which may reflect opportunity cost determined by the resource owners. This is particularly relevant to hydro resources where the value of water is guided by long term planning consideration and opportunity cost. On the other hand offer, offer based cost components may still be subject to market monitoring and reverted to cost based default values when the market is deemed non competitive.

In some system like the Colombian system before 2009, dispatch was based only on simplified optimization using daily linear energy only offers ignoring Startup and No Load costs. **<u>The rules were amended in 2009 by CREG Resolution 51/2009 which introduced Stratup and No Load costs into the dispatch optimization</u>**. An economic analysis study by Alvaro, Bernal, Luciano and Oren[24] demonstrated efficiency gains resulting from the transition to central dispatch accounting for start up and no load costs. **Hence, we recommend that Startup and No Load cost be accounted for in the dispatch as common in SCUC.**

Another aspect worth noting is that it is possible to have cost based Startup and No Load cost subject to verification in combination with offer based energy. Such an approach has been adopted in the ERCOT market as a condition for eligibility for make whole payment that guarantee cost recovery for resources that are subject to central commitment. The result of implementing such cost verification at ERCOT has been a significant increase in self scheduling. In other systems like CAISO, the Startup and No Load cost are offer based but are subject to market monitoring oversight with prescribed bounds and the frequency of change in these offer component is controlled. As a transition from cost based to offer based we may consider an interim approach where the Startup and No Load cost remain cost based as in ERCOT with a sunset period that will allow the market to adjust to a full offer based approach.

Overall, whether the Startup cost and No Load cost are offer based or cost based the hourly day ahead offer has a three part format as illustrated in Figure 7-5. This offer structure which **we recommend** specifies Startup cost, No Load cost, Minimum Load and a convex incremental energy cost consisting of piecewise linear segments.



[24] Riascos, Alvaro, Miguel Bernal, Luciano de Castro and Shmuel Oren, "Transition to Centralized Unit Commitment: An Econometric Analysis of Colombia's Experience", Energy Journal, Vol 37, No. 3, (2015), pp. 271-291

Figure 7-5: Three Part Hourly Offer Format

Reliability (or Residual)Unit Commitment (RUC) is used by the system operator to commit additional resources when the day ahead market commitment appears not be sufficient to meet the system operator's load forecast. The RUC committed resources are selected so as to minimize a as-offered cost for the minimum load (consisting of start up, no load and Min load cost) although the RUC accounts for the full energy capability of these RUC committed resources.

7.10.4 Ancillary Services Offer Format

7.10.4.1 Regulation Service

Hourly regulation service can be offered as a bidirectional service or separated into Regulation Up and Regulation Down. These products are intended to provide load following energy and frequency control within each five minutes interval. In either case resources that can provide such service submit capacity offers consisting of quantity and price into a one sided uniform clearing price procurement auction with total procurement quantities determined by the system operator based on technical consideration. In addition Regulation offers should specify a "mileage" offer for the the sum of absolute deviation in energy deviation provided by the offered regulation resource. The millage settlement can be based on offer prices or a uniform clearing milage price.

7.10.4.2 Reserves

Resources submit resource offers according to their capabilities to meet requirements for the different reserve types that are classified by response time and possibly by rough geographic granularity that reflects possibly limited access due to systematic transmission constraints. The different reserves products are hierarchical so that higher products in the hierarchy can meet requirements for a lower product. The auction for all reserve products is simultaneous so as to meet all the reserve requirements at minimum "as offered" cost, with uniform clearing prices for each reserve product. Resources offers should specify an hourly capacity price, offered capacity, geographical location (with the same granularity as the reserve product characterization}, as well as the highest reserve product in the hierarchy that the resource can provide. Reserve offers should also specify energy offers for the energy that may be dispatched from the cleared reserve capacity. The downward substitutability of the products ensure that higher products in the hierarchy command highr clearing prices for the cleared reserve capacity.

The settlement received by each resource offered for reserves for the cleared capacity is based on the auction clearing price for the highest (in the hierarchy) reserves product the resource can serve regardless of how it is used. Furthermore, reserve resources that are deployed to produce energy are also entitled to payment for the energy they produce based on the corresponding LMP. Co-optimization of energy and reserves can further reduce dispatch cost by comingling energy offers with reserve capacity so that energy demand is served by the cheapest energy available regardless off how it is offered (i.e. as an energy offer or reserve offer). With co-optimization of energy and reserves, undeployed capacity of resources offered as either energy or reserves, which is in the money (i.e. with energy offer below the corresponding LMP) is entitled to opportunity cost payment for the undeployed capacity. Some systems like PJM employing such co-optimization with opportunity cost settlements require that reserve resources reduce their reserve capacity offer to zero. Such a restriction was also adopted by CAISO for flexible ramping reserves.

7.10.5 Cost-based vs. Offer-based

In all the offer formats described above offers can be set administratively to verifiable cost components or resources may be allowed to submit offers freely, which would reflect opportunity costs but still be subjected to exante market power mitigation. Munoz, Wogrin, Oren and Hobbs[25]. Provide an extensive discussion advocating the offer based approach. We repeat here the key arguments.

In a nutshell, cost-based spot market designs have two main features that make them inefficient. First, the exercise of market power is still possible in concentrated markets where there are barriers to entry, since firms have incentives to under-invest or to increase the share of the peaking technology, deviating from the socially-optimal generation portfolio. Thus, an offer-based design can be more efficient than a cost-based one even if firms can behave strategically in the spot market.

Second, auditing the true marginal costs of generation is difficult in a market environment when firms face important opportunity costs that are not directly attributable to expenses on fuel and other out-of-pocket operations and maintenance costs. Opportunity costs can be large in situations where, for example, generators face inflexible fuel contracts, firms are subject to environmental regulations or renewable generators can obtain additional revenues from tax credits or from the sales of RECs, and where there exist intertemporal generator constraints such as ramping limits or bounds upon the number of starts over a limited period. These become more relevant in systems with increasing shares of variable

---

1. [25] Francisco D. Munoz, Sonja Wogrin, Shmuel S. Oren, and Benjamin F. Hobbs, "Economic Inefficiencies of Cost-based Electricity Market Designs", The Energy Journal, Vol. 39, No. 3, (2018) pp 51-68

and unpredictable generation from renewable energy resources.

Finally, distortions in short-term dispatch schedules and prices due to disregarding opportunity costs could affect long-term contracting and investment choices. One would expect that short-term signals will inevitably influence longterm contract prices and, possibly, the choice of technologies, but the magnitude of this effect will depend on the risk attitudes of the generation firms present in the electricity market and the relative magnitudes of opportunity costs for different generation technologies.

Based on the above we recommend the offer-based approach with possible gradual transition pending the implementation of adequate market monitoring.

7.10.6 Recomendations

1) Implement an Offer-based three-part energy offers in the day ahead market including Startup cost, No Load cost and multi segment piecewise linear incremental energy cost, subject to ex-ante market power mitigation
2) Alternatively implement cost-based start-up and no-load costs (like in ERCOT) for the first three (3) and then transition to a full offer-based energy architecture for all three components
3) Energy offers should also include technical resource parameter such as Min Up time, Min Downtime, Ramp Rate, Min Load and Max Load which can be subject to verification and limited change frequency.
4) Regulation Offers should specify an hourly capacity cost and a "mileage" cost[26]
5) Reserves offers should specify an hourly capacity cost, an incremental energy price or price curve) as well as resource capability to meet reserve products requirement.

7.11 Impact of Low Carbon Assets on RES Operational Challenges

Variability and uncertainty are not new problems for grid operations – electric load is also variable and uncertain, and the output of thermal generators is not perfectly predictable due to unplanned outages, changes in atmospheric temperature and pressure and other factors. System operators have historically accommodated variability by committing and operating "regulating" reserves – resources that vary their output automatically based on the needs of the grid, either through direct detection of the grid frequency conditions or via automated signal from the control center.

What is different in a RES-based (or VRE-based) system is the scale of the challenge. In most studies ECCO has executed renewable penetrations above 50% that include large shares of wind and solar generation lead to variability and

[26] Panagiotis Andrianesis, Alex Papalexopoulos, "Performance-based Pricing of Frequency Regulation in Electricity Markets," IEEE Transactions on Power Systems, Special Section on Electricity Markets Operation, Vol. 29, no. 1, 2014, pp. 441-449.

forecast error that far exceed historical operating experience. In many studies, under some circumstances, RES systems may not have enough flexible capacity to meet the ramping requirements-the requirement to change output rapidly in the upward or downward direction-imposed by load, wind and solar.

Moreover, a RES-based system will frequently experience overgeneration conditions, i.e. hours when "must-run" generation (generation which nominally cannot be dispatched downward, including wind, solar, combined-heat-and-power generation, and thermal generation needed for local reliability) is greater than load.

Figure 7-6 below shows a sample operating day from a study we have executed that illustrates four distinct types of flexibility challenges and risks that the system will face under high renewable penetration. The dark black line near the top indicates the load that must be served during each hour. The shaded bands indicate the types of resources that are operating throughout the day.

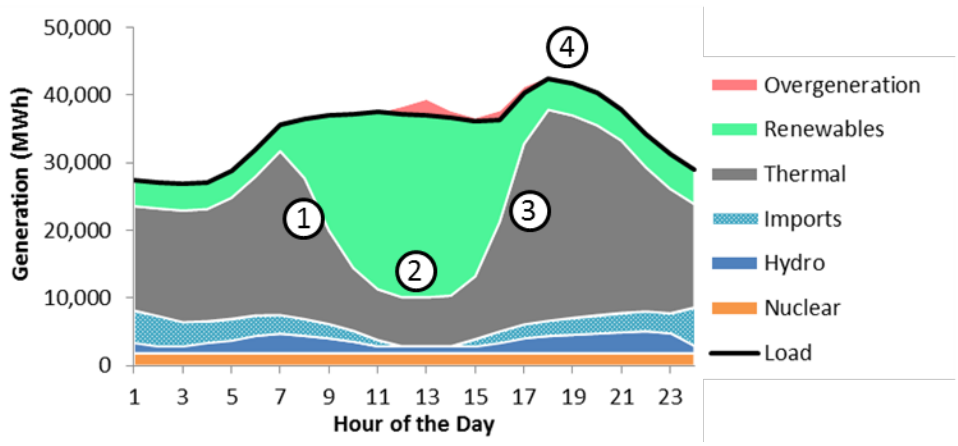


Figure 7-6: RES present Four Challenges

We can clearly observe four types of risks.

1. **Downward ramping capability:** Thermal resources operating to serve loads at night must be ramped downward and potentially shut down to make room for a significant influx of solar energy after the sun rises, in this case around 8:00 AM.
2. **Minimum generation flexibility**: **Overgeneration** may occur during hours with high renewable production even if thermal resources and imports are reduced to their minimum levels. A system with more flexibility to reduce thermal generation will incur less overgeneration.
3. **Upward ramping capability:** Thermal resources must ramp up quickly from minimum levels during the daytime hours and new units may be required to start up to meet a high net peak demand that occurs shortly after sundown.

4. **Peaking capability:** The system will continue to need enough resources to meet the highest peak loads with sufficient reliability.

A fifth flexibility challenge, not visible on the chart, involves ensuring that sufficient resources are operating to provide necessary real-time services such as primary frequency response, regulation, etc. This challenge can be exacerbated by RES resources, which tend to require more balancing resources and may not be able to provide these services as effectively as thermal generation.

Careful evaluation of Figure 7-6 gives credence to the claim that RES generation has a dramatic effect on the net load shape (i.e., Load profile MINUS load served by RES generation) which needs to be served by conventional generation, coal, gas, hydro, etc. In a RES-based system, the following system changes are predicted and are already present in some systems):

1. The peak net load period shifts, potentially to a smaller morning and evening peak, rather than a daytime peak period. Peak net load occurs when high load is coincident with low renewable generation. In the summer months, the effect of higher renewable penetration can shift the peak period from 3:00 PM with no renewables to 8:00 PM under high renewable penetration.
2. The net load that must be served with dispatchable generation shows steep ramp periods in the morning and evening, which will require the availability of flexible resources, like gas power plants.
3. In the middle of the day, conventional resources may need to be shut down to accommodate low, or negative, net load conditions. Low net load conditions will need to be accommodated while also maintaining enough flexibility on the system to accommodate the evening ramp up.
4. Negative net load conditions, in which renewable generation exceeds electricity demand will become typical in RES-based systems.

These observations create huge operational and financial problems. A comprehensive treatment of these issues is outside the scope of this Project but we offer the following potential solutions for solving the operational problems created by a RES dominated system.

- **Solution Category A: Enhanced Regional Coordination**

Enhanced regional coordination can help alleviate flexibility operational challenges in a RES-based system in at least two ways. **First, closer integration of operations between Chile and its neighbors may allow additional, latent generation flexibility that exists in other regions to be brought to bear to contribute to meeting flexibility needs in Chile (and vice-versa). This effect is approximated as a relaxation of hour-to-hour ramping constraints on power flows across the interties**.

Second, enhanced coordination can help Chile find markets in other neighboring states for its surplus energy that is available during hours with overgeneration conditions.

- **Solution Category B: Conventional Demand Response**

Curtailable load and other conventional demand response programs that result in reduced load during peak periods can help to provide flexibility by reducing the magnitude and frequency of **extreme ramping events**. If properly managed, the ramping capability of conventional demand response programs may also help with **overgeneration** by giving CEN more latitude to reduce generation from conventional resources during times of high renewable production. Cconventional demand response refers only to load curtailment.

- **Solution Category C: Advanced Demand Response/Flexible Loads**

Unlike conventional demand response programs, advanced demand response programs that provide **downward flexibility** directly contribute to mitigating the overgeneration problem by absorbing energy during times of surplus. Studies we have executed show that Advanced Demand Response is shown to provide both **the ramping contributions of conventional demand response and the downward flexibility benefits of the Enhanced Regional Coordination case.**

- **Solution Category D: Energy Storage**

Similar to the Advanced Demand Response solution described above, energy storage can contribute to mitigating operating flexibility problems so prevalent in RES-based systems **by providing both upward and downward flexibility**. Because overgeneration is identified as the most critical integration challenge, energy storage can be treated as a diurnal energy shifting technology, similar to the Advanced Demand Response solution, but also incorporating round-trip losses.

There are less critical solutions this mjor challenge, but clearly demand response, storage, etc., play a critical role soecuring the safe operation in a RES-based system.

8 Market Design Proposal Relation with the Current Chilean Energy Market Architecture

In this Section we analyze and present the current Chilean energy market architecture in relation with the analysis we presented on the chosen market elements of the proposed previous Sections. This analysis will provide clarity on the potential energy market changes on the existing energy market architecture the proposed recommendations of this Chapter will require. Our objective is to ensure that there is clear understanding of the market impact of the proposed changes.

For every market element analyzed in the previous Sections there is a direct analysis of the existing Chielan energy market architecture.

It is important to note that the key fundamental market elements of the existing Chilean energy market were not part of our analysis in Chapter 7. These include:

- the Day Ahead Market (DAM)
- the Real Time market (RTM) (not as a real time market but as an adjustment market from the DAM schedules to incorporate real-time information of generation and transmission assets to ensure reliability)
- the Physical Bilateral Contracts
- the Capacity Makret in the form of Capacity Payments
- the Nodal market architecture
- the LMP transmission pricing, and
- the co-optimization of energy with the Ancillary Services (AS) in the DAM

Note:

- A formal RTM is critically important and will be proposed as part of Task 3
- An important market element, not included in Chapter 7, is the two settlement system and an detailed analysis is included in Section 2.4.5.

8.1 Current Chilean Electricity Market Operation and Ownership Structure

The electricity demand in Chile is about 75 TWh per year, with a peak demand of around 11.5 GW (2023). On the supply side, there is 34.68 GW of installed capacity; which is owned by generation firms with different technologies. Up to October, 2023, there were 9.27 GW (26.7%) of solar PV, 7.5 GW (21.7%) of hydro, 5.38 GW (15.5%) of natural gas, 4.6 GW (13.2%) of wind, 4 GW (11.5%) of coal, 3 GW (8.6%) of diesel installed power capacity. Figure 8-1 depicts the energy mix per technology in the Chilean market.

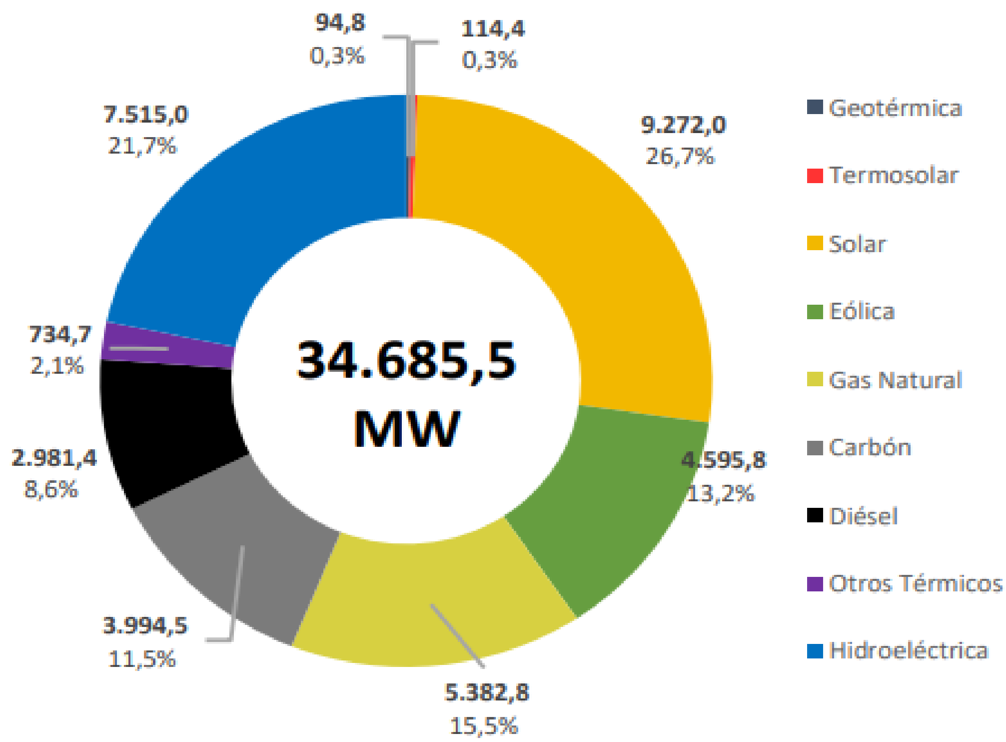


Figure 8-1: Energy Mix in the Chilean Market

Source: CEN, 2023. Reporte Energético. November, 2023. Available on-line at: https://www.coordinador.cl/wp-content/uploads/2023/11/CEN_Reporte_Energetico_SEN_Nov23.pdf

Up to October, 2023, there were 744 generation firms participating in the Chilean National Electric System (SEN). However, only 5 companies have more than 1 GW of installed capacity. **Enel Generación Chile S.A. owns 5.1 GW (15%) of capacity, Colbún S.A. owns 3.6 GW (10%) of capacity, Engie Energía Chile S.A. owns 2 GW (6%) of capacity, Aes Andes S.A. owns 1.8 GW (5%) of capacity, and Enel Green Power del Sur SpA owns 1.3 GW (4%) of capacity**.

This ownership share in the market gives credence to the claim that there is no dominant market participant in the market based on ownership shares. This means that the Chilean market will pass the structural tests (See detailed design of the Market Power Mitigation, in Task 3). However, we still need to apply a rigorous ex-ante market power mitigation based on the behavior of the players in the daily bidding process.

The short-term electricity market of the National Electric System (SEN) in Chile is a centralized, multi-nodal wholesale market (pool) based on audited costs and non-binding day-ahead operation scheduling. In this market, agents declare and support relevant information about their generation resources, such as variable operations and maintenance costs, minimum load, start-up costs, among other technical parameters, to the Coordinador (market operator). This information, along with the availability of power generation and transmission network resources, must be submitted before the realization of the DAM (day N-2).

At the same time (i.e., before the realization of the DAM, in day N-2), there is also an ancillary services (AS) market which is cost-based and clears in a co-optimized manner with the DAM using cost-based bids. The Coordinador makes a decision each hour after the cost-based DAM clearing whether the AS market is competitive in order to perform a follow up, competitive AS auction with competitive bids. In case the Coordinador determines the AS market does not present enough competition in the next hours, AS are directly instructed to some coordinated firms and paid based on an audited cost. **In the particular case of the frequency control services, which are bid-based only for the wear and tear costs of the power plants, existing side payments for opportunity costs and out of merit order generation takes place.**

In addition, each generator must periodically report fuel availability and its prices or generation forecast in the case of variable renewable energy (VRE) plants. However, there are specific regulations that allow certain technologies to be assigned a different variable operating cost from the declared cost. Such is the case of units that operate with LNG associated to long-term contracts for which the alternative cost of the energy determined by the Coordinador is utilized as the variable cost. In the case of the hydro reservoirs, the Coordinador determines the variable cost of the generators by estimating the shadow price of the water solving a medium-term (2-year-horizon) planning optimization problem, as explained later in Section 8.11.1.

It is important to note that hydro plants are owned by private companies which own the reservois and operate the turbines, but they are always following strict instructions from the CEN; that is, hydro power firms do not make any strategic decision or self-schedule regarding their real-time operations. We consider this as suboptimal and we’ll propose certain minor modifications later in this Section that offer some limited flexibility to hydro power owners to offer bids and control their dispatch decisions.

All the variables associated with costs and technical parameters are used to solve the day-ahead unit commitment problem (UCP) by co-optimizing energy and reserves for frequency control, modeling the system with a high degree of detail and technical constraints, such as constraints associated with thermal power plants, hydraulic series, and irrigation agreements, and using a 7-day horizon. As a result of this process, a merit order list is obtained, which is used to update the day-ahead UCP (pre-dispatch) operation. All these activities take place in the day ahead market (day N - 1).

There is no market operation in real time, but the Coordinador can make real-time reassignments or adjustments (using the merit order list obtained from the UCP) and submit instructions to coordinated firms to ensure power-system security and minimum-cost of operations, based on the latest information of generation and transmission assets, forecsts, etc.

Currently, only generators are allowed to participate in the spot market and all loads must have an energy supply contract with generators. These are private bilateral physical agreements. Generators can also enter into private financial bilateral contracts with industrial loads to hedge their market positions.

The energy market is based on locational marginal pricing (LMP pricing), **but these nodal prices are determined ex-post based on the actual system operation**.

It is important to mention that this short-term electricity market is complemented with long-term contracts both as bilateral contracts between generators and large industrial consumers and as auctions run by distribution companies to supply their estimated demand by end consumers. The bilateral contract of generators with the large industril loads could be physical or financial. **These long-term contracts are fundamental for providing stability to the market, so in all the recommendations for the short-term market we assume these long-term contracts remain as in the current system.**

In the Chilean electricity market, most regulation is detailed in the Law, which currently leaves little room for the CEN to make decisions on key issues. This is very important since an efficient offer-based market requires that the CEN, as the operator of the market, makes important market and operationl decisions always of course in adherence with the approved regulations and laws.

For example, it is very critical that the ex-ante market-power mitigation mechanisms introduced in the proposed bid-based market are implemented and managed by the ISO (CEN), consistent the new regulation rules encoded in the Tariffs approved by the regulator. This is the standard international practice worldwide. Any deviation from such practice will be to the detriement of the benefits of the transition to a bid-based market. For example, it makes no sense and it will be very inwefficient for the regulator (or an institution such as the "Panel de Expertos") to undertake such a market activity due to lack of daily operational and market information.

Of course the regulator can have access to the market data, if requested, and administer any ex-post investigation of market power abuse.

Next, we present an analysis of each market element of the proposed market design identified in Chapter 7 and a comparison of these options with the current Chilean market design architecture.

8.2 Price Formation (Pay-as-Clear or Marginal vs Pay-as-Bid)

As explained in the previous Section, our recommendation is to implement a pay-as-clear pricing mechanism (or uniform price formation) because this mechanism incentivizes bidding the true marginal cost of each firm. The Chilean electricity market already has a **single settlement pay-as-clear pricing mechanism** (with the particularity that it is based on audited costs and non-binding day-ahead operation scheduling). As such the market design philosophy with respect to this important market element remains the same, which provides a clear advantage for the transition to the proposed offer-based system.

Nevertheless, it is important to remark that there is no formal Real-Time Market (RTM) in the Chilean electricity market. Currently, the CEN solves the day-ahead UCP and, as the result of this process, **a merit order list** is obtained. In real time, no optimal dispatch is solved, but the CEN makes real-time reassignments (or adjustments when needed) based only on this merit order list obtained from the UCP, without re-computing optimal dispatch. These adjustments are required to ensure the dispatch is physically realizable based on the latest reall-time information. This implies that CEN submits instructions to coordinated firms, to ensure power-system security and minimum-cost operations, based on its own intra-day forecasts under the current operation conditions. As such, no pay-as-clear pricing mechanism is implemented in Chile in the real time dispatch.

The existence of a Real-Time Market (in which the optimal dispatch is computed every 5 or 15 minutes and prices are formed using a pay-as-clear pricing mechanism) is fundamental in any electricity market to avoid inefficient dispatches. Therefore, the reliance on adjustments based on the DAM merit order list as a proxy of an optimization-based RTM is suboptimal. **This is independent of the implementation of an offer-based or a cost-based market.**

In Task 3 we propose the design and implementation of a formal bid-based RTM to deal with un-anticipated events as they materialize close to the real-time operations.

Another feature of the Chilean electricity market is that prices are computed at a nodal level. The short-term electricity market in Chile is a centralized, multi-nodal wholesale market (pool) based on audited costs. Since we are proposing a locational marginal pricing system for the offer-based market, we view the required market modifications consistent with the current Chilean energy market.

8.3 DAM Market Arquitecture

In the Chilean electricity market, the CEN solves the day-ahead UCP by co-optimizing energy and reserves for frequency control, modeling the system with a high degree of detail and technical constraints, such as constraints associated with thermal power plants, hydraulic series, and irrigation agreements, and using a 7-day horizon. This UCP is solved in a unified manner; that is, with no splitting by technology. **<u>This is an important advantage of the current Chilean system that we recommend to maintain, as our recommendation is to implement a DAM unified, without splitting UCP by generation technology. The benefits of technology agnostic, bid-based DAM auctions are fully analyzed in Section 7.3.</u>**

Nevertheless, as mentioned, it is important to remark that there is no formal otpimiztion-based Real-Time Market in the Chilean electricity market and that it is fundamental to implement such Real-Time Market, independently of the implementation of an offer-based or a cost-based market architecture. The current use of the merit order list (obtained from the UCP), without re-computing the optimal dispatch based on updated information for un-anticipated events, gives strong credence to the claim that CEN produces instructions to coordinated firms that are suboptimal, as compared to the optimal dispatch using real-time information in an optimization formulation of the real-time operations.

8.4 Long-term Contract PPAs

The Chilean electricity market already allows long-term bilateral contract PPAs. The Chilean system has two types of long-term bilateral contracts. On the one hand, bilateral contracts between generators and large industrial consumers occurs spontaneously in the market. These long-term contracts are private agreements between parties, so only the amount of energy and capacity contracted (and their specific requirements) are informed to the CEN. **<u>These contracts could be physical or financial.</u>** On the other hand, distributions companies are required by Law to auction the energy they estimate their customers will need for the next 10, 15 or more years in public tenders. These public tenders are run in a centralized manner and some particular rules are applied to some tenders to facilitate the offers of VRE (for example, there are some tenders in which generation firms can offer energy provision only during day hours). **<u>The contracts between the generators and the distribution compnaies are physical contracts.</u>**

These long-term contracts are fundamental in the Chilean market for providing price stability and finantial certainty. As mentioned in previous Sections, our recommendation is to maintain these long-term physical contracts as in the current Chilean system and complement it with the short-term offer-based market to be proposed. **<u>We also propose financial contacts with CfDs as fully analyzed in Section 7.4.</u>**

It is important that long-term contracts in Chile maintain the following two principles that they currently have: (i) The regulatory authority must perform all tasks needed to ensure the compliance of supply obligations and balance payments (contract responsabilities) and (ii) the regulatory authority must encourage technology neutrality.

However, there are some improvements to long-term contracts that may be implemented. **For instance, the regulatory authority may consider financial instruments, such as Finantial Transmission Rights, that allow market participant to manage congestion risks. This would allow a better management of long-term contracts in Chile.**

8.5 Financial Markets vs Physical Markets

In Chile, currently, only generators are allowed to participate in the spot market and all loads must have an energy supply contract, which are private bilateral physical agreements. Generators can also enter into private financial bilateral contracts with large industrial loads to hedge their market positions.

We plan to expand in the proposed bid-based energy market design the scope of the financial markets. **This will include a standardized financial forward energy market with a central underwriter which is open to financial participants with regulatory oversight. The benefits of such market are fully analyzed in Section 7.5**. We also recommend to institute credit worthiness and contract enforcement for the standardized forward energy contracts with possible verification of fuel supply hedging.

In terms of lhysical markets, we recommend the implemention of a **physical resource adequacy mechanism** with a capacity or firm energy must offer obligation with a high price cap. A mixed approach with a capacity obligation for thermal resources and firm energy obligation for hydro power is also possible. We recommend this mechanism to be restricted to physical participants or investors in future resources and load and should be subject to verification.

We also recommend to impose a forward energy contract procurement requirement on load serving entities serving regulated load and on resources with capacity or energy obligated under the resource adequacy mechanism. With respect to load serving entities we recommend the elimination of spot price pass through provisions for entities serving regulated load, exposing them to spot price risk, and instituting incentives for managing price risk of regulated load. The detaild of these important changes are included in section 7.5.

8.6 Energy only Markets with Scarcity Provisions vs Resource Adequacy Mechanisms (or Both)

The short-term electricity market in Chile is a centralized, multi-nodal wholesale market, based on audited costs, which co-optimizes energy and reserves for frequency control considering all technical constraints and using a 7-day horizon.

To solve the missing-money problem, the Chilean electricity market has implemented a centralized capacity payment. The capacity payment is a fixed amount that is distributed to all generation firms participating in the market at pro-rate of their "firm capacity". The regulatory authority is the one in charge of defining the capacity that is considered firm and, thus, that is subject to capacity payments.

Since the current market is cost-based, there is no price cap in the spot market in Chile because, currently, prices are based on audited costs and generation firms are the only ones allowed to participate in the spot market. Also there is no scarcity pricing of any kind.

The transition to a bid-based market architecture, however, requires the enforcement of price caps. Further we plan to recommend a scarcity prcing mechanism in addition to a Resource Adequancy mechanism. All US market and many EU market have a combination of both such mechanisms now. The details of the benefits of the scarcity mechanism is provided in Section 7.6.

8.7 Capacity Market Mechanisms

There is no organized capacity market in the Chilean electricity market. The Chilean electricity market has implemented a **centralized capacity payment**. **The capacity payment is a fixed amount that is distributed to all generation firms participating in the market at pro-rate of their "firm capacity". The regulatory authority is the one in charge of defining the capacity that is considered firm and, thus, this capacity is subject to capacity payments**. This issue (the definition of firm capacity) has been controversial in Chile, especially in the cases of VRE and storage services, due to their intermittency.

In Section 7.7 we fully analyzed the benefits and the details of a capacity market and its variants with the introduction of a bid-based energy market. **Specifically, we recommend a centralized single-buyer capacity auction with forward generation capacity commitments like in most US markets. We plan to define the capacity auctions as Reliability Options, as the best way of risk management. Firm energy options with a regulated strike price can also be adopted for energy constrained resources such as hydro resources as an alternative or in parallel with capacity-based reliability options.**

Finally, as we discussed in Section 8.5 we also recommend the implementation of a scarcity pricing mechanism to complement the CRM mechanism.

8.8 Market Power Mitigation (Ex-ante vs. Ex-Post)

In Chile, generation electricity market is based on market competition, which allows free (although limited) consideration of market opportunities and risks by generation firms. **Within this operating and regulatory environment, open-access to the transmission system and generation technology neutrality have been two important pillars of the Chilean market operation. These features of the Chilean electricity market must be maintained**.

In an bid-based market, ex-ante market power mitigation is obviously more challenging than in a cost-based market. However, some characteristics of the Chilean system, such as the open-access to transmission system, the generation technology neutrality, the locational marginal pricing mechanism, and the obligation of loads to have long-term energy supply contracts, facilitate the implementation of ex-ante market power mitigation, as recommended in the Section 7.8. This Section clearly demonstrates the benefits of ex-ante market power mitigation as opposed to ex-post market power mitigation which can always be exercised by the regulator if market outcomes warantee such an action.

In this context it is important to mention that the international standard market practice (for efficiency purposes) is that the ex-ante market-power mitigation mechanisms are designed, implemented and administered daily by the ISO, based on ex-ante market power mitigation transparent rules which have been approved by the regulator.

In Chile, given the detailed nature of Chilean regulation, this might be a major change to the current market standard practice. **It is crucial to understand that the CEN should design the details of the implementation of the ex-ante market-power mitigation measures and propose the mitigation rules to the regulator for approval**.

Any deviation from international standard market mitigation practices will create inefficiencies to the detriment of introducing a bid-based energy market architecture. CEN operates the energy market daily and has all the technical information of the market participants on a daily basis. If the regulator (or an institution such as the "Panel de Expertos") intents to manage the ex-ante market-power mitigation mechanism we can easily foresee major ineficiencies, delays, and potentially inaccurate outcomes dues to lack of information and daily data submitted to the CEN for operating the market.

In summary, we strongly recommend the implementation of an ex-ante market mitigation approach especially given the fact that the Chilean market is higly concentrated. This is especially important given the ownership structure of the reservoirs. The risk of collusion is real and this makes the immediate

implementation of the ex-ante market power mitigation even more critical from the startup of the the bid-based energy market architecture. We also recommend that this market mechanism should be daily administered by the CEN.

Finally, we recommend the application of **ex-post** measures (in the form of ex-post investigations, penalties, fines, etc.) administered by the regulator to deter harmful market behavior, in addition to the ex-ante market power mitigation architecture as it practices in all international energy markets in the US and the EU regions.

8.9 Virtual Bidding Markets vs Physical Markets Only

Currently, there is no virtual bidding markets allowed in Chile.

As we discussed iin Section 7.9 Virtual Bidding yields the desired outcome of price convergence between the DAM LMPs and the RTM LMPs. Price convergence is regarded as a benefit to the DAM and RTM markets. It reduces the incentives for market participants to defer their physical resources to the RTM market in expectation of favorable RTM LMPs. The improved stability of the DA market is beneficial from reliability perspectives as well. Extensive experience has shown that absent a visible Virtual Bidding market, market participants may rely on implicit virtual bidding to achieve price convergence. Howeve, such practice can also lead to reliability problems and jeopardize the efficiency of the DAM and RTM markets. Without the revelation of the true economic costs and benefits of physical resources, it is difficult for the system operator to allocate resources efficiently and optimally. In addition, the prices at which market participants bid their physical resources largely depend on their own anticipation of DAM and RTM LMPs, and this introduces uncertainty into the DAM market.

In some cases, the system operator can either over-schedule physical supply in the DAM that has to be sold back in the RTM market, or under-schedule physical supply in the DAM that relies on the procurement in the RUC to balance. These variations decrease the stability of the DAM market, and could undermine reliability of the power grid.

In summary we recommend the design and establishment of a) a formal RTM market as we discussed in previous Sections, b) a two settlement system, c) a Virtual Bidding market, and d) allow financial entities to participate in the virtual bidding market.

8.10 Unit Commitment Costs & Ancillary Services Bidding Format

To solve the day-ahead UCP, the CEN periodically receives information from market agents about their variable operation and maintenance costs, minimum load, no load, and start-up costs, among other technical parameters. This information, along with the availability of power generation and transmission network resources and other technical constraints (such as constraints associated with thermal power plants, hydraulic series, and irrigation agreements), is incorporated in the clearing of the day-ahead market. **In this context, the Chilean DAM already considers unit commitment costs in the clearing of its DAM. Thus, a three-part bid (start-up, no load, and incremental energy) submission by market participants as proposed in the bid-based energy market we propose is not considered a major challenge.**

The UCP solved by the CEN co-optimizes energy and reserves for frequency control and models the system with a high degree of technical details. However, ancillary services (AS) are currently provided by two different alternative mechanisms available in the Chilean system. On the one hand, there is a day-ahead bid-based AS market, where market agents can submit their offers for AS provision. In the particular case of the frequency contol services, these are bid-based only for the wear and tear costs of the power plants. Existing side payments for opportunity costs and out of merit order generation also take place. On the other hand, AS can also be provided by some coordinated firms if they are directly instructed by the CEN to provide those AS. In this case, the coordinated firms are paid based on an audited cost.

The CEN daily decides whether the AS market for the following day is competitive. In case the AS market is competitive, market agents can submit their offers for AS provision for the next day and CEN can clear the day-ahead bid-based AS market. In case the CEN determines the AS market is not competitive, CEN directly instructs specific coordinated firms to offer AS services and pays them based on their audited costs. Otherwise it uses the results of the competitive AS market.

This aspect of the current market should change with the trnstion to a bid-based market architecture because the day-ahead AS market should be fully compatible with day-ahead energy market.

The rule for distributing the costs of the AS provision is a very critical issue under the current Chilean AS market operation. The rule of causality among generators with pro rata of energy withdrawals is usually applied in the current Chilean system. However, if implementing an offer-based AS market that is fully compatible with day-ahead energy market, then there is no need for the CEN to implement such a cost-distribution rule (at least not when the offer-based market is fully implemented). The reason is that in a bid-based system each market agent will make its own estimation of the AS provision cost and the associated opportunity cost.

Specifically as we discussed in detail in Section 7.10 we recommend to continue strengthening the participation of market agents in the Chilean day-ahead bid-based AS market with the following characteristics: a) full offer-based three-part energy offers in the day ahead market including Startup cost, No Load cost and multi segment piecewise linear incremental energy cost, subject to ex-ante market power mitigation, or b) a cost based Startup cost, No Load cost initially for three (3) years and then transitioning to a full offer-based energy architecture for all three components, and c) regulation offers with an hourly capacity cost and a "mileage" cost[27] and reserves offers should with an hourly capacity cost, (an incremental energy price or price curve) as well as resource capability to meet reserve products requirement.

In summary, the already existence of a day-ahead bid-based AS market in Chile is an advantage for the implementation of an offer-based electricity market, although some improvements would be needed.

8.11 Hydro Power Treatment in a Bid-Based Energy Market Architecture

8.11.1 Current Hydro Power Energy Market Treatment

In a bid-based energy market architecture, hydro-power units can submit bids incorporating their own estimations of the opportunity cost of water (future value of water, FVW) to the ISO or, alternatively, the opportunity cost of water can be centrally computed, as it occurs currently in the Chilean market.

Currently, in the Chilean power market, the CEN determines the variable cost of the hydro generators by estimating the shadow price of the water solving a medium-term (2-year-horizon) planning optimization problem. This problem is solved using the software PLP, a domestic version of the Stochastic Dynamic Dual Programming (SDDP) software developed by PSR. The PLP not only incorporates the SDDP optimization (as originally incorporated by PSR), but also incorporates piece-wise linear linearizations of the irrigation agreements that have implications on the dispatch of the water used from the reservoirs.

The medium-term (2-year-horizon) planning optimization problem solved using PLP is a cost minimization problem where the fuel cost is minimized. Specifically, the objective function to be minimized is the sum of all fuel costs minus a term corresponding to the sum of the FVW of each hydro basin multiplied by the volume of water existing in that basin in the last period considered. In this manner, the PLP accounts for the benefits of saving water until the last period.

In the UCP, the Future Cost Function (FCF) of the dammed water is considered in the objective function, which represents the savings in the operating cost, according to the level of the reservoir at the end of the evaluation period. The FCF is obtained via the PLP model. **Subsequently, the shadow price of the energy balance restriction in each**

[27] Panagiotis Andrianesis, Alex Papalexopoulos, "Performance-based Pricing of Frequency Regulation in Electricity Markets," IEEE Transactions on Power Systems, Special Section on Electricity Markets Operation, Vol. 29, no. 1, 2014, pp. 441-449.

reservoir is obtained from the UCP, which is used as a variable cost for the purpose of the economic dispatch in the real-time operation. These computed variable costs and all the variables associated with costs and technical parameters are used to solve the day-ahead UCP by co-optimizing energy and reserves for frequency control, modeling the system with a high degree of details and technical constraints, such as constraints associated with thermal power plants, hydraulic series, and irrigation agreements, and using a 7-day horizon. **As a result of this optimization process, a merit order list is obtained, which is used to update the day-ahead UCP (pre-dispatch) operation.**

In addition, each generator must periodically report fuel availability and its prices or generation forecast in the case of VRE plants. However, there are specific regulations in the Chilean market that allow certain technologies to be assigned a different variable operating cost from the declared cost. Such is the case of units that operate with LNG associated to long-term contracts for which the alternative cost of the energy determined by the CEN is utilized as variable cost. All these activities take place in the day-ahead market (day N - 1).

It is important to mention that, in the Chilean market, the CEN centrally determines the dispatch of all hydro units (based on the FVW); that is, the solution of the Day-Ahead UCP is mandated by the CEN. This means that hydro units do not make any dispatch decisions and have no control of their operation. We view this as sub-optimal.

In this context, there is some ambiguity about the opportunity cost of water used for the compliance of irrigation agreements. Questions such as whether hydro plants should be compensated when obligated to generate due to irrigation agreements or not remain unsolved. In such cases, it is not obvious whether the value of water should be considered zero or should be considered equal to its future value. Furthermore, in a scenario where most thermal plants are decommissioned and only natural gas plants, owned by the same companies that own the reservoirs, remain on-line, the consideration of the impact on electricity generation of long-term water resource management becomes more relevant.

In theory, it is not needed to calculate the FVW too frequently, but only when hidrological conditions are changing. However, one of the problems with this formulation in the PLP software is that the FCF, and consequently the variable cost of hydro units used in the day-ahead UCP, varies significantly depending on the irrigation contracts, the hydrological conditions considered and/or the updated information of the water availability from flows. For this reason, and because the FCF is a very important variable in the central dispatch of the hydro units, the CEN runs this medium-term (2-year-horizon) planning optimization problem in the PLP twice a week, to obtain a clear notion of the degree of variability of the resulting FCF. Then, the CEN uses an updated FCF when solving the day-ahead UCP each day.

The current ownership structure of hydro units in Chile is that private companies own the reservoirs and operate the turbines, but always following strict instructions from the CEN;

that is, hydro firms do not make any strategic decision or self-schedule in the market operation. It is very important to consider the benefits of offering bids for reservoir owners who do not have any control over dispatch decisions.

Alternatively, the market could be designed in such a way to provide the hydro power plant owners some flexibility to submit daily offers incorporating their own estimations of the opportunity cost of water (future value of water, FVW or using the CEN's centrally obtained FCF) and other forecasting information around the reservoir levels produced by the long term PLP cost minimization problem. We propose this market option in the next Section.

8.11.2 Bid-Based Hydro Power Energy Market Treatment

The option to have CEN control the dispatch decisions of hydro power plants (based on the methodology presented in the previous Section) is restrictive and not market-based. Alternatively, CEN could provide some flexibility to hydro power owners and allow them to bid in the market based on the FCF (produced by CEN's PLP long-term cost minimization problem).

In the simplest case many hydro power owners choose to self-schedule the hydro power generaiton consistent with infow and outflow constraints, etc.

Alternatively, in some markets' like in Greece, hydro power owners participate in market with offers within a percentage (say 10%) over or under the monthly value. The purpose of this approach is to account for the inherent uncertainty associated with water availability and to enable hydropower operators to adjust their bidding strategies according to changing water conditions.

In general, hydropower owners typically participate in organized wholesale energy markets with bids which often depend on various factors, including the current water inflow, reservoir levels, market prices, demand forecasts, and regulatory constraints. Here is a general outline of the processes that take place in constructing their bids.

1. **Assessment of Reservoir Levels and Water Inflow**: Hydropowners assess the current levels of water in their reservoirs and predict future water inflows. This information is crucial as it directly impacts their ability to generate electricity in the coming days.
2. **Demand Forecast and Market Price Analysis**: Hydropower owners consider the forecasted demand of electricity and analyze the current market prices. Understanding the demand-supply dynamics helps them determine the most appropriate times to sell their energy at optimal prices.
3. **Operating Constraints and Regulatory Obligations**: Hydropower owners take into account any operating constraints, such as environmental regulations or water

release requirements for maintaining downstream ecosystem health. These constraints can affect the timing and volume of their electricity generation.

4. **Generation Cost Considerations**: Hydropower owners evaluate the costs associated with hydropower generation, including maintenance, operational expenses, and any other relevant expenses. These costs help them determine the minimum acceptable bid prices that cover their expenses and ensure profitability.
5. **Risk Management Strategies:** They may also deploy risk management strategies to mitigate price and operational risks. This could involve hedging strategies, such as forward contracts or financial derivatives, to protect against price fluctuations and uncertainties in the energy market.
6. **Optimization of Bidding Strategy:** Based on the above factors, hydropower owners develop a bidding strategy that optimizes their revenue generation while considering their operational limitations and market conditions. This strategy could involve submitting bids that reflect their available water resources and the prevailing market conditions to maximize their profits.

A typical sophisticated approach for optimizing their energy offers/bids within that range is provided below. Countries with large hydropower capacity such as Norway, Canada, and New Zealand, among others, are known to have sophisticated bidding strategies that consider water availability and its future value.

Typically, hydro power owners usually deploy a Mixed Integer Linear Programming package that maximizes the profit from trading in the co-ordinated energy and ancillary service market while respecting hydro and thermal unit operating constraints and hydro reservoir constraints.

The objective is to maximize the profit from trading in the market, using as input the forecast Locational Marginal Prices (LMPs) for the different commodities and unit constraints, start up costs and operating costs and unit initial conditions, with particular emphasis on co-ordinating the energy and ancillary services bids within unit constraints to maximize the overall profit. In this short-term optimization hydro power owners are taking into account the monthly or hourly water values obtained from the long-term hydro scheduling process and the hydro inflow forecast and storage levels.

The market results include the optimal bids for each unit for energy, spinning reserve, non-spin reserve, regulation up and regulation down.

Hydro unit costs can be represented by applying a monthly water value in $\$/Mm^3$ to each reservoir. These water values are obtained from a long-term hydro-thermal simulation (like the PLP optimization software) and reflect the relative values of water based on the long-term storage strategy.

A water balance equality constraint is applied to each hydro node to ensure that the change in volume of the reservoir is equal to the net flow from upstream reservoirs and inflows less the flow to downstream reservoirs. Travel times between nodes are incorporated by using the flow in the incoming branch from the previous time increment which corresponds to the current time increment less the number of time increments corresponding to the branch travel time. For those branches with travel times, it is also necessary to specify flows in branches prior to the start of the optimization.

Further, irrigation restrictions (i.e., mandatory hydro releases) are modeled as explicit hard constraints in the short-term optimization, which means that the MMS software should use penalty function prices with very high values to respect these hard constraints. This is one of the hard constraints that the software needs to explicitly model. Other are minimum technical generation constraints, Must Run units for established reliability reasons, other existing contracts, etc.

The objective of the short-term optimization model is to maximize the profit from trading in the market where the profit is calculated from the forecast prices times the dispatched generation less the generation dispatch costs which in the case of hydro units are derived from the long-term monthly water values in each storage.

Optimal bids are then determined from the optimal dispatch schedule by selecting a bid level at a configurable tolerance below the forecast clearing prices where the optimal commitment and dispatch has scheduled the unit to be on-line for the interval.

In this option CEN will post the monthly future value of water (FVW) by weekly and hourly based on the CEN's PLP long-term cost minimization problem. Hydro power owners will have the flexibility to start bidding with 5% deviation allowance. This allowance will increase annually by 5%. After the first 3 years only monthly centrally computed value of water need to be published. This option can be implemented from the commencement of the bid-based energy market or after a delay of three (3) years.

8.11.3 Recommendation

1. Continue to deploy the current hydro power energy market treatment where hydro power owners have no control of the dispatch decisions of their plants
2. After a period of three (3) years transition to a bid-based approach to allow hydro power owners to control the dispatch decisions of their plants based on their bids. Start with a deviation allowance of 5% and increase it annually by 5%.

8.12 Summary

The analysis contained in the previous Section of Chapter 8 leads us to the following important conclusions regarding the major changes which will be activated during the implementaiton of the transition to a bid-based energy market. These conclusions, presented in Table 8-1 are based on a detailed evaluation of the current energy market in Chile and the mapping of the important new market element recommendations (presented in Chapter 7) on the existing maret architecture.

The activation of the market changes will be grouped for implementation in difference phases in the subsequent Tasks of this Project. For now all market changes are considered without any consideration of the phased-in approach which will be analyzed and propsoed in the subsequent Tasks in this Project.

Table 8-1: Impact of Proposed market Changes on the Existing Market

	Existing Energy Market	New Bid-based Energy Market
Price Formation (Pay-as-Clear vs Pay-as-Bid)	Yes	Yes
DAM Market Design Architecture (Unified vs. Split DAM by Technology)	Yes	Yes
Long Term Contracts PPAs	Yes	Yes
Financial Markets	Yes	Yes (expanded)
Energy Only Markets with Scarcity Provisions vs. Capacity Adequacy Mechanisms (or Both)	No scarcity	Yes
Capacity Market with ROs	No	Yes
Ex-Ante Market Power Mitigation	No	Yes
Virtual Bidding Market	No	Yes
Unit Commitment Costs Bidding	No	Yes
Hydro Power Bidding	No	Yes
Bid-based market clearing	No	Yes
Two settlement system	No	Yes
Reat-Time Market (RTM)	Adjustment Market	Fully Fledged RTM

Careful evaluation of the conclusions of this Chapter gives credence to the claim that the existing energy market in Chile already has the fundamentals of a solid energy market design including; a) co-ptimization of energy and Ancillary Services, b) a nodal market architecture, c) LMP pricing, d) Long Term Physical and Financial contracts, and e) a technology agnostic auction.

Therefore, the key market changes for a bid-based market architecture include the following implementations: a) bid submission, b) a formal RTM, c) a two settlement system, d) an organized capacity market, e) ex-ante market power mitigation, f) a Virtual Market, g) Unit Commitment cost bidding, h) Hydro power bidding.

The implementation of these changes will be phased-in gradually as we'll analyze in subsequent Tasks.

9. Appendix A: Theoretical Foundation of Marginal Cost Pricing and Uniform Capacity Payment

From the perspective of classical economic theory, the “gold standard” for remuneration of energy cost and the provision of adequate generation capacity in a competitive electricity market is through reliance on uniform energy prices set to marginal cost and scarcity rents. In other words, as it is common for most commodities, producers are paid uniform market prices for the energy they sell and recover their fixed costs for capacity, start-up costs and no-load costs from their inframarginal profits and scarcity rents. Such profits are accrued by the seller whenever the market-clearing price for energy exceeds the marginal cost of production (when generators can also sell ancillary services such as operating reserves, the income from such sales also contributes to the inframarginal profits). The basic principles of such a market are:

- Energy is priced at marginal cost with demand side setting the price during scarcity hours
- Competitive forces drive generation capacity, technology mix and prices toward a long term equilibrium where the total amount and technology mix of generation capacity is optimized with respect to supply and demand preferences for reliability and cost
- Fixed costs of generation capacity at long run equilibrium are exactly covered by inframarginal costs and scarcity rents
- Forward markets and hedging instruments enable parties to manage their risk exposure.

Figure 9-1 bellow illustrates the profits for inframarginal units during hours of the day when the marginal price is set by more expansive units and the scarcity rents that accrue to all the operating units when the marginal price is set by unserved load.

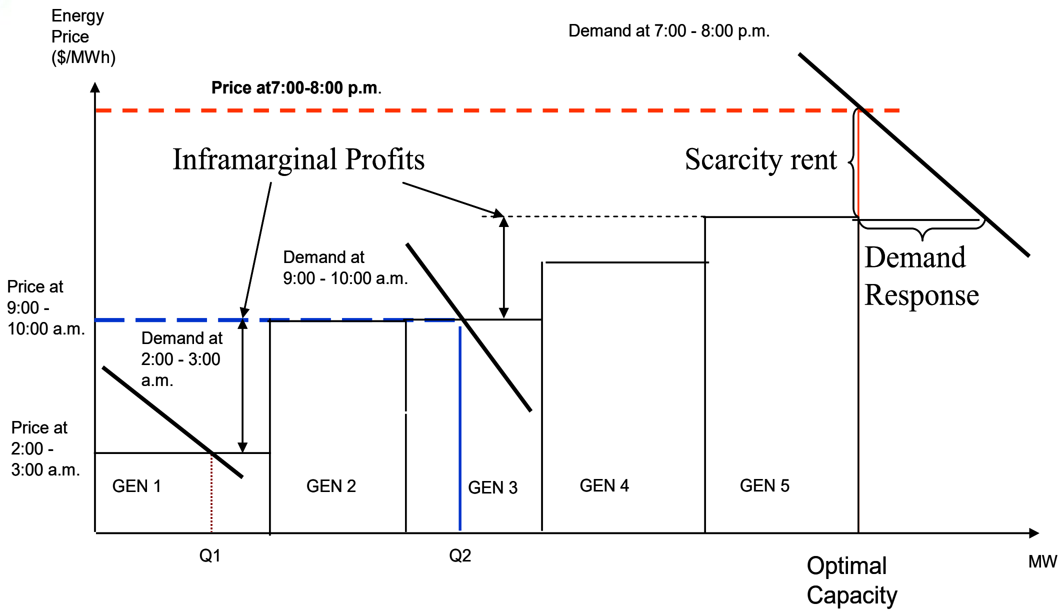


Figure 9-1: Illustration of Inframarginal profits and Scarcity Rent under Marginal Cost Pricing

In theory, when the total amount of generation capacity and the capacity mix are optimal (i.e., minimize total cost of energy, capacity and lost load) then marginal cost pricing including scarcity prices (with prices set to the value of lost load-VOLL) during shortage periods will produce inframarginal profits that exactly cover the fixed costs so that on average sellers break even. We will now demonstrate this result which supports marginal cost pricing in a multi-technology production mix.

Figure 9-2 below illustrates on the left panels, the classical approach to determining the optimal capacity mix to serve a variable load represented by a load duration curve (LDC). Each generation technology is characterized by an amortized fixed cost per MW and a variable (energy) cost per MWh. The load duration curve can be interpreted as a stack of load slices of different duration and each load slice can be assigned to a technology that minimizes the combined fixed and variable cost for serving its duration[28]. Projecting the duration boundaries, within which each technology is optimal, up to the LDC enables us to determine the optimal capacity for each technology and the duration range within which each technology is at the margin under merit order dispatch.

The total cost of each generation unit when the generation portfolio is optimally planned and dispatched can be recovered by charging each load slice the fixed and variable cost corresponding to its assigned technology and paying it to the corresponding generator.

[28] This approach assumes that generation capacity is infinitely divisible and ignores startup costs associated with multiple starts during the period represented by the load duration curve.

So the one MW load slice marked in the upper left panel of Figure A2 costs

F3+C3*(T0+T1+T2+T3) which can be paid directly to a generator of type GN3 serving it[29]. Alternatively, the cost of serving that load slice can be recovered through a marginal cost charge for each MWh produced (i.e., the variable cost of the unit on the margin) supplemented by a capacity payment equal to F1. This will produce a total payment of F1+C1*(T0+T1)+C2*T2+C3*T3 which is shown in Figure A2 to be equal to F3+C3*(T0+T1+T2+T3). However, as illustrated in Figure 9-3, since under optimal capacity planning F1=T0*VOLL, the capacity payment F1 can alternatively be recovered as a scarcity price on energy set to VOLL for the duration T0 during which some load is not served either through demand response or involuntary curtailment. Such an approach is preferable to a capacity payment since it produces incentives for demand response but in case that demand response is technologically infeasible a scarcity price of VOLL combined with any form of load shedding may be regarded as an appropriate proxy to demand response. The above analysis also demonstrates that either capacity payments or scarcity rents for energy should be paid to all generation units uniformly and not just to the peaking units as it is often suggested.

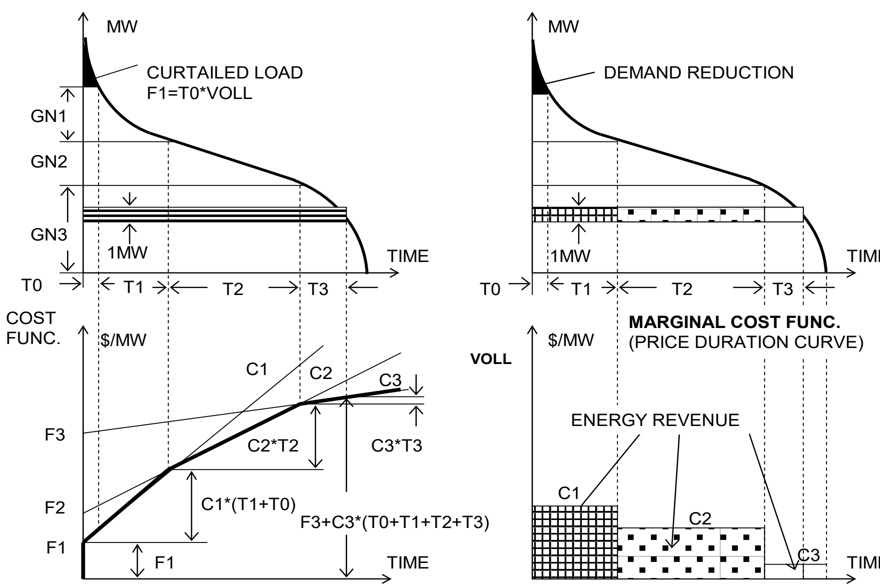


Figure 9-2: Capacity Planning and Cost Recovery

[29] This approach underlies what has been known as the Wright Tariff in used in the US in the early years of the industry and in France circa 1990.

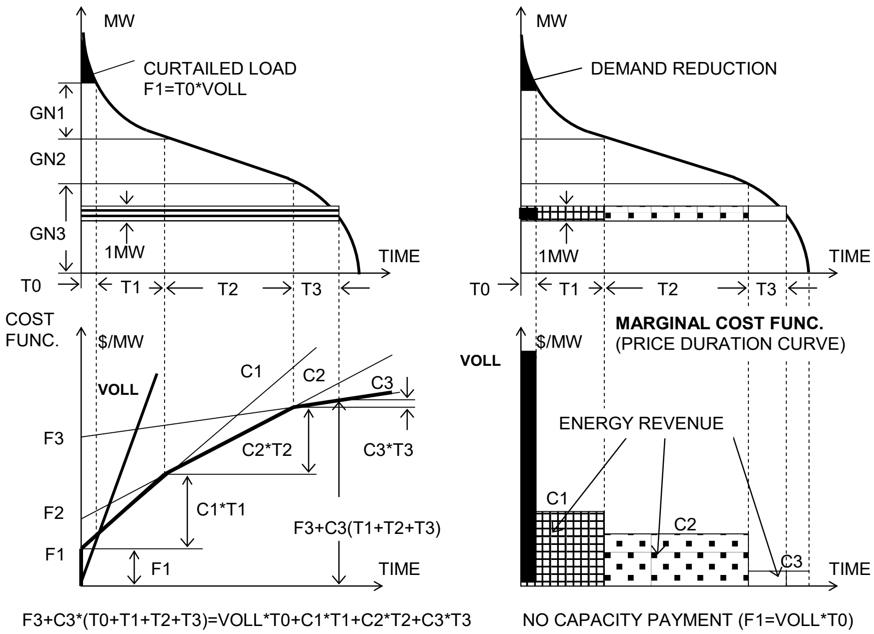


Figure 9-3: Fixed cost recovery through scarcity pricing during load curtailment

In the above analysis we have shown that when generation capacity and technology mix are at their optimum, marginal cost pricing with energy prices set to VOLL during load shedding periods, will lead to full cost recovery by generators. However, with changing fuel costs and uncertain load growth, neither the total capacity nor the capacity mix is likely to be at their optimal levels. Thus some technologies will experience excess profits while other will sustain losses. Such profits and loss scenarios are the correct economic signals that will drive the industry toward the desired equilibrium through new entry (of the profitable technologies) and retirement (of the loosing technologies). Regulatory intervention aimed at rectifying such temporary windfalls or deficits through differential payments for capacity availability or subsidies, is misguided since it mutes the economic signals for entry and exit that a market based system is designed to provide.

Following is a simple example that illustrates the process of exit and entry that a market based system induces. Figure 9-4 below illustrates a supply function for a system with two types of generation units labeled G1 and G2. The following, Table 8-1, summarize the relevant data for the supply system

Table 9-1: Data Table

Generator type	No of units	Unit capacity	Fixed cost	Marginal cost
G1	50	80MW	$926,400/Month	$15/MWh

G2	100	60MW	$288,000/Month	$25/MWh

The demand is characterized by two demand functions for peak and off hours (P=price, Q=Quantity)

Off-Peak: 420 Hrs./Month P=30-Q/1000   Peak: 300 Hrs./Month P=50-Q/1000

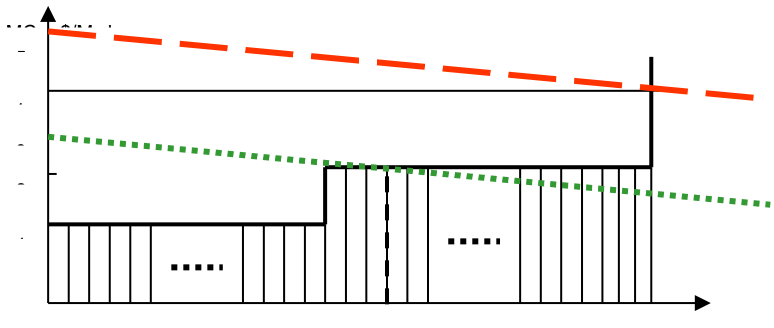


Figure 9-4: Short run equilibrium

In the short run equilibrium the price during off peak is $25/MWh with demand at 5000MW while during the peak hours total supply of 10000 MW is exhausted at a scarcity price of $40/MWh The corresponding net income for G1 and G2 generators is as follows:

G1: 80*[(40-15)*300+(25-15)*420)] - 926,400 = $9,600/Month (excess profit)

G2: 60*(40-25)*300 - 288,000 = ($18,000/Month) (deficit)

Under such circumstances one expects entry by G1 generators and retirements of G2 generators that are loosing money. This process will reach a long run equilibrium with 2000MW of new G1 capacity and exit of 3000MW of G2 capacity. Figure 9-5 - Long run equilibrium illustrates the new capacity mix and resulting long run equilibrium.

Peak price $41/MWh Off-Peak price $24/MWh

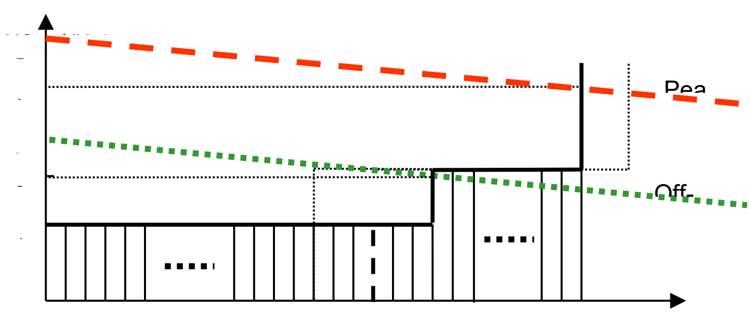


Figure 9-5: Long run equilibrium

The new equilibrium will have an off-peak scarcity price of $24/MWh with corresponding demand of $6000MW and peak scarcity price of $41/MW with all 9000MW capacity being used. The resulting net income for both generator types is summarized below.

G1: 80*[(41-15)*300+(24-15)*420] – 926,400 = 0 Breakeven

G2: 60*(41-25)*420 – 288,000 = 0 Breakeven

Thus, the capacity mix has reached a long run equilibrium where all generators break even and any addition capacity will loose money.

In reality, a long-run equilibrium as described above, is achieved, if at all, through a series of “boom and bust” cycles of capacity expansion. A capacity shortage drives energy prices up, resulting in excess profits that attract new investment in capacity, which in turn drives energy prices down. Likewise, excess capacity results in low energy prices that reduce inframarginal profits to the point that some generators are not able to cover their fixed capacity costs. That results in early retirement of older or less efficient units and a reduction in capacity, which in turn raises energy prices. A similar process takes place when the generation mix is suboptimal, resulting in adjustments in the generation mix through retirements of some plants and construction of new ones.

There are several unique characteristics of electricity markets that differentiate them from other commodity markets and induce higher volatility of spot energy prices. These include the “instant perishability” and non-storability of the product, the steep “hockey stick” shape of the supply function (which is a direct consequence of the optimal technology mix and the load-duration profile) and the high level of uncertainty about demand and available supplies. Such uncertainty is induced by weather that affects demand and availability of inexpensive hydro resources, and by outages of plants and transmission facilities. Price volatility alone, however, should not impede the efficiency of energy markets as a mechanism for inducing adequate investment in generation. On the contrary, high price volatility will, in theory, motivate risk averse buyers and sellers to enter into long-term forward contracts that will hedge their risk exposure. Such contracts enable investment in new generation when needed. High spot prices and anticipated shortages increase forward prices and stimulate entry. In such an ideal setup, suppliers and buyers are free to select the level of risk they want to assume and they will use financial hedges and contractual arrangements that allocate risk efficiently.

The natural inclination of market designers with an economic background is to favor energy-only markets. Such markets have been established in Australia, New Zealand, Alberta, Canada, ERCOT (Texas). This approach functions adequately in systems with abundant reserve capacity – like ERCOT – and in systems where extremely high spot prices reflecting temporary scarcity are tolerated – as in Australia where spot prices can and have risen occasionally to $8,000/MWh and in Texas where energy bids can and have been as high as $9000/MWh.

10. Appendix B: When is Pay-As-Bid Justified

As we have discussed in Chapter 7 most market designers agree that at least in the case of a homogeneous commodity, like electricity, a uniform clearing price auction is superior to a pay-as-bid rule.

As we discussed in uniform price auctions bidders have an incentive to reveal their true cost and therefore, the dispatch will be efficient. However, one must recognize that if demand is uncertain a uniform clearing price will reflect that entire uncertainty and will be more volatile than the average procurement price in a pay-as-bid auction. The lower volatility results from the fact that in a pay-as-bid setting the markup function which balances bidders' desire to get a higher price against their fear of not being selected, tends to absorb part of the demand uncertainty. In principle the ISO conducting the auction should be risk neutral and not care about price volatility. ECCO argues that suppressing price volatility is undesirable since it will also suppress demand response when possible. In other words, price volatility, up to a point is desirable. deploy Nevertheless, reduced price volatility is one of the arguments used, for instance by the UK NETA proponents, to advocate the pay-as-bid settlement rule in the RTBM market.

In actuality in the power markets the auctioned products are not completely homogeneous. Consequently, winner determination is often based on attributes such as location (in the US and other Nodal LMP markets), ramp rate, reactive power capability, etc. that are not explicitly priced in the auction. In some cases, such heterogeneity leads to product fragmentation where the distinct products are procured through separate auctions conducted in a coordinated fashion in series, in parallel or simultaneously. Designing such auctions so as to take into consideration the partial substitutability among the different products is challenging. Furthermore, such fragmentation reduces the liquidity in each of the separate auctions to the point where some of the underlying assumptions favoring the uniform clearing price approach may no longer be valid.

In some cases, reserve markets are good examples of such market fragmentation where reserve capacity is categorized based on response time (e.g. Regulation, Spin, Non-Spin, etc.) and auctioned as separate products in sequential (in the past) or currently in simultaneous auctions. In the US markets where reserve requirements are also location based, the reserve products are further categorized by location. In the CAISO for instance, the reserves are identified according to seven geographic categories resulting in 42 distinct reserve products. Under these circumstances the assumption of liquidity, homogeneity may no longer be valid and one must reexamine the wisdom of having a uniform clearing price auction for each separate product as opposed to a pay-as bid approach. Even in this case the reserve products are settled with marginal pricing in all US markets.

In summary, if market and operational conditions result in auctioned products that are non-homogeneous that necessitate a high degree of product fragmentation, a pay-as-bid settlement approach with optimized assignment based on requirements and multi-attribute specifications of the tender, may be justified (even rarely).

