PJM Reserve Markets: Operating Reserve Demand Curve Enhancements

William W. Hogan, Harvard University

Susan L. Pope, FTI Consulting Inc.

March 21, 2019

Table of Contents

PJM Reserve Markets: Operating Reserve Demand Curve Enhancements

William W. Hogan[i], Harvard University

Susan L. Pope, FTI Consulting Inc.

March 21, 2019

Overview

Organized electricity markets require attention to the details of market design, i.e., the tariff rules and software models used to determine the quantities clearing in the markets and the prices used for settlements. The requirement for open and non-discriminatory access to the transmission grid makes it essential to get the prices right to provide incentives for coordinated but largely voluntary actions by market participants that are economically efficient and reinforce reliable grid operations. The basic market design for transmission access and energy market clearing in place in PJM addresses this fundamental challenge of non-discriminatory open access. Refinements for pricing of operating reserves are an important opportunity to enhance the PJM market design to better value reserves and improve pricing under scarcity conditions.

This paper addresses the underlying economic efficiency elements of PJM's proposed Operating Reserve Demand Curve reforms that provide a basis for just and reasonable rates. As summarized here and discussed further by PJM[1], the existing PJM operating reserve market design fails to support economic efficiency in a number of ways. Elements of the current operating reserve market construct in PJM were developed in stages in prior years and, as a result, are not explicitly founded in a principled theory connecting the composite operating reserve market design to sustainable, just and reasonable rates. In addition, changing circumstances, such as increasing variability in net load due to intermittent resources, have revealed problems where the prices for different types of reserves do not reflect their underlying economic values or provide the needed support for efficient, reliable operation. The existing operating reserve markets in PJM can and should be replaced by an enhanced design that better meets the standard of just and reasonable rates.

---

1 Affidavits of Adam Keech, Christopher Pilong and Patricio Rocha Garrido contained as Attachments D, E, and F to the Reserve Market Pricing Reform Package submitted by PJM to the Federal Energy Regulatory Commission on March 29, 2019

The PJM proposal is to consolidate and redefine the types of operating reserves it schedules in its day-ahead and real-time markets and to introduce Operating Reserve Demand Curve (ORDC) reforms that reflects the value of incremental operating reserves of different types and in different locations. This will enable energy and reserve pricing that better reflect the value of reserves under different operating conditions and will also support consistency in operating reserve pricing between PJM's day-ahead and real-time markets. Implementation of the desired enhancements requires policy choices and modeling decisions to enable a workable representation of the demand for operating reserves.

The discussion in this paper provides a theoretical basis for the formulation of operating reserve demand curves and the co-optimization of energy and reserve markets given these reserve demand curves. The resulting clearing prices for reserves and energy are connected to economic efficiency because they are derived from a model of co-optimized economic dispatch. The initial model presented, for energy and a single type of reserves, is extended to incorporate further features relevant to the scheduling of operating reserves, including multiple types of reserves, the activation of emergency response, locational operating reserve requirements, and day-ahead markets.

PJM's proposed design choices for its operating reserve demand curves have been made with purpose to align with the underlying theory to the extent possible and implement a market design that better meets the standard of just and reasonable rates. PJM's ORDC reform proposal is an important step toward improving the efficiency of PJM's electricity markets and achieving just and reasonable rates relative to the current problematic and unjust operating reserve market design.

PJM Operating Reserve Demand Curve Proposal

The existing PJM operating reserves design includes two classes: Tier 1 and Tier 2 reserves. Tier 1 reserves have an assumed 10-minute response time provided by flexible generation resources available after the real-time economic dispatch. This includes available capacity on partially dispatched synchronized resources. Put simply, Tier 1 is the amount of available capacity that is assumed to be available in ten minutes from a resource that is online for the purposes of producing energy. These reserves are not formally procured and have been treated as being free. By contrast, Tier 2 consists of generating resources whose real-time dispatch is reduced from their economic set point, so the Tier 2 resources incur opportunity costs in providing reserves. Tier 2 resources are paid these opportunity costs during the same dispatches when Tier 1 resources provide the same reserve product for free.

The operational requirements and payments for Tier 1 and Tier 2 reserves are not consistent, leading to unjust and unreasonable prices. Tier 1 resources do not have an obligation to respond and are not paid the market-clearing price of reserves set by the opportunity cost of the marginal Tier 2 resource. By contrast, Tier 2 resources are obligated to respond if activated and are paid a market-clearing opportunity cost price for reserves regardless of deployment. As a result of this inconsistency in pricing and operational requirements, PJM has found the average Tier 1 response rate has been 60.1%, whereas the Tier 2 average response rate has been 87.6%. The inconsistency in the prices of the reserves is not just and reasonable, and the inconsistency in the performance of the reserves is a matter of concern for reliable system operation.

A further problem is that PJM's current pricing structure for operating reserves does not derive from an assessment of the value of operating reserves. There is a minimum reserve requirement (MRR) derived from North American Electric Reliability Corporation (NERC) standards, and PJM sets "penalty" factors used to set reserve prices when reserves would fall below this level absent various unpriced operator actions. The current maximum penalty factor is $850/MWh, which is below the costs of many of the actions operators must take to avoid violating the MRR. Because the penalty factor is lower than the cost of these actions, it will not provide a market price signal to call forth voluntary provision of additional reserves at prices lower than the cost of the mandatory emergency actions. This results in operator interventions that might not be necessary with a higher penalty price. This impacts settlements across the energy and reserves markets, because when interventions occur they can deploy more reserves than are necessary to resolve a shortage (e.g., by activation of a block of demand response), which can suppress energy and reserve prices and increase uplift costs. The $850/MWh maximum penalty factor is also far less than the Value of Lost Load (VOLL) that would be relevant in determining the marginal value of additional reserves in the event of involuntary load curtailment that might be required to avoid cascading system failures.

Implementation of the existing design, especially given changing operating conditions, yields energy and operating reserve prices in PJM that do not align with economic principles. The prices of incremental reserves and energy can deviate from incremental cost and are not consistent with the implications of first principles for determining the value of operating reserves when supply is constrained. These inconsistencies with economic principles support the conclusion that the existing system for operating reserves and their associated pricing is not just and reasonable.

The main elements of the PJM reform proposal include consolidation, by removing the Tier 1 synchronized reserve product, and replacing the Tier 2 reserves with a new structure of reserve requirements to align better with operating requirements. The proposal is for real-time markets for 10-minute synchronized reserves, primary (non-synchronized) operating reserves, and 30-minute reserves, and corresponding day-ahead markets for the same products: 10-minute synchronized and primary reserve, as well as 30-minute reserves.

These reserve product markets will include an enhanced ORDC anchored by an increase of the maximum "penalty" factor from $850/MWh to $2,000/MWh to improve scarcity pricing and replace operator intervention with market response to higher prices. The design proposal includes a cascaded hierarchy in the requirements for reserves with different response times, and this hierarchy flows through into their relative prices. For example, synchronized reserves can meet the primary requirement, and the 30-minute requirement. Hence, as discussed in the Appendix, the price for synchronized reserves is the sum of the clearing prices for all three of these reserves types and the levels that can be reached are consistent with reasonable estimates of the range of VOLLs.

The PJM ORDC reform addresses the value of incremental operating reserves that are both below and above the MRR. The value of reserves of different amounts is the probability that, with a given level of scheduled reserves, PJM's actual reserves in real-time will fall below the MRR and require some form of operator intervention. PJM's proposed Probability of Reserves Falling Below the Minimum Reserve Requirement (PBMRR) is similar to the familiar loss of load probability (Lolp) and produces a downward sloping demand curve for additional operating reserves starting from the MRR quantity, at which the price is equal to the penalty factor. The penalty factor anchoring the ORDC is the maximum price paid for incremental reserves when PJM is taking emergency action. The downward sloping PBMRR recognizes that the scarcity value of reserve capacity does not suddenly drop to zero just beyond the MRR, as PJM's ORDC effectively does now. The ORDC reform recognizes the value of scheduled reserves greater than the MRR, which in practice contribute to PJM's reserves not falling below the MRR during actual real-time operation.

When combined with simultaneous optimization of the energy and reserve dispatch, the ORDCs proposed by PJM enable internally consistent prices for energy and reserves at the margin. The improvements will avoid PJM taking emergency action when it faces uncertainty about the actual response of (Tier 1) reserves, because its prices are too low to call forth voluntary supplies of economic reserves, or because of an inability to co-optimize and reconfigure how available resources are assigned to provide reserves versus energy in real-time.

The existing PJM market design includes the basic elements of an ORDC through two levels of penalty factor when operating reserves fall to certain reserve thresholds. But this design is inadequate. It does not recognize the true value of reserves along a continuum derived from the probabilistic representation of the expected need for additional reserves in real-time. Additionally, the maximum penalty price falls short of the cost of actions the PJM system operator currently takes in order to restore operating reserves. This leaves market participants exposed to the impact of out of market decisions which undermines confidence by PJM market participants that the prices in the markets will be the result of competitive market forces. PJM's market design must be enhanced to ensure reserves are priced consistently with their value, and consistently across energy and reserve products.

The details of the ORDC reform implementation are important, especially those that connect to the fundamentals of efficient electricity market design. The proposed ORDC structure builds on a foundation of electricity market design principles and adopts workable implementations that conform to its energy dispatch models. By building internally consistent connections between its reserves and energy markets, which are founded in system operation and flow through into pricing, the PJM ORDC design will greatly improve the justness and reasonableness of its energy and operating reserve prices and provide the framework for future improvements.

Fundamentals of Electricity Market Design

The electricity system presents three unusual characteristics that have a profound effect on any electricity market design. First, unlike other commodities, electricity supply and demand must maintain constant balance under rapidly changing conditions. Second, power flows along parallel paths through the transmission system as dictated by physics and the locations of generation and load. Third, the speed of response of the physical system necessitates a variety of reliability constraints to ensure that secure power balance continues under sudden changes in the system conditions.

One implication of these characteristics is that power dispatch requires explicit coordination. With current technology it is not possible to maintain security or achieve efficiency with a purely decentralized system. Hence, a ubiquitous feature of electricity systems is the need for and presence of a system operator, such as PJM.

To maintain secure operation, the system operator must enforce the various reliability constraints to ensure that in the event of a contingency the system response will be such that the physical power flows will remain within the specified limits. This requires maintaining reserves of generation and load-response, while monitoring pre-and-post contingency power flows.

To achieve economic efficiency, the coordinated dispatch must maximize social welfare, i.e., the benefits of load minus the costs of generation, subject to the various reliability constraints. This is a challenging problem, and over the decades the industry has steadily advanced its capabilities to represent and optimize over the inter-related physics, economics, and policy requirements (Ott, 2003).

In many regions, coordinated electricity markets have been introduced to reduce or eliminate the role of monopoly operation, procurement and resource planning in achieving the benefits of coordinated dispatch. In these regions, such as PJM, energy and ancillary services markets are designed to work together to provide price signals for operations and help support investment in an electricity system configuration capable of delivering intended levels of reliability to customers at an economically efficient cost.

The accuracy of the real-time price signal, in terms of sending a high price when the supply (i.e., energy plus reserves) is constrained, or a low price during times of abundant supply, is a lynchpin of efficient electricity market design. Real-time dispatch and pricing corresponding closely to the physical requirements of reliable system operation reduce the need to devise alternative, non-market mechanisms to guarantee that the Independent System Operator (ISO) has the needed resources at the right times. It is essential to continue to improve the real-time coordinated dispatch and associated settlements by providing strong real-time price signals for resources that are most economic, because they provide energy and reserves when and where they are most needed and most valuable in real-time.

Security-Constrained Economic Dispatch with Locational Prices

The electricity market design has successfully confronted the challenges of non-discriminatory open access and the simultaneous need for coordination of power flows on the grid. It requires explicit modeling of physical operating constraints and specification of the welfare-maximizing economic objective function, yielding both the economic dispatch (for energy and operating reserve schedules) and prices supportive of the desired optimized dispatch. The market cannot solve the problem of market design. The system operators and regulatory oversight authorities have the problem of making and improving the design choices to build this model. Application of the basic theory of economic dispatch, and the experience of costly mistakes, reinforce the same conclusion: the foundation of successful market design is bid-based, security-constrained, economic dispatch with locational marginal prices and financial transmission rights (Hogan, 2002). This is the basic framework in PJM.

A criterion imposed on the design of markets for energy and operating reserves is that the dispatch, schedules and prices should be ruled by the practices for secure system operation and the decisions made by the system operator. The bids and offers come from load and generation and present the system operator with a set of choices for dispatch and schedules (with different prices) to meet the many constraints, including the transmission constraints. Differences in the marginal impact on the cost of losses and the marginal impact on the cost of respecting physical limits cause the marginal value of generation and marginal cost of serving an increment of load to differ across locations. Efficient dispatch produces the marginal costs of losses and the marginal cost of respecting physical constraints; the marginal values of generation supply and load consumption are identically the same thing as the locational prices. Contracts in forward markets can be settled at these locational prices and financial transmission rights provide a mechanism for hedging the locational differences in prices (Hogan, 1992).

An important property of this market design, based on economic optimization constrained by the physical constraints imposed for secure operation of the transmission system, is that the resulting locational prices support the actual economic dispatch. This means that, with limited exceptions, all that is required to incent suppliers and loads to follow their dispatch signal is to pay or charge them the market-clearing price. Given certain usual simplifying assumptions about the underlying cost structure, which includes convex total cost offers, generators and loads who acted as price-takers would have no incentive to deviate from the economic dispatch. This basic model of security-constrained economic dispatch and locational pricing is the only pricing model that is consistent both with efficient dispatch and with the principles of open access and non-discrimination. Any other pricing model would create incentives to deviate from the dispatch and would require some departure from open access or non-discrimination in order to maintain dispatch efficiency.

This basic outline summarizes the key reasons for the importance of getting the prices right. Now, after two decades of experience, all the organized markets in the United States build on these foundations in their market design.

PJM has been a pioneer and has moved steadily to adopt industry best practice for dispatch and pricing in their real-time energy market in order to harness market forces to support minute-by-minute reliability. The PJM implementation of Locational Marginal Pricing (LMP) was a central step in recognizing that economically efficient prices supportive of minute-by-minute reliability must vary to reflect locational differences in the value of energy when there are active transmission constraints. Similarly, the ORDC reform proposal is needed to recognize that efficient energy and reserve prices supportive of minute-by-minute reliability must consistently reflect differences in the value of operating reserves with different responsive lead-times and located in different locations.

Co-Optimization of Energy and Reserve Markets

As an improvement on the basic market design, co-optimized dispatch of energy and reserves is recognized as the market design standard in order to reliably and economically position resources of all types, including generation, load, imports and exports. PJM utilizes co-optimization and it is central to their ORDC reform proposal. As intermittent energy production increases, it will increase the frequency with which PJM will need to direct resources to ramp up or down to balance the system's net load and maintain required levels of reserves. PJM uses a co-optimized security-constrained economic dispatch (SCED) to look at all of the bids and offers for energy and operating reserves at the same time, while also taking into account all operational constraints (such as ramping constraints, transmission system contingencies, etc.) and issue simultaneous real-time dispatch instructions for energy and schedules for operating reserves. The objective function for the co-optimized SCED is the welfare maximization defining economic efficiency.

As intermittent production increases, co-optimized dispatch of energy and operating reserves, with ORDCs reflecting the economic value of incremental reserves, automatically will efficiently adjust the output of the resources that can increase or decrease production at least cost to balance load, taking into account ramping constraints, reserve requirements and possible contingencies. Co-optimized real-time dispatch of energy and operating reserves markets, coupled with dispatch interval settlements and improvements to the ORDC, mean suppliers will be paid a price that better reflects the value of the services that they are providing. Under this market design, resources make bids and offers into the real-time dispatch based on their estimated costs of operation. The co-optimization dispatches and/or schedules their capacity for its highest-value use (whether reserves or energy, in particular) from the standpoint of maximizing social welfare. The supplier will be paid the market-clearing price for its real-time supply of energy and operating reserves, not its offer or bid price, thus supporting the least-cost dispatch. Importantly, co-optimization produces consistent prices for energy and reserves. In particular, the demand for reserves represented by the ORDC provides a market-clearing scarcity price for reserves that also enters into the formation of the clearing-price for energy.

Economic efficiency improves because the co-optimized SCED with ORDC provides the incentive for bids and offers to enable the least-cost combination of resources to be used to provide energy and ancillary services (such as operating reserves) in real-time, and because improved prices for these products will more efficiently incent market entry and exit.

Day-Ahead Markets

The basic market design principles described for real-time also affect any coordinated forward markets operated by the system operator, such as the day-ahead market in PJM. The forward markets create financial contracts for settlement against real-time prices. The benefits include hedging for supply, demand and transmission constraints, and the opportunity for a centralized look-ahead to insure the unit commitment will ensure real-time reliability. In addition, a day-ahead market allows for broader participation by financial entities, known as virtual bidders, who do not plan physical delivery in real-time but provide added liquidity and therefore competition in the day-ahead markets (Hogan, 2016).

The complete design of electricity markets based on security-constrained economic dispatch and locational prices is a challenging task. If the industry had not already developed many of the essential economic dispatch tools, electricity restructuring as we know it may not have been possible. The basic model focused on pricing real power flows, and there have been many simplifications required to achieve workable and auditable approximations for the dispatch and associated prices consistent with the many complexities of electricity system operation. For example, the treatment of reactive power flows, ancillary services such as black start and frequency response, unit commitment and other dynamic constraints, is a partial list of the details that must be considered by system operators but which are treated with approximations in the economic optimization and pricing model. The process has been to work steadily to improve price formation, and to address as needed the related problems that arise with the successful implementation of the necessarily imperfect approximations.

Scarcity Pricing

Scarcity pricing is an important example where an extension of the basic market design is needed to address a pricing problem arising with increasing frequency. There is a need to enhance the basic market design to provide prices that better reflect the cost of scarcity when the system is pressing against its capacity limits. This section focuses on the underlying principles of scarcity pricing and how the objective of implementing scarcity pricing drives the proposed PJM enhancements of the characterization of the short-term value of operating reserves through the ORDCs.

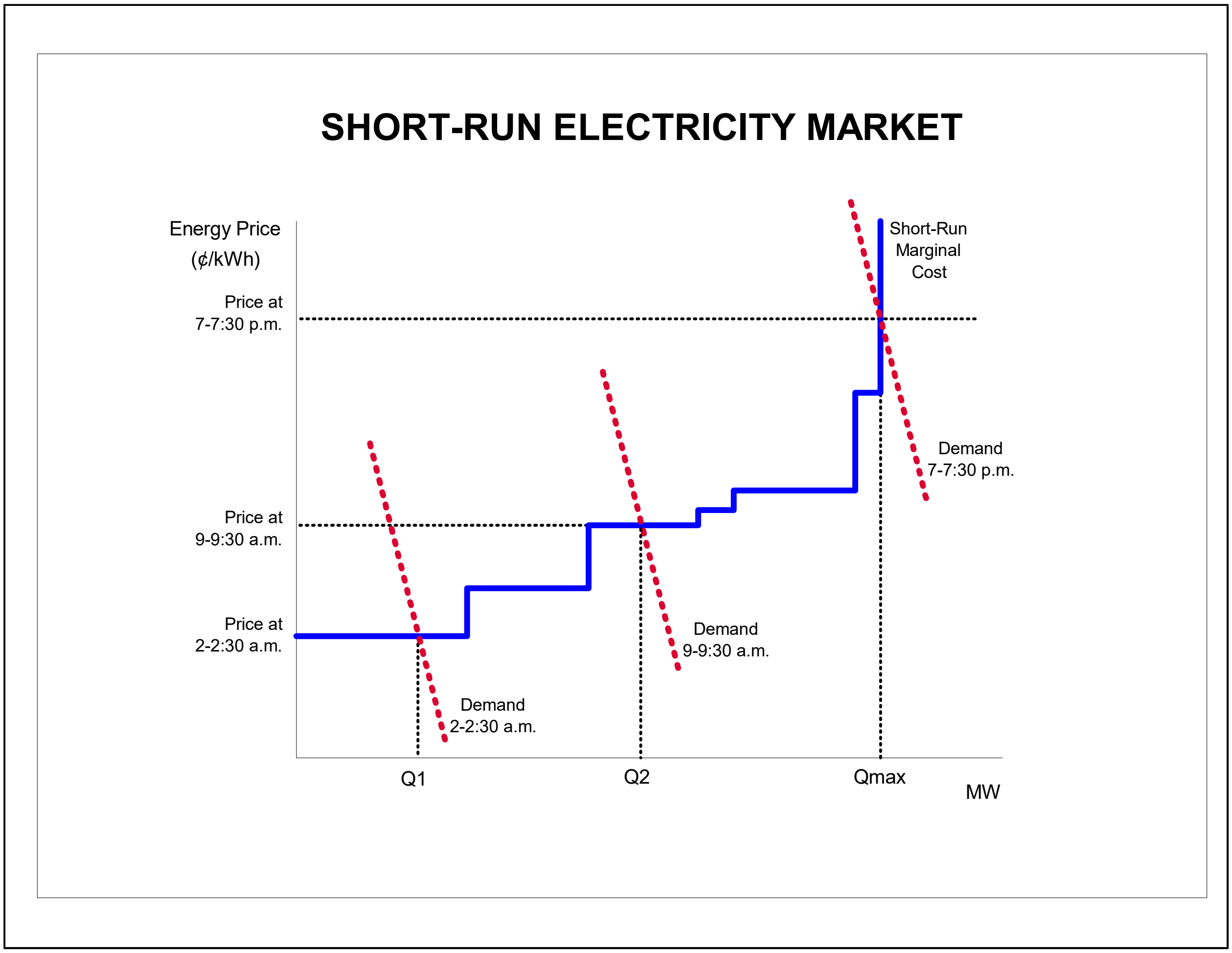
Early on in the development of workable market designs it was clear that prices did not fully reflect the marginal value of generation during scarce conditions (Joskow, 2008), and this is a continuing concern (Joskow, 2019). This impact was captured under the label of the "missing money" (Shanker, 2003). This problem created an early debate about the need to improve the energy and reserves price formation versus the approach of creating another mechanism for providing incentives to meet the investment requirements of resource adequacy. An approach adopted in PJM and elsewhere was to create a related but separate capacity market focused on long-term incentives and implemented under the Reliability Pricing Model (RPM) and the related Regional Transmission Expansion Plan (RTEP).

To be clear, short-term operating conditions and long-term investment requirements are related, but notwithstanding the existing long-term adequacy requirements of the reliability rules, something more is needed in order to connect the solution to the missing-money requirement to short-term dispatch operations and actions that maintain minute-by-minute reliability. Although important, the capacity market with RPM and RTEP is not the focus of the present discussion.

In the current PJM framework, the long-term resource adequacy planning requirements are separate from the short-term requirements for operating reserves and other ancillary services. The long-term resource adequacy market innovations do not in themselves deal with the problem of improving the representation of scarcity – and its pricing – in the real-time and day-ahead markets. For example, the existing capacity performance mechanism does provide a scarcity signal for generators but has no corresponding impact on demand or other sources that could substitute for generation at the margin and does not replace the need for the current ORDC or its reform. In order to incent real-time provision of reserves and energy when and where there is scarcity, the value of this additional capacity needs to be included in the energy and reserve prices paid to *all* resources able to respond and reduce the scarcity and paid by *all* loads consuming during these particular intervals of the dispatch. Better scarcity pricing – applied in specific locations and during specific times – will call forth supply and responsive load when and where it is needed, and through the two-settlement system, those who do not respond economically will bear the cost or opportunity cost. Furthermore, the prices for scarcity will be directly tied to the locational and temporal marginal value or marginal cost as determined through the SCED optimization.

The basis for defining consistent scarcity pricing begins with an elementary static model of energy supply and demand. Consider in the first instance a characterization of a single location with energy bids for demand and offers for generation at the variable cost of production. As shown in Figure 1, taken from materials prepared more than two decades ago, market equilibrium occurs at the efficient solution that balances generator supply and load demand and sets the market-clearing price (Hogan, 1993).

Figure 1



During periods of low demand, the equilibrium price is low and the market-clearing price equates the marginal benefit of load with the marginal cost of generation and is equal to the variable cost of production for the marginal plant. As demand increases, represented by a shift in the demand curve to the right, the equilibrium defines a higher market-clearing price, where in the illustration the price that equates marginal benefits and marginal cost is again at the variable cost of generation. Finally, in the highest price case in the illustration, the efficient solution balances supply and demand, but the pricing condition is somewhat different. The efficient price is equal to the marginal benefit of load, but the efficient price is higher than the variable cost of the most expensive generator.

The difference between the efficient price based on marginal conditions in the highest price case, and the variable cost of the last generator, leads to a definition of the scarcity price. The scarcity price appears whenever the available capacity is constrained and at that dispatch level prices rise to clear the market without an accompanying increase in supply. A similar scarcity component would appear on any of the vertical steps of the representation of the generator supply curve.

The illustration is not drawn to scale, and the scarcity component could be very large when the system is operating at its capacity limits. The assumption in the example that supply and demand for energy are at a single location is not important here. The same analysis generalizes to a network with locational marginal pricing to reflect generation costs, marginal losses, and congestion differing across locations. Under the basic pricing principles, scarcity pricing would arise in a natural way and could be an important part of market-clearing prices and settlements to pay for energy.

This basic model of scarcity pricing derives from the principles of economic efficiency and provides a number of benefits. Higher prices during critical periods facilitates dispatch response by transmission connected loads and generation (i.e., resources settled at the ISO's real-time locational prices) when it is most needed. This supports both economic efficiency and system reliability. Additionally, better scarcity pricing would reduce the "missing money" in energy markets and the corresponding challenges of operating good capacity markets and ensuring long-term resource adequacy. All resources providing energy or reserves during scarce conditions would be paid the market-clearing price for the value of their capacity during the scarce intervals. So, for example, instead of PJM paying uplift to resources dispatched at prices of up to $2000 to alleviate a reserve shortage and recovering the cost through a socialized charge to loads, all resources would be paid the market price for the energy or reserves they provide during scarce intervals, which would be higher due to including a scarcity component.

Better scarcity pricing would improve incentives for dispatchable resources to be available to complement increased supply from intermittent renewable energy. This is because flexible resources needed to balance changing renewable generation would see better price incentives to respond to the changing system needs. And better scarcity pricing would interact with transmission congestion by providing scarcity signals on a locational basis, providing better signals for transmission investment by clearly indicating through significant price separation where transmission investments would be cost-beneficial.

This approach to scarcity pricing would already be in place if the actual market matched the theoretical assumption of complete demand bidding to represent the marginal value of increased load. However, for a variety of reasons, demand bidding has been the exception rather than the rule. In part, the absence of demand bidding is a symptom of the imperfections in the determination of scarcity prices. If prices do not reflect scarcity conditions, there is not as much reason to invest in the flexibility to dispatch load, creating the self-fulfilling prophecy that scarcity will not be well-reflected in the implementation of market pricing. Without significant demand-side bidding, there is currently no market mechanism that will increase prices to indicate exactly where and when there is a potentially very high value for additional capacity.

Better scarcity pricing of energy would go beyond the improvements described above by addressing a closely related challenge in pricing operating reserves, which has received less attention. In fact, the simple energy demand and supply model outlined to introduce the importance of scarcity pricing is silent on the implications for operating reserves. Better representation of scarcity pricing through ORDC reform would complement increased participation of demand bidding in the dispatch. While the absence of demand bidding makes better scarcity pricing models more important, there is no conflict with improving on both dimensions.

Given the development of the basic energy market design in PJM, better scarcity pricing should be a high priority for enhancing the market design so as to improve the efficiency of the price signal during conditions of scarcity, stimulate more active participation by the demand side of the market, and stimulate efficient investments in supply side resources that can respond quickly to provide more reserves or energy. PJM's current market design does not fully price scarcity at all times because it only values reserves at the NERC-driven minimum reserve requirement plus one step. But in reality, operator actions (with costs greater than the maximum reserve penalty price) are taken to maintain reserves both before the economic dispatch model is run, through biasing, and after, through out of market actions.[2] These actions skew and depress market price signals. PJM's ORDC reform proposal is designed to better value scarcity and it is, therefore, a much-improved approach to ensure energy and reserves are appropriately valued.

Addressing scarcity through ORDCs requires an expanded discussion that goes beyond the basic static model of energy supply and demand to address multiple reserve types, locations and emergency actions.

Operating Reserve Demand Curve

Operating reserves are an essential element of security-constrained, economic-dispatch although the simple static energy model pushes them into the background. Operating reserves are required in the real dynamic system that the operator must coordinate in order to ensure that the system reliability is maintained in the event of material changes in conditions. A complete dynamic model for the dispatch of energy and operating reserves would incorporate multiple look-ahead periods with continuous adjustment of the actual dispatch and operating reserve schedules to match changing conditions and respect physical constraints such as ramping and transmission limits. Alternatively, if all generation and load could respond instantaneously over their full capacity, then the dynamic dispatch problem could be simplified to a series of static one-period optimizations that would return the discussion to the simple model described for energy supply and demand. Scarcity pricing would arise naturally through demand bidding, and there would be no need to define or maintain operating reserves.

The reality is, however, that generation redispatch can take some time. The typical circumstance is to characterize the ramping rate of generators. This creates interdependencies across periods,

2 See Pilong Affidavit.

because the economic dispatch that can be achieved in one period depends on the level of the dispatch in previous periods. It also will take into account constraints on the amount of change needed in the dispatch between successive future periods. Modeling the dynamics of the inter-dependencies of this period's dispatch on the dispatch in future and later periods is challenging, given the constraints, but there is nothing fundamentally new in the principles of economic dispatch across multiple periods under *known* conditions, and the basic insights of the static model would continue to apply. There would be no separate requirement for operating reserves in the absence of uncertainty and the need to plan for contingencies.

But we do not have perfect foresight: the need for operating reserves arises from the uncertain future supply and demand conditions. The extension of the static model to a dynamic model with known supply and demand conditions, coupled with ramping limits, is conceptually simple. But adding a large element of uncertainty transforms this into a problem with a different degree of complexity. Uncertainty implies that this period's forecast for generation and load balance will not be the same as the actual realized system requirements. Also, the net change in load balance within a dispatch interval may be positive or negative. As a result, the system operator must create some necessary flexibility to be able to ramp up or down as needed to meet the actual load conditions. The need for excess available capacity in any period to ensure reliable system operations defines the operating reserves and related ancillary services that must be set aside to meet the uncertain conditions that unfold.

The expanded optimization could be handled in theory with an appropriate extension of the economic dispatch optimization model. In place of a deterministic optimization across known future conditions, the model would now become a stochastic dynamic optimization model. This is easy to define, but hard to implement. Continuous solution of a full stochastic dynamic optimization model defined over the relevant look-ahead period of immediate concern to the system operator would be beyond present computational capabilities, because of the curse of dimensionality. In each future period multiple "states" could exist, and the number of possible paths through the future grows rapidly like the branches of a decision tree. After very few dispatch periods, the number of possible paths would overwhelm the capabilities of existing software and super computers.

The most immediate evidence of this computational problem is found in the basic formulation of security-constrained economic dispatch. The formulation includes contingency constraints which are defined by credible events that might occur with some non-trivial probability. However, representing the full probability distribution for these events would overwhelm the system dispatch. The compromise has long been to take the conservative position to protect simultaneously against one of these events occurring (the well-known "N-1" condition), without accounting for the complete joint probabilities of one or more contingences at the same time. This is difficult enough, but it is computationally feasible within the requirements of the actual dispatch timing.

The requirement to maintain operating reserves addresses a similar problem of maintaining enough excess capacity and ramping capability to meet the reasonable possible deviations from forecast load and generator outage conditions without resorting to emergency actions up to and including involuntary load curtailment.

The resulting modeling approach is a blend of deterministic and stochastic representation of the supply, demand and system constraints. To illustrate, take the limiting case of a single period, say an hour, where the system operator has a forecast of supply and demand conditions, and a set aside of operating reserves to address the uncertain conditions over the hour. For the moment, the operating reserves are treated as fixed by rule. The operator would choose the economic dispatch ex ante based on the forecast and send instructions to the generators and loads. The exante selection of which generators to dispatch for energy and which generators to provide reserves, would be part of a co-optimization problem. Then, as the actual conditions unfold, the available operating reserves would be utilized to address the net deviations from the forecast. This simplified deterministic model utilizes a fixed quantity of operating reserves to address uncertain conditions within the hour.

For this problem, the co-optimization results would determine an implied price of operating reserves. This is conceptually the marginal cost of meeting the last megawatt of the reserve requirement, as measured by the marginal change in the value of the objective function of the exante economic dispatch. This implied price would capture the impact of the fixed reserve requirement. However, absent a good representation of scarcity values, such as through demand bidding, this implied reserve price would inherit all the problems of inadequate scarcity pricing, i.e., without demand bidding the price determined from the change in the objective function would not fully reflect the willingness to pay for additional reserve to avoid scarcity. Furthermore, the co-optimization would not capture the benefits of having higher or lower reserves than the fixed requirement.

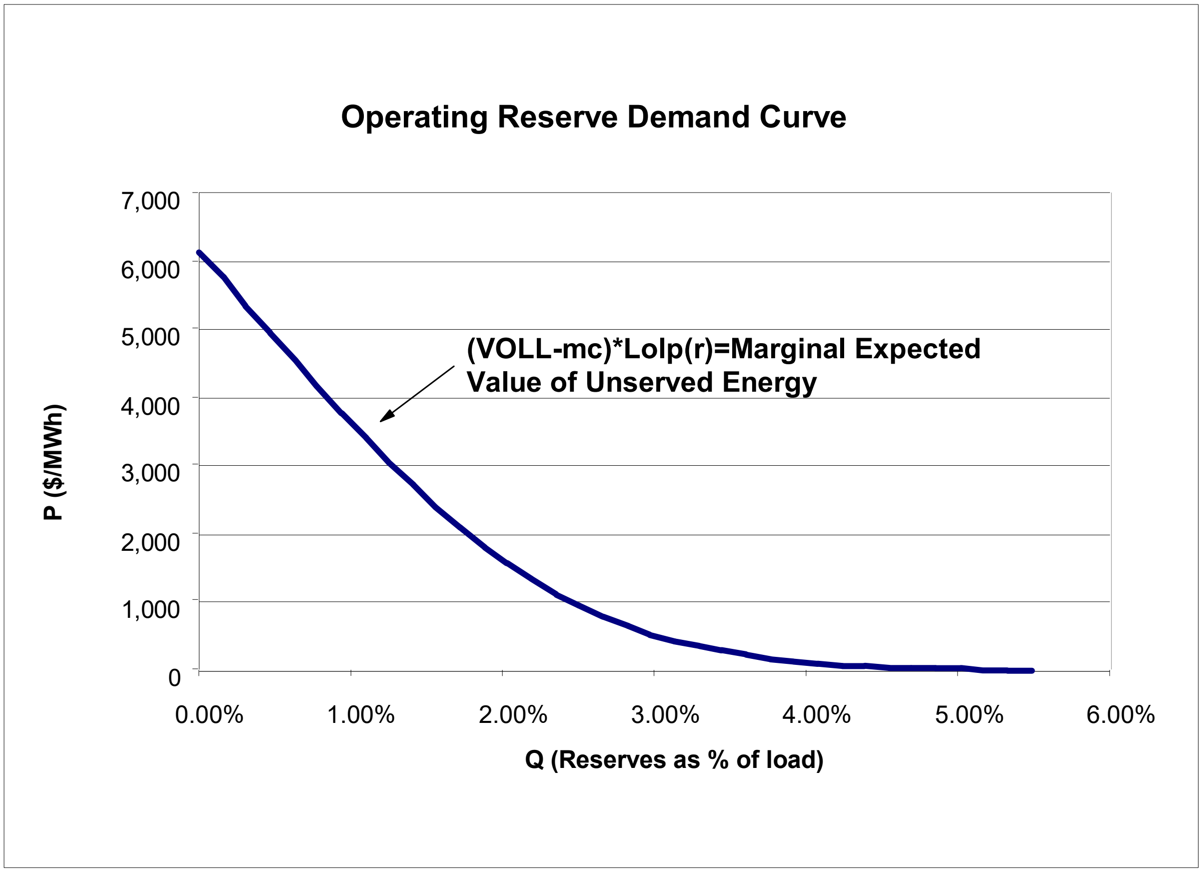
An ORDC arises to proxy for the absence of demand bidding, which as described above has not to this point been sufficiently incentivized to occur naturally in the PJM market. The essential idea of the operating reserve demand curve is to replace the fixed reserve requirement with the variable value of different levels of operating reserves. This is analogous to replacing a fixed load requirement with a bid-in load value in the energy market dispatch.

Operating reserves are a system requirement. They are a product that provides reliability simultaneously to all users of the system. The requirement to procure reserves cannot be efficiently assigned to individual loads or generators, because the need to procure a given megawatt of reserves cannot be attributed to particular market participants. For this reason, operating reserve requirements are determined and applied at the system or zonal level. There is no decentralized mechanism like demand bidding available that would provide a better market solution to efficiently schedule and provide a market-based value for operating reserves. As a result, the demand for operating reserves is an administrative construct that requires a determination by a central coordinator such as the system operator.

ORDC Structure

In a simple, one period static model with energy and operating reserves, the role of operating reserves suggests a method for approximating the ex ante value of operating reserves of different quantities. Suppose that the ex post choices for the system operator are simply to meet the net change in load through use of the available operating reserves, or to involuntarily curtail load at the cost of the value of lost load (*VOLL*) net of the variable cost of generation at the margin (*mc*). Additionally, suppose the system operator has an estimated distribution of deviations from forecasted net load during actual system operations. With this information it can calculate the loss of load probability for any given level of reserves (*Lolp(r)*). By multiplying by the (*VOLL-mc)*, this approach yields the expected cost of marginal load curtailment during actual operation corresponding to any given level of scheduled reserves. In Figure 2 below, the horizontal axis is the quantity of scheduled reserves and the curve shows the estimated incremental value of scheduling additional reserves, given the probability of loss of load. As expected, the value of additional reserves rises as the scheduled quantity falls.

Figure 2



The figure assumes that zero reserves is the reference point at which load curtailment will start to occur. The *VOLL-mc* defines the marginal willingness to pay for an increment of operating reserves at the moment of load curtailment. For this simple model, the product of the (*VOLL-mc*)**Lolp(r)* defines the ORDC, as illustrated in Figure 2 with representative parameter values. This methodology based on the value of lost load and the loss of load probability is familiar from many decades of use in resource adequacy planning studies. The only innovation here is in the use of the probability distribution over the next hour rather than over the much longer horizon of the planning studies. The extension to include minimum contingency reserves (*i.e.,* the Minimum Reserve Requirement) is described in the Appendix.

The formulation of the ORDC described here is conceptual; there are logical variations and enhancements corresponding to differences in how ISOs define the points at which they will curtail load or take emergency actions, as described in more detail below. In particular, PJM is proposing a Penalty Factor based on the maximum cost at which resources could be procured based on market offers, rather than the *VOLL*, as the anchor for its ORDCs. Consistently, the ex post choice implicit in PJM's ORDC anchor is to invoke operator actions to maintain reserves at or above the MRR at a price of $2,000, rather than the choice illustrated here, which is to curtail load at a price equal to *VOLL*. Further, the PJM proposal does not subtract the marginal cost of energy from its $2,000 penalty price in determining the anchor for the ORDC. There are other market rules in PJM which preclude use of any energy offers for price setting that rise above a current system-wide offer cap. This should preclude conditions where including the variable cost of energy in the operating reserve price would cause price rises to or above the *VOLL*. Should this change in the future, the corresponding adjustment of the representation of the scarcity price could be necessary.

PJM's formulation of its ORDCs is consistent with the theory presented here about how the value of incremental reserves will vary with the probability of loss of load during actual operation, but is anchored around PJM-specific assumptions about the actions that will be taken as the level of reserves declines below the MRR.

This simple single period example suggests the basic design criteria for an operating reserve demand curve implementation. To reach a supported result in light of the benefits of – but also the complexities of – stochastic dynamic optimization, it is appropriate to approximate the underlying stochastic dynamic optimization model using a simpler representation of uncertainty. The basic approach is to take the existing practice in defining a deterministic dynamic model with some degree of anticipation of future conditions and optimizing dispatch based on expected values of load and generation in future periods, without explicitly modeling probabilities of different outcomes in future periods. The basic structure of this standard deterministic model is retained but with the addition of an operating reserve demand curve that approximates a current estimate of the future expected value of the use of the reserves. The description should accommodate periods of different lengths and possible multiple emergency actions that could be invoked in response to the realization of net load deviations from the forecast. There could be different reserve types that have different lead times and ramping capabilities.

One design choice will be the selection of the degree to look forward in describing the uncertainty applicable in estimating the value of operating reserves. To illustrate, suppose we take the simple single period model with a one hour forward look for both the deterministic energy dispatch forecast and the period of uncertainty to be met with operating reserves. Now suppose that the actual practice is to use this model on a rolling basis to update the forecast used for the dispatch based on new information, say every five minutes. The rolling update for the forecast and deterministic optimization still looks one hour ahead. But we know that after the next five-minute interval we will update this forecast and the optimization.

How should this affect the look-ahead period for the uncertainty that applies to the operating reserve loss of load probability calculation? One suggestion might be that the uncertain period should now be matched to the five-minute period, recognizing that the operator can change both the dispatch and the selection of reserves at the start of the next five-minute interval. Although appealing, this approach would understate the importance of uncertainty and undervalue operating reserves.

Recall that the intent is to approximate prices that would capture the value in the full stochastic dynamic optimization model. This expanded modeling framework would attribute two different sources of value for operating reserves. First would be the reserves needed to meet the uncertainty in the first five-minute interval. The second element would be the value at the end of the five-minute interval for the remaining unused reserves to meet the commitments in all future intervals. The shorter the dispatch interval, the less important would be the uncertainty in the immediate interval and more would be contributed by the future value of reserves.

However, the computational barrier of the full stochastic dynamic model extends to the estimation of the value of reserves for all future periods. Hence, the actual implementation cannot depend on calculating this elusive estimate. The outline of the ORDC relies on calculation of the contribution in the relevant look-ahead interval, and implicitly assumes that the future value of the remaining reserves is effectively zero. This implies that the horizon for the uncertainty estimation should be long enough to make the residual value close to zero. This modeling requirement provides that the period selected for the estimation of the *Lolp* used in the ORDC approximation need not, and probably should not, be the same as the dispatch interval in the rolling update of the ex ante dispatch. For instance, a reasonable choice would be to have a rolling five-minute dispatch interval, with a multi-period look ahead, with each included ORDC obtained from an estimate of the uncertainty over the next hour.

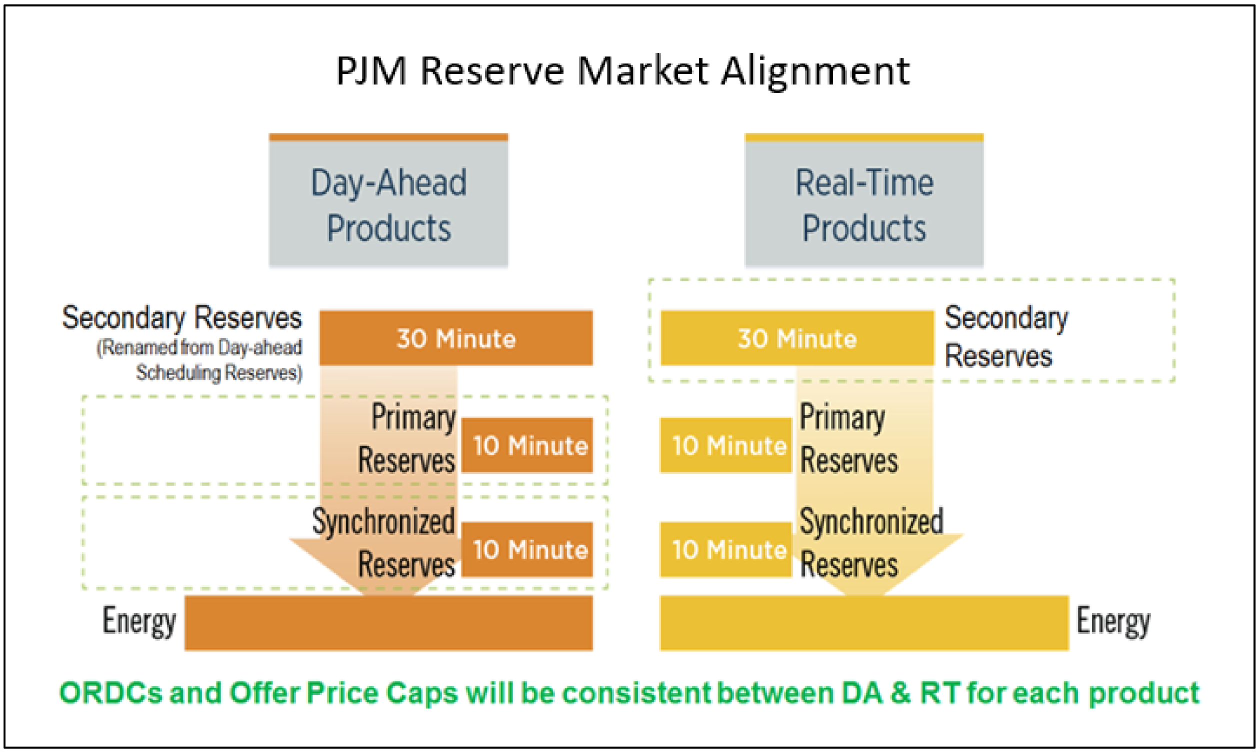
PJM proposes to base its ORDC reforms on a 30-minute look ahead for uncertainty for the Synchronized and Primary Reserve Requirements and a 60-minute look ahead for the 30-Minute Reserve Requirement. This approach is a reasonable way to account for the uncertainty in estimating the value of operating reserves because it appropriately accounts for the multi-period nature of forecast uncertainties.

Multiple Reserve Products

The basic ORDC design can accommodate multiple reserve products. The typical distinctions among types of operating reserves are based on the response time before the reserves would be available. Some reserve capability is always available, such as the capacity set aside for frequency regulation. This total regulating capacity is relatively small and must be maintained to address the need for continuous adjustments to maintain system balance and operations within set frequency tolerances. This frequency regulation capacity is not the focus of the present discussion.

In the proposed PJM design, as summarized in Figure 3, there will be three categories of operating reserve requirements denoted as “Synchronized”, “Primary”, and “30-Minute” reserves. These reserve capacities cascade. For example, Primary Reserve can also meet all the conditions that can be addressed by the 30-minute Reserves, but the reverse is not true. Hence, the price of Primary Reserve will always be at least as large as the price for 30-minute Reserve. The reserve prices in the proposed PJM model include minimum reserve requirements, penalty factors for falling below these minimums, and loss-of-load type calculations for the probability of falling below these minimums. The details for the resulting reserve products and prices are provided in the Appendix.

Figure 3



The deterministic dispatch interval for the rolling PJM economic dispatch is 5 minutes. As discussed above, this is different than the period used to set the relevant probability distribution for the uncertainty applied in valuing the operating reserves. The PJM plan can be described as a one-hour interval to determine the full range of the distribution of deviations from the expected net load. For the first half of that period, the 30-minute uncertainty applies only to the valuation of two types of 10-minute reserves. For the second half of the period, the full one-hour uncertainty applies to the combined levels of 10-minute and 30-minute reserves. The combination of uncertainties would define an ORDC that would be applied to each five minutes in the deterministic dispatch.

These modeling choices are reasonable and are consistent with similar applications in other organized markets (Hogan and Pope, 2017). The result would be a rolling dispatch of energy and reserves, co-optimizing the level and choice of generators to provide energy and those to provide reserves. Non-generator sourced reserves, such as through demand response, can be included as well. The corresponding prices for energy and reserves would be internally consistent and would reflect the scarcity value of reserves as expressed in the ORDC.

The results of the PJM approach are similar to those that would be obtained with full demand bidding of flexible load every five minutes assuming that the estimated scarcity price corresponds to the bids the loads would make and the PJM approach is, therefore, sound. Furthermore, there would be no change in the dispatch model design required to allow for the entry of more demand bidding because the co-optimization would take into account any price-sensitive demand when determining the optimal level of reserves at prices that respected the price-sensitive demand bids.

Defining and Valuing Emergency Response

System operators employ a number of actions to address imbalance problems when net load (i.e., load less generation) deviates from the dispatch forecast. These range from calls for voluntary load reductions to temporary voltage reductions and, as a last resort, involuntary load curtailments through rotating blackouts. The basic model for valuing operating reserves illustrated in Figure 2 illustrates the logic and the connection to only one action, which is to the curtailment of load

The extension to include a sequence of increasingly costly emergency actions would replace the single estimate of the *VOLL* with a series of emergency actions. Each action is available over a limited range and has an associated cost. The actions would be ordered according to increasing cost. Now the single estimate of the loss of load probability extends to an estimate of the probability that the net load change falls within the appropriate range for each emergency action. The implied value of an increment of operating reserves is the sum of the expected cost of the avoided emergency actions.

Although the refinements for multiple emergency actions are straightforward, the impact on the estimate of the ORDC over the range above the level requiring emergency actions may not be very large. Furthermore, there is a trade-off between representing the cost of the individual emergency actions and the range for the net load change that triggers the action. An example and further details on the treatment of the emergency actions are provided in the Appendix.

Analysis can support the selection of values for the various emergency actions. A guiding principle should be to set values and implied prices to reflect the actual choices of the economic dispatch as set by the system operator. However, the identification of a few representative emergency actions and their associated values also presents policy choices.

The focus is on the demand curve representing the willingness to pay to avoid the respective emergency actions. This demand perspective and the avoided cost is distinct from the cost of actions necessary to avoid a need to invoke the emergency action. The principal cost of alternatives is in the opportunity cost of providing the operating reserves, and this cost is already reflected in the generator supply offers. But in constrained situations, the measure of the cost of supply and the value of demand differ by the value of scarcity. The values needed for the ORDC are those that define the willingness to pay to avoid the emergency action.

This general perspective still leaves important choices. For example, in the case of involuntary load curtailment, the *VOLL* represents the appropriate concept to guide the decision. However, empirical studies will provide only a little guidance as to which value is appropriate across a range of estimates. The choice will depend in part on the actual curtailment policy applied by the system operator. For example, a rolling blackout will have priorities to exclude certain loads, such as for hospitals or other emergency facilities. For those who are curtailed, the relevant value for ORDC construction would be the average value of lost load for all those included, not the implied higher *VOLL* for those that would not be curtailed.

Similar issues would apply in defining other emergency actions. The balance between detail and workable approximations will be a choice. As discussed in the Appendix, in many cases it would be reasonable to aggregate the emergency actions into a small number of groups or have only one to serve as a representative proxy.

The choice in the PJM reform proposal is to employ the revised penalty factors, motivated by an objective to ensure that the individual penalties approximate at least the various costs of the emergency actions. In addition, the PJM proposal includes a cascade model that aggregates these penalty factors to produce higher operating reserve prices. With very low reserves the resulting prices would be within the range of reasonable estimates of the value of loss load (OFGEM, 2014) (Potomac Economics, 2017). The details appear in the Appendix.

Implementing an ORDC

The details of a workable computational implementation of an ORDC depend in part on the characteristics of the optimization model formulated for the bid-based, security-constrained, economic dispatch. The direct approach would be to emulate the treatment of bid-in load, where the price of the load varies with the quantity dispatched.

The price for the ORDC would be calculated as above to reflect the estimated probabilities of the changes in net load and the costs of emergency action. In principle, the scarcity component would be the marginal cost of the avoided emergency actions net of the energy cost saving from the dispatch. In most cases, this energy cost saving adjustment would be small relative to the cost of emergency actions. However, in circumstances where the variable cost of the marginal generator is close to the cost of emergency actions, the scarcity value should reflect the adjustment to ensure that the implied price of energy does not exceed the estimated costs of the emergency actions.

The basic model of the integrated real-time ORDC takes the commitment of units and related operator actions as given and independent of the scarcity conditions. In real systems, the system operator may take actions that bias the inputs into the dispatch model or after-the-fact through out-of-market (OOM) actions, for example to ensure reliability conditions that are not well represented in the dispatch model. These reliability commitments would have the effect of increasing the available system capacity and, therefore, increasing the estimated level of operating reserves.

Without some adjustment for OOM decisions, the implication would be that scarcity conditions could trigger OOM actions which in turn produce lower scarcity values for a given level of operating reserves, and therefore lower market prices. This unintended consequence should be mitigated by recognizing that the OOM actions imply a decision to pay for higher reserve requirements. The essence of the adjustment to offset the price impact of the OOM action would be to shift the ORDC by an amount that proxies for the change in capacity induced by the operator actions.

This integration of a system-wide ORDC which is carried through PJM's proposal allows for the simultaneous optimization of bid-in load, offered generation, and operating reserves. The price of operating reserve will be determined by the value from the ORDC and the tradeoff with the dispatch of generation and load. The price of energy will be the variable cost of the marginal generator plus the implied scarcity value of the generating capacity derived from the price of operating reserves.

The dispatch model can have multiple periods with a rolling update of the energy and reserve dispatch and the associated prices. In this formulation, the approximation of the underlying stochastic model employs the forecast of net load over the look-ahead horizon, and incorporates the expected marginal value represented by the penalty factors for the use of emergency actions across the range of the estimated probability distribution of the net load change over the look-ahead period.

The result is an economic dispatch model with the same deterministic structure as current dispatch models, with nothing more than the addition of the modeling equivalent of one more load subject to flexible dispatch. The incremental computational requirements would be de minimis.

Locational ORDC Design

A system-wide ORDC assumes that the location of operating reserves is not a constraint. Given the importance of transmission constraints in the basic energy dispatch, this may appear as a contradiction. The resolution of the apparent conflict rests on a characterization of transmission constraints as applied to the dispatch of energy and scheduling of reserves.

A simple approach would be to characterize the constraints under the assumption that the forecast load, generation and transmission conditions never changed. Then the economic dispatch would have the same steady-state power flows over an extended period of time. These sustained flows would face transmission limits on the ability to maintain the flows. For example, thermal limits and line sag do not happen instantaneously, but can be material limits on sustained power flows. In effect, therefore, the usual transmission limits are soft constraints in that they apply most directly to steady state conditions.

The time dependence of transmission constraints appears in the N-1 contingency limits for security constraints. Typically, the transmission power flow models are the same for the contingency conditions, but the transmission limits are higher to reflect the fact that the post-contingency flows will be active only for a limited period before the system is fully restored to normal steady-state operating conditions.

The implicit assumption in a single system-wide ORDC is that the level and period of deployment of operating reserves will be limited enough to ensure that transmission constraints need not inhibit the use of the reserves when needed.

However, there may be limits on the ability to match locational deviations in net load and the deployment of operating reserves. In effect, there could be transmission limits on the deployment of reserves. The same power flow models (i.e., lines and nodes) would govern for any given realization of the geographic dispersion of deviations in net load, but the power flow would be different in each realization. The full power flow description could be included in principle in a stochastic dynamic programming model, but this approach is computationally infeasible at present.

To remain within the deterministic framework of the existing economic dispatch optimization formats (i.e., the optimization does not explicitly consider the probabilities of different future outcomes), another approach that relies more on expected value formulations is required. Borrowing from the long experience with planning models, such as in the PJM RTEP process, the use of a zonal representation of reserve location and interface transmission limits is a feasible approach that would capture the main outlines of locational operating reserve demand curves.

Based on external analysis and simulation, the zonal approach assumes that the system operator can define nested operating reserve regions separated by a closed interface. Consider the case of the total system and one nested zone. The short-term energy transfer limit across the interface from the rest of the system to the nested zone would be a policy choice based on the external analysis. The probability distributions for net load change would be estimated for the rest of the system and for the nested zone. Since these distributions apply to the deviations from the respective forecast, independence of these distributions of forecast deviations is a simplifying but reasonable assumption.

In addition, the emergency actions that would apply in the nested zone and for the rest of the optimal dispatch of operating reserves would imply demand curves with a simple form reflecting the best use of the available reserves inside and outside of the nested zones and the interface limits. For any given realization of the net load deviation, the cost of the emergency action is assumed to be lower outside the nested zone. The cost of emergency actions outside of the nested zone applies within the nested zone up to the megawatt limit equal to the available interface capacity net of the scheduled flow from the ex ante energy dispatch. For incremental reserves within the nested zone above this amount, the more expensive emergency actions within the nested zone enter the optimal reserve schedules and determine the price of incremental reserves within the nested zone.

These rules apply for each realization of the deviation in net load. In the case of a single representation of emergency actions in each region, a corresponding probability tree allows direct calculation of the marginal value of the two types of reserves and of the interface capacity. The tree also determines the probability of the marginal values, which allows direct calculation of the expected marginal value and defines the three-element operating reserve demand curve for reserves inside and outside the constrained zone and the interface capacity. The details appear in the Appendix.

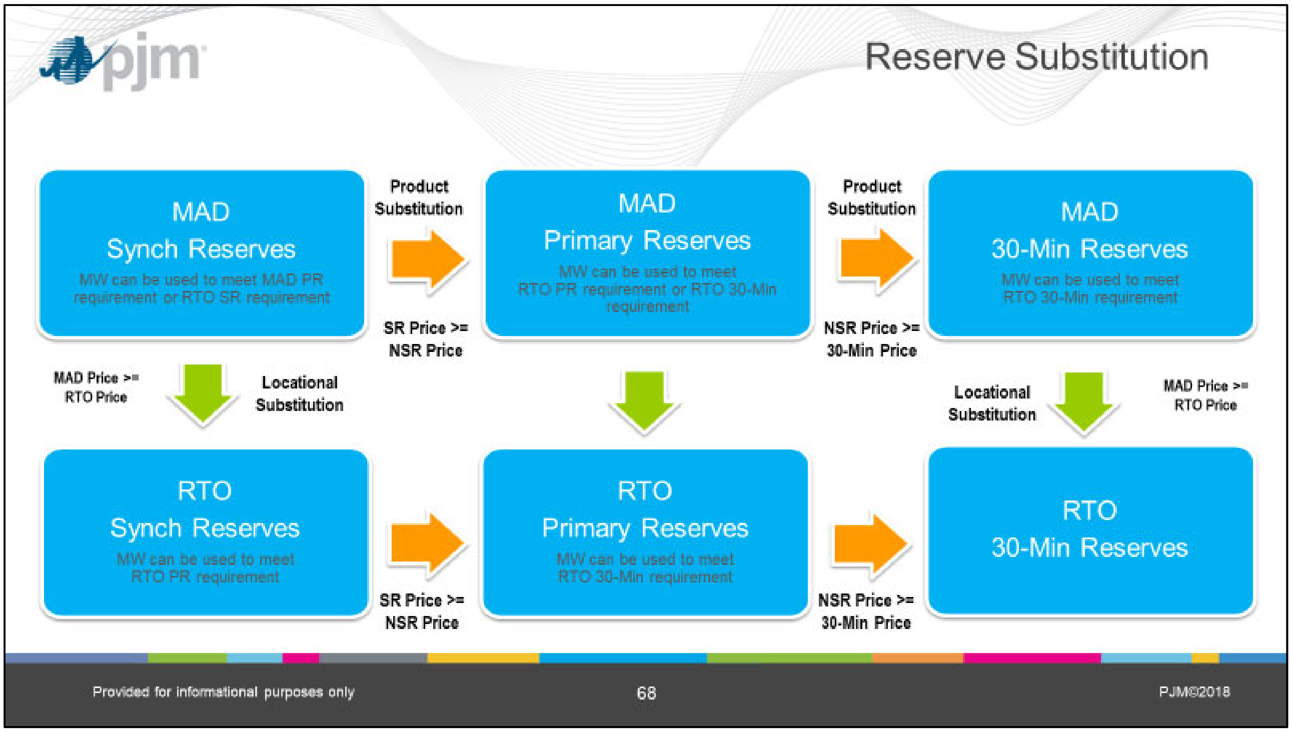
The interaction of reserves in the different regions creates a three-dimensional operating reserve demand curve interacting with load and generation in the dispatch to determine the clearing price of reserves inside of the zone, outside of the zone, and for the interface constraint. This would be similar to the conditions that would arise if bids were accepted for loads where the implied price for each load depended on the dispatch of related loads. Although possible in principle, this is not the normal assumption in existing economic dispatch models, where load bids are separable.

To maintain the goal of working within the framework of existing dispatch models, implementation of the locational ORDC must impose a separability condition, where the values are additive. One approach would set a benchmark estimate of the two levels of reserves and the interface capacity. For example, use of the co-optimized solution from the previous dispatch interval to serve as the benchmark for the current dispatch interval. Then varying each quantity, while holding the others fixed, would yield a separable set of three demand curves. This could be accepted as a workable approximation of the values from the multi-dimensional ORDC. These curves could be incorporated in the economic dispatch model in the same ways as described for the single ORDC.

If needed, this framework also would provide the information required to update the benchmark and converge to a combined solution that yields a consistent solution for the ORDC zonal prices. The co-optimization in the dispatch model includes the use of the interface limit for energy and the set aside for reserve deployment. All this would combine with the usual locational prices of energy. The details are in the Appendix.

The PJM proposed approach applies a similar but distinct approach in characterizing locational interactions. Suppose the immediate case, as shown in Figure 4, in which the Mid-Atlantic Dominion (MAD) region is constrained. The result recognizes six reserve products, the three types of real-time reserves in each of the MAD and the entire Regional Transmission Organization (RTO). The cascade model now has MAD reserves affecting the price in the RTO, but the RTO reserves do not affect the incremental price in the constrained MAD region.

Figure 4



The same approach would apply to any constrained zones. The principal simplification is to treat the entire market as a single zone, and then to represent the effects inside each constrained zone as additive to the impacts and prices for the entire region. By assuming this cascade model and additivity across locations, the PJM approach does not reflect interactions between the reserves in different regions, but has the advantage of greater simplicity in implementation. The results are illustrated in the detailed example in the Appendix.

Market Power and ORDC Pricing

A central problem in regulating power markets has been the concern with generators exercising market power. When a generator can control multiple facilities, or has related financial interests, it can manipulate price by withholding supply. The loss on its reduced production is more than made up by the increase in price received for the output from remaining units and from settlements on related financial interests. This market power can be implemented through physical withholding, by removing the plants from dispatch, or economic withholding, by increasing the plant's offer price.

Physical withholding can be addressed by must-offer requirements. More controversial has been the policy of setting accompanying offer caps to foreclose economic withholding. It is difficult to distinguish between legitimate high-cost offers and the exercise of market power. The difficulties have been compounded by the faulty assumption that higher market-clearing prices, needed to reflect scarcity and provide better incentives in operations, require high energy offers from generators. The resulting dilemma has been how to separately identify price increases resulting from scarcity from the exercise of market power.

The ORDC substantially mitigates this problem of identifying economic withholding by providing a clear distinction between a scarcity price, from the ORDC, and energy offers that reflect variable generation costs. Under conditions of reserve scarcity, when operating reserves are reduced, low energy offers can be fully compatible with high market-clearing prices, for example because the generators on line and providing energy have limited ability to ramp quickly. Reserves are limited and the ORDC produces a high scarcity price. This scarcity component adds to the variable cost of energy and produces a high market-clearing price for energy. Hence, generators do not need to inflate their variable offers in order to achieve higher prices reflecting the scarcity value of their capacity. Generous offer caps would prevent material exercise of market power through economic withholding without creating a conflict for the ORDC. Likewise, if needed, offer caps could be provided for operating reserves as well as for energy. Hence an ORDC reduces the cost of using offer caps to mitigate generator market power. This is a natural and beneficial feature of the PJM ORDC reform proposal.

Day-Ahead and Real-Time ORDC

The basic outline for an ORDC describes the design and implementation challenges within the framework of economic dispatch. The essential design elements apply both to the real-time and to day-ahead markets with a two-settlement system. As always, a general objective is to maintain consistency between the real-time design and the day-ahead design.

Consistency requires that the day-ahead representation of the ORDC reflects the uncertainty regarding the net load deviations from the forecast dispatch. The uncertainty regarding these deviations, and the short-lead times required of operating reserves, give rise to the estimated value of incremental reserves. This real-time deviation is not the same as the uncertainty between the day-ahead forecasts and real-time outcomes. The day-ahead dispatch will face both types of uncertainty.

The variation between day-ahead and real-time is important, but different from the real-time uncertainty. The treatment of this uncertainty depends on the availability of generation capacity and the timing of its commitment. If all the capacity were available for the real-time dispatch, and required no day-ahead commitment decision, then the uncertainty around the day-ahead forecast would be handled through the rolling change in the dispatch updated over the day. By contrast, since some relevant generating facilities must be committed day-ahead in order to be available for the real-time dispatch, the circumstance is analogous to the real-time problem of setting operating reserves as part of the real-time dispatch before the actual conditions are known.

The form of the day-ahead ORDC follows the same model as the real-time ORDC. The principal issue is the treatment of the day-ahead uncertainty. The details are in the Appendix in the discussion of day-ahead and real-time ORDCs.

The day-ahead settlements will be for energy and reserves. The resulting forward contracts create a financial obligation that can be met by providing the physical counterpart (generation injection, load withdrawal or operating reserve schedule) in the real-time dispatch or paying the associated market-clearing price in the real-time dispatch.

As discussed in the Appendix, the basic model for the day-ahead includes virtual transactions for both energy and reserves. Virtual bids and offers support price convergence by eliminating arbitrage opportunities between markets. In addition, with exact replication of the physical generator offers as in real-time, virtual transactions for energy and reserves can be shown to achieve full consistent equilibrium between day-ahead and the real-time dispatch and prices. See the Appendix for further details for the settlement of the operating reserve requirements.

The PJM proposal does not anticipate including virtual reserve products, but there will be virtual energy products and energy price arbitrage. The main impact of not including virtual reserve products should be to substitute "physical" reserves for virtual reserves. As always, the day-ahead reserve contracts will be settled against the real-time prices for reserves.

Summary

The PJM proposal for enhancement of operating reserve products and prices is a significant advance in the market design and will contribute to just and reasonable rates consistent with economic efficiency, reliability, open access and non-discrimination. Any approach to valuation of operating reserves, embedded within the framework of existing dispatch models, entails approximations. Economic connections between the value of emergency actions, including the value of loss of load, the probability distribution for changes in reserve requirements, and the interaction across constrained reserve zones provide a framework based on first principles and a guide to the development of implementable approximations of the interacting multiple product operating reserve demand curves. The PJM proposal moves far in this direction and the framework provides a guide for future enhancements.

Appendix: Formulation and Computation of Reserve Scarcity Prices through Operating Reserve Demand Curves

A representation of the value of operating reserves is essential for establishing prices for energy and reserves. This Appendix, adapted from (Hogan and Pope, 2017), provides further detail on the elements in the structure of an operating reserve demand curve based on first principles. The ORDC illustrated here provides an approximation of the value of operating reserves appropriate for inclusion in a single period representation of a dispatch model.

The full co-optimization framework, simultaneously considering both the multi-period dispatch of energy and reserves to meet forecast load conditions, could be important for some extensions of the ORDC.

Economic Dispatch and Operating Reserves

The assumption of the existence of an operating reserve demand curve simplifies the analysis. The demand curve gives rise to a reserve benefit function that can be included in the objective function for economic dispatch. The basic framework approximates the complex problem with a wide range of uncertainties and applies a pricing logic to match the actions of system operators. The main features include:

- **Single Period Model**. There is a static representation of the underlying dynamic problem. This static formulation is a conventional building block for a multi-period framework.
- **Deterministic Representation**. The single period dispatch formulation is based on bids, offers, and expected network conditions as in standard economic dispatch models. The operating reserve demand curve represents the value of uncertain uses of reserves without explicitly representing the uncertainty in the optimization model.
- **Security Constrained**. The economic dispatch model includes the usual formulation of N-1 contingency constraints to preclude cascading failures.
- **Ex-Ante Dispatch**. The dispatch is determined before uncertainty about net load relative to forecast is revealed.
- **Expected Value for Reserves**. The reserve benefit function represents the expected value of avoiding involuntary load curtailments and similar emergency actions.
- **Multiple Reserve Types**. The model of the operating reserve demand allows for a typical cascade model of different reserve types. Online spinning reserves and fast start standby reserves interact to provide complementary reserve prices.
- **Administrative Balancing**. Subsequent uncertain events are treated according to administrative rules to utilize operating reserves to maintain system balance and minimize load curtailments.
- **Consistent Prices**. The model co-optimizes the dispatch of energy and reserves and produces a consistent set of prices for the period.

The framework allows for a variety of implementations with multiple zones, forward markets and other common aspects of electricity markets.

Modeling Co-Optimized Economic Dispatch with Operating Reserves

The model presented below is a one-period "DC-load" model with co-optimization of reserves and energy. The notion is that the dispatch set at the beginning of the period must include some operating reserves that could deal with subsequent uncertain events. The emphasis is on the co-optimization of energy and reserves to illustrate the major interactions with energy prices. The initial approach assumes no locational constraints on reserves. The initial, simplified example assumes the existence of a separable non-locational benefit function for reserves.

Here the various variables and functions include:

d : Vector of locational demands

g_R : Vector of locational responsive generation

r_R : Vector of locational responsive reserves

r_{NS} : Vector of locational non-spin reserves

r_R^0 : Aggregate responsive reserves

r_{NS}^0 : Aggregate non-spin reserves

g_{NR} : Vector of locational generation not providing reserves

$B(d)$: Benefit function for demand

$C_k(g_k)$: Cost function for generation offers

K_k : Generation Capacity

$R_k(r_k)$: Reserve value function integrating demand curves

$r_k^{\max}$: Maximum Ramp Rate

H, b : Transmission Constraint Parameters

i : Vector of ones.

Assuming that unit commitment is determined, the stylized economic dispatch model is:

(1)
$$
\begin{array}{lcc}
\underset{d,g_R,g_{NR},r_R^0,r_R,r_{NS}^0,r_{NS}\geq 0;\hat{y}}{Max} \quad B(d)-C_R(g_R)-C_{NR}(g_{NR})+R_I(r_R^0)+R_{II}(r_{NS}^0) & & \\
d-g_R-g_{NR}=\hat{y} & \text{Net Loads} & \rho \\
i^t\hat{y}=0 & \text{Load Balance} & \lambda \\
H\hat{y}\leq b & \text{Transmission Limits} & \mu \\
g_R+r_R\leq K_R & \text{Responsive Capacity} & \theta_R \\
g_{NR}\leq K_{NR} & \text{Generation Only Capacity} & \theta_{NR} \\
r_{NS}\leq K_{NS} & \text{Non-Spin Capacity} & \theta_{NS} \\
i^t r_R = r_R^0 & \text{Responsive Reserves} & \gamma_R \\
i^t r_R + i^t r_{NS} = r_{NS}^0 & \text{Non-Spin Reserves} & \gamma_{NS} \\
r_R\leq r_R^{\max} & \text{Responsive Ramp Limit} & \eta_R \\
r_{NS}\leq r_{NS}^{\max} & \text{Non-Spin Ramp Limit} & \eta_{NS}.
\end{array}
$$

This formulation assumes that the non-spinning reserve generators are not spinning and, therefore, cannot provide energy for the dispatch. The Non-Spinning Reserve equation implements a cascaded model for reserves, where both responsive and non-spinning reserves contribute to the aggregate non-spinning supply. The cost for reserves is the opportunity cost in the tradeoff for providing energy.

For the present discussion, the pricing relationships follow from the usual interpretation of a convex economic dispatch model. This could be expanded to include unit commitment and extended LMP formulations (ELMP), but the basic insights would be similar (Gribik, Hogan and Pope, 2007).

An interpretation of the prices follows from analysis of the dual variables and the complementarity conditions. For an interior solution, the locational prices (ρ) are equal to the demand prices for load.

(2) $\rho=\nabla B(d).$

The same locational prices connect to the system lambda and the cost of congestion for the binding transmission constraints in the usual way.

(3) $\rho=\lambda i+\mu^t H.$

In addition, the locational prices equate with the marginal cost of generation plus the cost of scarcity.

(4) $\rho=\nabla C_R(g_R)+\theta_R.$

A similar relation applies for the value of non-reserve related generation.

(5) $\rho = \nabla C_{NR}\left(g_{NR}\right) + \theta_{NR}.$

The marginal value of responsive reserves connects to the scarcity costs of capacity and ramping limits.

$$(6) \qquad \theta_R + \eta_R = \gamma_R i + \gamma_{NS} i = \frac{dR_I\left(r_R^0\right)}{dr} i + \frac{dR_{II}\left(r_{NS}^0\right)}{dr} i.$$

The corresponding marginal value of non-spinning reserves reflects the scarcity value for capacity and ramping limits.

$$(7) \qquad \theta_{NS} + \eta_{NS} = \gamma_{NS} i = \frac{dR_{II}\left(r_{NS}^0\right)}{dr} i.$$

If there are no binding ramp limits for responsive reserves, then $\eta_R = 0$ and from (6) we have θ_R as a vector where every element is the price of responsive reserves. Similarly, for the ramping limits on non-spinning reserves, if these are not binding, then $\eta_{NS} = 0$ and from (7) we have θ_{NS} as a vector where every element is the price of non-spinning reserves.

Approximate Operating Reserve Demand Curve in a Co-optimized System

This co-optimization model captures the principal interaction between energy offers and scarcity value. The assumption of a benefit function, $R(r)$, for reserves simplifies the analysis. Here, a derivation of a possible reserve benefit function provides a background for describing the form of an ORDC. To simplify the presentation, focus on the role of one class of responsive reserves only. And consider only an aggregate requirement for reserves with no locational constraints.

To the various variables and functions add:

$f(x)$: Probability for net load change equal to x.

Again, for purposes of designing the ORDC take that unit commitment as given. The stylized economic dispatch model includes an explicit description of the expected value of the use of reserves. For the reserves here, only aggregate load matters. This reserve description allows for a one-dimensional change in aggregate net load, x, and an asymmetric response where positive net load changes are costly and met with reserves and negative changes in net load are ignored. This model is too difficult to implement but it provides an interpretation of a set of assumptions that leads to an approximate ORDC. Here we first ignore minimum reserve requirements to focus on the expected cost of the reserve dispatch.

The central formulation treats net load change x and use of reserve, δ_x, to avoid involuntary curtailment. This produces a benefit minus cost of $VOLL \cdot \left(i^t \delta_x\right) - \left(C_R\left(g_R + \delta_x\right) - C_R\left(g_R\right)\right)$ and this is weighted by the probability $f(x)$. This term enters the objective function summed for all non-negative values of x. The basic formulation includes:

$$
(8)\qquad
\begin{aligned}
&\underset{d,g_R,g_{NR},r_R,\delta_x\geq 0;\hat{y}}{Max} \quad B(d)-C_R(g_R)-C_{NR}(g_{NR})+\sum_{x\geq 0}\Big(VOLLi^t\delta_x-\big(C_R(g_R+\delta_x)-C_R(g_R)\big)\Big)f(x)\\
&\begin{array}{lll}
d-g_R-g_{NR}=\hat{y} & \text{Net Loads} & \rho\\
i^t\hat{y}=0 & \text{Load Balance} & \lambda\\
H\hat{y}\leq b & \text{Transmission Limits} & \mu\\
g_R+r_R\leq K_R & \text{Responsive Capacity} & \theta_R\\
i^t\delta_x\leq x,\ \forall x & \text{Responsive Utilization} & \gamma_x\\
\delta_x\leq r_R,\ \forall x & \text{Responsive Limit} & \varphi_x\\
g_{NR}\leq K_{NR} & \text{Generation Only Capacity} & \theta_{NR}.
\end{array}
\end{aligned}
$$

This model accounts for all the uncertain net load changes weighted by the probability of the outcome, and allows for the optimal utilization of reserve dispatch in each instance.

To approach the assessment of how to approximate reserves with a common scarcity price across the system, further simplify this basic problem.

1. Treat the utilization of reserves δ_x as a one-dimensional aggregate variable.
2. Replace the responsive reserve limit vector with a corresponding aggregate constraint on total reserves.
3. Utilize an approximation of the cost function, $\hat{C}$, for the aggregate utilization of reserves, and further approximate the change in costs with the derivative of cost times the utilization of reserves.

This set of assumptions produces a representation for the use of a single aggregate level of reserves for the system:

$$
(9)\qquad
\begin{aligned}
&\underset{d,g_R,g_{NR},r_R,\delta_x\geq 0;\hat{y}}{Max} \quad B(d)-C_R(g_R)-C_{NR}(g_{NR})+\sum_{x\geq 0}\Big(VOLL\delta_x-\partial\hat{C}_R\left(i^tg_R\right)\delta_x\Big)f(x)\\
&\begin{array}{lll}
d-g_R-g_{NR}=\hat{y} & \text{Net Loads} & \rho\\
i^t\hat{y}=0 & \text{Load Balance} & \lambda\\
H\hat{y}\leq b & \text{Transmission Limits} & \mu\\
g_R+r_R\leq K_R & \text{Responsive Capacity} & \theta_R\\
\delta_x\leq x,\ \forall x & \text{Responsive Utilization} & \gamma_x\\
\delta_x\leq i^tr_R,\ \forall x & \text{Responsive Limit} & \varphi_x\\
0\leq r_R, & \text{Explicit Sign Constraint} & \omega_R\\
g_{NR}\leq K_{NR} & \text{Generation Only Capacity} & \theta_{NR}.
\end{array}
\end{aligned}
$$

This formulation provides a reasonably transparent interpretation of the implied prices. Focusing on an interior solution for all the variables except r_R, we would have locational prices related to the marginal benefits of load:

(10) $\rho = \nabla B\left(d\right).$

The same locational prices connect to the system lambda and the cost of congestion for the binding transmission constraints.

(11) $\rho = \lambda i + H^t \mu.$

The locational prices equate with the marginal cost of generation-only plus the cost of scarcity when this generation is at capacity, which appears in the usual form.

(12) $\rho = \nabla C_{NR}\left(g_{NR}\right) + \theta_{NR}.$

The locational prices equate with the marginal cost of responsive generation and display the impact of reserve scarcity. First, the impact of changing the base dispatch of responsive generation implies:

$$\rho = \nabla C_R\left(g_R\right) + \sum_{x \geq 0}\left(\partial^2 \hat{C}_R\left(i^t g_R\right)\delta_x i\right) f\left(x\right) + \theta_R.$$

The second-order term captures the effect of the base dispatch of responsive generation on the expected cost of meeting the reserves due to deviation in net load leading to δ_x in reserves being needed. This term is likely to be small. For example, if we assume that the derivative $\partial \hat{C}_R$ is constant, then the second order term is zero.

When we account for the base dispatch of reserves, we have:

$$\theta_R = \sum_{x \geq 0} \varphi_x i + \omega_R .$$

When accounting for utilization of the reserves, we have:

$$\gamma_x + \varphi_x = \left(VOLL - \partial \hat{C}_R\left(i^t g_R\right)\right) f\left(x\right).$$

Let $r = i^t r_R$. Then for $x \leq r, \quad \varphi_x = 0; \quad x \geq r, \quad \gamma_x = 0$. Hence,

$$\theta_R = \sum_{x \geq r} \varphi_x i + \omega_R = \left(VOLL - \partial \hat{C}_R\left(i^t g_R\right)\right)\left(1 - F\left(r\right)\right) i + \omega_R .$$

Combining these, we can rewrite the locational price as:

(13) $\rho = \nabla C_R\left(g_R\right) + \sum_{x \geq 0}\left(\partial^2 \hat{C}_R\left(i^t g_R\right) i \delta_x\right) f\left(x\right) + \left(VOLL - \partial \hat{C}_R\left(i^t g_R\right)\right)\left(1 - F\left(r\right)\right) i + \omega_R.$

Equations (9) through (13) capture our approximating model for aggregate responsive reserves. Here $Lolp(r) \equiv 1 - F(r)$, the loss of load probability given reserve r.

The term $\left(VOLL - \partial\hat{C}_R\left(i^t g_R\right)\right)\left(1 - F(r)\right)$ in (13) is the scarcity price of the ORDC. If the second order terms in (13) are dropped, then the scarcity price is the only change from the conventional generation-only model. In practice, we would have to update this model to account for minimum reserve levels, non-spin, and so on, and include an estimate of $\overline{c} \approx \partial\hat{C}_R$ in defining the net value of operating reserves $v \approx VOLL - \overline{c}$.

Note that under these assumptions the scarcity price is set according to the opportunity cost of using generation for reserves rather than the increasing cost to produce energy, i.e., using $\hat{C}$ for the marginal responsive generator in the base dispatch. Depending on the accuracy of the estimate in $\hat{C}$, this seeks to maintain that the energy price plus scarcity price never exceeds the value of lost load.

Providing a reasonable estimate for $\hat{C}$ could be done either as an (i) exogenous constant, (ii) through a two-pass procedure, or (iii) approximately in the dispatch. For example, a possible procedure would define the approximating cost function as the least unconstrained cost of the responsive generation dispatch to provide $\hat{g}_R$ of reserves:

$$\hat{C}\left(\hat{g}_R\right) = Min\left\{C\left(g_R\right) \middle| \hat{g}_R = i^t g_R\right\}.$$

This information would be easy to evaluate before the dispatch.

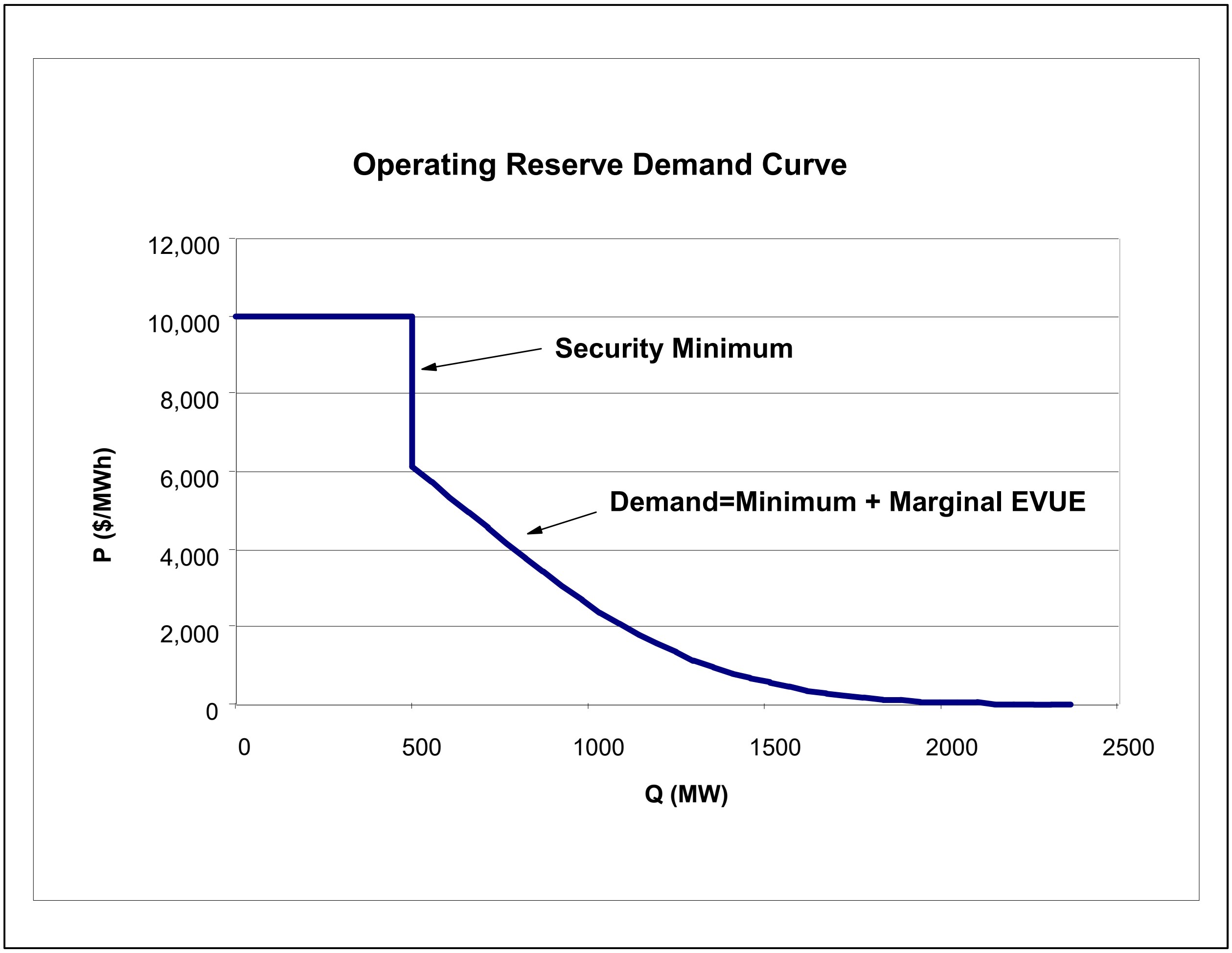
The loss of load probability calculation could reference zero reserves, or could require a minimum contingency level of reserves that provide the base level for the calculation. To construct the ORDC for responsive reserves that modifies (13) to incorporate the security minimum or last resort reserves X priced at v. Here, $Lolp(r) = \text{Probability}\left(\text{Net Load Change} \geq r\right)$. For a candidate value of the aggregate responsive reserves define the corresponding value on the operating reserve demand curve:

$$\pi_R\left(r_R\right) = \begin{Bmatrix} Lolp\left(i^t r_R - X\right), & i^t r_R - X \geq 0 \\ 1, & i^t r_R - X < 0 \end{Bmatrix} \tag{14}$$

$$P_R\left(r_R\right) = v\pi_R\left(r_R\right).$$

This defines the ORDC for responsive reserves with contingency minimum X, as illustrated in Figure 5. The corresponding reserve value is the area under the ORDC defined by the minimum level and the marginal expected value of unserved energy (EVUE).

Figure 5



With this definition, the price of energy is the marginal cost of energy plus the scarcity value, and is bounded by $VOLL$.

Multiple Emergency Actions

The basic logic extends to the case where there are multiple stages of emergency actions triggered by a low level of responsive reserves. The price of reserves is defined by the willingness to pay at the margin to obtain an additional unit of reserves. If emergency actions need be taken ex-ante, then the willingness to pay will be at least the cost of the emergency action. In addition, the value of reserves would be at least the ex-post value of an increment of reserves given the probability distribution of the net load change relative to the anticipated dispatch of the net load.

For example, suppose that we have three emergency actions, with limited capacity, where only the last requires involuntary curtailment of load at the full $VOLL$. Let the first two actions have values of emergency action $VEA_1 < VEA_2 < VOLL$, and available capacities KEA_1, KEA_2. Define the contingency minimum for reserves at X_3 where the $VOLL$ applies. Let the other breakpoints be:

$$\begin{aligned} X_2 &= X_3 + KEA_2 \\ X_1 &= X_2 + KEA_1 \end{aligned}.$$

Then define $\upsilon(s)$, including the minimum contingency levels and emergency actions, as the greater of the ex-ante cost and the expected cost of using the emergency action given the level of reserves in the event that there is a deviation for the forecast net load. Here *VOLL* and values of various emergency actions are set net of the marginal cost of energy dispatch $\bar{c}$.

$$(15) \qquad \upsilon(s) = \begin{cases} VOLL, & s \leq X_3 \\ Max\left(VOLL * Lolp(s - X_3), VEA_2\right), & X_3 \leq s \leq X_2 \\ Max\begin{pmatrix} VOLL * Lolp(s - X_3) + \\ VEA_2 * \left(Lolp(s - X_2) - Lolp(s - X_3)\right), VEA_1 \end{pmatrix}, & X_2 \leq s \leq X_1 \\ VOLL * Lolp(s - X_3) + VEA_2 * \left(Lolp(s - X_2) - Lolp(s - X_3)\right) & \\ \quad + VEA_1 * \left(Lolp(s - X_1) - Lolp(s - X_2)\right), & X_1 \leq s \end{cases}.$$

Hence, the ex-ante scarcity value for reserves is $P_R(r_R) = \upsilon(r_R)$.

If the two emergency values are high enough, then given an operating reserve level *r* above the total $X + KEA_1 + KEA_2$, the marginal value of an increment of responsive operating reserves would be:

$$\begin{aligned} P_R(r) = {} & VEA_1\left[Lolp(r) - Lolp(r + KEA_1)\right] + VEA_2\left[Lolp(r + KEA_1) - Lolp(r + KEA_1 + KEA_2)\right] \\ & + VOLL\left[Lolp(r + KEA_1 + KEA_2)\right]. \end{aligned}$$

This is the expected value component of the ORDC. The full ORDC in the dispatch would include the steps in the emergency response, and the probabilistic value of additional reserves, as in (15).

Figure 6

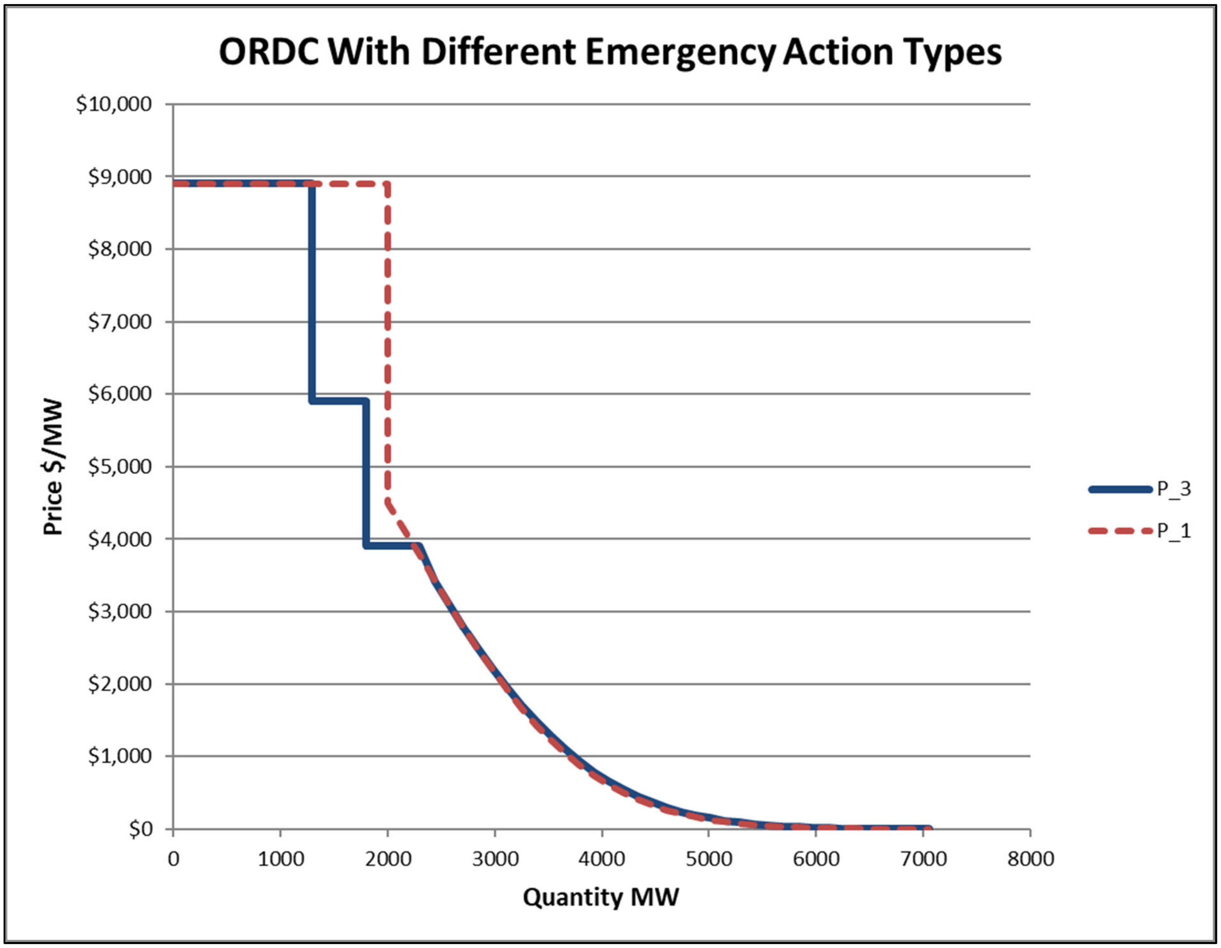


Figure 6 shows an illustrative case, P_3, with the first emergency action at $4000/MW for 500MW, the second at $6000 for 500MW, and the final X value of minimum contingency reserves at 1300MW with a *VOLL*=$9000/MW. The corresponding emergency action "X_1" value is then at 2300MW. The comparison is with the ORDC P_1 with only one emergency action implemented with X=2000MW and *VOLL*=$9000.[3]

---

3 The basic assumptions for the illustrative normal distribution of changes in net load are

Expected Total (MW)	16
Std Dev (MW)	1357.00
VOLL ($/MWh)	9000
Marginal Dispatch ($/MWh)	100

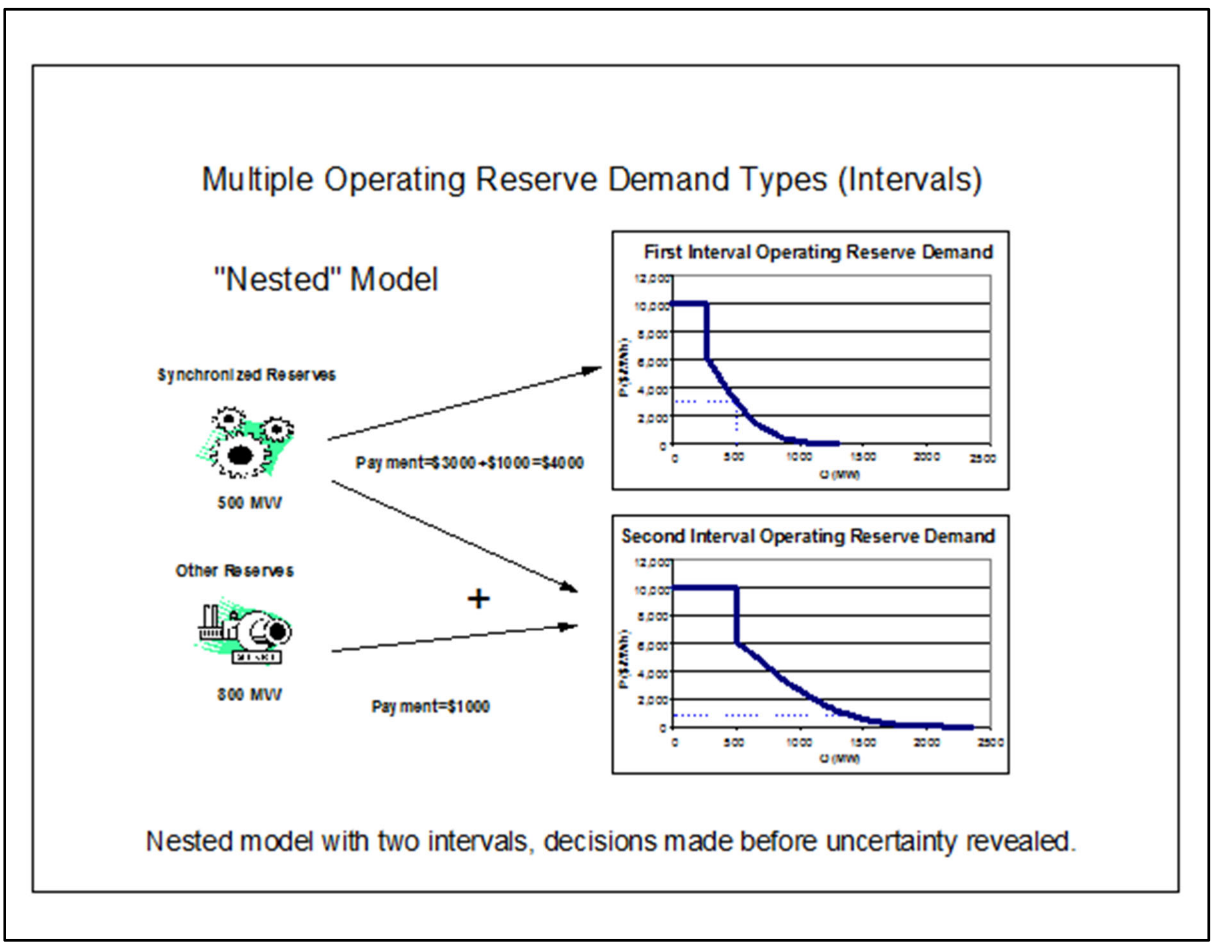
Multiple Reserve Types

The organized market in practice distinguishes several types of reserves. An approximation of the impacts of the different reserve types allows for a cascade model, reflecting different qualities of reserves, and ensure consistent prices that recognize these interactions among reserve types. A generic example with two types of reserves provides the general analysis. This serves as background for a comparison with different implementations.

Two Reserve Types

Setting aside regulation, the principal distinction is between "responsive" reserves (R) and "non-spin" reserves (NS). The ORDC framework can be adapted to include multiple reserves. This section summarizes one such modeling approach with two types of reserves and relates it to the co-optimization examples above. The main distinction is that "responsive" reserves are spinning and have a quick reaction time. These reserves would be available almost immediately and could provide energy to meet increases in net load over the whole of the operating reserve period. By comparison, non-spinning reserves are slower to respond and would not be available for the entire period.

Figure 7



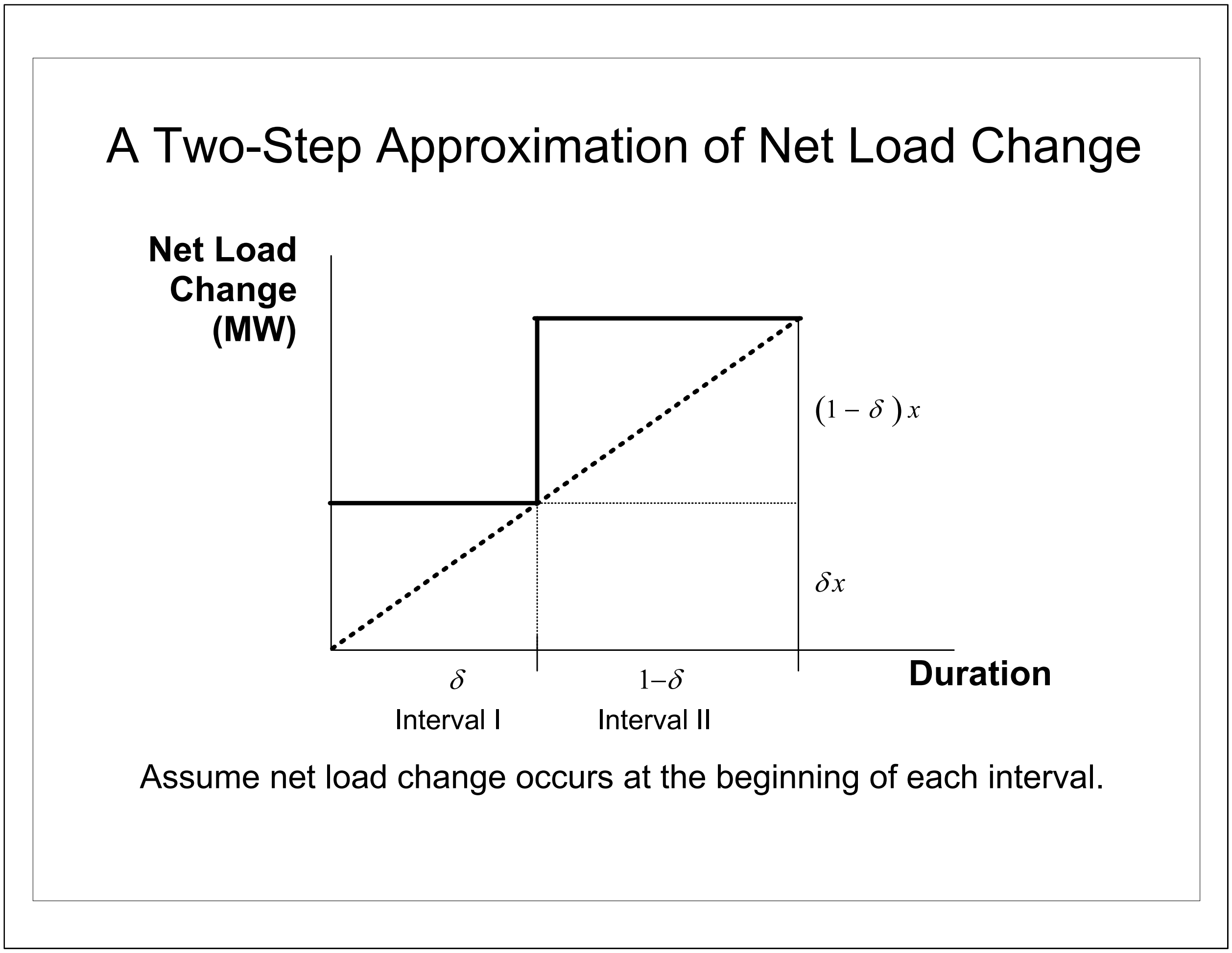
This formulation lends itself to the interpretation of Figure 2 where there are two periods with different demand curves and the models are nested. In other words, responsive reserves r_R can meet the needs in both intervals and the non-spinning reserves r_{NS} can only meet the needs for the second interval.

The proposed model of operating reserves approximates the complex dynamics by assuming that the uncertainty about the unpredicted change in net load is revealed after the basic dispatch is determined. The probability distribution of change in net load is interpreted as applying the change over the uncertain reserve period, say the next hour, divided into two intervals. Over the first interval, of duration (δ), only the responsive reserves can avoid curtailments. Over the second interval of duration (1-δ), both the responsive and non-spin reserves can avoid involuntary load shedding.

In order to keep the analysis of the marginal benefits of more reserves simple, there is an advantage of utilizing a step function approximation for the net load change. (This keeps the marginal value in an interval constant, and we don't have to compute expectations over the varying net load change possible in each period. We only need the total *Lolp* over that interval.)

The standard deviation of the change in net load is for the total over the period. If the change were spread out over the period, then on average it would be more like the diagonal dashed line in Figure 8. An alternative two-step approximation in Figure 8 is that the net load change in the first interval, when only responsive reserves can respond, is proportional to the total load change over the period relative to the length of the first interval, and the second step captures the total change at the beginning of the second interval.

Figure 8



During the first interval, only the responsive reserves apply. In the second interval, both responsive and non-spin reserves have been made available to help meet the net change in load. Suppose that there are two variables y_I, y_{II} representing the incremental net load change in each of the two intervals. Further assume that the two variables have a common underlying distribution for a variable z for total net load change but are proportional to the size of the interval. Then, assuming independence and with x the net load change over the full two intervals, we have:

$$E(y_I) = E(\delta z) = \delta E(z),$$
$$E(y_{II}) = E((1-\delta) z) = (1-\delta) E(z).$$
$$Var(y_I) = Var(\delta z) = \delta^2 Var(z),$$
$$Var(y_{II}) = Var((1-\delta) z) = (1-\delta)^2 Var(z).$$
$$E(z) = E(y_I + y_{II}) = E(x) = \mu.$$

Imposing the independence assumptions, with the hour ahead standard deviation estimated as (σ), we have:

$$Var(x) = Var(y_I + y_{II}) = Var(y_I) + Var(y_{II}) = \left(\delta^2 + (1-\delta)^2\right) Var(z).$$
$$Var(z) = \frac{Var(x)}{\delta^2 + (1-\delta)^2} = \frac{\sigma^2}{\delta^2 + (1-\delta)^2}.$$

The distinction here is that the implied variance of the individual intervals is greater compared to the one-draw assumption, even though the total variance of the sum over the two intervals is the same as the one draw. This is simply an impact of the square root law for the standard deviation of the sums of independent random variables.

Hence, for the first interval, the standard deviation is $\frac{\delta\sigma}{\sqrt{\delta^2 + (1-\delta)^2}}$, where σ is the standard deviation of the net change in load over both intervals.

Here the different distributions refer to the net change in load over the first interval, and over the sum of the two intervals. The distribution over the sum is just the same distribution for the whole period that was used above. Then $y_I \sim Lolp_I, y_I + y_{II} \sim Lolp_{I+II}$. A workable approximation would be to utilize the normal distribution for the net load change.

As before, there would be an adjustment to deal with the minimum reserve to meet the max contingency. The revised formulation would include:

(16)
$$\pi_R(r_R) = \begin{Bmatrix} Lolp_I\left(i^t r_R - X\right), & i^t r_R - X \geq 0 \\ 1, & i^t r_R - X < 0 \end{Bmatrix}$$
$$\pi_{NS}(r_R, r_{NS}) = \begin{Bmatrix} Lolp_{I+II}\left(i^t r_R + i^t r_{NS} - X\right), & i^t r_R + i^t r_{NS} - X \geq 0 \\ 1, & i^t r_R + i^t r_{NS} - X < 0 \end{Bmatrix}$$
$$P_R(r_R, r_{NS}) = v * \left(\delta * \pi_R(r_R) + (1-\delta) * \pi_{NS}(r_R, r_{NS})\right),$$
$$P_{NS}(r_R, r_{NS}) = v * (1-\delta) * \pi_{NS}(r_R, r_{NS}).$$

This representation produces different values for the responsive and non-spin reserves. Let *v* be the net value of load curtailment, defined as the value of lost load less the avoided cost of energy dispatch offer for the marginal reserve. The interpretation of the prices of reserves, P_R and P_{NS}, is the marginal impact on the load curtailment times $Lolp$, the probability of the net change in load being greater that the level of reserves, r_R and r_{NS}. This marginal value differs for the two intervals, as shown in the following table:

Marginal Reserve Values		
	Interval I	Interval II
Duration	δ	1-δ
P_R	$vLolp_I\left(r_R\right)$	$vLolp_{I+II}\left(r_R+r_{NS}\right)$
P_{NS}	0	$vLolp_{I+II}\left(r_R+r_{NS}\right)$

This formulation lends itself to implementation in the co-optimization model. For example, given benchmark estimates for each type of reserves, $\left(\hat{r}_R,\hat{r}_{NS}\right)$, the problem becomes separable in responsive and non-spin reserves. For example, the benchmark for the current dispatch interval could be set obtained from the co-optimized dispatch solution for the previous dispatch interval. Numerical integration of $P_R\left(r_R,\hat{r}_{NS}\right)$ and $P_{NS}\left(\hat{r}_R,r_{NS}\right)=P_{NS}\left(0,+\hat{r}_R+r_{NS}\right)$ produce separable functions that yield the counterpart benefit functions, $R_I\left(r_R^0\right),R_{II}\left(r_{NS}^0\right)$. With weak interactions between the types of reserves, the experience with this type of decomposition method suggests that the initial approximation will be good and updating the benchmark estimates in an iterative model could produce rapid convergence to the simultaneous solution (Ahn and Hogan, 1982).

PJM ORDC Proposal

The PJM proposal includes three types of reserve requirements. These are Synchronized (SR), Primary (PR), and 30-Minute (30) reserves. Each requirement has a penalty factor (PF), a minimum reserve requirement (MRR), and an associated probability distribution. For SR and PR, the probability distribution is taken from the empirical 30-minute forecast error look ahead distribution. For the 30-min reserves, the probability distribution is the empirical 1-hour ahead forecast error distribution.

Let the (increasing) MRRs be X_{SR},X_{PR},X_{30}. The loss-of-load probability distributions are $Lolp_{30},Lolp_{60}$. There is a cascade model for the prices and probabilities. The convention here is to distinguish the separate types of reserve products: Synchronized (SR), Non-Synchronized (NSR) and Secondary (SecR) $(r_{SR},r_{NSR},r_{SecR})$, recognizing that the SR product contributes to the SR, PR and 30-Minure reserves requirements, the NSR product contributes to the PR and 30-minute reserves requirements and the SecR product contributes to the 30-minute reserves requirement. The relevant probability factors for the three types of reserves become:

$$\tilde{\pi}_{SR}\left(r_{SR}\right)=\begin{Bmatrix} Lolp_{30}\left(r_{SR}-X_{SR}\right), & r_{SR}-X_{SR}\geq 0 \\ 1, & r_{SR}-X_{SR}<0 \end{Bmatrix}$$

$$\tilde{\pi}_{PR}\left(r_{SR},r_{NSR}\right)=\begin{Bmatrix} Lolp_{30}\left(r_{SR}+r_{NSR}-X_{PR}\right), & r_{SR}+r_{NSR}-X_{PR}\geq 0 \\ 1, & r_{SR}+r_{NSR}-X_{PR}<0 \end{Bmatrix}$$

$$\tilde{\pi}_{30}\left(r_{SR},r_{NSR},r_{\mathrm{Sec}R}\right)=\begin{Bmatrix} Lolp_{60}\left(r_{SR}+r_{NSR}+r_{\mathrm{Sec}R}-X_{30}\right), & r_{SR}+r_{NSR}+r_{\mathrm{Sec}R}-X_{30}\geq 0 \\ 1, & r_{SR}+r_{NSR}+r_{\mathrm{Sec}R}-X_{30}<0 \end{Bmatrix}.$$

With these definitions, the single region (RTO-wide) PJM proposed definition of the price approximation is as in:

$$\tilde{P}_{SR}\left(r_{SR},r_{NSR},r_{\mathrm{Sec}R}\right)=PF_{SR}*\tilde{\pi}_{SR}\left(r_{SR}\right)+PF_{PR}*\tilde{\pi}_{PR}\left(r_{SR},r_{NSR}\right)+PF_{30}*\tilde{\pi}_{30}\left(r_{SR},r_{NSR},r_{\mathrm{Sec}R}\right)$$

$$\tilde{P}_{NSR}\left(r_{SR},r_{NSR},r_{\mathrm{Sec}R}\right)=PF_{PR}*\tilde{\pi}_{PR}\left(r_{SR},r_{NSR}\right)+PF_{30}*\tilde{\pi}_{30}\left(r_{SR},r_{NSR},r_{\mathrm{Sec}R}\right)$$

$$\tilde{P}_{\mathrm{Sec}R}\left(r_{SR},r_{NSR},r_{\mathrm{Sec}R}\right)=PF_{30}*\tilde{\pi}_{30}\left(r_{SR},r_{NSR},r_{\mathrm{Sec}R}\right).$$

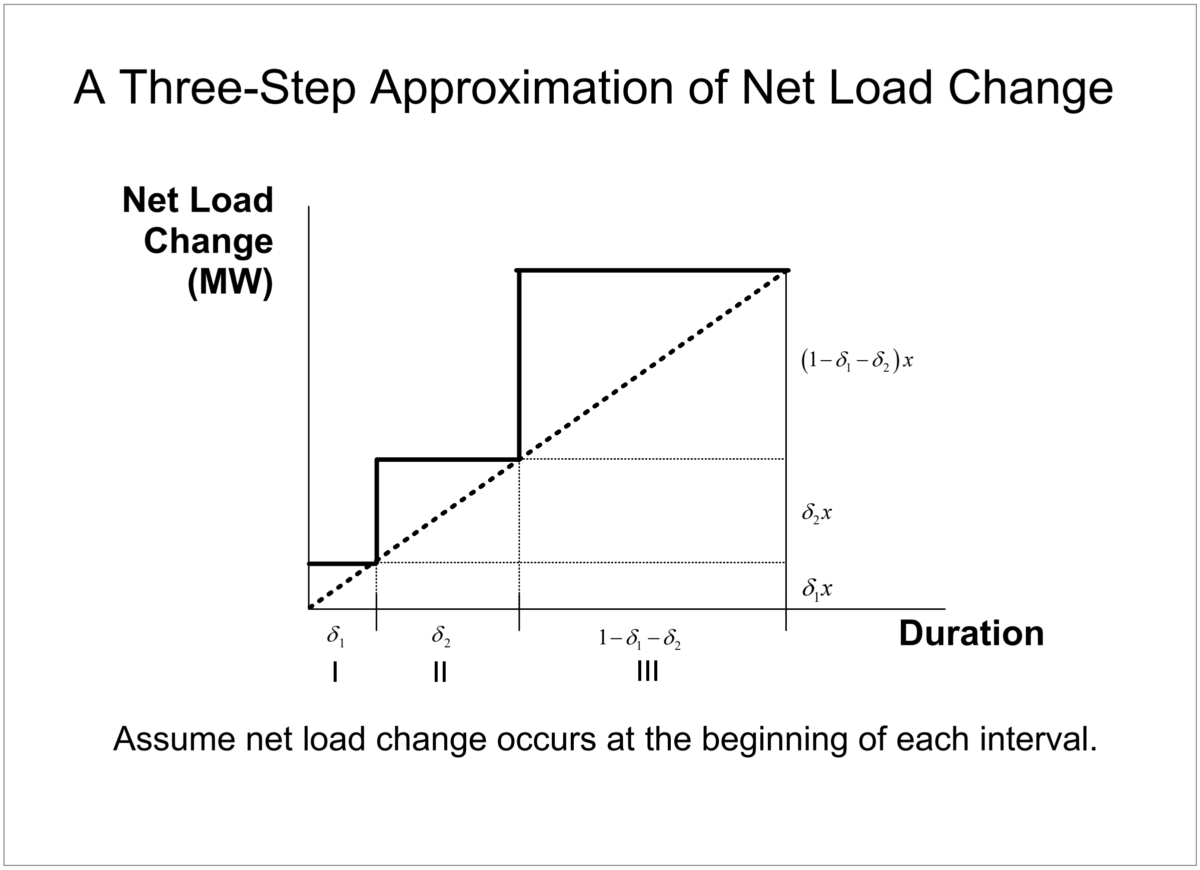
Hence, if all reserve levels are below the MRRs, then the SR Market Clearing Price is the sum of the three penalty factors, the NSR Market Clearing Price is the sum the PR and 30-min penalty factors, and the SecR Market Clearing Price is the 30-min penalty factor. Otherwise the prices incorporate the respective *Lolps*.

VOLL Approach with Three Reserve Types

The PJM approach uses three types reserves corresponding to three different response intervals. Comparison with the methodology above for the cascaded model, begins with the observation that the price approximation in the *VOLL* approach assumes that the different reserve products apply over different periods. This ties back to a characterization of the timing of and probability distribution for revealed information about the net load change.

The assumption here is that there are three periods defined as: I, 0-10 minutes; II, 10-30 minutes; and III, 30-60 minutes. The net load change approximation follows a step function as shown in the figure. The increment in reserve requirements occurs at the beginning of each interval, and that change persists over the rest of the one-hour period.

Figure 9



The first step SR reserves can meet the requirement for all three periods, but the marginal value of the reserves is different for each of the three periods. Similarly, for the other reserve types.

The *Lolp* approximation for each period starts with the hourly distribution of forecast errors. In order to treat the three periods as independent in adding up the prices, the estimated mean and variance for each period is scaled to be consistent with the variance for the full look ahead period. Suppose that there are three underlying variables y_I, y_{II}, y_{III} representing the incremental net load change in the three intervals. Further assume that the three variables have a common underlying distribution for a variable z but are proportional to the size of the interval. Then, assuming independence and with x the net load change over the full three intervals, we have:

$$
\begin{aligned}
&E(y_I) = E(\delta_1 z) = \delta_1 E(z), \\
&E(y_{II}) = E(\delta_2 z) = \delta_2 E(z), \\
&E(y_{III}) = E(\delta_3 z) = \delta_3 E(z). \\
&Var(y_I) = Var(\delta_1 z) = \delta_1^2 Var(z), \\
&Var(y_{II}) = Var(\delta_2 z) = \delta_2^2 Var(z), \\
&Var(y_{III}) = Var(\delta_3 z) = \delta_3^2 Var(z), \\
&E(z) = E(y_I + y_{II} + y_{III}) = E(x) = \mu.
\end{aligned}
$$

Imposing the independence assumptions, with the hour ahead standard deviation estimated as (σ), we have:

$$
\begin{aligned}
&Var(x) = Var(y_I + y_{II} + y_{III}) = Var(y_I) + Var(y_{II}) + Var(y_{II}) = \left(\delta_1^2 + \delta_2^2 + \delta_3^2\right) Var(z). \\
&Var(z) = \frac{Var(x)}{\sum \delta_i^2} = \frac{\sigma^2}{\sum \delta_i^2}.
\end{aligned}
$$

The distinction here is that the implied variance of the individual intervals is greater compared to the one-draw assumption, even though the total variance of the sum over the two intervals is the same. This is simply an impact of the square root law for the standard deviation of the sums of independent random variables.

Here the different distributions refer to the net change in load over the first interval, and over the sum of the relevant intervals. The distribution over the final sum is just the same distribution for the whole period that was used above. Then:

$$
y_I \sim Lolp_I, \quad y_I + y_{II} \sim Lolp_{I+II}, \quad y_I + y_{II} + y_{III} \sim Lolp_{I+II+III}.
$$

A workable approximation would be to utilize the normal distribution for the net load change.

As before, there would be an adjustment to deal with the minimum reserve to meet the max contingency. The revised formulation would include:

$$
\begin{aligned}
&\pi_{SR}(r_{SR}) = \begin{Bmatrix} Lolp_I(r_{SR} - X_{SR}), & r_{SR} - X_{SR} \geq 0 \\ 1, & r_{SR} - X_{SR} < 0 \end{Bmatrix} \\
&\pi_{PR}(r_{SR}, r_{PR}) = \begin{Bmatrix} Lolp_{I+II}(r_{SR} + r_{PR} - X_{PR}), & r_{SR} + r_{PR} - X_{PR} \geq 0 \\ 1, & r_{SR} + r_{PR} - X_{PR} < 0 \end{Bmatrix} \\
&\pi_{30}(r_{SR}, r_{PR}, r_{30}) = \begin{Bmatrix} Lolp_{I+II+III}(r_{SR} + r_{PR} + r_{30} - X_{30}), & r_{SR} + r_{PR} + r_{30} - X_{30} \geq 0 \\ 1, & r_{SR} + r_{PR} + r_{30} - X_{30} < 0 \end{Bmatrix}.
\end{aligned}
$$

With the *VOLL* (adjusted for marginal generation cost) of v, the corresponding prices would be:

$$P_{SR}\left(r_{SR},r_{PR},r_{30}\right)=v*\left(\delta_1*\pi_{SR}\left(r_{SR}\right)+\delta_2*\pi_{PR}\left(r_{SR},r_{PR}\right)+\delta_3*\pi_{30}\left(r_{SR},r_{PR},r_{30}\right)\right)$$
$$P_{PR}\left(r_{SR},r_{PR},r_{30}\right)=v*\left(\delta_2*\pi_{PR}\left(r_{SR},r_{PR}\right)+\delta_3*\pi_{30}\left(r_{SR},r_{PR},r_{30}\right)\right)$$
$$P_{30}\left(r_{SR},r_{PR},r_{30}\right)=v*\delta_3*\pi_{30}\left(r_{SR},r_{PR},r_{30}\right).$$

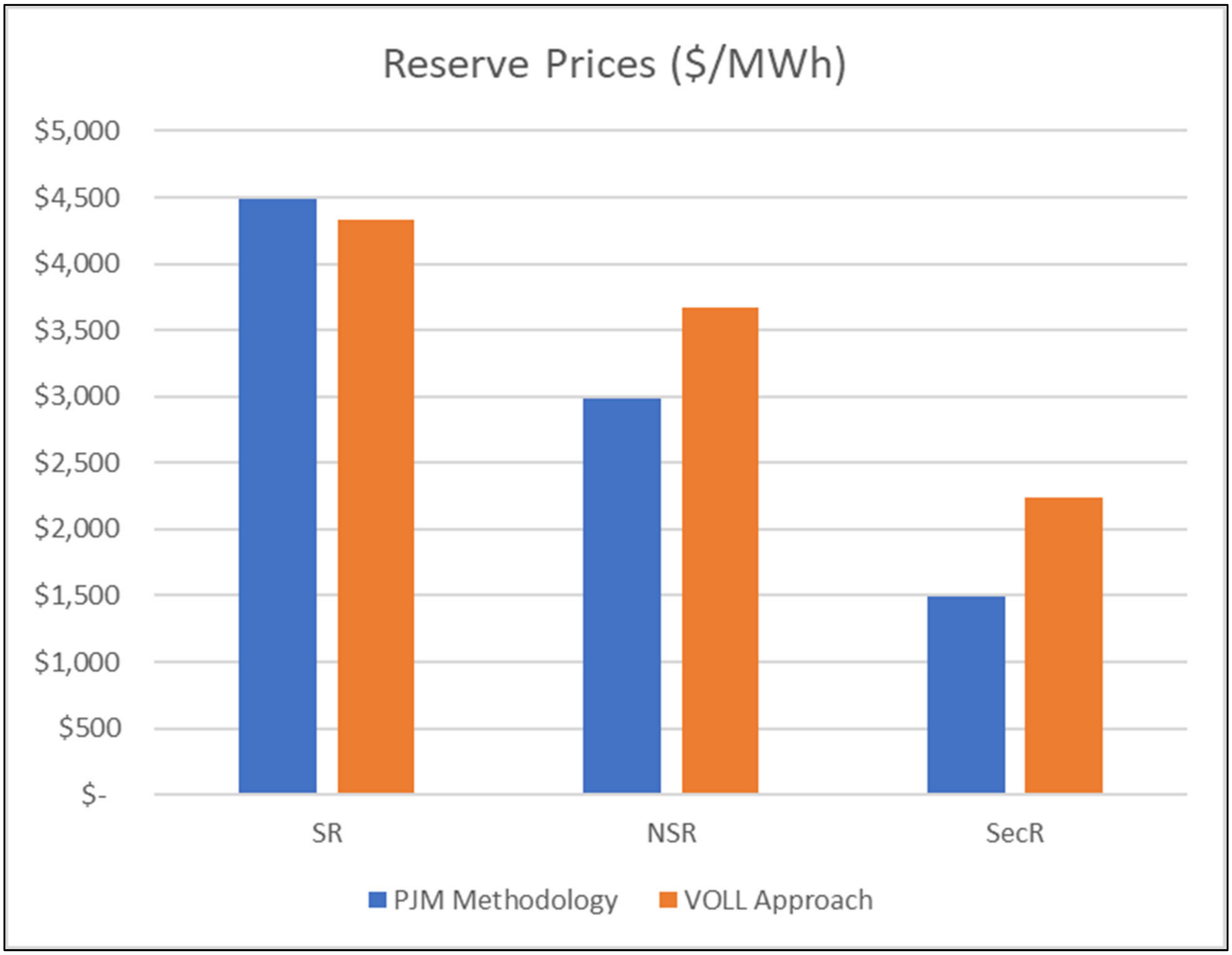
Here $v*\delta_i$ looks similar to the PF_i in the PJM proposal; however, there is also a difference in the approximation of the probabilities and the duration of the period for avoided emergency action.

Illustrative Comparison

For purposes of an illustrative calculation, consider an empirical distribution for actual reserves, viewed from the standpoint of a 30-minute look-ahead, as having approximately a mean of 400 MW and a standard deviation of 600 MW. For the illustration here, double these values and assume this defines the actual reserve distribution with a 60-minute look-ahead forecast, using the Normal Distribution approximation to simplify the calculation. Assume the three MRR values are given, one for each type of reserves. The price for anchoring the demand curve for each of the three individual reserve products at the level equal to the corresponding MRR, would be the price corresponding to the highest probability of this type of reserves falling below its MRR. Therefore, these would be the highest reserve prices for reserves at or above the MRR levels. The corresponding *VOLL* is $6000, equal to the sum of the three assumed penalty factors of $2000.

With these assumptions, Figure 10 summarizes the three reserve prices compared between the two approximation methods. The prices are different principally due to differences in the scaling assumptions for the three different periods. Both approximations are consistent with a significant estimate of the *VOLL*, compared to the existing PJM $850/MWh maximum penalty factor.

Figure 10



These are the unconstrained prices for the RTO region. If there is a constrained region, the PJM increment in prices follows the same methodology. Hence, if the regional parameters for load, probability distributions, and MRR scale proportionally, then for the same level of product reserves at the MRR in the constrained zone, the resulting product reserves prices would double the levels shown in Figure 10, consistent with a *VOLL* in the constrained zone of $12,000/MWh.

Multiple Zones and Locational Operating Reserves

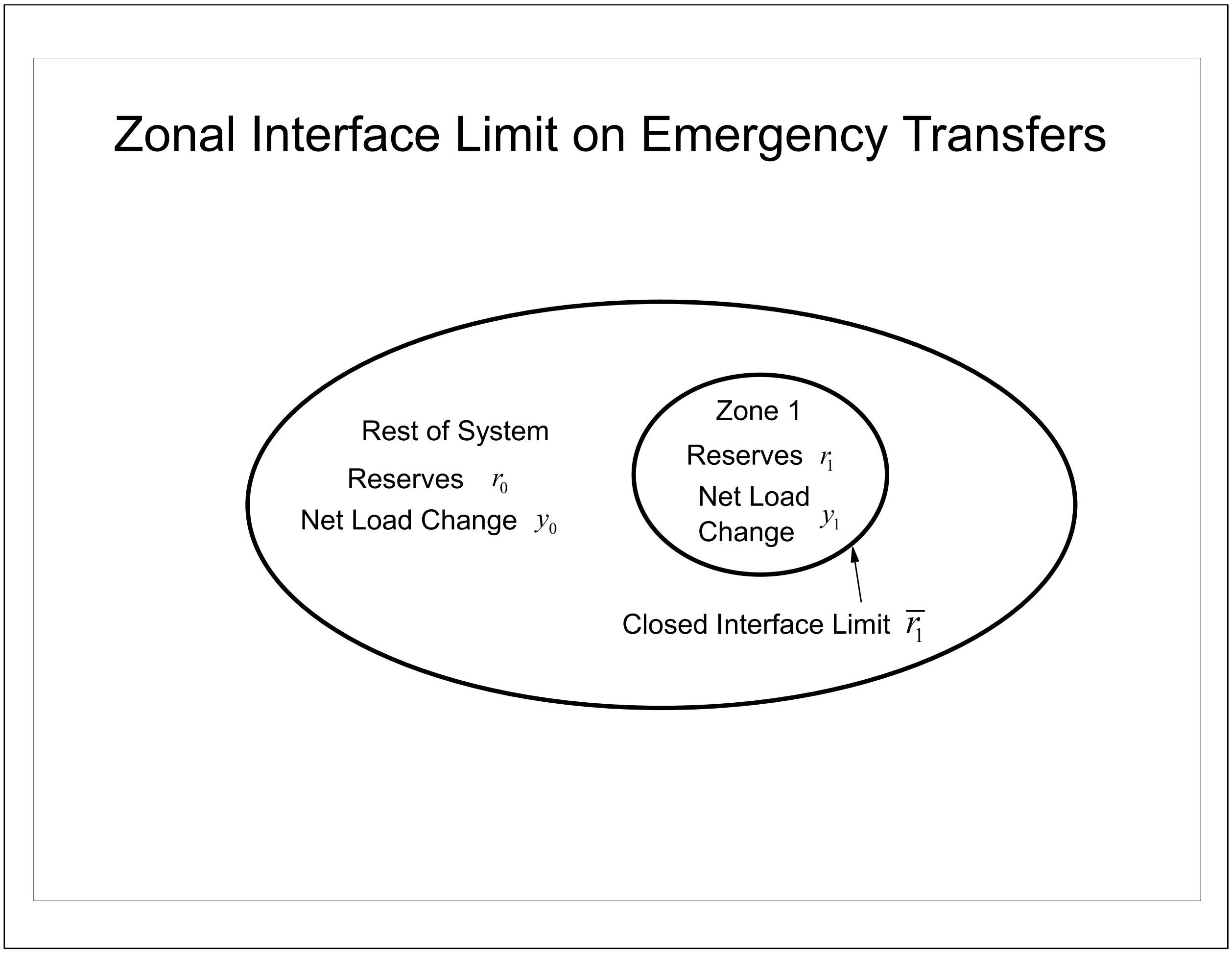
The assumption that there is a single system-wide operating reserve benefit may need to be modified. The steady-state constraints of transmission limits and loop flows apply to the base dispatch. These constraints need not apply necessarily to the short-term use of operating reserves in a stressed situation. However, it is possible a set of transmission limits includes locational constraints on operating reserves. An approach for modeling locational operating reserves is to define a nested zone and the associated interface constraint that limits the emergency movement of power. This constraint then separates the reserves inside and outside the constrained region and defines their interaction.

The task is to define a locational operating reserve model that approximates and prices the dispatch decisions made by operators. To illustrate, consider the simplest case with one constrained zone and the rest of the system. The reserves are defined separately and there is a known transfer limit for the closed interface between the constrained zone and the rest of the system. This zonal interface constraint would be analogous to the Capacity Emergency Transfer Limit in PJM planning models (PJM, 2016). The probability distribution for net load changes would be estimated separately for locations inside and outside the zone. The zonal requirements for operating reserves interact with energy and economic dispatch, incorporate local interface constraints, and provide compatible short-term prices for operating reserves and interface capacity. This basic argument leads to a simple numerical model that can incorporate multiple embedded zones and interface constraints and be implemented with the co-optimization framework for energy and reserves.

An outline of the basic framework illustrates the representation of locational operating reserve demand curves. Adaptation of a single system ORDC to address locational reserve requirements raises additional issues.

To illustrate, consider the simplest case with one constrained zone and the rest of the system, as in Figure 11. The regions are nested, meaning that the locational requirement is a subset of the system requirement. The reserves are defined separately for the system and within the local region, but they interact and there is a known transfer limit for the closed interface between the constrained zone and the rest of the system.

Figure 11



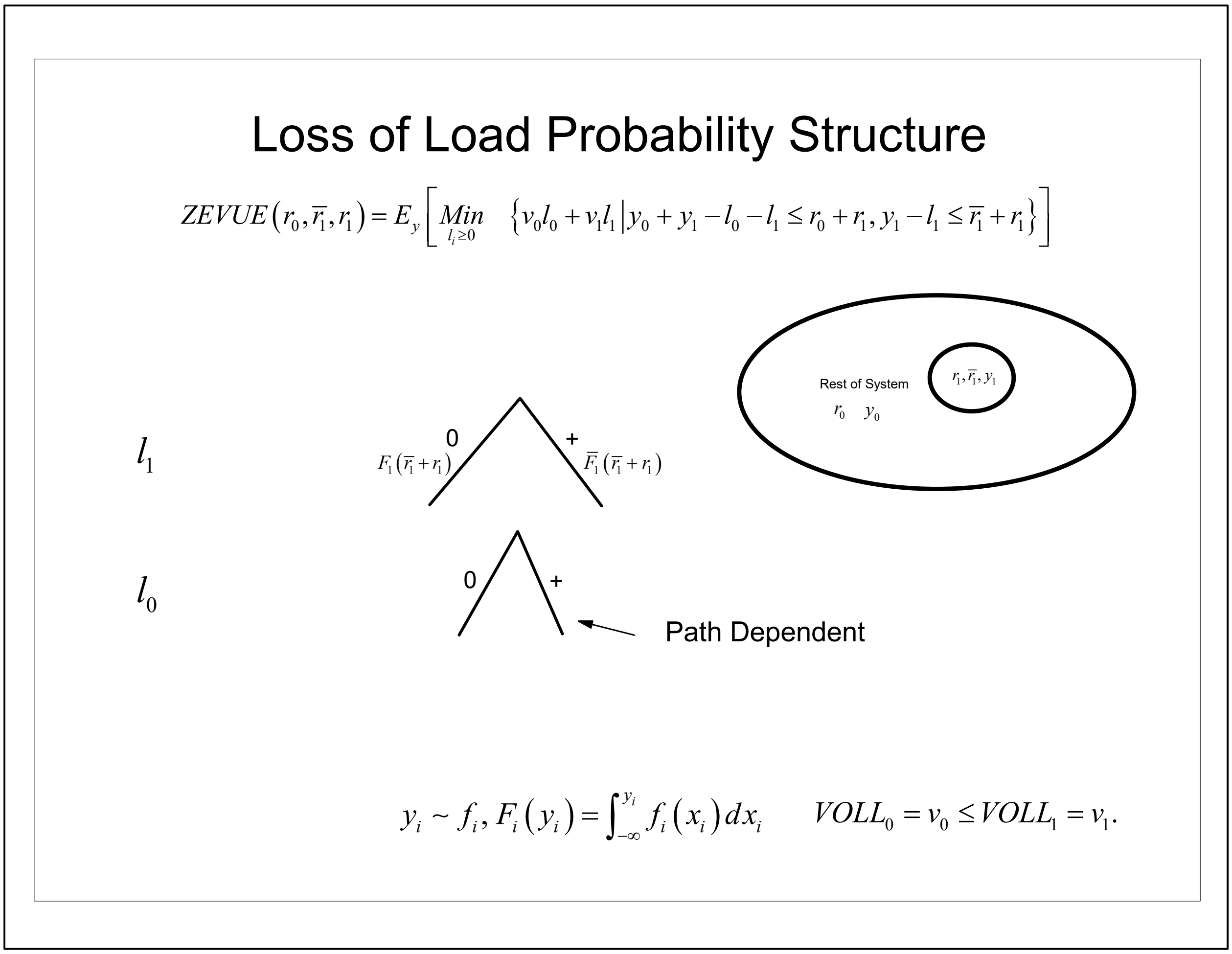
The interior Zone 1 has a known level of reserves r_1. The distribution of net load changes within the zone is $y_1 \sim f_1$. The closed interface defines the interior zone by a limit $\bar{r}_1$ on the aggregate power flow from the rest of the system into the local zone. This limit will interact with the dispatch power flow. The rest of system has a known level of reserves r_0 and a distribution of net load changes outside of the interior zone, $y_0 \sim f_0$. These are treated as independent distributions. Independence is not a strong assumption. The dispatch load forecast might be strongly interacting across the zones, but the unanticipated deviations from the forecast can be viewed as approximately independent across the zones.

The distributions for each net load change have corresponding cumulative distributions.

$$y_0 \sim f_0, y_1 \sim f_1, \quad F_o\left(y_0\right) = \int_{-\infty}^{y_0} f_0\left(x_0\right) dx_0, \quad F_1\left(y_1\right) = \int_{-\infty}^{y_1} f_1\left(x_1\right) dx_1.$$

The zonal expected value of unserved energy (ZEVUE) would be an added component of the objective function in economic dispatch. A simplifying assumption is only one type of emergency action in each zone. Further, assume that the $v_1 = VOLL_1 - \overline{c}_1$ is at least as great as the corresponding value in the rest of the system, $v_0 = VOLL_0 - \overline{c}_0$. With this assumption, we adopt the protocol that gives priority to meeting load deviations inside the constrained zone relative to the rest of the system, and the ex-post dispatch will have a simple structure. In Figure 12, the first priority is to meet the net change of load within the interior zone. The unserved load l_i will be penalized at the respective value of loss load.

Figure 12



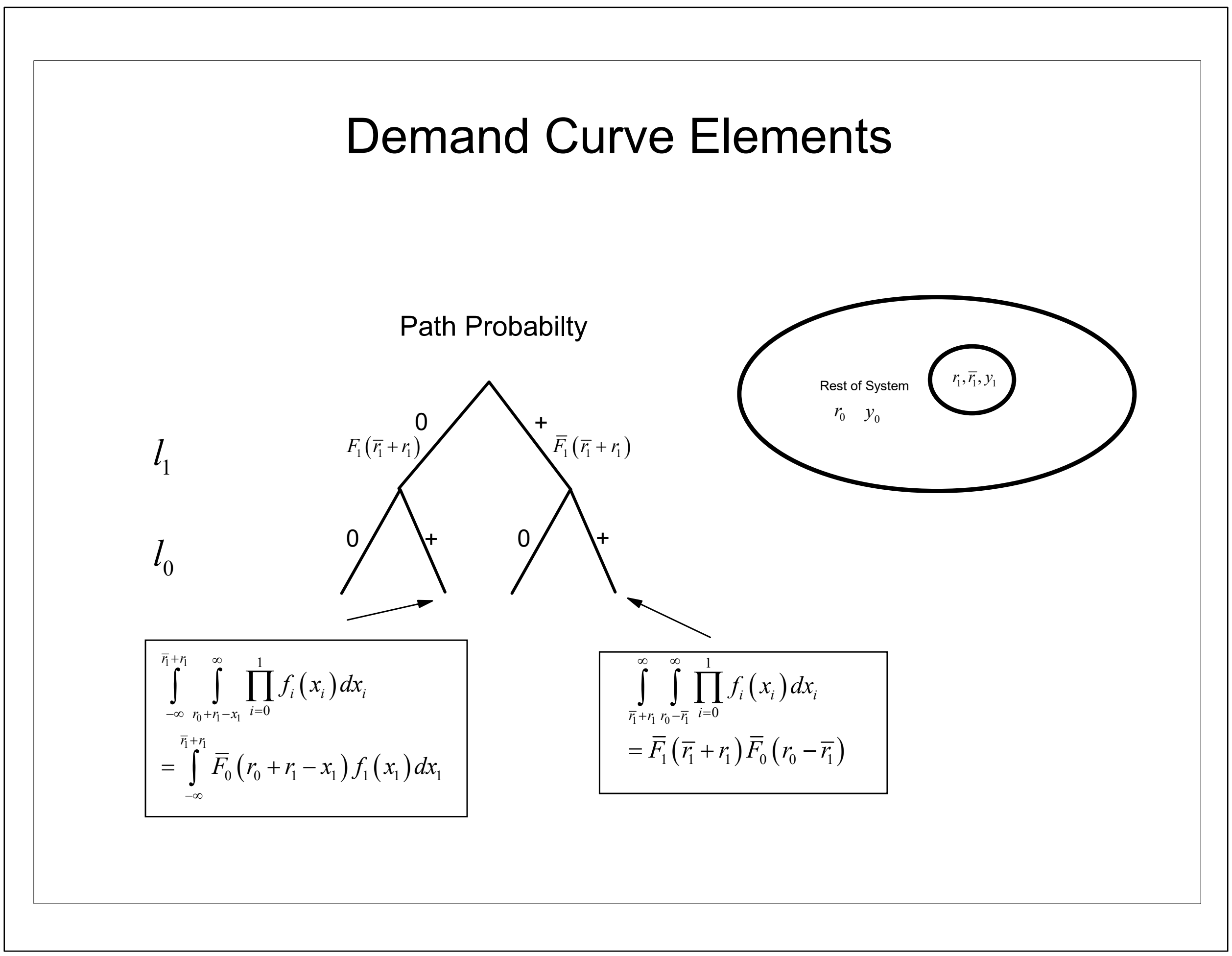
The basic problem determines the configuration of lost load and the ZEVUE.

$$ZEVUE(r_0, \overline{r}_1, r_1) = E_y \left[\underset{l_i \ge 0}{Min} \quad \left\{ v_0 l_0 + v_1 l_1 \middle| y_0 + y_1 - l_0 - l_1 \le r_0 + r_1, y_1 - l_1 \le \overline{r}_1 + r_1 \right\} \right].$$

The derivatives of ZEVUE define the demand curves for operating reserves. Given the simplifying assumptions, the tree structure in Figure 13 illustrates the steps to construct these prices for reserves and the interface constraint. At the top of the branching is the amount of lost load in region 1. This is either zero or positive, and the probabilities on the branches apply for these conditions. The key is the limit on internal reserves and the interface limit. If the net change in load inside the zone is greater than $\overline{r}_1 + r_1$ then all the reserves inside the region and all that could move from outside the region would be utilized, and there would be loss of load inside the region. This occurs with probability $\overline{F}_1(\overline{r}_1 + r_1) = 1 - F_1(\overline{r}_1 + r_1)$. Likewise, the left branch with $l_1 = 0$ has probability $F_1(\overline{r}_1 + r_1)$.

The probabilities for the next level down are path dependent, but the calculation is conceptually straightforward.

Figure 13



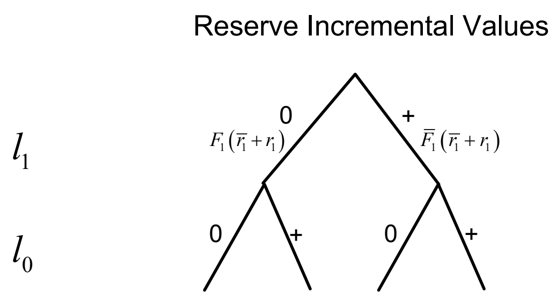
For example, in Figure 13, given that we are on the path with $l_1 \geq 0$, the reserves available for the rest of the region must be the total rest-of-system reserves minus the interface capacity, because the interface capacity is being used to meet requirements in the constrained zone. The conditional probability of this case is $\overline{F}_0\left(r_0-\overline{r}_1\right)$. Hence, the probability for the full path is $\overline{F}_1\left(\overline{r}_1+r_1\right)\overline{F}_0\left(r_0-\overline{r}_1\right)$, as shown in Figure 13. A similar argument applies to the other paths.

The full ZEVUE is difficult to characterize and calculate. However, inspection of the possible configurations of outages reveals the marginal zonal values of unserved energy, which define the locational demand curves for operating reserves.

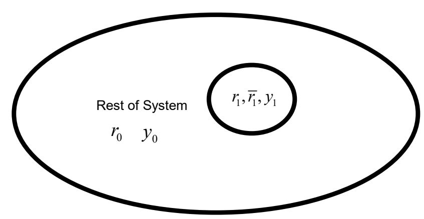
Figure 14

Loss of Load Values

$$ZEVUE\left(r_0,\overline{r}_1,r_1\right)=E_y\left[\underset{l_i\geq 0}{Min}\quad\left\{v_0l_0+v_1l_1\left|y_0+y_1-l_0-l_1\leq r_0+r_1, y_1-l_1\leq\overline{r}_1+r_1\right.\right\}\right]$$



r_1	0	v_0	v_1	v_1
$\overline{r}_1$	0	0	v_1	v_1-v_0
r_0	0	v_0	0	v_0

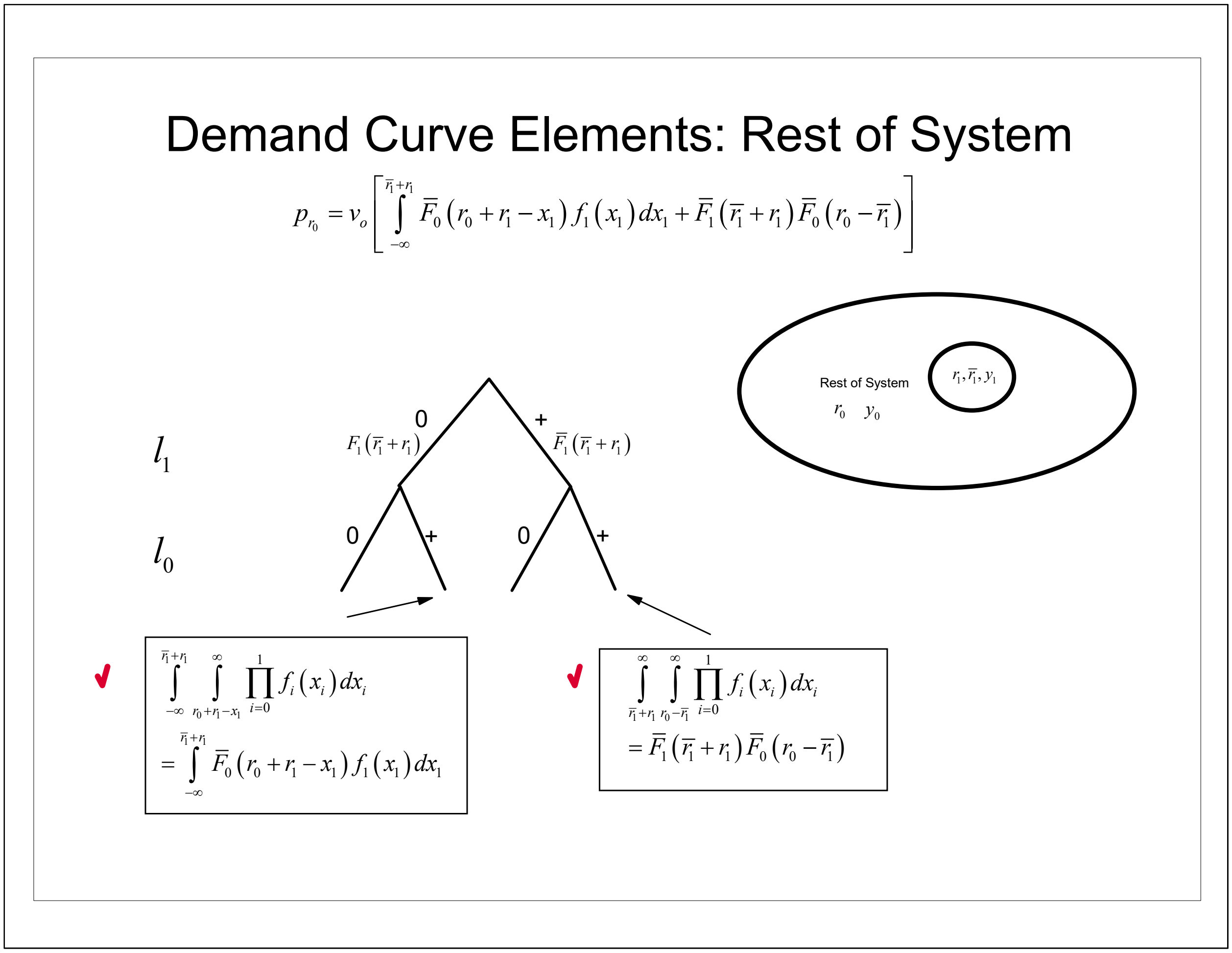


The table in Figure 14 illustrates the reserve incremental values on each of the paths. For example, on the right-most path, the marginal value of reserves inside the region is v_1 and the marginal value in the rest of the region is v_0, because there are load losses in both regions. On this same path, the marginal value of incremental interface capacity is the increased flow from outside to inside, which would produce net benefit v_1-v_0. Similar arguments apply to the other elements of the table. And with this table we see the paths where values are non-zero and we need the associated path probabilities.

Combining the marginal values and probabilities for each path in the tree yields the corresponding value which defines the expected marginal value of the increment of reserves or interface capacity.

For example, Figure 15 shows the demand curve for the price of reserves in the rest of the system, with the check marks showing the relevant paths.

Figure 15



The demand is a function of all three elements and the associated probability distributions.

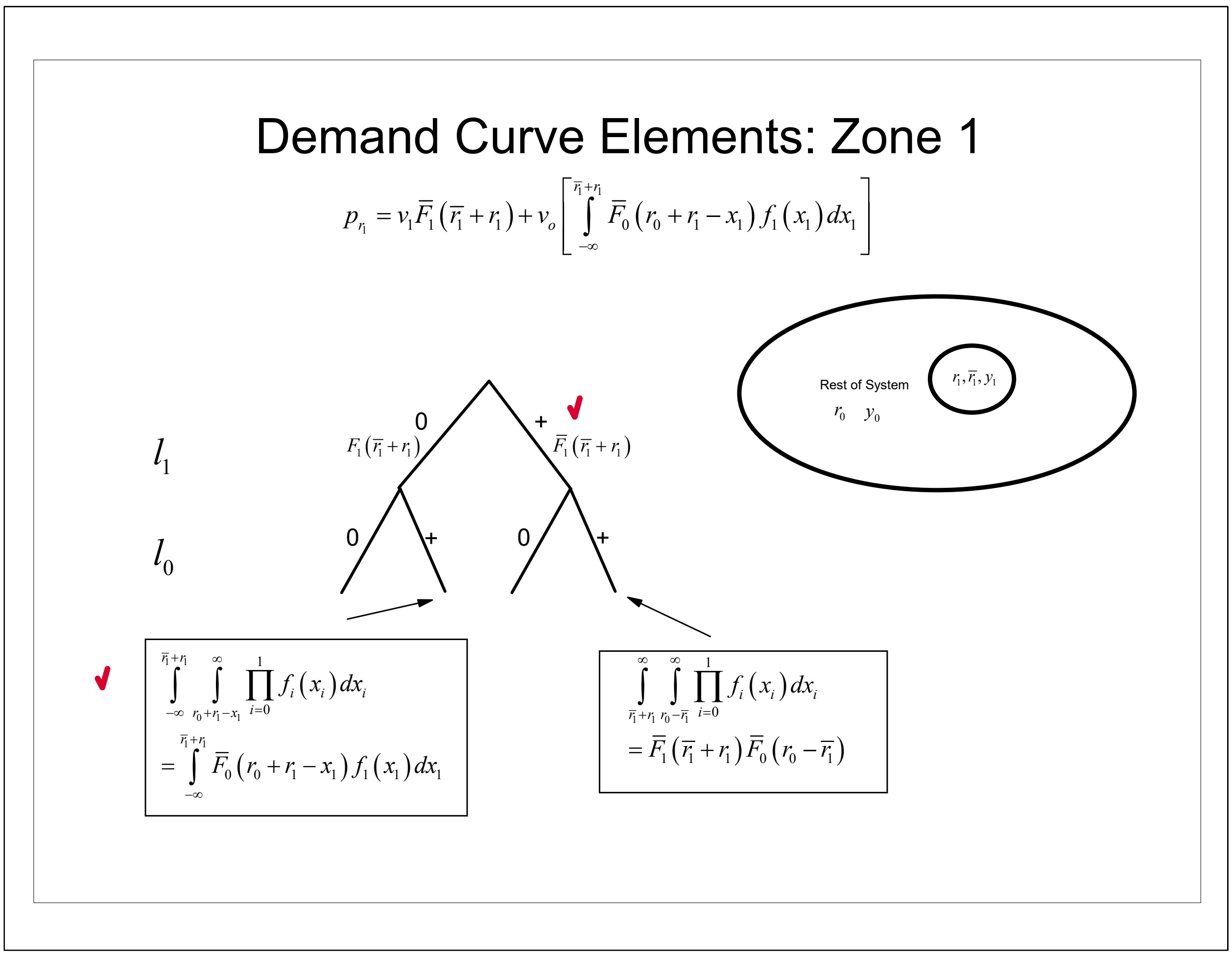
(17) $$p_{r_0} = v_o \left[\int_{-\infty}^{\bar{r}_1 + r_1} \bar{F}_0 (r_0 + r_1 - x_1) f_1 (x_1) dx_1 + \bar{F}_1 (\bar{r}_1 + r_1) \bar{F}_0 (r_0 - \bar{r}_1) \right].$$

Since all these elements are known, it is a simple calculation to trace out the elements of the demand curve to include in the dispatch objective function and solve for energy and reserves.

There is a similar story for the price of reserves inside the local zone in Figure 16.

(18) $$p_{r_1} = v_1 \overline{F}_1(\overline{r}_1 + r_1) + v_o \left[\int_{-\infty}^{\overline{r}_1 + r_1} \overline{F}_0(r_0 + r_1 - x_1) f_1(x_1) dx_1 \right].$$

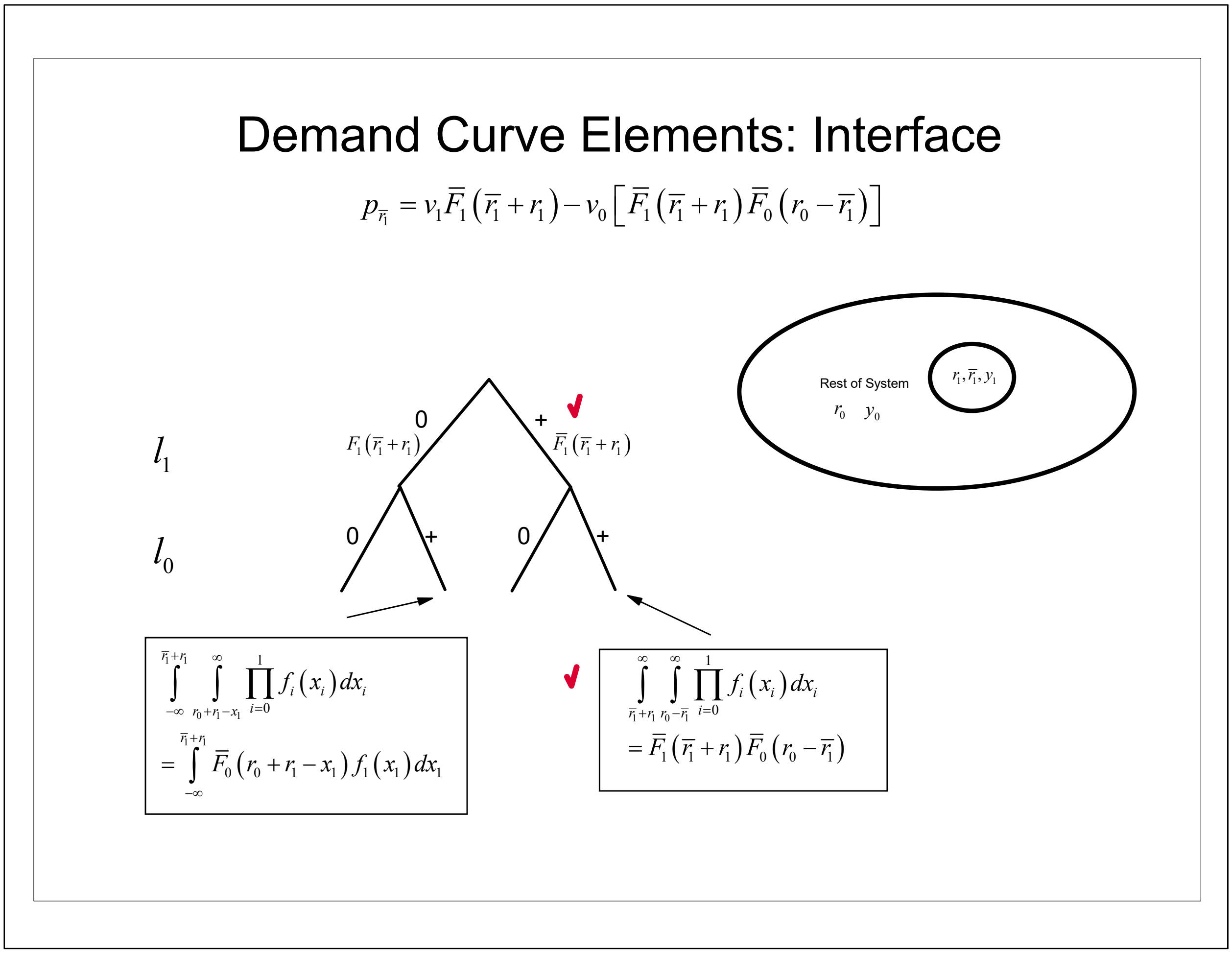
Figure 16



Finally, the analysis extends to the demand curve for interface capacity in Figure 17.

$$p_{\bar{r}_1} = v_1 \bar{F}_1\left(\bar{r}_1 + r_1\right) - v_0\left[\bar{F}_1\left(\bar{r}_1 + r_1\right)\bar{F}_0\left(r_0 - \bar{r}_1\right)\right].$$

Figure 17

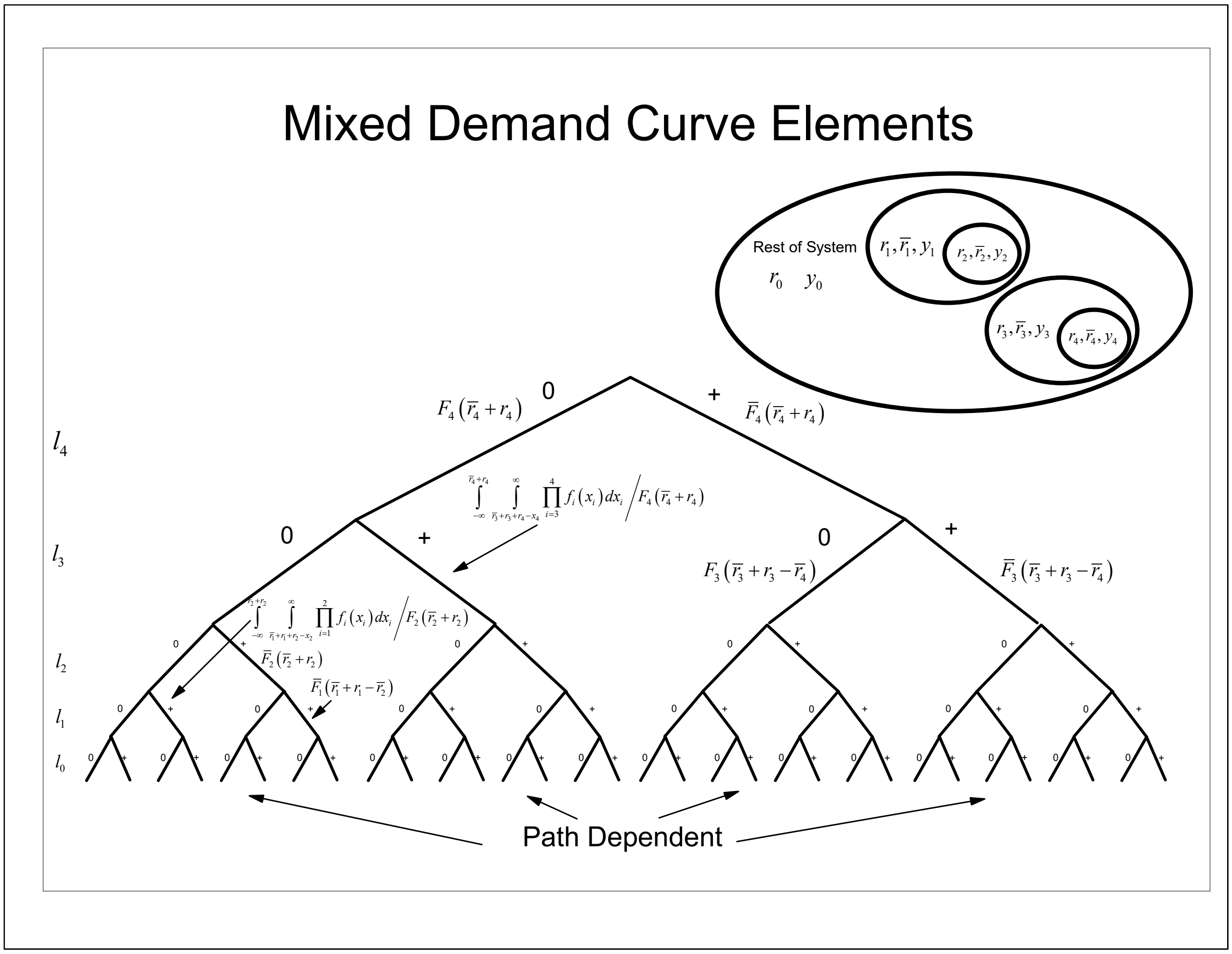


Although the values for each reserve differ in each case on the tree, the expected values defining the reserve prices satisfy $p_{r_i} = p_{r_0} + p_{\bar{r}_1}$.

The extensions to include multiple zones or further nested zones would follow a similar logic. At some stage the "curse of dimensionality" would make the size of the probability tree too large to maintain computational tractability. However, the simple structure could well accommodate a few zones.

The illustration in Figure 18 suggests the basic structure with parallel and nested zones. On each path there would be an algorithm for numerically integrating the probabilities to obtain the path weights. And there would be a corresponding table of marginal values of each zonal reserve and interface constraint (Hogan, 2010). The resulting demand curves could be included in the dispatch logic.

Figure 18

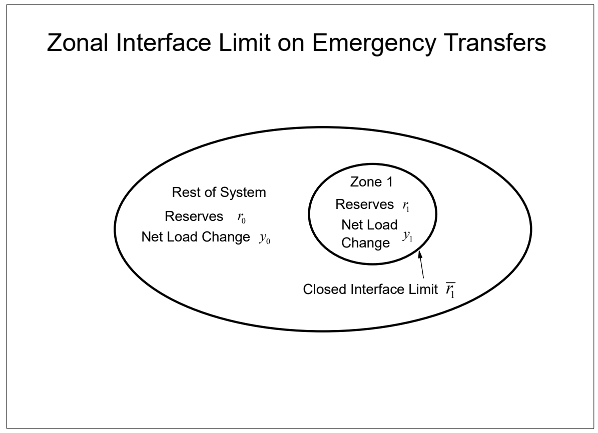


Constrained Zone Example

An example illustrates the separable implementation of three locational reserve-related demand curves. The parameter assumptions and an assumed benchmark provide the components for the approximation.

	ROS	Zone
Expected Total (MW)	107.1	45.90
Std. Dev (MW)	488.99	209.57
VOLL ($/MWh)	7000	10000

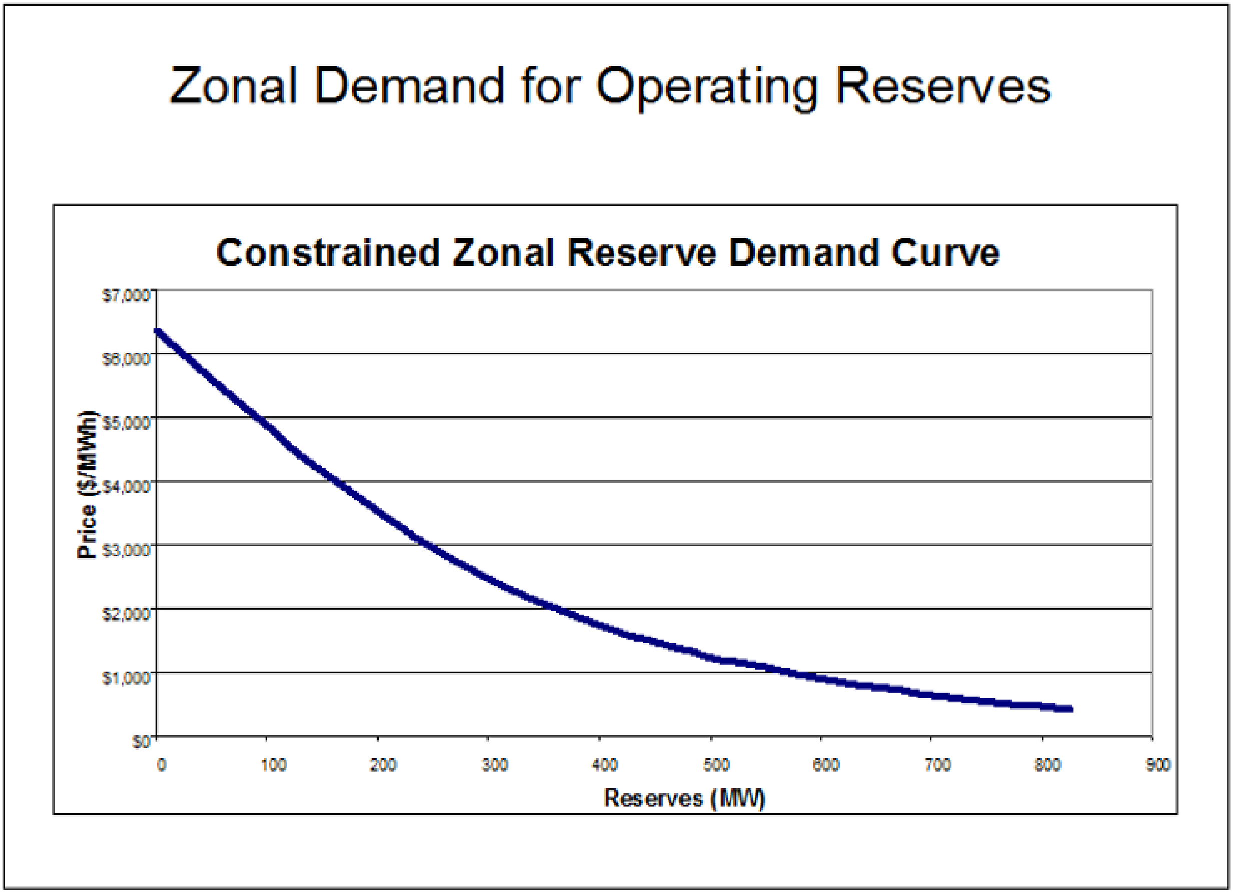
	ROS	Zone 1	Interface
Benchmark (MW)	160.65	68.85	45.90



With these assumptions, we can use the normal approximation of the net load changes to calculate the corresponding probabilities on each path and the resulting estimates of the reserve-related prices. For the price in the constrained zone we have:

$$p_{r_1} = v_1\left(1 - F_1\left(\overline{r}_1 + r_1\right)\right) + v_0\int_{-\infty}^{\overline{r}_1 + r_1}\left[1 - F_0\left(r_0 + r_1 - x_1\right)\right] f_1\left(x_1\right) dx_1.$$

Figure 19

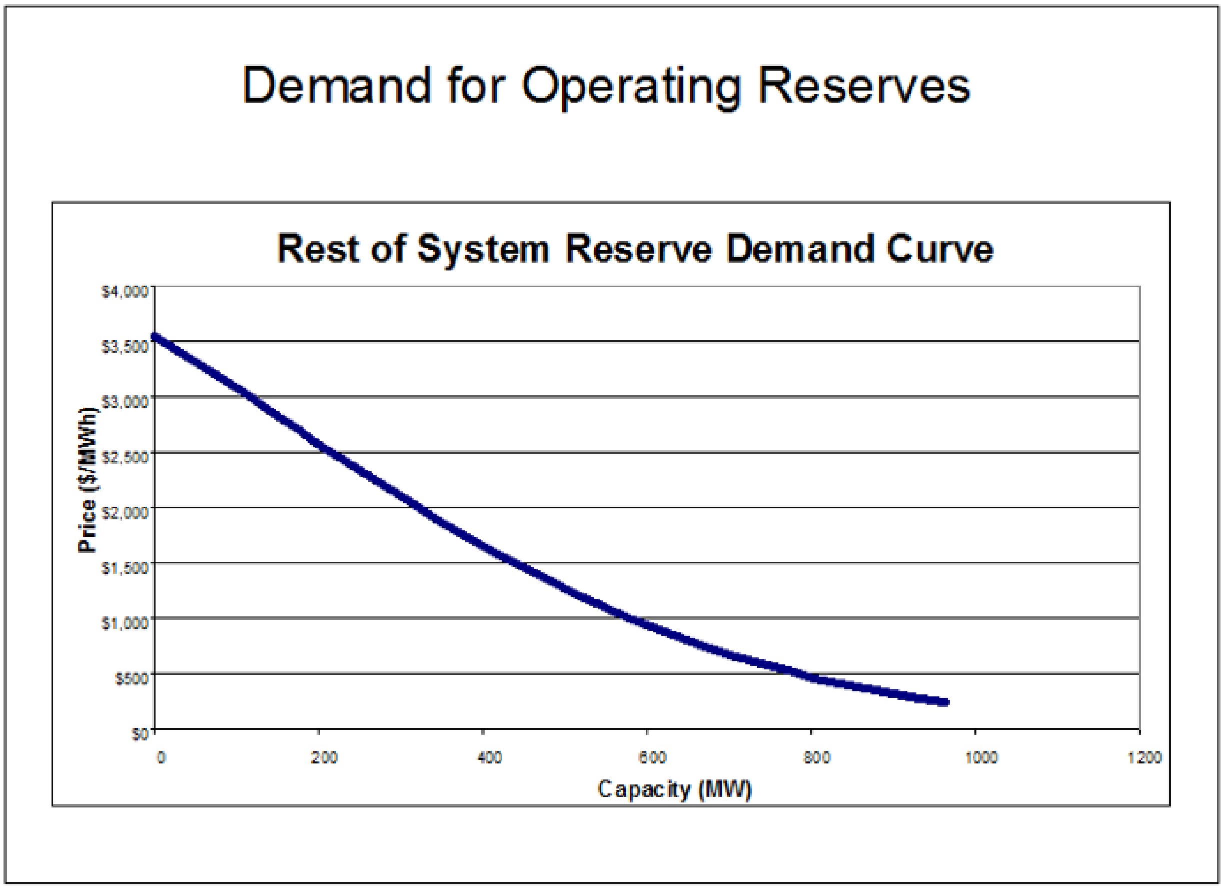


The maximal zonal price in Figure 19 is slightly over $6,000/MWh, determined by the value of lost load and the loss of load probability.

For the price of reserves on the rest of the system we have:

$$p_{r_0} = v_0 \int_{r_0 - \overline{r}_1}^{\infty} \left[1 - F_1 \left(r_0 + r_1 - x_0 \right) \right] f_0 \left(x_0 \right) dx_0 .$$

Figure 20

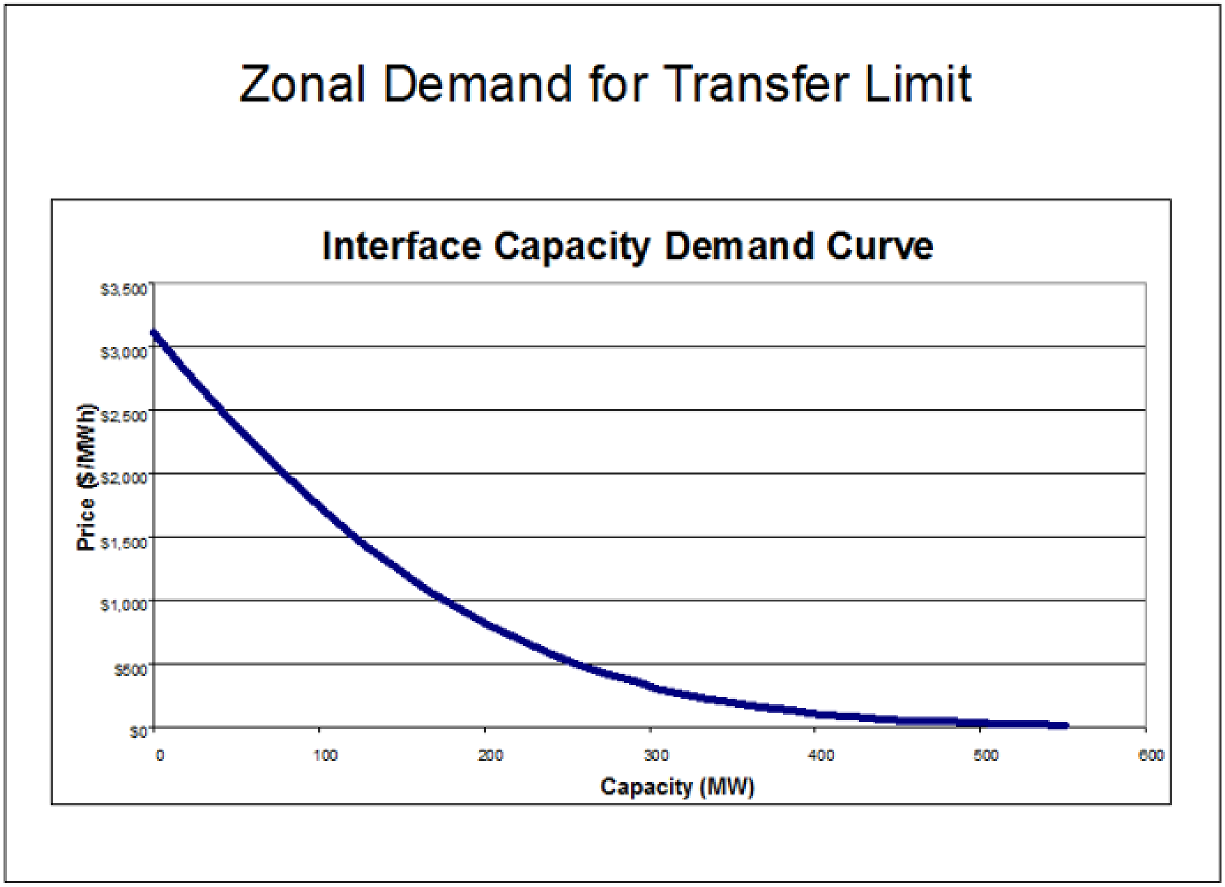


The maximal rest of the system price in Figure 20 is slightly over $3,500/MWh, determined by the lower value of lost load and the loss of load probability.

For the interface constraint, the price is:

$$p_{\overline{r}_1} = v_1 \left(1 - F_1 \left(\overline{r}_1 + r_1 \right) \right) - v_0 \left(1 - F_0 \left(r_0 - \overline{r}_1 \right) \right) \left(1 - F_1 \left(\overline{r}_1 + r_1 \right) \right).$$

Figure 21



The maximal interface capacity scarcity price in Figure 21 is slightly over $3,000/MWh, determined by the differences in the values of lost load and the loss of load probability.

Separable Demand Curve Approximation

In all cases, the price for reserves and the interface constraint are functions of all the reserve components. Furthermore, with constrained zones the relationships are not separable. One implication is that the scarcity prices are not simply additive, and the highest price in a region can never be higher than the value of lost load for that region. However, unlike the case of a single regional ORDC, the construction of the counterpart of $R_k\left(r_k\right)$ requires more than simply integrating under the prices along a single dimension.

A requirement to construct a counterpart of (1) is to have integrated functions $\hat{R}_k\left(r_o,\overline{r}_1,r_1\right)$ such that at the optimal solution $\left(r_0^*,\overline{r}_1^*,r_1^*\right)$ the derivatives equal the respective prices. For the single constrained zone and rest of the system, an example of a separable version of such a function would be:

$$
\begin{aligned}
\hat{R}(r_0) &= \int_0^{r_0} p_{r_0}\left(x, \overline{r}_1^*, r_1^*\right) dx \\
\hat{R}(\overline{r}_1) &= \int_0^{\overline{r}_1} p_{\overline{r}_1}\left(r_0^*, x, r_1^*\right) dx \\
\hat{R}(r_1) &= \int_0^{r_1} p_{r_1}\left(r_0^*, \overline{r}_1^*, x\right) dx
\end{aligned} \quad . \tag{19}
$$

Implementation of these approximations utilizes a benchmark estimate of a reasonable version of the dispatch solution. The better the estimate, the better the approximation. Iteration on the estimate could be combined with the dispatch search algorithm in a manner that would implement the path model numerically without significant computation difficult.

At the equilibrium solution, the expected prices of reserves and interface constraints would satisfy the condition that $p_{r_i} = p_{r_0} + p_{\overline{r}_1}$. However, this is not true if separable demand curves employ a benchmark that is not equal to the equilibrium solution for the co-optimized dispatch. To address this property, an alternative approach would keep the separability assumptions in terms of intermediate variables, while enforcing the marginal condition for the relationship of the reserve and interface prices.

A. Select a representative benchmark, $\left(r_0^*, \overline{r}_1^*, r_1^*\right)$.
B. Use the ORDC model off-line to produce the initial "separable" approximations: $p_{r_0}\left(x, \overline{r}_1^*, r_1^*\right)$ and $p_{r_1}\left(r_0^*, \overline{r}_1^*, x\right)$.
C. Construct a new variable $\Delta\overline{r}_1 = \overline{r}_1 - \overline{r}_1^*$, and these "incremental interface" variables are included in the energy co-optimization model along with the various reserve values.
D. Implement the reserve component as the pricing model: $p_{r_0}\left(r_0 - \Delta\overline{r}_1, \overline{r}_1^*, r_1^*\right)$ and $p_{r_1}\left(r_0^*, \overline{r}_1^*, r_1 + \Delta\overline{r}_1\right)$.

In other words, the co-optimization objective function drops the explicit value for the interface constraint and constructs the values for the two types of reserves as:

$$
\begin{aligned}
\hat{R}_0\left(r_0 - \Delta\overline{r}_1\right) &= \int_0^{r_0 - \Delta\overline{r}_1} p_{r_0}\left(x, \overline{r}_1^*, r_1^*\right) dx \\
\hat{R}_1\left(r_1 + \Delta\overline{r}_1\right) &= \int_0^{r_1 + \Delta\overline{r}_1} p_{r_1}\left(r_0^*, \overline{r}_1^*, x\right) dx
\end{aligned} \quad . \tag{20}
$$

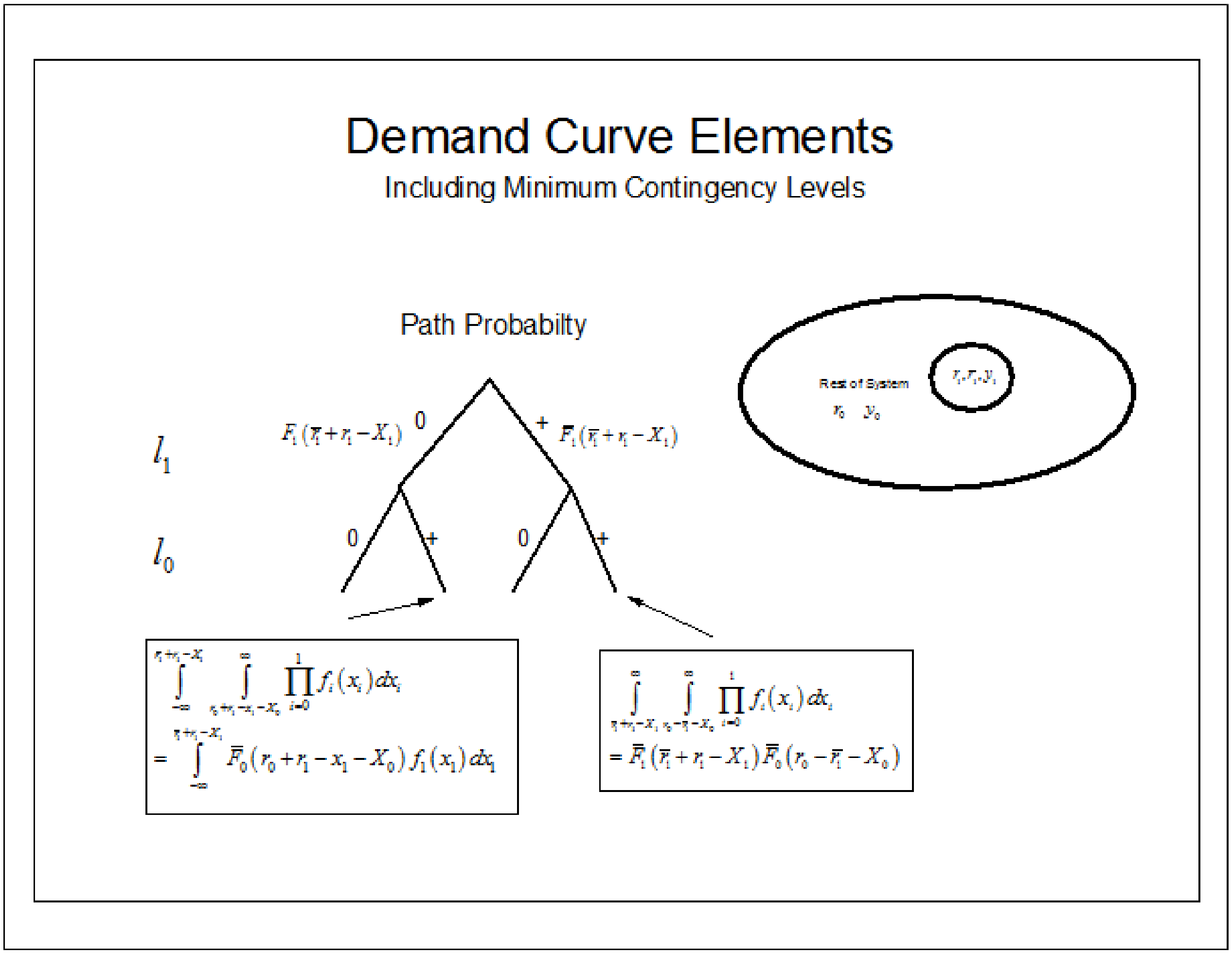
These approximate functions coincide with the reserve benefits when $\Delta\overline{r}_1 = 0$, and provide prices that imply an interface value equal to the difference of the constrained zone prices.

Zonal Contingency Requirements

The zonal demand curves would be modified to include minimum contingency requirements for emergency action such as a curtailment of load at the respective $VOLL$. The impact would be to change the path probability calculation to reflect the effect of the minimum contingency level.

For example, the revised version of the two critical path probabilities in Figure 13 would appear as in Figure 22.

Figure 22



This provides a calculation of the marginal values for ZEVUE with minimum contingency levels. The prices in the dispatch for these ex ante reserves will be the greater of the expected marginal value and the regional scarcity value v_i. In other words, the final price

$$p_{r_0} = \begin{cases} v_0 \int_{r_0 - \overline{r}_1}^{\infty} \left[1 - F_1\left(r_0 + r_1 - x_0\right)\right] f_0\left(x_0\right) dx_0, & \text{if } r_0 + r_1 \geq X_0 + X_1 \\ v_0, & \text{if } r_0 + r_1 < X_0 + X_1 \end{cases},$$

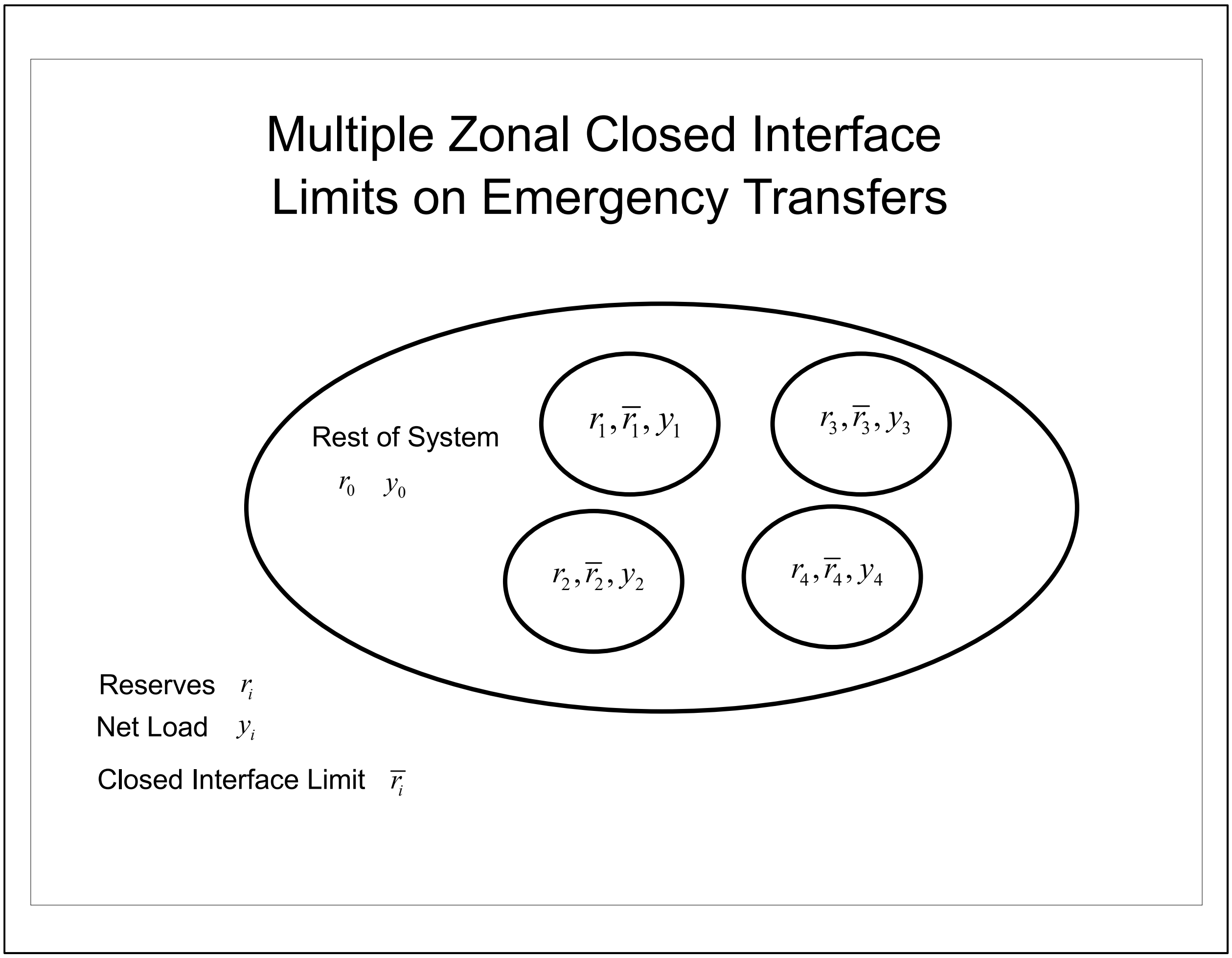
$$p_{r_1} = \begin{cases} v_1\left(1 - F_1\left(\overline{r}_1 + r_1\right)\right) - v_0\left(1 - F_0\left(r_0 - \overline{r}_1\right)\right)\left(1 - F_1\left(\overline{r}_1 + r_1\right)\right), & \text{if } \overline{r}_1 + r_1 \geq X_1 \\ v_1, & \text{if } \overline{r}_1 + r_1 < X_1 \end{cases}.$$

This method of incorporating the contingency minimum levels is a generalization of the simple shift of the ORDC in Figure 5.

Multiple Locations and Emergency Actions

The probability tree approach provides an analytical derivation of an ORDC. Adding more regions is straightforward. The associated probability tree grows, but the simplifying assumptions about relative values of loss load in regions preserve the basic structure of the tree.

Figure 23



Extension to include multiple types of emergency actions would require expansions of the event tree to incorporate different event combinations. In particular, the simple paths in the various probability trees arise because of the protocol that loss of load inside the constrained region takes precedence over that outside the region. If there are many emergency steps modeled, the optimization assumption could upset this protocol.

Adding multiple emergency actions within each zone is possible, but the relationships between the costs of emergency actions may allow some actions inside a constrained zone to be less costly than actions in the rest of the system. Hence, the probability tree grows through introduction of more possible system states implying different binding constraints.

The underlying logic remains for calculation of ZEVUE as a function of all the reserve and interface levels. An alternative computational approach would apply Monte Carlo simulation to estimate the expected values of the prices of interest. A generalization of the *ZEVUE* with parallel constrained zones would incorporate multiple zones ($i = 0, \ldots, n$), multiple emergency actions ($em_{ij}, j = 1, \ldots, m_i$) with costs net of the marginal energy dispatch level defining scarcity as ($v_{ij}, j = 1, \ldots, m_i$), and minimum contingency reserves (X_i). The constrained zones are all parallel to each other, as in Figure 23. The "rest of the system" is zone *0*. The emergency action upper bound is infinite for the first emergency action type ($j = 1$) in each region, which is assumed to be the most expensive action in each region and represents involuntary load curtailment at the region's net *VOLL*. The interface limits are on flows into constrained regions, but there is no limit modeled for the flows out of a constrained region.

Let:

r_i	: reserves , $i = 0, \ldots, n$
$\overline{r}_i$	: available transfer limit , $i = 1, \ldots, n$
em_{ij}	: emergency action , $i = 0, \ldots, n; j = 1, \ldots, m_i$
v_{ij}	: value of emergency action , $i = 0, \ldots, n; j = 1, \ldots, m_i$
X_i	: minimum required reserves , $i = 0, \ldots, n$
Kem_{ij}	: maximum of emergency action , $i = 0, \ldots, n; j = 1, \ldots, m_i$
$\hat{r}_i$	: utilized reserves , $i = 0, \ldots, n$
$\hat{\overline{r}}_i$	: utilized available transfer limit , $i = 1, \ldots, n$
y_i	: realized net load change , $i = 1, \ldots, n$

The zonal expected value of unserved energy is defined as before according to the program to minimize the cost of emergency action given each realization of a deviation in the net loads:

$$(21) \qquad ZEVUE(r, \overline{r}) = E_y \left\{ \begin{array}{ll} \underset{em, \hat{r}, \hat{\overline{r}} \geq 0}{Min} & \sum_i \sum_j v_{ij} em_{ij} \\ & \sum_i y_i - \sum_i \sum_j em_{ij} \leq \sum_i \hat{r}_i - \sum_i X_i \\ & y_i - \sum_j em_{ij} \leq \hat{\overline{r}}_i + \hat{r}_i - X_i, \quad i = 1, \ldots, n \\ & em_{ij} \leq Kem_{ij}, \quad i = 0, \ldots, n; j = 2, \ldots, m_i \\ & \hat{r} \leq r \qquad : p_r \\ & \hat{\overline{r}} \leq \overline{r} \qquad : p_{\overline{r}} \end{array} \right\}$$

For any give realization of the random change in the net load (y) that must be met by reserves, the dual variables from the reserve availability constraints define the realized marginal value of operating reserves and reserve transfer limits, $p_r, p_{\bar{r}}$. The expected values of these prices define the loss-of-load probability estimates in (17) and (18). This corresponds as well to the loss-of-load probability derivation of the single region ORDC. A Monte Carlo application for the optimization provides an offline method to estimate the respective expected values.

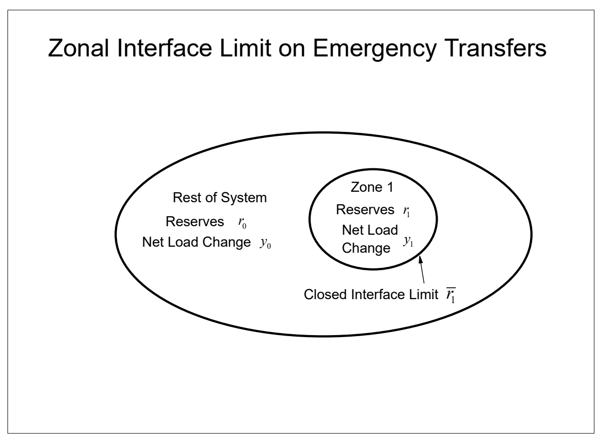
If the available reserves violate the minimum contingency levels ex ante, then emergency action would be needed to restore the reserves in the ex-ante dispatch. This corresponds to the minimum contingency step function derivation of the single region ORDC. The marginal value of this ex ante minimum condition can be solved by solving the value of unserved energy problem (21) with the load deviation set to zero ($y = 0$).

The resulting operating reserves values for the ORDC is the maximum of each of the two estimated reserve prices for each type of reserve. The marginal value of the interface capacity is the difference in the respective reserve prices.

The example in Figure 19 to Figure 21 illustrates an estimated ORDC with one constrained zone and only one type of emergency action. The parameter assumptions and an assumed benchmark provide the components for the approximation.

	ROS	Zone
Expected Total (MW)	107.1	45.90
Std. Dev (MW)	488.99	209.57
VOLL ($/MWh)	7000	10000

	ROS	Zone 1	Interface
Benchmark (MW)	160.65	68.85	45.90



Using the same input assumptions, the Monte Carlo replication produces prices at the benchmark as:

	ROS	Zone 1	Interface
Benchmark (MW)	2808	5245	2437

These are consistent with corresponding values estimated from the probability tree in Figure 19 to Figure 21.

As with the probability tree, the implied ORDC is multivalued, with the changing values of all of the elements affecting all the reserve prices. A separable approximation for the combined energy-reserve dispatch model (1) would apply integrated functions such that at the optimal solution $\left(r_0^*, \bar{r}_1^*, r_1^*\right)$ the derivatives equal the respective prices. For the single constrained zone and rest of the system, an example of a separable version of such a function would be as in (20).

This approach requires an estimate of the reserve solution, and each of the separable functions in (20) depends on all the elements of the benchmark reserve estimate. With a good benchmark estimate, the separable reserve value functions would provide a good approximation. If necessary, the reserve solution from the dispatch model would update the benchmark estimate as part of the iterative solution of the combined energy and reserve dispatch model.

Co-Optimization with Locational Demand Curves and Interface Constraints

The design of the ORDC allows for co-optimization with the energy dispatch. A modified version of the dispatch co-optimization problem in (1) would include these reserve functions in the objective and add a constraint that captures the interaction of energy and reserves in the locational transfer limit.

Extending this to incorporate different types of reserves, such as responsive and non-spin, with the corresponding approximation of separable reserve value functions, allows for simultaneous dispatch of energy and reserves. The constrained zone interface limit ($Kint_i$) must accommodate both the energy flow into the constrained needed to meet the net load ($i^t \hat{y}_i$) and the reservation for operating reserves ($\bar{r}_i$). With these additions, a representative version of the combined dispatch model becomes:

$$
\begin{array}{lll}
\underset{\substack{d,g_R,g_{NR},r_R^z,r_R,r_{NS}^z,r_{NS},em\geq 0;\\ \hat{r}_R^z,\hat{r}_R,\hat{r}_{NS}^z,\hat{r}_{NS},\Delta\overline{r},\hat{y}}}{Max} \quad B(d)-C_R(g_R)-C_{NR}(g_{NR})+\sum_{i=0}^{n}\left[R_{Ii}\left(\hat{r}_{Ri}^z\right)+R_{IIi}\left(\hat{r}_{NSi}^z\right)\right]-\sum_i\sum_j v_{ij}em_{ij} & & \\
d-g_R-g_{NR}=\hat{y} & \text{Net Loads} & \rho \\
i^t\hat{y}=0 & \text{Load Balance} & \lambda \\
H\hat{y}\leq b & \text{Transmission Limits} & \mu \\
g_R+r_R\leq K_R & \text{Responsive Capacity} & \theta_R \\
g_{NR}\leq K_{NR} & \text{Generation Only Capacity} & \theta_{NR} \\
r_{NSi}\leq K_{NSi} & \text{Non-Spin Capacity} & \theta_{NS} \\
i^t r_{Ri}=r_{Ri}^z & \text{Zonal Responsive Reserves} & \gamma_R \\
i^t r_{Ri}+i^t r_{NSi}=r_{NSi}^z & \text{Zonal Non-Spin Reserves} & \gamma_{NS} \\
r_R\leq r_R^{\max} & \text{Responsive Ramp Limit} & \eta_R \\
r_{NS}\leq r_{NS}^{\max} & \text{Non-Spin Ramp Limit} & \eta_{NS} \\
i^t\tilde{y}_i+\overline{r}_i\leq Kint_i & \text{Zonal Interface Limit, } i>0 & \gamma_{K\,\mathrm{int}_i} \\
\overline{r}_i^{\,*}+\Delta\overline{r}_i=\overline{r}_i & \text{Zonal Reserve Interface , } i>0 & \gamma_{\overline{r}_i} \\
r_{R0}^z-\sum_{i=1}^{n}\Delta\overline{r}_i=\hat{r}_{R0}^z & \text{Adjusted Responsive Reserves} & \omega_{R0} \\
r_{Ri}^z+\Delta\overline{r}_i=\hat{r}_{Ri}^z & \text{Adjusted Responsive Reserves , } i>0 & \omega_{Ri} \\
r_{NS0}^z-\sum_{i=1}^{n}\Delta\overline{r}_i=\hat{r}_{NS0}^z & \text{Adjusted Non-Spin Reserves} & \omega_{NS0} \\
r_{NSi}^z+\Delta\overline{r}_i=\hat{r}_{NSi}^z & \text{Adjusted Non-Spin Reserves , } i>0 & \omega_{NSi} \\
-\sum_{i=0}^{n}\sum_{j=1}^{m}em_{ij}\leq\sum_{i=0}^{n}r_{Ri}^z-\sum_{i=0}^{n}X_i & \text{System Contingency Minimum} & \phi_S \\
\sum_{j=1}^{m}em_{ij}\leq\overline{r}_i+r_i-X_i, & \text{Zonal Contingency Minimum, } i>0 & \phi_i \\
em_{ij}\leq Kem_{ij} & \text{Emergency Action Limits,} & \kappa_{ij}.
\end{array}
\tag{22}
$$

The result of co-optimization of reserves and energy would induce scarcity prices for reserves and the interface constraint that affect the locational price of energy.

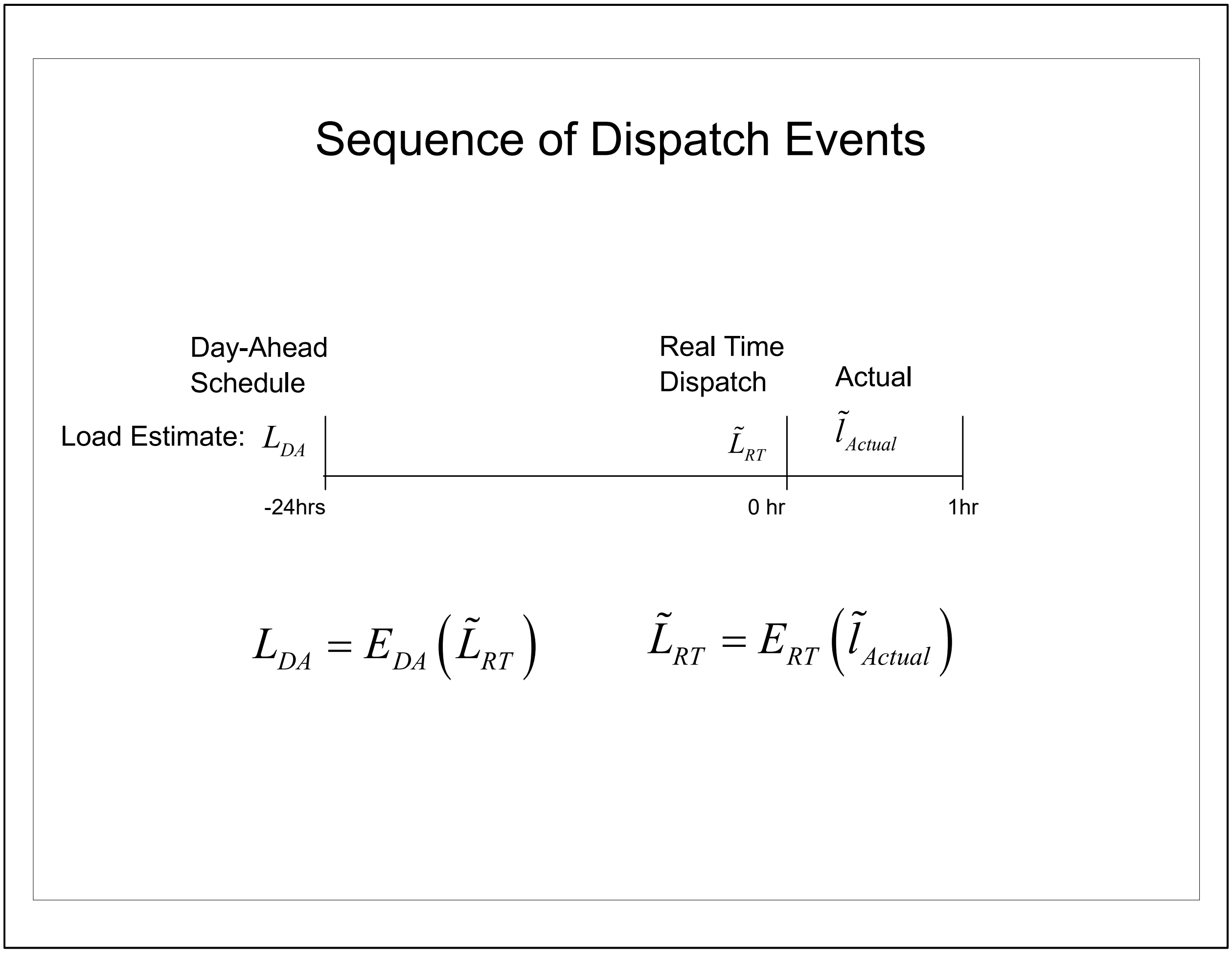
Day-Ahead ORDC

A real-time ORDC with reserve products provides a workable approximation of scarcity pricing and incentives for efficient operation. The basic model extends to forward markets where generation, load and reserves create forward commitments and there is a multi-settlement resolution of these forward financial obligations and imbalances. Extension of the ORDC to the forward market includes elements that are similar to but not identical with the application of the

ORDC in the real-time market. The case of a day-ahead market coupled with a real-time market illustrates derivation of a day-ahead ORDC from the underlying principles.

The timeline in Figure 24 provides the connections with the main variables to support differences in expectations of uncertainty from the perspective of the real-time dispatch versus 24-hours ahead at the time of the day-ahead scheduling.

Figure 24



The actual load is defined as $\tilde{l}_{Actual}$, which will be observed over the actual dispatch interval, taken here for illustrative purposes to be one hour. Just before the start of the dispatch interval, shown as hour 0, the system operator sets a dispatch schedule L_{RT}. The actual load is not known at the time of this real-time dispatch decision. For illustrative purposes, assume that the dispatch is set at the expected value of the actual load given the available information at the start of the real-time interval, so in real-time dispatch we have:

$$L_{RT} = E_{RT}\left(\tilde{l}_{Actual}\right).$$

The actual net load changes over the dispatch period. These changes can be deviations in load and the full range of deviations of dispatch generation, all interpreted as an uncertain change in net load that must be met by reserves or emergency action. The real-time ORDC defines the marginal willingness-to-pay for an increment of reserves. If we have real-time reserves of r_{RT}, then the focus is on the deviation $\tilde{l}_{Actual} - L_{RT}$. Given the real time reserves, the loss of load probability is:

$$Lolp_{RT}\left(r_{RT}\right) = Prob_{RT}\left(\tilde{l}_{Actual} - L_{RT} - r_{RT} \geq 0 \middle| r_{RT}\right). \tag{23}$$

This is used to define the real-time ORDC.

Although L_{RT} is set at the time of the real-time dispatch, it is a random variable from the perspective of the day-ahead schedule, denoted as $\tilde{L}_{RT}$. The same could be true for the real-time reserves, $\tilde{r}_{RT}$, as seen from the day-ahead scheduling.

The day-ahead schedule L_{DA} is set at time -24 hours, where the illustrative schedule is set to the expected value of the real-time dispatch based on the information available day-ahead. Hence,

$$L_{DA} = E_{DA}\left(\tilde{L}_{RT}\right).$$

The day-ahead scheduling decision faces these uncertain variables, and other complications such as lumpy commitment decisions. For the sake of the present discussion, the assumption here is that all day-ahead variables are continuous and thus the focus is on the marginal conditions. But these day-ahead commitment decisions retain the property that they are inflexible in the sense that the decision must be made in the day-ahead schedule, otherwise the associated generation or reserve capacity would not be available in real-time.

In real-time it is natural to assume that the controllable capacity is inflexible in the sense that it must be obtained prior to its use, and before the uncertainty is revealed. The real-time capacity is all physical and there are no strictly financial or "virtual" elements in real-time. These are underlying assumptions of the design of the real-time ORDC.

In the day-ahead setting, the situation differs. There is some capacity that is inflexible in the sense that it requires a day-ahead physical commitment. But other capacity is flexible in that it can be dispatched in real-time without requiring a day-ahead commitment decision. In addition, there can be virtual transactions that are not connected to a physical schedule or dispatch.

Another complication of the day-ahead falls under the rubric of "reliability unit commitment," (RUC) taken here to represent additional and separate commitment decisions that are not necessarily implicated by the bid-in load. For example, RUC commitments may be used to deal with special transmission contingency issues outside the day-ahead model, or to allow for conservative operator load forecasts that differ from the bid-in load.

The treatment of RUC commitments and virtual transactions is set aside here to focus on the marginal value of "physical" commitments, including those for reserves, as part of the economic solution given day-ahead loads bids and generation offers. The discussion considers first the case of a single reserve product, and then introduces the case of more than one reserve product. Again, the approach is to approximate the marginal benefits of the underlying stochastic, dynamic optimization that is too hard to implement.

One Reserve Product

In the simplest case, where there is only reserve product, r_{DA}, the day-ahead ORDC would be defined by the willingness-to-pay for an increment of these day-ahead reserves. The analysis is similar to the real-time problem of finding a workable approximation without requiring explicit solution of the full dynamic, stochastic optimization problem.

In the approach here, there is a simplifying choice in the treatment of the day-ahead reserve schedules and the interaction with the changing load and dispatch in real-time. The range would be between no interaction and strong interaction.

With a single type of reserves, the strong interaction assumption would be that changes in load and associated generation in the real-time dispatch would imply a corresponding change between the day-ahead reserve schedule and the real-time reserves. Hence, the reserve in real-time, $\tilde{r}_{RT}$, would follow as:

(24) $$\tilde{r}_{RT} = r_{DA} + \left(L_{DA} - \tilde{L}_{RT} \right).$$

From the perspective of the day-ahead, the real-time dispatch is uncertain, and hence the real-time reserves are uncertain. Therefore, the marginal value of real-time reserves, derived using (23) but seen from the day-ahead perspective, is uncertain. The *Lolp* expected value depends in part on the uncertainty between day-ahead and real-time. The marginal event is still seen as the need to revert to emergency actions during the actual dispatch. The focus remains on the use of reserves and the probability of the event of reverting to emergency actions, as in:

$$\tilde{l}_{Actual} - \tilde{L}_{RT} \geq \tilde{r}_{RT}.$$

The strong interaction assumption provides the probability of this occurrence by connecting the day-ahead and the real-time. In particular, assuming (24) holds, we have:

(25) $$\tilde{l}_{Actual} - \tilde{L}_{RT} - \tilde{r}_{RT} = \tilde{l}_{Actual} - \tilde{L}_{RT} - r_{DA} - \left(L_{DA} - \tilde{L}_{RT} \right) = \tilde{l}_{Actual} - L_{DA} - r_{DA}.$$

From the perspective of the day-ahead conditions, the real-time emergency action condition is:

$$\tilde{l}_{Actual} - \tilde{L}_{RT} - \tilde{r}_{RT} \geq 0.$$

This includes random variables from the perspective of the day-ahead decisions. However, using (25) this is equivalent to:

$$\tilde{l}_{Actual} - L_{DA} - r_{DA} \geq 0.$$

Therefore, the relevant loss of load probability for defining the marginal value of reserves scheduled day-ahead is:

(26) $$Lolp_{DA}\left(r_{DA}\right) = Prob_{DA}\left(\tilde{l}_{Actual} - L_{DA} - r_{DA} \geq 0 \middle| r_{DA}\right).$$

This loss-of-load probability would be used with the *VOLL* and any threshold minimum contingency reserve requirements to define the day-ahead ORDC using the same analysis and approximations as the real-time ORDC, but with the different probability distribution.

By way of comparison, the case of no interaction would be relevant with multiple reserve products.

Multiple Reserve Products

The strong interaction assumption between day-ahead reserve schedules and real-time changes in net load may not be appropriate when there are multiple reserve products. For instance, with two reserve products, say responsive and non-spin as above with $\left(r_R, r_{NS}\right)$, an alternative assumption would be that the total of the two types of reserves would follow the strong interaction assumption, but the higher quality r_R reserves would have no interaction required by any physical constraint. In effect, the assumption is that the random variation between day-ahead and real-time is absorbed by the lower quality reserves, r_{NS}. With the added assumption that there is an interior solution, so that $r_{NS} > 0$, the marginal conditions for the lower quality reserves must account for the uncertain interactions over the day, but the higher quality reserves would be preserved and require a *Lolp* contribution calculation based only on real-time uncertainty.

Hence, the no interaction assumption replaces (24) for the responsive reserves and maintains that:

(27) $$\begin{aligned} \tilde{r}_{NS,RT} &= r_{NS,DA} + \left(L_{DA} - \tilde{L}_{RT}\right) \\ r_{R,RT} &= r_{R,DA}. \end{aligned}$$

From this perspective, the analysis of the corresponding loss-of-load probabilities would be different for the two types of reserves. The non-spin type reserve would by valued based on the day-ahead probability distribution as in (26). The responsive reserve contribution to the total reserve price would be valued using the real-time probability distribution as in (23).

Hence, following the approach in (16), with the *VOLL* approach for multiple reserves types to define the day-ahead reserve prices, the sub-period *Lolp* relevant for each type of reserve product would be different for the responsive and non-spin reserves, and based on the real-time or day-ahead *Lolp* as in:

(28)
$$\pi_{R,DA}\left(r_{R,DA}\right)=\begin{Bmatrix} Lolp_{I,RT}\left(i^t r_{R,DA}-X\right), & i^t r_{R,DA}-X\geq 0 \\ 1, & i^t r_{R,DA}-X<0 \end{Bmatrix}$$
$$\pi_{NS,DA}\left(r_{R,DA},r_{NS,DA}\right)=\begin{Bmatrix} Lolp_{I+II,DA}\left(i^t r_{R,DA}+i^t r_{NS,DA}-X\right), & i^t r_{R,DA}+i^t r_{NS,DA}-X\geq 0 \\ 1, & i^t r_{R,DA}+i^t r_{NS,DA}-X<0 \end{Bmatrix}$$
$$P_{R,DA}\left(r_{R,DA},r_{NS,DA}\right)=v*\left(\delta*\pi_{R,DA}\left(r_{R,DA}\right)+\left(1-\delta\right)*\pi_{NS,DA}\left(r_{R,DA},r_{NS,DA}\right)\right),$$
$$P_{NS,DA}\left(r_{R,DA},r_{NS,DA}\right)=v*\left(1-\delta\right)*\pi_{NS,DA}\left(r_{R,DA},r_{NS,DA}\right).$$

In principle, if reserves did not interact with generation and load changes, so that the available reserves in real-time would not be affected by uncertainty, everything would reduce to the case of day-ahead prices being determined solely by the real-time *Lolp*. This special case would be the same result that would obtained from (28) if there were no uncertainty between day-ahead scheduling and real-time dispatch.

The cascade structure of (28) ensures that the day-ahead price of the higher quality reserve is greater than the price of the lower quality reserve, and the difference in prices reflects only the fast response benefit that appears in real time and is not affected by the uncertainty between day-ahead and real-time.

The true uncertain environment of the stochastic, dynamic optimization would fall somewhere between the assumptions of strong interaction and no interaction. A reasonable choice between the day-ahead and real-time probability distributions would depend in part on an estimation of the strength of the interaction effect.

Day-Ahead and Real-Time Settlements with an ORDC

An ORDC in the day-ahead and real-time gives rise to energy and reserve imbalance charges (Hogan, 2013). There are several issues, but here the focus is on how the payments should work in the case of examples where market-clearing energy offers are at the value of lost load (*VOLL*). This is an extreme case, but the discussion clarifies the approximations in the ORDC and the interactions between day-ahead and real-time settlements. Does this necessarily give rise to a double payment above *VOLL*? Is it necessary to use the day-ahead prices in determining the real-time settlements? The answers discussed below are “no on average” and “no.”

The issue is closely related to what has been described as "discounting" in the ORDC, where the implicit scarcity price is defined by the net value,

$$v = \left(VOLL - \partial \hat{C}_R \left(\hat{g}_R \right) \right) = \left(VOLL - \text{marginal energy offer} \right),$$

combined with the *Lolp*. Hence, the scarcity price is $v * Lolp$. The discounted value applies for the energy offers from the marginal plants in the dispatch. In constructing the ORDC, this implies that with sufficiently high energy offers, at the *VOLL*, the implicit value of scarcity (of the marginal dispatched capacity) is zero. The discussion below describes an approach that integrates this idea with a settlements process. The examples are for a single ORDC and one type of reserves, but the same issues appear in the interacting zonal ORDC models.

Two-Settlement Model

The concern with examples that show apparently anomalous results when there are very high energy offers is not something that appears to have arisen in other applications of ORDCs. For example this concern could have arisen in the New York Independent System Operator (NYISO) although the scarcity values in an earlier version the NYISO model were low (max about $1800/MWh), along with a low energy offer cap ($1000/MWh). This means that the combined price never approaches the *VOLL*. However, if the NYISO model were applied without these lows offer caps, the energy price could be driven to 2**VOLL*. In fact, with the additive regional zonal model for reserves, the limit in principle could be 3**VOLL*.

A principal reason for offer caps is to mitigate market power, and as a safety valve to avoid unforeseen consequences that produce unbounded prices. For these purposes, there is no need for any link between the offer cap and the *VOLL*. All that is required is that the offer cap be above the true variable generation costs. Furthermore, for a generator without market power, the uniform equilibrium price with an ORDC provides a strong incentive to offer energy at variable cost. The generator does not need to offer high in order to capture a scarcity rent. In general, we would expect $VOLL > Offer\ Cap$, but this is not required and in some cases there could be a high offer for generation that could provide both energy and reserves. (There could also be high energy offers for generation capacity that does not provide reserves, but this would be easy to accommodate and is ignored in the illustrations). The market design does not assume low energy offers, and the example here shows how high energy offers affect the outcomes.

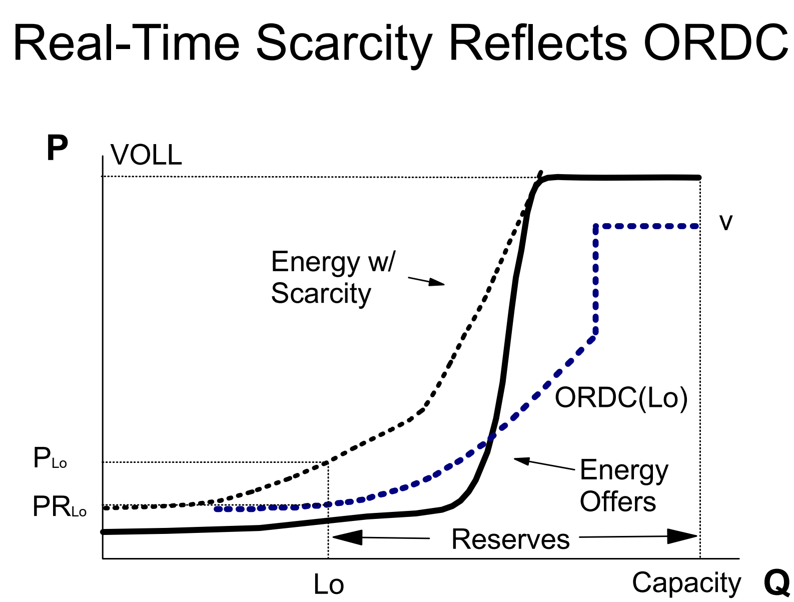
Equilibrium Prices

There are many moving pieces. The illustration here uses simple graphics with continuous functions, only one type of reserves, and market participants who are acting like risk-neutral competitive generators and loads that do not manipulate the markets or prices.

The presentation uses an equilibrium formulation, with uncertainty about the real time outcomes when the day-ahead market clears. The design features include that there is no arbitrage at the margin in the equilibrium solution. Hence, at the margin, no generator has an incentive to deviate from the dispatch of energy and reserves. Further, the day-ahead equilibrium prices are equal to the expected value of the real-time prices for energy and reserves. Virtual bids and offers in the day-ahead market are then settled at the real-time price. For simplicity, the physical loads and generators are treated as hedgers who bid for expected demand and offer the full generation available. Virtual bidders are all alike and assumed to be risk neutral and not liquidity constrained. This assumption for virtual trading implies a horizontal virtual offer curve day-ahead at the expected real-time prices.

In order to focus on the impact of very high energy offers, the illustration in Figure 25 assumes that there is a very nonlinear generation offer curve that caps out at the *VOLL*. There are multiple dimensions to this problem: the energy quantity is measured from left to right; physical reserves are measured from right to left; there is a different ORDC for every energy dispatch, but the figure shows only the ORDC for the given dispatch.

Figure 25

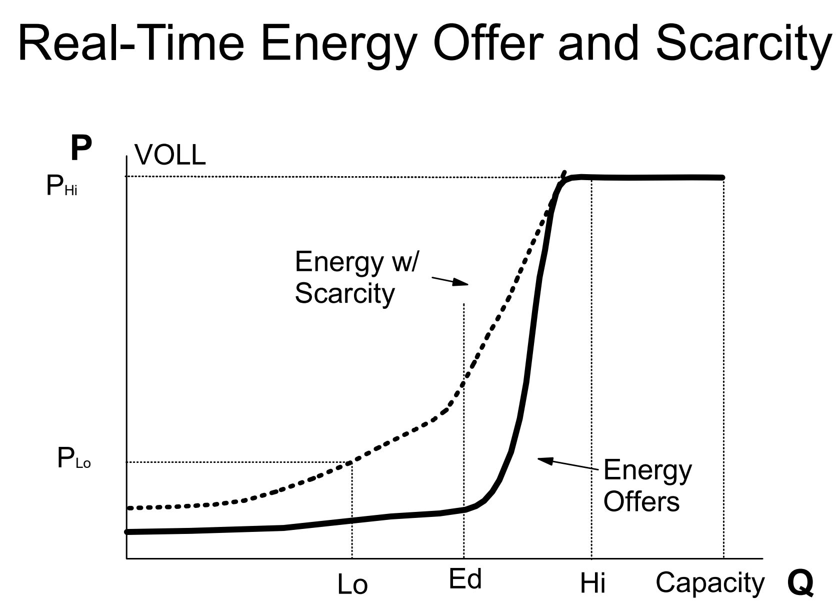


Capacity above the dispatch point provides reserves. Each dispatch produces a different ORDC to reflect "v=VOLL-Marginal Energy Offer" and the adder to the marginal energy offer.

At the real-time "Lo" level of load in Figure 25, energy is dispatched to meet load and the remaining capacity is treated as reserves. The ORDC superimposed on the reserves determines the scarcity adder $v * Lolp$ to the marginal energy offer at the "Lo" quantity, and the total determines the energy price. For each dispatch, there is a different value of v to reflect the changing marginal energy offer. The dashed line traces out the resulting combination of "energy with scarcity" price. If the dispatch load is high enough, the energy offer is at the *VOLL* and there is no scarcity adder for this marginal capacity. This real-time dispatch always leaves the marginal generator indifferent between providing energy and providing reserves.

In Figure 26 the illustration sets out two possible outcomes in the real time dispatch. To simplify, assume that there are two load levels, "Lo" and "Hi," that are equally probable.

Figure 26

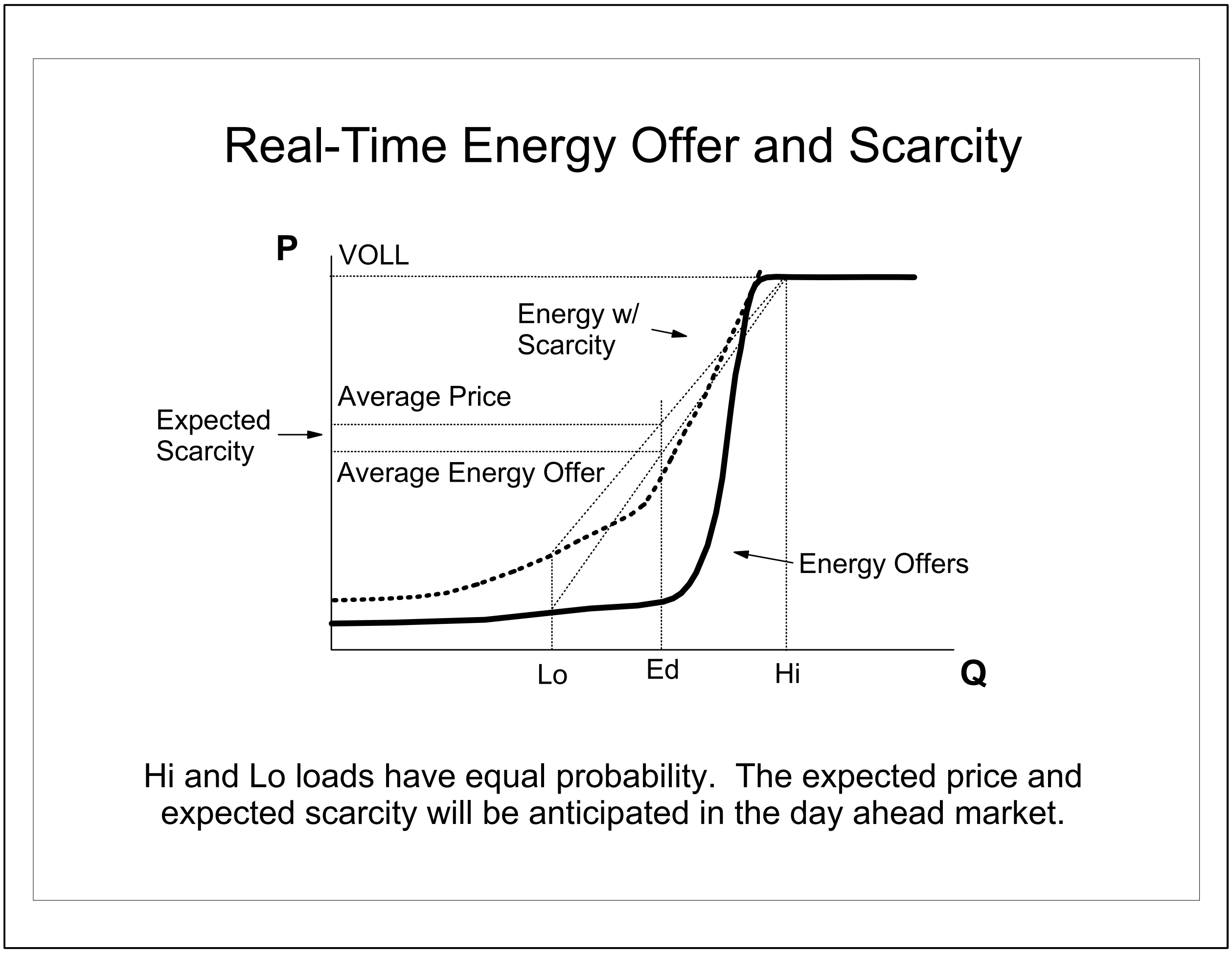


Hi and Lo loads have equal probability. Scarcity derived from the ORDC is the marginal value of incremental capacity net of energy costs. Offers at VOLL have no implicit scarcity value.

The expected value of demand or average load is "Ed." When there is Lo load the dispatch produces an energy price that includes a positive scarcity adder. When the load is Hi, the dispatch produces a high energy price at the *VOLL*, but there is no scarcity adder and the ORDC(Hi) is just the zero axis. The expected load at Ed is never realized in real time, but it will be relevant in the day-ahead schedules.

The illustration in Figure 27 takes this input and illustrates the average energy price, the average energy offer, and the difference between these two which is the average scarcity price.

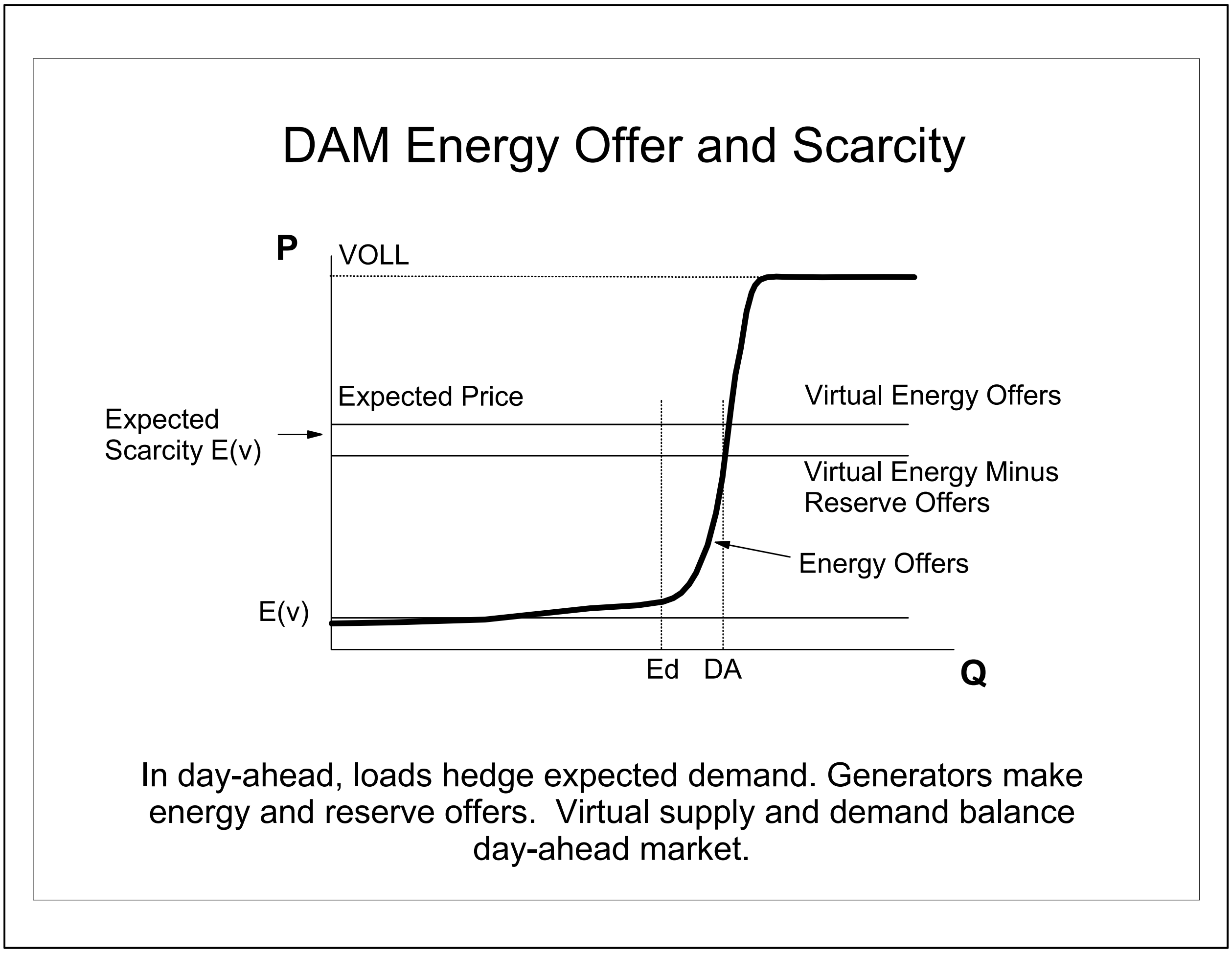
Figure 27



These average prices deviate significantly from the energy offers and scarcity prices that would apply in real time at the average load. This arises from the non-linear generation offer curve and the assumption that the Lo and Hi values are materially different. In actual application the real-time range could be relatively small, and the offer and scarcity curves would be approximately linear. However, the extreme case in Figure 27 is useful in illustrating the connections with the day-ahead market and the large variation in the real-time outcomes.

In Figure 28 the illustration summarizes the day-ahead market with expected load and energy offers to hedge physical generation, and virtual offers and bids that will be cleared at the real-time prices.

Figure 28



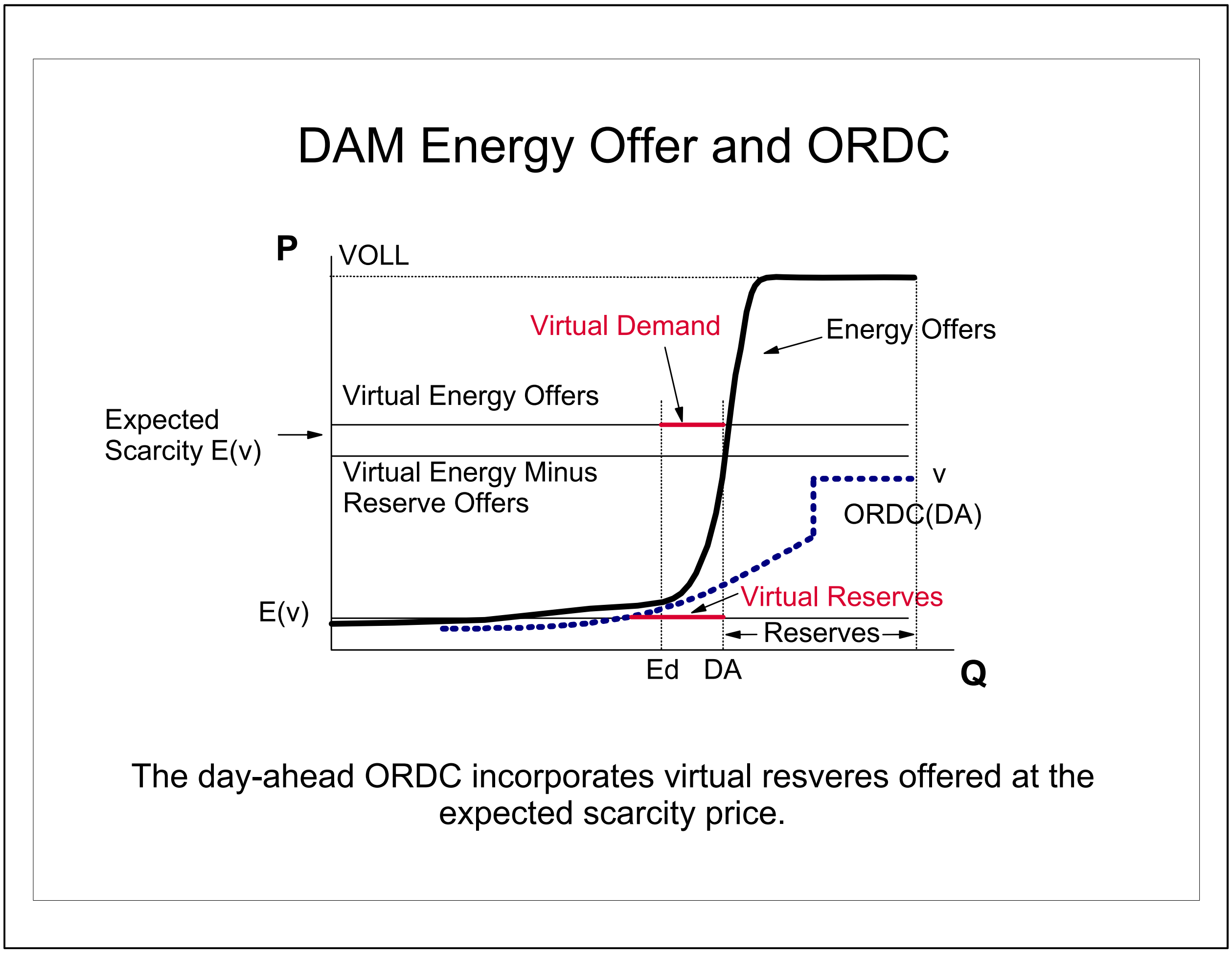
The assumption is that load hedges the expected value of real-time load, Ed. Generators make offers that are at the same offer costs as the real-time. The equilibrium conditions, identical information, and risk neutrality assumptions imply that the virtual offers and bids day-ahead are at the expected real-time prices.

The day-ahead ORDC reflects the expected conditions, bids and offers. In this case, the difference between the day-ahead energy dispatch, DA, and the expected real-time load, Ed, arises from the cleared virtual demand bids. The dispatch of energy at DA implies a marginal energy cost that is higher than the Lo value and lower than the Hi value from real time. As shown in Figure 29, this produces a value of v and results in an ORDC that includes positive scarcity prices but with a lower implicit scarcity adder than the ORDC illustrated in Figure 25 for the real-time Lo case. Furthermore, virtual offers for reserves can shift the ORDC, and both effects result in an equilibrium scarcity price that equals the expected value of the real-time scarcity adder.

To simplify the illustration, the day-ahead ORDC applies the case with no interaction between day-ahead schedule and real-time dispatch in affecting the availability of the real-time reserve.

Hence, the day-ahead ORDC uses the same *Lolp* as the real-time ORDC. The more general case would not affect the main conclusions here.

Figure 29



The illustration in Figure 29 indicates how the day-ahead ORDC is affected by the virtual offers and remains consistent with the basic model, counting the virtual reserves as part of the reserve supply, in addition to the "physical" reserves provided by the difference between day-ahead dispatch of physical energy and physical capacity.

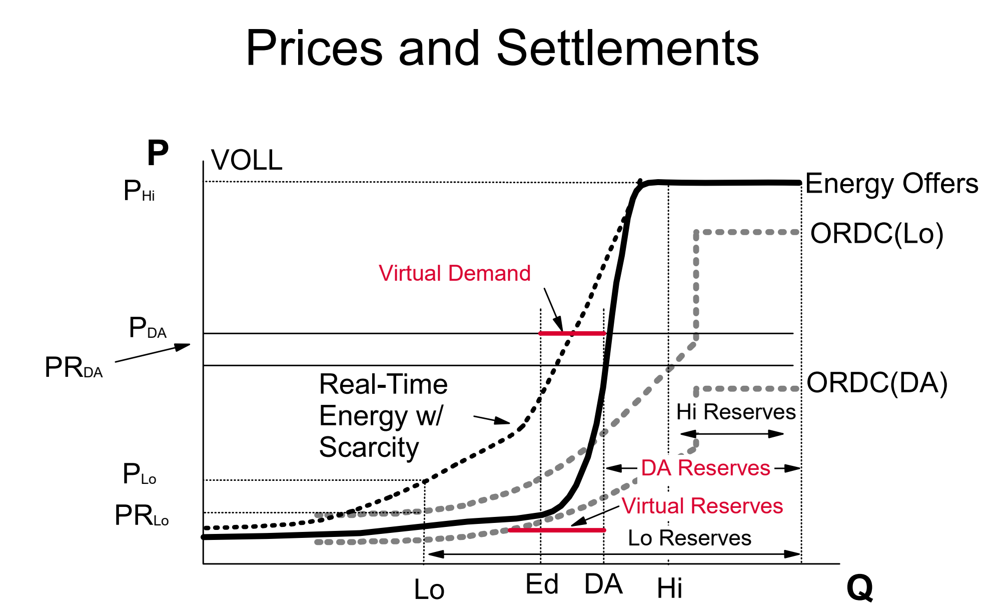
The assumption of the same generation offers in day-ahead and real-time simplifies the presentation and illustrates the various component transactions. The more realistic case, with inflexible generators that need a day-ahead commitment decision would imply a more elastic offer curve day-ahead. In addition, the choice of different probability distributions for the day-ahead and real-time ORDCs would change the graphics somewhat but would not affect the basic settlements logic. The assumed presence of virtual transactions day-ahead would assure the equilibration of the day-ahead and expected real-time prices for energy and reserves.

Settlements

One perspective for describing the day-ahead and real-time settlements is that all cleared transactions in the day-ahead market are financial obligations, not just the virtuals. This is sometimes referred to the "gross pool" approach. The results would be the same as for the "net pool" approach where we net out real-time deliveries relative to day-ahead schedules and make settlement payments only for the imbalances. However, the gross pool interpretation is more direct because we avoid the necessity of doing the accounting to net out the imbalances between day-ahead schedules and real-time deliveries.

The cleared prices for the day-ahead are shown in Figure 30. In the first settlement the day-ahead awards are paid or pay for the cleared quantities at the day-ahead prices. In the second settlement, one of the load conditions appears in real time, either Lo or Hi, and we see the appropriate corresponding real-time prices in Figure 30. In the gross pool interpretation, all the day-ahead schedules are purchased back at the real-time prices, and all the real-time physical transactions are settled at the real-time prices.

Figure 30



Hi and Lo loads have equal probability. Day-ahead and real-time energy and reserve prices apply in settlement system.

For energy transaction cleared in the day-ahead, the real-time price is either P_{Lo} or P_{Hi}. The revenues day-ahead are determined by P_{DA}. Hence, under the gross pool, the net payment is either $P_{DA} - P_{Lo}$ or $P_{DA} - P_{Hi}$. The difference may be positive or negative, but on average the payment is $P_{DA} - 0.5P_{Lo} - 0.5P_{Hi} = 0$, and the equilibrium has zero net profit. A similar result applies for the payments for reserves.

In real time, the physical dispatch applies the usual rules and results of economic dispatch. The gross payments for the physical dispatch in real time reflect the efficient economic solution, and have the standard incentive properties.

When viewed from a net pool perspective, for a particular real-time dispatch, the payments may include a "double" amount. For example, if a generator clears to provide reserves in the day-ahead at price PR_{DA}, and the real-time outcome is the Hi case with the generator providing energy, the generator receives $\left(PR_{DA} - 0\right)q + P_{HI}q = \left(PR_{DA} + VOLL\right)q$. This is a type of "double" payment, but it is not a problem. In the alternative case, where the generator is not dispatched but provides reserves at the Lo load, then the payment is $\left(PR_{DA} - PR_{Lo}\right)q + PR_{Lo}q = PR_{DA}q$. The average payment across the two real-time outcomes is:

$$0.5\left(PR_{DA} + VOLL\right)q + 0.5 * PR_{DA}q = \left(PR_{DA} + 0.5 * VOLL\right)q = \left(0.5 * 0 + 0.5 * PR_{Lo} + 0.5 * VOLL\right)q .$$

This is the same dispatch result and the same as the average payment in the real-time without the day-ahead market. Hence, on average there is no double payment. This condition is an inherent characteristic of the equilibrium.

Hence, within the equilibrium framework, any apparent double payments are not a problem. They average out across the different real-time conditions. The equilibrium principles follow from maintaining the right incentives, with the marginal generator in the dispatch always indifferent between providing energy or reserves, or between sales in real-time or day-ahead.

The extreme examples of the very nonlinear energy offer curve topping out at *VOLL* would be rare in practice. The actual real-time load and the day-ahead energy dispatch should be closer. The actual virtual bids would not individually be so perfect as to be at the expected price. If there are no virtual offers and bids, generators will have an incentive to adjust day-ahead offers to reflect the same information. For example, in equilibrium a generator could set its offer to the maximum of its variable cost or the expected variable cost. But on average across all the virtual bids the market design assumption is that the day-ahead price will be a reasonable estimate of the expected real-time price (i.e., there should be price convergence between real-time and day-ahead).

Two Settlement Summary

The ORDC does require an estimate of the appropriate "discount" relative to the *VOLL*, to reflect the estimate of the marginal energy offer in the relevant dispatch. This produces different energy and scarcity prices depending on the possible outcomes in the real-time market. The day-ahead and real-time clearing energy prices never exceed the *VOLL*. The day-ahead market equilibrium condition would be at the expected value of the real-time prices, with virtual bids and offers cleared to balance the market. The prices in day-ahead and real-time are designed to be in equilibrium where the marginal generator in the dispatch is indifferent between providing energy or reserves. The day-ahead settlement is at the day-ahead prices. In the gross pool interpretation, all the day-ahead cleared transactions are financial and settle in real-time at the real-time prices. The day-ahead prices do not play a role in the real-time settlements. In any particular case, the real-time settlement payments may be greater or less than the day-ahead price. But on average, the net of day-ahead and real time settlements for all financial transactions nets out to zero. The net of all payments day-ahead and real-time for both real-time physical transactions and day-ahead financial transactions would, on average, be the same as for the physical market alone.

References

Ahn, B. and Hogan, W. W. (1982) 'On convergence of the PIES algorithm for computing equilibria', *Operations Research*, 30(2), pp. 281–300. Available at: http://or.journal.informs.org/content/30/2/281.short (Accessed: 4 April 2013).

Gribik, P. R., Hogan, W. W. and Pope, S. L. (2007) 'Market-Clearing Electricity Prices and Energy Uplift'. Available at: http://www.hks.harvard.edu/fs/whogan/Gribik_Hogan_Pope_Price_Uplift_123107.pdf.

Hogan, W. W. (1992) 'Contract networks for electric power transmission', *Journal of Regulatory Economics*, 4(3). doi: 10.1007/BF00133621.

Hogan, W. W. (1993) 'A Competitive Electricity Market Model'. Harvard University. Available at: https://sites.hks.harvard.edu/fs/whogan/transvis.pdf.

Hogan, W. W. (2002) 'Electricity market restructuring: Reforms of reforms', *Journal of Regulatory Economics*, 21(1). doi: 10.1023/A:1013682825693.

Hogan, W. W. (2010) 'Scarcity Pricing and Locational Operating Reserve Demand Curves'. Available at: http://www.hks.harvard.edu/fs/whogan/Hogan_FERC_060210.pdf.

Hogan, W. W. (2013) 'Electricity scarcity pricing through operating reserves', *Economics of Energy and Environmental Policy*, 2(2). doi: 10.5547/2160-5890.2.2.4.

Hogan, W. W. (2016) 'Virtual bidding and electricity market design', *The Electricity Journal*. Elsevier Inc., 29(5), pp. 33–47. doi: 10.1016/j.tej.2016.05.009.

Hogan, W. W. and Pope, S. L. (2017) *Priorities for the Evolution of an Energy-Only Electricity Market Design in ERCOT*. Available at: https://www.hks.harvard.edu/fs/whogan/Hogan_Pope_ERCOT_050917.pdf.

Joskow, P. L. (2008) 'Capacity payments in imperfect electricity markets: Need and design', *Utilities Policy*, 16(3), pp. 159–170. Available at: http://www.sciencedirect.com.ezp-prod1.hul.harvard.edu/science/article/pii/S0957178707000926/pdf?md5=141854e71a805b11bea5d3510c9fbc6f&pid=1-s2.0-S0957178707000926-main.pdf.

Joskow, P. L. (2019) 'Challenges for Wholesale Electricity Markets with Intermittent Renewable Generation at Scale: The U.S. Experience'. Available at: https://economics.mit.edu/files/16650.

OFGEM (2014) 'Electricity Balancing Significant Code Review : Impact Assessment for Final Policy Decision'. Available at: https://www.ofgem.gov.uk/sites/default/files/docs/2014/05/electricity_balancing_significant_code_review_-_final_policy_decision.pdf.

Ott, A. L. (2003) 'Experience with PJM market operation, system design, and implementation', *IEEE Transactions on Power Systems*, 18(2), pp. 528–534. doi: 10.1109/TPWRS.2003.810698.

PJM (2016) *PJM Region Transmission Planning Process: Manual 14b*. Revision 36. Available at: http://www.pjm.com/~/media/documents/manuals/m14b.ashx.

Potomac Economics (2017) ‘Report on Best Practices in Wholesale Electricity Market Design’. Available at: https://www.potomaceconomics.com/wp-content/uploads/2018/10/Best-Practices-Report-to-Alberta-MSA-20171129.pdf.

Shanker, R. (2003) ‘Comments on Standard Market Design: Resource Adequacy Requirement’, *Federal Energy Regulatory Commission, Docket RM01-12-000*. Available at: http://elibrary.ferc.gov/idmws/common/opennat.asp?fileID=9619272.

[i] William W. Hogan is the Raymond Plank Professor of Global Energy Policy, John F. Kennedy School of Government, Harvard University. This work was supported by PJM. This paper draws on research for the Harvard Electricity Policy Group and for the Harvard-Japan Project on Energy and the Environment. The author is or has been a consultant on electric market reform and transmission issues for Allegheny Electric Global Market, American Electric Power, American National Power, Aquila, AQUIND Limited, Atlantic Wind Connection, Australian Gas Light Company, Avista Corporation, Avista Utilities, Avista Energy, Barclays Bank PLC, Brazil Power Exchange Administrator (ASMAE), British National Grid Company, California Independent Energy Producers Association, California Independent System Operator, California Suppliers Group, Calpine Corporation, CAM Energy, Canadian Imperial Bank of Commerce, Centerpoint Energy, Central Maine Power Company, Chubu Electric Power Company, Citigroup, City Power Marketing LLC, Cobalt Capital Management LLC, Comision Reguladora De Energia (CRE, Mexico), Commonwealth Edison Company, COMPETE Coalition, Conectiv, Constellation Energy, Constellation Energy Commodities Group, Constellation Power Source, Coral Power, Credit First Suisse Boston, DC Energy, Detroit Edison Company, Deutsche Bank, Deutsche Bank Energy Trading LLC, Duquesne Light Company, Dyon LLC, Dynegy, Edison Electric Institute, Edison Mission Energy, Electricity Authority New Zealand, Electricity Corporation of New Zealand, Electric Power Supply Association, El Paso Electric, Energy Endeavors LP, Exelon, Financial Marketers Coalition, FirstEnergy Corporation, FTI Consulting, GenOn Energy, GPU Inc. (and the Supporting Companies of PJM), GPU PowerNet Pty Ltd., GDF SUEZ Energy Resources NA, Great Bay Energy LLC, GWF Energy, Independent Energy Producers Assn, ISO New England, Israel Public Utility Authority-Electricity, Koch Energy Trading, Inc., JP Morgan, LECG LLC, Luz del Sur, Maine Public Advocate, Maine Public Utilities Commission, Merrill Lynch, Midwest ISO, Mirant Corporation, MIT Grid Study, Monterey Enterprises LLC, MPS Merchant Services, Inc. (f/k/a Aquila Power Corporation), JP Morgan Ventures Energy Corp., Morgan Stanley Capital Group, Morrison & Foerster LLP, National Independent Energy Producers, New England Power Company, New York Independent System Operator, New York Power Pool, New York Utilities Collaborative, Niagara Mohawk Corporation, NRG Energy, Inc., Ontario Attorney General, Ontario IMO, Ontario Ministries of Energy and Infrastructure, Pepco, Pinpoint Power, PJM Office of Interconnection, PJM Power Provider (P3) Group, Powerex Corp., Powhatan Energy Fund LLC, PPL Corporation, PPL Montana LLC, PPL EnergyPlus LLC, Public Service Company of Colorado, Public Service Electric & Gas Company, Public Service New Mexico, PSEG Companies, Red Wolf Energy Trading, Reliant Energy, Rhode Island Public Utilities Commission, Round Rock Energy LP, San Diego Gas & Electric Company, Secretaría de Energía (SENER, Mexico), Sempra Energy, SESCO LLC, Shell Energy North America (U.S.) L.P., SPP, Texas Genco, Texas Utilities Co, Tokyo Electric Power Company, Toronto Dominion Bank, Transalta, TransAlta Energy Marketing (California), TransAlta Energy Marketing (U.S.) Inc., Transcanada, TransCanada Energy LTD., TransÉnergie, Transpower of New Zealand, Tucson Electric Power, Twin Cities Power LLC, Vitol Inc., Westbrook Power, Western Power Trading Forum, Williams Energy Group, Wisconsin Electric Power Company, and XO Energy. The views presented here are not necessarily attributable to any of those mentioned, and any remaining errors are solely the responsibility of the authors. (Related papers can be found on the web at www.whogan.com).