PPL Corporation

FERC Technical Conference on PJM's Reliability Pricing Model

February 3, 2006

Problems With RPM

- PPL Corporation ("PPL") continues to support a market-oriented solution to resource adequacy for PJM.
- Unfortunately, RPM is a non-market, administrative solution.
- RPM will prevent the formation of an active, liquid bilateral market.
- RPM's base and up to 3 incremental auctions provide only limited ex post pricing points compared to a bilateral market that offers continuous trading and ex ante pricing.
- RPM's forward auction, while not necessary, will only provide one year of price certainty and will not itself encourage or assure a steady stream of investment.
- While PPL agrees with RPM's attempt to value capacity locationally, requirement for a forward commitment, and use of a sloped demand curve to provide a more stable revenue stream than a vertical demand curve, PPL believes that RPM's forward auction which guarantees one year of revenue four years in the future must be eliminated.

Proposed Fixes to RPM

- RPM's forward base and incremental auctions should be eliminated and replaced with a single mandatory auction just prior to the delivery year.
- Information transparency should be improved – Information about Generation Supply and Load Demand data, as well as information from the RTEP and information about local reliability constraints should be made readily available via the PJM website.
- Set Obligations Forward – To the extent necessary to provide additional transparency to the market, and to facilitate transmission or generation solutions, PJM could set obligations and establish Local Deliverability Areas on a forward basis without clearing an auction.

But whether RPM is adopted as proposed, or with these proposed changes, FERC should require PJM to adopt an "energy-only" construct to operate in parallel with RPM.

"Energy-Only" Market Construct

- Implementation of an "energy-only" market would require gradually increasing the effective energy price cap from $1000/MWH to the average value of lost load ("VOLL").
- Under this construct, there would be no "missing money". Scarcity rents will compensate peaking units, and encourage demand side response.
- Volatility in the spot price is a powerful incentive for load to hedge via long-term contracts, and for generators to perform well in real time.
- At the same time, PJM must co-optimize the economic dispatch of energy and operating reserves and begin to price operating reserves consistently with the pricing of energy.
- As the energy price cap is gradually lifted in PJM, the amount of price responsive demand will naturally increase.
- PJM's adequacy construct will then be consistent with MISO's reliability model should MISO move forward with the suggested energy-only market.

RPM and "Energy-Only" Can Coexist

- Many observers believe that they can coexist and that, in the end, an energy-only construct should replace RPM.
  - In his June technical conference comments, Dr. Roy Shanker acknowledges that RPM is designed to phase itself out as the energy market becomes more relevant. (Shanker, p. 2).
  - In its RPM filing letter, PJM agrees that RPM should diminish in importance as energy markets prove more effective at incenting construction of capacity resources. (PJM Filing Letter, p. 15).
  - In his September 23, 2005 paper on an "Energy-Only" Market design, Professor Hogan states that an improved spot market design is not mutually exclusive of an ICAP approach, and that with transparent scarcity pricing and an appropriate operating reserve demand curve, the net price of capacity would be zero and the role of the ICAP requirement would fade, or at least be substantially reduced. (Hogan Paper, p. 33).

RPM and “Energy-Only” Can Coexist (Continued)

- Dr. Shanker also explains that all that is necessary to phase out RPM is for PJM to gradually increase the energy price caps. (Shanker, p. 3).
- By implementing RPM and putting in place an energy-only market that programmatically increases the energy price caps up to the average value of lost load, the experts on all sides of the resource adequacy issue can agree that reliability will be preserved.
- Consumers will be protected by bilateral contracting, self-supplied generation, hedging and financial instruments, and increased amounts of demand responsive load.
- Forces of supply and demand will ensure the lowest, long-run costs for electricity rather than administrative mechanisms and unnecessary regulatory interference.

Ensure Reliability and Generation Adequacy

- An energy-only market will act as a backstop if RPM fails.
- RPM will act as a backstop if an energy-only market fails.
- Together, these mechanisms will provide all of the necessary incentives for new generation investment, bilateral contracting and hedging, and development of more demand responsive load.
- PPL is convinced that RPM **will fail**, and there must be a market already established in PJM to keep reliability at the desired level if it **does fail**.

PPL Corporation

RPM Technical Conference Presentation

February 3, 2006

Good afternoon. Thank you for the opportunity to be here today.

As you know, PJM continues to advocate RPM as its long-term adequacy solution. PPL has carefully analyzed RPM and concludes it is an administrative, non-market solution that will not work. RPM's mandatory four-year forward auction will interfere with or eliminate both short-term and long-term bilateral markets for capacity.

PPL has suggested several ways to fix RPM that would make it a more market-oriented approach, and would allow for bilateral contracting. But these suggested fixes haven't exactly been embraced by PJM. Like you, PPL has been searching for some way to resolve the ongoing debate between those favoring the PJM RPM proposal and those who believe it is seriously flawed and will fail. We think we have found a possible resolution.

The Commission should consider adopting an energy-only market in PJM alongside RPM. We have been studying the work of Professor William Hogan and the Midwest ISO and believe that work has a great deal of merit. Professor Hogan and MISO describe how energy-only markets will achieve the optimal level of investment in generation and demand response without the need for extensive administrative structures such as proposed in RPM.

It is important to note that an energy-only market is not a free-for-all with $10,000 price spikes and boom or bust cycles of investment, where consumers are left unprotected. Rather, an energy-only market, as described in the other materials I am submitting today, provides proper incentives for load serving entities to act responsibly to protect themselves and their end-use customers from the full impact of price volatility, by using long-term contracts, self-supplied generation, and financial hedges. State regulators can oversee reliability and hedging requirements. PPL believes these incentives are the key to creating the necessary environment for bilateral contracting and efficient new investment.

Proponents of capacity mechanisms like the existing ICAP market or RPM claim that an energy-only market will not provide adequate long-term resource adequacy. They do not trust market mechanisms to provide the appropriate incentives to both buyers and sellers to produce the right mix of resources in time to meet demand. They admit that the existing situation creates a “missing money” problem. Some of them argue that higher energy price caps expose consumers to higher spot market prices. But market proponents like PPL also recognize that higher spot market prices provide the incentives for load to hedge, and for generators to perform well.

PPL and other energy-only market proponents also believe that replacing the “missing money” through capacity-based systems actually creates a disincentive to short-run operational reliability. Further, if the capacity-based system of choice fails to replace all of the “missing money”, because of an improperly constructed demand curve, or because of excessive market mitigation, it will lead to insufficient investment in long-term resource adequacy.

To resolve this debate, FERC must permit an energy-only market to work **<u>in parallel with</u>** PJM's RPM. The Commission should require PJM to raise its $1,000 price cap gradually up to the value of lost load, simultaneously with implementation of RPM. If PJM is right and RPM produces sufficient new generation in the right places and of the right type, then the scarcity pricing signals that would be occasionally produced by an energy-only market will be absent or rarely experienced. Rather than relying on forward contracts to secure resource adequacy, resource providers and LSEs would rely upon the RPM auctions. The energy-only market can safely stay in the background with little impact on RPM. And if Professor Hogan, the authors of the MISO paper, and PPL are wrong about the ability of an energy-only market to encourage the necessary new investment, then

RPM will function as a backstop to ensure that adequate peaking generation is built to meet PJM's reserve requirement.

However, if PJM is wrong and if PPL, Professor Hogan and the authors of the MISO paper are right, RPM will fail to incent construction of new resources. Then, the higher energy market caps of an energy-only market will assure that there is no missing money, and will encourage both investment and long-term contracting between resource providers and load. Over time, RPM's net revenue offset will prevent overcompensation of resource providers by offsetting RPM payments with energy market revenues, which will make payments under the RPM system small or nonexistent.

Of course, there are many more details to be worked out. And we confess that we have not thought through every issue that could arise in making sure that the programs can work in parallel without creating unnecessary market distortions. In this regard, a vigorous and open stakeholder process to work through all the details would be helpful. However, because PPL places such a high value on reliability, we are not willing to trust the future to RPM alone. Nor should the Commission. PPL believes RPM **will fail**, and there must be a market established in PJM to pick up the pieces if it **does fail**. The Commission should ensure that reliability remains protected even if RPM falters. We look forward to working with the Commission, PJM and its stakeholders in developing an energy-only market for PJM and implementing it in parallel with RPM.

Thank you and I look forward to your questions.

ON AN "ENERGY ONLY" ELECTRICITY MARKET DESIGN FOR RESOURCE ADEQUACY

William W. Hogan

Center for Business and Government
John F. Kennedy School of Government
Harvard University
Cambridge, Massachusetts 02138

September 23, 2005

On an “Energy Only” Electricity Market Design for Resource Adequacy

Contents

On an "Energy Only" Electricity Market Design for Resource Adequacy

William W. Hogan[i]

An "energy only" market design could avoid the need for increasingly prescriptive regulations targeted at ensuring resource adequacy. Transparent scarcity pricing would create better incentives for both operations and investment. An improved electricity market design would not eliminate all need for regulatory prescriptions. However, it would change the nature of the remaining problems and allow for market-based approaches that would not overturn the market.

Introduction

Electricity resource adequacy programs often target the "missing money" problem. The missing money problem arises when occasional market price increases are limited by administrative actions such as price caps. By preventing prices from reaching high levels during times of relative scarcity, these administrative actions reduce the payments that could be applied towards the fixed operating costs of existing generation plants and the investment costs of new plants. The resulting missing money reduces the incentives to maintain plant or build new generation facilities. In the presence of a significant missing-money problem, alternative means appear necessary to complement the market and provide the payments deemed necessary to support an appropriate level of resource adequacy.

In the United States experience, resource adequacy programs designed to compensate for the missing money create in turn a new set of problems in market design. The resource adequacy approaches become increasingly detailed and increasingly prescriptive to the point of severing the connections between major investment decisions and energy market incentives. Consideration of these new unintended consequences prompts interest in seeking ways to operate an electricity market without any money missing. This interest in turn requires further specification of an "energy-only" market design and consideration of alternative modifications of such designs that would address underlying policy concerns without recreating the missing money problems. The reference in quotes emphasizes that the intent is not to create a completely unregulated market. Rather the intent is to eliminate prominent imperfections in the market design and thereby change the nature of the regulatory prescriptions to allow for market-based approaches that would not overturn the market. The purpose of the present paper is to highlight the critical conceptual features of such an energy-only market.

There is an analogy here to other contexts with regulatory policies to influence markets to achieve public purposes without prescribing the technology or investments. For example, in the control of sulfur emissions from coal burning power plants there has always been a tension between command-and-control approaches that dictated technological solutions such as requiring scrubbers on all plants versus market-based approaches like cap-and-trade programs for sulfur emission allowances. The market-based approach targeted the problem (e.g., total emissions) rather than dictating the solution (e.g., scrubbers). The technology prescriptive approach for controlling sulfur emissions would have created high costs and unintended consequences. The alternative market-based approach with tradable emission allowances provided lower costs and better incentives. A similar task in electricity markets would be to establish better market designs and more compatible market-based interventions. In an energy-only market, the potential problems and objectives would be different and the same resource adequacy policy prescriptions might not be required.[1]

Missing Money

The missing-money analysis begins with the load cycle over the day and the seasons.[2] Changing levels of generation supply matched with changing load levels produce volatile costs. These costs include a mixture of the direct variable costs of marginal generators, energy values for storage limited hydro facilities, the marginal value of incremental demand, and so on. At the margin, we refer to the opportunity cost as the cost of meeting an increment of demand by decreasing other load or increasing generation. If contemporaneous spot prices reflect these opportunity costs, these prices would provide market participants with strong incentives during periods of scarcity. During most periods, market prices would be at a relatively low level defined by the variable operating costs of mid-range or base load generating plants. However, in some periods prices would rise above the variable operating costs of peaking units that were running at capacity and would reflect scarcity under constrained capacity with the incremental value of demand defining the system opportunity cost.

Spot prices could be summarized over the year by a price duration curve depicting the cumulative number of hours when prices exceed a given level. As shown in Figure 1, under perfect dispatch generation would operate according to its variable cost of production. The most expensive peak generation (e.g., $85/MWh variable cost) would operate for relatively few hours. The payments in area **A** above the operating cost of $85 would be the returns to cover the fixed costs of the peak plant, including the investment cost needed to compensate new entrants.

---

1 For a discussion of the objectives of resource adequacy programs, see James Bushnell, "Electric Resource Adequacy: Matching Policies and Goals," University of California Energy Institute, CSEM WP 146, August 2005.

2 The characterization as "missing money" comes from Roy Shanker. For example, see Roy J. Shanker, "Comments on Standard Market Design: Resource Adequacy Requirement," Federal Energy Regulatory Commission, Docket RM01-12-000, January 10, 2003.

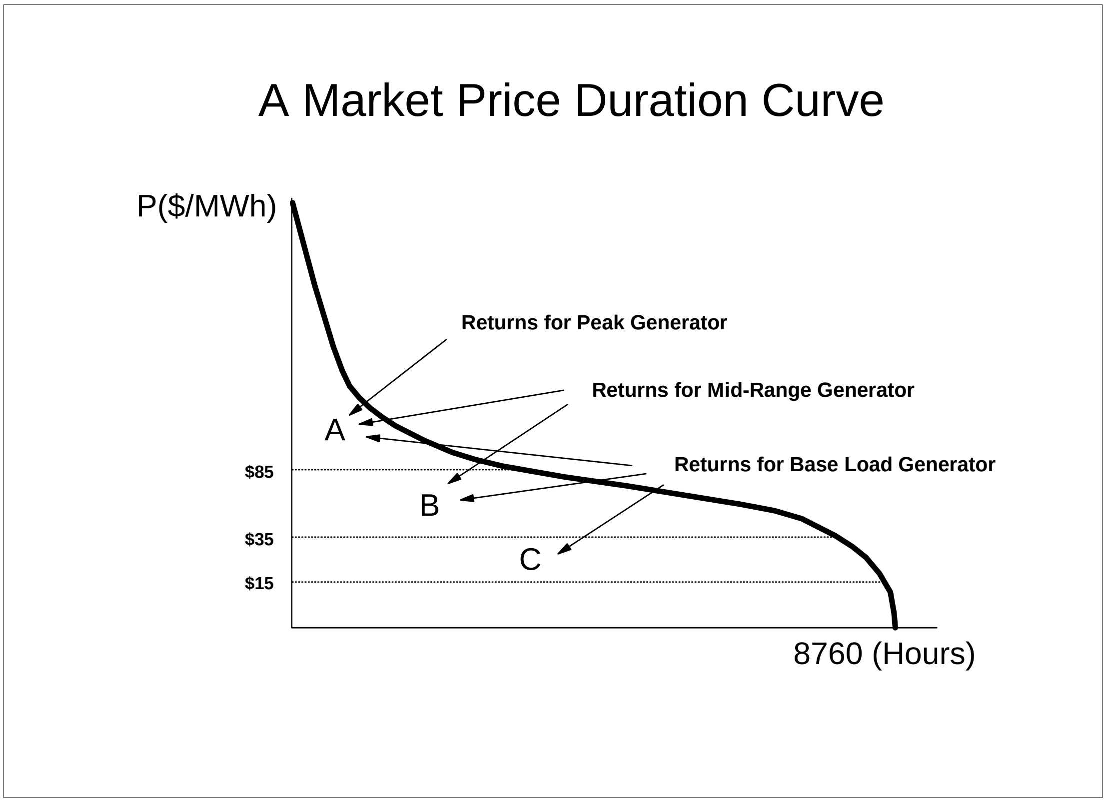


Figure 1

The mid-range generators (e.g., $35/MWh variable cost) would operate for many more hours, and the payments in area **A+B** would cover the larger fixed and investment costs of these plants. The lowest cost (e.g., $15/MWh variable cost) based-load generators would operate in virtually all hours but only rarely would the price fall to the lowest level. The combined area **A+B+C** would provide the payments for fixed and investment costs of base load generators.

The simplified graphic in Figure 1 illustrates the important point that in equilibrium the high prices during the peak hours provide part of the compensation needed for all generators, not just the peaking capacity. The magnitudes could be substantial. Although estimates vary, the approximate magnitude of area **A** would be on the order of $65,000/MW-year for a simple combustion turbine.[3] Hence, average peak prices of $1,000/MWh above operating cost would be needed sixty five hours a year in order to meet the payment requirements for peaking generation.

Introduction of administrative measures to limit the highest prices, such as through a price cap, would reduce payments available to all types of generating plants.

[3] PJM, State of the Market Report 2004, March 8, 2005, p. 82. Estimates for New York city could be twice as high.

As shown in Figure 2, the price cap creates the "missing money." This is the payment to generation that is eliminated as a result of the curtailment of prices. The administrative rules that produce the missing money include a variety of procedures. Explicit price caps are not the usual means of restraining prices. A more likely constraint on generator revenues would arise from an offer cap on generators to mitigate market power. In principle, mitigation through an offer cap need not create missing money. However, when the offer cap combines with a pricing rule that spot prices must be set by the highest (mitigated) offer for any plant running, the result can be capped prices and missing money. The problem would arise when the pricing rule does not fully account for the opportunity cost of incremental demand or incremental operating reserves during periods of scarcity. These opportunity costs at the margin could result in prices above the offer cap.

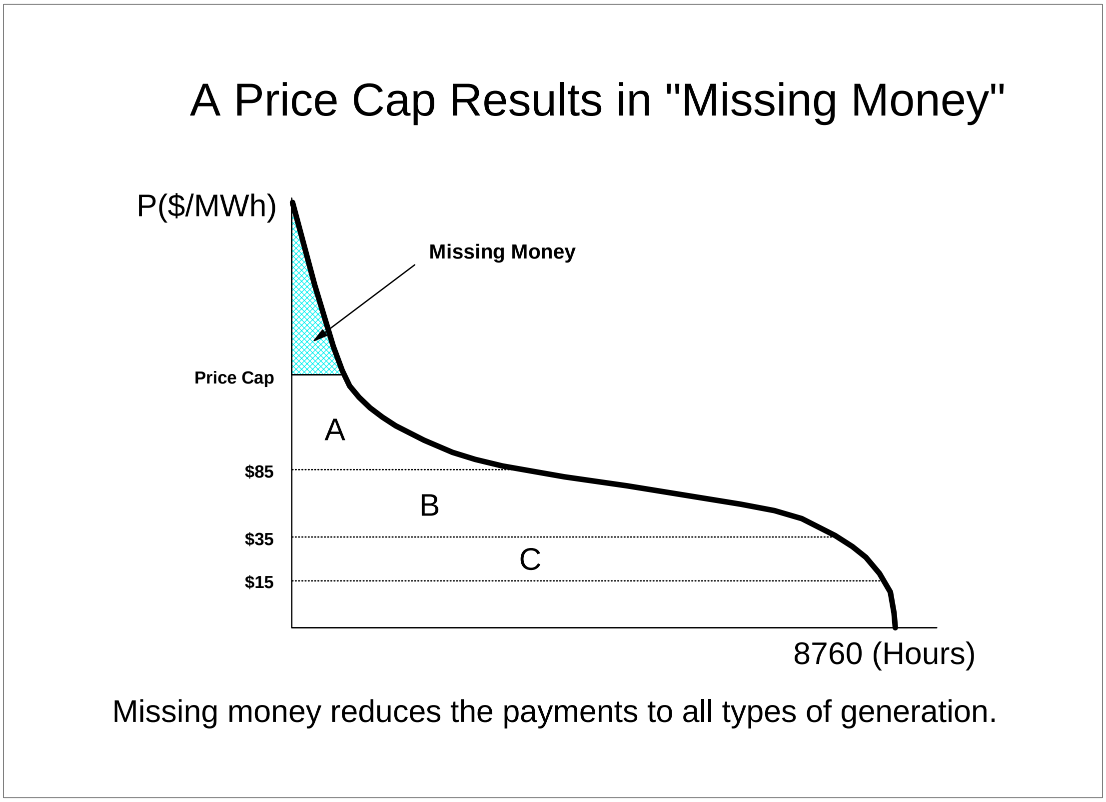


Figure 2

More indirect contributors to missing money would be reliability must run (RMR) units or out of market (OOM) dispatch, both of which involve use of expensive generating plants to meet load. However, various rules exclude these plants from consideration as part of the market and associated determination of prices. For example, the innovation of "soft" price caps institutionalized a device for keeping reported market prices low when real opportunity costs were higher. Collectively these rules produce the anomaly of constrained supplies and low prices.

The missing money goes hand-in-hand with "missing incentives." In the short run, the missing incentives complicate the task of the system operator in maintaining a secure dispatch. With prices disconnected from opportunity costs, market participants would have little incentive to act to enhance security and could face strong incentives to take actions, like leaning on the system to increase exports, which add to the security problems. A result is the need for ever more complicated rules to mandate behavior that is otherwise inconsistent with the incentives. As the rules mount up, the flexibility sought through markets can disappear. By now, it has been widely recognized that one of the principal advantages of locational marginal pricing systems is the alignment of participant incentives to support rather than oppose actions needed to manage congestion in the grid.[4] This role of reinforcing incentives to support reliability would be of greatest importance during periods of scarcity when the missing incentives arise.

The missing incentives extend to investment in new generation facilities or their substitutes. By some estimates the missing money amounts to a much as half or more of the $65,000/MW-yr payment required for new peak load generation investments.[5] It is this missing money that motivates interest in supplementary resource adequacy programs to provide a return to existing plants or support investment in new facilities. A direct approach to meeting the requirements of resource adequacy is to create a process beyond the energy market that provides added payments to generators who maintain or build needed generation facilities. Given an underlying assumption that administrative measures will necessarily be invoked to cap energy prices, and will always lead to insufficient payments in the energy market to maintain the desired level of capacity, the missing money leads inexorably to alternative approaches such as installed capacity (ICAP) requirements and related designs.[6]

The ICAP programs present a number of challenges.[7] For example, the assumption of a price-capped energy market implies that appropriate capacity choices must be identified by means other than market participants responding to the incentives provided through energy prices. This creates a need for central planning and greater prescription by regulators. Accordingly, regulators act on behalf of customers to take on more of the risks inherent in the long term investment decisions. When these details begin to emerge, participants recognize that this resource adequacy approach would recreate many of the features of electricity systems that were to be replaced by greater

---

[4] Phillip G. Harris, "Relationship between Competitive Power Markets and Grid Reliability: The PJM RTO Experience," Issue Papers on Reliability and Competition, US Department of Energy and Natural Resources Canada, August 2005, pp. 4-5. (www.energetics.com/meetings/reliability/papers.html)

[5] Federal Energy Regulatory Commission, State of the Market Report, Washington DC, June 2005, p. 60.

[6] For an extensive examination of this logic, see the workshops sponsored by the California Public Utility Commission, California Energy Commission and the California Independent System Operator at www.cpuc.ca.gov/static/industry/electric/installedcapacity/041004_instcapacity.htm. A summary of the arguments and subsequent recommendations appear in California Public Utility Commission, "Capacity Markets White Paper," Staff White Paper, CPUC Energy Division, August 25, 2005, (www.cpuc.ca.gov/word_pdf/REPORT/48884.pdf).

[7] Scott M. Harvey, "ICAP Systems in the Northeast: Trends and Lessons," LECG, Report to the California ISO, September 2005.

reliance on restructured electricity markets. Inherent in this integrated planning process would be to shift a substantial portion of the investment risk and stranded costs back to customers who would be required to pay under the force of regulation.[8]

The usual argument that leads to this resurrection of integrated planning and regulatory driven investment begins with an assertion that the pure "energy only" electricity market behind Figure 1 is not politically feasible. If low price caps are imposed on the energy market, then something like a significant missing-money problem must follow. However, when the ICAP cure for the missing money problem starts to look worse than the disease, analysts of ICAP-type mechanisms revise this judgment and look more closely at a market design that does not breed the disease and hence does not require the cure.[9] This second look at an energy-only market design should not translate immediately into ignoring the underlying motivations for the administrative measures that constrain prices, such as the need to mitigate market power. But it would require refocusing attention on the policy ends and not the administrative means.

The ubiquitous reference to the missing-money problem found in resource adequacy proposals usually pays homage to the appeal of Figure 1 and the so-called "energy only" market. If only electricity prices could reach market-clearing levels defined by the opportunity costs, where there would be no administrative measures to cap prices and no missing money, there would be no need for administrative measures to supplement the payments to generators.[10] But if the highest spot prices would be too

---

8 Bruce W. Radford, "Holes in the Market," Public Utilities Fortnightly, March 2005, pp. 19-21, 46-47.

9 For example, a series of workshops sponsored by the Public Utility Commission of Texas (PUCT) outlined this evolution of concern with installed capacity markets and return to interest in an energy-only approach. See www.puc.state.tx.us/rules/rulemake/24255/24255.cfm. A staff paper summarized many of the issues, Eric S. Schubert, "An Energy-Only Resource Adequacy Mechanism," Public Utility Commission of Texas, Staff White Paper, April 14, 2005. The staff subsequently recommended that the PUCT develop the details to "… provide for resource adequacy in ERCOT to be achieved through an energy only design." See Richard Greffe, Public Utility Commission of Texas, Wholesale Market Oversight, Memorandum, July 8, 2005. A similar direction appears in the Midwest Independent System Operator, "Discussion Paper on Resource Adequacy for the Midwest ISO Energy Markets," Draft, August 3, 2005.

10 Miles Bidwell, "Reliability Options," Electricity Journal, Vol. 18, Issue 5, June 2005, pp. 1. Eugene Meehan, Chantale LaCasse, Phillip Kalmus, and Bernard Neenan, "Central Resource Adequacy Markets for PJM, NY-ISO, and NE-ISO," NERA Final Report, New York, February 2003. Shmuel Oren, "Capacity Payments and Supply Adequacy in Competitive Markets," VII Symposium of Specialists in Electric Operational Systems Planning," Basil, May 2000. Shmuel Oren, "Ensuring Generation Adequacy in Competitive Electricity Markets," University of California at Berkeley, April 2004. Shmuel Oren, "Capacity Mechanisms for Generation Adequacy Insurance," CPUC-CEOB-CAISO Installed Capacity Conference, San Francisco, California, October 4-5, 2004. Roy J. Shanker, "Comments on Standard Market Design: Resource Adequacy Requirement," Federal Energy Regulatory Commission, Docket RM01-12-000, January 10, 2003. See also Roy J. Shanker, "Comments," Federal Energy Regulatory Commission Technical Conference on Capacity Markets in the PJM Region, June 16, 2005. Harry Singh, Call Options for Energy: A Market Based Alternative to ICAP and Energy Price Caps," PG&E National Energy Group, October 16, 2000. Steven Stoft, Testimony on Behalf of ISO New England on Locational Installed Capacity Market Proposal, Submitted to Federal Energy Regulatory Commission, Docket Number ER03-563-030; Direct August 31, 2004; Supplemental November 4, 2004; Rebuttal February 10, 2005. Carlos Vázquez, Michel Rivier, and Ignacio J. Pérez-Arriaga, "A Market Approach to Long-Term Security

high, so the argument goes, the exposure of customers would be unacceptable and the energy-only market could never work. Typically, the design of an energy-only market is dismissed without further consideration.[11] The discussion of an energy-only market ends there, and attention turns to ICAP design in its many forms.

When frustration with the implications of ICAP markets sets in, the possibility arises to give more serious attention to how an "energy only" market might operate and what would be needed to make the transition. To revisit this issue and consider the ends rather than the means requires a greater specification of such a market in order to understand what would be included in the basic elements and what problems might remain that could motivate an interest in modifications of an energy-only market approach to resource adequacy.

Energy-only Market Design

There are many appeals to an idealized vision of an energy-only market, but there is little description of the key features. The assumptions of the idealized vision do not describe real systems. For example, high spot prices raise the specter of market power, and it is difficult to step back from this reality and describe a market without the exercise of market power. Hence, the discussion of design elements quickly detours from consideration of the core features of an idealized energy-only market. The intent here is to avoid this detour long enough to sketch out the principal elements. The discussion here assumes competitive behavior and no transaction costs, but returns to these issues later in discussing potential problems with an "energy-only" market design.

A core idea of an electricity market that relies on market incentives for investment is that these incentives appear through the largely voluntary interactions of the participants in the market. A main feature of the market would be prices determined without either administrative price caps or other interventions that would depress prices below high opportunity costs and leave money missing. The real-time prices of electric energy, and participant actions, including contracting and other hedging strategies in anticipation of these prices, would be the primary drivers of decisions in the market. The principal investment decisions would be made by market participants, and this decentralized process would improve innovation and efficiency. A goal would be to avoid repeating the problem of leaving customers with stranded costs arising from decisions in which the customers had no choice. This change in the investment decision process and the associated reallocation of risks would arguably be the most important benefit that could justify greater reliance on markets and the costs of electricity restructuring. If this were not true, and if it would be easy for planners and regulators to

of Supply," IEEE Transactions on Power Systems, Vol. 17, No. 2, May 2002. Carlos Vázquez, Carlos Battle, Michel Rivier, and Ignacio J. Pérez-Arriaga, "Security of Supply in the Dutch Electricity Market: the Role of Reliability Options," Report IIT-03-084IC, Comillas, Universidad Pontifica, Madrid, Draft Version 3.0, Madrid, December 15, 2003.

[11] "Energy-only" markets of different designs exist in Alberta, Australia, New Zealand and Europe. The lessons there are relevant, and the resource adequacy issue is not fully resolved. However, the differences in context and details would take the discussion further afield from the U.S. setting.

lay out the trajectory of needed investment for the best portfolio of generation, transmission and demand alternatives, then electricity restructuring would not be needed.

A central concern of the growing doubts about the direction of development in administrative installed capacity markets is the loss of this critical reallocation of decisions away from regulators and towards market participants. The increasing scope, increasing detail and increasingly longer horizons of ICAP programs establish an evolutionary path with no end in sight.[12] It is not the capacity construct *per se* that should be of concern. For example, operating reserve capacity requirements are an essential part of electricity systems. Rather it is the scope, duration and detail of mandatory ICAP obligations that are imposed by the central planner, and the corresponding shift of the locus of investment risks decisions away from the market with the decisions made under the central plan and the risks assigned back to captive consumers.

The changes reverberate through the market design, transforming the original role of the system coordinators, the Independent System Operator (ISO) or Regional Transmission Organization (RTO): "the ISOs/RTOs are not *principals* ***in*** *the market* but rather *service providers* ***to*** *the market*. … Creating a forward looking capacity construct is therefore potentially not 'just' an incremental increase in responsibilities for an ISO or RTO; but rather it is a significant structural change to the role they currently perform."[13] In effect, the effort to compensate for capped spot prices recreates the regulatory integrated resource planning process that was to be replaced by more decentralized market decisions. This movement to create centralized investment decisions according to long-term requirements set by the ISO follows directly from the missing incentives of the price-capped energy market. Hence, any effort to support market decisions and avoid mission creep at the ISO must then provide the missing incentives through market-based prices rather than administrative substitutes. An essential feature of an energy-only market design would be efficient spot pricing to reflect opportunity costs.

Just as important as embracing market-driven spot pricing is a recognition that an energy-only market does not assume or require that all transactions be limited to the spot market. To the contrary, a robust energy market with realistic real-time prices would permit and encourage long-term contracting of a variety of forms to reflect the conditions and relative risk preferences of the market participants. The major economic decisions surrounding investment and most of the actual transactions might and could be made with long-term contracts voluntarily arranged by the parties. Market participants might choose voluntarily to enter into ICAP type contracts, but this would be only one of the possible contract forms. However, even if spot-market transactions were reduced to a small volume for balancing and congestion management, expectations regarding future spot prices created by uncapped energy pricing in the spot market would be essential and would be reflected in the terms of forward contracts.

---

[12] Craig Hart, "Capacity Markets: A Bridge to Recovery?", Public Utilities Fortnightly, May 2005, pp. 24-27.

[13] Midwest Independent System Operator, "Discussion Paper on Resource Adequacy for the Midwest ISO Energy Markets," (emphasis in the original), Draft, August 3, 2005, pp. 1-2.

Similarly, the emphasis on an "energy only" market does not mean that there would be nothing but spot deliveries of electric energy with a complete absence of administrative features in the market. Since the technology of electricity systems does not yet allow for operations dictated solely by market transactions with simple well-defined property rights, the system requires some rules to deal with the complex interactions in the network. To the contrary, there would of necessity be an array of ancillary services and associated administrative rules for such services. For instance, the existing technology for electricity requires that system security be met by providing operating reserves in generation in order to meet the possible contingencies. The required level and configuration of these operating reserves are not determined in a market. The operating reserve requirements are based on prior studies and experience of what resources would be needed and where they would be needed over the next minutes or hours of the dispatch in order to protect against both involuntary load shedding and more serious dangers of system collapse.

Other examples include the rules for providing contingency security constraints for transmission, and various ancillary services that are inherently administrative, even though they may be designed with an economic component included. The goal of the energy-only market is not to eliminate all administrative rules. Rather a goal is to design the rules, pricing and implied incentives to support operating and investment decisions made by market participants in response to the forces in the market.

In the discussion below, the emphasis is on providing electric energy with the critical operating reserves as the representative ancillary service. For the initial discussion, it is assumed that there are no locational issues and the design sketch applies as though everything occurred at a single location. The discussion illustrates the principles that would apply to implementation in a network with transmission constraints and locational requirements.

Demand

The energy-only market begins with the real-time demand for energy in the wholesale market. In order to capture some of the features of the real system, the demand for electricity shown in Figure 3 divides into two segments. The inflexible demand represents the customers that are assumed not to have individual real-time meters or real-time individual controls. It is not possible for these customers to receive or respond to the incentives provided by real-time prices. Typically these customers pay a fixed price over the period, here assumed to be $30/MWh. Of course, these customers have some implicit demand curve and the load that results at $30 is the load assumed to apply no matter what the spot market conditions.

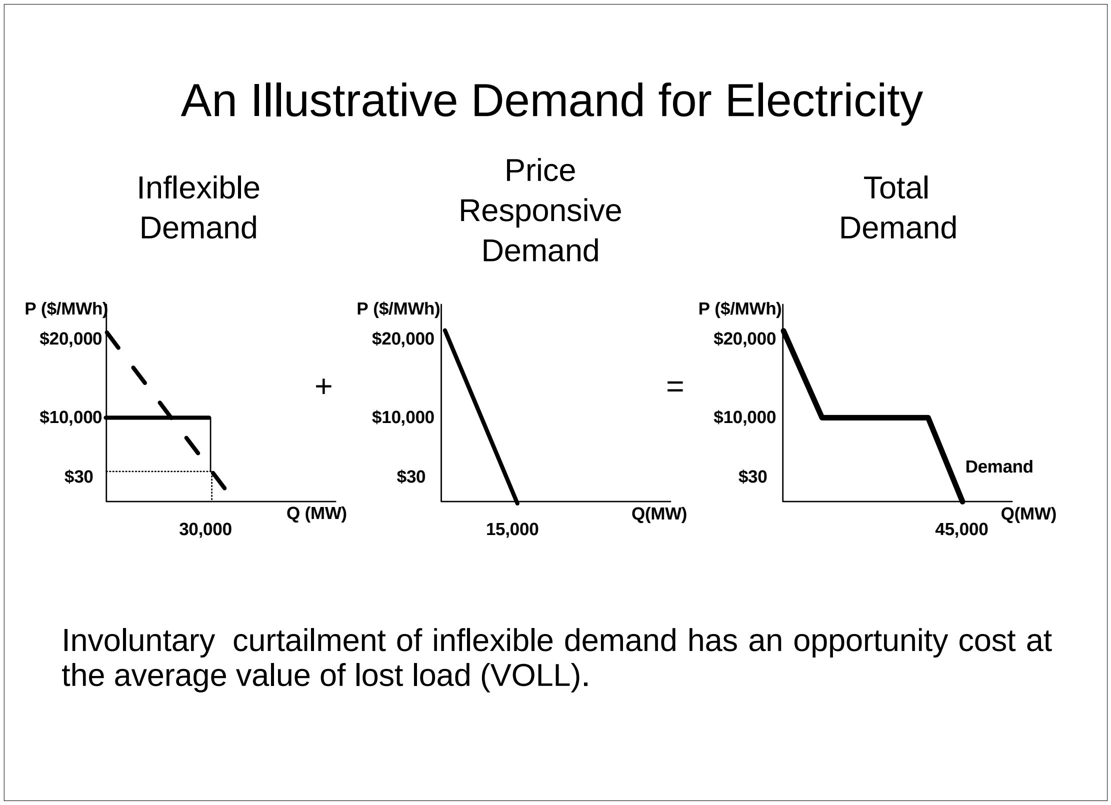


Figure 3

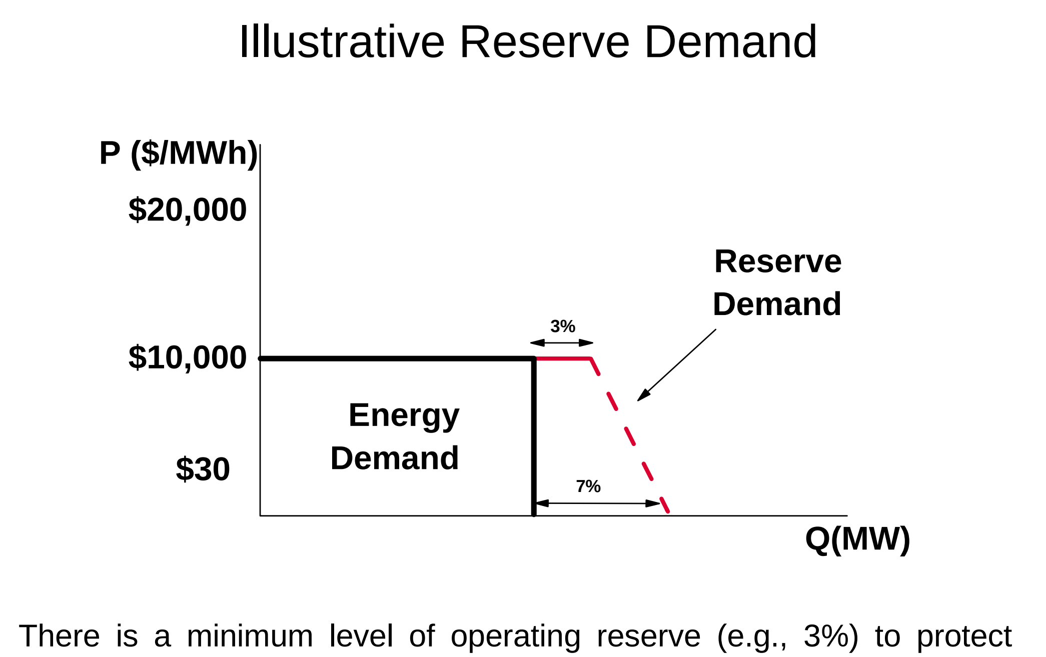
Inflexible demand would maintain this load level unless conditions became severe enough that the system operator must intervene. In this event there would be involuntary curtailment through rolling blackouts or brownouts. The simplifying assumption for the illustration is that the lack of metering and individual controls limits the system operator to random, or at least proportional reductions of inflexible loads. The individual inflexible customers could not be dispatched or metered separately, but the load of the group could be reduced by curtailing all the loads proportionally. Since the price the customers will be paying may still be $30/MWh, this would perforce be involuntary curtailment. The average opportunity cost of the involuntary curtailment would be the average "value of the lost load" (VOLL) defined by the implicit demand curve, which represents the correct estimate of the cost of curtailment given the limits on control of inflexible load. This average VOLL is assumed to be $10,000/MWh in the example. Of course, some of these customers would be willing to pay more, but some who are curtailed would also be prepared to be curtailed for much less than the average opportunity cost. Under the circumstances, the marginal cost of curtailment for the inflexible group is the average cost of the involuntary curtailment, $10,000/MWh.[14]

---

[14] Steven Stoft, Power System Economics, IEEE Press, 2002, p. 149-150.

Who provides this estimate of average VOLL? Absent some means of credible declaration by the customers, this implicit demand curve would be estimated by regulators or by the system operator. Hence, there would be an administrative determination of what should serve as the average VOLL. This average VOLL price and the proportional curtailment in effect convert the implicit demand curve of the inflexible load into a horizontal demand curve as shown in the left panel in Figure 3.

By contrast, the flexible load would have appropriate metering and control. The control might arise through the dispatch actions of the system operator or through the decentralized choices of the loads given their estimate of prices. These flexible customers would be able to bid demand in real-time and follow signals to reduce or increase load when prices were high or low. This demand curve would arrive naturally through the bids or choices of the load, and would require no further determination by the system operator. The combination of the two demand curves through horizontal addition as illustrated in Figure 3 results in the demand for the electricity to be delivered to the customers.



There is a minimum level of operating reserve (e.g., 3%) to protect against system-wide failure. Above the minimum reserve, reductions below a nominal reserve target (e.g., 7%) are price senstive.

Figure 4

A characterization of electric energy demand would not complete the demand story. In addition to the electric load at any moment, the system operator must obtain an appropriate level of operating reserves. For instance, Figure 4 illustrates operating reserves determined according to two rules that would combine to provide a reserve demand curve in this energy-only market. To isolate the reserve demand curve, take the level of energy demand as fixed. First assume there is a minimum level of reserve, set here at 3%. The criteria for determining this minimum requirement would be an issue, but is not the point of the discussion here. The constraint could be taken directly from established minimum reserve requirements set by the North American Electric Reliability Council (NERC). Some minimum is needed, for example, to prevent catastrophic failure through a widespread and uncontrolled blackout in the system. The exact level is not important for the illustration, and we return to this issue later. What is important for the energy-only market design is that this minimum level conforms to the actions of the system operator. The system operator would not go below this level of reserves even if this required involuntary curtailment of inflexible load. This level of operating reserves is shown as the solid red line in Figure 4.

Above this minimum level of operating reserves, there would be more flexibility. Other things being equal, it would be better to have more operating reserves. But if energy demand increased the operator would not impose involuntary curtailments in order to maintain a higher level of operating reserve. Under the traditional system, a system operator would reduce operating reserves and turn to load shedding only as a last resort. To do otherwise would be to impose involuntary load shedding with certainty in order to avoid a contingent probability that involuntary load shedding might be required. Hence, recognizing that there is some flexibility in operating reserves is nothing new. What would be different than the traditional approach would be to identify the cost of reduced reserves. [15]

This would provide the second piece of the operating reserve demand. There is a tradeoff that would consider the marginal change in the expected loss of load, valued at the average VOLL, to define the willingness to pay for incremental reserves. Beyond the minimum level of reserves there would be a nominal reserve target, which for the illustration in Figure 4 reaches 7% of capacity above load if the reserves were free. However, at reserve levels above the minimum requirement the operator would not be willing to institute involuntary load curtailments and would instead accept operating reserve levels less than the target level if load were high and reserves were not freely available and free. In this range, as reserves levels approached the minimum level, the price would be increasing for operating capacity needed to meet energy load plus reserves. This is the price sensitive part of the operating reserve demand illustrated by the dashed red line.[16]

Setting operating reserve schedules over the next minutes or hours is a regular activity of system operators. Since the configuration of load and installed generation would already be known or relatively easy to predict, the task of scheduling suitable

---

[15] Steven Stoft, Power System Economics, IEEE Press, 2002. p. 112-113.

[16] The reserve price would be the energy price net of avoided costs plus any spinning costs. There would be an interaction between the demand for energy and the demand for operating reserves. Strictly speaking, therefore, the simplified graphic with the dashed line captures the effect of reserve demand on energy price in the simultaneous clearing of energy and reserve markets.

operating reserves requires much less information than would parallel attempts to specify required installed capacity many months or years in advance. For operating reserves, there would be substantially less uncertainty and the dispatch could be updated quickly as conditions change. This is a familiar exercise for system operators. Equally familiar is the practice of accepting lower reserve levels during periods of generation scarcity relative to load requirements.

Less familiar would be the practice of translating these reduced reserve levels under increased scarcity into economic terms to reflect the implied higher opportunity costs.[17] Although reserve demand curves have been implemented by the NYISO, the experience with explicit demand curves linked to high average VOLL is limited. This additional step to price reserves according to a reserve demand curve would be included as part of implementation of the energy-only market sketched here. The reserve demand curve would be made explicit and included as part of the simultaneous implementation of the energy and reserve dispatch.

The absence of an appropriate operating reserve demand curve is one of the difficulties in market designs that result in *de facto* price caps and missing money. With little explicit energy demand bidding and no recognition of an operating reserve demand curve, the pricing rules default to the most expensive generator offer. With mitigated offers, this can result in a generator running at capacity and price being set at the variable cost with no scarcity rent, even when reserves are reduced. A reserve demand curve would improve the determination of prices in these scarcity conditions. However, if the reserve demand curve does not raise prices towards the average VOLL when operating reserve levels approach the minimum reserve level, then the demand curve is not capable of representing the effects of scarcity or capturing the true opportunity cost at the margin. If the pricing algorithms do not incorporate a demand effect, whether through participants' energy demand bids or through the simultaneous interaction with the operating reserve demand curve, the resulting price determination would be flawed in the periods of scarcity when it would be needed the most. If the price is always set at the running cost of the most expensive generator included in the dispatch, then either there is never any time when capacity is constrained or the pricing calculation is based on a misunderstanding of the meaning of short-run opportunity cost. The operating cost of the most expensive unit running is relevant when there is excess capacity. However, when capacity is short and generation supply is in effect fixed over the range, the demand curve for energy and operating reserve should set the price, and should be able to set the scarcity price at a high level.

[17] Steven Stoft, Power System Economics, IEEE Press, 2002. pp. 197-200.

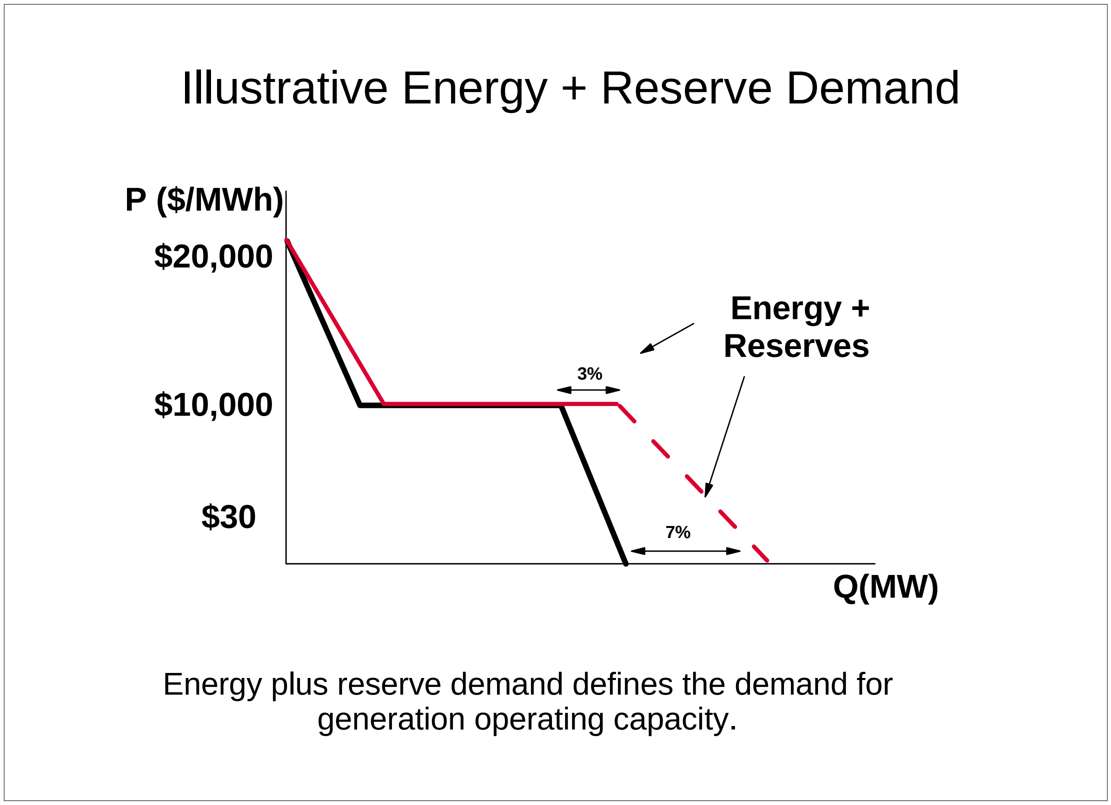


Figure 5

This price sensitivity for operating reserves provides an added slope for operating capacity demand between the minimum requirement and nominal target. When combined with the price sensitive energy demand, Figure 5 summarizes the total demand for energy and reserves. As the market tightens relative to available supply, reserves would be reduced and prices for energy and reserves would rise. When operating reserves reach the minimum level, the price reaches the $10,000/MWh average VOLL and involuntary load curtailments would be required

Supply and Pricing

On the supply side, the system operator would receive offers of available generating capacity with prices for energy and reserves. The system operator would use the load bids and generator offers to determine a bid-based, security constrained, economic dispatch (with locational prices) in the usual way. This would include both bilateral schedules and spot market imbalances priced at the equilibrium market price. During normal operating conditions with moderate load and a high degree of available generating capacity, the system operator would obtain the market equilibrium for energy at a relatively low price. Operating reserves would be well above the minimum level. This equilibrium condition is illustrated in Figure 6. All loads and all generation would clear imbalances at these energy prices. All generators providing operating reserves would be paid the market-clearing price for reserves at the energy price less the avoided variable costs of generation. Under the assumed design here, operating reserve requirements would not be attributable to individual customers, and all loads would be charged a contemporaneous uplift payment to cover the cost of operating reserves and other ancillary services.

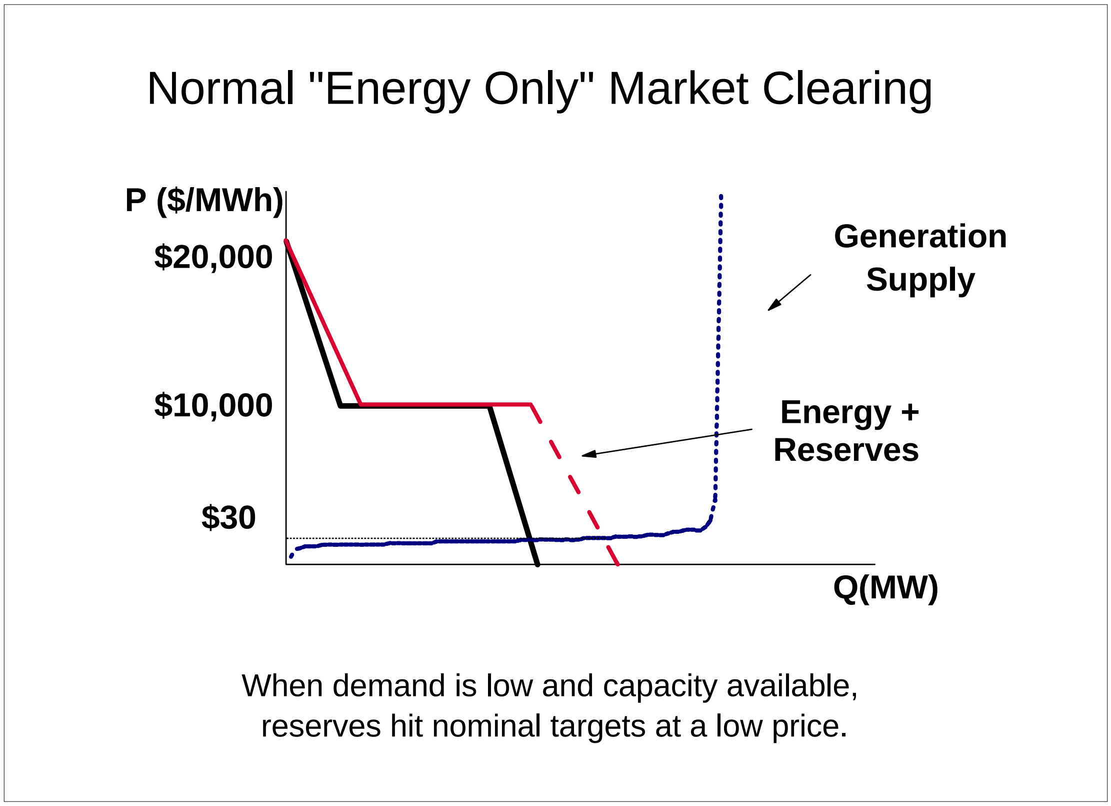


Figure 6

Under stressed conditions there would not be adequate capacity to meet all load and maintain the target nominal level of reserves. This would give rise to scarcity pricing determined by the capacities of the generation offered, the energy demand, and the administrative demand for operating reserves. As shown in Figure 7, the resulting price would be high, here illustrated as $7,000/MWh for energy and essentially the same market-clearing price for operating reserves. This would approach the average VOLL. Flexible customers with real-time metering would respond to the high price signals by reducing load. The system operator would make the decisions to reduce the level of operating reserves. The resulting equilibrium prices again would apply to all imbalances relative to bilateral schedules. Payments for operating reserves would be made to generators providing reserves and the cost would be applied to loads in a proportional uplift payment. All generators providing energy would receive the high energy price. All generators providing reserves would receive this high energy price less the variable cost of the marginal reserve capacity. Although scarcity conditions with very high prices would apply in relatively few hours, the payments to generators during these hours would include a large fraction of the total contribution to fixed and investment costs.

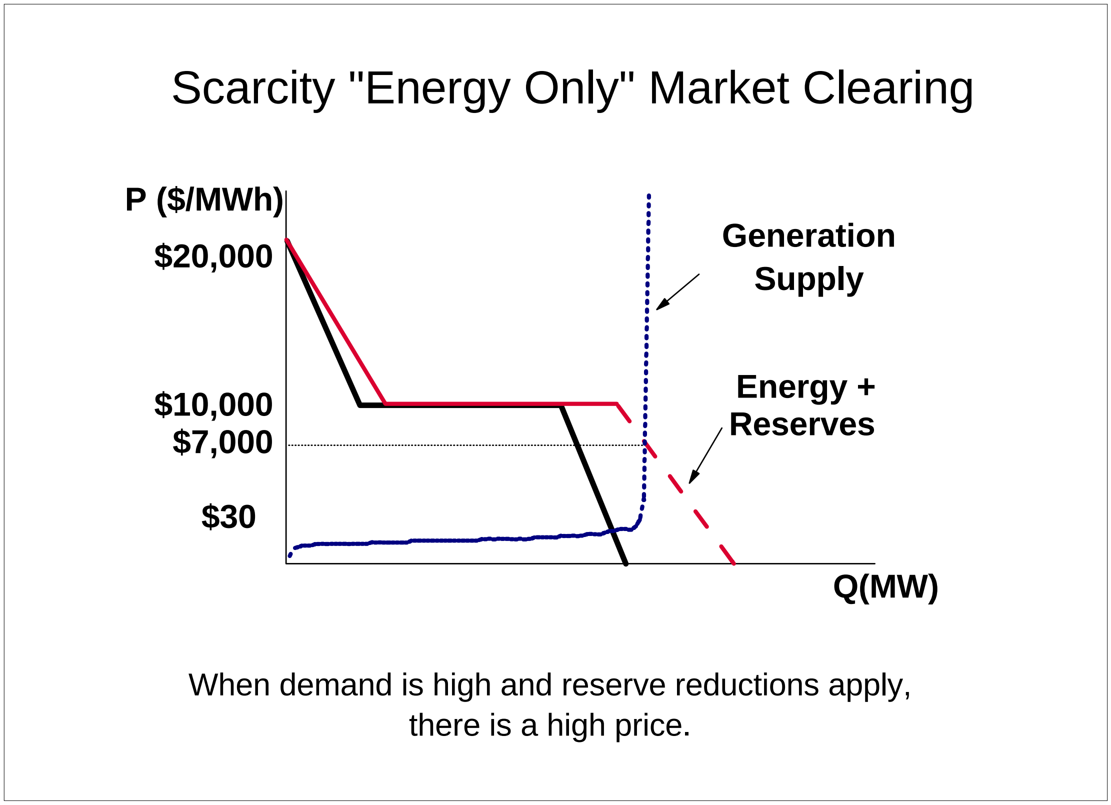


Figure 7

High energy prices during scarcity conditions would approach the average VOLL. If the degree of scarcity reaches the point that reserves are reduced to the minimum operating level, the system operator would turn to random or rotating involuntary load curtailments for the inflexible load. Under these conditions the price of energy would be at the average VOLL, with a corresponding price of reserves. This would continue over the full range of the inflexible load indicated by the horizontal segment of the demand curve in Figure 8. Customers with very high valuations (those above the average VOLL) would have the ability and the incentive to install the meters and controls allowing for real-time pricing to ensure they were not included as inflexible load. Thus the system could produce higher reliability levels for those who would be willing to pay above the average VOLL. However, except when *all* inflexible load would be curtailed, the equilibrium price would not rise above the administratively determined average VOLL. In real systems, curtailing all inflexible load with resulting prices going above the average VOLL would be highly unlikely.

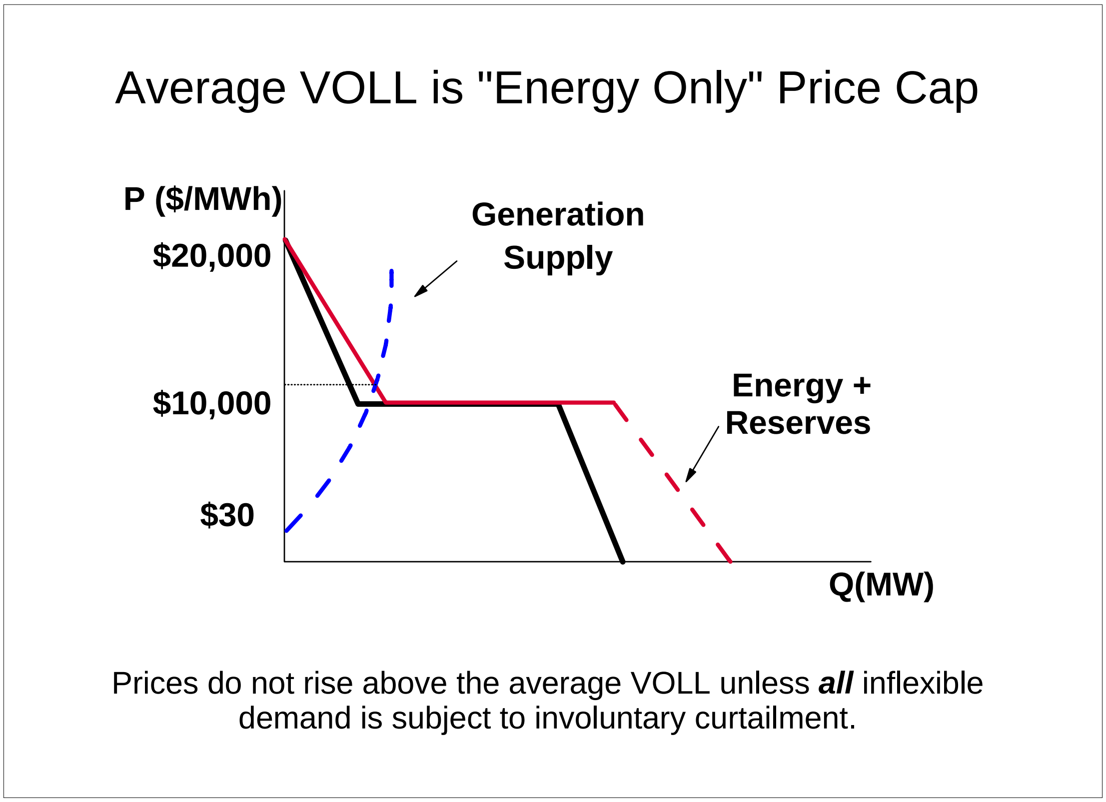


Figure 8

In this sense, it might be natural to refer to the average VOLL as a price cap. Although this is true in the sense of a prediction, it is worthwhile to emphasize that this would not be a price cap in the usual meaning. As discussed above, the average VOLL is the proper measure of the opportunity costs of the involuntary curtailments. Unlike with the usual price cap at a low price, the system operator would not have any offers for generation above this price that should be accepted. There would be no room for and no need for OOM purchases in their many forms to obtain additional energy or to reduce other load in order to avoid the involuntary curtailment. The average VOLL would set the appropriate market price when involuntary load curtailments were required.

The emphasis on connecting the energy and operating reserve demand curves to the average VOLL is not to ensure that this price will be reached often or that curtailments of inflexible load would be common. With an appropriate implementation of the operating reserve demand curve, the opposite would be true. Prices should be higher than in the capped market, but should seldom rise to the level of the average VOLL. The response in the market would obtain more demand response and investment in new capacity. It is the intermediate price responsive segment of the demand curve that would be important, and the connection to the average VOLL provides the anchor at the high end.

With the expected range of supply and demand conditions spanned by the illustration in Figure 6 and Figure 7, the "energy-only" market should produce a distribution of prices as illustrated in Figure 1. In long-run equilibrium, there would be no missing money. Market prices would provide the needed incentives for loads and generation. All loads and generators would settle imbalances relative to bilateral schedules at the market equilibrium price. Prices would at times be volatile, varying substantially over the day and over the seasons. For some hours, prices would be very high.

The anticipation of these uncertain and sometimes high prices would create strong incentives for market participants to contract forward. Under the idealized assumption of no transaction costs, contracts would arise to cover a substantial portion of all load and generation. The precise terms and prices embedded in these contracts would be determined according to the preferences of the market participants. In principle, neither the system operator nor the regulators need observe these contracts nor approve the terms.

Since there would be no missing incentives, there would be no need to devise operating rules to compel competitive actors to act against the incentives they face. Since there would be no missing money, there would be no need for resource adequacy programs designed to provide the missing money. Both generators and loads would be hedged through the forward contracts. Hence, there would be limited exposure to the volatile spot prices. These volatile prices and the price duration curve in Figure 1 would be critical to success of the energy-only market, but the limited exposure to high prices should not provide a critical mass of political pressure to induce further intervention in the market.

Energy-only Market in a Network

The simple energy market depicted in Figure 6 and Figure 7 illustrates the ideas in the context of a single period at a single location. The real system would involve many locations connected by a network and multiple periods. The resulting design would have the now familiar core features of the organized RTO markets.[18] The centerpiece would be a spot market organized as a bid-based, security-constrained economic dispatch with locational (nodal) prices as in Figure 9. The energy-only market includes bilateral schedules at the difference in the locational prices. Transmission hedges appear in the form of financial transmission rights designed congestion revenue rights (CRRs).

---

[18] See the CAISO Market Redesign and Technology Upgrade Program (MRTU), www.caiso.com/docs/2001/12/21/2001122108490719681.html.

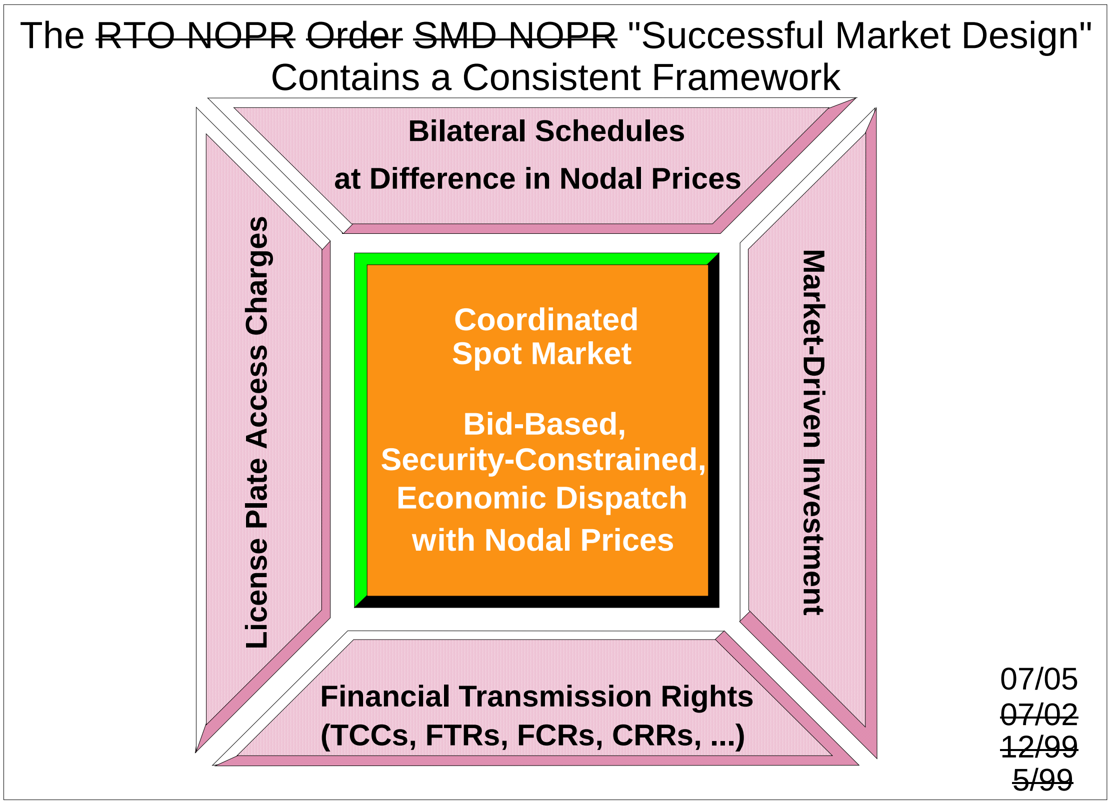


Figure 9

The real-time market could be combined with one or more forward markets such as the day-ahead market (DAM) with the same basic structure supplemented with virtual bids and schedules relative to the real-time market. This day-ahead market could include both a dispatch and a unit commitment process.

As is common, the unit commitment could be supported further by a reliability unit commitment (RUC) process in which the system operator's day ahead forecast would be checked and additional units committed as needed to ensure reliability. The RUC units would receive a bid-cost guarantee if the realized revenues in the energy market were not as large as the bid costs. The RUC is an example of a largely non-market administrative rule driven by reliability requirements but structured in a way to support the remainder of the energy market. The RUC would not involve forward procurement of energy beyond minimum output levels, and hence would not undermine the incentives in the DAM. The same arguments and rules that apply in the current RUC implementations would carry over in the energy-only market design.

The requirements for operating reserves, expressed as location specific operating reserve demand curves, would apply as is now done in the RTO markets. Typically there would be some regional aggregation at a greater level than the individual nodes, and operating reserves must meet certain requirements in that region. Hence, congestion management would require finer nodal specification and pricing, and there would be a different set of aggregation rules for different types of operating reserves. Although the resulting model does not lend itself to the simple graphical addition of demand curves in the stylized illustrations above, there is nothing unusual about including operating reserves in the model with the appropriate mix of aggregation and rules that link reserves to nodal load and generation activities in a network. The resulting locational energy and reserve prices would be obtained through simultaneous optimization of energy and reserve dispatch with pricing to minimize the reliance on uplift payments.[19] The main innovations of the energy-only market design would be in the configuration of the reserve demand curves, connection to the average VOLL, and elimination of *de facto* price caps. In addition, this should eliminate the need for most or all OOM purchases.

Energy-only Market Defects

The outline of an energy-only market, with the necessary administrative rules structured to support market decisions, would raise concerns with regulators and others regarding its promise of providing efficient results and adequate investment. To some analysts the absence of the missing money problem might signal the presence of other market design defects. Further, since the idealized assumptions would not hold exactly in a real system, potential problems could undermine the workability of any attempt to approximate an energy-only market. Here we consider first issues like demand response, reliability and missing markets that could be addressed in principle within the energy-only framework. Market power would be more difficult to mitigate without regulatory intervention, but it is possible to design compatible regulatory interventions that would not disrupt basic operation of the market. A difficult problem would be a lack of forward contracting given both transaction costs for customers and regulatory rules that limit the incentives of load serving entities. A concern with inadequate contracting touches directly on the resource adequacy issues associated with the missing money problem.

Demand Response

The generic demand characterization in Figure 3 includes flexible load that responds to price. A conclusion might follow that an energy-only market would require a substantial degree of formal bidding for dispatchable energy that a system operator could use to respond to scarcity conditions. If there were little or no flexible loads, then prices would bounce between the low operating costs of generators and the high average VOLL. Hence, there is often an assumption that substantial energy demand response would be a prerequisite of an energy-only market implementation.

It is a commonplace that electricity markets would work better with more demand response. This is true of the energy-only market design just as it is true in a price-capped design. Under an energy-only approach there would be more, perhaps much more, incentive for load to acquire the necessary meters and take action to reduce demand during periods of scarcity and the associated high prices. However, despite the benefits, greater demand bidding is not a prerequisite of an energy-only market design.

---

[19] William W. Hogan and Brendan J. Ring, "On Minimum-Uplift Pricing For Electricity Markets," Center for Business and Government, Harvard University, March 19, 2003, (www.whogan.com) .

Strictly speaking, for successful operation of market clearing it would not be necessary that load formally bid into the dispatch. With regular calculation and publication of locational prices, loads could follow the current conditions and make short term forecasts. Loads could adjust their consumption in response to actual or expected real time prices without actions being taken by the system operator. If the metering for real-time load is in place and consumption is settled at real-time spot prices, demand could respond to real time prices without receiving dispatch instructions from the ISO.

Furthermore, a large energy demand response is not a requirement to avoid the binary price outcomes under perverse examples with prices bouncing from low operating costs to the high average VOLL. Such outcomes require that there be no price response in the total demand for energy *and* the demand for reserves. However, as illustrated in Figure 5, even if there were no flexible energy demand response there would by construction be a price response associated with the demand for operating reserves. Of course, energy demand response would be valuable. For any level of capacity that provides a given level of reliability, there is some set of shortage prices that would produce generator revenue streams that if correctly anticipated would be sufficient to sustain that level of capacity.[20] If there is a potential for error in setting these shortage prices or for error by suppliers in translating these shortage prices into expected income, then the amount of energy demand response is a cushion against load shedding arising from errors in estimating future prices and the profitable level of capacity.

There is also less potential for error if the ISO can set an operating reserve demand curve, rather than just shortage prices for reserves, so that small errors in setting shortage prices do not drive the system too far from the expected equilibrium level of prices. And it doesn't take much of a response to move away from a system where prices bounce between the lowest to the highest level towards a system where the prices, while volatile still, trace out a more "normal" response to market conditions.

Hence, while it would be difficult to implement a pure energy-only market with no energy demand response and no pricing of reserves, once operating reserves are included in the "energy-only" market design the same conditions would not apply. More energy demand response would be better, but the energy demand response cart could come after the horse of efficient "energy-only" market design applying an operating reserve demand curve and simultaneous pricing of energy and operating reserves.

Reliability

An energy-only market design defines and prices reliability through the demand for operating reserves. It would be possible to estimate the resulting loss of load probability to compare with the traditional planning standard of one day in ten years for a loss of load event. But the expected loss of load probability would be an output of the market choices more than an input to system design. If the operating reserve demand curve properly reflected the value of reserves, the "correct" loss-of-load probability would follow automatically.

---

[20] Steven Stoft, Power System Economics, IEEE Press, 2002, pp. 167-168.

In a market context, the concept of reliability differs for the two types of demand. In the case of inflexible demand, any interruption of service would be involuntary. This would conform closely to the traditional situation where the planning criterion assumed a given and fixed load and the probability that adequate generation would be available could be calculated. However, for flexible price-responsive demand the same concepts would not apply. As the market tightens and prices rise, flexible demand reduces and this by definition would not be an involuntary loss of load. Reliability becomes reliability at a price for the flexible demand.

Different levels of installed capacity would imply different prices and a distribution of probabilities for any involuntary loss of load. This would be true in both an energy-only market and in a price-capped market.[21] Given the demand curve for operating reserves, the expected long-run equilibrium loss of load probability could be more or less than any given planning standard. To the extent that there is a difference, it might be argued that the demand for operating reserves should be adjusted to match the planning standard.

However, an alternative interpretation would be that the operating reserve demand curve provides the better representation of the appropriate level of reliability. The direct consideration of the value of operating reserves captures the tradeoff that could be included only indirectly in the traditional planning standard. For example, suppose the equilibrium loss of load probability that would result from a given operating reserve demand curve would be higher than the under the traditional standard at the target level of installed capacity. This might be argued to require an increase in the demand for operating reserves in order to increase reliability. In effect, this argument would say that the price of reserves should be higher than the tradeoff defined by the operating reserve demand curve. But this argument would contradict the definition of the demand curve. If the operating reserve demand curve captures the value of reliability, the resulting expected equilibrium loss of load probability should be the standard.

This stands in contrast to an alternative approach that would fix the installed capacity requirement and determine the operating reserve demand to produce enough revenue to support investment that would meet the installed capacity target. In an energy only market, or in a price capped market, it would always be possible to set the operating reserve requirement at a level that met the expected revenue requirement and eliminated the missing money problem for a given level of installed capacity.[22]

The indirect route of specifying the expected loss of load probability should not replace the direct determination of the willingness to pay for operating reserves. If the expected loss off load probability were less than the traditional planning standard, a

---

[21] For an analogous discussion in the context of the ISONE LICAP design, see Steven Stoft, Testimony on Behalf of ISO New England on Locational Installed Capacity Market Proposal, Submitted to Federal Energy Regulatory Commission, Docket Number ER03-563-030, Rebuttal February 10, 2005, pp. 35-36..

[22] The energy-only operating reserve demand curve is not derived from a revenue target. The combination of energy demand and reserve demand reflects the value of the energy and reserves. This is distinct from the illustrative dichotomy between VOLL pricing and operating reserve (OpRes) pricing described in Steven Stoft, Power System Economics, IEEE Press, 2002, pp. 108-200.

similar argument would give deference to the operating reserve demand curve. Hence, while an estimate of the distribution of loss of load probabilities might inform the design of the operating reserve demand curve in judging the adequacy of reserves, by forcing questions about the realism of the operating reserve demand curve, any conflict between the two would ultimately be resolved in favor of the operating reserve demand curve. Once the willingness to pay for operating reserves is made explicit and accepted by the regulators and system operator, there would be no need to further reconcile the expected loss of load probability with the operating curve demand curve.

Other things being equal, use of the energy-only market design should reduce the cost of meeting the reliability requirements. The flexible demand provides an important added tool to meet reliability standards. The demand response should be more than found in traditional utility systems or that could be provided in price-capped markets. Without the incentive of spot market prices that reflect the real scarcity conditions, it is more difficult to design market mechanisms that provide a flexible demand response. Hence, for any given cost level the energy-only market design should produce greater reliability in terms of reduced curtailments of inflexible demand.

A similar conclusion applies to a concern that an energy only market and the associated operating reserve demand curve would produce higher average costs that somehow might be avoided. The same issue arises under ICAP systems where the capacity payments appear to raise costs to customers. However, if reliability is to be maintained, some of these costs cannot be avoided. The traditional system included these costs in the rate base for the portfolio of generation owned by the utility. The costs were there, but they were less visible than explicit capacity payments or market-clearing energy prices. The costs would not be created by the ICAP requirements or operating reserve demand curve. It is the reliability requirement that gives rise to the costs. An ICAP requirement or an energy-only market would be different means to provide the payments and achieve the reliability objective.

Implementation of a demand curve for operating reserves would require attention to adapting the standards and information from the traditional model under the NERC rules. "Grid reliability is a difficult issue to discuss objectively, because few metrics describe and measure bulk system reliability consistently across the nation."[23] While this would be important, there is nothing in principle that should deter use of an energy-only market design. The market should reinforce reliability. The reliability rules and definitions may have different impacts in a market context, but the goals of reliability are not add odds with the market design.

The NERC standards already follow a structure that includes minimum operating reserve levels that should not be breached and where controlled but involuntary load curtailments should be imposed. At the minimum levels, reserves would be purchased at any price up to the price cap. However, when the price cap is below the average VOLL, this creates a conflict that produces the missing incentives and the missing money. In the

[23] Phillip G. Harris, "Relationship between Competitive Power Markets and Grid Reliability: The PJM RTO Experience," Issue Papers on Reliability and Competition, US Department of Energy and Natural Resources Canada, August 2005, pp. 4. (www.energetics.com/meetings/reliability/papers.html)

energy-only market design, a similar rule could apply regarding the minimum operating reserve levels, but the maximum price for reserves becomes the average VOLL because generation that costs more is more expensive that involuntary load reduction.

Above the minimum operating reserve levels, there is still a value for incremental reserves but the value decreases at higher levels of reserves. Explicit pricing of these incremental levels would be part of the design to eliminate the missing incentives and missing money. The operating reserve values would be set by regulators and the system operator to capture expected impacts on the system. This should include the total change in system costs across the integrated grid if everyone followed the same principles.

Although the structure of the reserve requirements might be compatible with the existing NERC standards, the level of operating reserve requirements and associated prices might turn out to be different. The NERC standards were designed for a different setting, where the value of reliability was less explicit than would be the case with a realistic operating reserve demand curve. Explicit consideration of the tradeoffs would be required for the energy-only market design, and might change the approach that NERC takes in the new era of mandatory reliability standards. With the energy-only market design, NERC's enforcement problems would be reduced because the market incentives would be compatible with the reliability requirement.

Missing Markets

The missing money problem created by limiting scarcity pricing provides an example of a missing market. There could be a market for reliability, but the regulatory constraint prevents its operation. The implicit price caps on energy and reserves present the most significant problem of this type. As a practical matter, introducing an energy-only market design with explicit energy and operating reserve demand curves may be all that is really needed to provide adequate incentives for investment in generation and other resources.

However, there are other ancillary services that are essential for successful operation of energy systems. Black start capability, regulation services, and voltage support through reactive power management are prominent examples. These services are necessary and must at least be procured by the system operator. The compensation rules for providers of these services must be adequate. However, the total expenditure may be modest relative to the energy and operating reserves under the energy-only market design.

These other ancillary services may be amenable to targeted compensation schemes that do not much affect the remainder of the energy market design. To the extent that this is not true, then it would be important to extend the analysis of the energy-only market design to include efficient pricing and incentives for these additional services. For instance, it would be possible to consider spot-pricing and forward contracting for reactive support.

Market Power

A primary concern that drove the public policy decisions towards administrative measures like price caps was the possibility of market participants possessing and exercising market power. The immediate effect of market power was cast as high prices that could not be justified by input costs. It seemed logical that the most direct route to controlling high prices would be by imposing limits on prices. This in turn created an array of other measures such as out-of-market purchases and reliability must run (RMR) contracts. And the low prices created the missing-money problem with the attendant call for resource adequacy programs.

The concern with market power would remain in an energy-only market. With no limit on energy and operating reserve prices other than the average VOLL, there would be even greater fear about potential incentives to exercise market power. The exercise of market power would violate the assumption of competitive behavior that underpins the efficiency of energy-only pricing.

The ability of generators to enter the market with new capacity supported by voluntary contracts with consumers should make the long-term energy market workably competitive. Without artificial barriers to entry, no special policy would be required to address market power in forward contracting with a sufficiently long horizon that allows for entry.

The problem would then be in the short-term spot market, especially in the presence of transmission congestion that created load pockets where generators might have substantial market power and would be able to raise prices above competitive levels. This is a large topic with many details, but the essence of the relevant points for the energy-only market design is straightforward. The market design could include administrative intervention when and where there was a serious possibility of an exercise of (local) market power through physical or economic withholding. However, the interventions would be structured to emulate the results of competition to the greatest degree possible. These interventions would be in the form of offer caps and offer requirements for generators, with appropriate exemptions for all generators who are not in a position to exercise market power or who enter a market with new facilities.

Deciding on the level of the appropriate offer caps would be contentious, but the focus would be on preventing withholding and not on keeping prices low. With the full capacity of a mitigated plant in use, the market-clearing price would be determined by the opportunity costs reflected in the demand curves and seldom by the offer cap *per se*. Setting offer caps is even more contentious in a price capped market where the effect may be to reduce supply. Furthermore, the energy-only design does not necessarily increase any incentive to exercise market power. For the same demand response, a higher price increases the incentive to produce above any given level of output. Under these conditions, when the supplier hits its capacity constraint there is no incentive to exercise market power. To the extent that the energy-only market design increases the total energy and operating reserve demand response, the design would help mitigate market power.[24]

These types of market power interventions are familiar in the organized markets under ISOs/RTOs. The principal difference with the energy-only market design would not be in the form of the market power intervention. Rather, it would be in the treatment

[24] The common assertion that scarcity conditions increase market power depends more on an assumed movement to a nearly vertical residual demand curve than on the absolute level of price.

of scarcity pricing after mitigation to address market power. In particular, the operating reserve demand curve would allow for scarcity prices that could be very high, but would not arise from the exercise of market power. Even with offer caps in place, high demand and limited available generation capacity would create a shortage of reserves and higher prices for both energy and reserves. Hence, there would be no formal cap on prices, only limitations on economic and physical withholding for generators with market power. Market power mitigation would be targeted at the exercise of market power, not at high prices. Prices would be high during scarcity conditions and there would be no missing money.

Inadequate Contracting

If there were an adequate level of forward contracting, the contracts would reflect the expected prices going forward but customers would face relatively little exposure to the volatility of prices in spot markets. If regulators were confident voluntary contracting would suffice, the energy-only market with its voluntary forward contracts would suffice.

A principal concern of regulators could be that left to their own devices market participants might not select an appropriate level of contracting. In effect, this would be a consequence of a violation of the assumption of low transaction costs. There could be barriers to entry into contracting, particularly for small customers. Without sufficient hedges supplied through forward contracts the loads would be too exposed to volatile spot prices and this, in turn, would create inevitable pressures for regulators to intervene when scarcity appeared. This intervention would be a political challenge for regulators and would create associated regulatory uncertainty that would undermine investment.

To the extent that forward contracts would be needed by generators to maintain existing facilities or arrange the financing for new investment, insufficient demand for forward contracts could work against the intended incentives for market based resource investments. Hence, inadequate demand for forward contracts could translate into a resource adequacy problem relative to the level of investment that would occur if the transaction costs could be eliminated.

A concern with inadequate forward contracting might arise from an expectation of market failure with many small customers who are unable or unwilling to enter into forward contracts. The group illustrated above in Figure 3 as the inflexible load might have little incentive to contract. The price to the inflexible load might be fixed and probability of curtailment ignored. The load serving entities providing the inflexible load's power at fixed prices may have some incentive to contract forward, but this would depend on the regulatory design at the retail level even though the effects would be felt in the wholesale market.

Without elaborating every case, there are instances when regulators would be inclined to require forward contracting on behalf of some or all customers. In the markets where there is a missing money problem, this contracting directive moves almost immediately to contracting for installed capacity in the ICAP mode. However, in an energy-only market without a missing money problem there would be other ways to approach mandatory contracting. If the problem is forward contracting and not missing money, then the regulatory approach to forward contracting could focus on the objective and be less prescriptive about the means to achieve the objective.

Mandatory Load Hedge Contracting

The main targets for regulatory intervention would be the problems of market power and inadequate contracting. If market power could be contained with sufficient mitigation in the spot market, there would be no concern with the possibility of withholding, and the incentives of the energy-only market would provide a powerful force to make plant available when most needed. This need for mitigation is not unique to the energy-only design, and well-designed mitigation should address the market power problem.

If somehow adequate contracting could be arranged, then there would be protection for loads that would be hedged against high prices and suppliers that would avoid exposure to volatility. The challenge arises in specifying the requirement for and design of forward hedging contracts.

An "energy-only" market design could accommodate a mandatory load hedge (MLH) requirement. This would be a regulatory intervention to address the concern that there would be inadequate forward contracting. The details of MLH requirements would be important, but the critical issues introduced by an energy-only market approach would be relatively limited. The comparison with an ICAP design helps identify why removing the missing money problem would simplify the policy intervention.

ICAP and MLH

An ICAP requirement includes forward contracts with specific generators for installed capacity, sometimes with an explicit or implicit option on the energy that could be produced from the designated plant. Among the concerns with ICAP programs have been that the forward requirement horizons are not long enough to support investment, and not specific enough to support the right investments. This produces pressure to extend the horizons and specificity as to type and location.[25] Similar challenges would face any forward contracting requirement, but if the objective of MLH is framed relative to the energy-only market design, there would be important differences with the ICAP approach.

For both ICAP and MLH approaches, an initial step in specifying the requirement would be to identify the targeted load and the intended duration of the forward contracts. For the intended forward contracts, the procedure would determine the forecast load level, locations, and horizon. Although this would not be an easy task, assume that the profile of loads has been identified for each location. Under an ICAP program, this is the beginning of the process. To move load levels at specific locations to installed capacity at other locations, the load forecast must be converted into a description of the mix of transmission and generation capacity that would be required in order to meet this load requirement. The ICAP program may include specifications for demand-side alternatives with their own locational and operating characteristics. For a forecast many years ahead, this translation from demand to supply would be a complicated process with many uncertainties.

---

[25] John Chandley, "ICAP Reform Proposals in New England and PJM," LECG, Report to the California ISO, September 2005.

Under an MLH approach, however, the load forecast could be enough. There would not be a requirement to convert the forecast into a prescription for production capacity by location. The system operator might be in the best position to describe correlated generator outage risk, transmission outages, weather volatility that would determine expected prices in the energy-only market. These studies could support analysis by investors deciding on where and how much generating or other capacity to build given the resulting expectation of prices. But the assumption of the energy-only market design would be to rely on the investment decisions by the market participants even when these may differ from the choices of the regulators or system operators.

For simplicity in making the distinction with ICAP, assume the regulator requires customers to arrange for energy forward contracts that met the same forecast load used to drive an ICAP requirement. The MLH contracts could be arranged through either direct negotiations or through a formal auction process. The critical distinction relative to the ICAP approach is that these MLH forward contracts would be based on prices and delivery at the load location. In effect, these would be equivalent to financial "contracts for differences" relative to the real-time locational price at the load point. To support these contracts the supplier could arrange bilateral contracts to deliver the energy to the load at the load location, or the supplier would settle for imbalances at the locational price.[26]

Unlike with the ICAP programs, under the MLH there would be no need for the regulator or the load serving entity operating on behalf of the customer to arrange for transmission delivery or link the contract to any particular generating facility. These decisions would all be handled by the market participants who agree to be the suppliers under the contracts. The costs and risks of providing the MLH hedges would fall on the supplier and be reflected in the forward contract price. However, competition among

[26] Proposals to use financial contracts in resource adequacy programs are not new. For example, see Harry Singh, "Call Options for Energy: A Market Based Alternative to ICAP and Energy Price Caps," PG&E National Energy Group, October 16, 2000. However, most discussions of resource adequacy proposals employing forward contracts explicitly reject a strictly financial interpretation because the absence of efficient, energy-only pricing precludes the contracts from providing the right incentives for investment or operations. This missing-money problem dictates the need for the physical connection with specific resources that spawns the administrative complexity of ICAP programs. For these quasi-financial contract designs, see Miles Bidwell, "Reliability Options," Electricity Journal, Vol. 18, Issue 5, June 2005, pp. 11-25. Hung-Po Chao and Robert Wilson, "Resource Adequacy and Market Power Mitigation via Option Contracts," Electric Power Research Institute, Draft, March 18, 2004. Shmuel Oren, "Capacity Payments and Supply Adequacy in Competitive Markets," VII Symposium of Specialists in Electric Operational Systems Planning," Basil, May 2000. Shmuel Oren, "Ensuring Generation Adequacy in Competitive Electricity Markets," University of California at Berkeley, April 2004. Shmuel Oren, "Capacity Mechanisms for Generation Adequacy Insurance," CPUC-CEOB-CAISO Installed Capacity Conference, San Francisco, California, October 4-5, 2004. Roy J. Shanker, "Comments on Standard Market Design: Resource Adequacy Requirement," Federal Energy Regulatory Commission, Docket RM01-12-000, January 10, 2003. Carlos Vázquez, Michel Rivier, and Ignacio J. Pérez-Arriaga, "A Market Approach to Long-Term Security of Supply," IEEE Transactions on Power Systems, Vol. 17, No. 2, May 2002. Carlos Vázquez, Carlos Battle, Michel Rivier, and Ignacio J. Pérez-Arriaga, "Security of Supply in the Dutch Electricity Market: the Tole of Reliability Options," Report IIT-03-084IC, Comillas, Universidad Pontifica, Madrid, Draft Version 3.0, Madrid, December 15, 2003.

suppliers should drive innovation and efficiency to make the aggregate forward arrangements capture the expected benefits of the market.

Furthermore, compliance tracking under the MLH contracts would follow automatically in the energy-only market settlements of imbalances. There would be no need for the system operator or anyone else to monitor the generators or ensure the availability or deliverability of any particular generator's capacity. The market itself would provide strong incentives for the suppliers to make these arrangements in an efficient and least cost manner.

The myriad forecasting and monitoring requirements of the ICAP approach, which tend to recreate the problems of integrated resource planning, would be replaced by financial MLH contracts defined at the customer's location. A focus on financial contracts emphasizes the need for enforcement of the financial obligations. The enforcement provisions would be addressed by credit requirements that for the most part would also be present in ICAP markets. In this regard, the complexity of "physical" contracts in ICAP markets does not remove the need for enforcement of financial obligations. The typical ICAP design includes penalties for performance failure, and in the end the ICAP approach carries with it much of the baggage of the financial contract without the simplicity.

In the absence of transparent scarcity pricing of the type found in an energy-only market, a major problem for ICAP markets is to ensure compliance during periods of stress, and to make sure the requirement is broad enough and enforceable so that there is no leaning on the system. If generators could turn to a capped energy price during periods of stress, there would be strong incentives to withdraw from the ICAP obligations and pay the low damages at the capped price. Furthermore, parties outside the formal program would attempt to purchase from the capped spot market during periods of stress. It is a major challenge in ICAP markets to devise sufficient penalties, export controls, delivery obligations and enforcement mechanisms to ensure that the intended generation capacity, transmission deliverability or demand response is really available during periods of aggregate scarcity when real opportunity costs exceed the price cap. By contrast, the MLH requirement could be targeted to a part of the load, and regulators would not be faced with the problem of dealing with market participants outside their jurisdiction or outside the intended target of the program. In other words, the MLH requirement could focus in particular customer classes (e.g., residential and not industrial) and regions (e.g., inside the state and subject to state regulation).

The gold standard for ICAP programs is a set of penalties and enforcement rules that attempt to emulate the incentives of the energy-only market.[27] Of course, by construction the energy-only market creates these incentives naturally as part of its inherent design. There would be no need to have special monitoring and enforcement for generators during periods of scarcity because in the energy-only market this is precisely

---

[27] For example, the LICAP proposal of ISONE attempted to emulate as much as possible the incentives of an energy-only market through the design of monitoring and performance incentives. Steven Stoft, Testimony on Behalf of ISO New England on Locational Installed Capacity Market Proposal, Submitted to Federal Energy Regulatory Commission, Docket Number ER03-563-030, Direct August 31, 2004, Supplemental November 4, 2004, Rebuttal February 10, 2005.

when (very) high prices provide (very) strong incentives to perform. There would be no need for export limitations because the export loads not covered by the MLH obligation would be paying the real opportunity cost of their demand. Likewise there would be no concern with loads not covered by the MLH requirement leaning on the system because there would be no where to lean, only to stand straight and pay the full opportunity cost in the energy-only market when scarcity conditions arose. The MLH approach would provide the hedges through contracts that apply only to the parties that have contracted.

This structure would be consistent with retail access systems like the Basic Generation Service (BGS) in New Jersey.[28] The BGS involves an auction for forward procurement of energy delivered to the load. There is no specification in the contracts about how or where the supplier obtains the power or hedges its obligations. This is left to the supplier's activities in the market and the supplier's risks are internalized in the offers it makes in the auction. The rolling horizon is three years for the smallest customers and one year for larger commercial customers, with the alternative to opt out of the protection. The largest commercial and industrial customers are not included except for uplift and ICAP payments. In the context of an energy-only market design, where there would be no ICAP payments, it would be natural to ask if anything in addition to the BGS program would be required.

This outline of the relative simplicity of an MLH approach in an energy-only framework follows from its targeting of protections. In a price-capped energy market, the capped prices apply in theory to all spot transactions. In order to achieve the benefits of the forward ICAP contracts and associated investment, regulators would face the broad choice between either contracting enough so that spot prices naturally stay below the price cap for the free riders or restricting access to the capped prices through exclusionary rules that would apply during periods of scarcity. The free riders would create very high costs for those bearing the burden of forward contracting. The exclusionary rules would exacerbate the complexity and incentive problems that undermine the very purposes of the market.

By contrast, the energy-only market would not provide hedges through capped spot prices that would apply to everyone. Hence, forward contracting requirements could be limited to a targeted group. There would not be the great concern with free riders. Reaping the benefits of the forward contracts would not require exclusionary rules.

The focus on financial contracts provides great flexibility for suppliers to craft generation, transmission and demand efficiency packages to support their commitments under the MLH contract. There need be no rules to overcome missing incentives or decide on the tradeoffs between supply and demand alternatives.

In short, an MLH requirement would be an intervention by regulators to address a concern that, despite the strong incentives of the energy-only market, loads would not have sufficient interest or incentive to arrange the long-term hedges that would eventually prove necessary. But the intervention would be tailored to fit the intention of using

---

[28] For a description of the New Jersey forward contracting requirement, see http://www.bpu.state.nj.us/home/bgs.shtml

market pricing to drive investment decisions. The required contracts would be financial instruments without explicit connection to any particular resources.

MLH and Other Contracts

Energy-only market pricing would be essential in relying on these relatively simple MLH instruments. Absent credible scarcity pricing of the energy-only market, any MLH requirement would confront the same perverse conditions that would compromise incentives and generator performance precisely when needed the most, during times of scarcity. An energy-only market approach could come as part of a package, with some form of MLH contracts to provide the needed hedges and the forward contracts underpinning investment in new generation (and transmission) capacity.

The generic outline of an MLH approach leaves open many details about the particulars of contract design. The essential features would be an (i) energy-only market design, (ii) specification of the obligation in financial rather than physical terms, and (iii) linking the financial terms to the prices at the load location. Within this framework there could be a variety of contract requirements that would satisfy the limited objectives of hedging sufficiently to provide the protection sought by regulators and to support the intended investment. The MLH approach would not face the additional demands of assuring operational performance or mitigating market power. These would be handled by the incentives and mitigation rules in the spot market, respectively.

The MLH form of financial contract would have much in common with the familiar "portfolio" contracts with liquidated damages (LD) found in electricity markets. The LD contracts do not identify particular resources, and the obligation is to deliver the energy from somewhere or pay the liquidated damages. The damages are often determined as the spot-market prices. In the context of a spot market, the LD contracts are financial instruments, and in a price-capped market the damages are capped. Hence, a principal complication under the price-capped energy markets is that the liquidated damages are too low, and the supplier has an incentive to lean on the spot market and pay the capped spot price during times of scarcity. As a result, the price hedging value of the LD contract remains for the customer but the contract does nothing to eliminate the missing money or the risk of involuntary curtailment. With only the obligation to deliver the energy or pay the low penalty, the equilibrium price for LD contracts would not substitute for the missing money in payments to generators. The money would still be missing from the price-capped energy market and would not be provided through the LD contract. In a price-capped market, LD contracts would be part of the problem.

By contrast, under the energy-only market approach these LD contracts might well be compatible and workable hedging instruments to substitute for some or all of the MLH requirements. To the extent that the LD requirement specifies the payment obligation as it appears in the market, movement to an energy-only market would increase the *de facto* penalty payments in the LD contracts. If these payments were also keyed to the locational price of the load, then the LD contracts could be fully included in meeting the MLH requirement. In an energy-only market, LD contracts could be part of the solution.

Under the energy-only framework, the function of the MLH requirement is more limited and the scope of the contracts more flexible than for the capacity contracts in a comparable ICAP design. Under the energy-only market design, there is no missing money and forward contracts would not carry the burden of providing additional payments above the forward price of energy. Under the energy-only market design the focus of market power mitigation would be on the spot market and not be imposed as a design constraint for the forward contracts. Under the energy-only market design spot prices would provide the incentives for generator availability during periods of scarcity and would not require performance features on the contracts.

The test of the adequacy of the MLH requirement would be in the regulator's judgment that the contracts provided sufficient hedging on average and were of long enough duration to support investment in generation and other resources. For example, it might suffice to specify the MLH requirement in terms of peak and off-peak energy blocks that follow the common pattern of bilateral arrangements. The total energy over the month for peak periods would be set, with a separate requirement for off-peak periods. The load would have the discretion as to when to exercise the contract within the period. Suppliers might contract to provide the hedge for some or all the energy and for some or all the period. This flexibility would avoid the need to specify in advance exactly which hours would apply for availability of particular quantities, avoiding some of the complications of the ICAP programs.

In addition, regulators may wish to leave some customers with a certain amount of discretion while ensuring that there is a minimal level of hedging. Hence, the MLH requirement might be specified as a call option at a high price for the contracted energy over the period. Hedging contracts with lower strike prices would suffice to meet the requirement. Further, the MLH specification could have different requirements for different groups. For example, commercial customers might be required to arrange at least the call options at a high price, while regulators might require smaller residential customers to have full requirements energy contracts at a fixed price. The challenge would be to devise a set of acceptable MLH requirements that did not require extensive review in substituting alternative contracts to meet the requirement.

The New Jersey BGS system employs flexibility in both duration and application to different customer groups. The rolling horizon for the residential load means that one-third of the forecast load is contracted anew three years in advance. Large customers have no forward obligation and can seek their own hedging arrangements or rely in part or all on the spot market.

An alternative type of flexibility might be to formulate the MLH contracts as synthetic tolling arrangements. In effect, the contract would set a price for delivery of electric energy to the load and a price for delivery of a mix of input fuels to the supplier at a standardized location. This would provide the load with a long term hedge for the non-fuel costs ("capacity") but exclude the hedge for fuel costs. This would allow for a separate decision by the load to hedge the fuel costs, perhaps closer to the time of delivery. The suppliers would avoid taking on the complete long-term fuel price risk. This might be a more attractive allocation of risk for both parties.

Since the economic incentives for investment and for operations flow from the energy-only market design and transparent scarcity pricing, the direct purpose of MLH contracts would be hedging and need not address the operations problems under the ICAP approach. There would be few limits on the flexibility of MLH requirements. The details would address the matters of duration, amount, strike price, credit requirements, and substitution rules. But as long as the requirements were specified as financial contracts relative to the load location and real-time price, the contracts would be largely independent of the remainder of the market design and operation. The contracts would impose minimal requirements on the system operator.

Transition

The sketch of an "energy-only" market design sets a destination but does not define the path. It is common in discussions of ICAP proposals to accept this destination as a goal but not consider how to get there, partly because the goal is not adequately defined. With a common understanding of the objective, it would be easier to make choices along the way.

The transition would be important and it would depend in part on the starting point. If there is no existing ICAP program and no imminent capacity shortage, the focus would be on implementing the critical reforms of the spot market design to include transparent scarcity pricing. Similarly, if there is no load hedging program in place, then this would be a focus of a regulatory decision to either accept reliance on voluntary forward contracting as politically sustainable or turn attention to mandatory load hedging for some of the customer classes.

In regions with ICAP reforms underway, and real fears of immediate capacity shortages, the attractions of an energy-only market may require a period of confidence building. However, there appears to be nothing that dictates that an improved spot market design is mutually exclusive of an ICAP approach. The absence of transparent scarcity pricing makes an ICAP program necessary and more difficult to implement. But transparent scarcity pricing for energy and operating reserves would simplify many of the monitoring and performance problems that come with an ICAP approach.

A goal of moving to an energy only market design should influence the design of any ICAP program. The better the scarcity pricing the less burden there would be on designing performance standards, the easier it would be to develop demand alternatives, and the easier it would be to set criteria for phase out of the ICAP system. For example, recent ICAP reform proposals such as the Reliability Pricing Model in PJM set the demand for installed capacity with the price determined net of an estimate of the revenues that should be earned in the spot market.[29] With transparent scarcity pricing and an appropriate operating reserve demand curve (plus a compatible installed reserve planning target and a few other simplifying assumptions) this net price of capacity would be zero and the role of the ICAP requirement would fade, or at least be substantially reduced.

[29] John Chandley, "ICAP Reform Proposals in New England and PJM," LECG, Report to the California ISO, September 2005.

A common step would be to address directly the interaction between reliability standards and market design. The Energy Policy Act of 2005 requires FERC to propose rules to establish a new Electric Reliability Organization (ERO), set mandatory reliability standards and provide for enforcement.[30] These new standards could be compatible with the principles of the operating reserve demand curve, or not. Everyone would benefit if the standards and rules were written in explicit recognition of the charge to have markets reinforce reliability.[31] Poorly designed markets and poorly designed reliability standards would make everything harder.

An explicit consideration of the destination and the transition path would be important for mitigating regulatory risk. Market observers regularly identify uncertainty about regulatory rules and pricing as a principal obstacle to investment. Hence, despite any promises that the rules will be stable, an unsustainable system will be seen for what it is and the confidence required to support investment will be impossible to mandate. The basic elements of a successful market design are not a mystery, and experience shows that deferring attention to market design increases the likelihood and cost of the failures, requiring ever more interventions and changes of the rules. It is reasonable to conclude that moving quickly to a successful market design is a necessary condition for mitigating regulatory risk.

Conclusion

The missing money problem reflects a view that market design imperfections suppress electricity prices in spot markets. This produces inadequate incentives to invest in infrastructure resources such as generation capacity and its substitutes. A common policy response is to mandate purchases of installed capacity as a resource adequacy requirement. When this proves inadequate, more prescriptive reforms arise to compensate for the missing incentives in the market. An alternative approach would be to address the imperfections in the market design, provide the missing incentives, and eliminate the missing money. The resulting "energy only" market would not remove the need for regulatory interventions, but it would substantially change the character of those interventions. A sketch of such a market design illustrates how to address the market imperfections without overturning the market.

---

[30] Federal Energy Regulatory Commission, "Rules Concerning Certification of the Electric Reliability Organization; and Procedures for the Establishment, Approval, and Enforcement of Electric Reliability Standards," 18 CFR Part 38, Docket No. RM05-30-000, September 1, 2005.

[31] U.S.-Canada Power System Outage Task Force, "Final Report on the August 14, 2003 Blackout in the United States and Canada: Causes and Recommendations," April 2004, p. 140.

References

Miles Bidwell, "Reliability Options," Electricity Journal, Vol. 18, Issue 5, June 2005, pp. 11-25.

James Bushnell, "Electric Resource Adequacy: Matching Policies and Goals," University of California Energy Institute, CSEM WP 146, August 2005.

John Chandley, "ICAP Reform Proposals in New England and PJM," LECG, Report to the California ISO, September 2005.

Hung-Po Chao and Robert Wilson, "Resource Adequacy and Market Power Mitigation via Option Contracts," Electric Power Research Institute, Draft, March 18, 2004.

California Public Utilities Commission (CPUC) workshop on resource adequacy: www.cpuc.ca.gov/static/industry/electric/installedcapacity/041004_instcapacity.htm

California Public Utilities Commission, "Capacity Markets White Paper," San Francisco, August 25, 2005.

Peter Cramton, Hung-po Chao, and Robert Wilson, "Review of the Proposed Reserve Markets in New England," University of Maryland, January 18, 2005.

Federal Energy Regulatory Commission, State of the Market Report, Washington DC, June 2005.

Federal Energy Regulatory Commission, "Rules Concerning Certification of the Electric Reliability Organization; and Procedures for the Establishment, Approval, and Enforcement of Electric Reliability Standards," 18 CFR Part 38, Docket No. RM05-30-000, September 1, 2005.

Richard Greffe, Public Utility Commission of Texas, Wholesale Market Oversight, Memorandum, July 8, 2005.

Phillip G. Harris, "Relationship between Competitive Power Markets and Grid Reliability: The PJM RTO Experience," Issue Papers on Reliability and Competition, US Department of Energy and Natural Resources Canada, August 2005. (www.energetics.com/meetings/reliability/papers.html)

Scott M. Harvey, "ICAP Systems in the Northeast: Trends and Lessons," LECG, Report to the California ISO, September 2005.

William W. Hogan and Brendan J. Ring, "On Minimum-Uplift Pricing For Electricity Markets," Center for Business and Government, Harvard University, March 19, 2003, (www.whogan.com) .

Edward Krapels, Paul Fleming, and Stephen Conant, "The Design and Effectiveness of Electricity Capacity Market Rules in the Northeast and California," Electricity Journal, October 2004, Vol. 17, Issue 8, pp. 27-32.

Midwest Independent System Operator, "Discussion Paper on Resource Adequacy for the Midwest ISO Energy Markets," Draft, August 3, 2005.

Eugene Meehan, Chantale LaCasse, Phillip Kalmus, and Bernard Neenan, "Central Resource Adequacy Markets for PJM, NY-ISO, and NE-ISO," NERA Final Report, New York, February 2003.

Shmuel Oren, "Capacity Payments and Supply Adequacy in Competitive Markets," VII Symposium of Specialists in Electric Operational Systems Planning," Basil, May 2000.

Shmuel Oren, "Ensuring Generation Adequacy in Competitive Electricity Markets," University of California at Berkeley, April 2004.

Shmuel Oren, "Capacity Mechanisms for Generation Adequacy Insurance," CPUC-CEOB-CAISO Installed Capacity Conference, San Fransciso, California, October 4-5, 2004.

Public Utilities Commission of Texas (PUCT) workshop on resource adequacy: www.puc.state.tx.us/rules/rulemake/24255/24255.cfm

Bruce W. Radford, "Holes in the Market," Public Utilities Fortnightly, March 2005, pp. 19-21, 46-47.

Eric S. Schubert, "An Energy-Only Resource Adequacy Mechanism," Public Utility Commission of Texas, Staff White Paper, April 14, 2005.

Roy J. Shanker, "Comments on Standard Market Design: Resource Adequacy Requirement," Federal Energy Regulatory Commission, Docket RM01-12-000, January 10, 2003.

Roy J. Shanker, "Comments," Federal Energy Regulatory Commission Technical Conference on Capacity Markets in the PJM Region, June 16, 2005.

Harry Singh, "Call Options for Energy: A Market Based Alternative to ICAP and Energy Price Caps," PG&E National Energy Group, October 16, 2000.

Steven Stoft, Power System Economics, IEEE Press, 2002.

Steven Stoft, Testimony on Behalf of ISO New England on Locational Installed Capacity Market Proposal, Submitted to Federal Energy Regulatory Commission, Docket Number ER03-563-030; Direct August 31, 2004; Supplemental November 4, 2004; Rebuttal February 10, 2005.

U.S.-Canada Power System Outage Task Force, "Final Report on the August 14, 2003 Blackout in the United States and Canada: Causes and Recommendations," April 2004.

Carlos Vázquez, Carlos Battle, Michel Rivier, and Ignacio J. Pérez-Arriaga, "Security of Supply in the Dutch Electricity Market: the Role of Reliability Options," Report IIT-03-084IC, Comillas, Universidad Pontifica, Madrid, Draft Version 3.0, Madrid, December 15, 2003.

Carlos Vázquez, Michel Rivier, and Ignacio J. Pérez-Arriaga, "A Market Approach to Long-Term Security of Supply," IEEE Transactions on Power Systems, Vol. 17, No. 2, May 2002.

Endnotes

i William W. Hogan is the Lucius N. Littauer Professor of Public Policy and Administration, John F. Kennedy School of Government, Harvard University and a Director of LECG, LLC. This paper was prepared for the California Independent System Operator. This paper draws on work for the Harvard Electricity Policy Group and 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, Australian Gas Light Company, Avista Energy, Brazil Power Exchange Administrator (ASMAE), British National Grid Company, California Independent Energy Producers Association, California Independent System Operator, Calpine Corporation, Central Maine Power Company, Comision Reguladora De Energia (CRE, Mexico), Commonwealth Edison Company, Conectiv, Constellation Power Source, Coral Power, Detroit Edison Company, Duquesne Light Company, Dynegy, Edison Electric Institute, Edison Mission Energy, Electricity Corporation of New Zealand, Electric Power Supply Association, El Paso Electric, GPU Inc. (and the Supporting Companies of PJM), GPU PowerNet Pty Ltd., GWF Energy, Independent Energy Producers Assn, ISO New England, Luz del Sur, Maine Public Advocate, Maine Public Utilities Commission, Midwest ISO, Mirant Corporation, Morgan Stanley Capital Group, 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 IMO, Pepco, Pinpoint Power, PJM Office of Interconnection, PP&L, Public Service Electric & Gas Company, Reliant Energy, Rhode Island Public Utilities Commission, San Diego Gas & Electric Corporation, Sempra Energy, SPP, Texas Utilities Co, TransÉnergie, Transpower of New Zealand, Westbrook Power, Western Power Trading Forum, Williams Energy Group, and Wisconsin Electric Power Company. The author has benefited from comments by Jim Bushnell, Keith Casey, John Chandley, Steve Greenleaf, Scott Harvey, Lorenzo Kristov, Yakout Mansour, Mark Rothleder, Roy Shanker, and Anjali Sheffrin, among others. The views presented here are not necessarily attributable to any of those mentioned, and any remaining errors are solely the responsibility of the author. (Related papers can be found on the web the web at www.whogan.com).

AN ENERGY-ONLY MARKET FOR RESOURCE ADEQUACY IN THE MIDWEST ISO REGION

November 23, 2005

An Energy-Only Market for Resource Adequacy
In the Midwest ISO Region
Table of Contents

AN ENERGY-ONLY MARKET FOR RESOURCE ADEQUACY IN THE MIDWEST ISO REGION

November 23, 2005

INTRODUCTION

This paper continues the discussion of resource adequacy issues initiated in the Midwest through working groups organized by the Midwest Independent System Operator ("ISO") and the Organization of MISO States (OMS).[1] It follows a White Paper prepared by the Midwest ISO and issued on August 3 and September 9 (revised) of this year and comments submitted by stakeholders.[2] In its earlier White Paper, the Midwest ISO expressed a desire to examine the feasibility and merits of using what is commonly called an "energy-only market" approach to ensure resource adequacy. Comments by stakeholders, including important questions asked by the OMS,[3] seek additional information about how an energy-only market would function. In this paper, we attempt to address those questions that can be answered at this time, while providing a framework for thinking about other questions that should be answered as this discussion moves forward.

The principal reason for considering an energy-only market approach to achieving resource adequacy is the expectation that it would allow market incentives, rather than centralized administrative direction, to drive investment decisions. The rationale is consistent with the principal reasons for developing markets and departing from the traditional regulatory structure in the first place. In the words of William Hogan:

> A main feature of the [energy-only] market would be prices determined without either administrative price caps or other interventions that would depress prices below high opportunity costs and leave money missing. The real-time prices of electric energy, and participant actions, including contracting and other hedging strategies in anticipation of these prices, would be the primary drivers of decisions in the market. The principal investment decisions would be made by market participants, and this decentralized process would improve innovation and efficiency. A goal would be to avoid repeating the problem of leaving customers with stranded costs arising from decisions in which the customers had no choice. This change in the investment decision process and the associated reallocation of risk would arguably be the most important benefit that could justify greater reliance on markets and the costs of electricity restructuring. If this were not true, and if it would be easy for planners and regulators to lay out the trajectory of

---

[1] This paper was prepared by and at the direction of Midwest ISO staff, with significant assistance from John Chandley and Robert Borlick. Helpful comments were provided by William Hogan and Mike Cadwalader.

[2] Midwest ISO, "Discussion Paper on Resource Adequacy for the Midwest ISO Energy Markets," posted August 3, 2005; revised, September 9, 2005. http://www.midwestmarket.org/publish/Document/25228f_10631e11216_-7fac0a48324a. Hereafter referred to as "White Paper."

[3] http://www.midwestmarket.org/publish/Document/2b8a32_103ef711180_-77cd0a48324a.

needed investment for the best portfolio of generation, transmission and demand alternatives, then electricity restructuring would not be needed.[4]

From this perspective, the goals of an energy-only market are to improve innovation and efficiency, avoid the problems of stranded costs, and shift the risks and rewards of prudent investment from consumers to investors.[5] This would be achieved by moving the primary responsibility for investment decision-making from regulated planning and centralized administrative processes to the decentralized, voluntary decisions made by market participants responding to prices set by the market. But these are not the only reasons for considering an energy-only market.

One need not be convinced of the innate superiority of competitive markets to recognize other, very practical considerations that favor an energy-only approach to delivering an essential service to consumers without compromising reliability. The structure of an energy-only market directly defines the incentives that induce resources and loads to take actions consistent with reliable operations in the short run. Hence a further justification for giving serious consideration to an energy-only market approach is that it offers a consistent set of incentives that directly support both real-time reliability and resource adequacy. In contrast, current installed capacity (ICAP) structures and reform efforts make clear that achieving equally consistent and effective incentives in a capacity construct is extremely difficult and getting agreement on these administrative mechanisms is even harder, generally resulting in less than adequate incentives in both the short-run and long-run. These are not trivial concerns; they go to the heart of a very difficult problem we are trying to solve.

This paper expands the energy-only discussion by providing a more detailed description of how an energy-only market would work and what the ISO and others would need to do to implement such a market. The paper describes and discusses each element of that market and how it serves in promoting resource adequacy as well as ensuring operational reliability.

Defining An Energy-Only Market

As broadly defined in the earlier White Paper, an energy-only market explicitly pays resources only for the energy and ancillary services they deliver. It does not pay for installed capacity (ICAP). There is no requirement for utilities or load-serving entities to acquire or contract for "capacity" *per se*, or would anyone need to administer markets for

---

[4] William W. Hogan, *On an "Energy-Only" Electricity Market Design for Resource Adequacy*, paper prepared for the California ISO, September 23, 2005. See http://ksghome.harvard.edu/~whogan/.

[5] The term, energy-only market, is a misnomer and actually refers to a series of closely linked sub-markets for spot energy, operating reserve, other related ancillary services and bilateral contracts. The common bond is that all depend on cost-reflective, transparent spot energy prices in order to function efficiently and effectively.

such "capacity" because they wouldn't exist.[6] This does not mean that energy-only markets do not compensate generators for their "capacity" costs (i.e., their fixed operating costs, start-up and no-load costs, plus the recovery of, and return on invested capital). However, generators recover these costs through the enhanced profit margins (scarcity rents) they earn from selling energy and ancillary services, rather than through direct payments earmarked to recover those costs.

That said, it is worth emphasizing that an energy-only market deals with more than just energy; rather it is a convenient label for a set of market rules that govern the ISO's day-ahead scheduling and real-time dispatch for energy *and* operating reserve.[7] The mandatory capacity rules of ICAP markets presume that these spot prices will not reflect the true system conditions, thus turn to alternative regulatory requirements to provide adequate investment incentives. By contrast, the energy-only market approach presumes that spot prices can be made to reflect operating conditions and provide the right incentives. The expected stream of hourly spot prices for energy and operating reserve provide a foundation for contracts and investment decisions occurring in a series of interdependent markets (*e.g.,* bilateral contracts, derivative markets, *etc.*) complementary to the ISO run spot markets, that will yield the desired level of resource adequacy.

It may seem strange that the real-time spot market should be the key to talking about long-run resource adequacy. Typically, in discussions of resource adequacy or "capacity" constructs these topics are separated from discussion of the real time spot markets. But they are *not* separate and unrelated and it is a serious mistake to separate them. A coherent discussion of resource adequacy must consider how the ISO conducts real-time dispatch and prices energy.

When these topics become separated, the result will likely fall well short of solving the resource adequacy problem and may also undermine short-run system reliability. Indeed, virtually every problem that the Eastern power markets are having with their current ICAP mechanisms can ultimately be traced to the mistake of ignoring or deemphasizing this critical linkage.[8] The hourly spot prices that signal the need for investments to keep existing plants operational and to build new capacity also induce generators to make their existing units available to the dispatch to ensure safe and reliable operations in real time.

---

[6] The ISO leaves open the question of whether it should administer a voluntary forward market for energy, where market participants that are long or short could adjust their positions.

[7] Of the various ancillary services needed to support power system operation, operating reserve is the service most intimately linked with energy production. To maximize economic efficiency the markets for these two products must be closely coordinated.

[8] Scott M. Harvey, "ICAP Systems in the Northeast: Trends and Lessons," September 19, 2005, available at www.caiso.com. The New England LICAP filing (FERC Docket No. ER03-563-030) and the PJM RPM filing (FERC Docket Number EL05-148 *et al.*) both include substantial and remarkably candid critiques of the problems each ISO has experienced with current ICAP approaches, thus necessitating their reform efforts.

Getting the spot price signals right, and making sure that those who make operational and investment decisions are exposed to those prices, is essential. Even if the ISO ultimately adopts a capacity-based construct, that mechanism is unlikely to work very well if the designers lose sight of this basic truth.

The central question raised by the ISO in this paper, and in its original White Paper, is whether an energy-only market would be better than one that includes an explicit capacity mechanism. Would it be easier to implement? Could it do a better job of stimulating investment and ensuring resource adequacy? Would it be easier to design? This paper discusses why the central question is so important and explains why consideration of an energy-only market framework is both necessary and worthwhile no matter which path the Midwest region pursues.

Energy-Only Market And The Need For Intermediate Or Backstop Mechanisms

The energy only market has been characterized by some as an "end state" – a position that the market should eventually evolve into, but one that is not immediately feasible. This raises two issues that must be addressed. First, if an energy only market is the desired end state, then the short and long term impediments need to be identified and either overcome or mitigated. Second, it must be recognized that there is a fundamental difference between a "backstop" mechanism, *i.e.* one that is hopefully never used and is triggered only when a set of pre-defined criteria are met, and an "interim" mechanism which is a step along the way to an energy-only market. In either case it will be necessary to design structures that are incentive compatible with the development of an energy only market. That is, both the "backstop" or "interim" mechanisms cannot provide disincentives for either reaching the end state or ensuring the potential failure of the end state once implemented.

This paper explores the critical issues that need to be addressed to determine the need for an interim capacity mechanism. The paper also summarizes and provides references to recent papers that describe alternative capacity mechanisms.

The Role Of State Regulators

A purely market-based energy-only wholesale market can work without changes in the way electricity is consumed and paid for by end-users but the task becomes much more difficult to achieve. In particular, markets – for any service or commodity – function best when the costs and benefits of specific actions are either implicitly or explicitly transparent. When actions are linked to costs and benefits then parties have the incentive to manage risk in ways that allow for socially optimal outcomes. Under current rules ensuring that adequate capacity is installed is the responsibility of the States. Therefore, to assure both short-run reliability and long-term resource adequacy, wholesale market rules and retail regulation need to work in mutually supportive ways. As a result, state regulators play a critical role in a well functioning energy-only market.

In today's MISO energy markets, market participants are credited or charged based on the locational marginal price (LMP) at their injection or offtake node. Market participants that serve load (LSEs) with sufficient contract cover can insulate themselves from the potential for real-time LMP price volatility. LSEs without sufficient cover through contracts with generation may be able to make up the shortfall through contracts with their customers having price responsive demand. Without either, these LSEs can be exposed to significant real time price volatility. As discussed later, state regulators will have access to LMP data for LSEs under their jurisdiction and will have knowledge of their contract cover, including demand-side contracts. The state regulators can use this information to guide LSE procurement strategies.

Over time retail rate designs can be changed and real-time, interval metering and billing can be implemented. However, decisions need to be made regarding how fast this should be done and for which customers. Because these decisions reside with the state utility commissions an energy-only market cannot be designed and implemented exclusively by the ISO. Important elements, particularly with respect to retail rate design, demand-side response and forward contracting would benefit from significant attention from state regulators. [9]

Regardless of the approach taken to resource adequacy, the energy charge component of any retail rate design should reflect, as closely as practicable, the real-time spot price where the load is located. Any retail rate designs that reflect this locational energy charge, coupled with appropriate metering and billing systems, would charge customers on the basis of cost causality and would facilitate economically efficient demand-side response, which in turn would help define and mitigate prices, discourage market power, and significantly reduce the need for involuntary curtailments (rolling or tailored blackouts). We expand on these concepts in later chapters.

An important issue is whether contracting should be promoted or mandated by state regulators to complement an EOM. In addition to self-supply by utility-owned generation, power contracts of varying terms could play a dominant role in the prices retail customers would face. Such contracts could protect most consumers from the spot price volatility that is essential for an energy-only market to assure operational reliability and long-run resource adequacy. This contracting would not diminish the importance of spot prices in providing the right incentives but it would redistribute the risks of volatile spot prices.

In states where utilities are regulated and have an obligation to serve their franchised customers, such price hedges are a natural consequence of utility plant ownership and power supply contracting. In states that allow retail choice, or where utilities have divested their generation, state regulators would want to consider how best to structure "default service."

The degree of support the ISO receives from state regulators, as well as the degree of cooperation among them (such as by acting collectively through the OMS) would largely determine how successful an energy-only market would be. It is unlikely that the

[9] These structural flaws also interfere with the efficient functioning of capacity-based resource adequacy mechanisms so correcting them is important regardless of what the ISO decides to do.

ISO can establish and implement a fully functional energy-only market design without substantial state cooperation, although the market would still function better if the ISO implemented the key elements of that market in any event. Of course, the same need for support and cooperation would apply for any "capacity-based" mechanism that the ISO might pursue.

Compatibility With PJM's Reliability Pricing Mechanism

The Midwest ISO market shares interfaces with several other markets, including PJM. Six states in the Midwest have utilities and transmission systems that overlap both the Midwest ISO and PJM footprints. In addition, all of the utility systems in either RTO would be affected through their interconnection with the other RTO. Thus, all stakeholders within the two footprints have a vested interest in "seamless" trading between the two RTOs, and most support the goal of achieving a "Joint and Common Market."

Several questions arise from the Midwest ISO-PJM interconnection. First, is it necessary for the two RTOs to employ similar resource adequacy mechanisms to assure seamless trading? Second, if their market designs proceed along different paths how can they be made compatible? Third, would inter-RTO trading suffer if they were different, and if so, how could those problems be mitigated?

PJM has recently proposed to address resource adequacy issues in its footprint through what it calls a "Reliability Pricing Mechanism" (RPM), which is now before the Federal Energy Regulatory Commission. RPM is a very different mechanism for achieving resource adequacy than an EOM; it is a capacity-based construct employing an innovative "capacity demand curve." This novel feature has diverted attention away from the important linkage between real-time spot prices and resource adequacy.

Given the tentative status of the RPM filing, this paper cannot fully answer all of the questions raised earlier; however, it does provide an analytical foundation for assessing the issue.

I. THE ENERGY-ONLY MARKET

The defining characteristic of an energy-only market is that prices are set by supply and demand while minimizing the use of administrative caps or other means that artificially suppress those prices below competitive levels. Although some administrative intervention may be needed to deal with missing markets, such as those for some ancillary services or other possible forms of "market failure," one would choose a solution that would least distort spot energy prices.

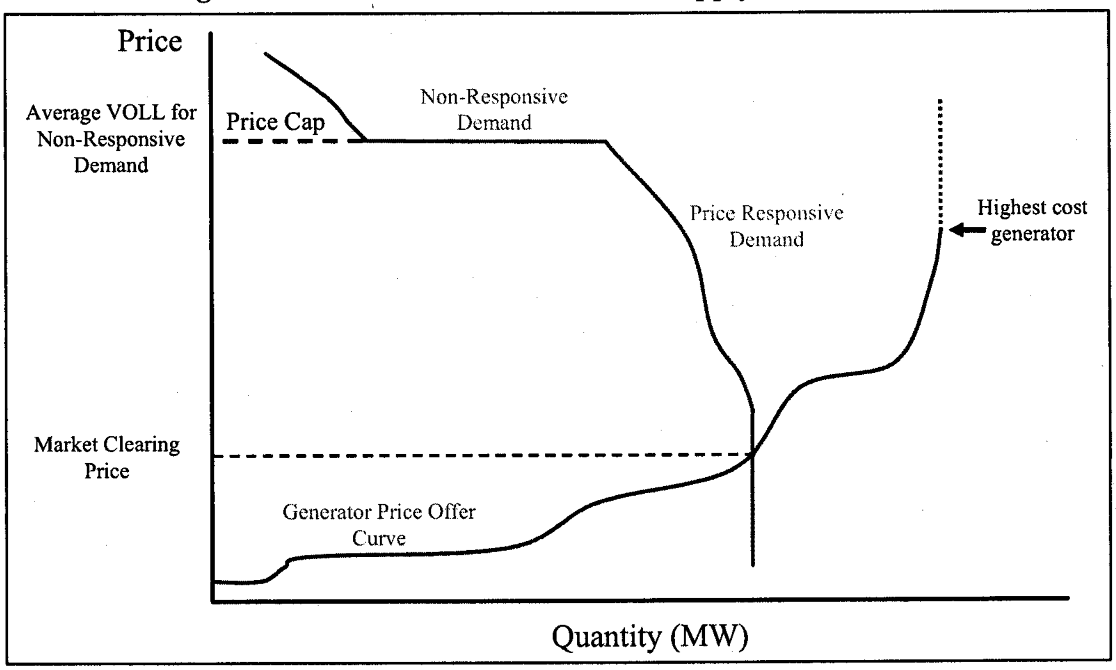
How Energy Prices Are Determined

The common notion that an energy-only market equates to *unbridled pricing and "let-the-market-rip" is **not** what is meant here by an energy-only market.* Nor does it mean that prices can reach infinite or indeterminate levels during supply shortage conditions. In an energy-only market, there are constraints on prices.

One constraint is competition among generators, i.e., market power is minimal or mitigated (this issue is addressed in a later chapter). A second constraint is the ability of consumers with flexible loads to reduce their consumption in response to high spot market prices. Finally there is an overall cap on spot energy prices reflecting the willingness to pay of those consumers who cannot respond in real time to the spot price. These three price constraints are illustrated in the figures that follow.

Figure 1 shows how the spot energy price is determined when available generating capacity is sufficient to satisfy demand.

Figure 1. Price Determination When Supply is Sufficient [10]



The spot energy price is set by the intersection of the generation price offer (supply) curve with the demand curve. Assuming there is sufficient competition among the generators the supply curve will closely track the generators' short-run marginal costs of production, which is almost all fuel expense. Thus, this first constraint is embedded in the generation price offer curve.

The second constraint on energy prices is the price-responsiveness of customer loads, represented by the downward sloping portions of the demand curve. The rightmost

[10] The portion of the demand curve labeled "Non-Responsive Demand" indicates the amount of load that end-use customers have that is either not sensitive to price, or represents load where customers may not see the wholesale price. Usually economists indicate demand not sensitive to price as vertical segments; here it is drawn horizontally to indicate prices are capped at the average VOLL.

segment of this curve represents loads that value electricity at less than the "average VOLL." The leftmost downward sloping segment of the demand curve represents price-responsive loads that value electricity by more than the "average VOLL" price.

The horizontal portion of the demand curve represents the loads of customers that are not price-responsive, primarily because they lack real time meters thus are not billed on an hourly basis. Although their loads value electricity across a broad spectrum of electricity prices, they are all represented here at the average valuation for the entire load group because the customers cannot directly reveal the valuations of their individual loads by turning them off or on in response to spot energy prices. Thus, "average VOLL" price is defined to be average value of lost load (VOLL) for this group of spot price-insensitive customers when electric service to them is randomly interrupted.

This brings us to the third constraint on spot energy prices. As illustrated in Figure 1, the market design imposes an overall cap on real time spot energy prices equal to the average VOLL. This absolute price ceiling prevents the market from inefficiently purchasing electric energy at prices that would exceed the prices the customers incrementally served by that energy would be willing to pay.[11]

It is likely that the vast majority of customers, including virtually all residential and most small C & I customers, would be part of the horizontal segment of the demand curve. A defining characteristic of these non-responsive consumers is that they lack real time metering, which reflects the realities of how these customers are metered and billed today. Furthermore, in most cases it would not be cost-beneficial to install these meters because the savings they could realize through voluntary load curtailments would not recover the cost of the requisite real time metering and billing.

For this large middle group of customers, it can be assumed that changes in spot energy prices would have essentially no effect on their short-run usage – hence their portion of the demand curve is flat. They would continue to make their energy consumption decisions based on their tariff rates, without regard to the level of spot prices, even if those spot prices occasionally rose to very high levels.[12]

[11] The incrementally served customers are those whose loads would not be curtailed if additional electric energy were available at prices above average VOLL and were purchased to serve them. Such high cost energy would likely come from generators that have marginal costs well below average VOLL price but that seldom run, thus can only recover their fixed costs by charging very high prices during the few hours they do run. These generators need not be technically inefficient (i.e. high heat rates, low availability, etc); indeed, they could be state of the art plant. However, if there are too many of them, relative to customer demand, each will only run a small number of hours per year, which means they are economically inefficient and everyone would be better off if they had not been built. One function of the price cap it to ensure that consumers do not subsidize the construction of excess capacity.

[12] However, if spot energy prices remained high for extended periods, as they did in the 2000-01 California electricity crisis, consumers billed under fixed tariffs would still see high monthly bills which would in the longer-run would induce them to inefficiently reduce their usage in both high priced and low priced hours.

As stated earlier, we don't know how much the price-insensitive customers would be willing to pay to avoid involuntary curtailments. Nonetheless we must assume some value for their average VOLL in order to set the price cap. In resource adequacy discussions in the US, the estimated averages range from about $2,000 to $10,000 per MWh or more, depending on which categories of customers are assumed to be curtailed.

The prices paid by price-responsive customers would never exceed their willingness to pay, as determined by their own choices to consume or not. In principle, the prices paid by non-responsive demand would also never exceed their willingness to pay if the market price cap accurately reflects their average VOLL.

Customers with price-responsive loads would be served under retail rate designs employing energy charges that closely track the real-time hourly wholesale spot energy prices.[13] In addition, their usage would be metered on at least an hourly interval basis and their responses to the hourly spot prices would be properly reflected in their bills. The price-responsive loads need not be dispatchable by the ISO or by the customers' load serving entities. All that is essential is that these customers be able to respond to the spot prices by voluntarily curtailing their consumption or shifting it to lower-priced hours.

There is no strict, minimum requirement for the amount of price-responsive load needed to achieve optimal resource adequacy. This is an issue that must be empirically assessed, as discussed in a later chapter. Obviously, the more price-responsive load there is the more economically efficient the market will be and the lower the total costs.[14] However, a subset of the largest industrial customers facing real-time energy prices is likely to be all that is required to control market power and reduce costs for all customers. During conditions of scarcity, even a modest amount of price-responsive demand could dramatically limit the level of electricity prices for all customers while also forestalling rotating blackouts.

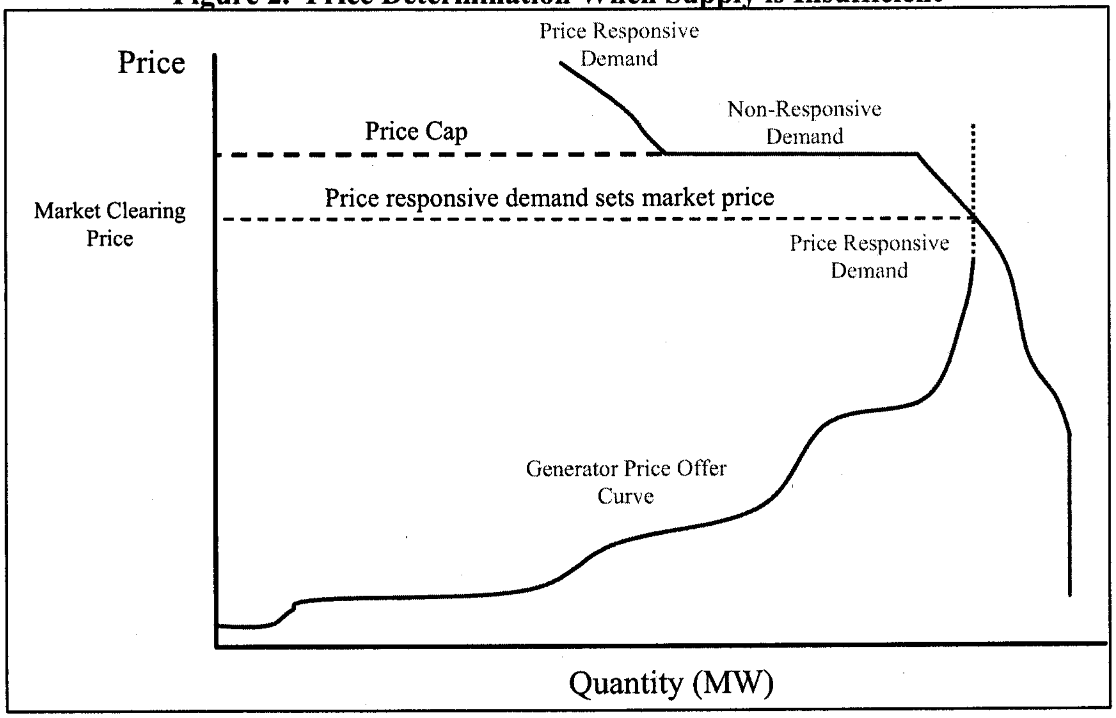
The earlier figure illustrated how the energy price is determined when available

[13] In fact it is quite likely that most price responsive customers have loads in both value categories.

[14] If more demand is price responsive, less generation capacity will be needed to meet the desired loss-of-load expectation adopted for the non-price responsive loads. Conversely, if few consumers are price responsive, the total costs of reliably meeting all loads will be higher. In deciding how much price-responsive demand to encourage, state regulators, regardless of their views on "restructuring," should keep this fundamental tradeoff in mind.

generating capacity is abundant. Figure 2 below illustrates what happens when generating capacity is in short supply and price responsive demand fills the shortfall.

Figure 2. Price Determination When Supply is Insufficient



In this situation the spot energy price is determined by the demand curve intersecting the generation supply curve at maximum available capacity, which can occur at a market price substantially higher than the short-run marginal cost of any generator. This pricing rule is referred to as "scarcity pricing" or "shortage cost pricing" and it is a critically important feature of the energy-only market design.

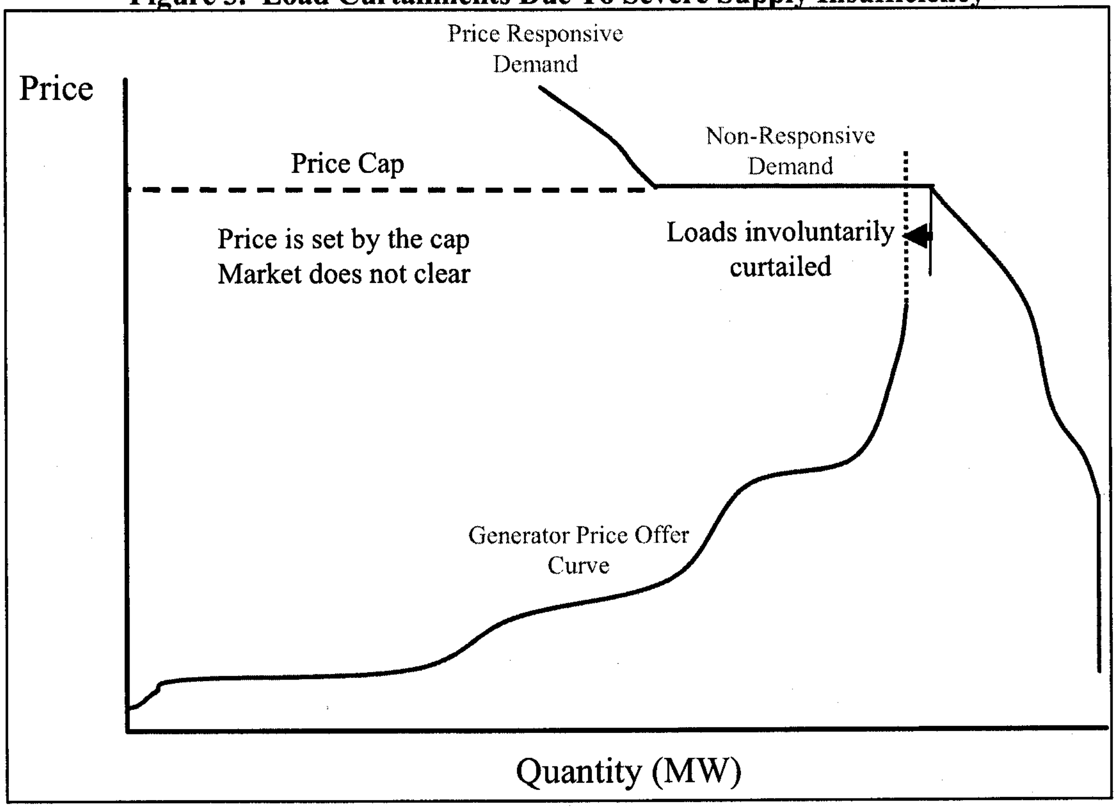
Figure 3 illustrates what happens when supply insufficiency is so severe that involuntary curtailments result. As shown there, the spot energy price rises to the cap and some non-price responsive customers must be curtailed, perhaps through rolling blackouts or based on which customers had inadequate contract cover prior to the operating day. From a societal perspective the first customers to be interrupted should be those that place the lowest monetary value on electricity. Value of service studies consistently suggest that residential customers value electricity far less than commercial and industrial customers do.[15] While it may appear inequitable for residential customers to always be interrupted first, we must remember that the customers remaining on the system pay very high prices for their consumption during hours of stress and substantially contribute to the fixed costs of the generating capacity serving them. Their payments

---

[15] Lawton, Sullivan, VanLiere, Katz and Eto, A Framework and Review of Customer Outages: Integration and Analysis of Electric Utility Outage Cost Surveys, Report LBLN-54365, Lawrence Berkeley National Laboratory, November 2003.

make it possible for that generating capacity to be economically viable and therefore available to serve the curtailed customers at other times.

Figure 3. Load Curtailments Due To Severe Supply Insufficiency



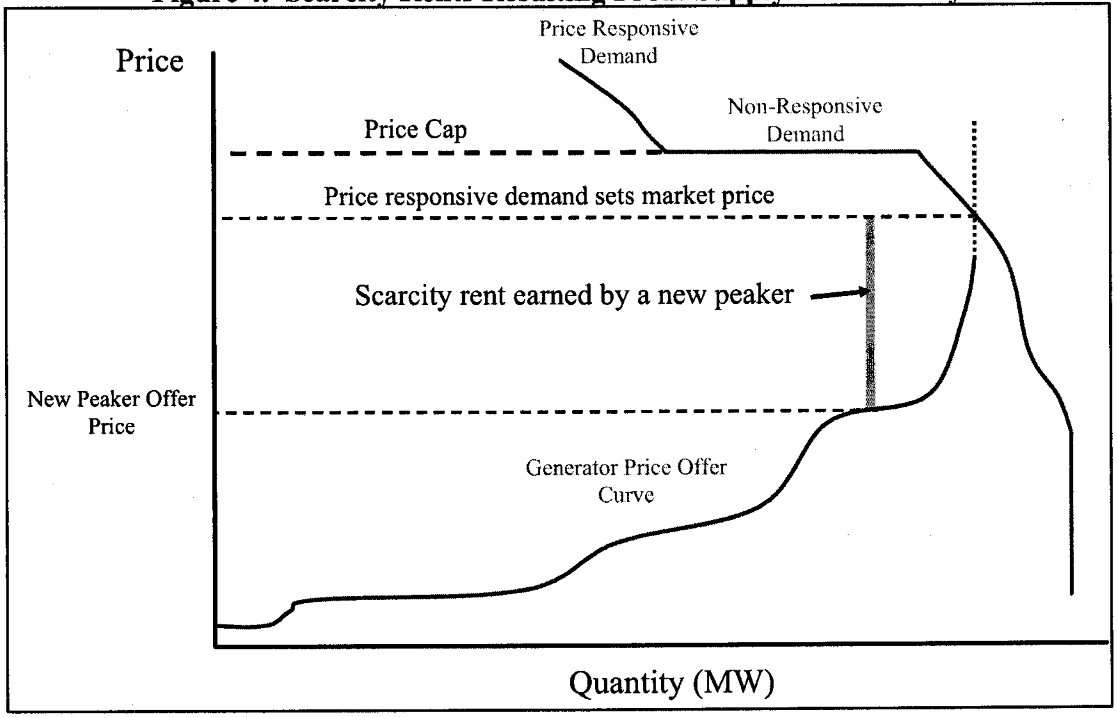
The Role of Scarcity Rents

Let us return to the question of how scarcity pricing in an energy-only market can assure resource adequacy without the need for capacity payments.

Figure 4 illustrates how scarcity pricing produces "scarcity rents" for all generators operating in hours when such pricing occurs and these rents contribute to the generators' fixed costs, capital recovery and profits.[16] For clarity of explanation the figure only highlights the rents for one plant – a new peaker – but all of the other plants producing energy also earn scarcity revenues. It is the expectation of earning such scarcity rents that produces the incentive for owners of existing generating plants to make those plants available for dispatch when the market needs them most, and also the incentive for developers to invest in new generating plants.

[16] As used here the term, "scarcity rent" means the difference between a generator's sales revenues and its variable operating costs (which exclude any startup or no-load costs) during a given time period.

Figure 4. Scarcity Rents Resulting From Supply Insufficiency



The latter incentive is depicted by the shaded area depicting the scarcity rent that a new peaking unit would earn. If developers expect a new peaking unit to earn sufficient scarcity rents to cover its fixed costs, recover its capital investment and earn a compensatory rate of return on that capital, revenue adequacy will be assured. Thus, in an energy-only market scarcity rents are the genesis of resource adequacy as well as operational reliability.

The specific process through which scarcity rents work to promote resource adequacy is as follows. Assume that we begin with the power system in a state of overcapacity (as is the case today throughout the MISO footprint). As demand grows, the generation reserve margin shrinks and price-responsive demand will be called on with increasing frequency to compensate for shortfalls in generating capacity. The more frequently price-responsive demand sets the energy price the greater will be the scarcity rents and the more profits generators will earn, *and will expect to earn in the future.* This process will continue until investors' expectations regarding future scarcity rents reaches a level that they accept as sufficient to justify constructing new peaking capacity. As the new plants enter service they will arrest further declines in the generation reserve margin and maintain it at that level.

For purposes of achieving generation resource adequacy all that is required is the economic viability of new peaking units. However, once such viability is assured, investors will also have incentives to build other types of plant, driving the power system toward the optimal mix of generation fuels and technologies. This will happen through the following process.

As load grows and new peaking units are added to meet peak demand, the most fuel-efficient units will be dispatched to run more and more hours each year. Eventually the point will be reached where a different generation technology, such as a combined-cycle unit, can produce energy at a lower average cost when run at these high capacity factors, producing higher profits for its owners. This same argument also extends all the way down to base load unit additions.

The level of resource adequacy associated with the equilibrium condition just described depends on a number of factors, primarily how high the energy price cap is set, how much price-responsive demand exists, how much are the capital and fixed operating costs of a new peaker, and investors' risk aversion.[17] This issue cannot be resolved through qualitative arguments alone; consequently, we intend to quantitatively address it in future documents.

DEMAND FOR AND PRICING OF OPERATING RESERVE

The supply-demand analyses presented in Figures 1 through 4 are useful (albeit simplified) abstractions. In real power systems generators must supply not only energy to customer loads but also reserve capacity (aptly termed "operating reserve") that can quickly deliver energy in response to a "contingency", i.e., a sudden, unexpected failure of a major generator, transmission line, transformer or other key facility. Thus, the available generation must satisfy a combined demand consisting of customer loads plus the required operating reserve.

From this it follows that the market price of energy is more closely approximated by the intersection of the generation supply curve and the combined demand curve.[18] Even so, the concepts presented earlier remain valid.

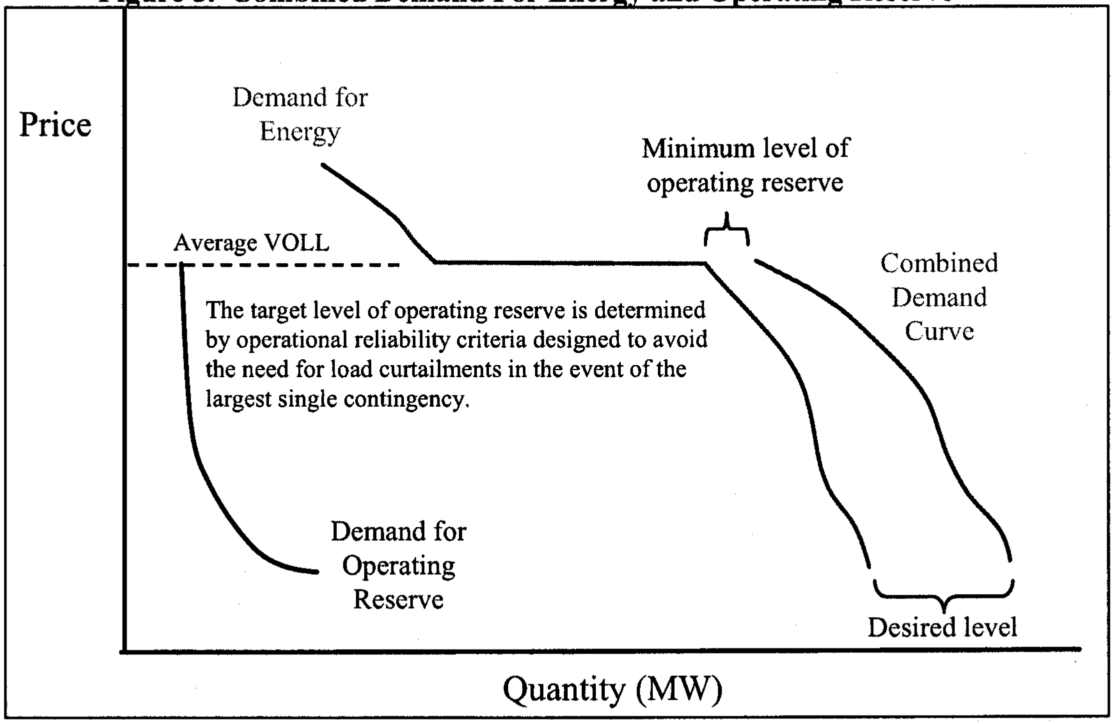
---

[17] Even with no price-responsive demand an energy-only market would deliver an equilibrium level of resource adequacy but it would likely be unacceptably low by current industry standards. Without price-responsive demand load curtailments would be more frequent, driving the spot energy price to the price cap whenever they occur. The more hours during which this occurs, the more scarcity rents generators earn. Eventually equilibrium would be reached when these scarcity rents are sufficient to attract new generation.

[18] This graphical method is still only approximate but it is useful. Only the effect of "in the money" operating reserve, i.e., that which could profitably produce and sell energy if it were not providing reserve, affects energy prices. Often there are units operating at minimum load in off-peak hours that can provide spinning reserve at no opportunity cost. Also quick-start units, such as pumped storage hydro, can provide non-spinning reserve at no cost. Neither of these effects are captured in simple supply-demand curve diagrams.

Figure 5 illustrates the combined demand curve. As shown there, total demand is displaced to the right of the energy demand curve by the amount of operating reserve required. It also shows that as the energy price increases the system operator will trade operating reserve for energy until some operating reserve minimum is reached, an important concept discussed below.

Figure 5. Combined Demand For Energy and Operating Reserve



Operating reserve comes in two forms:

- spinning reserve and
- non-spinning reserve.

Spinning reserve is that portion of an operating generator's capacity that is not delivering power to the system. This unloaded capacity can begin delivering replacement energy almost instantly in response to a contingency and the amount of lost capacity it can replace is determined by the thermo-mechanical ramp rate of the generator.[19] Spinning reserve is very important in arresting the decay of system frequency and returning it to its target 60 cycle value. A large decline in system frequency would trip

[19] In US power systems operating reserve is defined as the total output that a provider is capable of delivering within 10 minutes following a contingency. For small systems with few or no connections, such as Singapore or New Zealand, the deliverability criteria are measured in seconds, not minutes.

the circuit breakers of operating generators, further compounding the under-frequency problem and ultimately leading to uncontrolled, cascading blackouts.

Non-spinning reserve is provided by generating plants that are off-line but can quickly start up. Hydro units with storage are ideal for this type of service. The only requirement is that they must be able to deliver substantial amounts of energy within the prescribed 10 minute period.

An important feature of an energy-only market is the trade off between energy and operating reserve alluded to earlier when introducing Figure 5. As a general rule, with current technology the demand for operating reserves cannot be defined solely by the market; it is an administrative construct determined by the ISO or by some entity responsible for setting reliability standards, such as the Regional Reliability Council.[20]

Figure 6 shows how the demand for operating reserve is determined as a function of energy price. The curve is concave, reflecting the fact that the marginal value of operating reserve rapidly declines as more and more is available. The ISO should substitute operating reserve for energy as the marginal value of energy to the system increases. However, there is some minimum level of operating reserve which the system operator will not violate, illustrated here as 3 percent of the contemporaneous energy load.[21] Rather than encroach upon this minimum level the system operator will shed load in order to preserve this minimum requirement.

The amount of operating reserve consumed by a contingency is randomly distributed and can be small or large. Obviously the less operating reserve the system has the greater the probability that the system operator will need to shed load. Thus, this probability directly translates into a loss of load expectation (LOLE) for the system. Multiplying this LOLE by the average VOLL applicable to the customers likely to be curtailed, produces the concave expected cost function shown in Figure 6.

Although the system operator would like to have some target level of operating reserve, *e.g.*, 7 percent, it would not maintain that target at any price. As supply tightens, increasing the spot energy price, the price of operating reserve will rise in lockstep and the system operator will trade operating reserve for energy until the minimum operating reserve requirement is reached, at which point the spot energy price will equal average VOLL.

The economic tradeoff between energy and operating reserve produces a combined demand curve that is more steeply sloped than the energy demand curve, thereby producing greater scarcity rents for generators when loads are not curtailed than

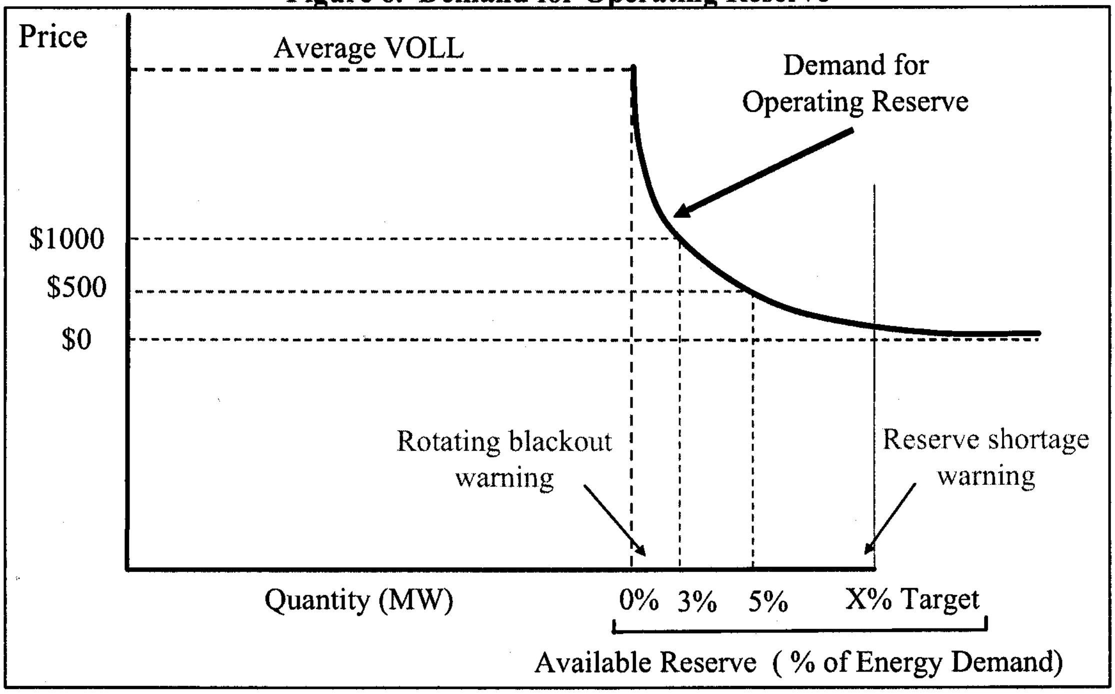
---

[20] In the Midwest ISO, the operating reserve standard varies by RRO; for example, in ECAR it is 3%.

[21] However, for a power system that has few internal transmission constraints and has substantial import capacity, operating reserve can actually be allowed to drop to zero without endangering operational reliability. If done frequently, such free rider behavior would be frowned upon.

would be the case if this operating reserve was held constant. This assists an energy-only market in achieving resource adequacy.

Figure 6. Demand for Operating Reserve



If there's an insignificant amount of price-responsive demand the demand for operating reserve will still produce a downward sloping combined demand curve (albeit a steep one) that produces additional scarcity rents for generators so long as energy prices are determined by the intersection of the supply curve with the combined demand curve. *Again, we emphasize that this approach to pricing spot energy will not occur unless explicit provision for doing so is included in the market rules.*

OPTIMIZING ENERGY AND OPERATING RESERVE PRODUCTION

The Midwest ISO's economic dispatch of generation currently does not jointly optimize ("co-optimize") the production of energy and operating reserve. Consequently, spot energy prices currently do not reflect the effect of operating reserve shortages if and when they occur. This would change with the adoption of the energy-only market. The economic dispatch process would direct generators to produce energy and/or operating reserve based on their respective energy offer prices, to provide customers with the needed energy or operating reserve at the lowest bid-based cost.[22]

Through such co-optimization, the spot prices paid for energy and operating reserve would be mutually consistent. A generator that is available to provide energy at a given offer price might be directed to partially load its unit in order to provide some operating reserve. The generator would then be paid the spot energy price for its energy and the spot operating reserve price for its operating reserve. The market price for operating reserve would be set by the generator with the largest opportunity cost (i.e., foregone profit from energy sales) thereby ensuring that all other generators would be adequately compensated.

When generating capacity is in ample supply, spot energy prices would be low to modest and prices for operating reserves would be even lower. However, when generating capacity is in short supply, spot energy prices would be very high and the spot prices for operating reserves, though generally still lower, would track energy prices upward. Ultimately, if loads must be curtailed both spot prices would equal the energy price cap.

II. OPERATIONAL RELIABILTY, RESOURCE ADEQUACY AND ENERGY MARKET FAILURE

THE MISO SPOT MARKETS

Since April 1, 2005, the Midwest ISO has operated two spot markets for energy – The Day-Ahead Market and the Real-Time Market. These two markets arise as a natural consequence of its real-time, security-constrained economic dispatch of the generating plants under its control.[23] The Day-Ahead Market sets *ex ante* energy prices for generators and loads bidding into that market. The Real-Time Market sets *ex post* energy prices for generators and loads that do not participate in the Day-Ahead Market or that depart from their commitments scheduled previously in that market. Along with the ISO-coordinated provisions for operating reserve and regulation service by market participants, the ISO's security-constrained economic dispatch ensures safe and reliable system operations every moment of every day across the region.[24]

---

[22] See the Appendix for a more detailed discussion of the co-optimization of energy and operating reserve markets, with examples of how prices would be determined for energy and operating reserves.

[23] This economic dispatch process adjusts the outputs of all generators under MISO control at five minute intervals in order to supply all customer loads while minimizing the total payments made to generators and also satisfying system security constraints. These constraints prescribe that certain generators and transmission facilities remain partially loaded in order to provide reserve energy in the event of generation or transmission contingencies.

[24] Regulation is the ancillary service that matches generation to load within each hour. It is provided by generators that continuously adjust output in response to small changes in system frequency.

Like the Eastern ISOs, the MISO spot markets employ locational marginal pricing (LMP) to signal the value of energy at each injection and offtake node on the transmission grid. LMPs in the Day-Ahead Market are forecasts of the next-day values produced by iterative computer simulations of the entire MISO system. These prices are inputs to the generation scheduling and unit commitment process. LMPs in the Real-Time Market are *ex post* values reflecting the actual real-time dispatch. Both sets of prices provide accurate signals to generators about the actions they need to take at each location to satisfy customer loads, maintain voltage and manage transmission congestion. In addition, these same prices also lay the foundation for resource adequacy.

The economic dispatch process and its associated spot markets are not just tools for ensuring short-run reliability; they also play an important role in the assurance of long-term resource adequacy. If these spot markets are properly designed the resulting spot prices will reflect the value of the last MW of load served. These price signals will, in turn, encourage generators to make their capacity available when needed and investors to finance the new generation needed in the future. Thus, these two spot markets play a pivotal role in assuring both short-run reliability and long-run resource adequacy.

SPOT PRICES EFFECT GENERATOR AND CUSTOMER BEHAVIOR

For existing plants, the hourly LMP prices for energy (and operating reserve when these markets exist) signal plant owners to take the actions needed to assure short-run operational reliability. The spot prices provide incentives for such actions as performing or postponing maintenance, securing adequate fuel supplies, deciding between spot purchases or firm fuel contracts, procuring fuel storage, maintaining operational crews during key periods and other activities needed to keep the existing plants operable and available for dispatch.

For new plants, the expected future prices signal to investors when and where new capacity is likely to be needed, what types of and sizes to build, what fuels to use, and what capabilities the plant should have (such as quick start, cycling capability, and ramp rates). Although many factors influence investment decisions, expectations of future spot prices are the key driver. Even if investment decisions are premised on long-term contracts, particularly when regulatory and/or market uncertainty is high, those contracts will reflect both parties' expectations regarding spot prices over time.

Spot prices are important regardless of whether the participants decide how to operate their plants or are responding to instructions from the ISO. When such choices are independent, it is critical that the spot prices signal the actions that best support reliability and efficiency. When plants follow instructions from the ISO, consistent spot pricing ensure that the instructions do not act against participants' self interests.

Energy spot prices are just as important on the demand side as they are on the supply side. In the short-run spot prices can signal both consumers and their load-serving entities (LSEs) about whether and when to use more or less energy. In the long run, these same prices influence consumers' decisions to invest in efficiency improvements and other demand-side measures that can change their future energy usage. These actions, in turn, affect not only peak demands but also load shapes.

IMPORTANCE OF SCARCITY PRICING

The concept of scarcity pricing and the scarcity rents it produces was introduced and illustrated in Chapter I. During times when generation is in short supply it is important that spot energy prices be allowed to rise to levels that reflect this scarcity condition. During such times spot prices should be determined by the willingness of consumers to pay for energy, rather than by the energy offer prices of generators. Such relatively high spot prices will yield scarcity rents for generators and cost savings for price-responsive customers while efficiently rationing the available energy to those end uses that value it most highly.

The extraordinary efforts taken by both suppliers and consumers in responding to these high prices will combine to obviate the need for involuntary load curtailments in all but the most extreme capacity-short situations. *The lights will stay on, and no customer will be involuntarily curtailed given their willingness to pay the high spot energy prices.*

Over time, the level of investment induced by scarcity pricing will tend to fluctuate around the level of resource adequacy for which consumers (or their LSEs) have indicated a willingness to pay. But the key feature that allows this mechanism to work is the ability of spot prices to occasionally rise to levels reflecting the actual shortage conditions. *Without such occasional shortage-cost prices the mechanism will fail to produce the desired result.* Some generators will not recover their fixed costs, and over time, investors will support a lower level of investment than that which satisfies the region's adequacy criterion.[25]

MARKET FAILURE FROM SUPPRESSING ENERGY PRICES

Despite the critical importance of scarcity pricing, spot prices are routinely suppressed in every US electricity market. This may occur through caps on generator offer prices or through more general caps on market energy prices. In the Eastern ISO markets and the Midwest ISO, the cap on generator offer prices is currently $1,000/MWh. In California the cap is even lower – currently $250/MWh. It is not the offer caps that are the problem, but rather how they interact with other market rules for determining market prices. Spot prices can exceed the offer caps through other means but the offer caps applied under shortage conditions tend to limit prices below market-

[25] The discussion in this section is only a summary of the economic argument. For more detailed explanations of the economic theory, a useful source is Steve Stoft, *Power System Economics, Designing Markets for Electricity,* Wiley–IEEE Press, 2002.

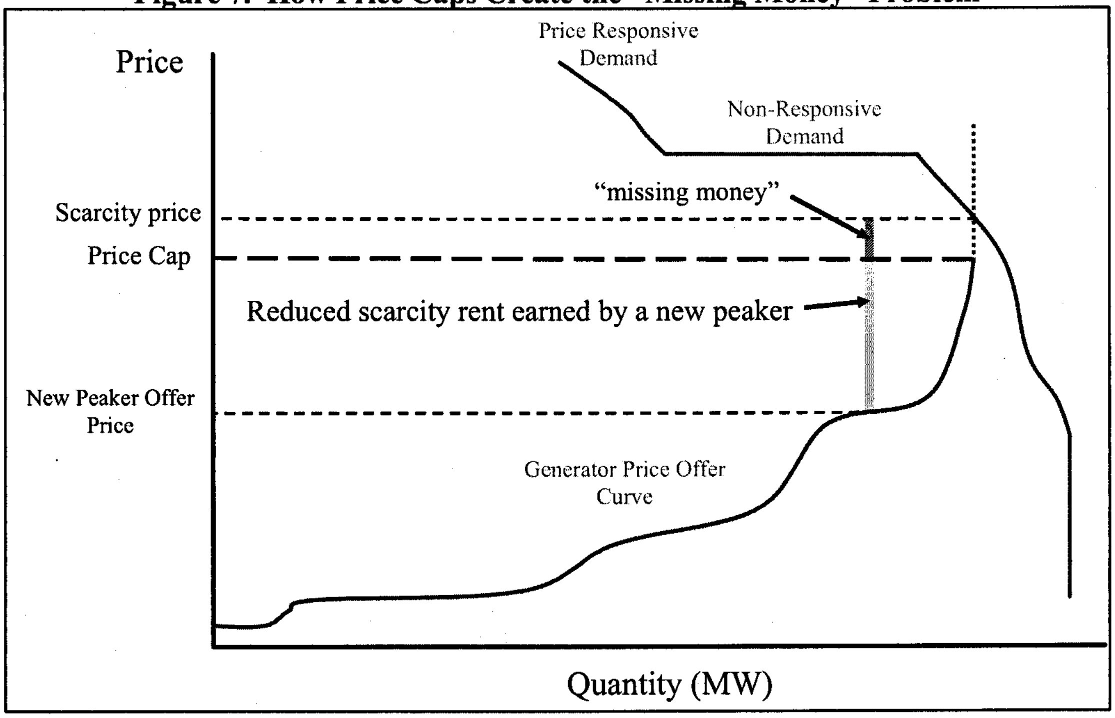
based levels.[26]

Suppressing spot energy prices through price caps and other market flaws, creates the "missing money" problem illustrated in Figure 7. This figure is the same as Figure 3 in Chapter 1 except that the price cap is set lower than the market clearing price.

Shown here by the black diagonal lines is the scarcity rent eliminated by the price cap that a new peaker would otherwise have earned at the market clearing price. This lost scarcity rent is termed the "missing money."

Without the missing money investors will not build or maintain enough capacity to meet the region's resource adequacy goals. Thus, the rationale underlying all capacity payment mechanisms is to replace the missing money.

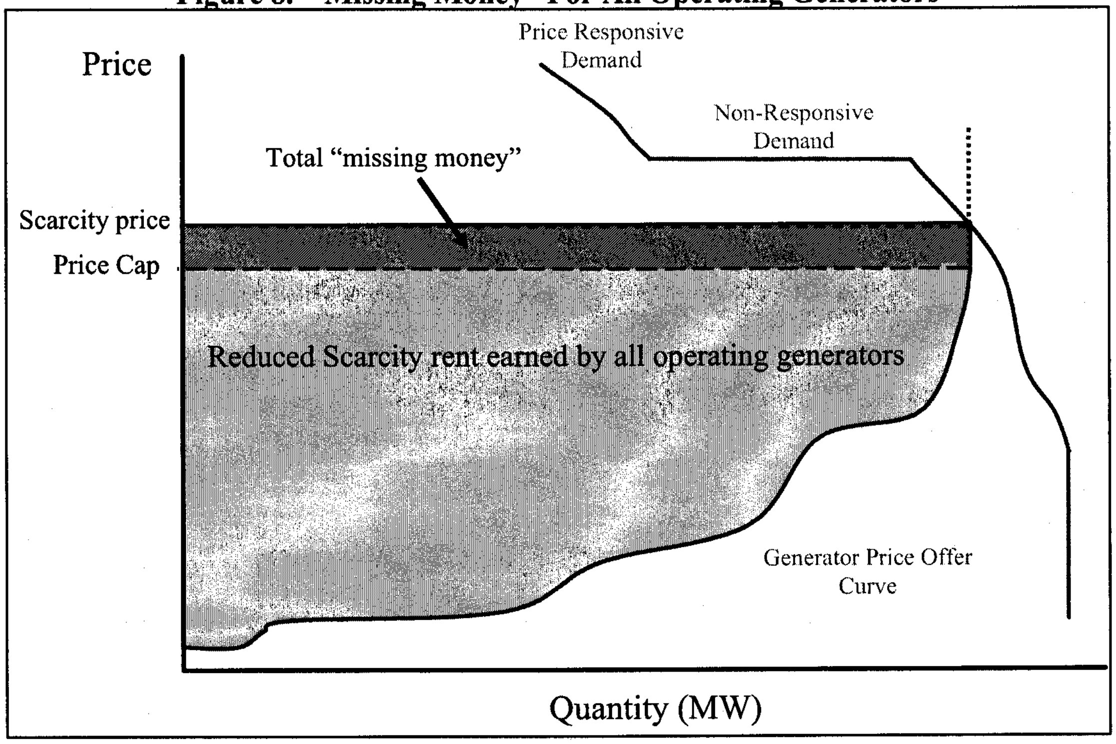
Figure 7. How Price Caps Create the "Missing Money" Problem



[26]The New York and New England ISOs have a modified version of scarcity pricing, which raises the spot prices above the generators' nominal marginal operating costs whenever the ISO experiences shortages in operating reserves; the greater the shortage, the higher the price rise, up to a cap. The shortage-cost adder is determined from an administratively defined demand curve for operating reserves. Such a mechanism would likely be a component in an energy-only market construct, as discussed in the previous chapter.

While figure 7 only illustrates the missing money problem for the one peaker, the price cap reduces the scarcity rents of *all operating* generators as shown in Figure 8.

Figure 8. "Missing Money" For All Operating Generators



Most ISOs, including the Midwest ISO, employ focused mitigation rules to limit generation offer prices when conditions suggest the possibility that market power could be exercised, such as in local areas with significant transmission constraints. Also, generators needed for out-of-merit dispatch to resolve congestion typically have their price offers mitigated. Even when market power is not a concern, spot prices may be suppressed because of political opposition to high spot energy price spikes.

Occasional price spikes are normal and are not *by themselves* indicative of market failure; rather they are natural outcomes in electricity markets that reflect the effects of varying demand, generating plant operating constraints and transmission constraints. Indeed, price spikes are a necessary signal for generators to make extraordinary efforts to be available or to increase output to the maximum. On the other hand, sustained prices at very high levels are not normal; more likely they indicate presence of market power or some market design or regulatory flaw.[27]

[27] During the California energy crisis, several thousand megawatts of existing capacity refused to schedule or offer their power for dispatch because state regulatory decisions prevented the utilities from paying either suppliers under QF contracts or suppliers that might otherwise have sold spot energy through the dispatch. The resulting shortages were regulatory artifacts which contributed to rotating blackouts in early 2001.

Spot prices can also be artificially depressed by ISO dispatch and pricing rules that have little to do with market power or politics. For example, an ISO may commit too many high-cost units to run at minimum levels, or hold high cost units in contingency reserve, while not allowing those plants to affect the LMP spot prices used for settlements.

Another common example is the overuse of "reliability must run" (RMR) contracts, which are often used in locally constrained regions to maintain a plant "needed for local reliability" that might otherwise retire because of depressed spot prices. RMR plants tend to be high cost, but they typically are not allowed to set market prices, which keeps down the market prices in constrained regions that are already short of supply. These, and other administrative mechanisms, suppress spot prices below the levels needed to provide the very incentives to solve the problem.

An energy-only market would require the system operator to carefully review its rules to ensure they do not artificially depress spot prices below competitive levels. If spot prices fail to give the right operational incentives, the system operator will be forced to impose additional administrative rules and sanctions to ensure operational reliability.

Capacity Payment Mechanisms

It is now conventional wisdom that the main purpose of capacity payment mechanisms is to replace the "missing money" that scarcity rents would have provided if spot prices (or contracts based on expectations regarding spot prices) were not artificially suppressed. When an ISO gives up on getting spot market prices right through better pricing of reserves, scarcity pricing, and other market design features, it is forced to solve the missing money problem with capacity payments. In addition, it must turn to non-market payments and administrative rules and sanctions to replace the operational incentives that efficient spot market prices would have provided. This inexorably leads to the need for further administrative rules and sanctions, along with demands for increasingly complex capacity markets administered by the ISO.

Capacity payment mechanisms can take different forms. The most common form is to impose an installed capacity (ICAP) requirement on LSEs (as found in the Northeast and proposed in California). Another approach is to require the LSEs to purchase a prescribed amount of capacity through ISO-coordinated capacity auctions (as in Eastern ISOs). These ISO-administered mechanisms may be short-run (such as monthly, as in New York and New England) or long-run (e.g., for delivery four-years out, as in the PJM RPM proposal). As the acquisition process looks further out, there must be periodic true-ups and interim mechanisms to account for forecast errors or other changes in conditions such as shifting LSE loads and capacity obligations under retail choice.

Once the ISO introduces a forward auction for generation resources, the natural instinct is to allow alternatives, demand-side resources and transmission upgrades, to compete on the same playing field. In fact, without appropriate scarcity pricing in the spot-market, explicit subsidies to demand-side and transmission alternatives would be logical. The tradeoffs between these options are quite complex, particularly the relationship between transmission upgrades and locational generation choices. Sorting it all out inevitably leads to the demand for some form of long-run planning that encompasses the entire topic of traditional integrated resource planning. As the only operational *regional* entity, the ISO must also become the regional planner. And this planner would be driven to doing more than simply forecasting and analysis. The ISO would perforce be assuming the role of decision maker mandating investments and directing subsidies.

The six existing ISOs in the US are at different places along this path. But the trend toward increasingly complex administrative mechanisms, focused around ISO-coordinated planning, auctions and procurement mechanisms is unmistakable. In light of all this, the question that should be asked is whether the initial decision to abandon the quest for better spot price signals was correct, especially now that we have a better understanding of where the alternative paths take us.

PROBLEMS CREATED BY CAPACITY MECHANISMS

To understand why capacity-based mechanisms are becoming increasingly complex and administrative, it is essential to recognize that the "missing money" caused by suppressing spot prices creates two very serious problems for a market-based electricity system. First, it reduces incentives for promoting operational reliability. Second, it reduces incentives for investing in generating plant thereby threatening resource adequacy. Virtually all of the administrative complexity is designed to "fix" these two problems.

The operational problem should be obvious but it is often overlooked. If spot prices provide the signals for plants to make themselves available for dispatch, then suppressing those prices when supplies are tight and demand is high will consistently fail to signal the need for the generators to do everything they can to make their units available for dispatch. The more serious the shortage, the greater the incentive problem becomes, undermining operational reliability.

During hours of shortage when spot prices *should be* very high are precisely the times when *all* plants *should be* available in order to justify a substantial portion of the revenues need to cover their fixed costs.[28] Generators must recover these costs just to remain operational, let alone make a profit. If spot prices could rise to reflect the actual shortage conditions, generators would have strong incentives to take whatever actions were most economic to improve and maintain the reliability of their units.

If the "missing money" is made up through monthly capacity payments, generators have even less incentive to make their plants available for dispatch in shortage

[28] Actual payments to generators could be a combination of spot market and contract payments.

hours and loads have less incentive to reduce consumption during these hours.[29]

Suppressing spot prices also has an adverse effect on investment incentives. The amount of investment that will occur in response to prices will only reach a level consistent with the level of revenues provided by the markets. If the suppressed spot markets reduce the total level of expected revenues then investment in new plant will be reduced also and resource adequacy will suffer.

These causal relationships should not come as a surprise; direct analogies can easily be found in the cost-of-service regulated environment. In the past investor-owned utilities resisted building new generation to meet their service obligations when their expectations were that their future revenue requirements, would not be covered through their rates. In the late 1970s electric rates were generally held down for political reasons, causing many utilities to defer the construction of new generation projects that were projected to be needed.[30] Looking back we now know that those plants were not needed because load growth slowed in response to due to significantly higher electricity prices, but that effect was not anticipated by the industry resource planners.

III. PRICE RESPONSIVE DEMAND

A key driver of resource adequacy in an energy-only market is the existence of significant price responsive demand. But how much is "significant?" Does the MISO system have enough price responsive demand potential to meet that requirement? Assuming it does, what needs to be done to convert this potential into an actual resource? These questions are now addressed.

HOW MUCH PRICE RESPONSIVE DEMAND IS REQUIRED?

The generally accepted US industry standard for resource adequacy is a loss of load expectation (LOLE) of one day in ten years. To achieve this, a typical US power system typically requires a generation reserve margin in the range of 12 to 18 percent of peak demand. This provides operating reserve (spinning and non-spinning) and also insurance against generating unit forced outages and peak load forecasting errors.

---

[29] This incentive problem is worse if the capacity credit assigned to a unit is defined by its average availability based on forced outages over time (EFORd), as is true in Eastern ISOs. In the "LICAP" proceeding, the ISO-New England proposed to discard the familiar "Unforced Capacity" (UCAP) availability metric and replace it with an availability metric that measures whether a unit was actually producing energy or capable of providing operating reserves in real time during those hours in which the ISO experiences shortages in operating reserves. This approach in effect redefines the "capacity" product to be more like energy and operating reserves, in an attempt to make the "capacity" payment mechanisms work more like an "energy-only" market. One implication of this proposal is that a capacity payment mechanism can best solve the operational incentive problem by attempting to mimic an energy-only market with unsuppressed spot prices.

[30] This disincentive effect is known as the "Aversh-Johnson effect" and has been studied ad nauseam since introduced by two RAND Corporation economists. See Aversh H. and Johnson L.L., "Behavior of the Firm under Regulatory Constraint, " American Economic Review, Vol. 52, December 1962, pp 41-54.

Price responsive demand is defined as load the customer interrupts in response to price increases. This is load not under the direct control of the system operator, thus not providing operating reserve.[31] Since operating reserve represents about 4 to 8 percent of peak demand and will be supplied by other resources, the role of price responsive demand in directly contributing to resource adequacy is necessarily limited to reducing peak demand by about 10 percent. However, it also indirectly contributes to resource adequacy by creating the scarcity rents that make new generation profitable, thereby attracting investment in new plant.

The exact amount of price responsive demand needed to facilitate sufficient investment in new generation is determined by the following factors:

- amount of flexible load (in MW)
- price elasticity of the flexible load
- level of energy price cap
- construction and fixed operating costs of a new peaking plant
- rates of return investors require to invest in generating plant.

We intend to quantitatively investigate the effect of these factors through computer modeling of the Midwest ISO system to determine what combination will ensure that an energy-only market will produce an acceptable level of resource adequacy. The results of these efforts will be released in the near future.

PRICE RESPONSIVE DEMAND POTENTIAL IN THE REGION

The Midwest ISO staff currently estimates the price responsive demand potential within its footprint to be in the range of 5,000 MW to 10,000 MW. This estimate is based in part on information provided by market participants to the Midwest ISO to fulfill its Module E requirements.

In addition, the above estimate of price responsive demand potential is based on the magnitudes of the large C&I loads served currently by MISO participants, combined with the price elasticities estimated in recent studies done of the real-time pricing

[31] However, price responsive demand can simultaneously participate in emergency programs controlled by the ISO, as is currently done in New York and PJM. If the customer can guarantee interruption within 10 minutes of the system operator's command it can also provide non-spinning operating reserve, as is done in New Zealand. In addition, California and international system operators have experimented with allowing customer load to provide spinning reserve as well, as long as it meets reserve specifications (*e.g.*, water pumping loads).

programs of Niagara Mohawk Power Corporation and Georgia Power Company.[32] In addition, survey data produced by EIA's most recently available Manufacturing Energy Consumption Survey (MECS), are consistent with these estimates of price responsive demand potential. Future efforts will provide our detailed analysis of price responsive demand potential within the MISO footprint.

FACILITATING PRICE RESPONSIVE DEMAND DEVELOPMENT

The incentive for retail customers to modify their electricity usage in response to price requires the Midwest ISO and state regulators to take the following actions:

- adopt scarcity pricing
- expose all generators and flexible loads to the spot market
- gradually increase the energy price cap to average VOLL for residential customers.

Scarcity Pricing

Scarcity pricing is introduced in Chapter I and discussed further in Chapter II. It reduces the frequency of involuntary curtailments in two ways. First, it encourages price responsive demand development by allowing the market price to rise above the cap on generator offer prices when generating capacity is in short supply.[33] Second, it provides incentives for generators to operate during hours of capacity when their output is needed most and for developers to add new generating capacity.

Expose Generators And Flexible Loads To The Spot Market

For generators to have the correct incentive to be available when needed, and for developers to add capacity when supply shortfalls loom, all generators should be exposed to the market-clearing spot energy price.[34] Unfortunately, price caps designed to mitigate market power violate this principle and create the "missing money" problem described earlier. For this reason the cap on market energy price should be set as high as

[32] Braithwaite, S. and O'Sheasy, M., "RTP Customer Demand Response – Empirical Evidence on How Much Can You Expect," in *Electricity Pricing in Transition (Topics in Regulatory Economics and Policy),* edited by A. Faruqui and K. Eakin, Kluwer Academic Publishers, 2002.

[33] In a LMP market some nodal prices can exceed the highest accepted generator offer even without scarcity pricing. This will occur if an increase in load forces a low-cost generator to be backed down in order to relieve a transmission constraint. The lost energy must then replaced by a higher-cost generator elsewhere on the grid. Thus, the marginal cost of serving this load increment includes not only the marginal cost of the incremental energy delivered to the load but also the increase in system cost caused by redispatch needed to relieve the constraint.

[34] One can argue that generators with bilateral contracts, or those owned by LSEs and dedicated to serving native load, need not be exposed to spot market prices. In a strict sense this is true; however, these same generators are still affected by spot market prices, albeit on a different time scale, because their alternative use before entering into a contract, or after being transferred to an unregulated subsidiary, is to sell into the spot market

practicable – ideally equal to the average VOLL for the loads that will be curtailed during hours of capacity insufficiency.[35]

Flexible loads are those that have the capability to respond to price signals on a short-term (e.g. day ahead or less) basis. These are the only loads for which exposure to real-time prices makes sense.

In theory, all loads are flexible at high enough prices, thus should also be exposed to the full spot energy prices in order for the demand curve to capture their individual valuations of electricity in any given hour. If that were so then (by definition) there would be no involuntary curtailments and (by definition) resource adequacy would be assured. However, for the foreseeable future at least, it is not cost-beneficial to meter all customers in real time (or for them to respond to those price signals even if so metered). Thus involuntary curtailments of customers lacking real-time meters will still occur on occasion and the assessment of resource adequacy will turn on the frequency and duration of such curtailments. Most of the price responsive demand response will come from a small subset of large industrial and institutional customers so the limit on meter installations is not particularly troublesome.

Exposing market participants to spot energy prices will increase their financial risk, however, most large customers will have opportunities to hedge their risk thorough forward contracting. However, residential and small C&I customers are not likely to have such opportunities available, at least not at modest cost, so it may be desirable for LSEs to serve these customers under fixed tariffs that are hedged through wholesale supply contracts or other financial instruments.

Gradually Increase The Energy Price Cap To Average VOLL

The primary reasons for capping energy prices are to mitigate generator market power and to reduce spot price volatility. As price responsive demand develops it will provide an alternate vehicle for achieving both of these objectives. This suggests a gradual relaxation of the energy price cap as price responsive demand takes over these mitigation functions. This progressive process takes advantage of the fact that as price responsive demand emerges at different price levels, suppressing market power and price volatility at each level, the price cap can be further relaxed to induce the next stage of development and price response, and so forth. In effect, the system pulls itself up by its own "bootstraps."[36]

[35] Under scarcity pricing the cap on market energy prices is conceptually different from the cap on generator offers.

[36] Numerous examples of "bootstrapping" processes are seen in the physical sciences. A thermonuclear explosion is one such example. An electric charge sets off a primer charge - which then triggers a conventional explosive - driving two masses of fissionable material together and triggering a nuclear fission explosion - which heats and compresses fusible material, triggering a thermonuclear fusion reaction.

The price cap should continue to be relaxed until it equals the average VOLL for the lowest-valued loads that are not on real-time pricing.[37] The reason for not raising the cap any further is that these customers will prefer to have their service interrupted rather than pay more for electricity than the value they place on it. In effect the spot market is acting on their behalf when it refuses to purchase energy for their use at prices above the cap.

To pursue this strategy it will be necessary to periodically assess the degree of market power remaining as price responsive demand develops so that the price cap can be relaxed as residual market power is observed to wane.

IV. MARKET POWER MITIGATION

Market power is usually an artifact of market structure rather than market design and some degree of market power will always exist in any market regardless of its design. Market power must be mitigated if reasonably competitive outcomes are to result - and allowed to produce "just and reasonable" rates.

Market power can arise if transmission constraints create pockets of captive load by limiting imports from external sources needed to restrain prices to competitive levels. In any transmission-constrained situation the potential for exercising market power is increased regardless of the market design, but in an energy-only market contrived shortages can result in spot energy prices reaching very high (shortage cost) levels. This potential for high prices provides temptation for generators to drive up prices by withholding supply.

The freedom of spot prices to reach shortage cost levels is an essential element of an energy-only market design because those prices provide the correct incentives for short-run responses from generators and flexible loads and for long-run investment in new generation and other facilities. For this reason, it is important to distinguish between real shortages and those that are contrived. The goal of market power mitigation is to preclude suppliers from profiting from contrived shortages while still allowing prices reach shortage-cost levels when there are genuine shortages.

The energy-only market design mitigates market power in several different ways. The four principal approaches are

- demand response
- limits on offer prices and must-offer rules
- reducing barriers to new entry
- contracting for power.

[37] Most likely these will be the residential customers based on the results obtained in numerous value-of-service studies conducted since the 1970s.

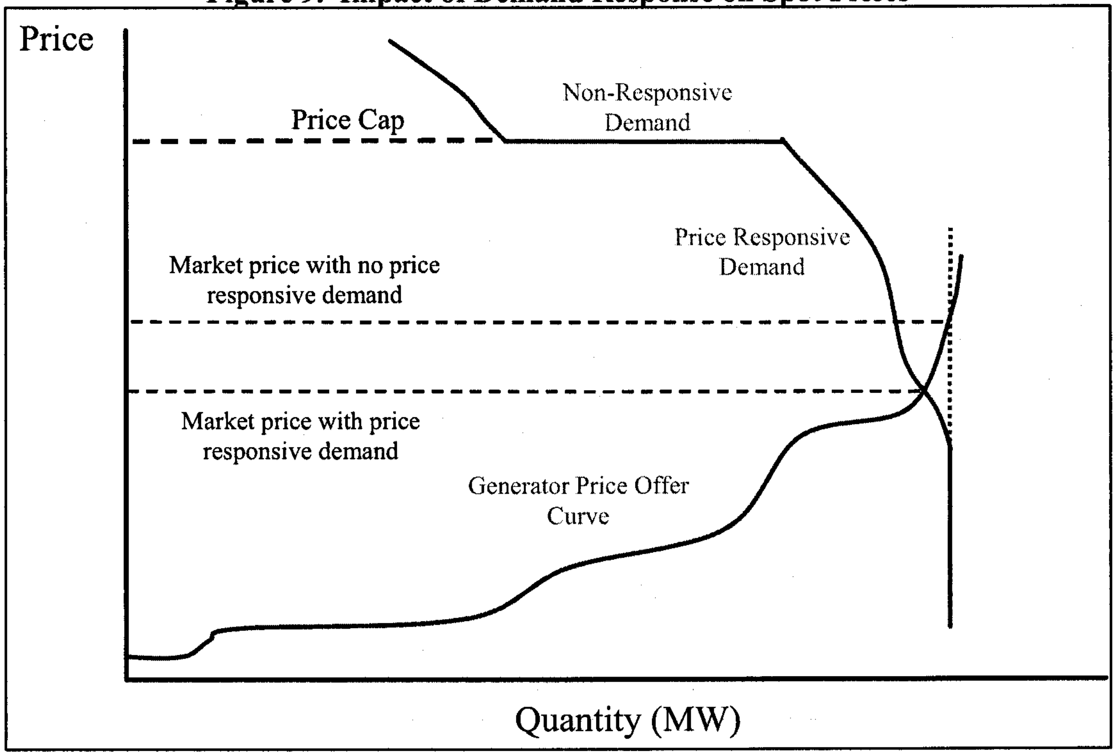
DEMAND RESPONSE

The energy-only market design allows prices to reach levels determined by customers' willingness to pay even when there is insufficient generation to meet all loads plus operating reserve requirements at lower price levels. Hence, price responsive demand is an important mechanism for limiting prices while also forestalling involuntary curtailments. As discussed in previous chapters, the price level at which involuntary curtailments occur should be the average VOLL for non-responsive customers.

This aspect of the energy-only market design stands in stark contrast to that of alternative designs that do not permit scarcity pricing. In these alternative designs prices are limited by an assumed maximum marginal cost of generation. A common misconception is that in LMP markets the spot prices at any location on the transmission grid should be determined by the marginal generator located there, and in zonal markets, by the marginal generator within the pricing zone. In contrast, the energy-only market design allows spot prices to be determined by the willingness of loads to pay prices higher than any generator offer price. In a shortage condition all generators would be operating at full output but none would be setting the spot energy price.

A relatively small amount of demand-side response can go a long way in mitigating prices when supplies are tight. As shown in Figure 9, price responsive customers will tend to reduce their loads when prices move up. The Figure shows that a small amount of demand response can produce a relatively large reduction in prices.

Figure 9. Impact of Demand Response on Spot Prices



This simple analysis clearly reveals that customers and state regulators concerned about the exercise of market power should be very interested in facilitating the development of price responsive demand.

LIMITS ON OFFER PRICES AND MUST-OFFER RULES

Supply can be withdrawn from the market either through physical withholding, whereby the generator declares some of its capacity as physically unavailable, or through economic withholding, whereby the generator submits price offers that substantially exceed the marginal cost of the capacity. In any market the temptation to withhold capacity arises from the potential that the resulting higher net revenues paid to capacity not withheld more than offset for the net revenues lost by the withheld capacity.

Mitigating Physical Withholding

In markets other than an energy-only market, the inflated spot prices resulting from physical withholding are set by the offer prices of marginal generators. In the energy-only market the spot prices could go much higher because scarcity pricing allows market prices to be set by demand-side responses up to the energy price cap set at average VOLL.

To mitigate physical withholding the market design would need to include "must-offer" rules for all units that could significantly affect market prices through physical withholding. Such rules would be reinforced with contractual incentives that discourage market power in the spot market, as discussed below.

Mitigating Economic Withholding

The market rules employed in some markets mitigate economic withholding by imposing limits on offer prices of those generating units that are needed to maintain reliable operations. Such mitigation rules would *continue* in an energy-only market, with the focus on ensuring the rules were only applied at the right times and to the right units.

Existing units needed for reliability in transmission constrained areas ("load pockets") would be the primary targets for offer price mitigation. An offer price cap would be defined for each such unit to reflect its marginal operating cost. The cap would apply whenever the plant's output was needed for reliability, such as when it is dispatched out of merit in order to relieve congestion and its plant's original offer would set the market price. These mitigated offer prices would be eligible to set the market price at each unit's respective location if the unit were on the margin.

Generating units operating at their capacity limits would not be allowed to set spot energy prices; those prices would be set by the demand side. The offer caps would serve to prevent economic withholding but prices would still reflect actual shortages in the constrained region as they should to send the correct signals. Note, however, that these units would still be paid the market price for their energy.

Because mitigated units would still receive the market energy prices, they would earn scarcity rents during periods of actual shortage, thereby avoiding the "missing money" problem. These scarcity rents would minimize the need for reliability must-run (RMR) contracts, which are used extensively in markets that are not energy-only, to keep generating units in service that might otherwise retire because of an inability to cover their fixed operating costs.

Once the costs of RMR contracts are accounted for, the total cost of the energy-only market design is not higher than those of alternate market designs. Indeed, the total cost of serving load could be lower in the energy-only market because its pricing structure provides better incentives for all generating plant to be available when most needed and for demand-side response to occur.

In any case it would be important that when units are not positioned to exercise market power they should not be subjected to these types of mitigation. Under current rules in existing markets, units in unconstrained "competitive" regions are generally not mitigated because the presence of competitors encourages availability and drives offer prices close to marginal costs. This same principle applies equally applies in an energy-only market.

REDUCING BARRIERS TO NEW ENTRY

In any market high prices attract entry of new supply. Thus, in the absence of barriers to entry, the exercise of market power sows the seeds of its own destruction. History has shown that anytime market power is exercised over extended periods there are barriers sustaining it. Reducing or eliminating such barriers is a critically important mitigation measure.

An LMP-based, energy-only market sends the right price signals to both the supply-side and the demand-side regarding where additional demand response and/or new capacity would be the most valuable. Transmission constrained areas in which LMPs are higher will attract new entry, assuming that any siting issues can be overcome.

New supply entry can come in the form of traditional utility generation, which is difficult to site in transmission constrained urban areas, or it can come in the form of smaller, strategically located distributed generation, which are generally be easier to site but which often face difficulties in getting the host utility to allow interconnection.

Distributed generation includes on-site generation owned by large customers that are exposed to real-time spot prices and it offers these customers a physical option for hedging against spot price volatility. From outside the facility these investments look like just another form of price responsive demand.[38]

[38] In Georgia Power Corp's real-time pricing (RTP) program, most of the demand response that occurs at relatively low energy prices is due to program participants running their back-up generators. See: Braithwaite, S. and O'Sheasy, M., "RTP Customer Demand Response – Empirical Evidence on How Much Can You Expect," at 32.

Streamlining the regulatory process for gaining interconnection approval, combined with net-metering rules to allow proper credit for energy produced on-site, would go far to stimulate distributed generation. In addition to mitigating market power within urban load pockets distributed generation would make the spot energy market more competitive and more efficient. A similar argument applies to small municipal systems embedded within larger utility systems.

CONTRACTING FOR POWER

The market power mitigation measures discussed above would reduce or preclude the exercise of market power by directly addressing physical or economic withholding. In contrast, contracting is a fundamentally different mitigation measure that removes the ability of a generator to profit from the withholding activity.

If a generator has a substantial portion of its capacity committed to deliver power at a fixed contract price, the revenues from the power sale are unaffected by spot prices. Furthermore, if it physically withholds any of the contracted capacity, or it suffers an outage, the generator must purchase replacement power from the spot market at the higher prices in order to meet its contract obligations. Thus, a substantially contracted generator has little or no interest in withholding supply to drive up spot market prices.

From the buyer's perspective, having its loads covered by power contracted at fixed prices also hedges the buyer (and its ultimate customers if it is an LSE) from spot price uncertainty. If spot prices occasionally reach shortage-cost levels, LSEs and their customers have strong incentives to minimize their spot market exposure. Consequently, the incentives to contract are much greater in an energy-only market than they are in other market designs that administratively cap spot prices. These enhanced incentives have positive market mitigation consequences.

Are Mandatory Contracts Needed?

The potential for spot prices to reach shortage cost levels would provide strong incentives for parties to contract voluntarily. Still, there might be reasons why state regulators would consider some form of mandatory contracting to cover customers lacking the ability to effectively hedge on their own.

Retail Choice Not Allowed

In states that do not allow retail choice regulators could provide incentives for their regulated LSEs to hedge themselves against exposure to the uncertainties of spot prices. For example, the pass-through of fuel and purchased power costs to customers could be eliminated. Or, as is currently done in California, the LSEs might be required to conduct explicit *ex ante* assessments of customer risk that are subject to prudence reviews as part of their resource planning and procurement processes. The LSEs could hedge themselves by owning their own resources or by purchasing most of their residual power requirements under fixed-price contracts. The degree and nature of utility hedging would be subject to state regulatory oversight.

Retail Choice Allowed

Regulators in states that allow retail choice would need to distinguish between two types of customers:

- those likely to choose alternative retail suppliers
- those either ineligible or unlikely to choose.

For the first customer group, it would make sense to consider whether some portion of these presumably larger customers should be subject to real-time pricing and exposure to spot prices. This exposure would provide the right incentives both for demand-side responses and for contracting to hedge exposure to spot prices. Voluntary contracting would appear sufficient for this group.

For the second customer group, the case for mandatory contracting is more compelling. Assuming an LSE has the default supply obligation, it would be expected to prudently hedge a substantial portion of its potential exposure to spot prices. It would do this through owned generation (if any) supplemented with contracts to cover any residual exposure. There are successful models for utility hedging of default supply obligations, such as the Basic Generation Service auctions sponsored by the New Jersey Board of Public Utilities, in which default supply obligations are covered by contracts acquired through competitive auctions closely monitored by independent auctioneers and state regulatory staff.

In an energy-only market it is likely that most loads would be hedged, either through contracts or through self-supplied generation, leaving a small residual amount fully exposed to the spot market. Nonetheless, this residual, together with the flexibility of loads and generators to buy from, or sell to the spot market some of their hedged energy, would provide the correct incentives for efficient behavior. Furthermore, contract prices would reflect the contracting parties' expectations of future spot prices over the contracted periods. The volatile spot energy prices would do what they are supposed to do: *provide the right incentives for short-run dispatch operations and long-run investment, while still providing most customers with relatively stable and certain electricity prices.*

V. ROLE OF STATE REGULATORS

The success of the Midwest ISO's energy-only market in ensuring resource adequacy depends not only on gaining FERC approval of its market design but also on the approval and guidance of the state regulatory commissions having jurisdiction over the load-serving entities (LSEs) within the MISO footprint. In particular, the decisions state regulators make regarding retail rate design and the supply adequacy of LSEs under its jurisdiction are critical.

State regulators are important in part because they control retail electricity rates. For an energy-only market to be successful it needs significant price responsive demand (PRD) participation. But PRD works best when at least a subset of retail customers are exposed to the hourly spot prices.[39] The state regulatory commissions control the method of billing through their approval authority over retail rate designs.

To support efficient PRD development, state regulators need to require LSEs to directly pass through their wholesale energy prices on an hourly basis to customers with real-time meters. In addition, the regulators can:

- establish minimum reserve margins to be maintained by the LSEs
- allow LSEs to count their customer PRD capacity as part of their reserve margin obligation
- require LSEs to exempt PRD capacity from charges for generating capacity.

Pass-Through Of Wholesale Energy Prices

Retail energy prices for real-time metered customers should equal the LSE's wholesale energy cost, grossed up for distribution losses. In addition, to maximize allocative efficiency the pass-through prices should reflect to the greatest extent practicable, the hourly LMPs at the retail customers' off-take nodes, rather than some load-weighted average price. State regulators would need to consider this for at least the large C&I customers that take service directly off the transmission grid.

Establish Minimum Reserve Margins

The purpose of establishing minimum reserve margins is to guarantee an acceptable level of resource adequacy for all non-PRD customers. These reserve margins would need to be separately set for each transmission constrained sub-region within the MISO footprint and they would likely differ significantly among the sub-regions. When the energy market reaches maturity these reserve margins would seldom, if ever, be binding and the interim resource adequacy vehicle could benignly reside in the background as a "safety net." Or it could simply be dismantled.

Count PRD In The Reserve Margin

PRD should be allowed to compete with generation on the proverbial "level playing field." This means that it be allowed to substitute for generation when calculating each LSE's reserve margin - at least up to the point where the remaining reserve margin, consisting of generating plant, is just sufficient to supply the entire operating reserve requirement, including provision for forced outages. The Midwest ISO's current Module E requirement reflect this 'level playing field', so this would not be

[39] Price responsive demand can work where the end users receive a lump sum fixed credit to curtail demand and the LSE, as the market participant in the Midwest ISO energy market, receives the hourly LMP reflecting the full credit for the load reduction. However, studies have suggested that persistence rates for end users in these programs diminish if the end users do not receive the full hourly LMP credit.

a change from existing practice. Considering the critical role of operating reserve it seems wise to initially err on the side of conservatism.

A megawatt of PRD need not be simply treated the same as a megawatt of generation. Although the availability of either resource is uncertain, those uncertainties are hardly equal. Further study would be needed to establish the appropriate capacity-equivalent of a MW of price responsive demand compared to a MW of peaking plant.

EXEMPT PRD FROM CAPACITY CHARGES

The surpluses of generating capacity that currently exist within the MISO footprint means that PRD may initially contribute little economic value to the system. This is because the capacity surpluses will dampen both the average spot energy prices as well as their volatilities. Thus, PRD loads will have little incentive to respond because their savings from doing so will be too small. State regulators should consider exempting PRD loads from paying any capacity charges for generation, even during this early period, because they derive no benefit from that capacity.[40] More importantly, such exemption would provide an immediate incentive for customers to implement PRD programs.[41] Until new generating capacity is needed the cost of generation capacity surpluses would need to be borne by non-PRD loads and treated as an investment in a more economically efficient future power system that will benefit all customers.

In addition to the above retail rate-related actions, state regulators can encourage PRD development in other important ways. They can:

- mandate real-time metering for all customers whose demands exceed some threshold level
- take advantage of the newly enacted EPAct – mandated requirements to fully investigate the appropriateness of real-time pricing for each customer class
- offer incentives to LSEs that promote PRD through collaborative programs
- require LSEs to hedge the price risk for small customers that cannot easily do this by themselves.

[40] This is not to say that the PRD loads do not benefit from the investment in, or the fixed operating costs of generation - they do. However, capacity charges are not intended to recover those costs; they are designed to reflect the marginal value of peaking capacity. PRD loads do not benefit from this capacity because they get off the system when such capacity becomes scarce. At other times PRD loads may benefit from the energy produced by peaking plants but they pay for that benefit through the spot energy market.

[41] Indeed, until spot energy prices increase significantly this could be only real incentive driving PRD development. Thus, this measure is an important part of the bootstrap strategy described earlier in the paper.

Mandate Real-Time Metering

The State of California recently invested $35 million to install real-time meters on all loads 200 KW or larger. While the concept is sound, the threshold level adopted by the Golden State may not be appropriate for the MISO states. This is an important topic for investigation by each state regulatory commission, the objective being to determine what threshold levels are cost-beneficial based on each LSE's customer price elasticities and metering costs.

State regulators have the authority to designate retail customers as either as "price responsive" or "non-price responsive," and to order the LSEs to be meter and bill the price responsive classes on an hourly basis. Initial (or pilot) programs could begin with the largest customers for whom the cost of the metering and billing systems would likely be justified. These programs could then be expanded to determine what other customer groups might be cost-beneficial candidates for real-time metering.[42]

Take Advantage Of EPAct To Fully Investigate Real-Time Pricing

Section 1252 of the recently enacted EPAct directs each state regulatory commission to conduct an investigation of the appropriateness of offering customers various types of time-based rates.[43] This same section also directs the US Department of Energy to provide technical assistance to state regulators and multi-state organizations. State regulators should take full advantage of this federal resource in their investigations of real-time pricing and PRD development.

Provide LSEs With Incentives To Promote PRD

State regulators can provide their LSEs with strong incentives to promote PRD development. Such incentives can be performance-based measures such as higher allowed rates of return or regulatory vehicles through which the LSE to share the savings with the customer. LSEs have considerable data describing the consumption patterns and behavior of their large C&I customers. Typically the LSEs have marketing representatives that establish personal relationships with the managers of these facilities. These relationships can be very valuable in influencing large C&I customers to evaluate the economic attractiveness of PRD. In addition, the LSEs could efficiently provide these customers with the know-how regarding the methods and technologies that have been successfully demonstrated elsewhere.

---

[42] Recent surveys and studies of real-time pricing using approximations of hourly spot prices suggest that the customer size for cost-effectiveness may be much lower than previously thought. See the presentations by Chuck Goldman and Roger Levy at the October 28, 2005 Restructuring Roundtable available at http://www.raabassociates.org/main/roundtable/asp?sel=65.

[43] The state regulatory commission must complete these investigations and reach decisions by January 2007.

REQUIRE LSEs TO HEDGE SMALL CUSTOMERS' PRICE RISK

Regulators should consider requiring LSEs to contract forward for the power needed to serve small customers that lack real-time meters and cannot respond to spot energy prices. These same customers are unlikely to have the means to efficiently (if at all) hedge against electricity price risk. Although they are typically served under published tariff schedules that are fixed, sometimes for years, most are still subject to *expost* fuel and purchased power adjustments which transfer market risk directly to them with little delay. It is this risk that LSEs can hedge through forward contracting.

Although mandatory hedging on behalf of small customers is not needed for the success of an energy-only market, it does remove a potential barrier to efficient risk management while also increasing the liquidity of the forward contracts market. In addition, it reduces the potential for small customers to launch political opposition to the energy-only market design.

VI. COMPATIBILITY WITH PJM MARKET

This is an issue that will likely be resolved by the FERC. Nonetheless, all parties should be asking how the two regions would interact if the Midwest ISO adopts an EOM while PJM adopts a capacity mechanism.

Without further details about how RPM would work it is not clear that generators, given the choice of making their plants available to the Midwest ISO or PJM, would choose to give up the scarcity pricing offered by the Midwest market. Even if a generator sold its capacity in a forward PJM market, in return for monthly capacity payments, it might de-list its capacity and sell its energy into the Midwest market during actual shortages. The only way to discourage this would be for PJM to impose severe non-performance penalties on any capacity previously committed under RPM that did not offer all of its energy to PJM.

However, there is nothing about improved spot energy pricing, including the introduction of scarcity pricing, and efforts to improve demand-side response that would harm the Midwest regardless of what the FERC may ultimately decide. Moreover, it seems clear that an EOM approach in the Midwest would improve short-run reliability in this region by providing generators with the right incentives to make their plants available for dispatch when most needed.

VII. THE TRANSITION PATH

For the Midwest ISO to adopt an energy-only market a first step will be to develop a detailed roadmap specifying the tasks it must complete and the time schedule for doing so. Included in this roadmap is a menu of complementary state regulatory actions that would greatly contribute to the success of the effort. This menu includes improved retail rate designs, real-time metering and billing, and possibly mandatory contracting for energy by LSEs. Finally, the roadmap would need to address four basic issues:

- Scarcity pricing
- Operating reserve
- Market power mitigation
- Resource adequacy mechanism

SCARCITY PRICING

A high priority for the Midwest ISO is to design and implement scarcity pricing with a cap on spot energy prices at the average value of lost load (VOLL) to non-price responsive customer loads. This would require reviewing available studies followed by consultations between the ISO, state regulators and market participants. Existing market rules would need to be reviewed and possibly revised to fully integrate demand-side bidding into the real-time and day-ahead markets.

As soon as these initial elements were in place, the Midwest ISO region would have the basic elements of scarcity pricing. This would provide the basic tools for the energy-only market to bring about resource adequacy. The "missing money" problem would be substantially solved, thereby encouraging investment in new supply resources. Regulatory certainty, such as a clear commitment by state regulators and market participants to the development of the energy-only market would greatly facilitate new investment. In addition, the prospect of scarcity prices would provide an incentive for buyers and sellers to enter into contracts, which in turn would provide collateral needed to support debt financing of generating plant, further spurring investment.

OPERATING RESERVE

The ISO also needs to develop and implement software to co-optimize the economic dispatch of energy and operating reserve and to price operating reserve consistent with the pricing for energy.

Another key task is to develop a demand curve for operating reserve that is consistent with spot energy prices, as discussed in Chapter I. This demand curve plays a particularly important role in introducing some price elasticity into the spot energy market before PRD develops sufficiently to take over that function.

Eventually, it is desirable for the Midwest ISO to have bid-based regional markets for operating reserves to allow owners of resources to decide whether or not to provide operating reserve and at what price. As previously discussed, the possibility of procuring reserves through ISO-administered markets is being addressed in a parallel track through the Ancillary Services Task Force (ASTF). However, a first priority is to specify the demand curve for operating reserves that would produce energy prices consistent with any shortages in operating reserves.

Market Power Mitigation

Additional efforts focus on reviewing the adequacy of offer price caps and other rules designed to prohibit market power. The Market Monitor's review will attempt to ensure that the rules are able to identify and discourage both economic and physical withholding, and to apply mitigation selectively while not preventing prices from reflecting genuine scarcity conditions when they occurred. An assessment of the effectiveness of different amounts of price responsive demand in mitigating market power will also be conducted.

Need For A Resource Adequacy Mechanism?

A question that arises is whether the Midwest ISO should develop a capacity-based mechanism to assure resource adequacy - at least until the energy-only market is fully functional. The rationale for a dual approach to ensuring resource adequacy generally tracks the reasons often given for maintaining capacity constructs in general. Here, we discuss some of these reasons and suggest questions that parties in the region may wish to ask in deciding this issue.

Is An Interim Resource Adequacy Mechanism Needed?

The energy-only market could take some time to implement and it could take even longer for price responsive demand to reach significant levels. In the interim some mechanism would be needed to ensure that commitments to install new generation would be made in time to add the capacity when needed. Whether such a formal mechanism will be needed, and what form it should take, is best developed through extensive discussions among the stakeholders.

In any event, the resource adequacy mechanism would not need to be a capacity construct of the form adopted in the Eastern markets. A coherent proposal would need to be developed and submitted to FERC, where it would undoubtedly still face considerable opposition. Assuming some proposal could be approved, the ISO would then need to develop the software and monitoring systems to support the capacity-based mechanism. This could take longer than putting in place an energy-only market and developing the significant amount of price responsive demand needed to facilitate resource adequacy.

Is it credible to believe that such a process would take less time in the Midwest, which has no history of capacity mechanisms, than it has taken in the East? The New England LICAP process is now in its third year, is still incredibly contentious, and is still not resolved. The PJM RPM process has taken more than two years to develop and will likely take at least another year or so to litigate. Since the resource adequacy mechanism will need to be in place within the next year or two, any capacity construct appears to be infeasible. Furthermore, given the innumerable issues that would need to be resolved to develop a capacity mechanism, it seems fair to ask whether the effort would consume so much of the Midwest ISO's limited resources that the development of the energy-only market would be delayed.

Is A Capacity Requirement *Politically* Necessary To Avoid High Spot Prices?

One argument for imposing regional, or sub-regional, capacity requirements would be to reduce energy price volatility because state regulators and other public officials will not tolerate the price spikes that typically occur in an energy-only market. In response to that we need to ask whether this commonly held view is credible, given the conditions in the Midwest.

As pointed out earlier in this report, in an energy-only market most customers would not be fully exposed to *price spikes* because they would be partially or completely hedged. For example, most utilities in the Midwest still own generation and/or contract for power to serve their native load under traditional cost of service regulation. Although the utilities themselves could face volatile wholesale spot energy prices, *their customers would not see that price volatility*. This is because spot market prices would instead be averaged into the fixed retail rates customers pay, just as all supply costs are today.

For utilities that have divested their generation and/or serve under retail choice, most small and medium-size customers are still covered under state regulated rules governing default service. Those rules generally fix default retail rates over extended periods – from several months to several years – which shield customers from volatile spot prices. Furthermore, their LSEs are exposed to these spot prices only to the extent that they have chosen not to be hedged. In an energy-only market only those customers eligible for retail choice could be exposed to spot prices and most would have the ability to hedge their exposure on their own.

If so few customers would actually be exposed to the volatile spot energy market how much force does the political vulnerability argument have?

Does A Regional Capacity Requirement Make Sense For The Midwest?

There are historic reasons why the Eastern ISO markets have capacity-based requirements, and that history partly explains why those ISOs have retained capacity markets despite their obvious shortcomings. Before deregulation the Eastern markets were tight power pools that facilitated the efficient trading of economy energy among the pool members. In that environment a mechanism was needed to ensure that every pool member met its own resource adequacy requirements, because the pool's pricing rules did not penalize members that took advantage in real time of the resources built and paid for by others.[44] When bid-based markets were introduced in PJM, New York and New England, the capacity requirements were simply carried over. But after they adopted low energy price caps to mitigate market power, creating the missing money problem, the capacity requirement (or some capacity payment mechanism) became a necessity.

[44] For example, members of PJM would buy energy from each other at prices that equally "split the savings" produced by the transaction. However, if the pool ran short of energy the deficit would be physically shared among all pool members while the spot price of energy would remain limited to the pool "running rate," set by the operating unit with the highest marginal cost. PJM proudly referred to this sharing arrangement as its "three musketeers' agreement."

The history of resource adequacy in the Midwest is different from that of the East, and it differs within the Midwest. In the absence of regional power pools the need for capacity requirements to avoid free rider issues did not arise. Each utility maintained its own planning reserve requirements and in some areas no central entity allocated capacity obligations nor imposed deficiency charges for utilities that were short. The capacity requirements vary by RRO; in MAPP, for example, capacity requirements are rigorous with significant deficiency charges.

Given this very different history it is worth asking whether the usual reasons for maintaining capacity mechanisms in the East apply in the Midwest.

Is An Energy-Only Market More Complicated Than A Capacity Market?

In concept, the mechanisms of an energy-only market are straightforward and focus on getting the spot market prices to reflect the ISO's real-time economic dispatch solutions. As long as generators have a choice the price incentives they receive must support reliable operations. Consequently, getting the prices right is worth doing regardless of whether or not a mechanism also exists for compensating generators for their available capacity.

Similarly, an energy-only market works best if there is a significant amount of price-responsive demand. But there is almost universal agreement that such demand response improves any market regardless whether or not it has a capacity mechanism. However, it is also clear that price responsive demand would develop more quickly and be more effective if customers were exposed to spot energy prices that were not artificially limited by low price caps. In the energy-only market framework proposed here states would be encouraged, but not required, to expand the number of customers served under real-time tariffs to stimulate such demand-side response. Such efforts would still be worthwhile, though much less effective, if a capacity requirement were used to solve the missing money problem.

Thus, most elements of an EOM design would be worthwhile implementing whether or not the region also adopted a capacity mechanism.

If a capacity mechanism is used in lieu of appropriate spot prices, the ISO must solve three difficult problems:

- how much "missing money" must be given back to generators to achieve the region's resource adequacy goals
- how should the "missing money" be given back such that it provides generators incentives to make their plants available for real-time dispatch when most needed.
- how to induce the right amount of demand-side response - at the right times and at the right locations.

How Much To Give Back

Though generally sound in theory, the idea of giving "missing money" back to generators is a red flag for many load representatives, and the debates over this topic have been highly contentious. As long as this issue remains unresolved, it undermines all other efforts to provide regulatory certainty, and adversely impacts new plant investment.

What Availability Incentives To Provide

New England's LICAP proposal made an attempt at this, but it has been rejected at the FERC. The default is to use the UCAP mechanism proposed by PJM. But UCAP only measures availability on a time-averaged basis which does not provide the correct incentives for generator availability.

The existence of the two "missing money" problems indicate that spot energy prices are being artificially held down, which means that price responsive demand will not be strongly encouraged. This suggests the need for subsidies to correct the market pricing defect, thereby creating another potential uneconomic incentive as well as grounds for endless disagreement and opposition.

In the New York, New England and PJM reforms efforts, the "missing money" problems have exploded into an extensive and expanding list of controversial issues:[45]

- What is the reliability standard for the region? Is it a uniform reserve margin for the entire region? An LOLE standard? Different standards for different sub-regions?
- How is the standard interpreted? Is it a minimum target? An average target? A target that should not be violated more than some percentage of the time? And if the latter, how does the region decide select the data to determine the variability over time?
- How will the mechanism assure fixed cost recovery for the desired level of adequacy? What is the benchmark capacity resource, and how can we agree on its costs? Are the costs different for different sub-regions or zones?
- If the point is to restore the missing money, how does the ISO account for fixed cost recovery each generator receives from the energy and operating reserve markets? Is this contribution determined ex ante or ex post? Is it averaged, and if so, over what period? Given transmission constraints, does it vary by location? How many locations?

[45] For a lengthy description of the many design issues at only the conceptual level, see Chandley, John, *ICAP Reform Efforts in PJM and New England*, September 23, 2005, a paper prepared for the California ISO (www.caiso.com).

- How will the capacity requirement apply in different areas separated by transmission constraints? How many LICAP or RPM zones are needed? How are they defined? How can they be changed? How often will they be changed? What happens to long-run contracts when they are changed?
- How will the ISO (or FERC) define and allocate transmission rights for limited imports into constrained areas? Are the rights physical or financial? How can they be traded? What happens to these rights when the zonal boundaries change?
- What is the shape of the demand curve? Sloped or vertical? Who decides this?
- If vertical, how is the strong incentive to exercise market power dealt with?
- If the curve is sloped, what are its parameters? What is the maximum amount of capacity that warrants a payment? What is the maximum payment? What is the target for the break-even point? Who decides this? Is FERC the ultimate decision maker in deciding the curve's parameters, and thus deciding how much capacity is represented by the demand curve?
- Is capacity procured through an auction? How often? Monthly? Yearly? Four-years forward? If forward, how often must the results be updated for errors in forecasts, project delays, changed circumstances?
- Is the price determined in each auction through supplier bids? If so, how are they mitigated for market power? How is withholding prevented?
- Is the capacity price determined from the demand curve? If so, which capacity is counted? All of it? Mothballed units? De-listed units?
- How are exports handled? Can exports result in withholding? How is this checked and mitigated?
- How are imports handled?
- How is availability measured? UCAP? EFORd? Reserve shortage hours?
- If availability is measured by availability during reserve shortage hours, are there exceptions for non-availability that are not in the generator's control? Which ones? How does the ISO or its market monitor tell?

These are not hypothetical issues; they are some of the actual issues that have been or will be litigated before the FERC. Based on these observations one can only conclude that implementing a capacity construct would not simplify the market design or market operations; indeed, it would substantially complicate it.

Can Mandatory Contracting Substitute For A Capacity Mechanism?

Contracts play an important role in an EOM. Their principal purpose is to provide price certainty to buyers and sellers and hence a way to manage the risks of volatile spot prices. However, they also represent a commitment on the seller's part to have capacity available to produce the contracted energy.

Contracts for capacity can be used to circumvent the missing money problem, thereby ensuring resource adequacy. However customers will not voluntarily give generators the missing money through contracts; *they must be required to do so during the interim phase.* If contracting is voluntary, customers will be biased towards reliance on spot markets that cap prices at low levels and the region will fall short of meeting its adequacy goals.

If contracts are to be the vehicles for ensuring resource adequacy they must be mandatory. All customers (or their agents) must be required to contract for enough firm energy to cover their maximum loads or for enough capacity to cover their maximum expected loads plus a target reserve margin designed to meet the region's resource adequacy criterion. Then the ISO must ensure that the contract obligations are fairly allocated to all utilities and LSEs based on their forecasted loads, the "firmness" of their contracts (e.g., the expected forced outage rates of the plants supplying the contracted energy) obligations and load. In addition, the ISO must have some mechanism for enforcing the allocated contract obligations, such as penalties for non-performance.

Stakeholders need to consider whether mandatory contracting is really much better than other capacity mechanisms. Many of the same issues involved in designing and implementing a capacity mechanism may surface in designing and allocating the mandatory contracting obligations. If retail choice is allowed, with its potential for creating stranded costs, that will introduce further complexities.

VIII. NEXT STEPS

The Midwest ISO has initiated a work program designed to answer the primary questions surrounding an energy only market. In particular, the question of what type of market design is more likely to increase the elasticity of the real time demand curve is an important aspect of the work program. As part of this analysis, the question of what specific rules foster increased demand side participation will be evaluated. Another aspect of the work program will look at the empirical evidence related to investment in generation under different market structures.

The initial terms of reference for this work program, as well as the intermediate and final results will be presented and discussed with Stakeholders through the Working Group process. It is anticipated that the results of the analysis will not only provide guidance to the ongoing discussion on the topic but may also point out specific design elements that are necessary for a successful market.

Finally, as has been discussed in this paper, the interrelationship between the energy market and other ancillary services, in particular operating reserves, means that efforts and work programs focused on long-term capacity needs must be tightly integrated with similar efforts focused on the treatment of operating reserves within the Midwest ISO market.

Appendix: Co-Optimized Markets for Energy and Operating Reserves

While the largest and most important electricity market is the energy market, the ISO also requires ancillary services. A vital component of any energy-only market is ensuring that these providers of ancillary services are compensated appropriately. Failure in this regard can present short-term operational problems for the ISO, because it may give generators incentives not to follow operator instructions. In addition, it can undermine the market's ability to provide sufficient financial incentives to induce development of the economically efficient amount of generating capacity without the need to introduce separate revenue streams. Since prices that actually reflect the marginal value of energy or ancillary services at each location and point in time will lead to development of the economically efficient amount of generating capacity, on average, suppressing ancillary services prices below those levels will have just the same effect as would suppressing energy prices below those levels: each will reduce incentives to invest, and will generally result in less than the economically efficient amount of capacity being built.

In co-optimized markets for energy and operating reserves, the market operator simultaneously considers offers to provide energy and operating reserves submitted on behalf of each resource, and minimizes the total cost of meeting load given a set of operating reserve requirements. The price of energy then reflects the marginal cost of supplying additional energy to meet a small increase in load at each location, taking into account all of the effects of re-dispatching generation on the system to meet that increase in load at that location, including any effect of that re-dispatch on operating reserves costs. Therefore, if meeting an increment of load at that location results in an increase in operating reserves costs, that increase will be reflected in the energy price at that location. Similarly, the price of each category of operating reserves will reflect the marginal cost of supplying additional operating reserves to meet a small increase in that operating reserves requirement, taking into account all of the effects of re-dispatching generation on the system to meet that increased requirement, including any effect of that re-dispatch on energy costs. Therefore, if providing more of that category of operating reserves results in an increase in energy costs, that increase will be reflected in the price of that category of operating reserves.

As the following examples will demonstrate, settling energy and operating reserve markets using prices determined in this matter ensures that generators that are scheduled to produce a given amount of energy and provide a given amount of operating reserves do *not* have a financial incentive to disregard ISO instructions and produce more energy while providing less operating reserve, or provide more operating reserve while producing less energy. This means that the markets provide incentives for generators to follow ISO instructions, significantly reducing the need for the ISO to employ sanctions, penalties and the like to ensure that generators follow ISO instructions and provide the requested amounts of energy and operating reserve.

EXAMPLES

The following assumptions will apply to both of the examples below:

- There are three generators: A, B and C.
    - Each has 1000 MW of generating capacity.
    - Each has a minimum generation level of zero.
    - Each has a start-up offer of zero.
    - Each has submitted a single offer to produce energy using any or all of its generating capacity:
        - A has offered to produce energy for $60/MWh.
        - B has offered to produce energy for $65/MWh.
        - C has offered to produce energy for $75/MWh.
- There is only one operating reserve requirement, which states that at least 100 MW of 10-minute reserve must be maintained.
- There are no losses or transmission congestion, or requirements for other ancillary services.

Additionally, both of these examples will illustrate the calculation of prices in the real-time market. In the real-time market, there are no costs associated with making oneself available to provide operating reserves in the real-time market (as all related short-run decisions, such as whether to staff a given facility or to purchase fuel in the anticipation of being dispatched) will have been made by the time of the real-time market. Availability offers for operating reserve should therefore be zero and have been ignored in these examples.[46]

Example 1: Marginal Provider of 10-Minute Reserves is Capacity-Constrained

In addition to the assumptions above, the following additional assumptions apply to this example:

- There are 2500 MW of load.

[46] In the day-ahead market, these costs are not zero, so resources should be permitted to make availability offers, which then should be reflected in day-ahead prices. While an illustration of procedures for calculating these prices is outside the scope of this appendix, the procedures are similar to those illustrated here and produce prices that have the same properties as the prices calculated herein—namely, they ensure that each entity scheduled to provide energy and operating reserves would not be financially better off if it had been scheduled to provide more energy and less operating reserve or vice versa.

- Each generator can ramp at 4 MW/minute—meaning that each generator can provide a maximum of 4 x 10 = 40 MW of 10-minute reserve.
- The 10-minute reserves requirement is fixed, and does not depend on price.

The least-cost real-time dispatch given those assumptions is shown in Table A-1 below.

Table A-1: Least-Cost Dispatch to Meet 2500 MW of Load While Providing 100 MW of 10-Minute Reserve

Generator	Capacity (MW)	Energy Offer ($/MWh)	Energy Schedule (MW)	10-Min. Res. Schedule (MW)	Dispatch Cost ($/hr)
A	1000	60	980	20	58,800
B	1000	65	960	40	62,400
C	1000	75	560	40	42,000
Total			**2500**	**100**	**163,200**

A resource is the marginal provider of a category of operating reserves if a small increase in the amount of that category of operating reserves that must be maintained would have caused the ISO to schedule that resource to provide more of that category of operating reserves. In this case, A is the marginal provider of 10-minute reserve, as B and C are already providing as much 10-minute reserve as they can, given their ramp rates. Moreover, A is capacity-constrained, meaning that all of its capacity is used to provide energy or operating reserves, as all 1000 MW of its capacity are scheduled either to produce energy or to provide 10-minute reserve. Therefore, it faces a trade-off between providing 10-minute reserve and producing energy, as it can use each MW of its capacity to provide energy or operating reserves, but not both. If it is scheduled to provide additional 10-minute reserve, the 10-minute reserve price must recognize this trade-off, and ensure that A does not have an incentive to ignore ISO instructions.

If the 10-minute reserve requirement were to increase by a small amount, to 101 MW, the dispatch would change as shown in Table A-1A below.

Table A-1A: Least-Cost Dispatch to Meet 2500 MW of Load While Providing 101 MW of 10-Minute Reserve

Generator	Capacity (MW)	Energy Offer ($/MWh)	Energy Schedule (MW)	10-Min. Res. Schedule (MW)	Dispatch Cost ($/hr)
A	1000	60	979	21	58,740
B	1000	65	960	40	62,400
C	1000	75	561	40	42,075
Total			**2500**	**101**	**163,215**

The additional 10-minute reserve must come from A, but that reduces the amount of energy that A can provide, since A is capacity-constrained. This energy would have to be made up by C, which is the least expensive unit with capacity available. Therefore, the increase in the total cost of the dispatch resulting from an incremental change in the

10-minute reserve requirement consists of the increase in energy costs resulting from shifting one MWh of output from A to C. This is \$75 – \$60 = \$15, so the price of 10-minute reserves is \$15/MW.

The energy price in this example is simply C's \$75/MWh bid, since C is the marginal provider of energy (and is not capacity-constrained). We can therefore see that A would not be made better off by producing more energy, since it realizes a margin of \$75 – \$60 = \$15/hour on each MW used to produce energy, and it is paid \$15/MW for each MW used to provide 10-minute reserve. We can also see that B and C actually *prefer* providing 10-minute reserve to generating energy, since their margins on generating energy are less than \$15/MWh. Consequently, each would like to provide the maximum amount of operating reserve it is capable of providing, and each has been instructed by the ISO to do so, consistent with its wishes.

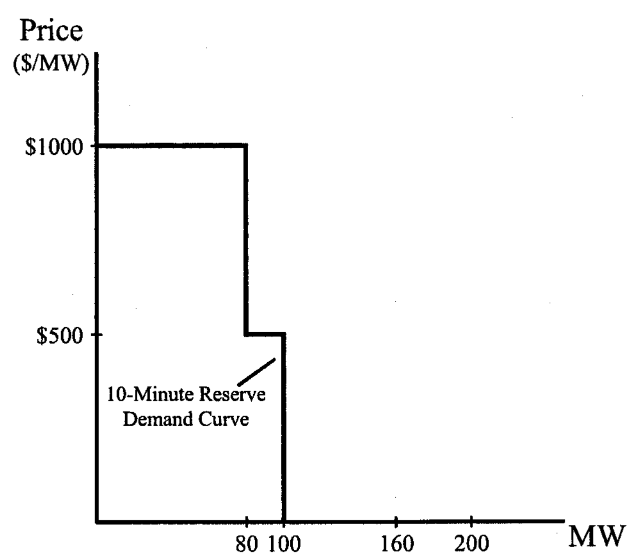
Example 2: Marginal Provider of Energy is Capacity-Constrained

In place of the assumptions above, the following additional assumptions apply to this example:

- There are 2910 MW of load.
- Each generator can ramp at 6 MW/minute—meaning that each generator can provide a maximum of 6 x 10 = 60 MW of 10-minute reserve.

Additionally, assume that the 10-minute reserves requirement is set using a demand curve, which is set at \$1000/MW for the first 80 MW of reserve meeting the requirement and \$500/MWh for the last 20 MW of reserve meeting the requirement, as shown in Figure A-1. In other words, the 10-minute reserve should normally be 100 MW, but it may be reduced to as little as 80 MW if the price of 10-minute reserve reaches \$500/MW, and it may be reduced to zero if the price of 10-minute reserve reaches \$1000/MW.

Figure A-1: Demand Curve for 10-Minute Reserves



One can consider the reserve demand curve (RDC) to be a "provider" of 10-minute reserve. Of course, it does not actually provide 10-minute reserve, but reducing the requirement using the RDC has the same effect on prices and schedules as an actual generator that is offering to provide up to 20 MW of 10-minute reserve at $500/MW and up to another 80 MW of 10-minute reserve at $1000/MW. For example, if the ISO only procures 90 MW of 10-minute reserve because 10-minute reserve prices have risen to $500/MW, then the actual generators would be scheduled to provide 90 MW of 10-minute reserve, just as they would have been scheduled to provide 90 MW of 10-minute reserve if an actual generator offering to provide 10 MW of reserve at $500/MW had suddenly appeared.

The least-cost real-time dispatch, given those assumptions, is shown in Table A-2 below.

Table A-2: Least-Cost Dispatch to Meet 2910 MW of Load with 10-Minute Reserve Requirement Determined Using Demand Curve

Generator	Capacity (MW)	Energy Offer ($/MWh)	10-Min. Reserve Cost ($/MWh)	Energy Schedule (MW)	10-Min. Res. Schedule (MW)	Dispatch Cost ($/hr)
A	1000	60	0	1000	0	60,000
B	1000	65	0	970	30	63,050
C	1000	75	0	940	60	70,500
RDC	20		500		10	5,000
RDC	80		1000		0	-
Total				**2910**	**100**	**198,550**

B is the marginal provider of energy in this example, since A, the unit with the lowest energy offer, is already producing as much energy as it can. B is also capacity-constrained, since all 1000 MW of its capacity are scheduled either to produce energy or to provide 10-minute reserve. Therefore, if the 10-minute reserve price is high (as it will be if the RDC is being used to reduce the 10-minute reserve requirement), and B is scheduled to provide additional energy in response to a small increase in load, the energy price must be calculated in a manner than ensures that B has an incentive to produce energy instead of providing 10-minute reserve, if B is to provide energy voluntarily.

If load were to increase by a small amount, to 2911 MW, the dispatch would change as shown in Table A-2A below.

Table A-2A: Least-Cost Dispatch to Meet 2911 MW of Load with 10-Minute Reserve Requirement Determined Using Demand Curve

Generator	Capacity (MW)	Energy Offer ($/MWh)	10-Min. Reserve Cost ($/MWh)	Energy Schedule (MW)	10-Min. Res. Schedule (MW)	Dispatch Cost ($/hr)
A	1000	60	0	1000	0	60,000
B	1000	65	0	971	29	63,115
C	1000	75	0	940	60	70,500
RDC	20		500		11	5,500
RDC	80		1000		0	-
Total				**2911**	**100**	**199,115**

In order to meet an incremental MWh of load at the lowest cost, the system operator would dispatch B to generate an additional MWh of energy while reducing the 10-minute reserve by an additional 1 MW. The change in the total cost of the dispatch that results from the need to meet an increment of load consists of A's energy offer plus the RDC price, since the RDC is "providing" an additional MW of 10-minute reserve. This is $65 + $500 = $565, so the price of energy is $565/MWh.

The 10-minute reserve price in this example is simply the $500/MWh price on the lower step of the RDC, since that portion of the RDC is the marginal provider of 10-minute reserve. B would not be made better off by producing more energy, since it realizes a margin of $565 – $65 = $500/hour on each MW used either to produce energy, which is equal to the $500/MWh it is paid to provide 10-minute reserve. A prefers generating energy to providing 10-minute reserve, since it realizes a $565 – $60 = $505/MWh margin when it generates energy, which exceeds the 10-minute reserve price, so it will be willing to follow the ISO's instructions to produce energy with all of its capacity, despite the high 10-minute reserve prices. C prefers providing operating reserve to generating energy, since the margin it realizes when it generates energy is only $565 – $75 = $490/MWh, so it will be willing to provide the maximum amount of operating reserve it is capable of providing, just as the ISO has asked it to do.